




Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular
Institut de Biotecnologia i Biomedicina

HUMAN PRION-LIKE PROTEINS AND THEIR RELEVANCE IN DISEASE

Doctoral thesis presented by Cristina Batlle Carreras for the degree of PhD in Biochemistry, Molecular Biology and Biomedicine from the Universitat Autònoma de Barcelona.

The work described herein has been performed in the Department of Biochemistry and Molecular Biology and in the Institute of Biotechnology and Biomedicine, supervised by Prof. Salvador Ventura i Zamora.

Cristina Batlle Carreras

Prof. Salvador Ventura i Zamora

Bellaterra, 2020

Protein Folding and Conformational Diseases Lab.

This work was financed with the fellowship “Formación de Profesorado Universitario” by “Ministerio de Ciencia, Innovación y Universidades”.

This work is licensed under a Creative Commons Attributions-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License. The extent of this license does not apply to the copyrighted publications and images reproduced with permission.

(CC BY-NC-SA 4.0) **Batlle, Cristina**: Human prion-like proteins and their relevance in disease. Doctoral Thesis, Universitat Autònoma de Barcelona (2020)



ENGLISH SUMMARY

Prion-like proteins have attracted significant attention in the last years. The interest in these polypeptides owes to: i) their implication in degenerative diseases; ii) the presence of disease-causing mutations that increase the risk to develop the disease; iii) their ability to establish protein-protein interactions; and iv) their participation in membraneless organelles formation.

Yeast prions are proteins with the ability to switch between a soluble and an amyloid state as an adaptive mechanism to confront environmental changes. Most of the yeast prions have a disordered prion forming domain (PFD) of low complexity, enriched in Gln and Asn, and depleted of hydrophobic and charged residues. In order to discover novel yeast prions, computational tools were developed and further used to uncover proteins with domains similar to yeast PFDs in other proteomes, including human. These proteins were named prion-like, and their identified regions prion-like domains (PrLDs). PrLDs are similar in composition to yeast PFDs, and they share the ability to aggregate, replicate, and propagate.

Cytoplasmic deposits of prion-like proteins have been observed in many degenerative diseases, and genetic mutations in their PrLD have been described. Most of the prion-like proteins are nuclear, but they translocate to the cytoplasm in disease conditions. This phenomenon has been associated with prion-like proteins aggregation into amyloid fibrils. Therefore, in this thesis we identified potential new prion-like proteins' candidates and experimentally validated the region responsible for aggregation of seven of them. We also improve predictions for disease-causing mutations impact in prion-like proteins aggregation process.

It was observed that prion-like proteins have the ability to form dynamic and reversible membraneless organelles, depending on the cellular conditions, for example, under stress. When these changes are too persistent, membraneless formation is not reversible anymore and they result in the appearance of the solid deposits recurrently observed in patients. We are still far from deciphering the exact mechanism of membraneless organelles formation. For this reason, in this thesis, we have studied the human prion-like protein hnRNPD_L to try to expand the knowledge of this phenomenon and its connection to disease. We characterized hnRNPD_L isoforms self-assembly properties demonstrating the importance of alternative splicing in controlling protein phase separation events, and show that D378N/H disease mutations accelerates hnRNPD_L aggregation dramatically reducing its solubility in the muscle of *Drosophila*.

Prion-like proteins are hubs of large protein-protein interaction networks through their PrLDs. Interestingly, a significant number of yeast PFDs have coiled-coil regions (CC) that participate in their aggregation process. We demonstrate here that this is also the case for human PrLDs and characterize the MED15 protein as an example, showing that its PrLD forms amyloid fibrils in a process mediated by its CC, which can account for its role in disease.

In general, this thesis is an extensive study of prion-like proteins and their properties.

RESUM EN CATALÀ

Les “prion-like proteins” han anat adquirint una atenció cada cop major en els últims anys. El seu interès recau en: i) la seva implicació en malalties degeneratives; ii) la quantitat d'elles que presenten mutacions genètiques que augmenten el risc a patir la malaltia; iii) la seva habilitat d'establir interaccions proteiques; i iv) la seva participació en la formació dels orgànuls sense membrana.

Els prions de llevat són proteïnes amb la capacitat de canviar d'un estat soluble a un amiloide com a sistema adaptatiu per fer front a canvis ambientals. La majoria dels prions de llevat tenen un “prion forming domain” (PFD) desordenat i de baixa complexitat, enriquit en Gln i Asn, i empobrit en residus hidrofòbics o amb càrrega. Per tal de descobrir nous prions de llevat, es van desenvolupar eines computacionals que després es van usar per buscar proteïnes amb dominis similars als PFDs de llevat en altres proteomes, incloent l'humà. Aquestes proteïnes es van anomenar “prion-like”, i les seves regions identificades “prion-like domains” (PrLDs). Els PrLDs tenen una composició similar als PFDs de llevat, i comparteixen la capacitat d'agregar, replicar-se i propagar-se.

En moltes malalties degeneratives s'han trobat dipòsits citoplasmàtics compostos de “prion-like proteins”, de les quals s'han descrit mutacions genètiques en el seu PrLD. La gran majoria de “prion-like proteins” són nuclears, però es transloquen al citoplasma quan provoquen la malaltia. Aquest fenomen s'ha associat a la seva agregació en forma amiloide. En aquesta tesis hem identificat nous candidats “prion-like” potencials, i hem validat experimentalment per set d'ells la regió responsable de la seva agregació. També hem millorat les prediccions de l'impacte de les mutacions genètiques en el procés d'agregació de les “prion-like proteins”.

S'ha vist que les “prion-like proteins” tenen la capacitat de formar orgànuls sense membrana dinàmics i reversibles en funció de les condicions cel·lulars, com per exemple en situacions d'estrès. Quan aquests canvis cel·lulars són massa persistents, aleshores els orgànuls sense membrana deixen de ser reversibles donant lloc als dipòsits sòlids trobats en pacients. Encara estem lluny de conèixer el mecanisme exacte de la formació dels orgànuls sense membrana. Per aquesta raó, en aquesta tesis hem estudiat la “prion-like protein” hnRNPD human a fons per intentar expandir el coneixement sobre aquests fenòmens i la seva connexió amb la malaltia. Hem caracteritzat les propietats d'autoassemblatge de les isoformes de hnRNPD demostrant la importància del splicing alternatiu en controlar els esdeveniments de separació de fase, i ensenyem que les mutacions D378N/H involucrades en malaltia acceleren dramàticament l'agregació de hnRNPD, reduint la seva solubilitat en el múscul de *Drosophila*.

Les “prion-like proteins” són nuclis de grans xarxes d'interaccions proteiques a través del seu PrLD. S'ha vist que un nombre significatiu de PFDs de llevat tenen regions “coiled-coil” que participen en el seu procés d'agregació. En aquesta tesis demostrarem que això també passa pels PrLDs humans i hem caracteritzat la proteïna MED15 com a exemple, mostrant que

el seu PrLD forma fibres amiloides en un procés mediat pel seu “coiled-coil”, fet que pot explicar el seu rol en malaltia.

En general, aquesta tesis és un estudi exhaustiu de les “prion-like proteins” i les seves propietats.

CONTENTS

1. LIST OF PUBLICATIONS:	6
2. ABBREVIATIONS	7
3. INTRODUCTION	8
4. OBJECTIVES	13
5. PUBLICATIONS	
5.1. PUBLICATION I	15
5.2. PUBLICATION II	33
5.3. PUBLICATION III	41
5.4. PUBLICATION IV	51
5.5. PUBLICATION V	69
5.6. PUBLICATION VI	77
5.7. PUBLICATION VII	97
6. MEDIATOR COMPLEX SUBUNIT 15 (MED15)	101
SUMMARY	102
INTRODUCTION	102
RESULTS	103
DISCUSSION	109
MATERIALS AND METHODS	112
REFERENCES	114
7. GENERAL DISCUSSION	117
8. CONCLUSIONS	122
9. REFERENCES	124
10. ANNEX. SUPPLEMENTARY MATERIAL.	
10.1. PUBLICATION IV	129
10.2. PUBLICATION VI	153
10.3. MED15	165

1. LIST OF PUBLICATIONS:

This thesis is composed of the following published works (in order of appearance):

- I. **Batlle C.**⁺, Iglesias V.⁺, Navarro S.⁺, Ventura S. **Prion-like proteins and their computational identification in proteomes.** Expert Rev Proteomics (2017) DOI: 10.1080/14789450.2017.1304214.
- II. Sabaté R., Rousseau F., Schymkowitz J., **Batlle C.** and Ventura S. **Amyloids or prions? That is the question.** Prion (2015) DOI: 10.1080/19336896.2015.1053685.
- III. Fernández MR., **Batlle C.**, Gil-García M., Ventura S. **Amyloid cores in prion domains: key regulators for prion conformational conversion.** Prion (2017) DOI: 10.1080/19336896.2017.1282020.
- IV. **Batlle C.**, de Groot NS., Iglesias V., Navarro S. and Ventura S. **Characterization of soft amyloid cores in human prion-like proteins.** Sci Rep (2017) DOI: 10.1038/s41598-017-09714-z.
- V. **Batlle C.**, Fernández MR., Iglesias V and Ventura S. **Perfecting prediction of mutational impact on the aggregation propensity of the ALS-associated hnRNPA2 prion-like protein.** FEBS Lett. (2017) DOI: 10.1002/1873-3468.12698.
- VI. **Batlle C.**, Yang P., Coughlin M., Messing J., Pesarrodoná M., Szulc E., Salvatella X., Kim HJ., Taylor JP. and Ventura S. **hnRNPD phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation.** Cell Rep. (2020). DOI: 10.1016/j.celrep.2019.12.080.
- VII. **Batlle C.**^{*} and Ventura S.^{*}. **Prion-like domain disease-causing mutations and misregulation of alternative splicing relevance in limb-girdle muscular dystrophy (LGMD) 1G.** Neural Regen Res. (2020). Under publication.

Other articles co-authored which are not included in this thesis (in chronological order):

- I. Sant'Anna R., Fernández MR, **Batlle C.**, Navarro S, de Groot NS, Serpell L, Ventura S. **Characterization of amyloid cores in prion domains.** Sci Rep (2016) DOI: 10.1038/srep34274.
- II. Medina-Carmona E, Betancor-Fernández I, Santos J, Mesa-Torres N, Grottelli S, **Batlle C**, Naganathan AN, Oppici E, Cellini B, Ventura S, Salido E, Pey AL. **Insight into the specificity and severity of pathogenic mechanisms associated with missense mutations through experimental and structural perturbation analyses.** Hum Mol Genet (2018) DOI: 10.1093/hmg/ddy323
- III. Iglesias V, Conchillo-Sole O, **Batlle C** and Ventura S. **AMYCO: evaluation of mutational impact on prion-like proteins and aggregation propensity.** BMC Bioinformatics (2019) DOI: 10.1186/s12859-019-2601-3.

2. ABBREVIATIONS

Alzheimer's disease	AD
Amyotrophic lateral sclerosis	ALS
Aggregation-prone regions	APR
Aggregation-prone sequences	APS
Coiled-coil	CC
Creutzfeldt-Jakob disease	CJD
Defined interval amino acid numerating algorithm	DIANA
Fluorescence recovery after photobleaching	FRAP
Frontotemporal lobar degeneration	FTLD
Aggregation hot-spots	HS
Inclusion-body myopathy	IBM
Intrinsically disordered region	IDR
Limb girdle muscular dystrophy 1G	LGMD1G
Liquid-liquid phase separation	LLPS
Lowest-probability subsequences	LPS
Mediator complex subunit 15	MED15
Membrane-less organelles	MLO
Multisystem proteinopathy	MSP
Prion aggregation prediction algorithm	PAPA
Parkinson's disease	PD
Prion-like amino acid composition	PLAAC
Prion domain	PrD
Prion-like domain	PrLD
Prion forming domain	PFD
Prion Protein	PrP
Protein-protein interaction	PPI
Post-translational modification	PTM
RNA recognition motif	RRM
Short linear motif	SLiM
Transmissible spongiform encephalopathy	TSE

3. INTRODUCTION

Protein aggregation

Proteins are large biomolecules that are made, in general, from a combination of 20 different amino acids (Creighton, 1993). They are involved in a wide range of cellular functions, and in order to carry out these processes, they require to be correctly folded in their native conformation (Dobson, 2003). Protein folding is a complex process, and errors during this process can occur thus driving proteins to misfolded states, and eventually to aggregation (Hartl et al., 2011).

Amyloids are a type of aggregate characterized by resistance to detergents (i.e., SDS) and a regular β -sheet fibrillar structure, with β -sheets arranged perpendicular to the longitudinal axis of the fibril, with the ability to bind amyloid-dyes, such as Thioflavin-T and Congo Red, inducing a spectral change (Chiti and Dobson, 2006). Amyloid formation is suggested to obey the short-stretch model, according to which aggregation is initiated by short regions of the amino acid sequence, usually 5-10 residues in length, with a predominant hydrophobic character and low net charge (De Groot et al., 2005; López De La Paz and Serrano, 2004). These short regions are named aggregation-prone sequences (APS), aggregation hot-spots (HS) or aggregation-prone regions (APR). Amyloid fibrils are usually formed by a nucleation-elongation reaction that comprises three sequential steps: the lag or nucleation phase, which is thermodynamically disfavored; the elongation or exponential phase, when mature fibrils are formed; and the stationary phase, when most monomers are consumed (Invernizzi et al., 2012).

In humans, protein misfolding and aggregation into amyloid fibrils have been linked with a broad range of human diseases, ranging from neurodegenerative conditions such as Alzheimer's disease (AD), Parkinson's disease (PD) and Creutzfeldt-Jakob disease (CJD), to non-neuronal disorders such as type II diabetes and cataracts (Chiti and Dobson, 2006; Invernizzi et al., 2012). In most of these disorders, there is the formation of insoluble protein deposits composed of the above-described amyloid fibrils, located either in the extracellular or intracellular space. Therefore, the understanding of what causes misfolding and aggregation of these proteins is critical for diagnostic and therapeutic purposes.

Yeast prions

Prions are proteins able to adopt multiple structural conformations, from which at least one has infectious and self-propagating properties, usually an amyloid state, with the ability to seed the endogenous soluble protein (Aguzzi and Calella, 2009; Kraus et al., 2013).

The term prion has been generally related to transmissible spongiform encephalopathy (TSE) diseases, caused by Prion Protein (PrP) misfolding and aggregation (Prusiner, 1982). PrP is the only proved protein to have self-propagating and self-templating properties across individuals. Nevertheless, additional proteins exhibit similar templating mechanisms, and prion classifications and terminologies according to their prionic properties have emerged. A more in-depth description of the different classes of prions and their properties are reviewed in the

attached manuscript entitled *Prion-like proteins and their computational identification in proteomes (Publication I)*. In this section, we will focus only on yeast prions, the best-characterized set of functional prions.

In *Saccharomyces cerevisiae*, prions have been reported to behave as epigenetic elements of inheritance necessary to stress adaptation and survival, indicating that not all proteins with prion properties are detrimental, but they can also play beneficial functions (Halfmann et al., 2010; Newby and Lindquist, 2013). Thereby, prions are an intriguing group of proteins with important biological functions in both health and disease.

Yeast prions are characterized by the presence of a prion domain (PrD) or prion forming domain (PFD) located in intrinsically disordered regions (IDRs) enriched in Gln and Asn residues and depleted in hydrophobic and charged residues (Alberti et al., 2009; Liebman and Chernoff, 2012). They encode the prion protein ability to switch between the soluble and aggregated states, regulating in that way protein function.

There exist two hypotheses accounting for yeast prions aggregation by means of its PFD: the compositional model and the short-stretch model. The first one suggests that aggregation is driven by a large number of weak interactions along the PFD (Toombs et al., 2012). The second one suggests that aggregation is driven by specific, short, amyloidogenic stretches of the PFD (Sabate et al., 2015). A deeper description of the two hypotheses can be found in the attached manuscript entitled *Amyloids or prions? That is the question (Publication II)*. Our group proposed the short-stretch model for yeast prion aggregation, and we further validated it with four of the best-characterized yeast prions, Sup35, Ure2, Swi1 and Mot3 (Sant'Anna et al., 2016). We demonstrated that the 21 residues long amyloid core of these proteins was enough for aggregation and propagation of the prion state. A description and discussion about these results are in the attached manuscript entitled *Amyloid cores in prion domains: key regulators for prion conformational conversion (Publication III)*.

Identification of prion-like proteins

The common properties shared by yeast PFDs stimulated the development of bioinformatics algorithms to uncover similar domains in other proteomes such as: DIANA (Michelitsch and Weissman, 2000), LPS (Harrison and Gerstein, 2003), PAPA (Toombs et al., 2012), PLAAC (Lancaster et al., 2014), pWaltz (Sabate et al., 2015) and prionW (Zambrano et al., 2015). These bioinformatics tools allowed the identification of many proteins with similar disordered Q/N-rich domains in other proteomes such as: *Dictyostelium discoideum* (Malinowska and Alberti, 2015; Malinowska et al., 2015), *Plasmodium falciparum* (Muralidharan et al., 2012; Singh et al., 2004), bacteria (Pallarès et al., 2016; Yuan and Hochschild, 2017), plants (Chakrabortee et al., 2016) and, of course, humans (King et al., 2012). These proteins are termed prion-like proteins, and their domains resembling yeast PFDs are referred to prion-like domains (PrLDs). Databases, such as Prionhome (Harbi et al., 2012) and Prionscan (Espinosa Angarica et al., 2014), storing the identified prion-like proteins, were also developed as information repositories.

A deeper description of the prion-like predictors and databases can be found in the attached manuscript entitled *Prion-like proteins and their computational identification in proteomes (Publication I)*. In this review, there is also information about identified prion-like proteins in the above-mentioned proteomes. In this thesis, we will only focus on human prion-like proteins.

Human prion-like proteins and their implication in diseases

Prion-like proteins in humans represent less than 1% of the proteome (King et al., 2012). Most of these proteins are nucleic acid-binding proteins, with a globular domain, such as the RNA recognition motif (RRM), and a disordered PrLD, being involved in transcription and translation regulation processes. Importantly, many emerged prion-like proteins, for example, FUS, TDP43, hnRNPA1, hnRNPA2, HTT, TAF15, EWRS1 and TIA1, have been related with the pathology and genetics of devastating neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD), inclusion-body myopathy (IBM), multisystem proteinopathy (MSP) or Huntington (Harrison and Shorter, 2017; Ito et al., 2017; King et al., 2012). All these proteins have been reported to aggregate into amyloid fibrils by its PrLD. Moreover, the associated diseases are characterized by the presence of pathologic protein aggregation in the cytoplasm of affected neurons, whereas in normal physiological conditions, most of these proteins are nuclear. Therefore, the understanding of the aggregation process of these proteins is crucial to find therapeutic approaches.

The analysis and identification of novel prion-like proteins could help to understand their biological function and implication in disease. We suspected that more prion-like proteins remained to be discovered. Therefore, we developed a strategy combining the compositional and short-stretch yeast prion models to identify novel human proteins with PrLDs (Batlle et al., 2017). We identified 535 polypeptides, encoded by 336 genes, as putative prion-like protein candidates. Moreover, we were interested in further validating our short-stretch model for human prion-like proteins, as we performed previously with yeast prions. From our computational analysis, we selected seven candidates involved in disease to evaluate their putative nucleating cores aggregation potential. We provided experimental evidence indicating that specific sequences inside PrLDs have the potential to trigger the conformational conversion of PrLDs in human proteins. The results and discussion of this study are in the attached manuscript entitled *Characterization of soft amyloid cores in human prion-like proteins (Publication IV)*.

Most of the identified prion-like proteins bear disease-causing mutations in their PrLDs, which are reported to increase protein aggregation and result in the formation of cytoplasmic inclusions in patients (Harrison and Shorter, 2017). Moreover, some of these mutations confer both gain- and loss-of-function phenotypes (Hutten and Dormann, 2016; Martinez et al., 2016; Winklhofer et al., 2008). Accurate predictions of the impact of this kind of mutation would help to uncover novel disease-associated proteins. Therefore, we developed a strategy that combining both yeast prion aggregation models, the compositional and the short-stretch models, permitted

us to predict with high accuracy the aggregation impact of mutations in prion-like proteins. The explanation of this strategy is in the attached manuscript entitled *Perfecting prediction of mutational impact on the aggregation propensity of the ALS-associated hnRNPA2 prion-like protein* (Publication V). From this study, we further developed the webserver AMYCO to evaluate the effect of mutations on the aggregation properties of any prion-like protein (Iglesias et al., 2019).

It is worth mentioning that despite extensive investigation, we still miss a clear description of the molecular forms of these proteins that causes the disease. For example, it has been shown that mutations in TDP43 that increase aggregation propensity reduce cellular toxicity in yeast, instead of increasing it as one would have expected (Bolognesi et al., 2019). We are still far away to completely understand prion-like proteins and their implication in disease, thus their study should be continued in order to uncover novel therapeutic strategies.

Membraneless organelles

Eukaryotic cells contain numerous compartments or organelles with specialized functions, which can be delimited by a membrane (i.e., mitochondria, lysosomes) or not (i.e., nucleolus, stress granules). Membraneless organelles (MLOs) are composed of proteins, nucleic acids, and other molecular components, and their functions are still under debate, but some suggestions are: concentration of biochemical reactions; signal amplification; sequestration of harmful components; and storage of biomolecules. An essential shared feature between prion-like proteins is their recruitment to MLOs, such as stress granules after cellular exposure to environmental stresses (Anderson and Kedersha, 2008; Elbaum-Garfinkle, 2019; Kim et al., 2013; Li et al., 2013; Mackenzie et al., 2017; Nedelsky and Taylor, 2019; Nott et al., 2015). This led to hypothesize that the self-assembly properties of human prion-like proteins might be not only responsible for disease but also play functional roles in cells.

The first insight of a connection between prion-like proteins and MLOs was provided by a serendipitous discovery using the chemical b-isox (Han et al., 2012; Kato et al., 2012). This chemical triggered the precipitation of a large number of proteins involved in RNA granule formation, specifically prion-like proteins, because b-isox binds to their PrLDs inducing hydrogel formation. These hydrogels were similar to amyloid fibrils, but they were more labile, being dynamic, reversible and detergent-soluble. This led the authors to hypothesize that the hydrogel state was the organization principle of membraneless subcellular structures.

Prion-like proteins incorporation to MLO is mediated through a liquid-liquid phase separation (LLPS) process in which PrLDs act as critical scaffolds (Boeynaems et al., 2018; Harrison and Shorter, 2017). LLPS results in the demixing of one liquid from another, resulting in two separate phases, one enriched in protein and one depleted of it, as occurs by mixing water with oil. It is characterized by a phase diagram between the one-phase regime and the two-phase regime delimited by a boundary, which is protein concentration dependent (Boeynaems et al., 2018; Wang et al., 2018). The boundary between the two regimes depends on the ability of a protein to phase separate. The first insight about the liquid nature of MLOs

was provided by Brangwinne *et al* in 2009 when they showed that P-bodies in *Caenorhabditis elegans* embryos have liquid-like properties and form by phase separation (Brangwynne *et al.*, 2009). Two years later, they did similar observations for nucleoli of *Xenopus leavis* (Brangwynne *et al.*, 2011).

MLOs are dynamic assemblies that undergo rapid internal rearrangement, as observed by fluorescence recovery after photobleaching (FRAP), with spherical shape and that can fuse with one another relaxing to a new sphere (Boeynaems *et al.*, 2018). They are maintained by transient weak contacts, such as π -cation, π - π , hydrophobic, electrostatic, dipole-dipole, or hydrogen bonding interactions. Their physical traits and composition depend on environmental factors. Many MLO can contain nucleic acids, which facilitate additional multivalent interactions for the formation of the complex. However, excessive nucleic acids concentration can also trap the protein in an unassembled state (Maharana *et al.*, 2018). Posttranslational modifications, such as methylation and phosphorylation, can also alter LLPS behavior and RNA granule dynamics (Hofweber and Dormann, 2018; Owen and Shewmaker, 2019). Salt, pH and ATP are also conditioning factors that regulate prion-like proteins phase separation (Franzmann *et al.*, 2018; Patel *et al.*, 2017).

Uncontrolled aggregation by mutations, prolonged stress or changes in protein concentration, cause an aberrant transition from a liquid to a more solid state exhibiting amyloid-like properties and suggested to be responsible for the disease (Franzmann and Alberti, 2018; Patel *et al.*, 2015). Nevertheless, the molecular mechanisms governing this transition and the exact pathogenic state are still not well understood. For example, inappropriate liquid-liquid demixing has also been suggested to cause cellular toxicity contributing to disease (Bolognesi *et al.*, 2016). Moreover, the Balbiani body in oocytes is solid-like, and it is dependent on stable amyloid-like interactions, but its formation is reversible (Boke *et al.*, 2016; Woodruff *et al.*, 2018). Consequently, a deeper study of these transitions and the interactions involved in these assemblies is required to understand their implications in human disorders.

In order to gain more insights about LLPS and disease-association mechanisms, we performed an exhaustive study on the human prion-like hnRNPD L self-assembly features. We analyzed the properties of the three existing isoforms and the effect of the disease-causing mutations associated to limb girdle muscular dystrophy 1G (LGMD1G) in both LLPS and aggregation. The results and discussion of this study are in the attached manuscript entitled *hnRNPD L phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation (Publication VI)* and *Prion-like domain disease-causing mutations and misregulation of alternative splicing relevance in limb-girdle muscular dystrophy (LGMD) 1G (Publication VII)*.

4. OBJECTIVES

The general objective of this thesis is the study of human prion-like proteins. We intend to identify novel human proteins with PrLDs and characterize their specific regions responsible for their aggregation and propagation. Moreover, we aim to establish a prediction method to elucidate whether a mutation increases aggregation propensity in these proteins and experimentally validate it with reported disease-causing mutations. Finally, we pursue to use PrLDs as a mechanism to sequester endogenous prion-like proteins and inactivate them specifically.

The detailed objectives of this thesis are:

- * Identify the specific region of a prion-like protein responsible for aggregation and propagation;
- * Identify novel human prion-like proteins by bioinformatics analysis;
- * Design short peptides corresponding to the nucleating core of human prion-like proteins;
- * Characterize the aggregation properties of new putative prionogenic cores *in vitro*;
- * Demonstrate that the short amyloid core of a PrLD is sufficient for promoting its aggregation;
- * Improve the predictions of mutational impact on prion-like proteins aggregation propensity;
- * Characterize the self-assembly properties of hnRNPD L isoforms *in vitro* and in cells;
- * Characterize the disease-causing mutations effect in hnRNPD L aggregation propensity;
- * Understand the relationship between polyQ regions, PrLDs, coiled coil propensity and amyloid formation;
- * Characterize the aggregation and propagation potential of MED15 PrLD *in vitro* and in cells.

5.1. PUBLICATION I

Prion-like proteins and their computational identification in proteomes

Battle C.⁺, Iglesias V.⁺, Navarro S.⁺, Ventura S.

Expert Rev Proteomics (2017)

DOI: 10.1080/14789450.2017.1304214.

5.2. PUBLICATION II

Amyloids or prions? That is the question.

Sabaté R., Rousseau F., Schymkowitz J., **Batlle C.** and Ventura S.

Prion (2015)

DOI: 10.1080/19336896.2015.1053685

Amyloids or prions? That is the question

Raimon Sabate,^{1,#} Frederic Rousseau,^{2,3,#} Joost Schymkowitz,^{2,3,#} Cristina Batlle,⁴
and Salvador Ventura^{4,*}

¹Departament de Físicoquímica; Facultat de Farmàcia; and Institut de Nanociència i Nanotecnologia (IN²UB); Universitat de Barcelona; Barcelona, Spain;

²VIB Switch Laboratory; VIB; Leuven, Belgium;

³Department for Cellular and Molecular Medicine; KU Leuven; Leuven, Belgium;

⁴Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular; Universitat Autònoma de Barcelona; Barcelona, Spain

ABSTRACT. Despite major efforts devoted to understanding the phenomenon of prion transmissibility, it is still poorly understood how this property is encoded in the amino acid sequence. In recent years, experimental data on yeast prion domains allow to start at least partially decrypting the sequence requirements of prion formation. These experiments illustrate the need for intrinsically disordered sequence regions enriched with a particularly high proportion of glutamine and asparagine. Bioinformatic analysis suggests that these regions strike a balance between sufficient amyloid nucleation propensity on the one hand and disorder on the other, which ensures availability of the amyloid prone regions but entropically prevents unwanted nucleation and facilitates brittleness required for propagation.

KEYWORDS. amyloids, neurodegenerative diseases, prions, protein intrinsic disorder, Q/N-rich domains, yeast

ABBREVIATIONS. AD, Alzheimer's disease; TSE, transmissible spongiform encephalopathy; PD, Parkinson's disease; CJD, Creutzfeldt-Jakob disease; fALS, familial amyotrophic lateral sclerosis; PFD, prion forming domain

In the cell, proteins attain the native structure through a delicate and balanced network of interactions, where protein folding and aggregation exert as competing pathways.^{1,2} In a protein energy landscape, amyloid-like aggregates represent an energy minimum, being usually thermodynamically more stable than the native conformation. This has led to the hypothesis

*Correspondence to: Salvador Ventura; Email: salvador.ventura@uab.es

Received April 22, 2015; Revised May 12, 2015; Accepted May 17, 2015.

[#]These authors contributed equally.

Extra View to: Sabate R, Rousseau F, Schymkowitz J, Ventura S. What makes a protein sequence a prion? PLoS Comput Biol 2015; 11:e1004013; PMID:25569335; <http://dx.doi.org/10.1371/journal.pcbi.1004013>

Color versions of one or more figures in this article can be found online at www.tandfonline.com/kprn.

that the amyloid structures reflects a universal mode of assembly of polypeptide chains and that native protein structures are evolutionary selected metastable states.² Amyloids are aggregates displaying fibrillar structure, which is constituted by repetitions of a specific protein in a regular β -sheet conformation that runs perpendicular to the fibril axis.³ In humans, amyloids are linked to diseases ranging from neurodegenerative conditions such as Alzheimer's disease (AD), Parkinson's disease (PD) and Creutzfeldt-Jakob disease (CJD), to non-neuronal systemic and localized disorders.³ On the other hand functional amyloids, i.e. proteins that exploit the amyloid fold for evolutionary selected biological functions, have been discovered in diverse species, including human.⁴ The roles fulfilled by these functional amyloids range from obligate amyloid structures required for scaffolding and/or movement to conditional amyloids such as the yeast prions that can be triggered by environmental factors.⁵ Whether obligate or conditional, the natural selection of amyloid structure as a functional motif indicates that these properties are likely sequence specific. Whereas the attainment and sustainment of the native structure relies on cooperative interactions involving most, if not all, of the sequence of a protein domain,^{6,7} the now widely accepted 'short stretch hypothesis' states that amyloid formation in contrast is nucleated by short regions of the amino acid sequence named aggregation hot-spots (HS), Aggregation Prone Regions (APR) or Aggregation Prone Sequences (APS).^{8,9} The short stretch model led to the development of over 20 algorithms that more or less successfully predict protein aggregation and amyloid formation based on the identification of specific β -aggregation and amyloid-prone regions in the polypeptide sequences.¹⁰⁻¹² In disease-associated amyloids these regions are generally between 5 and 10 residues in length.¹³

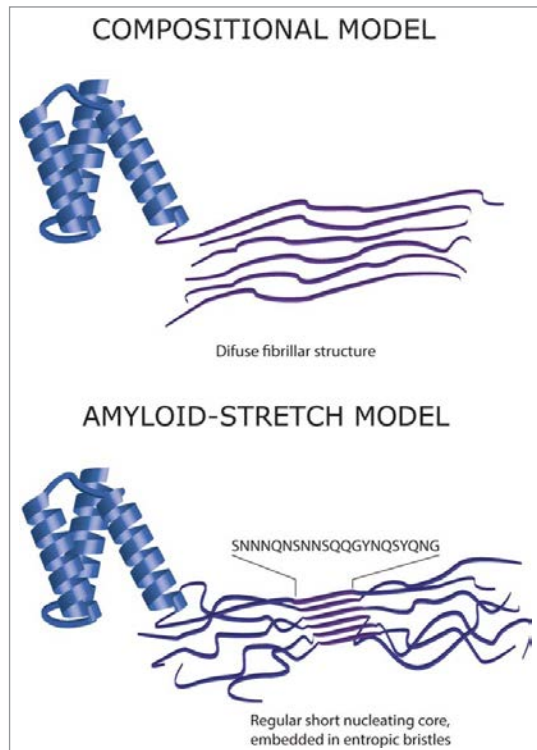
Prions are considered a subclass of amyloids in which protein aggregation becomes self-perpetuating and infectious. The phenomenon is known mostly as a neuronal pathology in mammals but in fungi prions play a crucial role in epigenetic inheritance.¹⁴⁻¹⁶ Importantly, despite the overlapping conformational properties of amyloids and prions, only a handful of amyloids

are currently considered to display at least partial prion capacity under natural conditions.¹⁶ As a result, β -aggregation and amyloid predictors are still a long way from correctly detecting prion sequences in proteomes.¹⁷ In fact, the sequence characteristics that make a protein sequence a prion have been elusive for years. Moreover, at first glance, the sequence features conferring prion capacity to prion protein in mammals (PrP) appear to differ remarkably from those determining prion behavior in fungi.

Yeast prions are the best characterized transmissible amyloids, thus being excellent model systems to address the determinants of concomitant amyloid formation and propagation.¹⁸ In these proteins, prion formation from an initially soluble state involves a structural amyloid conversion driven by specific, relatively large, unstructured domains enriched in glutamine/asparagine (Q/N) residues.¹⁸ Interestingly, protein domains displaying this sequence signature are over-represented in eukaryotic proteomes relative to prokaryotes, suggesting that prion-like conformational transition might have evolved as a mechanism for regulating gene function at the protein level in eukaryotes.¹⁹ It should be mentioned, however, that PrP, the archetypical mammalian prion, lacks these sequential features.

In order to explore the repertoire of prion proteins in *Saccharomyces cerevisiae* the Lindquist group conducted a genome-wide bioinformatics survey using a hidden Markov sequence model to identify putative candidates on the basis of their compositional similarity to known prion forming domains (PFDs)²⁰ and used experimental validation to identify the bonafide PFDs in their predictions. These results are at the core of several algorithms for prion domain prediction, all relying on the analysis of amino acid sequences.^{17,21-24} These programs are constructed on 2 alternative models for amyloid formation by prion-like domains (**Figure 1**): (1) The compositional model relying on the establishment of a large number of weak interactions¹⁷ and (2) our model, which proposes 'classical' nucleation by short amyloidogenic stretches, whose amyloid propensity is modulated by the structural context.²⁴ Despite the mechanistic difference between algorithms,

FIGURE 1. Two models for amyloid structure formation in Q/N-rich prion-like domains. The compositional model relies on the establishment a large number of weak interactions whereas the amyloid-stretch model proposes the existence of a preferential short nucleating sequence whose amyloid propensity is modulated by its structural context.



the advent of accurate computational tools to detect yeast prion domains opens new and exciting possibilities, allowing the exploration of proteomes for the discovery of novel and hitherto unexpected Q/N-enriched domains that may drive conformational conversion in novel prion proteins. Indeed, recent studies have revealed that over 250 human proteins display prion-like stretches in regions with high presence of uncharged polar residues and glycine, including several heterogeneous nuclear ribonucleoproteins (hnRNPs) related to neurodegenerative diseases such as familial amyotrophic lateral sclerosis (fALS).²⁵

In the light of these advances, the requirements for a polypeptide sequence to act as a prion begin to be defined. In our view, for a

protein sequence to become a Q/N enriched prion, 3 essential conditions appear to be required:

Requirement 1: A short amyloid-prone region able to trigger amyloid formation in a sequence specific manner. These amyloid cores should, however, possess distinctive features, since both a high aggregation rate and an elevated fragmentation capacity are necessary in prions in order to attain the number of propagons or seeds required for spreading and propagation.^{26,27} Thus, while a certain amyloid nucleation capacity favoring a sufficiently high aggregation rate is absolutely necessary, the final amyloid aggregate should at the same time display brittleness, a property that facilitates an increase in the number of nucleation events per cell. Accordingly, in contrast to most amyloids, the aggregation reaction should not be nucleated in PFDs by an extremely strong and highly hydrophobic amyloid core.

Requirement 2: The amyloid-prone region has to be located in a structurally disordered region, that permits its self-assembly without the necessity of conformational unfolding. The PFDs of all known Q/N enriched yeast prions display this property.^{18,20} The location of the amyloid core in large unstructured regions favors the acquisition of the β -cross motif without large conformational rearrangements and may at the same time promote the brittleness mentioned in requirement 1. Moreover, the disordered region may act as a so-called 'entropic bristle',^{27,28} which would reduce the overall aggregation propensity and could allow for a better biological control of the nucleation event, which is discussed more in detail in requirement 3.

Requirement 3: PFDs have to possess an amino acid composition allowing the protein to remain in a soluble state under physiological conditions while keeping intact a cryptic amyloid capacity. Stress situations promoting increased local

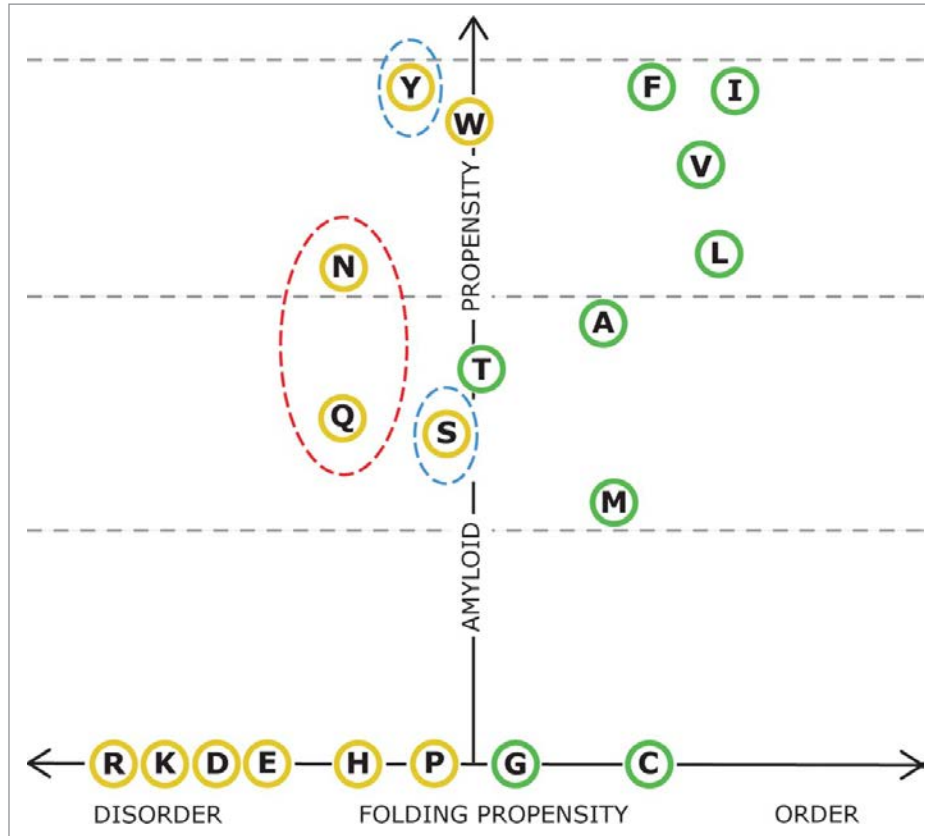
protein concentration, as well as the presence of preformed amyloid seeds, might alter the delicate equilibrium between native/soluble and amyloid/insoluble states, providing means to control the nucleation of amyloid aggregation and hence the onset of the prion phenotype. If we recapitulate requirements (1) and (2), clustering in the same sequence region amino acid residues with a significant amyloid propensity together with residues promoting structural disorder would favor prion capacity. Only five amino acid residues seem to unite these 2 essential properties, *i.e.*, amyloid propensity and structural disorder, according to FoldIndex²⁹ and Waltz³⁰ algorithms: N, Q, Y, S and W. (**Figure 2**). Interestingly enough N, Q, Y, S are, in this order, the most over-represented residues in bona-fide prion domains, relative to their frequency in the protein universe,²² with odds ratios of 5.70, 4.13, 1.72 and 1.66, respectively. In this context, N and Q residues show medium amyloid propensity, allowing the formation of amyloids with moderate strength, while at the same time are the amyloidogenic residues that more benefic disorder. This provides a rational basis for the strong over-representation of these particular residues in the PFDs of yeast. In good agreement with its higher frequency, N is the residue that best balances amyloid and disorder propensity and thus the preferred residue to support prion behavior. It is important to point out, however, that poly-N or poly-Q, these later being involved in a number of ataxias,³¹ are not expected to display a prion-like behavior since they lack requirement 1, that is, a specific region able to selectively nucleate ordered amyloid formation.

Hydrophobic amino acids are under-represented in Q/N-rich yeast PFDs, likely because a high proportion of these residues would render the protein excessively aggregation-prone and/or result in too strong amyloid assemblies. Despite the presence of a reduced number of

hydrophobic residues in PFDs has been shown to bust prion formation and amyloid formation,³² Y is the only hydrophobic residue over-represented in these domains. It has been proposed that this might respond to the fact that aromatic residues might facilitate both prion formation and chaperone dependent prion propagation.³³ However, F is indeed under-represented in PFDs with an odds ratio of 0.72 and the Y/F relationship between odds ratios in PFDs is 2.4, suggesting that the additional hydroxyl group in Y should provide a certain advantage, which in our opinion is allowing a better balance of amyloid propensity and intrinsic disorder. Despite its aromatic character, W is one of the most under-represented residues in prion domains with an odd ratio of 0.091, only C, which is able to crosslink covalently polypeptide chains, being less frequent.²² The absence of W in PFDs is best explained by its particular structure, wherein the indole group may not be easily placed in β -cross structures due to of steric impediments, being indeed depleted relative to F and Y in functional and pathogenic amyloids.³⁴

The two alternative models used to identify prion domains (**Figure 1**) coincide in the requirement of a relatively large disordered region in yeast PFDs. However, one prion model support the view of amyloid formation in PFDs resulting from a bias in sequence composition favoring a large number of weak interactions over a wide sequence stretch¹⁷ whereas the alternative model supports prion behavior to emerge from the preferential nucleation by specific and localized short amyloid-prone stretches embedded in the wider disordered region.²⁴ Despite the apparent contradiction between these 2 views, indeed the second model just pursues to delimitate the aggregation driving force of the amyloid cores embedded in the prions domains defined by the compositional model. In this way, the first composition based methods to predict potential yeast prions proposed a minimal core of 60 residues.²⁰ This further evolved into a method employing a 41 amino acid sliding window for compositional analysis, denoting that the initial 60 residues window was larger than actually required.¹⁷ We proposed to reduce this size

FIGURE 2. Balance between amyloid and structural propensities in natural amino acids. Residues rendering ordered and disordered 21-residues long homo-polymers according to FoldIndex¹⁴ are shown in green and yellow circles, respectively. The amyloid propensity of these stretches was calculated with Waltz.³⁰ The four more over-represented residues in yeast PFDs are circled by discontinuous lines, red indicates odd ratios > 4.0 and blue odd ratios > 1.5, relative to the composition of the protein universe.²²



even further to account for a 21 amino acid core, based on the length of the core of HET-s PFD, the unique protein for which an atomic-resolution structure of the infectious fibrillar state is available to date,^{24,35} which displays a β -arcade conformation.³⁶ The excellent performance of our method, based on a preferential nucleation by short amyloid-prone stretches, lead us to believe not only that a 21 residues core is indeed sufficient for prediction, but also that the ‘classical’ short stretch nucleation model applies to prions in a similar manner as it does for ‘classic’ amyloids, the main differences being that, in prions, amyloid nucleating stretches might fold into β -strand-turn- β -strand

elements and that their potency is strongly modulated by the entropy of the sequence context in which they are embedded, i.e. the degree of structural disorder will determine both the sensitivity for amyloid nucleation as well as the ability of formed fibrils to break up and provide additional propagons.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

ACKNOWLEDGMENTS

We thank Pol Ventura for his comments on **Figure 2**.

FUNDING

This work was supported in part by Ministerio de Economía y Competitividad, Spain [BFU2013-44763 to S.V.]; by SUDOE, INTERREG IV B, FEDER [SOE4/P1/E831 to S.V.]; by ICREA [ICREA Academia 2009 to S.V.]; by the Ramón y Cajal Programme from Ministerio de Ciencia e Innovación [RYC-2011-07987 to R.S.]. The Switch Laboratory was supported by grants from VIB, University of Leuven, the Funds for Scientific Research Flanders (FWO), the Flanders Institute for Science and Technology (IWT) and the Federal Office for Scientific Affairs of Belgium (Belspo, IAP network P7/16).

REFERENCES

- Jahn TR, Radford SE. The Yin and Yang of protein folding. *Febs J* 2005; 272:5962–70; PMID:16302961; <http://dx.doi.org/10.1111/j.1742-4658.2005.05021.x>
- Jahn TR, Radford SE. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys* 2008; 469:100–17; PMID:17588526; <http://dx.doi.org/10.1016/j.abb.2007.05.015>
- Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 2006; 75:333–66; PMID:16756495; <http://dx.doi.org/10.1146/annurev.biochem.75.101304.123901>
- Fowler DM, Koulov AV, Balch WE, Kelly JW. Functional amyloid—from bacteria to humans. *Trends Biochem Sci* 2007; 32:217–24; PMID:17412596; <http://dx.doi.org/10.1016/j.tibs.2007.03.003>
- Falson A, Falson SF. Legal but lethal: functional protein aggregation at the verge of toxicity. *Fron Cell Neurosci* 2015; 9:45; PMID:25741240; <http://dx.doi.org/10.3389/fncel.2015.00045>
- Levinthal C. Are there pathways for protein folding?. *J Chim Phys* 1968; 65:44–5
- Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. *Proc Natl Acad Sci U S A* 1992; 89:20–2; PMID:1729690; <http://dx.doi.org/10.1073/pnas.89.1.20>
- Lopez de la Paz M, Serrano L. Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci U S A* 2004; 101:87–92; PMID:14691246; <http://dx.doi.org/10.1073/pnas.2634884100>
- Sanchez de Groot N, Pallares I, Aviles FX, Vendrell J, Ventura S. Prediction of “hot spots” of aggregation in disease-linked polypeptides. *BMC Struct Biol* 2005; 5:18; PMID:16197548; <http://dx.doi.org/10.1186/1472-6807-5-18>
- Belli M, Ramazzotti M, Chiti F. Prediction of amyloid aggregation in vivo. *EMBO Rep* 2011; 12:657–63; PMID:21681200; <http://dx.doi.org/10.1038/embor.2011.116>
- Castillo V, Grana-Montes R, Sabate R, Ventura S. Prediction of the aggregation propensity of proteins from the primary sequence: aggregation properties of proteomes. *Biotechnol J* 2011; 6:674–85; PMID:21538897; <http://dx.doi.org/10.1002/biot.201000331>
- Espargaro A, Busquets MA, Estelrich J, Sabate R. Predicting the aggregation propensity of prion sequences. *Virus Res* 2015; S0168-1702(15)00119-7; PMID:25747492
- Rousseau F, Serrano L, Schymkowitz JW. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* 2006; 355:1037–47; PMID:16359707; <http://dx.doi.org/10.1016/j.jmb.2005.11.035>
- Aguzzi A, Calella AM. Prions: protein aggregation and infectious diseases. *Physiol Rev* 2009; 89:1105–52; PMID:19789378; <http://dx.doi.org/10.1152/physrev.00006.2009>
- Chien P, Weissman JS, DePace AH. Emerging principles of conformation-based prion inheritance. *Annu Rev Biochem* 2004; 73:617–56; PMID:15189155; <http://dx.doi.org/10.1146/annurev.biochem.72.121801.161837>
- Sabate R. When amyloids become prions. *Prion* 2014; 8; PMID:24831240; <http://dx.doi.org/10.4161/pri.29238>
- Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A, Ross ED. De novo design of synthetic prion domains. *Proc Natl Acad Sci U S A* 2012; 109:6519–24; PMID:22474356; <http://dx.doi.org/10.1073/pnas.1119366109>
- Uptain SM, Lindquist S. Prions as protein-based genetic elements. *Annu Rev Microbiol* 2002; 56:703–41; PMID:12142498; <http://dx.doi.org/10.1146/annurev.micro.56.013002.100603>
- Michelitsch MD, Weissman JS. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A* 2000; 97:11910–5; PMID:11050225; <http://dx.doi.org/10.1073/pnas.97.22.11910>
- Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell*

- 2009; 137:146–58; PMID:19345193; <http://dx.doi.org/10.1016/j.cell.2009.02.044>
21. Espinosa Angarica V, Angulo A, Giner A, Losilla G, Ventura S, Sancho J. PrionScan: an online database of predicted prion domains in complete proteomes. *BMC Genomics* 2014; 15:102; PMID:24498877; <http://dx.doi.org/10.1186/1471-2164-15-102>
 22. Espinosa Angarica V, Ventura S, Sancho J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC Genomics* 2013; 14:316; PMID:23663289; <http://dx.doi.org/10.1186/1471-2164-14-316>
 23. Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* 2014; 30(17):2501–2; PMID:24825614; <http://dx.doi.org/10.1093/bioinformatics/btu310>
 24. Sabate R, Rousseau F, Schymkowitz J, Ventura S. What makes a protein sequence a prion? *PLoS Comput Biol* 2015; 11:e1004013; PMID:25569335; <http://dx.doi.org/10.1371/journal.pcbi.1004013>
 25. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, MacLea KS, Freibaum B, Li S, Mollie A, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 2013; 495:467–73; PMID:23455423; <http://dx.doi.org/10.1038/nature11922>
 26. Tanaka M, Collins SR, Toyama BH, Weissman JS. The physical basis of how prion conformations determine strain phenotypes. *Nature* 2006; 442:585–9; PMID:16810177; <http://dx.doi.org/10.1038/nature04922>
 27. Derdowski A, Sindi SS, Klaips CL, DiSalvo S, Serio TR. A size threshold limits prion transmission and establishes phenotypic diversity. *Science* 2010; 330:680–3; PMID:21030659; <http://dx.doi.org/10.1126/science.1197785>
 28. Grana-Montes R, Marinelli P, Reverter D, Ventura S. N-terminal protein tails act as aggregation protective entropic bristles: the SUMO case. *Biomacromolecules* 2014; 15:1194–203; PMID:24564702; <http://dx.doi.org/10.1021/bm401776z>
 29. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL, et al. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 2005; 21:3435–8; PMID:15955783; <http://dx.doi.org/10.1093/bioinformatics/bti537>
 30. Maurer-Stroh S, Debulpaep M, Kuemmerer N, Lopez de la Paz M, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, et al. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Meth* 2010; 7:237–42; PMID:NOT_FOUND; <http://dx.doi.org/10.1038/nmeth.1432>
 31. Paulson HL, Perez MK, Trottier Y, Trojanowski JQ, Subramony SH, Das SS, Vig P, Mandel JL, Fischbeck KH, Pittman RN. Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron* 1997; 19:333–44; PMID:9292723; [http://dx.doi.org/10.1016/S0896-6273\(00\)80943-5](http://dx.doi.org/10.1016/S0896-6273(00)80943-5)
 32. Ross ED, Toombs JA. The effects of amino acid composition on yeast prion formation and prion domain interactions. *Prion* 2010; 4:60–5; PMID:20495349; <http://dx.doi.org/10.4161/pri.4.2.12190>
 33. Gonzalez Nelson AC, Paul KR, Petri M, Flores N, Rogge RA, Cascarina SM, Ross ED. Increasing prion propensity by hydrophobic insertion. *PLoS One* 2014; 9:e89286; PMID:24586661; <http://dx.doi.org/10.1371/journal.pone.0089286>
 34. Gazit E. A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J* 2002; 16:77–83; PMID:11772939; <http://dx.doi.org/10.1096/fj.01-0442hyp>
 35. Wasmer C, Lange A, Van Melckebeke H, Siemer AB, Riek R, Meier BH. Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science* 2008; 319:1523–6; PMID:18339938; <http://dx.doi.org/10.1126/science.1151839>
 36. Kajava AV, Baxa U, Steven AC. Beta arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils. *FASEB J* 2010; 24:1311–9; PMID:20032312; <http://dx.doi.org/10.1096/fj.09-145979>

5.3. PUBLICATION III

Amyloid cores in prion domains: key regulators for prion conformational conversion.

Fernández MR., **Batlle C.**, Gil-García M., Ventura S.

Prion (2017)

DOI: 10.1080/19336896.2017.1282020.

Amyloid cores in prion domains: Key regulators for prion conformational conversion

María Rosario Fernández, Cristina Batlle, Marcos Gil-García, and Salvador Ventura

Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain

ABSTRACT. Despite the significant efforts devoted to decipher the particular protein features that encode for a prion or prion-like behavior, they are still poorly understood. The well-characterized yeast prions constitute an ideal model system to address this question, because, in these proteins, the prion activity can be univocally assigned to a specific region of their sequence, known as the prion forming domain (PFD). These PFDs are intrinsically disordered, relatively long and, in many cases, of low complexity, being enriched in glutamine/asparagine residues. Computational analyses have identified a significant number of proteins having similar domains in the human proteome. The compositional bias of these regions plays an important role in the transition of the prions to the amyloid state. However, it is difficult to explain how composition alone can account for the formation of specific contacts that position correctly PFDs and provide the enthalpic force to compensate for the large entropic cost of immobilizing these domains in the initial assemblies. We have hypothesized that short, sequence-specific, amyloid cores embedded in PFDs can perform these functions and, accordingly, act as preferential nucleation centers in both spontaneous and seeded aggregation. We have shown that the implementation of this concept in a prediction algorithm allows to score the prion propensities of putative PFDs with high accuracy. Recently, we have provided experimental evidence for the existence of such amyloid cores in the PFDs of Sup35, Ure2, Swi1, and Mot3 yeast prions. The fibrils formed by these short stretches may recognize and promote the aggregation of the complete proteins inside cells, being thus a promising tool for targeted protein inactivation.

KEYWORDS. amyloids, prion forming domains, prion-like proteins, protein intrinsic disorder, Q/N-rich domains, yeast prions

Protein misfolding and aggregation is associated with a broad range of human disorders, including Parkinson's and Alzheimer's diseases.¹ The common underlying cause behind these pathologies is the conversion of specific soluble proteins into insoluble and highly ordered fibrillar aggregates, collectively referred as amyloid fibrils. Analysis by X-ray

Correspondence to: Salvador Ventura; Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193-Bellaterra (Barcelona), Spain; Email: salvador.ventura@uab.es

Received December 7, 2016; Revised December 29, 2016; Accepted January 4, 2017.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/kprn.

Extra View to: Sant'Anna R, Fernández MR, Batlle C, Navarro S, de Groot NS, Serpell L, Ventura S. Characterization of amyloid cores in prion domains. *Sci Rep* 2016; 6:34274; <http://dx.doi.org/10.1038/srep34274>

diffraction of these fibrils indicates that, in most cases, the polypeptide chains are embedded in an extended cross- β -conformation running perpendicular to the fibril long axis.² The folding of globular proteins into native structures relies on the establishment of an extensive network of interactions involving most of the protein sequence.³ In contrast, amyloid self-assembly seems to obey, in many cases, the “short-stretch hypothesis,” according to which, the formation of this supramolecular structure is first nucleated by the intermolecular contacts formed by a reduced number of specific short regions in the protein, named aggregation hotspots (HS) or short aggregation-prone regions (APRs)⁴ (Fig. 1A). These stretches are generally around 5–10 residues in length, with a predominant hydrophobic character and a low net charge.⁵

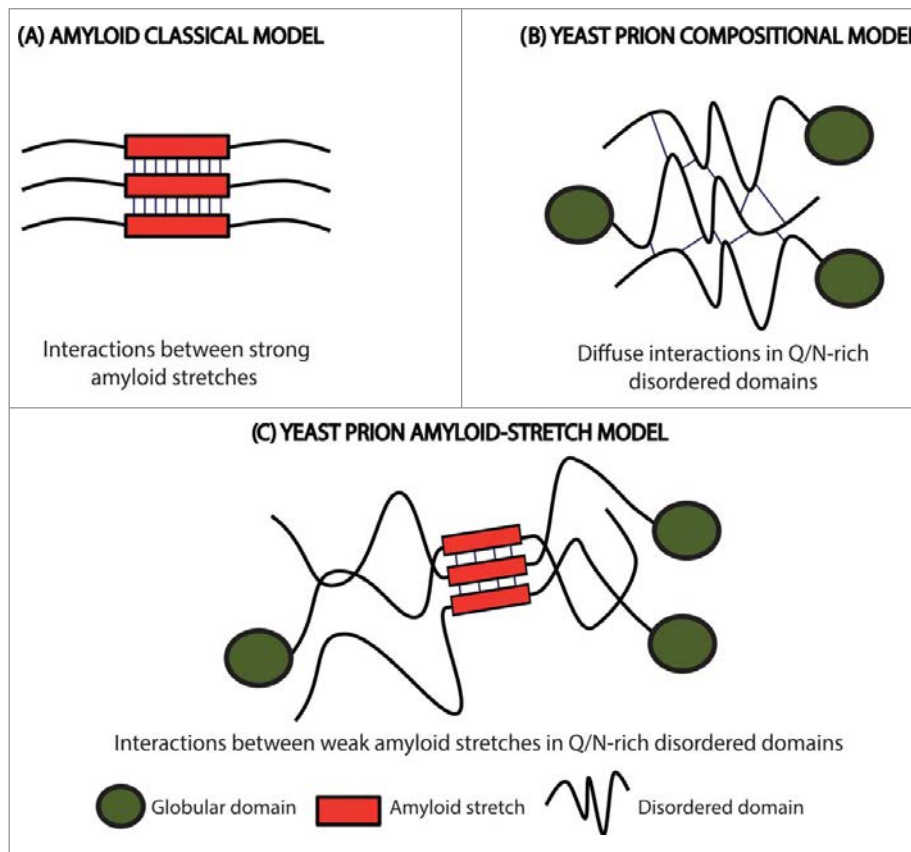
Prions are considered a subclass of amyloids with the ability to propagate their aggregated conformation and thus, to become potentially infectious. The prion phenomenon is best known by its association with spongiform encephalopathies in mammals,⁶ but increasing evidence indicates that similar templating mechanisms are exploited by nature for a variety of functional purposes,⁷ being protein-based epigenetic inheritance in yeast the best characterized of these processes.^{8,9} Like their mammalian counterparts, yeast prions undergo a self-perpetuating conversion into an amyloid conformation that shift the function of these proteins, which, in the case of yeast, might promote the expression of novel and, eventually, beneficial phenotypic traits.¹⁰ A common feature of many of these yeast prions is the presence of a long and intrinsically disordered region, which is enriched in glutamine/asparagine (Q/N) residues and usually depleted in hydrophobic ones, named prion-forming domain (PFD). This domain is both sufficient and necessary for prion conversion.¹¹ However, it should be mentioned that the presence of a low complexity Q/N-rich region is not a mandatory requirement for prion formation, since an increasing number of polypeptides able to promote protein-based inheritance are being discovered in fungi and other organisms that lack this specific compositional signature.

What it seems to be common to all PFDs is their intrinsically disordered nature in the soluble state of the prion protein.¹²

Whereas almost any protein bears the potential to form amyloid structures,¹³ only a small set of amyloid assemblies shows prion behavior in natural environments. Indeed, despite the structural similarity between the aggregated states of amyloids and prions, classical amyloid prediction algorithms fail to identify prion propensities in protein sequences.⁴ Actually, in spite of the particular sequential features of Q/N-rich PFDs being known since long time ago, we have little information on how they encode for a prion-like behavior. Traditionally, the self-assembly of these polypeptides have been thought to ultimately depend on their particular amino acid composition, with amyloid formation relying on the establishment of a large number of weak interactions distributed over the long PFD¹⁴ (Fig. 1B). A number of prion protein prediction programs have been developed according to this rationale. These algorithms score protein sequences according to their compositional similarity with the PFDs of a few, well-characterized, yeast prions. They have allowed to search for new PFDs candidates in previously unexplored proteomes.¹⁵ These predicted regions have been named prion-like domains (PrLDs), and the proteins holding them, prion-like proteins. Interestingly enough, the analysis of the human proteome revealed that a significant number of proteins displaying PrLDs are enriched in DNA or RNA binding domains and involved in regulatory functions, with some of them being associated with neurodegenerative diseases.¹⁶ Examples of these algorithms are DIANA,¹⁷ LPSs,¹⁸ PAPA,¹⁹ PrionScan,^{20,21} and PLAAC.²²

The success of composition-based predictors in the identification of new yeast proteins behaving like *bona fide* prions,¹⁵ together with the failure of classical amyloid prediction algorithms to do this task, has led to assume that the “short-stretch hypothesis” does not apply for prion or prion-like proteins.¹¹ This assumption was apparently consistent with the observation that, due to their disordered nature, PFDs are devoid of high local concentrations of hydrophobic residues.

FIGURE 1. Scheme of the initial intermolecular contacts leading to amyloid formation according to different models. (A) The amyloid classical model relies on the initial establishment of interactions (thin lines) between short APRs of about 5–10 residues with a predominant hydrophobic character (red boxes). (B) The prion compositional model relies on the initial establishment of a large number of diffuse weak interactions (thin lines) along the PFD. (C) In the new model for prion and prion-like proteins, the first contacts (thin lines) are formed between short amyloidogenic regions (red boxes) embedded into a Q/N-rich disordered region. The amyloid core is longer than those in the classical amyloid model, and, due to their particular composition, display less amyloidogenic potential. This allows the protein to remain soluble until needed and results in assemblies with a significant degree of brittleness.



The excellent hydrogen bonding capability of Q and N residues¹⁶ has led to propose that the formation of a diffuse network of hydrogen bonds would initiate the self-assembly reaction in PFDs. However, because hydrogen bonds between these residues are only slightly more stable than the ones they establish with the solvent, it is difficult to envision how the initial formation of delocalized and globally weak contacts can compensate enthalpically for the high entropic cost of immobilizing a long disordered and presumably highly flexible prion

domain. Therefore, we decided to re-explore the possibility that the presence of short stretches with a significant amyloid propensity embedded in the disordered Q/N-rich regions of PFDs might act as nucleating regions for amyloid assembly.¹⁷ Two reasons argued that the amyloid potential of these APRs, if they exist, should be weaker than those found in disease-linked amyloid polypeptides. First, despite PFDs should keep a certain amyloid capacity that permits the switch toward the insoluble state, their physicochemical characteristics

should be compatible with the protein being soluble under physiological conditions, at least during part of its lifetime. Fully exposed, highly aggregation-prone stretches will shift irreversibly the equilibrium toward the insoluble state, precluding any functional transition. Second, the amyloids formed by yeast prions should display a certain degree of brittleness (Fig. 1C), since this feature is crucial for their chaperone-assisted fragmentation and subsequent propagation between mother and daughter cells.¹⁸ The presence of a very strong amyloid core would prevent, or at least decrease, the fragmentation rate and thus reduce prion propagation potential.

Amyloid stretches in PFDs should fulfill two apparently contradictory properties: allow the domain to remain soluble and disordered and provide the nucleation force for amyloid formation. It is known, that polar and charged residues favor disorder, whereas hydrophobic amino acids favor aggregation. Then, how these 2 properties can be encoded at the same time in a short sequence stretch? By computationally analyzing the ability of the 20 proteino-genic amino acids to promote disorder and aggregation, we discovered that N and Q are the residues that best balance amyloid and disorder propensities.¹¹ We proposed that this unique property accounts for the over-representation of these residues in yeast PFDs.¹¹ Tyrosine (Y) is the most abundant hydrophobic residue in yeast PFDs.¹⁵ Our analysis indicated that Y is clearly superior to the rest of apolar residues in terms of disorder propensity, appearing thus as the best residue to endorse prionogenic Q/N-rich amyloid cores with increased amyloid potential without disturbing significantly the PFDs disorder properties. However, computational simulations soon demonstrated that a 5–10 residues long stretch based on Q/N residues would not have enough amyloid potential to drive by itself the transition toward the fibrillar state, even if it contains a certain number of Y residues. Thus, we hypothesized that the amyloid cores of PFDs should be longer, in such a way that the amyloid potential would be more distributed than in classical amyloid stretches; each residue having an average lower contribution, but with more

residues contributing to assembling driving force. We implemented this notion in a novel PFDs prediction algorithm named pWALTZ.¹⁷ pWALTZ predicts the 21 residues long sequence stretch with the average highest amyloidogenic potential into a Q/N-rich context and classifies proteins into the prion/non-prion categories according to the presence and the potency of these stretches. Interestingly enough, this approach displays better accuracy in discriminating prion propensity than algorithms relying only in the compositional model.¹⁷ Later on, this algorithm was incorporated into the publically accessible PrionW webserver.¹⁹

The fact that an algorithm based on the “short-stretch amyloid hypothesis” can infer the prion propensity of a sequence suggests, but does not prove, that these stretches really exist. Recently, we have provided experimental evidence for the presence of short amyloid cores in the PFDs of some of the most representative canonical yeast prions: Sup35, Ure2, Swi1 and Mot3, which are all implicated in regulatory functions relevant for cell adaptation to the changing environment. Sup35 is a component of translation termination complex. When Sup35 is converted to its prion form [*PSI*⁺], stop codons can be read-through increasing the phenotypic variability and therefore multiplying the chances to attain optimal yeast fitness.²⁰ Ure2 is a negative regulator of enzymes involved in nitrogen metabolism. In its prion form, named as [*URE3*], the repression is removed and the yeast cell can import ureido-succinate.²¹ Swi1, a subunit of the SWI/SNF ATP-dependent chromatin-remodeling complex implicated in the expression of 6% of the yeast genome, can become a prion named [*SWI*⁺]. Cells containing the prion [*SWI*⁺] show a phenotype indicative of partial loss of function of SWI/SNF²². Mot3 transcriptional factor in its prion form [*MOT3*⁺] regulates the facultative acquisition of multicellular structures in response to natural environmental signals.²³

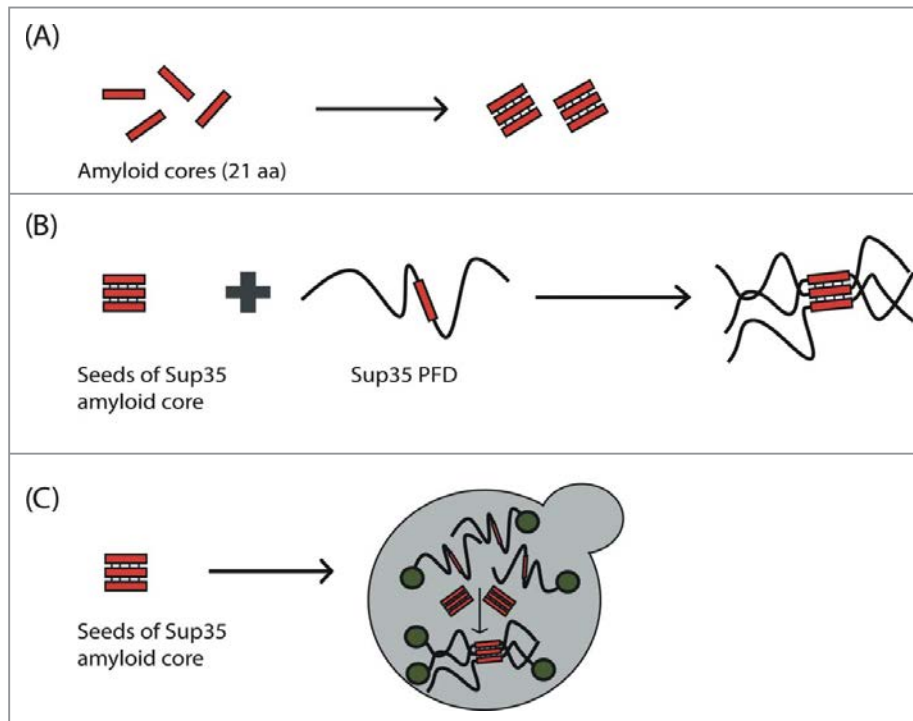
We first searched for the presence of amyloid cores into the PFDs of the above described 4 proteins using the pWALTZ formalism. In all the cases, we could identify at least one sequence that fulfills the algorithm

requirements.²⁴ As expected, classical aggregation predictors like TANGO,²⁵ AGGRES-CAN²⁶ or PASTA²⁷ were unable to identify these amyloid stretches, since the abundance of Q/N residues in these sequences is interpreted by these programs as a signature of low amyloid potential. This highlights how, as hypothesized, despite the “short-stretch amyloid hypothesis” would apply for both classical amyloids and prions, the specific features of their cores differ significantly.

We synthesized the peptides corresponding to the 4 predicted cores and show that all of them were able to experiment a transition between an initially disordered and soluble conformation and a β -sheet rich state, as determined by Far-UV circular dichroism and infrared spectroscopy. Moreover, analysis by Thioflavin T binding, transmission

electron microscopy and X-ray diffraction, all converged to indicate that these β -sheet rich conformations correspond to archetypical amyloid fibrils exhibiting a cross- β structure (Fig. 2A). The fibrils formed by these amyloid cores were able to seed the aggregation of their corresponding soluble form, but, despite all them share a similar composition, with $\sim 1/3$ of their residues being Q/N, no cross-seeding effect between cores could be observed. This indicated that, at least for these particular PFDs segments, they were sequence-specific contacts that drove their aggregation. Importantly, in the case of Sup35, the fibrils formed by the peptide were able to promote the accelerated aggregation of the complete PFD *in vitro* (Fig. 2B).²⁴ Indeed, they seeded the reaction with higher efficiency than the fibrils formed

FIGURE 2. Scheme of the experimental evidence for the presence of amyloid cores in canonical yeast prions. (A) Amyloid cores from Sup35, Ure2, Swi1 and Mot3 are able to make the transition from the soluble and disordered state to amyloid fibrils with cross- β structure. (B) The Sup35 amyloid core is able to promote (seed) the aggregation of the complete Sup35 PFD *in vitro*. (C) Introduction of Sup35 amyloid core seeds into yeast is able to induce the aggregation of the endogenous Sup35 protein *in vivo* and, consequently, the expression of the prionic phenotype.



by the entire PFD, which strongly argues that this stretch was able to nucleate the reaction by recognizing the homologous sequence in the soluble and disordered PFD. This will explain why, when the fibrils formed by the amyloid core of Sup35 are introduced in living yeast cells, they are able to promote the shift of a fraction of the population to the prionic state (Fig. 2C),²⁴ behaving thus as autonomous propagation-competent entities. This suggests that they identified regions may act as key regulators of the phenotypic conversion induced by yeast prions. Chaperone-mediated fragmentation is indispensable to propagate yeast prions; despite speculative, it is tempting to propose that, upon chaperone cleavage, fragments containing the identified amyloid cores might act as preferential seeds in the conversion of the soluble prion proteins present in daughter cells toward the amyloid state.

Overall, our results indicate that the sustained mechanism of functional prion assembly would resemble more than initially thought to that of disease-linked amyloids. This view is not in opposition to the previous compositional model, but it goes a step further by defining a region with a higher probability to initiate the conformational conversion. The Q/N compositional bias would define the structurally disordered context, while providing a cryptic and distributed aggregation propensity that can zipper the complete PFD upon amyloid core positioning and subsequent nucleation. In a way, the reaction would resemble that occurring in polyQ diseases.²⁸ In these disorders, the expansion of an intrinsically disordered polyQ tract flanking a folded globular domain results in protein aggregation. Despite the expanded polyQ stretch is critical for amyloid assembly, it is clear that its lack of sequence specificity makes difficult that it could encode an accurate, in register, disposition of the polypeptide chains at the beginning of the aggregation reaction. Accordingly, increasing evidence indicates that, in these proteins, aggregation is initiated by contacts between specific APRs in, or close to, the globular domain and only latter it is propagated to the polyQ segment.^{29,30} This sequence-specific initiation of amyloid

formation would allow a proper pairing and hydrogen bonding of the residues in the monotonic polyQ track. In a similar manner, in addition to contribute to the initial force for aggregation, amyloid cores in PFDs would allow an accurate positioning and intermolecular interaction of the disordered Q/N-rich regions that flank them. The same scheme will apply for seeding reactions, making cross-seeding between different PFDs less probable, despite they share very similar composition.

We want to make clear that the selected size of 21 residues for amyloid cores is arbitrary. It was the length that performed best in discriminating prion *versus* non prion behavior in Q/N-rich domains¹⁷ and also because this length has been shown to coincide with the minimal motif accounting for HET-S prion transmission.³¹ However, we assume that amyloid cores of different lengths would indeed exist, the size depending on both their composition and the sequential context in which they are embedded. In addition, we do not propose that the identified amyloid cores are the exclusive players in the aggregation process of the studied PFDs, they can clearly exist other short regions that collaborate either in a cooperative or additive way to the process. This is clearly the case of Sup35, for which a region adjacent to the N-terminus of the assayed amyloid-core has been shown to play an essential role in the aggregation and propagation of this protein.³² In this sense, we are developing a version of pWALTZ able to identify and compare all the putative amyloid cores in a Q/N-rich sequence. The preliminary analysis indicates that the average number of amyloid cores in PFDs is rather small.

It is interesting to notice that the characteristics of amyloid cores in PFDs resemble that of Short Linear Motifs (SLiMs)³³ found in intrinsically disordered proteins. SLiMs are short stretches of adjacent amino acids that mediate protein-protein association, also known as linear motifs (LMs) and molecular recognition features (MoRFs). SLiMs are usually constituted by a short stretch of contiguous amino acids, and their binding can be modulated by residues outside this region. As it happens with yeast prion amyloid cores, these SLiMs are

generally placed in unstructured regions and, when their binding partners are absent, they lack stable tertiary structure.³⁴ It has been proposed that the short length and the small number of essential residues defining these SLiMs make them much more easy to evolve through point mutations than equivalent structural binding motifs in the context of globular domains.³⁴ The new rudimentary motif created by these point mutations can be consequently selected positively or negatively by evolutionary pressures, to produce a functional SLiM or to eliminate a pernicious interaction, respectively.³⁴ It is feasible that a similar mechanism might apply to the amyloid cores we have identified in yeast prions. Genetic mutations generating new amyloid cores would be selected as long as they result advantageous for the cell and/or the population fitness. From an evolutionary point of view, the fact that sequences fully exposed to solvent and with a significant amyloid potential, as those we have identified in PFDs, have not been purged out by natural selection can only be explained if they serve for functional purposes,³⁵ since there is a strong selective pressure to reduce the amyloidogenic potential of protein sequences.³⁶ The presence of these regions is inherently risky, since mutations that increase their amyloid potential can shift the equilibrium to an aggregated and potentially toxic state. This is exactly what happens in the case of human hnRNPA1 and hnRNA2, two prion-like RNA-binding proteins which genetic mutation is associated with multisystem proteinopathy and amyotrophic lateral sclerosis.³⁷ The analysis of these pathogenic mutations with pWALTZ indicated that they map at the predicted amyloid cores, increasing their amyloid propensity.¹⁷ This mutation-induced pro-aggregational effect might also occur in other human prion-like proteins and explain why, as a group, they appear to be linked to neurological disease.³⁸ pWALTZ might turn to be useful in the detection and prediction of pathogenic mutations in this subproteome.

The ability of Sup35 amyloid core to fibrillate and to induce the intracellular aggregation of the homologous yeast prion protein, suggests that a similar strategy can be used for the

targeted inactivation of human prion-like proteins. A related approach has been recently shown to be effective to target the vascular endothelial growth factor 2 (VEGFR2), a protein containing amyloidogenic segments, but which is not known to aggregate in either pathological or normal conditions.³⁹ The internalization inside cells of a synthetic peptide, corresponding to a tandem repeat of a short amyloidogenic region of VEGFR2, induced its aggregation *in vivo* through a direct amyloid interaction, inactivating the protein and, accordingly, inhibiting VEGFR2-dependent tumor growth in a mouse tumor model. Interestingly enough, despite VEGFR2 exhibited a significant number of amyloid stretches, those belonging to the signal peptide exhibited the best performance. This is likely because in contrast to the ones mapping in the globular domain, they are already exposed to solvent, before the signal peptide is processed. Computational predictions indicate that amyloid cores in human PrLDs are placed in a similar structural context, which anticipates that they can be potentially targeted by homologous fibrillated peptides to inactivate the function of selected prion-like proteins. Work in this direction is ongoing in our lab.

ABBREVIATIONS

APRs	short aggregation-prone regions
PFD	prion-forming domain
PrLD	prion-like domain
SLiM	short linear motifs
VEGFR2	vascular endothelial growth factor 2

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

FUNDING

This work was supported by the Spanish “Ministerio de Economía y Competitividad” [BFU2013-44763-P] to S.V and by ICREA, ICREA-Academia 2015 to S.V.

REFERENCES

- [1] Invernizzi G, Papaleo E, Sabate R, Ventura S. Protein aggregation: Mechanisms and functional consequences. *Int J Biochem Cell Biol* 2012; 44 (9):1541-54; PMID:22713792; <http://dx.doi.org/10.1016/j.biocel.2012.05.023>
- [2] Sabaté R, Ventura S. Cross- β -sheet supersecondary structure in amyloid folds: techniques for detection and characterization. *Methods Mol Biol* 2012; 932:237-57.
- [3] Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973; 181(4096):223-30; PMID:4124164; <http://dx.doi.org/10.1126/science.181.4096.223>
- [4] Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A, Ross ED. De novo design of synthetic prion domains. *Proc Natl Acad Sci U S A* 2012; 109 (17):6519-24; PMID:22474356; <http://dx.doi.org/10.1073/pnas.1119366109>
- [5] de Groot NS, Pallarés I, Avilés FX, Vendrell J, Ventura S. Prediction of “hot spots” of aggregation in disease-linked polypeptides. *BMC Struct Biol* 2005; 5:1-15; PMID:15663787; <http://dx.doi.org/10.1186/1472-6807-5-18>
- [6] Collins SJ, Lawson VA, Masters CL. Transmissible spongiform encephalopathies. *Lancet* 2004; 363 (9402):51-61; PMID:14723996; [http://dx.doi.org/10.1016/S0140-6736\(03\)15171-9](http://dx.doi.org/10.1016/S0140-6736(03)15171-9)
- [7] Si K. Prions: What are they good for? *Annu Rev Cell Dev Biol* 2015; 31(1):149-69.
- [8] Chien P, Weissman JS, DePace AH. Emerging principles of conformation-based prion inheritance. *Annu Rev Biochem* 2004; 73(1):617-56; PMID:15189155; <http://dx.doi.org/10.1146/annurev.biochem.72.121801.161837>
- [9] Liebman SW, Chernoff YO. Prions in yeast. *Genetics* 2012; 191(4):1041-72; PMID:22879407; <http://dx.doi.org/10.1534/genetics.111.137760>
- [10] Wickner RB, Edskes HK, Gorkovskiy A, Bezsonov EE, Stroobant EE. Yeast and fungal prions: amyloid-handling systems, amyloid structure, and prion biology. *Adv Genet* 2016; 93:191-236; PMID:26915272
- [11] Sabate R, Rousseau F, Schymkowitz J, Batlle C, Ventura S. Amyloids or Prions? That is the Question. *Prion* 2015; 9:200-6; PMID:26039159; <http://dx.doi.org/10.1080/19336896.2015.1053685>
- [12] Chakrabortee S, Byers JS, Jones S, Garcia DM, Bhullar B, Chang A, She R, Lee L, Fremin B, Lindquist S, et al. Intrinsically disordered proteins drive emergence and inheritance of biological traits. *Cell* 2016; 167(2):369-81; PMID:27693355; <http://dx.doi.org/10.1016/j.cell.2016.09.017>
- [13] Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, Dobson CM. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci U S A* 1999; 96(7):3590-4; PMID:10097081; <http://dx.doi.org/10.1073/pnas.96.7.3590>
- [14] Ross ED, Edskes HK, Terry MJ, Wickner RB. Primary sequence independence for prion formation. *Proc Natl Acad Sci* 2005; 102(36):12825-30; PMID:16123127; <http://dx.doi.org/10.1073/pnas.0506136102>
- [15] Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A Systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 2009; 137(1):146-58; PMID:19345193; <http://dx.doi.org/10.1016/j.cell.2009.02.044>
- [16] Stapley BJ, Doig AJ. Hydrogen bonding interactions between glutamine and asparagine in α -helical peptides. *J Mol Biol* 1997; 272(3):465-73; PMID:9325104; <http://dx.doi.org/10.1006/jmbi.1997.1262>
- [17] Sabate R, Rousseau F, Schymkowitz J, Ventura S. What makes a protein sequence a prion? *PLoS Comput Biol* 2015; 11(1):e1004013; PMID:25569335; <http://dx.doi.org/10.1371/journal.pcbi.1004013>
- [18] Tanaka M, Collins SR, Toyama BH, Weissman JS. The physical basis of how prion conformations determine strain phenotypes. *Nature* 2006; 442 (7102):585-9; PMID:16810177; <http://dx.doi.org/10.1038/nature04922>
- [19] Zambrano R, Conchillo-Sole O, Iglesias V, Illa R, Rousseau F, Schymkowitz J, Sabate R, Daura X, Ventura S. PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res* 2015; 43(W1):W331-7; PMID:25977297; <http://dx.doi.org/10.1093/nar/gkv490>
- [20] Serio TR, Lindquist. The yeast prion [PSI⁺]: molecular insights and functional consequences. *Adv Protein Chem* 2002; 59:391-412.
- [21] Wickner RB. [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* 1994; 264:5669; PMID:7909170; <http://dx.doi.org/10.1126/science.7909170>
- [22] Du Z, Park KW, Yu H, Fan Q, Li L. Newly identified prion linked to the chromatin-remodeling factor Swi1 in *Saccharomyces cerevisiae*. *Nat Genet* 2008; 40(4):460-5; PMID:18362884; <http://dx.doi.org/10.1038/ng.112>
- [23] Holmes DL, Lancaster AK, Lindquist S, Halfmann R. Heritable Remodeling of Yeast Multicellularity by an Environmentally Responsive Prion. *Cell* 2013; 153(1):153-65; PMID:23540696; <http://dx.doi.org/10.1016/j.cell.2013.02.026>
- [24] Sant'Anna R, Fernández MR, Batlle C, Navarro S, de Groot NS, Serpell L, Ventura S. Characterization of amyloid cores in prion domains. *Sci Rep* 2016;

- 6:34274; PMID:27686217; <http://dx.doi.org/10.1038/srep34274>
- [25] Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 2004; 22(10):1302-6; PMID:15361882; <http://dx.doi.org/10.1038/nbt1012>
- [26] de Groot NS, Castillo V, Graña-Montes R, Ventura S. AGGRESKAN: method, application, and perspectives for drug design. *Methods Mol Biol* 2012; 819:199-220; PMID:22183539
- [27] Walsh I, Seno F, Tosatto SCE, Trovato A. PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res* 2014; 42:W301-7; PMID:24848016; <http://dx.doi.org/10.1093/nar/gku399>
- [28] Petrakis S, Schaefer MH, Wanker EE, Andrade-Navarro MA. Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. *Bioessays* 2013; 35(6):503-7; PMID:23483542; <http://dx.doi.org/10.1002/bies.201300001>
- [29] Lupton CJ, Steer DL, Wintrode PL, Bottomley SP, Hughes VA, Ellisdon AM. Enhanced Molecular Mobility of Ordinarily Structured Regions Drives Polyglutamine Disease. *J Biol Chem* 2015; 290(40):24190-200; PMID:26260925; <http://dx.doi.org/10.1074/jbc.M115.659532>
- [30] Scarff CA, Almeida B, Fraga J, Macedo-Ribeiro S, Radford SE, Ashcroft AE. Examination of Ataxin-3 (atx-3) Aggregation by Structural Mass Spectrometry Techniques: A Rationale for Expedited Aggregation upon Polyglutamine (polyQ) Expansion. *Mol Cell Proteomics* 2015; 14(5):1241-53; PMID:25700012; <http://dx.doi.org/10.1074/mcp.M114.044610>
- [31] Wan W, Stubbs G. Fungal prion HET-s as a model for structural complexity and self-propagation in prions. *Proc Natl Acad Sci U S A* 2014; 111(14):5201-6; PMID:24706820; <http://dx.doi.org/10.1073/pnas.1322933111>
- [32] Osherovich LZ, Cox BS, Tuite MF, Weissman JS. Dissection and design of yeast prions. *PLoS Biol* 2004; 2(4):e86; <http://dx.doi.org/10.1371/journal.pbio.0020086>
- [33] Mészáros B, Dosztányi Z, Simon I, Blow N, Jones S, Thornton J, Diella F, Haslam N, Chica C, Budd A, et al. Disordered binding regions and linear motifs—bridging the gap between two models of molecular recognition. *PLoS One* 2012; 7(10):e46829; PMID:23056474; <http://dx.doi.org/10.1371/journal.pone.0046829>
- [34] Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ, Davey NE. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev* 2014; 114(13):6733-78; PMID:24926813; <http://dx.doi.org/10.1021/cr400585q>
- [35] Malinowska L, Kroschwald S, Alberti S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim Biophys Acta* 2013; 1834(5):918-31; PMID:23328411; <http://dx.doi.org/10.1016/j.bbapap.2013.01.003>
- [36] Buck PM, Kumar S, Singh SK. On the role of aggregation prone regions in protein evolution, stability, and enzymatic catalysis: insights from diverse analyses. *PLoS Comput Biol* 2013; 9(10):e1003291; PMID:24146608; <http://dx.doi.org/10.1371/journal.pcbi.1003291>
- [37] Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, MacLea KS, Freibaum B, Li S, Mollieux A, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 2013; 495(7442):467-73; PMID:23455423; <http://dx.doi.org/10.1038/nature11922>
- [38] An L, Harrison PM. The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biol Direct* 2016; 11:32; PMID:27457357; <http://dx.doi.org/10.1186/s13062-016-0134-5>
- [39] Gallardo R, Ramakers M, De Smet F, Claes F, Khodaparast L, Khodaparast L, Couceiro JR, Langenberg T, Siemons M, Nyström S, et al. De novo design of a biologically active amyloid. *Science* 2016; 354(6313):aah4949; PMID:27846578; <http://dx.doi.org/10.1126/science.aah4949>

5.4. PUBLICATION IV

Characterization of soft amyloid cores in human prion-like proteins.

Battle C., de Groot NS., Iglesias V., Navarro S. and Ventura S.

Sci Rep (2017)

DOI: 10.1038/s41598-017-09714-z

SCIENTIFIC REPORTS



OPEN

Characterization of Soft Amyloid Cores in Human Prion-Like Proteins

Cristina Batlle¹, Natalia Sanchez de Groot^{2,3}, Valentin Iglesias¹, Susanna Navarro¹ & Salvador Ventura¹

Prion-like behaviour is attracting much attention due to the growing evidences that amyloid-like self-assembly may reach beyond neurodegeneration and be a conserved functional mechanism. The best characterized functional prions correspond to a subset of yeast proteins involved in translation or transcription. Their conformational promiscuity is encoded in Prion Forming Domains (PFDs), usually long and intrinsically disordered protein segments of low complexity. The compositional bias of these regions seems to be important for the transition between soluble and amyloid-like states. We have proposed that the presence of cryptic soft amyloid cores embedded in yeast PFDs can also be important for their assembly and demonstrated their existence and self-propagating abilities. Here, we used an orthogonal approach in the search of human domains that share yeast PFDs compositional bias and exhibit a predicted nucleating core, identifying 535 prion-like candidates. We selected seven proteins involved in transcriptional or translational regulation and associated to disease to characterize the properties of their amyloid cores. All of them self-assemble spontaneously into amyloid-like structures able to propagate their polymeric state. This provides support for the presence of short sequences able to trigger conformational conversion in prion-like human proteins, potentially regulating their functionality.

A broad range of human pathologies, ranging from neurodegenerative conditions such as Alzheimer's and Parkinson's diseases, to non-neuronal disorders such as type II diabetes and cataracts, are associated with protein misfolding and aggregation into amyloid-like structures¹. The self-assembly of proteins into β -sheet-enriched amyloid conformations appears to obey, in many cases, the so-called amyloid-short-stretch hypothesis, according to which, aggregation is first nucleated by the intermolecular contacts formed by a reduced number of specific short regions in the protein². In pathogenic proteins, these stretches are generally around 5–10 residues in length, with a high aggregation propensity and a predominant hydrophobic character³.

Prions are proteins able to adopt multiple structural conformations, from which at least one has self-propagating properties, usually an amyloid state⁴. Prions have been traditionally associated with the onset of mammalian neurophatologies⁵. Nevertheless, there are evidences that prion-like mechanisms are not always deleterious and instead they can be used for beneficial purposes⁶. The best characterized set of functional prions has been found in yeast, where they can behave as epigenetic elements, facilitating adaptation to fluctuating environments^{7,8}. The conformational promiscuity of yeast prions is encoded in Prion Forming Domains (PFDs)⁹. PFDs are both sufficient and necessary for prion conversion and usually correspond to long and intrinsically disordered segments of low complexity⁹.

The information on the common features shared by yeast PFDs has fueled the development of algorithms aimed to identify similar prion-like domains (PrLDs) and the proteins that contain them at the proteome level^{9–17}. Despite PFDs bear the capacity to shift to an amyloid state, classical amyloid prediction algorithms fail to identify them¹². This observation led to suggest that PFDs assembly is governed by the low complexity and compositional bias common to these domains^{9,12}. Examples of algorithms exploiting these features are DIANA¹⁰, LPS¹¹, PrionScan¹³, PLAAC¹⁶ and PAPA¹². Remarkably, these computational approaches predict the existence of PrLD-containing proteins in a wide variety of organisms, from prokaryotes to higher eukaryotes^{18–23}. In humans, this sub-proteome is enriched in nucleic acid-binding proteins^{18,19}. A fraction of these proteins seem to be involved in the formation of membraneless intracellular compartments, like RNA granules, through PrLDs

¹Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain. ²Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003, Barcelona, Spain. ³Universitat Pompeu Fabra (UPF), Barcelona, Spain. Correspondence and requests for materials should be addressed to S.V. (email: salvador.ventura@uab.es)

Received: 19 April 2017

Accepted: 28 July 2017

Published online: 21 September 2017

mediated liquid-liquid phase separation^{24,25}. Mutations in these domains have been shown to promote an aberrant transition to an aggregated amyloid-like state, the formation of which might lead to the onset of neurodegenerative diseases^{26,27}.

We have recently proposed that, in addition to composition, the presence of soft amyloid cores inside PrLDs and PrLDs could be important for their assembly^{28,29}. We rationalized that these assembly-nucleating regions should be longer than classical amyloid stretches, in such a way that the amyloid potential would be more diffusely distributed; each residue having an average lower potency, but with more residues contributing to the assembling force. This will make their aggregation sensitive to protein concentration and seeding. Thus, in our view, the aggregation of prion-like domains shares mechanistic features with those of classical amyloidogenic proteins. This notion was implemented in pWALTZ, an algorithm that predicts the 21-residues long sequence stretch with the highest average amyloid potential in PrLDs^{14,15}. This concept was further validated experimentally by demonstrating the existence of such soft amyloid stretches in the prion domains of four of the best characterized yeast prions³⁰, as well as in the predicted PrLD of the Rho termination factor of *Clostridium botulinum*³¹, which later led to the discovery of the first bacterial prion-like protein³².

Here, we study the presence of soft amyloid cores in novel putative human prion-like proteins. For this purpose, we performed a stringent computational analysis of the human proteome in the search of domains that, while fulfilling the compositional bias characteristic of PrLDs, would also exhibit a sequence stretch that can potentially nucleate their self-assembly. From this set, we selected seven nucleic acid-binding proteins associated to disease (DDX5, EYA1, ILF3, MED15, NCOA2, PHC1 and TIA1) to structurally characterize the nature of their putative nucleating cores. The results herein indicate that the PrLDs of all these proteins include 21-residues long stretches able to self-assemble spontaneously into non-toxic, β -sheet enriched, Th1 flavin-T positive amyloid-like structures displaying self-seeding activity. Therefore, the present work provides compelling experimental evidence for the existence of specific sequences with the potential to trigger the conformational conversion of PrLDs in human proteins.

Results

Identification of PrLD soft amyloid cores in human prion-like proteins. The analysis of the 70940 protein sequences in the human proteome was initially performed with PAPA¹² and further refined with pWALTZ¹⁴. Both PAPA and pWALTZ algorithms were trained on top of yeast prions; however, they are based on radically different concepts, a suitable composition of the PrLD and the presence of an embedded soft amyloid core, respectively. Sequences identified by these two orthogonal approaches are expected to recapitulate the conformational properties of yeast prions. A total of 663 human proteins were identified by PAPA and later short-listed using pWALTZ to render a total of 535 polypeptides, encoded by 336 different genes (Table S1).

We wanted to test whether the soft amyloid cores predicted inside the PrLDs of these putative human prion-like proteins could spontaneously self-assemble into amyloid-like conformations and propagate their aggregated state, as we observed before for yeast prions³⁰ and the prion-like *C. botulinum* Rho factor³¹. We focused on nucleic acid-binding proteins, both because this molecular function is enriched in our dataset and because most of the experimentally validated yeast prions act in translational or transcriptional regulation. We selected six proteins associated to disease whose prionogenic properties have not been reported before: DDX5, EYA1, ILF3, MED15, NCOA2 and PHC1. We also included in the analysis TIA1, an RNA-binding protein identified by the orthogonal approach for which a prion-like behavior has been already suggested³³ (Tables 1 and 2).

DDX5 (p68) (Fig. 1A) is a member of the DEAD box family of RNA helicases involved in transcriptional regulation³⁴ and it is overexpressed in various types of cancers such as those of prostate, breast and colon; promoting cell proliferation and metastasis^{35,36}. The C-terminal region containing the predicted PrLD seems to play a role in the interaction of the protein with other components of the transcriptional machinery³⁷.

EYA1 (Fig. 2A) is a transcriptional coactivator and a protein phosphatase with regulatory roles in nephrogenesis³⁸. The protein is overexpressed in breast cancer³⁹ and mutations in EYA1 gene are associated with branchio-oto-renal syndrome⁴⁰. The N-terminal region containing the PrLD has been reported to function as a transactivation domain⁴¹.

ILF3 (Fig. 3A) works in RNA metabolism, from transcription to degradation and it appears to be essential for cellular development⁴². Interestingly, ILF3 participates in ribonucleoprotein (RNP) granules assembly⁴³ and it interacts with FUS, a well-characterized prion-like protein⁴⁴.

MED15 (Fig. 4A) is one part of the Mediator complex involved in the transcription of RNA-polymerase II dependent genes⁴⁵. The identified PrLD corresponds to a Q-rich region similar to those accounting for conformational conversion in yeast prions. Indeed, its yeast homolog has been already classified as a prion⁴⁶ with the ability to form amyloid-like structures *in vivo* under stress conditions⁴⁷.

NCOA2 (Fig. 5A) is a transcriptional coactivator for steroid receptors and nuclear receptors acting in the upregulation of DNA expression⁴⁸. It has a tissue-specific role in tumorigenesis, acting as an oncogene in prostate cancer⁴⁹ and as a tumor suppressor in liver cancer⁵⁰. Moreover, NCOA2 is a key player in glucose homeostasis, being involved in Mediator recruitment for glucokinase expression⁵¹.

PHC1 (Fig. 6A) is one component of the Polycomb complex responsible for cellular differentiation during development. This complex is constituted by gene silencing proteins that repress important developmental regulator genes, including homeotic (HOX) genes⁵². PHC1 is associated with primary microcephaly⁵³.

TIA1 (Fig. 7A) is an RNA-binding protein and a component of stress granules required to regulate alternative splicing and mRNA translation and turnover. The predicted PrLD lies in a Q-rich region. TIA1 has already been suggested to be a functional prion-like protein whose dysregulation is involved in Amyotrophic Lateral Sclerosis (ALS)³³. However, the presence of an amyloid core inside its PrLD has not been addressed before.

To further confirm the presence of PrLDs in these proteins and to define more precisely their boundaries we used PLAAC, yet another composition-based predictor, in which, in contrast to PAPA, the length of the predicted

PROTEIN	FUNCTION	DISEASE
DDX5	RNA helicase protein involved in transcriptional regulation	Prostate, breast and colon cancer, leukemia, hepatitis and others.
EYA1	Transcriptional coactivator and protein phosphatase	Branchio-oto-renal syndrome, breast cancer, cataract and others.
ILF3	Facilitates gene expression regulation from transcription to degradation	Cancer, myocardial infarction, chronic kidney disease and others.
MED15	Component of the mediator complex involved in the transcription of RNA-pol II dependent genes	Epicondylitis, prostate cancer, prostatitis and others.
NCOA2	Nuclear receptors and steroid receptors coactivator	Prostate cancer, mesenchymal chondrosarcoma and others.
PHC1	Component of a Polycomb group involved in the maintenance of the transcriptionally repressive state of HOX genes	Microcephaly.
TIA1	Regulates alternative splicing, associated with apoptosis and 3'UTR mRNA binding protein	Amyotrophic Lateral Sclerosis, Hodgkin's Lymphoma, myopathy, Spinal Muscular Atrophy and others.

Table 1. Function and implication in disease of the selected human PrLD-containing proteins. Uniprot⁷⁷ and Malacards Human Disease database⁷⁸ were used to determine the function and disease-association for each human PrLD-containing candidate, respectively. All of them are involved in transcriptional or translational regulatory functions and have been associated to common diseases such as cancer or degenerative disorders.

PROTEIN	UNIPROT ID	PrLD AMYLOID CORE	pWALTZ SCORE	Q/N/S/G (%)
DDX5	P17844	530-TQNGVYSAANYTNGSFGSNFV-550	71.62	52.38
EYA1	Q99502	184-MQGSSFTTSSGIYTGNNSLTN-204	74.69	57.14
ILF3	Q12906	672-YGSYGYGGNSATAGYSQFYNS-692	71.01	57.14
MED15	Q96RN5	184-QQQQQFQAQQSAMQQQFQAVV-204	70.58	61.90
NCOA2	Q15596	1374-HFGQQANTSMYSNNMNINVS-1394	70.39	52.38
PHC1	P78364	382-QQQQIHLQKQVVIQQQIAIH-402	74.58	47.62
TIA1	P31483	331-AYGMYGQAWNQQGFNQSSA-351	67.81	57.14

Table 2. Selected human PrLD amyloid cores. For each PrLD-containing protein it is shown its Uniprot ID, its 21 residues-long amyloid core, the pWALTZ score for this protein region and its Q/N/S/G content.

PrLD also depends on the protein composition. PLAAC detected PrLDs, overlapping with the regions previously identified by PAPA, in all the above described proteins (Figs 1A to 7A).

The predicted soft amyloid cores for these proteins are shown in Table 2. Remarkably, well-validated aggregation predictors like Aggrescan⁵⁴, Tango⁵⁵ and Zyggregator⁵⁶ failed to classify these stretches as aggregation-prone, the exception being the PHC1 core, which was identified by Tango (Table S2). The underlying reason explaining why these amyloid predictors fail to score properly the putative cores is likely their much lower hydrophobicity, when compared with the amyloid stretches present in pathogenic proteins like A β 42 or α -synuclein (Table S3). Indeed, the predicted cores are enriched in polar residues like Gln, Asn and Ser as well as in Gly (Table 2), all amino acids considered conferring low aggregation propensities to protein sequences.

An analysis of the structural context in which these soft amyloid cores are embedded in their respective sequences using the prediction algorithms FoldIndex⁵⁷, IUPRED⁵⁸, PondR-FIT⁵⁹ and RONN⁶⁰ indicates that they are preferentially located in disordered protein segments (Table S4).

Predicted human PrLD soft amyloid cores assemble into β -sheet rich structures. We synthesized 21-residues long peptides correspondent to the detected amyloid cores and analyzed their secondary structure content by far-UV circular dichroism (CD) immediately after their dilution at 100 μ M in sodium phosphate buffer at pH 7.4 (Figs 1B to 7B). Five out of the seven peptides (DDX5, EYA1, ILF3, MED15 and TIA1) exhibited spectra consistent with a mostly disordered conformation. In ILF3, the high content in Tyr residues (23%) renders a characteristic aromatic signal at 230 nm. The other two peptides (NCOA2 and PHC1) already exhibited a β -sheet CD spectrum immediately after dilution, thus suggesting that they experience a very fast assembly in aqueous buffer.

Next, we incubated the peptides at 100 μ M for 2 days at 37 $^{\circ}$ C and monitored their ability to form macromolecular structures using synchronous light scattering, bis-ANS binding, infrared spectroscopy (ATR-FTIR) and far-UV CD.

The formation of high-order assemblies after incubation was confirmed for all peptides by measuring the light scattering of the correspondent solutions. All of them exhibited significant scattering signal after 2 days (Figure S1). Despite the amyloid cores in this study are less hydrophobic than those of pathogenic amyloids (Table S3), they still exhibit non-polar residues that might contribute to the initial assembly. We explored the presence of exposed hydrophobic clusters in the detected aggregated material by measuring their binding to bis-ANS (Figure S2A), a dye that increases its fluorescence emission upon interaction with these regions⁶¹. The bis-ANS fluorescence emission maximum increases and blueshifts from 530 nm in the absence of peptides to 510 nm in their presence, indicating the existence of hydrophobic-patches in the surface of all these assemblies. The binding of these assemblies to bis-ANS was, however, much lower than those of the amyloid fibrils formed by the Parkinson's associated α -synuclein protein at the same concentration (Figure S2B), consistent with the lower hydrophobicity of PrLDs amyloid stretches.

DDX5

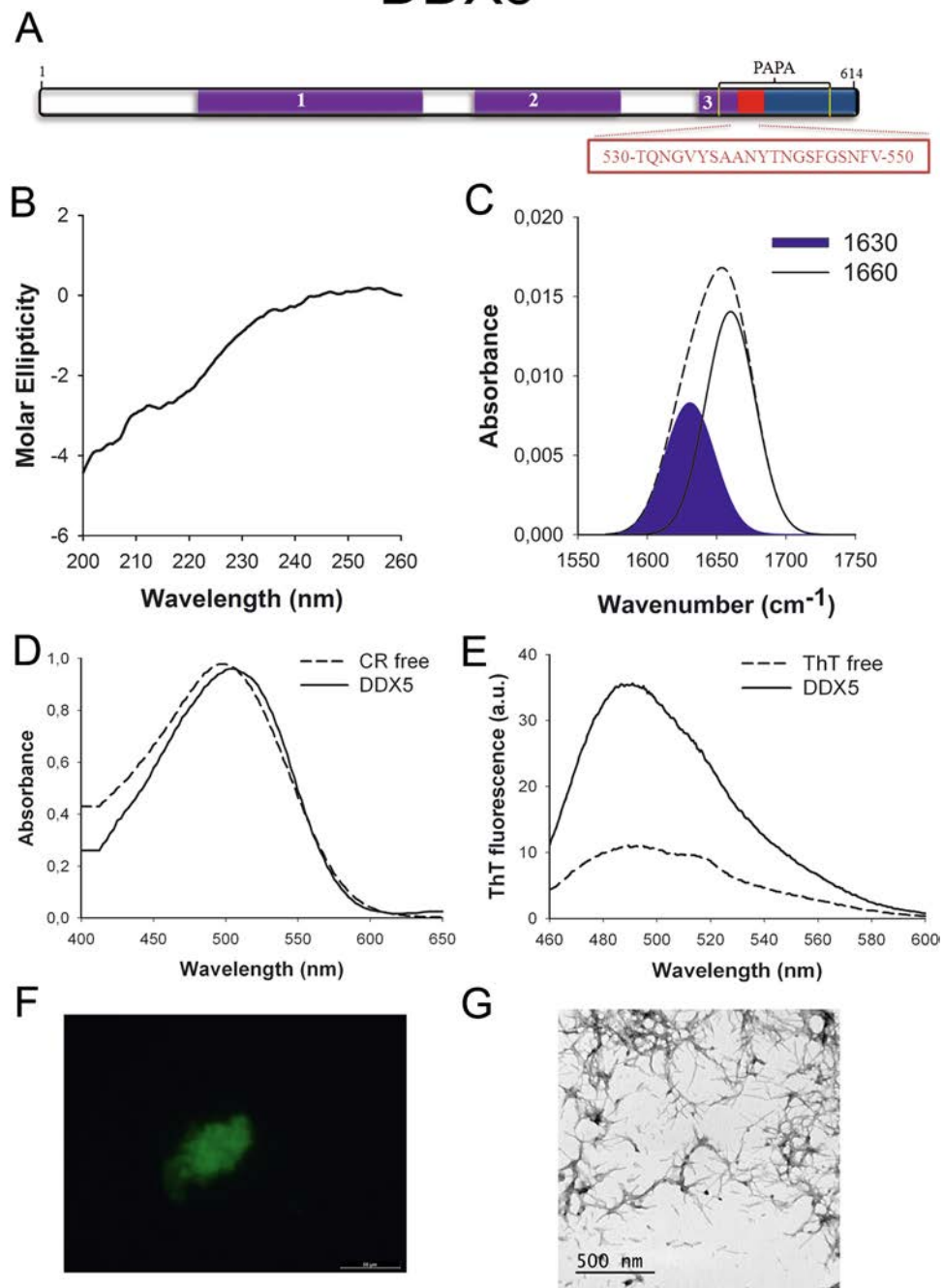


Figure 1. DDX5 PrLD amyloid core. (A) DDX5 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = DEAD domain. 2 = Helicase C domain. 3 = p68-like RNA helicase domain. The sequence of the amyloid core is shown in the box. (B) CD spectrum in the far-UV region of 100 μ M DDX5 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (C) DDX5 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β -sheet signal to the total area upon Gaussian deconvolution. (D) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of DDX5 peptide. (E) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of DDX5 peptide. (F) DDX5 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (G) DDX5 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of DDX5 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 $^{\circ}$ C.

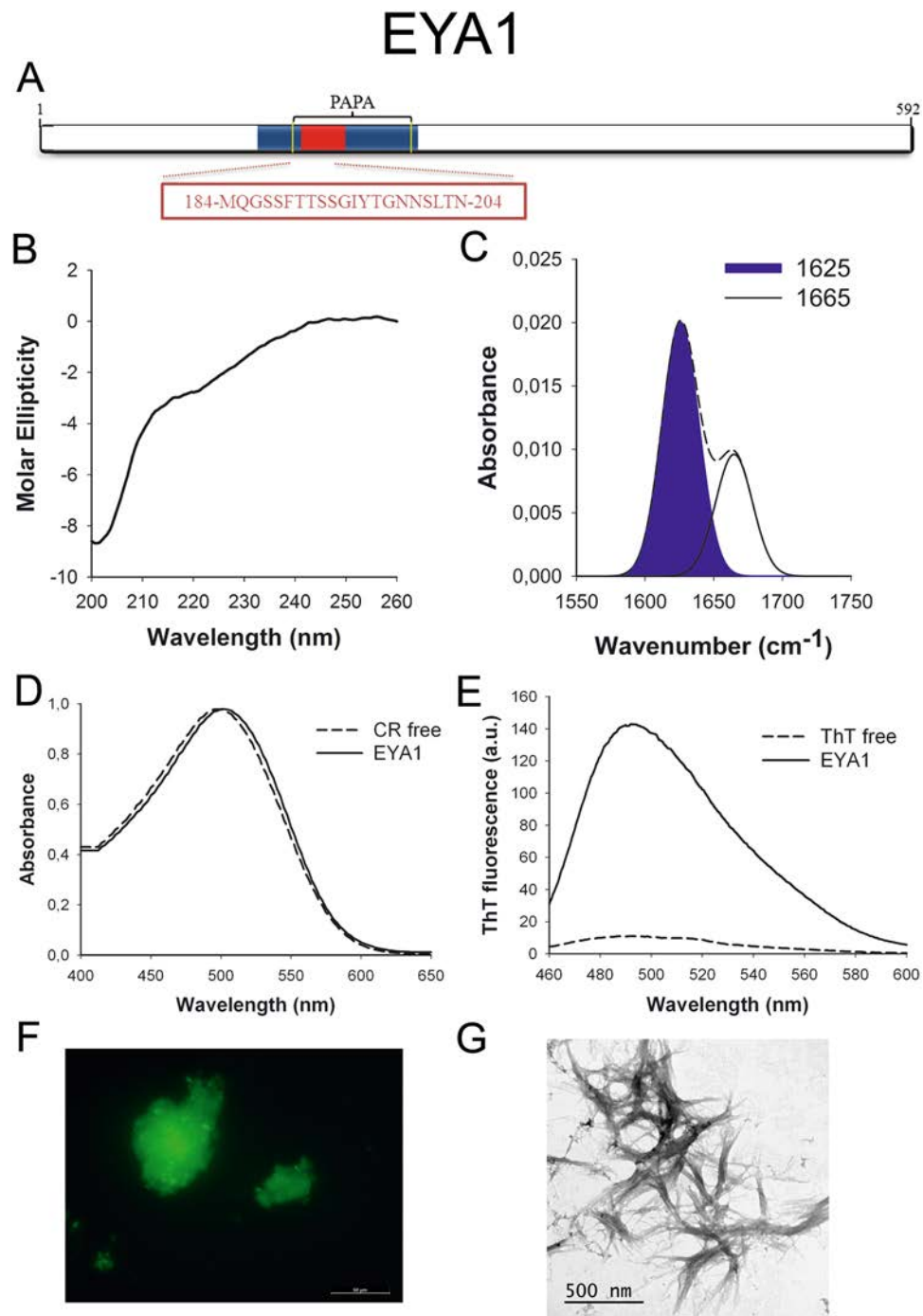


Figure 2. EYA1 PrLD amyloid core. **(A)** EYA1 diagram showing the location of the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). The sequence of the amyloid core is shown in the box. **(B)** CD spectrum in the far-UV region of 100 μM EYA1 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. **(C)** EYA1 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β -sheet signal to the total area upon Gaussian deconvolution. **(D)** CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of EYA1 peptide. **(E)** Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of EYA1 peptide. **(F)** EYA1 peptide stained with Th-Sand observed at 40X magnification using fluorescence microscopy. **(G)** EYA1 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of EYA1 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 $^{\circ}\text{C}$.

ILF3

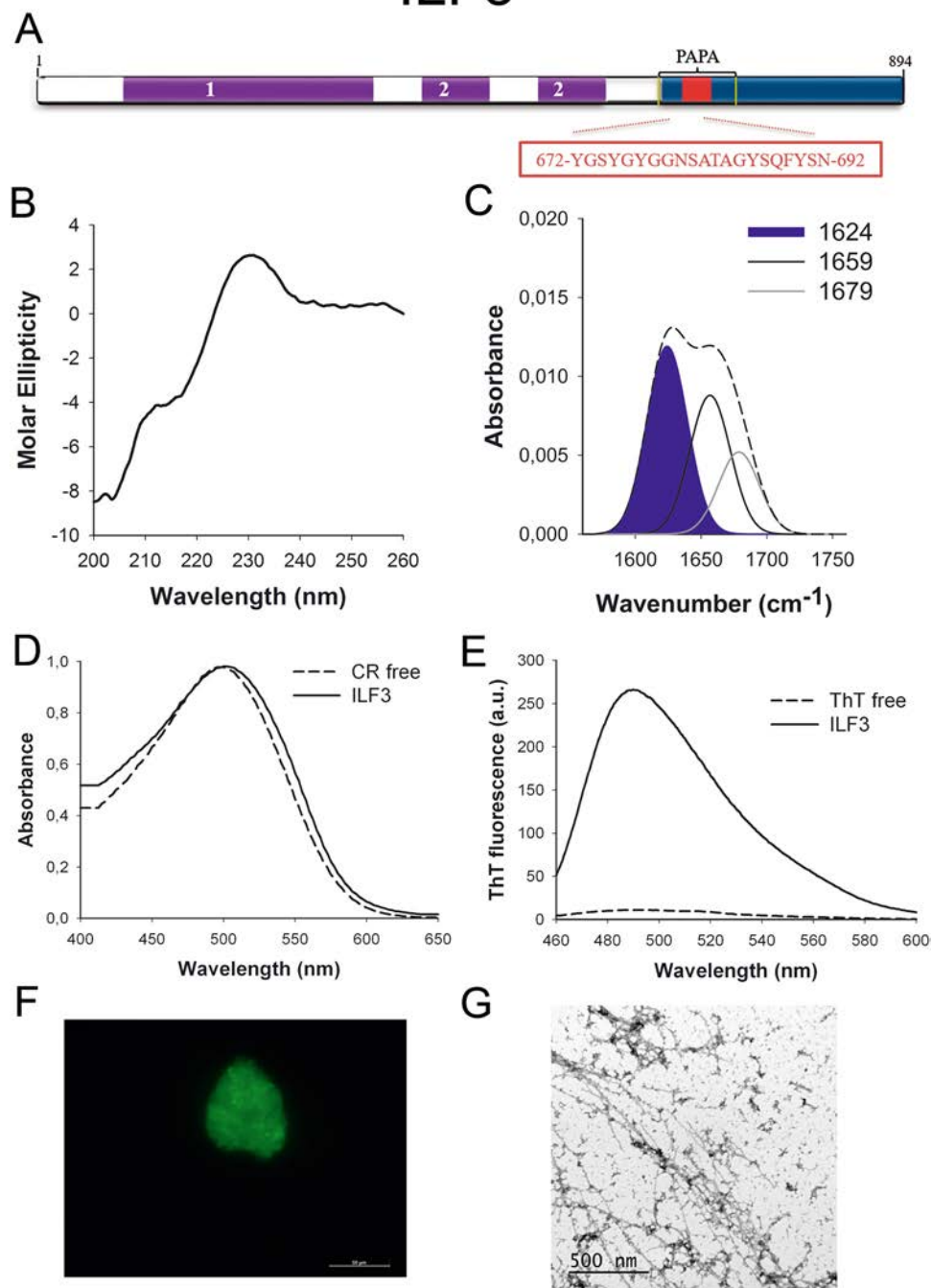


Figure 3. ILF3 PrLD amyloid core. (A) ILF3 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = DZF domain. 2 = dsRNA binding motif. The sequence of the amyloid core is shown in the box. (B) CD spectrum in the far-UV region of $100\mu\text{M}$ ILF3 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (C) ILF3 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β -sheet signal to the total area upon Gaussian deconvolution. (D) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of ILF3 peptide. (E) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of ILF3 peptide. (F) ILF3 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (G) ILF3 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of ILF3 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37°C .

MED15

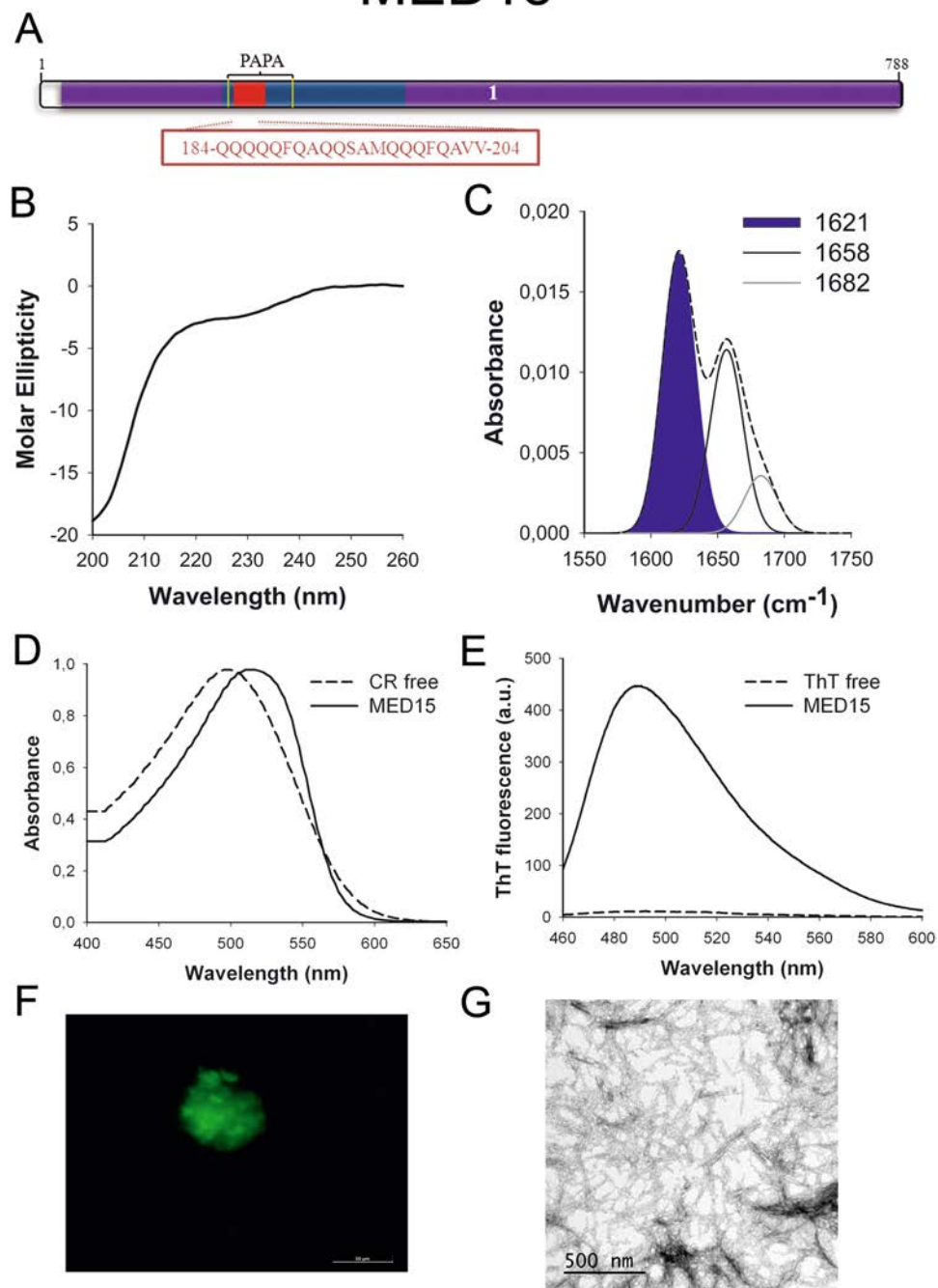


Figure 4. MED15 PrLD amyloid core. (A) MED15 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = MED15 domain. The sequence of the amyloid core is shown in the box. (B) CD spectrum in the far-UV region of 100 μ M MED15 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (C) MED15 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β -sheet signal to the total area upon Gaussian deconvolution. (D) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of MED15 peptide. (E) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of MED15 peptide. (F) MED15 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (G) MED15 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of MED15 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 $^{\circ}$ C.

NCOA2

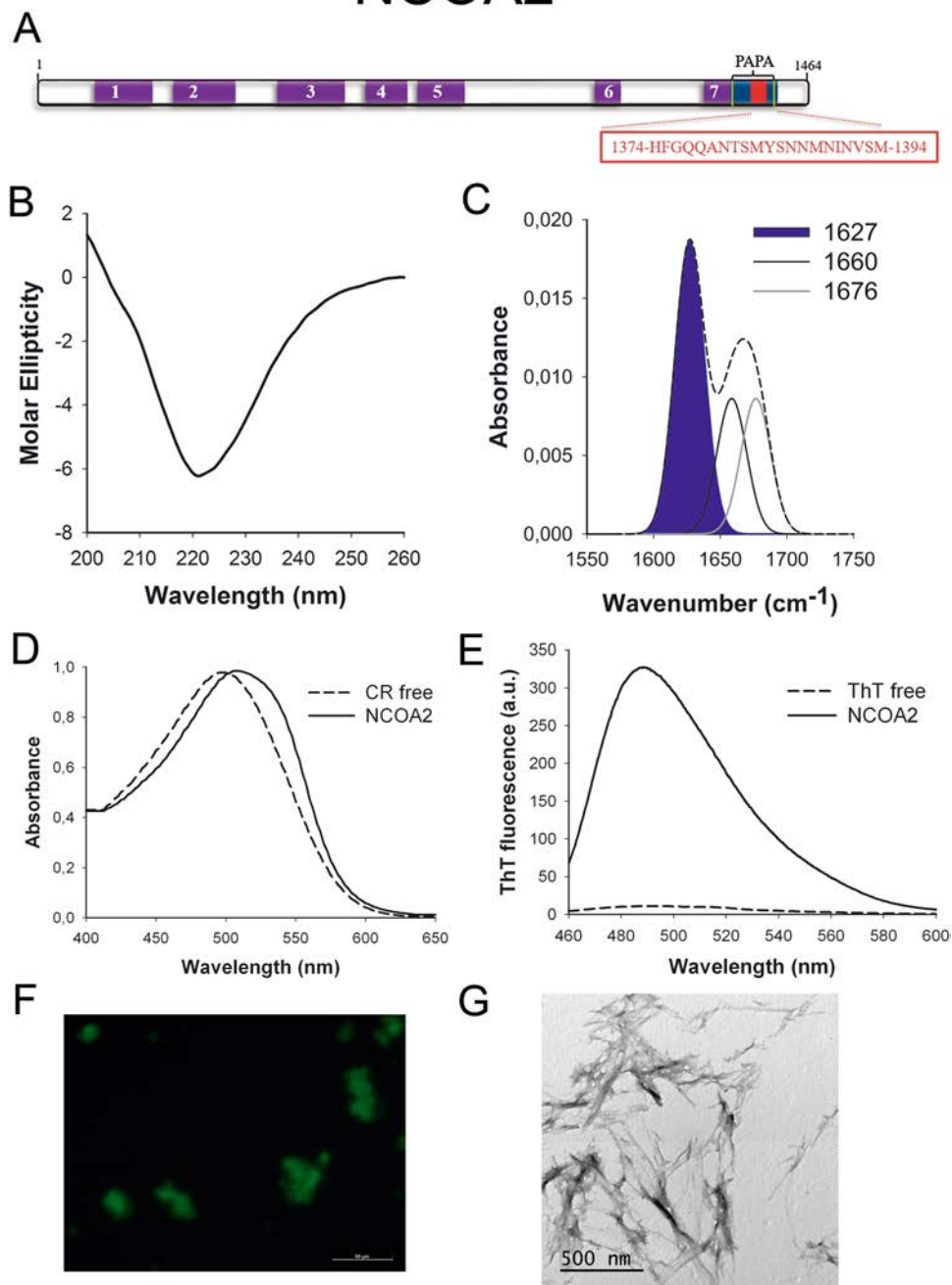


Figure 5. NCOA2 PrLD amyloid core. (A) NCOA2 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = PAS domain. 2 = PAS11 domain. 3 = NCOA_u2 domain. 4 = SRC1 domain. 5 = Duf4927 domain. 6 = Nucleocytosolic co-act domain. 7 = DUF1518. The sequence of the amyloid core is shown in the box. (B) CD spectrum in the far-UV region of 100 μ M NCOA2 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (C) NCOA2 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β -sheet signal to the total area upon Gaussian deconvolution. (D) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of NCOA2 peptide. (E) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of NCOA2 peptide. (F) NCOA2 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (G) NCOA2 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of NCOA2 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 $^{\circ}$ C.

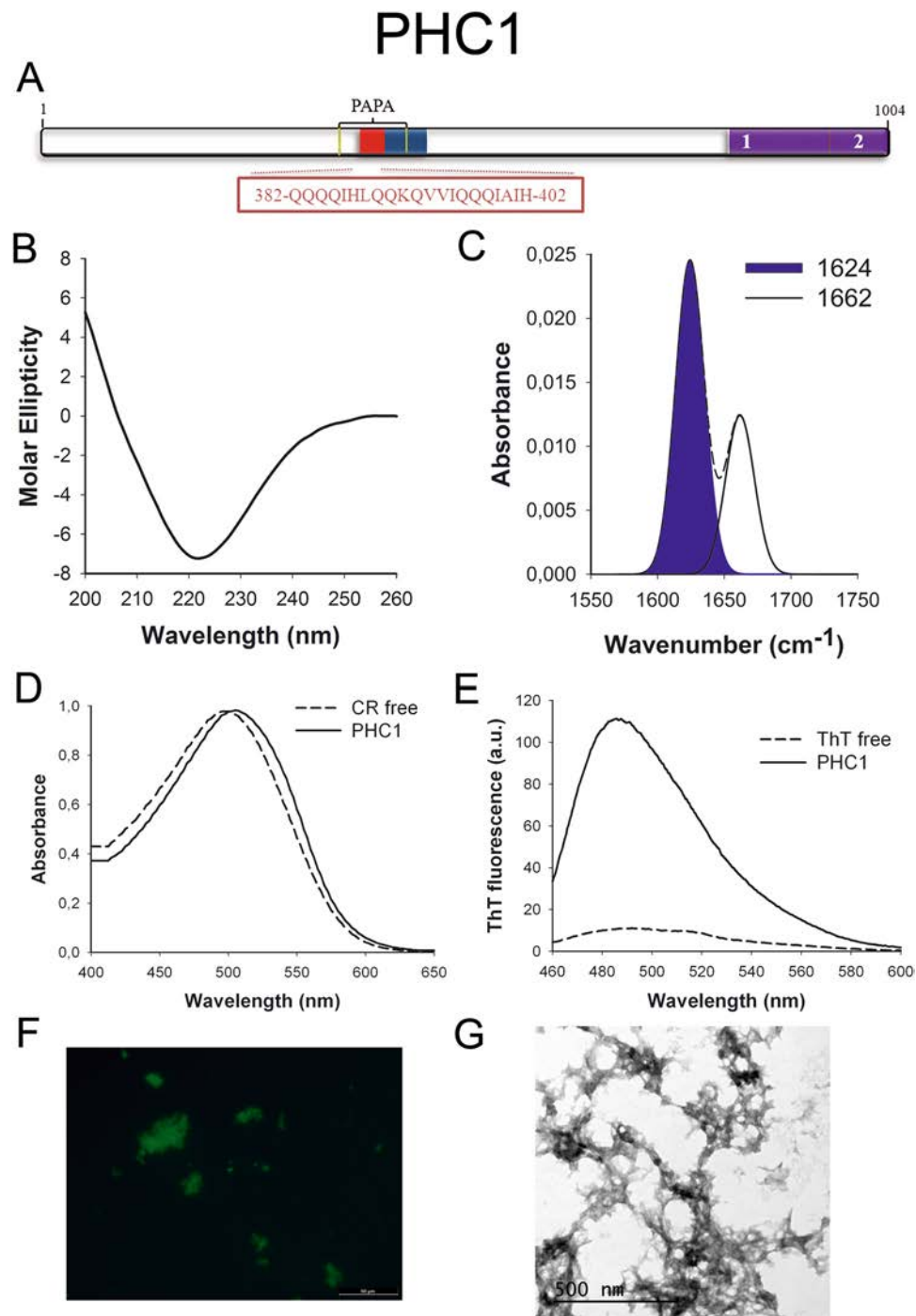


Figure 6. PHC1 PrLD amyloid core. (A) PHC1 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = PHC2 SAM assoc domain. 2 = SAM1 domain. The sequence of the amyloid core is shown in the box. (B) CD spectrum in the far-UV region of 100 μ M PHC1 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (C) PHC1 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β -sheet signal to the total area upon Gaussian deconvolution. (D) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of PHC1 peptide. (E) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of PHC1 peptide. (F) PHC1 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (G) PHC1 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of PHC1 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 $^{\circ}$ C.

TIA1

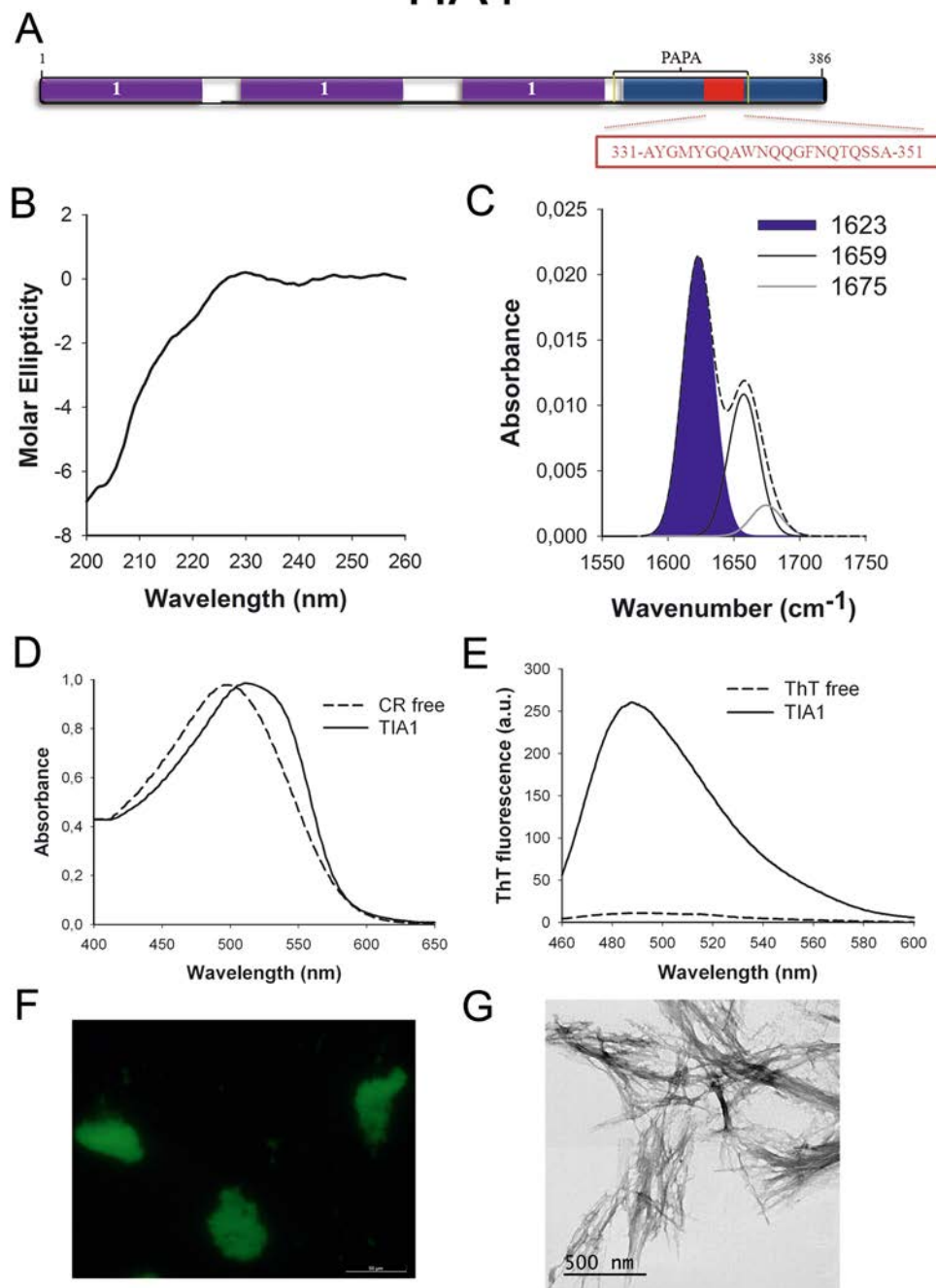


Figure 7. TIA1 PrLD amyloid core. (A) TIA1 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = RNA recognition motif (RRM). The sequence of the amyloid core is shown in the box. (B) CD spectrum in the far-UV region of 100 μ M TIA1 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (C) TIA1 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β -sheet signal to the total area upon Gaussian deconvolution. (D) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of TIA1 peptide. (E) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of TIA1 peptide. (F) TIA1 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (G) TIA1 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of TIA1 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37°C.

Next, we recorded the amide I region of the FTIR spectrum (1700–1600 cm^{-1}) for these aggregates (Figs 1C to 7C). This region corresponds to the absorption of the carbonyl peptide bond group of the protein main chain and it is conformation sensitive. Deconvolution of the spectra allowed us to assign the individual secondary structure elements of incubated peptides and their relative contribution to the main absorbance (Table S5). In all the cases we could identify a strong band at 1620–1630 cm^{-1} , usually assigned to the presence of inter-molecular β -sheets. This signal is the largest contributor to the absorbance spectrum in all peptides, except for DDX5, where it contributes 37% of the area. Interestingly, no anti-parallel β -sheet band was detected ($\sim 1690 \text{ cm}^{-1}$) in any of the samples; thus suggesting that the detected β -strands in self-assembled peptides would adopt preferentially a parallel disposition. The other detected structural elements are associated with disordered structure and turns (Table S5). We also monitored the secondary structure of the incubated peptides using far-UV CD (Figure S3). In all cases we could detect a band at 215–220 nm consistent with the population of a β -sheet enriched conformation, despite in some cases the ellipticity was low, indicating that a significant proportion of the peptide was aggregated, therefore out of the solution and not detectable.

Overall, our data are consistent with the spontaneous assembly of the predicted human PrLD amyloid cores into supramolecular β -sheet enriched structures.

Predicted human PrLD amyloid cores form non-toxic amyloid-like fibrillar structures. We used the amyloid-specific dyes Congo red (CR), Thi flavin T (Th-T) and Thi flavin-S (Th-S) to confirm that the detected β -sheet enriched aggregates were organized into amyloid-like suprastructures.

The absorbance spectra of CR red shifts in the presence of amyloid aggregates⁶². Incubation of CR in the presence of aggregated peptides resulted in a red shift of its spectrum in all the cases (Figs 1D to 7D); despite for DDX5, EYA1 and ILF3 the spectral shift was small. To confirm the ability of these three peptides to bind CR they were incubated under the same conditions at 500 μM final concentration. The peptides in these three solutions promoted a clear shift of the absorption maximum of the dye (Figure S4).

Th-T fluorescence emission is enhanced in the presence of amyloid fibrils⁶³. All the peptides promoted an increase in the intensity of Th-T fluorescence spectral maximum at 488 nm (Figs 1E to 7E). Furthermore, binding of Th-S to the aggregates could be visualized by fluorescence microscopy for all incubated peptides. Areas rich in fibrous material were stained with Th-S to yield green-yellow fluorescence against a dark background in all cases (Figs 1F to 7F).

The dye binding results indicate that incubated peptide solutions might contain detectable amounts of amyloid-like structures. To confirm this extent, the morphological features of the peptide assemblies in these samples were analyzed using transmission electron microscopy (TEM). Negative staining indicated that all peptides effectively assemble into supramolecular structures (Figs 1G to 7G). The aggregates formed by DDX5, EYA1, MED15, NCOA2, and TIA1 correspond to amyloid-like fibrillar arrangements, without any significant accumulation of amorphous material. The fibrils exhibit a diameter that varies from 5 to 10 nm and a length that ranges from 2 to 10 μm . In the ILF3 sample long amyloid-like fibrils coexist with small aggregates that appear to attach to the fibrils, whereas, for PHC1, despite its amyloid-like tintorial properties, the material appears to be essentially protofibrillar. All the peptides were also able to form macromolecular aggregates when they were incubated at 10 μM (1/10 the concentration used in the previous assays) (Figure S5).

The above data indicate that the predicted amyloid cores exhibit a strong propensity to form amyloid-like assemblies. The amyloids formed by pathogenic protein fragments are usually highly cytotoxic. We tested if the aggregates formed by the PrLD amyloid cores display any toxicity when administered to neuroblastoma SH-SY5Y cells. All the aggregates were essentially innocuous when added to the cell cultures at up to 10 μM (Figures S6).

Predicted human PrLD amyloid cores form aggregates with self-seeding properties. Seeded protein aggregation is a well-established mechanism for *in vivo* amyloid fibril formation and underlies prion propagation⁶⁴. The nucleation step of the amyloid assembly is shortened in the presence of preformed amyloid fibrils of the same protein, that can act as nuclei for the subsequent polymerization reaction⁶⁵. Specific and short aggregation-prone regions have been shown to play a crucial role in this process^{66,67}. To test whether preformed PrLDs core amyloid-like assemblies can seed the aggregation of the correspondent soluble peptides, we followed the aggregation kinetics of the peptides at 100 μM in the presence and absence of 2% (w/w) of preformed aggregates (Fig. 8). We could not monitor the aggregation kinetics of NCOA2 or PHC1 because they exhibited very high Th-T signal from the very beginning of the reaction, consistent with a very fast assembly into β -sheet structures, as suggested by the far-UV CD spectra they exhibit immediately upon dilution in aqueous solution (Figs 5B and 6B). The rest of peptides exhibited characteristic sigmoidal aggregation kinetics with lag-phases ranging from 20 to 120 min in the absence of seeds. The addition of preformed aggregates strongly accelerated the formation of Th-T positive assemblies in all cases (Fig. 8), supporting a nuclei-dependent aggregation mechanism and raising the possibility that such specific interactions could also occur in the context of the complete proteins in which these short regions are embedded.

Discussion

The prion phenomenon is best known by its association with spongiform encephalopathies in mammals, but in the last years a growing list of non-pathological prion-like proteins are being discovered. Functional prions were originally identified in yeast⁶⁸. In these prion proteins, specific PFDs encode for their ability to switch between soluble and self-assembled states. Computational analysis are revealing the existence of polypeptides displaying similar domains in previously unexplored proteomes¹⁷. This suggests that this type of conformational conversion might be an evolutionary conserved mechanism exploited by different organisms, including humans, for beneficial purposes.

A characteristic feature of many pathogenic amyloids is the presence of short hydrophobic sequence stretches able to nucleate the assembly of these proteins into toxic aggregates both *in vitro* and *in vivo* under pathological conditions. These highly amyloidogenic regions seem to be absent in PFDs and PrLDs, likely because their

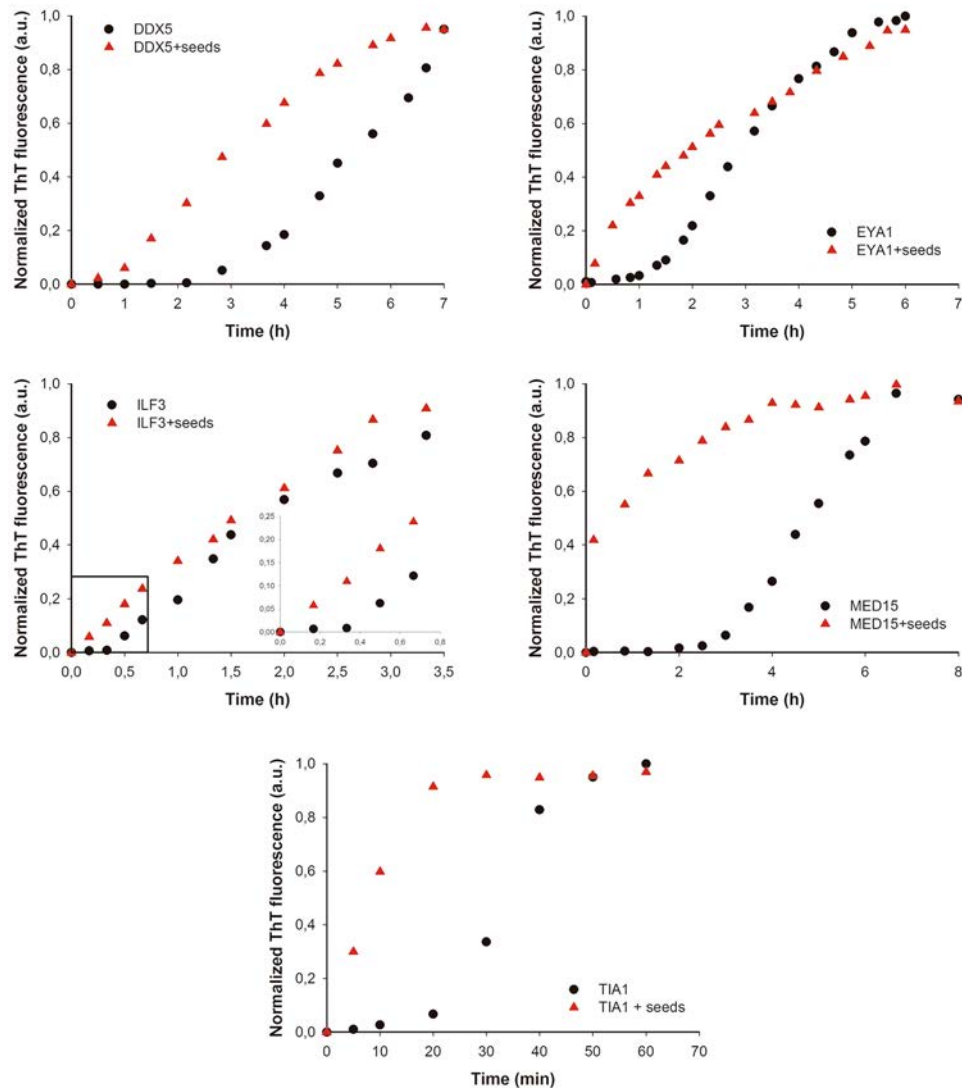


Figure 8. Human PrLD amyloid cores kinetics. Aggregation kinetics of DDX5, EYA1, ILF3, MED15 and TIA1 peptides at 100 μ M in the absence (black circles) and presence (red triangles) of 2% (w/w) of preformed aggregates (seeds) were monitored by tracking the changes in Th-T fluorescence emission spectra. All the peptides exhibited accelerated kinetics in the presence of seeds.

potency will unbalance the equilibrium between the soluble and self-assembled states of the proteins towards an irreversibly aggregated and potentially toxic state, even in the absence of stress. However, we have recently proposed that in addition to a special compositional bias, PFDs and PrLDs contain indeed cryptic soft amyloid cores that can play an important role at the early stages of assembly by restricting inter-molecular interactions to specific regions of these long domains²⁸. In these cores, the amyloid nucleating potency would be weaker and less concentrated than in pathogenic amyloids. This property would allow the protein to remain soluble in most physiological conditions, while being responsive to conditions that favor inter-molecular contacts or to the presence of preformed assemblies that target these specific segments. We have provided evidences for the presence of such regions in the PFDs of four of the best characterized yeast prions and for the ability of one of these soft amyloid cores to recruit the assembly of its correspondent full-length PFD *in vitro* and in the cell³⁰.

Using an approach analogous to the one described here, we previously screened the *C. Botulinum* proteome for the presence of regions displaying both compositional similitude to *bona fide* prions and containing a soft amyloid core, identifying a first bacterial PrLD in the Rho Terminator factor of this pathogen³¹. Later on, Yuan and Hochschild validated the prion-like nature of this protein, demonstrating that Rho can access alternative protein conformations in prokaryotes, including a self-assembled state with decreased Rho activity that results in genome-wide changes at the transcriptome level³². The seven human proteins selected in the present study for experimental characterization, exhibit predicted PrLDs with the same compositional properties than Rho or the yeast prions and are as well regulatory nucleic acid-binding proteins. We show here that they all contain a region able to autonomously self-assemble into amyloid-like structures displaying low cytotoxicity, a property that is likely linked to the weak hydrophobic patches in these assemblies, compared with those present in the

fibrils of pathogenic proteins like α -synuclein¹. In addition, for all the cases we could test, the fibrillar assemblies had the capacity to seed and accelerate the aggregation reaction of their soluble counterparts. Despite our assays are *in vitro* and just with a piece of the protein, the results reveal the self-assembly potential encoded in human prion-like proteins.

Similar to Short Linear Interaction Motifs (SLIMs) in intrinsically disordered proteins, the assembling properties of soft amyloid cores might contribute to mediate PrLDs functional protein-protein interactions (PPI)²⁸. Actually, it is known that the interaction between PrLDs of RNA granules proteins like TDP43, TAF15 and FUS is sufficient to induce liquid-liquid phase separation⁷, but also reversible hydrogels⁶⁹, which have a high β -sheet content, but are labile to dilution. Interestingly enough, mutations that severely impede the formation of intermolecular β -sheets also inhibit the ability of PrLDs to activate transcription, indicating that, in this particular context, the formation of soft amyloid-like assemblies plays a functional role⁷⁰. DDX5, ILF3 and TIA1 are also RNA granule proteins^{43,71-73}; thus it is tempting to propose that the identified PrLDs and embedded soft amyloid cores might likewise play a role in the formation of functional β -sheet-containing assemblies. The other four proteins in our study are either structural components of macromolecular complexes (MED15 and PHC1) or they are involved in protein-protein interactions (NCOA2 and EYA1). The detected PrLDs and the sticky nature of their cores might facilitate these contacts. It has been suggested that the role of PrLDs would be to promote a primary level of organization through homotypic interactions and that, once oligomerized, these assemblies would facilitate the establishment of novel lateral interactions with other proteins⁷⁴. Indeed, an analysis of the interaction networks of the seven proteins in this study using the STRING PPI network database⁷⁵ indicates that, as a trend, they tend to establish more PPI than the average human proteome (Table S6).

From an evolutionary point of view, the fact that the detected amyloid cores have been not purged out by natural selection support their functional role, since there is a strong selective pressure to reduce the amyloidogenic load of protein sequences, especially when they are located in disordered regions and thus exposed to solvent⁷⁶. The presence of this kind of regions is inherently risky, since mutations that would increase their amyloid potency, making them more similar to classical pathogenic amyloid stretches, can promote irreversible protein aggregation, generating a deleterious phenotype resulting from deregulation of the homo or heterotypic PPIs in which they are involved. This is the case of hnRNPA1 and hnRNPA2, two human prion-like RNA-binding proteins in which point mutations mapping exactly in the pWALTZ predicted soft amyloid core increase the aggregation propensity of their PrLDs, causing the loss of their regulatory activity and leading to the onset of multisystem proteinopathy and ALS^{14,27}. It is worth to explore whether the link to disease of the nucleic acid-binding human proteins in our subproteome and specifically of the seven polypeptides we studied here owes to a related mechanism. This would imply that a common process of mutation driven miss-assembly of cryptic amyloid cores in PrLDs might account for very different pathological phenotypes, from cancer to neurodegenerative disorders, depending on the affected protein and the pathways it regulates.

Methods

Computational identification of PrLDs in human proteins. The human reference proteome dataset was downloaded from Uniprot⁷⁷ and scanned for PrLDs using PAPA¹² with the default parameters. From the initial 70940 proteins in the proteome, 663 prion-like candidates were identified. Their putative PrLDs were further evaluated with pWALTZ¹⁴ with a threshold of 65 in order to identify those domains containing a soft amyloid core, which resulted in 535 final positive predictions. The protein sequences selected for experimental characterization were also analyzed with PLAAC¹⁶ to define the boundaries of their PrLDs.

Peptide preparation. We obtained the sequence of the 21 amino acid core region as predicted by pWALTZ¹⁴ for the seven protein candidates selected for experimental characterization. The correspondent peptides were purchased from CASLO ApS (Scion Denmark Technical University). The lyophilized peptides were solubilized at a final concentration of 5 mM in dimethyl sulfoxide (DMSO) or in hexafluoro-2-propanol for CD analysis, in order to avoid the large increase in voltage caused by residual DMSO. Right before each experiment, the stock solutions were diluted to 100 μ M in 5 mM sodium phosphate buffer pH 7.4. For aggregation assays the samples were incubated for 2 days at 37 °C with continuous agitation at 150 rpm in the presence of Teflon beads.

Synchronous light scattering. Synchronous light scattering was monitored using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 360 nm, emission range from 350 to 370 nm, slit widths of 5 nm, 0.5 nm interval and 1000 nm/min scan rate. The peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before measurement. 100 μ l of peptide solution was analyzed.

Bis-ANS (4,4-Dianilino-1,1-binaphthyl-5,5-disulfonate) binding. The fluorescent spectrum of bis-ANS was analyzed using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 365 nm, emission range from 440 to 640 nm, slit widths of 5 nm, 0.5 nm interval and 1000 nm/min scan rate. The peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before dye addition. 10 μ l of peptide solution was added to 100 μ l of 10 μ M bis-ANS in H₂O. A 10 μ M bis-ANS solution without peptide was used as a control.

Circular dichroism (CD) spectroscopy. CD experiments were performed using a JASCO J-715 spectropolarimeter. Measurements of the far-UV CD spectra (260–190 nm) were made by the addition of 200 μ l of the sample to a cuvette of 0.1 cm path-length. Spectra were recorded at room temperature, 1 nm band width and 100 nm/min scan rate. The resulting spectrum was the average of 10 scans. The contribution of the buffer was subtracted.

Fourier transform infrared (FT-IR) spectroscopy. FTIR experiments were performed using a Bruker Tensor 27 FT-IR spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Each spectrum consists of 16 independent scans, measured at a spectral resolution of 4 cm^{-1} within the $1800\text{--}1500\text{ cm}^{-1}$ range. All spectral data were acquired and normalized using the OPUS MIR Tensor 27 software. Data was afterwards deconvoluted using the Peak Fit 4.12 program. The buffer without peptide was used as a control and subtracted from the absorbance signal before deconvolution.

Transmission electron microscopy (TEM). The morphology of the aggregated peptides was evaluated by negative staining and using a JEOL JEM-1400Plus Transmission Electron Microscope. $5\ \mu\text{l}$ of peptide solution was placed on carbon-coated copper grids and incubated for 5 min. The grids were then washed and stained with $5\ \mu\text{l}$ of 2% w/v uranyl acetate for 5 min. Then, grids were washed again before analysis.

Congo red (CR) binding. CR binding to aggregated peptides was analyzed using a Specord® 200 Plus spectrophotometer (Analyticjena). The absorbance spectra were recorded from 400 to 650 nm. Spectra were acquired at 50 nm/sec scan rate. Peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before dye addition. $10\ \mu\text{l}$ of the sonicated aggregated peptide was added to $100\ \mu\text{l}$ of $5\ \mu\text{M}$ CR in 5 mM sodium phosphate buffer pH 7.4, and was incubated at room temperature for 5 min before the measurement. The same buffer with $5\ \mu\text{M}$ CR and without peptide was employed as a control.

Thioflavin-T (Th-T) binding. The fluorescence spectra of Th-T were recorded using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 440 nm, emission range from 460 to 600 nm, slit widths of 5 nm, 0.5 nm interval and 1000 nm/min scan rate. Peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before dye addition. $5\ \mu\text{l}$ of the sonicated aggregated peptide was added to $100\ \mu\text{l}$ of $25\ \mu\text{M}$ ThT in 5 mM sodium phosphate buffer pH 7.4. The same buffer with $25\ \mu\text{M}$ ThT and without peptide was employed as a control.

Thioflavin-S (Th-S) staining. First, $150\ \mu\text{l}$ of aggregated peptides were incubated for 1 h in the presence of $125\ \mu\text{M}$ of ThS in 5 mM sodium phosphate buffer pH 7.4. Then, the samples were washed two times with the same buffer. Finally, the precipitated fraction was resuspended in a final volume of $10\ \mu\text{l}$ and placed on a microscope slide and sealed. Images of the peptide aggregates bound to Th-S were obtained at 40-fold magnification in a Leica fluorescence microscope (Leica DMRB).

Aggregation kinetics and seeding assays. Reactions were carried out at of $100\ \mu\text{M}$ final soluble peptide concentration in a solution containing $25\ \mu\text{M}$ of Th-T at $37\ ^\circ\text{C}$ in the absence or presence of 2% seeds under quiescent conditions. The aggregation kinetics were followed monitoring the changes in Th-T fluorescence intensity at 488 nm over the time using a JASCO Spectrofluorometer FP-8200. Before each measure, the sample was mixed by pipetting up and down. The conditions of the spectra acquisition were: excitation wavelength of 440 nm, emission range from 460 to 600 nm, slit widths of 5 nm, 0.5 nm interval and 1000 nm/min scan rate. The seeds were prepared by sonicating the preformed aggregates of the corresponding peptide for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before addition.

Cell viability assay. Human SH-SY5Y cells were seeded into 96-well tissue culture plate with a density of 4,000 cells/well ($100\ \mu\text{L}/\text{well}$) in F-12 medium supplemented with 10% FBS, and maintained at $37\ ^\circ\text{C}$ and 5% CO_2 atmosphere. Cell cultures were incubated in the presence of different concentrations of peptides for 72 hours. To control cells, the same volume of PBS1x was added. Following incubation, cells were stained by adding $20\ \mu\text{L}$ PrestoBlue® Cell Viability Reagent (Invitrogen) directly to the sample wells. After 30 min of incubation, cell viability was determined by measuring fluorescence exciting at 531 nm and collecting emission at 615 nm in a Victor fluorescent plate reader (Perkin Elmer).

References

- Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry* **75**, 333–366 (2006).
- Ventura, S. *et al.* Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc. Natl. Acad. Sci. USA* **101**, 7258–7263 (2004).
- Rousseau, F., Serrano, L. & Schymkowitz, J. W. How evolutionary pressure against protein aggregation shaped chaperone specific ty. *J Mol Biol* **355**, 1037–1047 (2006).
- Aguzzi, A. & Calella, A. M. Prions: protein aggregation and infectious diseases. *Physiological reviews* **89**, 1105–1152 (2009).
- Sikorska, B. & Liberski, P. P. Human prion diseases: from Kuru to variant Creutzfeldt-Jakob disease. *Sub-cellular biochemistry* **65**, 457–496 (2012).
- Si, K. P. What Are They Good For? Annual Review of Cell and Developmental Biology **31**, annurev-cellbio-100913-013409 (2015).
- Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
- Chakrabortee, S. *et al.* Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits Article Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. *Cell* **167**, 1–13 (2016).
- Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* **137**, 146–158 (2009).
- Michelitsch, M. D. & Weissman, J. S. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11910–11915 (2000).
- Harrison, P. M. & Gerstein, M. A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. *Genome biology* **4**, R40 (2003).
- Toombs, J. a. *et al.* De novo design of synthetic prion domains. *Proceedings of the National Academy of Sciences* **109**, 6519–6524 (2012).

13. Espinosa Angarica, V. *et al.* PrionScan: an online database of predicted prion domains in complete proteomes. *BMC genomics* **15**, 102 (2014).
14. Sabate, R., Rousseau, F., Schymkowitz, J. & Ventura, S. What Makes a Protein Sequence a Prion? *PLoS Computational Biology* **11**, e1004013 (2015).
15. Zambrano, R. *et al.* PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Research*, 1–7, (2015).
16. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502 (2014).
17. Batlle, C., Iglesias, V., Navarro, S. & Ventura, S. Prion-like proteins and their computational identification in proteomes. *Expert Rev Proteomics* **14**, 335–350 (2017).
18. King, O. D., Gitler, A. D. & Shorter, J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Research* **1462**, 61–80 (2012).
19. An, L. & Harrison, P. M. The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biology direct* **11**, 32 (2016).
20. Malinowska, L., Palm, S., Gibson, K., Verbavatz, J.-M. & Alberti, S. Dictyostelium discoideum has a highly Q/N-rich proteome and shows an unusual resilience to protein aggregation. *Proceedings of the National Academy of Sciences* **112**, 201504459 (2015).
21. Singh, G. P. *et al.* Hyper-expansion of asparagines correlates with an abundance of proteins with prion-like domains in *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **137**, 307–319 (2004).
22. Espinosa Angarica, V., Ventura, S. & Sancho, J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC genomics* **14**, 1–17 (2013).
23. Chakrabortee, S. *et al.* Luminidependens (LD) is an Arabidopsis protein with prion behavior. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 201604478 (2016).
24. Kato, M. *et al.* Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
25. Decker, C. J. & Parker, R. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harbor perspectives in biology* **4**, a012286 (2012).
26. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).
27. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
28. Fernandez, M. R., Batlle, C., Gil-Garcia, M. & Ventura, S. Amyloid cores in prion domains: Key regulators for prion conformational conversion. *Prion* **11**, 31–39 (2017).
29. Sabate, R., Rousseau, F., Schymkowitz, J., Batlle, C. & Ventura, S. Amyloids or prions? That is the question. *Prion* **9**, 200–206 (2015).
30. Sant'Anna, R. *et al.* Characterization of Amyloid Cores in Prion Domains. *Scientific Reports* **6**, 34274 (2016).
31. Pallarès, I., Iglesias, V. & Ventura, S. The rho termination factor of *Clostridium botulinum* contains a prion-like domain with a highly amyloidogenic core. *Frontiers in Microbiology* **6**, 1–12 (2016).
32. Yuan, A. H. & Hochschild, A. A bacterial global regulator forms a prion. *Science* **355**, 198–201 (2017).
33. Li, X., Rayman, J. B., Kandel, E. R. & Derkatch, I. L. Functional Role of Tia1/Pub1 and Sup35 Prion Domains: Directing Protein Synthesis Machinery to the Tubulin Cytoskeleton. *Molecular Cell* **55**, 305–318 (2014).
34. Fuller-Pace, F. V. The DEAD box proteins DDX5 (p68) and DDX17 (p72): Multi-tasking transcriptional regulators. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1829**, 756–763 (2013).
35. Janknecht, R. Multi-talented dead-box proteins and potential tumor promoters: P68 RNA helicase (DDx5) and its paralog, p72 RNA helicase (DDx17). *American Journal of Translational Research* **2**, 223–234 (2010).
36. Mazurek, A. *et al.* DDX5 regulates DNA replication and is required for cell proliferation in a subset of breast cancer cells. *Cancer discovery* **2**, 812–825 (2012).
37. Fuller-Pace, F. V. The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochim Biophys Acta* **1829**, 756–763 (2013).
38. Xu, J. *et al.* Eya1 interacts with Six2 and Myc to regulate expansion of the nephron progenitor pool during nephrogenesis. *Developmental Cell* **31**, 434–447 (2014).
39. Wu, K. *et al.* EYA1 phosphatase function is essential to drive breast cancer cell proliferation through cyclin D1. *Cancer Res* **73**, 4488–4499 (2013).
40. Buller, C., Xu, X., Marquis, V., Schwanke, R. & Xu, P. X. Molecular effects of Eya1 domain mutations causing organ defects in BOR syndrome. *Human molecular genetics* **10**, 2775–2781 (2001).
41. Xu, P. X., Cheng, J., Epstein, J. A. & Maas, R. L. Mouse Eya genes are expressed during limb tendon development and encode a transcriptional activation function. *Proc Natl Acad Sci USA* **94**, 11974–11979 (1997).
42. Castella, S., Bernard, R., Corno, M., Fradin, A. & Larcher, J. C. Ilf3 and NF90 functions in RNA biology. *Wiley Interdisciplinary Reviews: RNA* **6**, 243–256 (2015).
43. Shiina, N. & Nakayama, K. RNA granule assembly and disassembly modulated by nuclear factor associated with double-stranded RNA 2 and nuclear factor 45. *The Journal of biological chemistry* **289**, 21163–21180 (2014).
44. Saunders, L. R. *et al.* Characterization of two evolutionarily conserved, alternatively spliced nuclear phosphoproteins, NFAR-1 and -2, that function in mRNA processing and interact with the double-stranded RNA-dependent protein kinase, PKR. *The Journal of biological chemistry* **276**, 32300–32312 (2001).
45. Wang, X. *et al.* Redefining the modular organization of the core Mediator complex. *Cell Research* **24**, 796–808 (2014).
46. Halfmann, R. *et al.* Opposing Effects of Glutamine and Asparagine Govern Prion Formation by Intrinsically Disordered Proteins. *Molecular Cell* **43**, 72–84 (2011).
47. Zhu, X. *et al.* Mediator tail subunits can form amyloid-like aggregates *in vivo* and affect stress response in yeast. *Nucleic Acids Research* **43**, 7306–7314 (2015).
48. Xu, J. & Li, Q. Review of the *in vivo* functions of the p160 steroid receptor coactivator family. *Molecular Endocrinology* **17**, 1681–1692 (2003).
49. Silva, M. P. *et al.* NCOA2 is a candidate target gene of 8q gain associated with clinically aggressive prostate cancer. *Genes Chromosomes and Cancer* **55**, 365–374 (2016).
50. Suresh, S. *et al.* SRC-2-mediated coactivation of anti-tumorigenic target genes suppresses MYC-induced liver cancer. *PLoS genetics* **13**, e1006650 (2017).
51. Fleet, T. *et al.* SRC-2 orchestrates polygenic inputs for fine-tuning glucose homeostasis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6068–6077 (2015).
52. Spemann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nature reviews. Cancer* **6**, 846–856 (2006).
53. Awad, S. *et al.* Mutation in PHC1 implicates chromatin remodeling in primary microcephaly pathogenesis. *Human Molecular Genetics* **22**, 2200–2213 (2013).

54. Conchillo-Sole, O. *et al.* AGGRESKAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC bioinformatics* **8**, 65 (2007).
55. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature biotechnology* **22**, 1302–1306 (2004).
56. Tartaglia, G. G. & Vendruscolo, M. The Zyggregator method for predicting protein aggregation propensities. *Chemical Society reviews* **37**, 1395–1401 (2008).
57. Prilusky, J. *et al.* FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438 (2005).
58. 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* **21**, 3433–3434 (2005).
59. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochimica et biophysica acta* **1804**, 996–1010 (2010).
60. 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* **21**, 3369–3376 (2005).
61. de Groot, N. S., Parella, T., Aviles, F. X., Vendrell, J. & Ventura, S. Ile-phe dipeptide self-assembly: clues to amyloid formation. *Biophys. J.* **92**, 1732–1741 (2007).
62. Klunk, W. E., Pettegrew, J. W. & Abraham, D. J. Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. *J. Histochem. Cytochem.* **37**, 1273–1281 (1989).
63. Sabate, R., Rodriguez-Santiago, L., Sodupe, M., Saupe, S. J. & Ventura, S. Thi flavin-T excimer formation upon interaction with amyloid fibers. *Chem. Commun. (Camb.)* **49**, 5745–5747 (2013).
64. Wickner, R. B. *et al.* Yeast prions act as genes composed of self-propagating protein amyloids. *Adv. Protein Chem.* **57**, 313–334 (2001).
65. Jarrett, J. T. & Lansbury, P. T. Jr. Seeding “one-dimensional crystallization” of amyloid: a pathogenic mechanism in Alzheimer’s disease and scrapie? *Cell* **73**, 1055–1058 (1993).
66. Pastor, M. T., Esteras-Chopo, A. & Serrano, L. Hacking the code of amyloid formation: the amyloid stretch hypothesis. *Prion* **1**, 9–14 (2007).
67. Sabate, R., Espargaro, A., Grana-Montes, R., Reverter, D. & Ventura, S. Native structure protects SUMO proteins from aggregation into amyloid fibrils. *Biomacromolecules* **13**, 1916–1926 (2012).
68. Liebman, S. W. & Chernoff, Y. O. in *Genetics* Vol. 191, 1041–1072 (Genetics, 2012).
69. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
70. Kato, M. & Mcknight, S. L. Cross- b Polymerization of Low Complexity Sequence Domains. 1–11, (2016).
71. Gilks, N. *et al.* Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Molecular biology of the cell* **15**, 5383–5398 (2004).
72. Kanai, Y., Dohmae, N. & Hirokawa, N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron* **43**, 513–525 (2004).
73. Elvira, G. *et al.* Characterization of an RNA granule from developing brain. *Molecular & cellular proteomics: MCP* **5**, 635–651 (2006).
74. Kwon, I. *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049–1060 (2013).
75. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–452 (2015).
76. Monsellier, E. & Chiti, F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* **8**, 737–742 (2007).
77. UniProt Consortium, T. U. UniProt: a hub for protein information. *Nucleic acids research* **43**, D204–212 (2015).
78. Rappaport, N. *et al.* MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res* **45**, D877–D887 (2017).

Acknowledgements

This work was funded supported by the Spanish Ministry of Economy and Competitiveness [BIO2016-783-78310-R to S.V.] and by ICREA [ICREA-Academia 2015 to S.V.]

Author Contributions

Conceived and supervised: S.V. Performed the experiments: C.B., V.I. and S.N. Design, analysis and interpretation of experiments: C.B., N.S.G. and S.V. Wrote the paper: C.B. and S.V.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-09714-z](https://doi.org/10.1038/s41598-017-09714-z)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

5.5. PUBLICATION V

Perfecting prediction of mutational impact on the aggregation propensity of the ALS-associated hnRNPA2 prion-like protein.

Battle C., Fernández MR., Iglesias V and Ventura S.

FEBS Lett. (2017)

DOI: 10.1002/1873-3468.12698

Perfecting prediction of mutational impact on the aggregation propensity of the ALS-associated hnRNPA2 prion-like protein

Cristina Batlle, María Rosario Fernández, Valentin Iglesias and Salvador Ventura

Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain

Correspondence

S. Ventura, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193-Bellaterra (Barcelona), Spain

Fax: +34 93 581 2011

Tel: +34 93 586 8956

E-mail: salvador.ventura@uab.es

(Received 14 March 2017, revised 7 April 2017, accepted 21 May 2017, available online 18 June 2017)

doi:10.1002/1873-3468.12698

Edited by Alfonso Valencia

An increasing number of human proteins are being found to bear a prion-like domain (PrLD) driving the formation of membraneless compartments through liquid–liquid phase separation. Point mutations in these PrLDs promote the transition to an amyloid-like state. There has been much debate on whether this aberrant aggregation is caused by compositional or sequential changes. A recent extensive mutational study of the ALS-associated prion-like hnRNPA2 protein provides a framework to discriminate the molecular determinants behind pathogenic PrLDs aggregation. The effect of mutations on the aggregation propensity of hnRNPA2 is best predicted by combining their impact on PrLD amino acid composition and sequence-based amyloid propensity. This opens an avenue for the prediction of disease causing mutations in other human prion-like proteins.

Keywords: amyloid; prion-like proteins; protein aggregation

Around 1% of human proteins contain a prion-like domain (PrLD) resembling the intrinsically disordered, low complexity, Q/N-rich regions present in many yeast prions [1,2]. A significant fraction of these proteins is also enriched in RNA-binding domains and seem to be involved in the formation of membraneless intracellular compartments through liquid–liquid phase separation [1,3,4]. Mutations in the PrLDs of several of these proteins have been shown to promote an aberrant transition to a solid amyloid-like state linked to the onset of degenerative diseases [3,4]. This is the case of the human heterogeneous nuclear ribonucleoprotein hnRNPA2B1, an ubiquitous RNA-binding protein involved in pre-mRNA processing [5]. HnRNPA2B1 contains a PrLD at its C terminus (Fig. 1A) with an intrinsic propensity to self-assemble into an amyloid-like hydrogel [3]. This protein can be present in two alternatively spliced isoforms, hnRNPA1 and

hnRNPA2, from which the shorter A2 isoform is predominant in most tissues. A single D290V mutation within its PrLD promotes hnRNPA2 aggregation and causes Amyotrophic Lateral Sclerosis (ALS) and multisystem proteinopathy [6].

Accurate predictions of the impact of mutations on the self-assembly of human PrLDs would help to uncover novel disease-associated proteins. In a recent and very detailed work, Ross and co-workers have examined the effect of a large set of mutations, in or close to position 290 of hnRNPA2, on the protein aggregation propensity *in vitro*, in yeast and in *Drosophila* [7]. In the yeast model, they replaced a region of the yeast Sup35 prion domain by that of wild-type (WT) hnRNPA2 PrLD or its mutants. The prion-promoting activity induced by the different mutations was easily tracked by monitoring the ability of Ade⁻ yeast cells (holding a premature stop codon in the *ade1*

Abbreviations

ALS, amyotrophic lateral sclerosis; AMYCO, combined AMYloid and COmposition-based prediction of prion-like propensity; PrLD, prion-like domain; WT, wild-type.

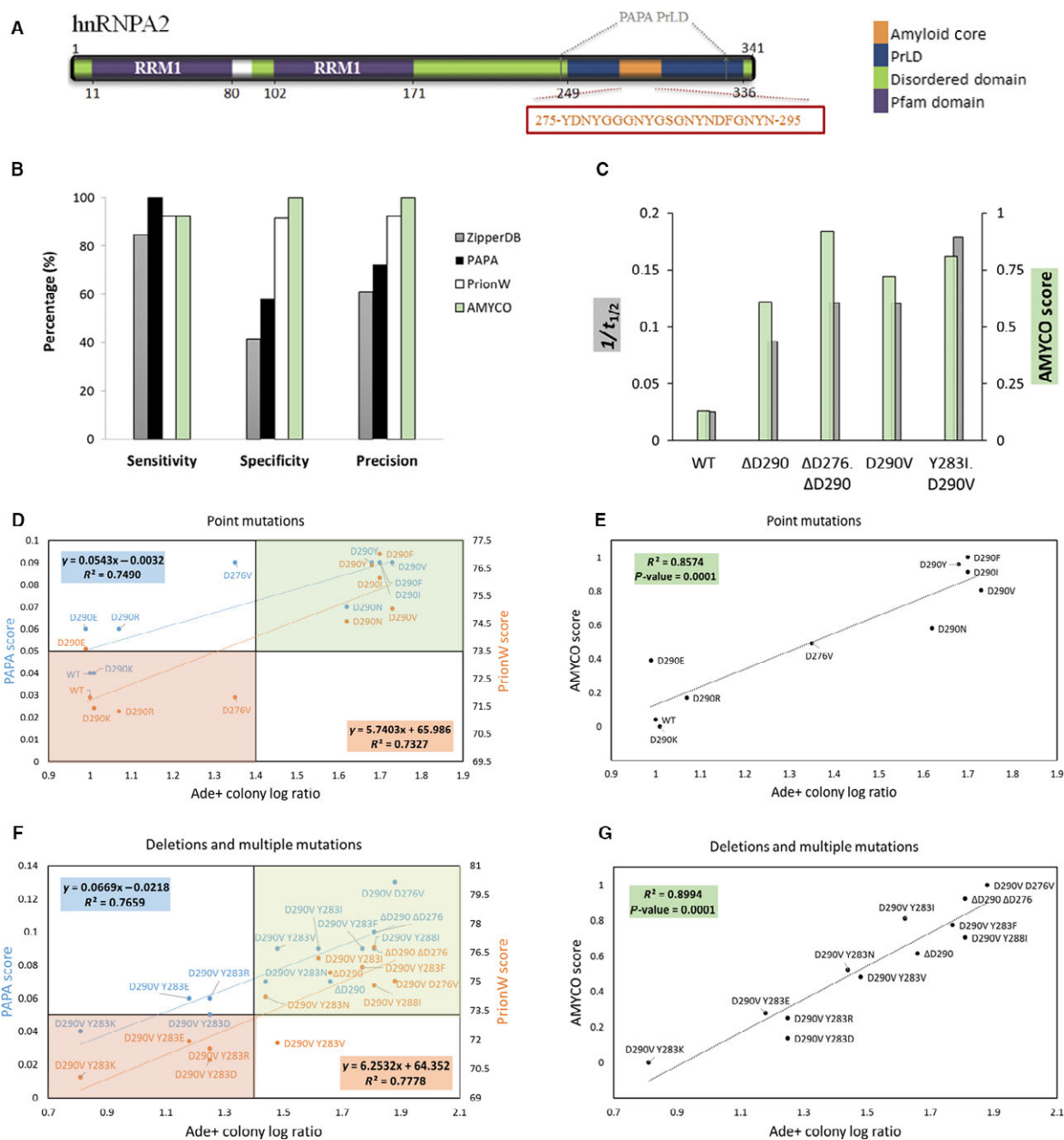


Fig. 1. Performance of ZipperDB, PAPA, PrionW, and AMYCO in the prediction of hnRNPA2 mutants. (A) Architecture of WT hnRNPA2: folded Pfam domains [16] are shown in purple. hnRNPA2 displays two RNA recognition motifs (RRM1). Disordered domains are shown in green. The PrionW predicted PrLD is shown in blue and that by PAPA flanked by arrows. The predicted PrionW soft amyloid stretch is shown in orange. (B) Graphic representation of Sensitivity, Specificity and Precision. (C) Relationship between the inverse of the half-time of *in vitro* aggregation ($t_{1/2}$) and AMYCO scores for 35-residue-long peptides containing the indicated mutations. (D) Correlation between PAPA and PrionW scores for single point mutants and their correspondent Ade⁺ colony log ratio (\log_{10} mutant Ade⁺/ \log_{10} WT Ade⁺). (E) Correlation between AMYCO scores for single point mutants and their correspondent Ade⁺ colony log ratio. (F) Correlation between PAPA and PrionW scores for deletions and multiple mutations and their correspondent Ade⁺ colony log ratio. (G) Correlation between AMYCO scores for deletions and multiple mutations and their correspondent Ade⁺ colony log ratio.

gene) to grow in adenine-defective media when they overexpress aggregation-prone hnRNPA2 PrLD-Sup35 fusions, rendering an Ade⁺ phenotype. For instance, the D290V mutant supports prion formation, whereas the WT does not [7]. Both ZipperDB [8] and PAPA [9] algorithms were shown previously to predict these relative prion propensities and thus the pathogenic impact of the D290V mutation [6]. ZipperDB and PAPA rely on very different conceptual basis. ZipperDB is a structure-based predictor that identifies short regions of six contiguous residues able to form a strong amyloid zipper [8]. In contrast, PAPA is a composition-based predictor aimed to identify long sequence stretches with a compositional bias similar to that of yeast prion domains [9]. When confronted with the complete set of mutants, ZipperDB could not predict steric zippers for a number of mutations promoting the same or higher level of aggregation than D290V and failed to foresee the inhibitory effect of compensatory mutations outside the main 6-residues steric zipper [7] (Table 1 and Fig. 1B). PAPA performed much better and this led the authors to propose that the changes in PrLD composition upon mutation would alone account for the observed phenotypes and thus, in contrast to what occurs in classical amyloids, the presence of strong steric zippers is neither necessary nor sufficient for the pathogenic aggregation of these kinds of human proteins [7]. Here, we reanalyze these data to provide a different view of the determinants underlying the pathogenic aggregation of human prion-like proteins.

Materials and methods

Computational analysis of hnRNPA2 mutations on aggregation propensity

Wild-type hnRNPA2 and the set of mutants generated by Ross and coworkers [7] were analyzed using ZipperDB [8], PAPA [9], and PrionW [10] algorithms. ZipperDB was used with the default setting and, as in the previous work [7], a $-25 \text{ kcal}\cdot\text{mol}^{-1}$ threshold was used to identify strong steric zippers. PAPA was used with the default settings and a 0.05 threshold was used to identify prionic sequences, as previously described [7]. PrionW was used with the standard settings, in which sequences with an amyloid core scoring above 73.55 threshold are considered prionogenic. The lower PrionW detection limit was set at 65.0. To develop the AMYCO function (combined AMYloid and COMposition-based prediction of prion-like propensity), the PAPA and PrionW scores for the different hnRNPA2 mutants (Table 1) were normalized between 0 and 1 (PAPA_n and PrionW_n). The AMYCO score for a given variant (AMYCO_s) was then calculated using the formula

$\text{AMYCO}_s = 0.5 \cdot \text{PAPA}_n + 0.5 \cdot \text{PrionW}_n$, that is, the arithmetic media of the normalized predictions of both algorithms. In Table 1, AMYCO_s were normalized again between 0 and 1 and the detection threshold for AMYCO set to 0.5.

Evaluation of yeast and *in vitro* experimental data

The number of Ade⁺ colonies in cells expressing the WT and the set of hnRNPA2 mutants was extracted from Ross and coworkers' results [7]. We calculated the ratio of Ade⁺ colonies between each individual mutant and the WT protein. In addition, it was calculated the ratio between the log₁₀ of the number of Ade⁺ colonies for each individual mutant and the log₁₀ of the number of Ade⁺ cells in WT hnRNPA2. The *in vitro* half-time of aggregation ($t_{1/2}$) for the peptides containing the different mutations was calculated from the kinetic assays reported by Ross and coworkers [7].

Performance analysis

The sensitivity, specificity, and precision of ZipperDB, PAPA, PrionW and AMYCO when predicting the experimental data was evaluated as follows: Sensitivity = True positives/(True positives + False negatives); Specificity = True negatives/(True negatives + False positives); and Precision = True positives/(True positives + False positives).

Results and Discussion

Driven by the idea that, in addition to composition, sequence should play a crucial role in PrLD assembly, we have recently proposed the soft amyloid stretch model, which postulates that these domains bear certain specific regions with significant amyloid potential and that mutations that increase the potency of these stretches would result in aberrant phase transitions [11,12]. We rationalized that in PrLDs, these regions should be longer than the classical amyloid stretches detected by algorithms like ZipperDB, in such a way that the amyloid potential would be more diffusely distributed; each residue having an average lower potency, but with more residues contributing to the assembling force. We implemented this notion in pWALTZ, an algorithm that predicts the 21-residue-long sequence stretch with the highest amyloid potential in a disordered Q/N-rich context and identifies PrLDs according to the presence and the potency of these stretches [13]. This concept was further included in the PrionW webserver [10], whose predictions have been validated experimentally by demonstrating the existence of such amyloid stretches in the prion domains of four of the best characterized yeast prions

Table 1. Correspondence between the observed prion behavior and the predicted effect of hnRNPA2 PrLD mutations.

Mutation	Ade ⁺ colony ratio ^a	Ade ⁺ colony log ratio ^b	Zipper DB	PAPA score	PrionW score	AMYCO score ^c
WT (290D)	1.00	1.00	–	0.04	71.83	0.13
D290V Y288P	0.30	0.83	–	0.04	–	–
D290V Y283P	0.80	0.96	+	0.05	–	–
D290P	0.80	0.96	–	0.04	–	–
D290V Y283K	0.31	0.81	+	0.04	70.07	0.00
D290E	0.95	0.99	+	0.06	73.58	0.40
D290K	1.10	1.01	–	0.04	71.43	0.10
D290R	1.55	1.07	–	0.06	71.32	0.23
D290V Y283E	3.09	1.18	+	0.06	71.92	0.28
D290V Y283D	4.90	1.25	+	0.05	70.95	0.14
D290V Y283R	4.90	1.25	+	0.06	71.54	0.25
D276V	8.71	1.35	+	0.09	71.83	0.48
D290V Y283N	15.50	1.44	+	0.07	74.22	0.52
D290V Y283V	19.50	1.48	+	0.09	71.84	0.48
D290V Y283I	48.96	1.62	+	0.09	76.21	0.81
D290N	48.96	1.62	+	0.07	74.58	0.55
ΔD290	63.08	1.66	–	0.07	75.46	0.61
D290Y	69.16	1.68	+	0.09	76.60	0.84
D290F	77.60	1.70	+	0.09	77.02	0.87
D290I	77.60	1.70	+	0.09	76.14	0.81
D290V	97.70	1.73	+	0.09	75.03	0.72
D290V Y283F	122.99	1.77	+	0.09	75.73	0.77
D290V Y288I	154.80	1.81	+	0.09	74.80	0.70
ΔD276 ΔD290	154.80	1.81	–	0.10	76.76	0.92
D276V D290V	245.40	1.88	+	0.13	75.03	1.00

^aAde⁺ colony ratio corresponds to the ratio between the number of Ade⁺ colonies in cells expressing a given variant and those in cells expressing the WT sequence (extracted from [7]).

^bAde⁺ colony log ratio corresponds to the log₁₀ of the number of Ade⁺ colonies for a given variant divided by the log₁₀ of the number of Ade⁺ cells in WT. The presence of a predicted steric zipper is indicated for ZipperDB [8]. PAPA [9] and PrionW [10] scores are indicated, the detection thresholds for these algorithms are 0.05 and 73.55, respectively. Proline mutants do not pass the lower score used in PrionW predictions (65.0). Mutations resulting in an Ade⁺ colony ratio below 10 are shown in red, and mutations above this ratio in green, as an arbitrary way to discriminate between potentially pathogenic and innocuous mutations.

^cAMYCO corresponds to a linear combination of normalized PAPA and PrionW scores. PAPA and PrionW scores were normalized between 0 and 1 (PAPA_n and PrionW_n). AMYCO score = 0.5*PAPA_n + 0.5*PrionW_n. The AMYCO scores were normalized between 0 and 1 for the complete mutation dataset and the detection threshold set at 0.5.

[12], as well as in the predicted PrLD of the Rho termination factor of *Clostridium botulinum* [14], which later led to the discovery of the first bacterial prion-like protein [15].

Interestingly enough, D290 is located at the highest-scoring 21-residue stretch in hnRNPA2 PrLD, as predicted by PrionW (Y275-N295) and all the mutations introduced by Ross and coworkers map within this protein region (Table 1 and Fig. 1A). The WT protein scores below the PrionW PrLD detection threshold, whereas the D290V mutation increases the amyloid potential of this stretch well above the cut-off (Table 1). Thus, as PAPA, PrionW is able to forecast the pathogenic impact of the D290V change.

To discriminate whether it is the sequence or the composition that accounts for PrLDs aggregation, we analyzed the correlation between the impact of all

assayed point mutations on hnRNPA2 aggregation propensity and the PAPA and PrionW scores for these sequences (Table 1 and Fig. 1D). The predictions of both algorithms correlated well with the experimental data, PAPA performing slightly better (Fig. 1D). In both cases, the D276V mutation behaves as an outlier. This mutation has a much lower impact on the experimental prion propensity of hnRNPA2 PrLD than D290V [7], which argues against composition alone accounting for the aggregation of this PrLD. Indeed, PAPA predicts the same aggregation potential for both mutations. PrionW, correctly predicts the trend, but it underscores the aggregation potential of D276V (Fig. 1D). In the light of these data, we rationalized that it could be the combination of compositional bias and sequence-specific amyloid potential that determines the prion/aggregation potential of a PrLD. Thus, we

normalized PAPA and PrionW scores and combined them in a linear function named AMYCO (combined AMYloid and COMposition-based prediction of prion-like propensity) (Table 1). As shown in Fig. 1E, the combined scores correlate much better with the experimental data than any of the individual algorithms. We repeated the analyses with the set of double mutations and deletions that Ross and coworkers introduced in the hnRNPA2 PrLD [7] to see whether AMYCO can also predict these larger sequential/compositional changes. The predictions of both individual programs correlated well with the experimental data, PrionW performing slightly better for this subset (Fig. 1F). However, AMYCO clearly outperformed them (Fig. 1G).

In yeast, the pathogenic D290V mutation increases the number of Ade⁺ colonies by almost two orders of magnitude, relative to the WT sequence [7] (Table 1); thus, we considered that a hnRNPA2 mutation should increase the Mutant/WT Ade⁺ ratio by at least by one order of magnitude to be considered potentially pathogenic. AMYCO exhibits higher precision and specificity than ZipperDB, PAPA, and PrionW algorithms in discriminating this subset of potentially toxic mutants from the rest of the variants in the complete dataset (Fig. 1B).

The AMYCO is also able to predict with good accuracy the impact of mutations on the *in vitro* aggregation kinetics of 35-residue-long peptides containing the PrionW soft amyloid core, as assayed by Ross and coworkers [7] (Fig. 1C), with a significant correlation between the half-time of *in vitro* aggregation ($t_{1/2}$) of the different variants and its AMYCO score ($R^2 = 0.91$, $P = 0.015$). However, as reported for ZipperDB and PAPA [7], AMYCO is an imperfect predictor of inclusion formation in *Drosophila*, where a few mutations behave as outliers, displaying a behavior different from the one they exhibit in yeast and *in vitro*. As stated by Ross and coworkers, the reasons for these discrepancies are unclear and should be further explored to clarify the mechanisms underlying pathogenic aggregation [7]. Nevertheless, it should be mentioned that all the four variants forming inclusions when expressed in *Drosophila* (D290V, D276V-D290V, Δ D290, and D290F) exhibit AMYCO scores at least 4.5-fold higher than the WT protein.

Overall, our analysis suggests that the ability of human hnRNPA2 PrLD to assemble into not only functional but also pathogenic macromolecular structures depends on both a special amino acid composition and on the presence of at least one specific soft amyloid stretch. The compositional bias would provide a cryptic and distributed aggregation propensity that can zipper the PrLD upon soft amyloid core driven nucleation. This mechanistic assumption provides a

tool to identify the occurrence of disease-associated mutations in other prion-like human proteins, as well as to rationally design the aggregation propensity of PrLD-containing proteins in order to study their role in human health and disease.

Acknowledgements

This work was funded by the Spanish Ministry of Economy and Competitiveness BIO2016-783-78310-R to S.V. and by ICREA, ICREA-Academia 2015 to S.V.

Author contributions

SV designed the research. CB, MRF, and VI performed the analyses. SV wrote the manuscript.

References

- King OD, Gitler AD and Shorter J (2012) The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res* **1462**, 61–80.
- An L and Harrison PM (2016) The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biol Direct* **11**, 32.
- Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J *et al.* (2012) Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753–767.
- Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoynev S, Mahamid J, Saha S, Franzmann TM *et al.* (2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077.
- He Y and Smith R (2009) Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol Life Sci* **66**, 1239–1256.
- Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, MacLea KS, Freibaum B, Li S, Molliex A *et al.* (2013) Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473.
- Paul KR, Molliex A, Cascarina S, Boncella AE, Taylor JP and Ross ED (2017) The effects of mutations on the aggregation propensity of the human prion-like protein hnRNPA2B1. *Mol Cell Biol* **37**, e00652–16.
- Goldschmidt L, Teng PK, Riek R and Eisenberg D (2010) Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci* **107**, 3487–3492.
- Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A and Ross ED (2012) De novo design of synthetic prion domains. *Proc Natl Acad Sci USA* **109**, 6519–6524.

- 10 Zambrano R, Conchillo-Sole O, Iglesias V, Illa R, Rousseau F, Schymkowitz J, Sabate R, Daura X and Ventura S (2015) PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res* **43**, W331–W337.
- 11 Sabate R, Rousseau F, Schymkowitz J, Batlle C and Ventura S (2015) Amyloids or prions? That is the question. *Prion* **9**, 200–206.
- 12 Sant’Anna R, Fernandez MR, Batlle C, Navarro S, de Groot NS, Serpell L and Ventura S (2016) Characterization of amyloid cores in prion domains. *Sci Rep* **6**, 34274.
- 13 Sabate R, Rousseau F, Schymkowitz J and Ventura S (2015) What Makes a Protein Sequence a Prion? *PLoS Comput Biol* **11**, e1004013.
- 14 Pallarès I, Iglesias V and Ventura S (2016) The rho termination factor of *Clostridium botulinum* contains a prion-like domain with a highly amyloidogenic core. *Front Microbiol* **6**, 1–12.
- 15 Yuan AH and Hochschild A (2017) A bacterial global regulator forms a prion. *Science* **355**, 198–201.
- 16 Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res* **40**, D290–D301.

5.6. PUBLICATION VI

hnRNPD L phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation.

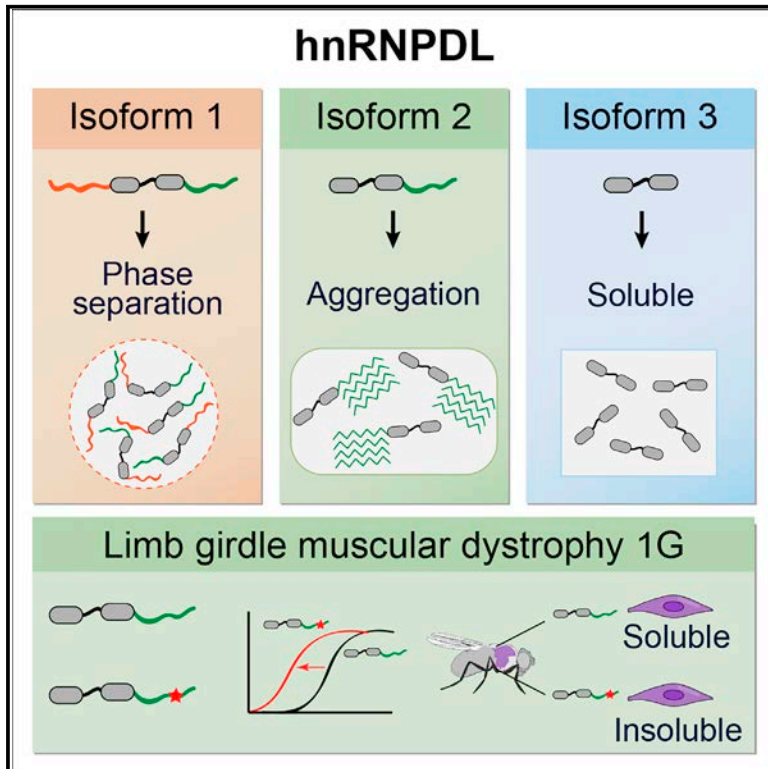
Battle C., Yang P., Coughlin M., Messing J., Pesarro dona M., Szulc E., Salvatella X., Kim HJ., Taylor JP. and Ventura S.

Cell Rep. (2020).

DOI: 10.1016/j.celrep.2019.12.080

hnRNPD L Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation

Graphical Abstract



Authors

Cristina Batlle, Peiguo Yang, Maura Coughlin, ..., Hong Joo Kim, J. Paul Taylor, Salvador Ventura

Correspondence

jpaul.taylor@stjude.org (J.P.T.), salvador.ventura@uab.cat (S.V.)

In Brief

Batlle et al. show that alternative splicing controls heterogeneous ribonucleoprotein D-like (hnRNPD L) phase separation, aggregation, and solubility. Mutations that cause LGMD1G accelerate hnRNPD L aggregation and promote insolubility in *Drosophila*.

Highlights

- hnRNPD L requires both N- and C-terminal IDRs to phase separate
- The absence of N-terminal IDR facilitates aggregation
- The unique combination of IDRs in each isoform determines its cellular behavior
- D378N/H mutations accelerate hnRNPD L aggregation and compromise its solubility



hnRNPD L Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation

Cristina Batlle,¹ Peiguo Yang,² Maura Coughlin,² James Messing,^{2,3} Mireia Pesarrodona,^{4,5} Elzbieta Szulc,^{4,5} Xavier Salvatella,^{4,5,6} Hong Joo Kim,² J. Paul Taylor,^{2,3,*} and Salvador Ventura^{1,7,*}

¹Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

²Department of Cell and Molecular Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

³Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

⁴Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain

⁵Joint BSC-IRB Research Programme in Computational Biology, Baldiri Reixac 10, 08028 Barcelona, Spain

⁶ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain

⁷Lead Contact

*Correspondence: jpaul.taylor@stjude.org (J.P.T.), salvador.ventura@uab.cat (S.V.)

<https://doi.org/10.1016/j.celrep.2019.12.080>

SUMMARY

Prion-like proteins form multivalent assemblies and phase separate into membraneless organelles. Heterogeneous ribonucleoprotein D-like (hnRNPD L) is a RNA-processing prion-like protein with three alternative splicing (AS) isoforms, which lack none, one, or both of its two disordered domains. It has been suggested that AS might regulate the assembly properties of RNA-processing proteins by controlling the incorporation of multivalent disordered regions in the isoforms. This, in turn, would modulate their activity in the downstream splicing program. Here, we demonstrate that AS controls the phase separation of hnRNPD L, as well as the size and dynamics of its nuclear complexes, its nucleus-cytoplasm shuttling, and amyloidogenicity. Mutation of the highly conserved D378 in the disordered C-terminal prion-like domain of hnRNPD L causes limb-girdle muscular dystrophy 1G. We show that D378H/N disease mutations impact hnRNPD L assembly properties, accelerating aggregation and dramatically reducing the protein solubility in the muscle of *Drosophila*, suggesting a genetic loss-of-function mechanism for this muscular disorder.

INTRODUCTION

Eukaryotic cells contain a variety of compartments or organelles with specialized functions. There are membrane-bound organelles like the nucleus or mitochondria, and membraneless organelles (MLOs) such as stress granules or P-bodies (Boeynaems et al., 2018). MLOs are enriched in a peculiar type of polypeptides known as prion-like proteins (March et al., 2016). These polypeptides consist of one or more globular domains with adjacent long intrinsically disordered regions (IDRs) of low

complexity. These IDRs are enriched in specific amino acids, such as glutamine, asparagine, serine, glycine, and tyrosine, being similar in composition to the disordered domains of yeast prions, and thus referred to as prion-like domains (PrLDs) (King et al., 2012). Interestingly, prion-like proteins often have the ability to phase separate into liquid droplets and this may contribute to the formation of MLOs in the nucleus or cytoplasm (Boeynaems et al., 2018). MLOs are dynamic structures and their formation is usually reversible, but these assemblies may become irreversible when proteins aggregate within MLOs due to mutations, prolonged stress, or changes in protein concentration. Protein aggregation is linked to the onset of a growing list of human disorders (Harrison and Shorter, 2017). Not surprisingly, increasing evidences indicate a connection between pathological states and MLOs proteins malfunction (Ito et al., 2017). ATX2 (Kato et al., 2019), EWSR1 (Maharana et al., 2018), FUS (Patel et al., 2015), hnRNPA1 (Kim et al., 2013; Mollieix et al., 2015), hnRNPA2 (Kim et al., 2013; Ryan et al., 2018), HTT (Peskett et al., 2018), TAF15 (Maharana et al., 2018), Tau (Wegmann et al., 2018), TDP43 (Babinchak et al., 2019), and TIA1 (Mackenzie et al., 2017) are well-characterized proteins involved in the formation of MLOs; their mutation being associated with age-related disorders such as amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), or inclusion-body myopathy (IBM) (Harrison and Shorter, 2017; Ito et al., 2017; Nedelsky and Taylor, 2019). Despite the increasing interest in MLOs, the molecular mechanisms that govern the transition between their functional and pathologic states are still not well understood.

Alternative splicing (AS) is an important mechanism underlying evolution complexity (Baralle and Giudice, 2017). Many MLOs proteins have AS isoforms with unknown functions (Gueroussov et al., 2017). Indeed, AS events are frequent in prion-like proteins, especially at their PrLDs, affecting their ability to establish multivalent interactions and to form higher-order complexes (Gueroussov et al., 2017). AS also alters the phase separation properties of prion-like proteins' isoforms as observed for FUS protein with or without exon 8 inclusion (Gueroussov et al., 2017).



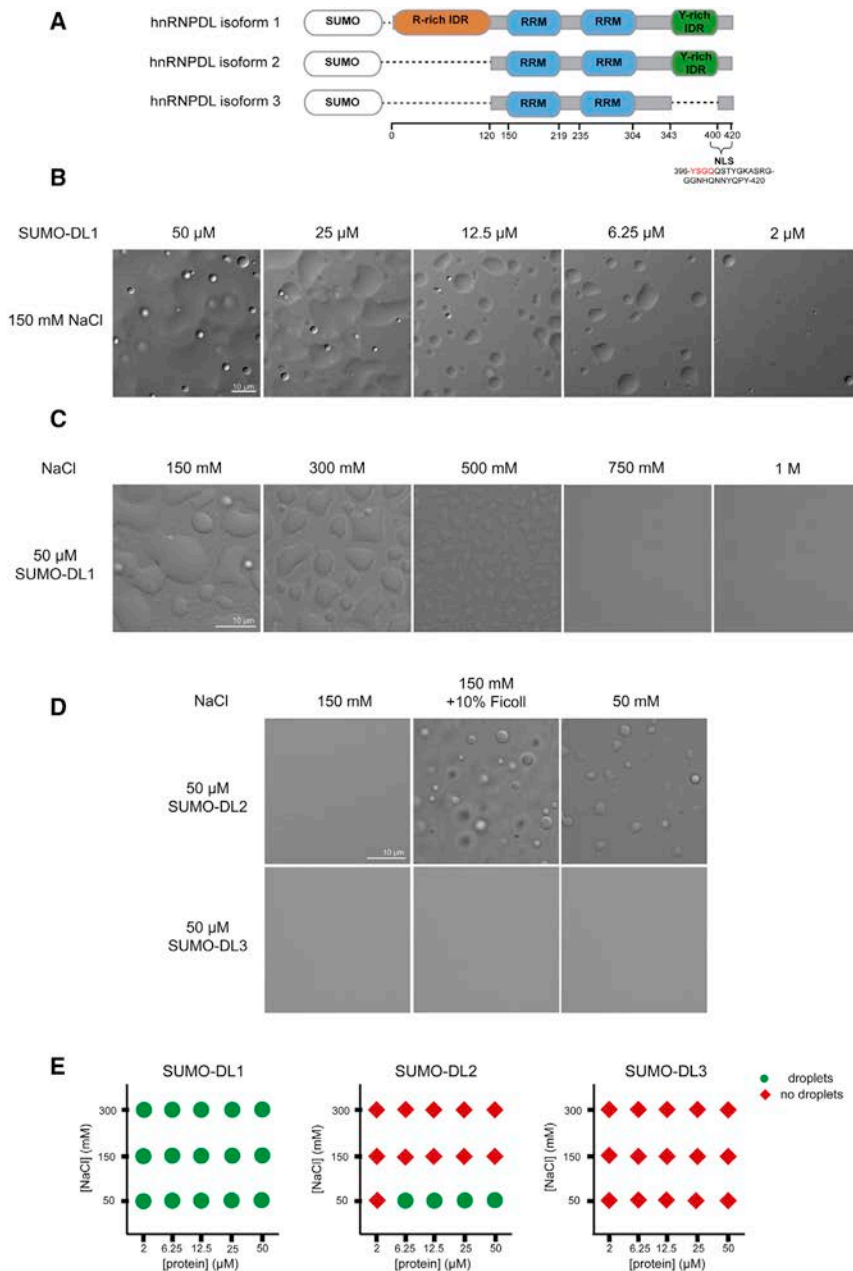


Figure 1. LLPS Propensity of hnRNPDL Isoforms

(A) Schematic diagram of hnRNPDL isoforms as SUMO fusion constructs. RNA recognition motifs (RRMs; blue) according to Pfam (El-Gebali et al., 2019), Arg-rich (orange), and Tyr-rich (green) IDR spliced regions according to Uniprot (Bateman et al., 2015) and their respective amino acid splicing positions are shown. hnRNPDL nuclear localization signal (NLS) sequence is as described in Kawamura et al. (2002).

(B) SUMO-hnRNPDL isoform 1 (DL1) LLPS at different protein concentrations in 50 mM HEPES pH 7.5 and 150 mM NaCl.

(C) 50- μ M DL1 LLPS at different salt concentrations.

(D) 50- μ M SUMO-DL2 and SUMO-DL3 LLPS in 150 mM salt with or without the presence of 10% Ficoll, and 50 mM salt.

(E) LLPS diagram of hnRNPDL isoforms in the absence of crowding agent. Green circles indicate positive and red diamonds indicate negative for the appearance of droplets at the indicated NaCl/protein concentration combinations.

(Tsuchiya et al., 1998) and it is the predominant isoform in all mouse and human tissues (Akagi et al., 2000). It is 301 amino acids long, constituted by two contiguous canonical RNA recognition motifs (RRMs) and one predicted PrLD at the C terminus, enriched in Gly and Tyr residues. DL1 is a longer isoform of 420 amino acids comprising an additional predicted to be disordered and Arg-enriched domain at the N terminus (Kamei et al., 1999). DL1 expression levels are 4-fold lower than those of DL2 and the transcript is mainly present in brain and testis (Akagi et al., 2000). DL3 is the shorter and minor isoform, with only 244 amino acids missing both N and C terminus disordered regions (Kawamura et al., 2002).

hnRNPDL bears a 25-residue C-terminal PY nuclear localization signal (NLS) and its transport is mediated by the M9-transportin-1 (TNPO1) pathway (Kawamura et al., 2002). Interestingly, only DL2 and DL3 are able to shuttle between the cytoplasm and the nucleus, whereas DL1 remains strictly nuclear. hnRNPDL isoforms share the same shuttling sequence, except for DL3, which misses four residues that are not required for TNPO1 interaction (Figure 1A; Kawamura et al., 2002). Therefore, the basis of this differential translocation is unknown, but it may indicate hnRNPDL isoforms playing different roles in cells.

HNRNPDL constitutes one of the four genes present in the smallest deletion of the 4q21 microdeletion syndrome, being associated with growth retardation and hypotonia (Hu et al., 2017). Moreover, hnRNPDL expression is upregulated in

We focus this study on the heterogeneous ribonucleoprotein D-like (hnRNPDL), an RNA-binding protein displaying AS isoforms, linked to disease and with motifs similar to those associated with liquid-liquid phase separation (LLPS) in well-characterized human prion-like proteins.

hnRNPDL is a highly conserved nuclear RNA binding protein involved in mRNA biogenesis located in the genomic position 4q21 (Kamei et al., 1999). The *HNRNPDL* gene contains nine exons and eight introns, and three isoforms are produced by AS, named here as hnRNPDL isoform 1 (DL1), hnRNPDL isoform 2 (DL2), and hnRNPDL isoform 3 (DL3) (Figure 1A). DL2 was the first isoform discovered as a JKT41 binding protein 1 (JKTBP1)

different types of cancers, such as prostate cancer, chronic myeloid leukemia, colon cancer, and hepatocellular carcinoma (Zhou et al., 2014; Liu et al., 2007; Wu et al., 2008; Zhang et al., 2018). Finally, genome sequencing of Brazilian, Chinese, Uruguayan, and Argentinian families affected by limb-girdle muscular dystrophy 1G (LGMD1G, or LGMDD3 in the new nomenclature; Straub et al., 2018) detected D378N and D378H point substitutions in *HNPNPD*, indicating a mutation hotspot for this disease (Berardo et al., 2019; Sun et al., 2019; Vieira et al., 2014). LGMD1G is an autosomal dominant inherited subtype of LGMD, the fourth most common muscular dystrophy, characterized by progressive weakness of hip- or shoulder-girdle muscles (Liewluck and Milone, 2018; Nigro and Savarese, 2014).

Interestingly, hnRNPDL D378N/H point mutations reside in the PrLD of hnRNPDL (Navarro et al., 2015). Disease-causing mutations in PrLDs have been discovered in other prion-like proteins, for example in hnRNPA1 and hnRNPA2 proteins (Kim et al., 2013), in both cases involving the replacement of a single Asp residue, as in hnRNPDL. These mutations have been reported to increase protein aggregation and result in the formation of cytoplasmic inclusions in patients (Harrison and Shorter, 2017; Kim et al., 2013; Mackenzie et al., 2017). Immunohistochemical analysis of LGMD1G patients identified nuclear condensates of hnRNPDL co-localizing with TNPO1 (Vieira et al., 2014). Congoophilic deposits have however not been detected in LGMD1G patients and it remains unknown whether hnRNPDL mutations impact the protein aggregation propensity.

In this study, we show how AS results in hnRNPDL isoforms with dramatically different self-assembling properties *in vitro* and *in vivo*. We also demonstrate how Arg and Tyr residues, segregated in two distant IDRs in hnRNPDL, act as crucial determinants for both LLPS and aggregation. This spatial segregation of multivalent interacting residues explains how AS controls the propensity to form high-order intranuclear assemblies in mammalian cells, likely accounting for the different shuttling properties of the hnRNPDL isoforms. Finally, we confirm that, as in other prion-like proteins, hnRNPDL disease-causing mutations accelerate protein aggregation, resulting in completely insoluble variants when expressed in *Drosophila* muscle.

RESULTS

hnRNPDL Alternative Splicing Isoforms Display Different Phase Separation Behavior

hnRNPDL has been shown to undergo LLPS *in vitro* (Wang et al., 2018). To evaluate the molecular determinants that govern this process, we took advantage of the different domain architectures of hnRNPDL isoforms and we tested their propensity to undergo phase transition. We expressed and purified hnRNPDL isoforms as fusions with solubility-enhancing His-SUMO tags (Figure 1A), hereinafter referred to as DL1, DL2, and DL3. As expected, the three hnRNPDL isoforms bear distinct LLPS propensity (Figures 1B–1E).

The DL1 isoform, containing Arg- and Tyr-enriched IDRs at the N- and C terminus, respectively, displays the strongest tendency to phase separate (Figure 1E). Upon salt dilution, the solution becomes turbid spontaneously, demixing from an

aqueous phase to form liquid-like protein droplets under physiological ionic strength (Figure 1B). DL1 undergoes LLPS in a protein concentration-dependent manner in the absence of any crowding agent even at low protein concentrations (2 μ M). For a given protein concentration, phase separation is enhanced with decreasing ionic strength (Figure 1C). DL2, missing the Arg-enriched disordered domain at its N terminus, also can undergo phase separation, but its propensity is much lower than that of DL1 (Figure 1E). At physiological ionic strength, DL2 requires the presence of a crowding agent (10% Ficoll) to phase separate (Figure 1D). When transferred to low ionic strength (50 mM NaCl), DL2 rapidly coalesces into micron-sized spherical structures without any crowding agent, but this process requires higher protein concentrations than for DL1 (Figures 1D and 1E). The fact that, in the absence of a crowding agent, DL2 only forms droplets at low ionic strength suggests that its LLPS depends on electrostatic interactions. Indeed, the C-terminal IDR is predicted to behave as a weak polyampholyte, a kind of molecule that displays reduced LLPS as the salt concentration increases (Das and Pappu, 2013). DL3, the isoform devoid of any IDRs, does not phase separate in any tested condition, neither at 50 mM NaCl nor after crowding agent addition (Figures 1D and 1E).

The radii of hydration of hnRNPDL isoforms were analyzed by dynamic light scattering (DLS) at physiological salt concentration without any crowding agent (Figure S1). The results correlated with the propensity to undergo LLPS (Figure 1E), with DL1 forming assemblies with a radius of hydration > 1,000 nm, which could not be observed in DL2 and DL3 proteins, that displayed average radii of hydration of 4.5 and 3.5 nm respectively, compatible with a monomeric state.

These data indicate that the absence of the N-terminus Arg-enriched IDR of hnRNPDL reduces LLPS propensity and the absence of both the N- and C-terminus IDRs completely abolishes LLPS, suggesting that these domains are required for hnRNPDL self-assembly. The results provide experimental support for the hypothesis that AS controls LLPS in hnRNPDL (Feng et al., 2019).

Interactions between Arg and Tyr Residues of DL1 IDRs Promote Its Phase Separation

Recent studies have reported the importance of Arg-rich domains in providing intermolecular interactions that contribute to LLPS (Boeynaems et al., 2017). Moreover, Arg residues have been shown to establish interactions with Tyr residues in FUS (Bogaert et al., 2018; Vernon et al., 2018; Wang et al., 2018). In the previous section, we have shown how the absence of the Arg- and Tyr-rich IDRs in hnRNPDL protein affects LLPS behavior. Therefore, we hypothesized that these residues might be interacting and promoting hnRNPDL phase separation. To confirm this hypothesis, we generated three hnRNPDL variants: (1) all N-terminus Arg mutated to Lys (R/K), (2) all C-terminus Tyr mutated to Phe (Y/F), and (3) both Arg to Lys and Tyr to Phe mutations (R/K+Y/F) (Figure 2A). These mutations were designed to maintain the aromatic and basic character of Tyr and Arg, respectively, while preserving the ability to establish π -cation contacts. The identity of the basic and positive residues determines the interaction strength, with Lys-Tyr, Arg-Phe, and

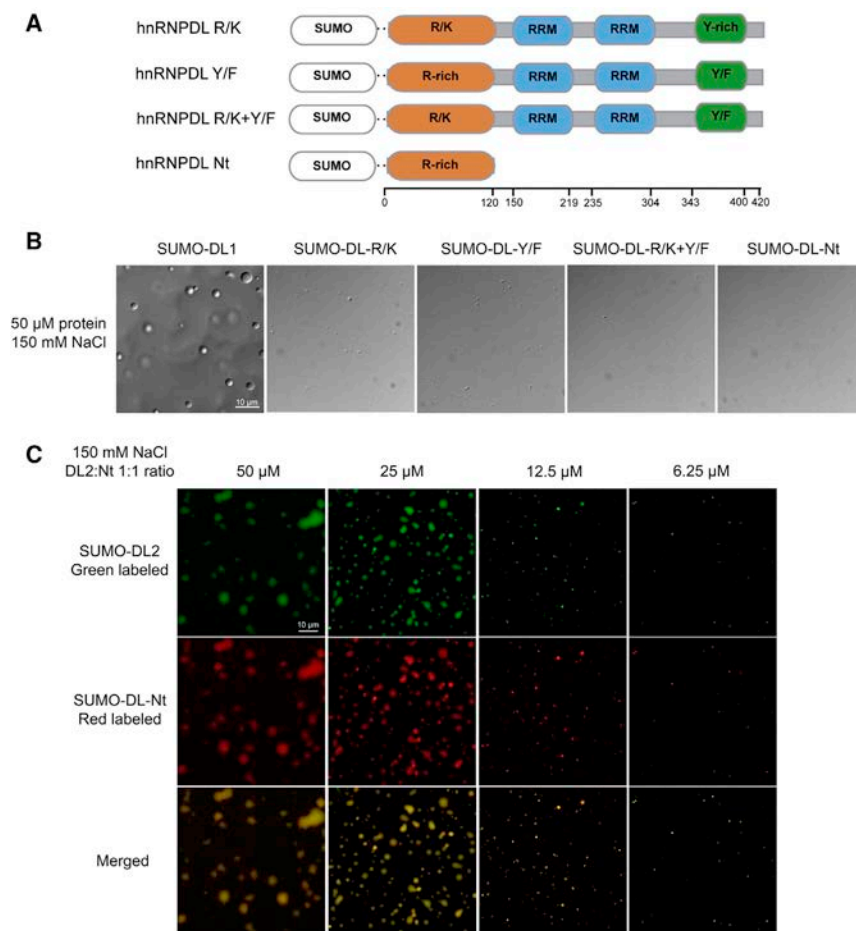


Figure 2. DL1 Mutants LLPS Behavior

(A) Schematic diagram of DL1 mutants as SUMO fusion constructs: (1) all N-terminus Arg mutated to Lys (R/K); (2) all C-terminus Tyr mutated to Phe (Y/F); (3) both Arg to Lys and Tyr to Phe mutations (R/K+Y/F); and (4) N-terminus (Nt) IDR alone. (B) LLPS of DL1 and mutants at 50 μM in 50 mM HEPES pH 7.5 and 150 mM NaCl. (C) LLPS of DL2 green labeled mixed with hnRNPDL-Nt red labeled, in a 1:1 ratio, at different protein concentrations and 150 mM salt.

with DL2 at a 1:1 ratio, we completely recover LLPS under physiological conditions and DL2 phase separated at concentrations as low as 6.25 μM in the absence of any crowding agent (Figure 2C).

The above results indicate that DL1 LLPS is likely governed by intermolecular interactions between the Arg residues at the N-terminal domain and Tyr residues at the C terminus.

DL1 Forms High-Order Complexes in the Nucleus of Mammalian Cells

In the previous sections, we observed how hnRNPDL isoforms bear different self-assembly propensities *in vitro*. hnRNPDL is located in the cell nucleus; consequently, we addressed their properties in the nuclear context. To this aim, we examined the isoforms localization in

Lys-Phe contacts all being weaker than Arg-Tyr (Wang et al., 2018). Therefore, these protein variants would allow us to evaluate the role of Arg and Tyr interaction strength in hnRNPDL LLPS.

The three hnRNPDL mutants showed a clear reduction of LLPS compared to DL1, being unable to phase separate at physiological conditions without crowding agents (Figures 2B and S2A). This supports the idea that hnRNPDL phase separation relies on the complementarity of Tyr and Arg contacts (Brady et al., 2017) and not on generic π -cation interactions, as observed also for FUS protein (Bogaert et al., 2018; Wang et al., 2018).

The reduced LLPS propensity of the R/K mutant (Figure S2A) is similar to that of the DL2 isoform (Figure 1E). The C terminus in DL2 is sufficient to promote LLPS, likely through Tyr-Tyr interactions (Burke et al., 2015; Murthy et al., 2019; Wang et al., 2018), but its ability to undergo LLPS under physiological ionic strength is weaker than that of DL1 (Figure 1E). To further confirm that the C-terminus Tyr-rich domain and the N-terminal Arg-rich domain can indeed form interactions responsible for the high LLPS propensity of DL1, we designed a construct consisting only of the Arg-rich domain of DL1 (hnRNPDL-Nt) (Figure 2A). hnRNPDL-Nt was unable to phase separate by itself (Figures 2B and S2A). However, when hnRNPDL-Nt was mixed

a HeLa hnRNPDL knockout (KO) cell line (HeLa^{DL-KO}) (Figure S3A). Cells were transfected with EGFP-tagged fusion constructs or EGFP alone, as control, and immunostained with anti-G3BP antibody as a cytoplasmic marker (Figure 3A).

All hnRNPDL isoforms are nuclear (Figure 3A). Both DL1 and DL2 are distributed throughout the nucleoplasm, but excluded from the nucleolus (Figure S4A). In contrast, DL3 was completely diffuse in the nucleus, suggesting that the hnRNPDL C-terminus IDR, the only region present in DL1 and DL2 and absent in DL3, might determine the intranuclear compartmentalization of the isoforms. DL1 and DL2 undergo LLPS *in vitro*, but they do not show nuclear puncta indicative of MLO formation. This behavior might be caused by the high RNA concentration in the nucleus, since high RNA/protein ratios have been shown to diminish LLPS in other prion-like RNA binding proteins, such as TDP43 and FUS (Maharana et al., 2018). In order to verify this idea, we added increasing concentrations of total RNA to DL1 *in vitro* and we observed a clear decrease in its LLPS propensity (Figures S2B and S2C). RNA completely dissolved DL1 droplets at 100 ng/μL (Figure S2B). This suggests that the nucleic acid titrates DL1 from the droplets. Accordingly, the droplets reappeared after RNase addition (Figure S2B).

We conducted fluorescent recovery after photobleaching (FRAP) analysis of a small region of the nucleus of HeLa^{DL-KO}

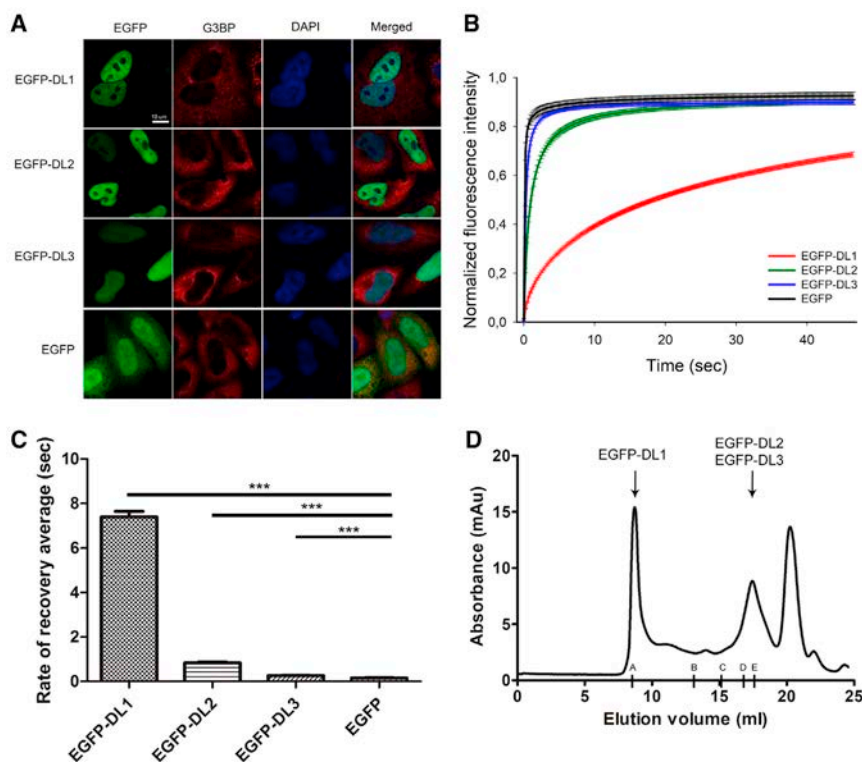


Figure 3. Cellular Localization and Mobility of hnRNPD L Isoforms in Mammalian Cells

(A) Cellular localization by immunofluorescence of EGFP-DL1, EGFP-DL2, EGFP-DL3, and unfused EGFP after expression in HeLa^{DL-KO}. Cells were stained with G3BP antibody (red) as cytoplasmic marker and DAPI (blue) as nuclear marker.

(B and C) Graph of normalized fluorescence intensity (B) and rate of recovery average (C) after FRAP in HeLa^{DL-KO} cells expressing EGFP-DL1, EGFP-DL2, and EGFP-DL3 and unfused EGFP (**p value < 0.0001).

(D) Example of a size exclusion chromatography elution pattern after individual EGFP-hnRNPD L isoforms expression in HeLa^{DL-KO} cells. Elution volumes of EGFP-DL1 (74 kDa), EGFP-DL2 (61 kDa), and EGFP-DL3 (55 kDa) are indicated by arrows. Letters along the x axis indicate the elution volumes upon column calibration: A, column void volume; B, ferritin (440 kDa); C, aldolase (158 kDa); D, conalbumin (75 kDa); and E, ovalbumin (44 kDa).

cells for all EGFP-tagged hnRNPD L isoforms. FRAP analysis showed that the three isoforms associate and dissociate within the nucleus on different timescales of seconds (Figure 3B). The fluorescence recovery half-times after photobleaching were 0.25, 0.84, and >7.0 s for DL3, DL2, and DL1, respectively (Figure 3C). The low mobility of DL1, relative to DL2 and DL3, suggests that it might be involved in the formation of larger or more stable complexes within the nucleoplasm. The DL2 and DL3 mobilities are also significantly different, indicating that these isoforms are also associated with nuclear complexes that differ in identity or stability, although the assemblies that they form are likely smaller and more dynamic than the ones formed by DL1.

To confirm that DL1, DL2, and DL3 could be involved in the formation of different complexes, we performed size exclusion chromatography (SEC) of individual HeLa^{DL-KO} cellular extracts after transfection with each of the different constructs (Figures 3D and S5). The DL1 isoform elution pattern differed significantly from that of DL2 and DL3, the protein being eluted in the void volume of the column corresponding to molecular size complexes larger than 5 million Da. Instead, DL2 and DL3 were eluted in volumes consistent with them being in their monomeric form. In order to prove that the EGFP-tag does not alter the elution pattern of the isoforms, the retention of endogenous hnRNPD L in a HeLa wild-type (WT) cell line was analyzed (Figure S5). The elution profile of endogenous DL2, the predominant isoform, is similar to that of EGFP-DL2. The levels of endogenous DL1 and DL3 are too low to be detected in this experiment.

These results correlate well with the observed hnRNPD L isoforms *in vitro* self-assembly behavior. The presence of

also for other proteins (Guerossov et al., 2017; Ying et al., 2017).

Transcription Inhibition Affects hnRNPD L Nuclear Localization

hnRNPs are predominantly located in the cell nucleus. Some hnRNPs, for example hnRNPA1, are known to shuttle between the nucleus and cytoplasm in a transcription-dependent manner, but others do not, for example hnRNPC (Dreyfuss et al., 2002; Piñol-Roma and Dreyfuss, 1991, 1992). Actinomycin D (ActD) is an anti-tumor chemical that inhibits transcription by intercalating into transcriptionally active regions, resulting in a significant reduction in RNA synthesis in the nucleus (Su et al., 2013). This treatment is usually performed to determine whether protein localization is dependent on active transcription, its inhibition resulting in protein translocation to the cytoplasm (Bounejah et al., 2014; Piñol-Roma and Dreyfuss, 1991, 1992; Zhang et al., 2005). Interestingly, in contrast to untreated cells (Figure 3A), after 5 μg/ml ActD treatment for 3 h, hnRNPD L shows different transport depending on the considered isoform. Only the DL2 and DL3 isoforms were translocated to the cytoplasm indicating nuclear import that is dependent upon active transcription; meanwhile, the DL1 isoform remained nuclear (Figure 4), consistent with previous observations (Kamei and Yamada, 2002; Kawamura et al., 2002). The three isoforms bear the same shuttling sequence (Figure 1A) and, thus, in principle, they could be transported in the same manner. In the previous section, we showed that DL1 forms higher- or more-stable-order complexes (Figures 3B–3D), larger than 5 million Da in the nucleus, suggesting extensive protein-protein or protein-nucleic

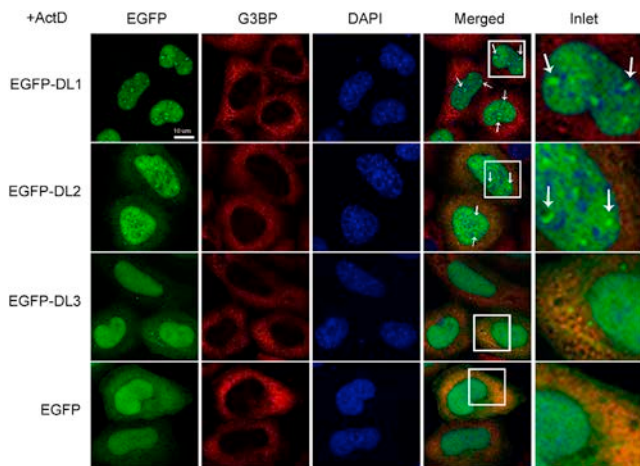


Figure 4. Transcription Inhibition Effects on hnRNPD Localization EGFP-DL1, EGFP-DL2, and EGFP-DL3 and unfused EGFP localization after their expression in HeLa^{DL-KO} cells and 5 μ g/ml Actinomycin D (ActD) treatment for 3 h. Cells were stained with G3BP antibody (red) as cytoplasmic marker and DAPI (blue) as nuclear marker.

acid interaction networks. This observation could provide an explanation for DL1 nuclear retention, which ultimately would depend on the presence of both IDRs.

ActD treatment disrupts the nucleolus (Figure S4B) and relocates DL1 and DL2 isoforms within the nucleus, now exhibiting a speckled pattern (Figure 4). Interestingly, DNA staining with DAPI also showed nuclear puncta, but they do not colocalize with DL1 and DL2 foci. Indeed, the DL1 and DL2 foci tend to coincide with nuclear areas exhibiting poor DAPI staining, suggesting that high local DNA concentrations might prevent DL1 and DL2 foci formation. In contrast to DL1 and DL2, DL3 remains diffusely distributed in the nucleus. Therefore, we can ascribe the observed DL1 and DL2 localization patterns in the absence of ActD to the presence of the C-terminal Tyr-rich IDR. Similar observations have been reported for FUS and TAF15 proteins (Marko et al., 2012; Zinszner et al., 1997).

The data show that only DL2 and DL3 isoforms can shuttle between nucleus and cytoplasm after transcription inhibition, whereas only DL1 and DL2 form nuclear foci. This indicates that the unique combination of IDRs in each particular variant is an important determinant of its localization.

hnRNPD L Alternative Splicing Isoforms Have Different Aggregation Propensity

Prion-like proteins are well known for their aggregation propensity, usually mediated by their PrLDs (March et al., 2016). hnRNPD L isoforms were highly insoluble in bacteria, which is why we studied them as His-SUMO fusions. Once purified, we proceeded to analyze their *in vitro* aggregation properties (Figure 5). We used the amyloid-specific dye Thioflavin-T (ThT) (Biancalana and Koide, 2010) to follow the kinetics of hnRNPD L isoform aggregation at 50- μ M protein concentration, 150 mM NaCl, and 37°C with agitation. Interestingly, after 2 days DL2 already exhibited a ThT signal in the plateau phase, DL1 started to bind ThT only after 3 days, and DL3 did not show any ThT

binding after 4 days (Figure 5A). We also used the alternative amyloid-specific dye Congo Red (CR) to confirm the presence of amyloid-like assemblies (Wu et al., 2012). The absorbance spectra of CR shifts in the presence of amyloid aggregates, and in agreement with the ThT results, only the DL2 isoform promoted this red shift in CR spectrum after 4 days (Figure 5B). These results suggest that, from the three hnRNPD L isoforms, only DL2 displays significant amyloid aggregation propensity. In fact, the DL2 aggregates also bind the amyloid dyes Thioflavin-S (ThS) and Proteostat, and exhibit CR birefringence under polarized light (Figure S6A). Finally, we observed by transmission electron microscopy (TEM) the morphological features of the three hnRNPD L isoforms after 4 days. Negative staining indicated that DL2 formed typical amyloid fibrillary structures without any significant accumulation of amorphous material, DL1 forms amorphous aggregates, and DL3 formed only small aggregates (Figure 5C).

These results suggest that the Tyr-rich IDR, including a predicted PrLD (Figure 6A), is responsible for hnRNPD L self-assembly into ordered amyloid-like structures, but also that this reaction only occurs in the absence of the positively charged N-terminal IDR, which would act as a kind of intramolecular chaperone. The data correlate with the relative solubility of the endogenous or transfected isoforms in HeLa WT or hnRNPD L KO cells, respectively, DL2 being the more insoluble of the three variants (Figure S3).

Disease-Causing Mutations Accelerate hnRNPD L Aggregation

Disease-causing mutations in PrLDs are common in prion-like proteins and they have been linked with an acceleration of the aggregation kinetics (Harrison and Shorter, 2017). Mutation of a specific Asp residue in the PrLDs of hnRNPA1 and hnRNPA2 proteins mapping to evolutionarily conserved regions of these IDRs are linked to ALS or MSP (Kim et al., 2013). The conserved Asp is involved in destabilizing electrostatic interactions and the removal of repulsion by mutation seems to be responsible for the increased propensity of the mutated PrLDs to self-associate and aggregate. Mutation of Asp378 of hnRNPD L to either Asn or His has been associated with LGMD1G (Berardo et al., 2019; Sun et al., 2019; Vieira et al., 2014). As in the case of hnRNPA1 and hnRNPA2, this Asp maps at the PrLD of hnRNPD L (Figure 6A; Navarro et al., 2015) and it is strictly conserved in vertebrates (Figure S7). Therefore, we examined how these mutations affect LLPS and the aggregation of hnRNPD L protein.

DL2 is reported to be the predominant isoform in tissues (Akagi et al., 2000), as observed by western blot of HeLa WT cells (Figure S3). Moreover, in the previous section we showed that DL2 is also the isoform with higher amyloid potential (Figure 5). Consequently, we focused the study on Asn and His mutants of the DL2 isoform (DL2N and DL2H) located at position D259; this position corresponds to D378 in DL1 because DL2 lacks the first 119 amino acids (Figure 6A).

We first checked how DL2 mutations affect LLPS behavior under physiological salt conditions (Figure 6B). In the absence of Ficoll, DL2H did not show detectable structures by light microscopy, whereas in DL2N small irregularly shaped and clustered structures were observed. In the presence of Ficoll, DL2H

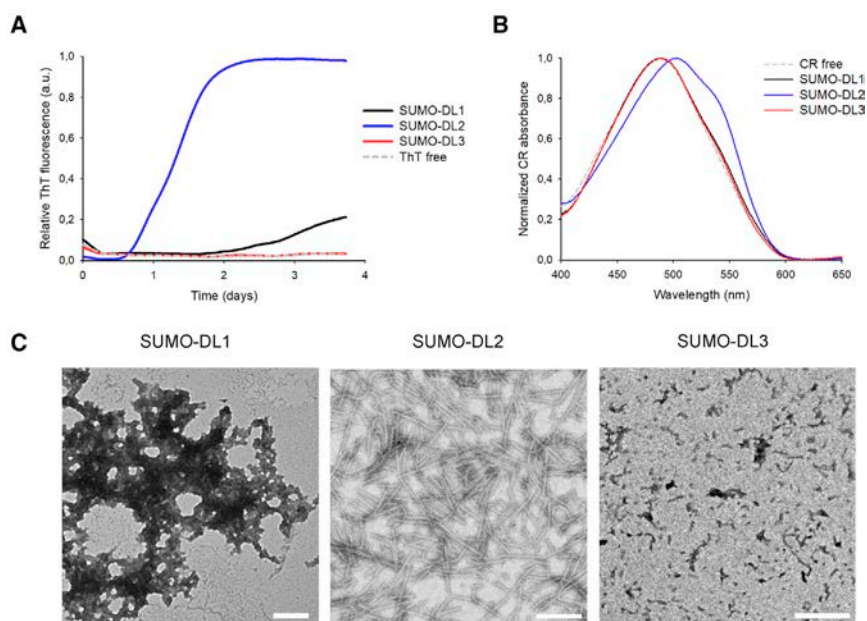


Figure 5. hnRNPD L Isoform Aggregation Propensity

Evaluation of Thioflavin T (ThT) binding over time (A), Congo Red (CR) binding (B), and transmission electron microscopy (TEM) (C) at final aggregation time point of 50- μ M SUMO-hnRNPD L isoforms in 50 mM HEPES pH 7.5 and 150 mM NaCl. Aggregation was conducted at 37°C and 600 rpm. Scale bars of TEM images in (C) represent 400 nm.

showed aggregated particles, whereas DL2N shows liquid droplets morphology similar to DL2 WT. These results indicate that the mutations in DL2 PrLD affect its LLPS propensity.

Comparison of DL2 mutants' *in vitro* aggregation propensity over time with that of the WT isoform confirmed the impact of the Asp mutation on aggregation. Both DL2 mutants exhibited faster aggregation kinetics, as monitored by ThT, displaying a shorter lag phase and reaching the plateau phase significantly before than the WT form (Figure 6C). Accordingly, TEM analysis indicates that DL2H and DL2N proteins have already assembled into amyloid fibrils after 24 h, whereas the DL2 WT remains protofibrillar (Figure 6D). This is in agreement with the predictions of ZipperDB, a structure-based threading algorithm that scores six-amino acid-sequence stretches according to their propensity to form "steric zippers" in the spine of amyloid fibrils (Goldschmidt et al., 2010). ZipperDB predicted higher steric zipper propensity for DL2H and DL2N, relative to DL2 WT (Table S1). The AMYCO algorithm, which evaluates the impact of mutations on the aggregation of prion-like proteins (Iglesias et al., 2019), also predicts increases in aggregation propensities of 7% and 10% for the DL2H and DL2N variants, respectively.

The acceleration of the aggregation reaction in the mutant variants resulted in smaller and less-ordered amyloid fibrils at the end of the reaction (4 days), as observed by TEM (Figure 6D), which displayed a lower, but still significant, binding to CR than DL2 WT fibrils (Figure 6E). The amyloid-like nature of the DL2H and DL2N aggregates was further confirmed by staining with ThS, Proteostat, and by CR birefringence (Figure S6A).

The secondary structure content of the aggregates was assessed using attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIR) in the amide I region of the spectrum (1,700–1,600 cm^{-1}). Deconvolution of the FTIR

absorbance spectra of DL2 WT, DL2H, and DL2N samples indicated the presence of a major band at 1,628 cm^{-1} , which can be assigned to the presence of an intermolecular β sheet (Figure S6B). This component accounts for 25%–26% of the spectral area in the three cases (Table S2), which sharply coincides with the proportion of residues mapping at the PrLD in the respective DL2 constructs (25.8%). No major secondary structure differences were observed between the

three amyloid assemblies, indicating that, *in vitro*, the impact of the mutations is mostly kinetic.

Overall, these results, provide evidences that mutation of Asp378 in hnRNPD L PrLD (position 259 in DL2) increases its aggregation propensity, as previously described for hnRNPA1 and hnRNPA2.

Disease-Causing Mutations Are Located in the Insoluble Fraction of *Drosophila* Muscle

To evaluate the effect of disease-causing mutations in hnRNPD L *in vivo*, we generated transgenic *Drosophila* expressing WT or mutant forms of human DL2 by using the UAS/GAL4 system. Multiple transgenic lines expressing a single copy of DL2 WT or mutant (DL2N or DL2H) were generated by site-specific ϕ C31 integrase-mediated transgenesis into *Drosophila* chromosome 3. This approach permits equal levels of expression between independent lines, as demonstrated empirically for the five independent lines expressing WT DL2 (Figure 7A). However, we observed that fly lines expressing mutant forms of DL2 (DL2N or DL2H) consistently exhibited lower levels of DL2 protein compared to flies expressing WT DL2 (Figure 7A).

Expression of either WT or mutant (DL2N or DL2H) forms of DL2 in *Drosophila* indirect flight muscle led to mild degeneration, showing disorganized muscle fibers as demonstrated by phalloidin staining (Figure 7B). Immunohistochemical analysis showed that both WT and mutant forms of DL2 localized predominantly to myonuclei (Figure 7B). Although we did not observe clear differences between WT and mutant proteins in their localization patterns, we did observe that the solubility of DL2 protein was strongly impacted by disease-causing mutations. Specifically, mutant DL2 (D259H or D259N) proteins were largely recovered from the detergent-insoluble fraction, whereas the WT DL2 protein is found both in the detergent-soluble and detergent-insoluble fractions (Figure 7C). This result is

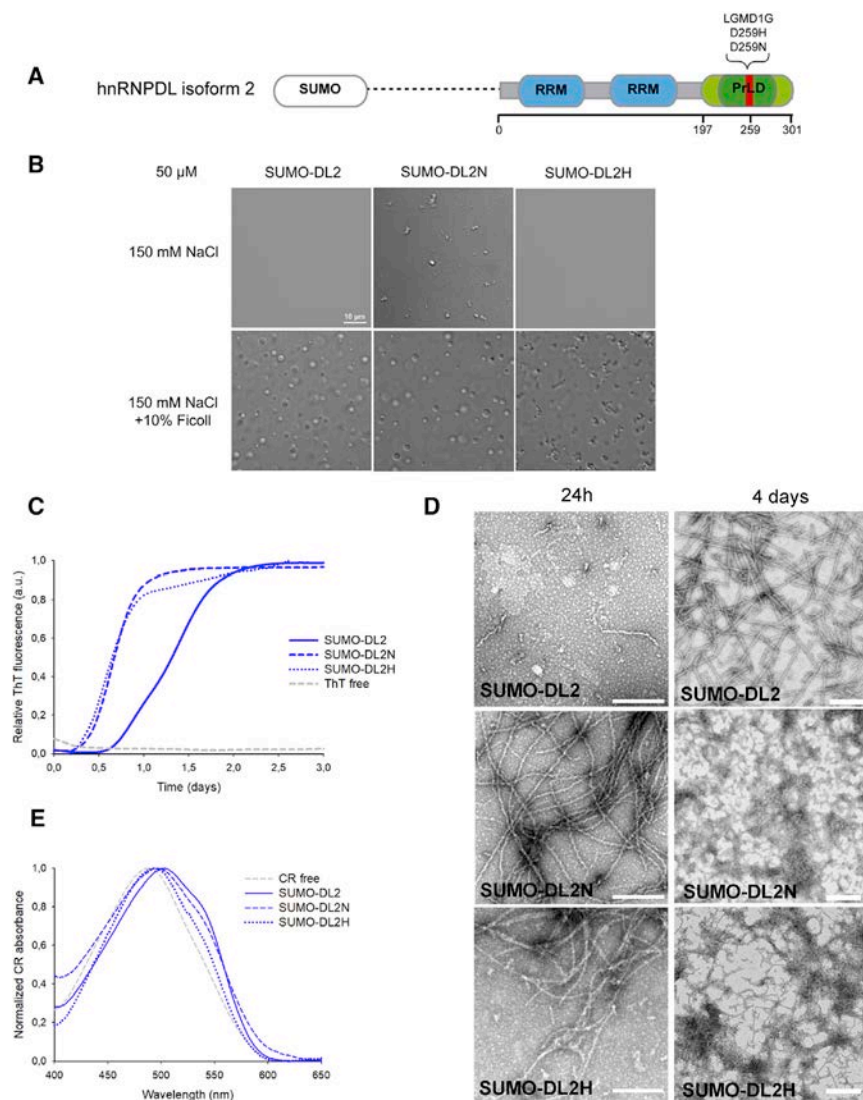


Figure 6. DL2 Disease-Causing Mutation Aggregation Propensity

(A) Schematic diagram of DL2, as a SUMO fusion construct. RRM (blue) according to Pfam (El-Gebali et al., 2019) and PLAAC (Lancaster et al., 2014) predicted PrLD (light green) are indicated. The Tyr-rich IDR (dark green) and the disease-causing mutations (red) within the PrLD are shown with their respective amino acid positions.

(B–E) LLPS behavior (B), Thioflavin T (ThT) binding over time (C), transmission electron microscopy (TEM) at 24 h and 4 days (D), and Congo Red (CR) binding (E) at final aggregation time point of 50- μ M DL2 and the disease-causing mutations D259N and D259H (DL2N and DL2H) in 50 mM HEPES pH 7.5 and 150 mM NaCl. Aggregation was conducted at 37°C and 600 rpm. Scale bars of TEM images in (D) represent 400 nm.

actions (Boeynaems et al., 2018). Recently, it was described that FUS LLPS was governed by interactions between Tyr residues of the PrLD and Arg residues of the RNA binding domain (Wang et al., 2018). We hypothesized that hnRNPDLLPS would follow the same mechanism than FUS, where Tyr residues from the C terminus and Arg residues from the N terminus would mediate interactions leading to phase separation. We experimentally validated this hypothesis generating hnRNPDLL variants that maintain the aromatic and cationic residue content but are predicted to establish weaker interactions. These variants could not undergo LLPS under physiological conditions, indicating that, in hnRNPDLL, Arg and Tyr residue interactions are required for phase separation. The absence of the N-terminus Arg-rich

consistent with *in vitro* data that indicate enhanced aggregation propensity by disease-causing mutations (Figure 6C). The lower levels of DL2 mutant proteins relative to the WT protein in fly muscle might respond to an attempt of the proteostatic machinery to degrade misassembled insoluble species.

DISCUSSION

In this work, we first characterized the molecular properties that govern the self-assembly of hnRNPDLL isoforms *in vitro*. The three AS isoforms exhibit different LLPS propensities according to their IDRs composition: DL1, with both Arg-rich and Tyr-rich IDRs, displays a strong LLPS behavior; DL2, with only the Tyr-rich IDR, has a mild LLPS propensity, requiring a crowding agent to phase separate at physiological ionic strength; and DL3, with none of the IDRs, does not phase separate. Protein self-assembly driving forces can be mediated by π -cation, π - π , hydrophobic, electrostatic, dipole-dipole, or hydrogen bonding inter-

IDR significantly reduces the DL2 LLPS propensity. However, DL2 can still undergo LLPS, most probably by Tyr-Tyr interactions of the C-terminal PrLD, as it occurs with purified PrLDs of proteins such as FUS (Wang et al., 2018).

Different self-assembly properties of the hnRNPDLL isoforms suggest that they might participate in distinct biological functions. We experimentally validated that hnRNPDLL isoforms exhibit a significantly different behavior in mammalian cells, the data indicating that they might be involved in the formation of distinct nuclear complexes. DL1 and DL2 are excluded from the nucleolus, while DL3 is diffusely distributed. This exclusion from the nucleolus has been also reported for other prion-like proteins like FUS (Yang et al., 2014), TAF15 (Marko et al., 2012), or EWSR1 (Tannukit et al., 2008). The only common region in DL1 and DL2 and absent in DL3 is the C-terminus Tyr-rich PrLD. Therefore, hnRNPDLL PrLD seems to determine the intranuclear compartmentalization of the isoforms. This is consistent with the observation that the deletion of the N-terminal PrLD of

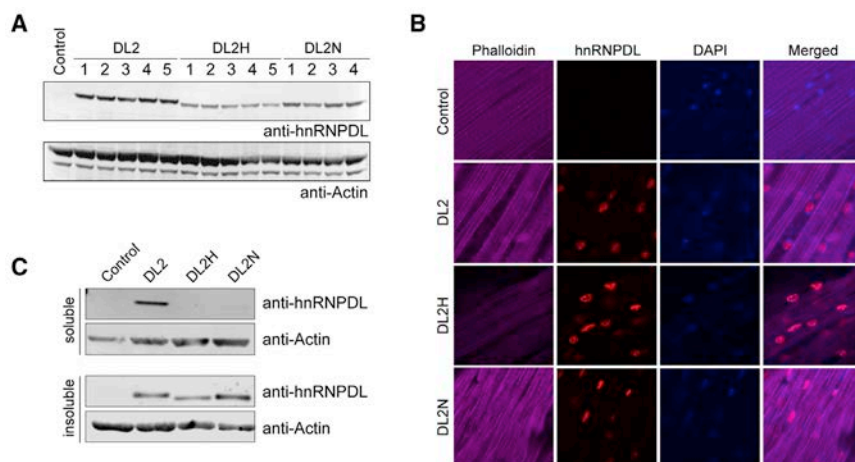


Figure 7. hnRNPD Isoform 2 Disease-Causing Mutation Effects in *Drosophila*

(A) Expression levels of DL2 and disease-causing mutations (DL2N and DL2H) in transgenic flies. Thoraxes of adult flies were processed for western blot analysis with an antibody against hnRNPD. Actin was blotted as a loading control.

(B) Adult flies were dissected to expose the dorsal longitudinal indirect fly muscle (DLM) and stained with Alexa Fluor 647-phalloidin (purple), hnRNPD (red), and DAPI (blue).

(C) Thoraxes of adult flies were dissected. Sequential extractions were performed to examine the solubility profile of hnRNPD.

FUS resulted in a protein variant evenly distributed in the entire nucleus (Yang et al., 2014), as it occurs with the natural DL3 isoform. hnRNPDs have LLPS propensity, but in the nucleus it is diffusely distributed, without the detection of nuclear puncta corresponding to MLOs. It has been shown that high RNA/protein ratios, as occur in the nucleus, prevent prion-like proteins from LLPS (Maharana et al., 2018). This could be the underlying reason of the diffuse distribution of the DL1 and DL2 variants in the nucleus. Accordingly, we experimentally validated that DL1 LLPS is strongly dependent on RNA levels. We were also interested in assessing whether hnRNPD isoforms exhibit different mobilities in cells due to their differential self-assembly properties. In excellent agreement with *in vitro* behavior, DL2 and DL3 form dynamic and small complexes in the nucleus, whereas DL1 forms more rigid and bigger complexes. This last behavior can be univocally ascribed to the DL1 Arg-rich N terminus IDR, which increases DL1 multivalency thus favoring protein-protein and protein-nucleic acids interactions and, consequently, the formation of the observed large complexes.

We further validated that DL2 and DL3 are the only two isoforms able to shuttle between the nucleus and cytoplasm, their nuclear import being dependent on ongoing transcription, as previously reported (Kamei and Yamada, 2002; Kawamura et al., 2002). Similar differences in isoform shuttling have been reported for hnRNPD, a member of the same heterogeneous ribonucleoprotein (hnRNP) family, where p37 and p40 isoforms shuttle between nucleus and cytoplasm but p45 and p42 isoforms remain in the nucleus (Arao et al., 2000). hnRNPD p37 and p40 result from AS and lack exon 7, which encodes for a region of 50 residues corresponding to the only Tyr-rich IDR region (28% Tyr) in its PrLD. The p45 and p42 isoforms' nuclear retention is suggested to be mediated through interaction with the nuclear SAF-B protein, p37 and p40 not interacting with it. SAF-B contains a large Arg-rich IDR region (26% Arg) and therefore it is plausible that π -cation contacts could contribute to the interactions with the p45 and p42 isoforms. DL1 possesses both types of IDRs and therefore can establish a potentially larger number of interactions. The fact that the protein is retained in the nucleus when RNA levels are decreased by ActD treatment suggests that these interactions are of proteic nature.

DL1 and DL2 form bright nuclear foci after RNA synthesis inhibition, while DL3 does not. This necessarily involves the Tyr-rich PrLD in the process. Interestingly, similar observations were described for hnRNPD, TAF15 and FUS after ActD treatment (Arao et al., 2000; Jobert et al., 2009; Marko et al., 2012; Zinszner et al., 1997). For the three proteins the formation of nuclear foci was depended on their PrLD (Marko et al., 2012; Zinszner et al., 1997). Nuclear foci after ActD may represent the retention of the protein in transient subnuclear compartments by interaction with proteins recruited to this site (Zinszner et al., 1997) thanks to the characteristic compositional bias of PrLDs.

Prion-like proteins are well known for their aggregation behavior and involvement in disease. Therefore, we evaluated the aggregation propensity of hnRNPD isoforms. Not surprising, DL2, the predominant isoform in humans, hence the most probable isoform being involved in disease, is the isoform with higher amyloidogenic propensity. The data indicate that the amyloidogenic potential concentrates in the PrLD, since DL3, containing the two RRM domains, remains soluble. A reasonable explanation for the amyloidogenicity of DL2 is that it lacks the N-terminus IDR Arg-rich domain, which may act as a mechanism of protection in DL1, precluding rapid aggregation through electrostatic repulsion between positively charged regions. Therefore, this domain plays a different role in LLPS and amyloid formation, promoting the first reaction and inhibiting the second one. Importantly, this suggests that, for hnRNPD, the process of phase separation plays a protective role against fibril formation.

As observed for other prion-like proteins, disease-causing mutations in the PrLD enhance the aggregation propensity of hnRNPD. We experimentally demonstrated a significant acceleration of DL2 aggregation after introduction of the D378N and D378H LGMD1G-associated mutations (DL1 nomenclature). In *Drosophila* muscle, hnRNPD mutants locate in the nucleus but show a clear reduction in solubility compared to the WT protein. Most pathogenic mutations in prion-like proteins are reported to promote the formation of cytoplasmic inclusions in patient tissues, a phenotype that can be recapitulated in cell-based assays and in animal models of the disease (Kim et al., 2013). This is not the case for hnRNPD. Analysis of muscle

tissues from LGMD1G patients bearing the hnRNPD L genetic mutations we studied here did not revealed the existence of congophilic deposits or any type of inclusions (Sun et al., 2019; Vieira et al., 2014), nor did we observe them in *Drosophila*. The mechanism by which hnRNPD L mutations elicits the LGMD1G phenotype is still unknown. Here, we uncover that DL2 is inherently aggregation prone and that the disease-causing mutations exacerbate this propensity, resulting in a protein that despite being diffusely distributed in the nucleus of *Drosophila* myocytes is totally insoluble. This evidence makes us think that LGMD1G is caused by a loss-of-function mechanism, in which DL2 function is lost because the protein aggregates or it is degraded by the cellular protein quality control machinery, in line with the lower levels of mutant proteins detected in *Drosophila*. Consistent with this view, blocking the translation of hnRNPD L in zebrafish to reduce the endogenous protein levels results in restricted and uncoordinated animal movements and disorganized myofibers (Vieira et al., 2014). Thus, it is feasible that LGMD1G patients would exhibit lower levels of soluble DL2 isoform in their muscular tissue, although this extent should be further confirmed. Importantly, like in the case of hnRNPA2 and hnRNPA1, aggregation in hnRNPD L is triggered by the mutation of a well-conserved Asp residue in the PrLD. This recurrent mutational change within three different hnRNP proteins indicates that these residues play a protective role against aggregation, likely by promoting electrostatic repulsion among side chains of neighboring monomers (Murray et al., 2018).

hnRNPD L protein is an RNA-binding protein involved in transcription and AS regulation (Li et al., 2019). The presence of the exon corresponding to its PrLD is responsible for the incorporation of hnRNPD L into multivalent hnRNP assemblies that, in turn, control the AS of other genes (Gueroussov et al., 2017). A recent study indicates that SRSF1 protein regulates AS of hnRNPD L, generating the isoforms containing the PrLD, and therefore the execution of the downstream splicing program (Feng et al., 2019). It was suggested that SRSF1 may serve as an upstream regulator of phase separation. We demonstrate here that this is the case. Indeed, we speculate that by regulating hnRNPD L AS, SRSF1 might also determine the size and dynamics of the complexes this protein forms in the nucleus, its nucleus-cytoplasm shuttling ability, and its amyloidogenic potential.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Bacterial Cell Culture
 - Mammalian cell Culture
 - *Drosophila* Culture
- **METHOD DETAILS**
 - Isoforms sequence
 - Bacterial molecular cloning
 - Protein expression and purification

- Phase separation
- Dynamic light scattering (DLS)
- Protein aggregation
- Transmission Electron Microscopy (TEM)
- Congo Red (CR) binding
- Thioflavin-T (Th-T) aggregation kinetics assay
- Thioflavin-S (ThS) and proteostat® staining
- Congo Red (CR) birefringence
- Fourier transform infrared (FT-IR) spectroscopy
- Mammalian molecular cloning
- Mammalian cell culture and transfection
- Immunofluorescence
- Fluorescence recovery after photobleaching (FRAP)
- Size exclusion chromatography (SEC)
- *Drosophila* stocks and culture
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND CODE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2019.12.080>.

ACKNOWLEDGMENTS

We are grateful to Prof. T. Mittag, Dr. C. Mathieu, and Dr. M. Purice from St. Jude Children's Research Hospital, USA, and UAB microscopy service of Barcelona for their technical advice. We acknowledge Leticia Nogueira, Antonio Fernando Ribeiro Jr., Mariz Vainzof, and Mayana Zatz from the Human Genome and Stem Cell Center, Biosciences Institute, University of São Paulo, Brazil, for their help with muscle samples. S.V. acknowledges funding from MINECO (BIO2016-78310-R) and ICREA (ICREA-Academia 2016). J.P.T. acknowledges the Howard Hughes Medical Institute, National Institutes of Health (R35NS097974), and the St. Jude Research Collaborative on Membrane-less Organelles. X.S. acknowledges funding from Obra Social "la Caixa," AGAUR (2017 SGR 324), MINECO (BIO2012-31043 and BIO2015-70092-R), and the European Research Council (CONCERT, contract number 648201). IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO (government of Spain). C.B. acknowledges funding from "Ministerio de Educación y Formación Profesional."

AUTHOR CONTRIBUTIONS

C.B. designed and performed the experiments, analyzed the data, and wrote the manuscript. P.Y., M.C., J.M., M.P., and E.S. contributed to specific experiments. P.Y. supervised the experiments. H.J.K. designed the *Drosophila* experiments. M.C. and H.J.K. wrote the *Drosophila* section and composed Figure 7. P.Y., M.P., E.S., X.S., H.J.K., and J.P.T. edited the manuscript. J.P.T. and S.V. designed the project. S.V. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 5, 2019
 Revised: November 12, 2019
 Accepted: December 19, 2019
 Published: January 28, 2020

REFERENCES

Akagi, T., Kamei, D., Tsuchiya, N., Nishina, Y., Horiguchi, H., Matsui, M., Kamma, H., and Yamada, M. (2000). Molecular characterization of a mouse

- heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its tissue-specific expression. *Gene* 245, 267–273.
- Arao, Y., Kuriyama, R., Kayama, F., and Kato, S. (2000). A nuclear matrix-associated factor, SAF-B, interacts with specific isoforms of AUF1/hnRNP D. *Arch. Biochem. Biophys.* 380, 228–236.
- Babinchak, W.M., Haider, R., Dumm, B.K., Sarkar, P., Surewicz, K., Choi, J.K., and Surewicz, W.K. (2019). The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.* 294, 6306–6317.
- Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451.
- Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., et al.; UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212.
- Berardo, A., Lornage, X., Johari, M., Evangelista, T., Cejas, C., Barroso, F., Dubrovsky, A., Bui, M.T., Brochier, G., Saccoliti, M., et al. (2019). HNRNPDL-related muscular dystrophy: expanding the clinical, morphological and MRI phenotypes. *J. Neurol.* 266, 2524–2534.
- Biancalana, M., and Koide, S. (2010). Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta* 1804, 1405–1412.
- Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., Guharoy, M., De Decker, M., Jaspers, T., Ryan, V.H., et al. (2017). Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol. Cell* 65, 1044–1055.e5.
- Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., et al. (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* 28, 420–435.
- Bogaert, E., Boeynaems, S., Kato, M., Guo, L., Caulfield, T.R., Steyaert, J., Scheveneels, W., Wilmans, N., Haeck, W., Hersmus, N., et al. (2018). Molecular Dissection of FUS Points at Synergistic Effect of Low-Complexity Domains in Toxicity. *Cell Rep.* 24, 529–537.e4.
- Boundedjah, O., Desforges, B., Wu, T.D., Pioche-Durieu, C., Marco, S., Hamon, L., Curmi, P.A., Guerquin-Kern, J.L., Piétrement, O., and Pastré, D. (2014). Free mRNA in excess upon polysome dissociation is a scaffold for protein multimerization to form stress granules. *Nucleic Acids Res.* 42, 8678–8691.
- Brady, J.P., Farber, P.J., Sekhar, A., Lin, Y.H., Huang, R., Bah, A., Nott, T.J., Chan, H.S., Baldwin, A.J., Forman-Kay, J.D., and Kay, L.E. (2017). Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. USA* 114, E8194–E8203.
- Burke, K.A., Janke, A.M., Rhine, C.L., and Fawzi, N.L. (2015). Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* 60, 231–241.
- Das, R.K., and Pappu, R.V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* 110, 13392–13397.
- Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* 3, 195–205.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47 (D1), D427–D432.
- Feng, H., Bao, S., Rahman, M.A., Weyn-Vanhenryck, S.M., Khan, A., Wong, J., Shah, A., Flynn, E.D., Krainer, A.R., and Zhang, C. (2019). Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. *Mol. Cell* 74, 1189–1204.e6.
- Goldschmidt, L., Teng, P.K., Riek, R., and Eisenberg, D. (2010). Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA* 107, 3487–3492.
- Guerousov, S., Weatheritt, R.J., O'Hanlon, D., Lin, Z.Y., Narula, A., Gingras, A.C., and Blencowe, B.J. (2017). Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell* 170, 324–339.e23.
- Harrison, A.F., and Shorter, J. (2017). RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417–1438.
- Hu, X., Chen, X., Wu, B., Soler, I.M., Chen, S., and Shen, Y. (2017). Further defining the critical genes for the 4q21 microdeletion disorder. *Am. J. Med. Genet. A.* 173, 120–125.
- Iglesias, V., Conchillo-Sole, O., Batlle, C., and Ventura, S. (2019). AMYCO: evaluation of mutational impact on prion-like proteins aggregation propensity. *BMC Bioinformatics* 20, 24.
- Ito, D., Hatano, M., and Suzuki, N. (2017). RNA binding proteins and the pathological cascade in ALS / FTD neurodegeneration. *Sci. Transl. Med.* 9, eaah5436.
- Jobert, L., Pinzón, N., Van Herreweghe, E., Jády, B.E., Guialis, A., Kiss, T., and Tora, L. (2009). Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep.* 10, 494–500.
- Kamei, D., and Yamada, M. (2002). Interactions of heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its domains with high-affinity binding sites. *Gene* 298, 49–57.
- Kamei, D., Tsuchiya, N., Yamazaki, M., Meguro, H., and Yamada, M. (1999). Two forms of expression and genomic structure of the human heterogeneous nuclear ribonucleoprotein D-like JKTBP gene (HNRNPDL). *Gene* 228, 13–22.
- Kato, M., Yang, Y.-S., Sutter, B.M., Wang, Y., McKnight, S.L., and Tu, B.P. (2019). Redox State Controls Phase Separation of the Yeast Ataxin-2 Protein via Reversible Oxidation of Its Methionine-Rich Low-Complexity Domain. *Cell* 177, 711–721.e8.
- Kawamura, H., Tomozoe, Y., Akagi, T., Kamei, D., Ochiai, M., and Yamada, M. (2002). Identification of the nucleocytoplasmic shuttling sequence of heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its interaction with mRNA. *J. Biol. Chem.* 277, 2732–2739.
- Kim, H.J., Kim, N.C., Wang, Y.-D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 495, 467–473.
- King, O.D., Gitler, A.D., and Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* 1462, 61–80.
- Lancaster, A.K., Nutter-Upham, A., Lindquist, S., and King, O.D. (2014). PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* 30, 2501–2502.
- Li, R.Z., Hou, J., Wei, Y., Luo, X., Ye, Y., and Zhang, Y. (2019). hnRNPDL extensively regulates transcription and alternative splicing. *Gene* 687, 125–134.
- Liewluck, T., and Milone, M. (2018). Untangling the complexity of limb-girdle muscular dystrophies. *Muscle Nerve* 58, 167–177.
- Liu, Y., Zhu, X., Zhu, J., Liao, S., Tang, Q., Liu, K., Guan, X., Zhang, J., and Feng, Z. (2007). Identification of differential expression of genes in hepatocellular carcinoma by suppression subtractive hybridization combined cDNA microarray. *Oncol. Rep.* 18, 943–951.
- Mackenzie, I.R., Nicholson, A.M., Sarkar, M., Messing, J., Purice, M.D., Pottier, C., Annu, K., Baker, M., Perkerson, R.B., Kurti, A., et al. (2017). TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics. *Neuron* 95, 808–816.e9.
- Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovskiy, A., Poser, I., Bickle, M., Guillén-boixet, J., Franzmann, T., Jahnel, M., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* 360, 918–921.
- March, Z.M., King, O.D., and Shorter, J. (2016). Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* 1647, 9–18.
- Marko, M., Vlassis, A., Guialis, A., and Leichter, M. (2012). Domains involved in TAF15 subcellular localisation: dependence on cell type and ongoing transcription. *Gene* 506, 331–338.
- Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase separation by low complexity domains

- promotes stress granule assembly and drives pathological fibrillization. *Cell* 163, 123–133.
- Murray, D.T., Zhou, X., Kato, M., Xiang, S., Tycko, R., and McKnight, S.L. (2018). Structural characterization of the D290V mutation site in hnRNP A2 low-complexity-domain polymers. *Proc. Natl. Acad. Sci. USA* 115, E9782–E9791.
- Murthy, A.C., Dignon, G.L., Kan, Y., Zerze, G.H., Parekh, S.H., Mittal, J., and Fawzi, N.L. (2019). Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* 26, 637–648.
- Navarro, S., Marinelli, P., Diaz-Caballero, M., and Ventura, S. (2015). The prion-like RNA-processing protein HNRPD L forms inherently toxic amyloid-like inclusion bodies in bacteria. *Microb. Cell Fact.* 14, 102.
- Nedelsky, N.B., and Taylor, J.P. (2019). Bridging biophysics and neurology: aberrant phase transitions in neurodegenerative disease. *Nat. Rev. Neurol.* 15, 272–286.
- Nigro, V., and Savarese, M. (2014). Genetic basis of limb-girdle muscular dystrophies: the 2014 update. *Acta Myol.* 33, 1–12.
- Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoyanov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162, 1066–1077.
- Peskett, T.R., Rau, F., O'Driscoll, J., Patani, R., Lowe, A.R., and Saibil, H.R. (2018). A Liquid to Solid Phase Transition Underlying Pathological Huntingtin Exon1 Aggregation. *Mol. Cell* 70, 588–601.e6.
- Piñol-Roma, S., and Dreyfuss, G. (1991). Transcription-dependent and transcription-independent nuclear transport of hnRNP proteins. *Science* 253, 312–314.
- Piñol-Roma, S., and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* 355, 730–732.
- Ryan, V.H., Dignon, G.L., Zerze, G.H., Chabata, C.V., Silva, R., Conicella, A.E., Amaya, J., Burke, K.A., Mittal, J., and Fawzi, N.L. (2018). Mechanistic View of hnRNP A2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol. Cell* 69, 465–479.e7.
- Straub, V., Murphy, A., and Udd, B.; LGMD workshop study group (2018). 229th ENMC international workshop: Limb girdle muscular dystrophies – Nomenclature and reformed classification Naarden, the Netherlands, 17–19 March 2017. *Neuromuscul. Disord.* 28, 702–710.
- Su, H., Kodiha, M., Lee, S., and Stochaj, U. (2013). Identification of novel markers that demarcate the nucleolus during severe stress and chemotherapeutic treatment. *PLoS ONE* 8, e80237.
- Sun, Y., Chen, H., Lu, Y., Duo, J., Lei, L., OuYang, Y., Hao, Y., Da, Y., and Shen, X.M. (2019). Limb girdle muscular dystrophy D3 HNRNPDL related in a Chinese family with distal muscle weakness caused by a mutation in the prion-like domain. *J. Neurol.* 266, 498–506.
- Tannukit, S., Wen, X., Wang, H., and Paine, M.L. (2008). TFIP11, CCN1 and EWSR1 Protein-protein Interactions, and Their Nuclear Localization. *Int. J. Mol. Sci.* 9, 1504–1514.
- Tsuchiya, N., Kamei, D., Takano, A., Matsui, T., and Yamada, M. (1998). Cloning and characterization of a cDNA encoding a novel heterogeneous nuclear ribonucleoprotein-like protein and its expression in myeloid leukemia cells. *J. Biochem.* 123, 499–507.
- Vernon, R.M., Chong, P.A., Tsang, B., Kim, T.H., Bah, A., Farber, P., Lin, H., and Forman-Kay, J.D. (2018). Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* 7, e31486.
- Vieira, N.M., Naslavsky, M.S., Licinio, L., Kok, F., Schlesinger, D., Vainzof, M., Sanchez, N., Kitajima, J.P., Gal, L., Cavaçana, N., et al. (2014). A defect in the RNA-processing protein HNRPD L causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum. Mol. Genet.* 23, 4103–4110.
- Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovskiy, A., Drechsel, D., et al. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 174, 688–699.e16.
- Wegmann, S., Eftekhazadeh, B., Tepper, K., Zoltowska, K.M., Bennett, R.E., Dujardin, S., Laskowski, P.R., MacKenzie, D., Kamath, T., Commins, C., et al. (2018). Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J.* 37, e98049.
- Wu, C., Scott, J., and Shea, J.E. (2012). Binding of Congo red to amyloid protofibrils of the Alzheimer A β (9–40) peptide probed by molecular dynamics simulations. *Biophys. J.* 103, 550–557.
- Wu, Y.Y., Li, H., Lv, X.Y., Wei, Q., Li, X., Liu, X.Y., Zhou, Q., and Wei, Y.Q. (2008). Overexpression of JKTBP1 induces androgen-independent LNCaP cell proliferation through activation of epidermal growth factor-receptor (EGF-R). *Cell Biochem. Funct.* 26, 467–477.
- Yang, L., Gal, J., Chen, J., and Zhu, H. (2014). Self-assembled FUS binds active chromatin and regulates gene transcription. *Proc. Natl. Acad. Sci. USA* 111, 17809–17814.
- Ying, Y., Wang, X.J., Vuong, C.K., Lin, C.H., Damianov, A., and Black, D.L. (2017). Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell* 170, 312–323.e10.
- Zhang, P., Ji, D., Hu, X., Ni, H., Ma, W., Zhang, X., Liao, S., Zeng, Z., Zhao, Y., and Zhou, H. (2018). Oncogenic heterogeneous nuclear ribonucleoprotein D-like promotes the growth of human colon cancer SW620 cells via its regulation of cell-cycle. *Acta Biochim. Biophys. Sin. (Shanghai)* 50, 880–887.
- Zhang, T., Delestienne, N., Huez, G., Krays, V., and Gueydan, C. (2005). Identification of the sequence determinants mediating the nucleo-cytoplasmic shuttling of TIAR and TIA-1 RNA-binding proteins. *J. Cell Sci.* 118, 5453–5463.
- Zhou, H., Ge, Y., Sun, L., Ma, W., Wu, J., Zhang, X., Hu, X., Eaves, C.J., Wu, D., and Zhao, Y. (2014). Growth arrest specific 2 is up-regulated in chronic myeloid leukemia cells and required for their growth. *PLoS ONE* 9, e86195.
- Zinszner, H., Immanuel, D., Yin, Y., Liang, F.X., and Ron, D. (1997). A topogenic role for the oncogenic N-terminus of TLS: nucleolar localization when transcription is inhibited. *Oncogene* 14, 451–461.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal to hnRNPD	Abcam	Cat#ab183136
Mouse monoclonal anti G3BP	BD Biosciences	Cat#611127; RRID: AB_398438
Alexa Fluor 555	Molecular Probes	Cat#A31572; RRID: AB_162543
Mouse monoclonal anti GAPDH	Santa Cruz	Cat#sc47724; RRID: AB_627678
Sheep polyclonal anti tubulin	Cytoskeleton Inc	Cat#ATN02; RRID:AB_10709401
Mouse monoclonal anti C23	Santa Cruz	Cat#sc-8031; RRID: AB_670271
Bacterial and Virus Strains		
One Shot TOP10 chemically competent <i>E. coli</i>	Life Technologies	Cat#C404003
HI-Control CL21(DE3) chemically competent cells (SOLOS)	Lucigen Corporation	Cat#60435-1
Chemicals, Peptides, and Recombinant Proteins		
1x TrypLE Express	Thermo Fisher Scientific	Cat #12604-013
Lipofectamine2000	Thermo Fisher Scientific	Cat# 11668027
His-SUMO-hnRNPD isoform 1	This study	N/A
His-SUMO-hnRNPD isoform 2	This study	N/A
His-SUMO-hnRNPD isoform 3	This study	N/A
His-SUMO-hnRNPD isoform 2 D259N	This study	N/A
His-SUMO-hnRNPD isoform 2 D259H	This study	N/A
Oregon Green Protein Labeling Kit	Molecular Probes	Cat #O10241
Texas Red Protein Labeling Kit	Molecular Probes	Cat#T10244
Q5 site directed mutagenesis kit	New England Biolabs	Cat#E0554S
Expresso T7 SUMO cloning and expression system	Lucigen Corporation	Cat#49003-2
NEBuilder HiFi DNA Assembly Master Mix	New England Biolabs	Cat#E2621S
Ficoll400	Sigma-Aldrich	Cat#F2637
Secure seal imaging spacers 1.9x0.12 mm	Grace Biolabs	Cat#654002
Proteostat® aggresome detection kit	Enzo life sciences	Cat#ENZ-51035-K100
Experimental Models: Cell Lines		
HeLa cells	ATCC	Cat #CCL-2; RRID: CVCL_0030
HeLa hnRNPD <i>knock out</i> cells	This study	N/A
Experimental Models: Organisms/Strains		
<i>Drosophila melanogaster</i> : cell line w1118	Bloomington Drosophila Stock Center	Cat#3605
<i>Drosophila melanogaster</i> : cell line Mhc-Gal4	Bloomington Drosophila Stock Center	Cat#55133
<i>Drosophila melanogaster</i> : cell line UAS-hnRNPD isoform 2	Bestgene Inc.	N/A
<i>Drosophila melanogaster</i> : cell line UAS-hnRNPD isoform 2 D259N	Bestgene Inc.	N/A
<i>Drosophila melanogaster</i> : cell line UAS-hnRNPD isoform 2 D259H	Bestgene Inc.	N/A
Oligonucleotides		
See Table S3 for the primer list		
Recombinant DNA		
pET28a-hnRNPD isoform 1	GenScript	N/A
pETite-His-SUMO Kan vector	Lucigen Corporation	Cat#49003-1
pETite-His-SUMO-hnRNPD isoform 1	This study	N/A
pETite-His-SUMO-hnRNPD isoform 2	This study	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pETite-His-SUMO-hnRNPD L isoform 3	This study	N/A
pETite-His-SUMO-hnRNPD L isoform 2 D259N	This study	N/A
pETite-His-SUMO-hnRNPD L isoform 2 D259H	This study	N/A
pEGFP-C3 vector	Clontech	Cat#6082-1
pEGFP-C3-hnRNPD L isoform 1	This study	N/A
pEGFP-C3-hnRNPD L isoform 2	This study	N/A
pEGFP-C3-hnRNPD L isoform 3	This study	N/A
pSpCas9(BB)-2A-Puro (PX459) V2.0	Addgene	Cat# 62988; RRID:Addgene_62988
pUASTattB	Drosophila Genomics Resource Center	Cat#1419
Software and Algorithms		
SigmaPlot 10.0	Systal Software	http://www.sigmaplot.co.uk
GraphPad Prism 5	GraphPad	https://www.graphpad.com/
ImageJ 1.51J Software	NIH	https://imagej.nih.gov/ij/download.html
Igor Pro 7.0	Wavemetrics	https://www.wavemetrics.com/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagent should be directed to and will be fulfilled by the Lead Contact, S. Ventura (Salvador.ventura@uab.es).

All unique/stable reagents generated in this study will be made available on request but we may require a payment and/or a completed Materials Transfer Agreement if there is potential for commercial application.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial Cell Culture

cDNA clones were transformed into One Shot TOP10 and BL21(DE3) SOLOS chemically competent *E. coli* cells (Thermo Fisher Scientific). Single colonies were grown overnight at 37°C and 220 rpm in LB media containing kanamycin selection antibiotic at a concentration of 50 mg/ml. All competent bacterial cells were stored at -80°C in 15% glycerol.

Mammalian cell Culture

HeLa cells (of female origin) were grown at 37°C and 5% CO₂, and maintained in DMEM High Glucose (Hyclone SH30022.01) medium supplemented with 10% fetal bovine serum. Cells were passaged and plated using 1X TrypLE Express (Thermo Fisher Scientific). Cells were authenticated by short tandem repeat (STR) profiling.

Drosophila Culture

All *Drosophila* stocks were maintained on standard diet in a 25°C incubator with a 12 h day/night cycle.

METHOD DETAILS

Isoforms sequence

hnRNPD L isoform 1, 2 and 3 (DL1, DL2 and DL3) sequences were obtained from Uniprot ([Bateman et al., 2015](#)):

MEVPPRLSHVPPPLFSPAPATLASRSLSHWRPRPPRQLAPLLPSLAPSSARQGARRAQRHVTAQQPSRLAGGAAIKGRRRRR
DLFRRHFKSSSIQRSAAAAAATRTARQHPPADSSVTMEDMNEYSNIEEFAEGSKINASKNQDDGKMFIGGLSWDTSKDKLTEYLSR
 FGEVVDCTIKTDPVTGRSRGFGFVLFKDAASVDKVLKELKEHKLKLDGKLIDPKRAKALKGKEPPKVFVGGGLSPDTSEEQIKEYFGAFGEIENI
 ELPMDTKTNERRGFCFITYTDEEVPKLLLESRYHQIGSGKCEIKVAQPKEVYRQQQQQQGGRGAAAGGRGGTRGRGRGQ^{ggnwnqgf}
 nnyydgqygnysayggdqnygygydytgynygnygyggyadysgqQSTY GKASRGGGNHQNNYQPY

Bold corresponds to the N terminus Arg-rich domain present in DL1; lowercase italic corresponds to the C terminus Tyr-rich domain present in DL1 and DL2.

Bacterial molecular cloning

pET28a-hnRNPD L vector was purchased from GenScript. This plasmid has assembled the synthetic human gene hnRNPD L isoform 1 (DL1) with optimized codon for bacterial expression. DL1 sequence was amplified by polymerase chain reaction (PCR) using

plasmid pET28a-hnRNPDL as a template. The PCR fragment was cloned into pETite-His-SUMO kan vector using the Expresso T7 SUMO cloning and expression system (Lucigen Corporation) obtaining the pETite-SUMO-DL1 vector. The final vector was transformed into One Shot TOP10 chemically competent *E. coli* cells (Thermo Fisher Scientific).

pETite-SUMO-DL1 was used as a template for isoform 2 construct (DL2), missing the first 119 amino acids, using the Q5 site directed mutagenesis kit (New England Biolabs).

hnRNPDL isoform 2 point mutations D259N/H (DL2N and DL2H) and isoform 3 construct (DL3), missing the first 119 amino acids and residues at position 343-399, were obtained using pETite-SUMO-DL2 plasmid as a template and the Q5 site directed mutagenesis kit (New England Biolabs).

DNA fragments of R/K and Y/F mutated regions were purchased from GenScript in order to obtain pETite-SUMO-DL-R/K, Y/F and R/K+Y/F plasmids. pETite-SUMO-DL1 was used as a template for these constructs using the Q5 site directed mutagenesis kit (New England Biolabs) and DNA fragments were inserted using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs).

hnRNPDL-Nt was obtained using pETite-SUMO-DL1 plasmid as a template and the Q5 site directed mutagenesis kit (New England Biolabs).

Once confirmed the sequences of the final vectors, they were transformed into chemically competent *E. coli* BL21 DE3 cells from Expresso T7 SUMO cloning and expression system (Lucigen Corporation) for protein expression.

Protein expression and purification

100 mL of Luria Broth (LB) with 50 mg/ml kanamycin (kan) were inoculated by single colony of BL21 cells with pETite-SUMO-DL1/2/2N/2H/3 and incubated overnight at 37°C and 220 rpm (New Brunswick Innova 44R shaker). 25 mL of saturated overnight culture was transferred into 1 L LB-kan and incubated at 37°C and 220 rpm. The protein expression was induced at $OD_{600} = 0.5$ by addition of isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The induced culture was incubated for additional 3 h at 37°C and 220 rpm. Cells were then harvested by centrifugation for 15 min at 4,000 g and 4°C (Sorvall LYNX 6000, Thermo Scientific). The cell pellet was frosted at -80°C .

Pellets from 2 L cell culture were resuspended in 40 mL binding buffer (50 mM HEPES pH 7.5, 1 M NaCl, 5% glycerol, 20 mM imidazole, 1 mM dithiothreitol (DTT)) supplemented with, 0.2 mg/ml lysozyme, 20 $\mu\text{g/ml}$ DNase and 1 tablet of protease inhibitor cocktail EDTA free (Roche). The suspension was incubated for 20 min at 4°C with slow agitation and then lysed by passing through a LM10 microfluidizer (Microfluidics) at 18,000 psi. Lysate cells were centrifuged for 30 min at 30,000 g and 4°C. The supernatant was filtered through a 0.45 μm PVDF membrane and loaded onto a HisTrap FF Ni-column (GE Healthcare) at a flow rate of 2 ml/min. Protein was eluted by an imidazole gradient over 10 column volumes starting from 0 to 100% of elution buffer (50 mM HEPES pH 7.5, 1 M NaCl, 5% glycerol, 500 mM imidazole, 1 mM DTT). His purified protein was treated with 0.2 mg/ml RNase (Thermo Scientific) for 15 min at 37°C and 1 mM PMSF was added to avoid protein degradation. Then, protein was filtered using 0.22 μm PVDF membrane, concentrated using 10 K Amicon (Millipore) to 5 ml, filtered again and subjected to size exclusion chromatography on a Superdex 200 16/600 column (GE Healthcare) equilibrated in 50 mM HEPES pH 7.5, 1 M NaCl, 5% glycerol and 1 mM DTT. The fractions were analyzed by SDS-PAGE, pooled, concentrated, filtered and stored in small volume aliquots at -80°C . Absence of RNA was confirmed by 260/280 absorbance ratio using Nanodrop 8000 (Thermo Scientific). Protein concentration was determined by OD280 absorbance using Nanodrop 8000 (Thermo Scientific).

For SUMO-hnRNPDL-Nt purification, with a calculated isoelectric point of 10.24, an ion exchange chromatography with a HiTrap SP column (GE Healthcare) using a salt gradient from 50 mM to 1 M NaCl was performed before the size exclusion chromatography.

Phase separation

Purified proteins were diluted to desired protein and salt concentrations in 50 mM HEPES pH 7.5. Dilution of salt from 1 M NaCl (storage buffer) induced phase separation. A secure seal imaging spacer (Grace Biolabs) was used between slide and coverslip to visualize protein droplets in a Leica TCS SP8 microscope. Ficoll400 (Sigma) was used in addition to induce phase separation for DL2.

For DL2:hnRNPDL-Nt colocalization experiments, DL2 was green labeled with Oregon Green and hnRNPDL-Nt was red label with Texas Red using molecular probes protein labeling kits (Invitrogen) as described in the commercial protocol. A stock at 100 μM in 150 mM NaCl was prepared for each protein and then proteins were mixed at 1:1 ratio making posterior serial dilutions from the 50 μM stock mixture.

For RNA experiments, DL1 was used at 6.25 μM final concentration to obtain droplets of 2-6 μm and ensure that the surface was not fully covered with droplets to facilitate imaging. Total RNA was obtained from HeLa cells following the TRIzol Reagent user guide (Invitrogen). RNA diluted in RNase-free water was added at the indicated concentrations and posterior addition of 5 ng/ μl of RNase (Thermo Scientific) was used to induce again LLPS.

Dynamic light scattering (DLS)

hnRNPDL isoforms protein size was determined using a DynaPro NanoStar (Wyatt technologies) in 50 mM HEPES pH 7.5 and 150 mM NaCl at different protein concentrations per duplicate. Protein radius was extracted from mass data.

Protein aggregation

Right before each experiment, the stock solutions were diluted to 50 μM in 50 mM HEPES buffer pH 7.4 and 150 mM NaCl. For aggregation assays the samples were incubated for 4 days at 37°C and 600 rpm.

Transmission Electron Microscopy (TEM)

The morphology of the aggregated proteins was evaluated by negative staining and using a JEOL TEM-1400Plus Transmission Electron Microscope, 80 KV. 50 μM aggregated protein solution was sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) and then diluted to 1 μM final concentration in H₂O. 5 μL of the diluted solution was placed on carbon-coated copper grids and incubated for 5 min. The grids were then washed and stained with 5 μL of 2% w/v uranyl acetate for 2 min. Then, grids were washed again before analysis.

Congo Red (CR) binding

CR binding to aggregated proteins was analyzed using a Specord® 200 Plus spectrophotometer (Analyticjena). The absorbance spectra were recorded from 400 to 650 nm. 10 μL of 50 μM aggregated protein was added to 90 μL of 20 μM CR in 50 mM HEPES buffer pH 7,4 and 150 mM NaCl and was incubated at room temperature for 5 min before the measurement. The same buffer with 20 μM CR and without protein was employed as a control.

Thioflavin-T (Th-T) aggregation kinetics assay

The fluorescence spectra of Th-T were recorded using a Perkin Elmer EnSpire Multimode plate reader. Reactions were carried out at 50 μM protein concentration in a solution containing 20 μM of Th-T in 50 mM HEPES buffer pH 7,4 and 150 mM NaCl at 37°C. The aggregation kinetics were followed monitoring the changes in Th-T fluorescence intensity at 495 nm every 5 min with prior shaking at 100 rpm for 15 s. The same buffer with 20 μM ThT and without protein was employed as a control.

Thioflavin-S (ThS) and proteostat® staining

100 μL of aggregates protein was incubated for 1 h in the presence of 150 μM ThS (Sigma) or 1/2000 Proteostat® dilution (Enzo Life Sciences) in 50 mM HEPES buffer pH 7,4 and 150 mM NaCl. Then, the samples were washed two times with the same buffer. Finally, the precipitated fraction was resuspended in a final volume of 10 μL and placed on a microscope slide and sealed. Images were obtained at 20x magnification in an Eclipse 90i fluorescence microscope.

Congo Red (CR) birefringence

Aggregated protein was first sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052). 10 μL sonicated aggregated protein was incubated for 1 h in the presence of 100 μM CR (Sigma). 5 μL sample was placed on a microscope slide and viewed at 10x magnification with a Leica stereoscopic microscope MZFLIII.

Fourier transform infrared (FT-IR) spectroscopy

100 μL aggregated protein was centrifuged and washed two times with H₂O to remove the presence of salts. The final pellet was resuspended in 10 μL H₂O. FTIR experiments were performed using a Bruker Tensor 27 FT-IR spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Each spectrum consists of 32 independent scans, measured at a spectral resolution of 4 cm^{-1} within 1800-1500 cm^{-1} range. All spectral data were acquired and normalized using the OPUS MIR Tensor 27 software. Data was afterward deconvoluted using the Peak Fit 4.12 program.

Mammalian molecular cloning

hnRNPD L isoform 1 sequence was amplified by PCR from cDNA extract of U2OS cells and assembled to pEGFP-C3 HindIII and BamHI digested plasmid using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). Similar to bacterial molecular cloning, pEGFP-C3-DL1 plasmid was used as a template for isoform 2 (DL2) construct using the Q5 site directed mutagenesis kit (New England Biolabs). hnRNPD L isoform 3 (DL3) was obtained using pEGFP-C3-DL2 plasmid as a template and the Q5 site directed mutagenesis kit (New England Biolabs). The final vectors were transformed into One Shot TOP10 chemically competent *E. coli* cells (Thermo Fisher Scientific).

Mammalian cell culture and transfection

HeLa hnRNPD L KO (HeLa^{DL-KO}) cell line was performed by CRISPR-Cas9 with following gRNA: gRNA1: ATTCTTGCTCGCGTTG ATCT; gRNA2: ACAGAGTACTTGTCTCGATT. Both gRNA target the common exons in all 3 isoforms. pSpCas9(BB)-2A-Puro (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid # 62988; <http://addgene.org/62988>; RRID:Addgene_62988). HeLa^{DL-KO} cells were grown and maintained in DMEM High Glucose medium (Hyclone) supplemented with 10% fetal bovine serum (FBS) (Hyclone) at 37°C and 5% CO₂. Cells were passaged and plated using 1x TrypLE Express (Thermo Fisher Scientific).

For cellular transfection, HeLa^{DL-KO} cells were seeded on 4-well or 8-well glass slides (Millipore) for immunofluorescence or 4-well lab-Tek chambered coverglass (Nunc) for FRAP. Cells were transfected 24h after seeding with 400 ng DNA using Lipofectamine2000 (Invitrogen). After 4 h of transfection, we changed cellular media to fresh one. Cellular stress was applied 24h post transfection.

Immunofluorescence

Transfected HeLa^{DL-KO} cells were stressed or not with 5 $\mu\text{g}/\text{ml}$ actinomycin D for 3 h. Cells were then fixed with 4% paraformaldehyde (Electron Microscopy Sciences, #15713-S), permeabilized with 0.2% Triton X-100 and blocked with 1% bovine serum albumin (BSA). Primary antibodies used were against G3BP (611127; BD Biosciences), nucleolin (sc-8031; Santa Cruz) and hnRNPD (ab183136; Abcam). For visualization, the appropriate host-specific Alexa Fluor 488 or 555 (Molecular Probes) secondary antibodies were used. Slides were mounted using ProLong Gold antifade reagent with DAPI (Invitrogen). Images were captured using a Leica TCS SP8 STED 3x confocal microscope (Leica Biosystems) with a 63x oil objective.

Fluorescence recovery after photobleaching (FRAP)

FRAP experiments were performed with the Opterra II Swept Field Confocal Microscope (Bruker) using Prairie View 5.5 Software. Immediately before imaging, the medium was changed to 500 μL complete phenol red-free DMEM medium (HyClone). During imaging, cells were maintained at 37°C and supplied with 5% CO₂ using a Bold Line Cage Incubator (Okolabs) and an objective heater (Bioptechs). Imaging was performed using a 60x Plan Apo 1.40NA oil objective and Perfect Focus (Nikon) was engaged for the duration of the capture.

For imaging, cells were selected based on level of intensity. Time lapses were taken using the 488-nm imaging laser set at 100 power and 100-ms exposure with acquisition set at max speed (0.5 ms period) for 100 frames. Photobleaching of the nucleus occurred 2 s into capture, using the 488-nm FRAP laser to bleach the green channel. Data was repeated in triplicate for each condition, with each replicate having at least $n = 10$ cells. Data was opened in ImageJ 1.51J (NIH) using the Prairie Reader plugin. ROIs were generated in the photobleach region, a non-photobleached cell, and the background for each timelapse, and the mean intensity of each was extracted. These values were exported into Igor Pro 7.0 (WaveMetrics), where photobleach and background correction were performed, and fit FRAP curves using Hill's equation were generated.

Size exclusion chromatography (SEC)

We analyzed individually the elution profile of the contents of cells expressing EGFP-hnRNPD isoforms, one isoform at a time. A total of 3 dishes of 10 cm per each hnRNPD isoform transfection were pooled for SEC analysis. Experiments were performed in triplicate for each isoform. HeLa^{DL-KO} cells were transfected 24 h after seeding with 3 μg DNA using Lipofectamine2000 (Invitrogen) and media was changed to fresh one after 5 h of transfection. Cells were collected 24 h post transfection, resuspended in 0.5 mL of 1xPBS supplemented with protease inhibitor cocktail (Roche) and lysed using 1 mL syringe (5-10 passes). Protein solubilization was confirmed by Bradford. Supernatant was collected by 5 min centrifugation at 1,000 g and 4°C, filtered with 0.45 μm PVDF membrane and loaded on a Superose 6 Increase 10/300 GL column (GE Healthcare) equilibrated with 1xPBS. The fractions were analyzed by Western Blot using a primary antibody against hnRNPD (ab183136; Abcam) and the appropriate host-specific secondary antibody (IRDye 680RD Goat anti-Rabbit IgG (H + L), LI-COR, P/N 926-68071). The elution pattern of endogenous hnRNPD in untransfected HeLa WT cells was also analyzed.

Drosophila stocks and culture

Gene sequences of hnRNPD isoform 2 (DL2) WT, DL2 D259H, and DL2 D259N were synthesized and subcloned into the pUASTattB *Drosophila* expression vector (BioBasic Inc.). Flies carrying the transgenes were generated by performing a standard injection through the ϕC31 integrase-mediated transgenesis technique (BestGene Inc.). All *Drosophila* stocks were maintained in a 25°C incubator with a 12 h day/night cycle. Eye phenotypes were imaged by light microscopy. The w1118 line was used as control.

To prepare adult fly muscle for immunofluorescence, the *Mhc-Gal4* driver was used to express the transgene in muscle at 25°C. Adult flies were embedded in a drop of OCT compound (Sakura Finetek) on a glass slide, frozen with liquid nitrogen and bisected sagittally by a razor blade. After fixing with 4% paraformaldehyde in PBS, hemithoraces were permeabilized with PBS containing 0.2% Triton X-100 and stained with anti-hnRNPD antibody (ab183136; Abcam). Hemithoraces were additionally stained by Alexa Fluor 647 (Life Technologies) and DAPI according to manufacturer's instructions. Stained hemi-thoraces were mounted in 80% glycerol, and the musculature was examined by STED (Leica SP8).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using GraphPad Prism v5. All data are shown as the mean \pm standard error of the mean (SEM). The statistical significance of each isoform compared to unfused EGFP was investigated by ordinary two-way ANOVA followed by Dunnett's multiple comparisons test. The number of samples analyzed per experiment is provided in the corresponding method details section. A p value of less than 0.05 was used to determine significance.

DATA AND CODE AVAILABILITY

This study did not generate any unique datasets or code.

5.7. PUBLICATION VII

Prion-like domain disease-causing mutations and misregulation of alternative splicing relevance in limb-girdle muscular dystrophy (LGMD) 1G.

Battle C.* and Ventura S.*.

Neural Regen Res. (2020)

Under publication



● PERSPECTIVE

Prion-like domain disease-causing mutations and misregulation of alternative splicing relevance in limb-girdle muscular dystrophy (LGMD) 1G

Human prion-like proteins often correspond to nucleic acid binding proteins, displaying both globular domains and long intrinsically disordered regions (IDRs) (Harrison and Shorter, 2017). Their IDRs are of low complexity and resemble in amino acid composition to the disordered yeast prion domains, being usually enriched in Gln and Asn residues and depleted in hydrophobic and charged residues (Batlle et al., 2017b). Accordingly, these sequence stretches are named prion-like domains (PrLDs). Prion-like proteins can aggregate into amyloid fibrils, which can accommodate incoming protein monomers, propagating thus the polymeric fold, both processes being driven by their PrLDs. Human prion-like proteins are attracting attention because they are found in the insoluble inclusions identified in an increasing number of neurodegenerative diseases (Harrison and Shorter, 2017). Some well-characterized examples are FUS, TDP43, TAF15, EWSR1, TIA1, hnRNPA1, and hnRNPA2 proteins. Importantly, mutations in the genes that encode for these polypeptides are connected with degenerative diseases, such as amyotrophic lateral sclerosis, frontotemporal dementia or multisystem proteinopathy. A significant proportion of these mutations map in the PrLD of the prion-like protein, and often they result in their accelerated aggregation, both *in vitro* and *in vivo*. These proteins shuttle between the nucleus and the cytoplasm and are involved in the formation of membraneless organelles, like stress granules, through liquid-liquid phase separation (LLPS) (Boeynaems et al., 2018). Their aggregation typically occurs after protein mislocalization to the cytoplasm, where they form the insoluble inclusions observed in patients. It has been hypothesized that LLPS, which creates a high local protein density, is the first step towards a liquid-to-solid state transition that initiates aggregation. Membraneless organelles are highly dynamic in order to sense environmental changes and generate adequate adaptive responses. This property obeys to the fact that prion-like proteins phase separate via transient and weak non-covalent interactions. Nevertheless, genetic mutations can strengthen LLPS interactions, increasing the kinetic barrier for dissociation, leading to the population of an irreversible aberrant state, which resolves into the aggregates observed in the above-described diseases. Thus, mutations in these prion-like proteins have been suggested to result in a gain of toxic function phenotype similar to the one occurring in the brain of patients with neurodegenerative diseases like Alzheimer's and Parkinson's diseases (Brundin et al., 2010; Harrison and Shorter, 2017; Soto and Pritzkow, 2018).

Limb-girdle muscular dystrophy (LGMD) is a rare genetic disease of late childhood to adult-onset, characterized by progressive pelvic or shoulder girdle muscle weakness and wasting (Liewluck and Milone, 2018). It is the fourth most common muscular dystrophy. There are autosomal dominant or recessive inherited forms of the disease, referred to as LGMD1 or LGMD2, respectively. LGMD1G is an autosomal dominant variant caused by mutations in the hnRNPD prion-like protein (Vieira et al., 2014; Berardo et al., 2019; Sun et al., 2019). hnRNPD is a ribonucleoprotein displaying two canonical RNA recognition motifs and is involved in mRNA biogenesis, including alternative splicing (AS) and transcriptional regulation (Batlle et al., 2020). LGMD1G-linked mutations involve a single conserved Asp residue in the PrLD of hnRNPD. In LGMD1G patients, this residue is substituted by either His or Asn. As observed for other prion-like proteins, these mutations accelerate the *in vitro* aggregation kinetics of hnRNPD (Batlle et al., 2020). It is important to note that the association between the mutation of a single and well-conserved Asp residue in a PrLD and disease

also occurs in the case of hnRNPA1, hnRNPA2, and hnRNPD prion-like proteins (Harrison and Shorter, 2017; Prakash et al., 2017). It has been suggested that these Asp residues act as gatekeepers, increasing the energy barrier for self-assembly employing electrostatic repulsion; therefore, their mutation would facilitate the transition towards aggregated states. Importantly, the impact of these mutations cannot be explained simply by a global reduction of the PrLD absolute net charge, but it is instead the specific positions of these Asp residues that determine the increase in aggregation propensities caused by their mutations (Batlle et al., 2017a).

Prion-like protein's associated diseases (i.e., amyotrophic lateral sclerosis, frontotemporal dementia, and multisystem proteinopathy) are characterized by the presence of cytoplasmic inclusions, as mentioned above (Harrison and Shorter, 2017; Batlle et al., 2020). This kind of deposits has, however, not been detected in LGMD1G patients. Batlle et al. (2020) demonstrated recently that the disease-associated mutations impact hnRNPD solubility *in vivo*, making the resulting variants utterly insoluble in the myocytes of a *Drosophila* disease model, without their coalescence into visible protein inclusions. Therefore, the authors proposed that a loss of function mechanism causes LGMD1G, instead of the toxic gain of function phenotype commonly caused by this kind of mutations (Figure 1A). Consistent with this view, hnRNPD knockdown in zebrafish results in restricted and uncoordinated movements, typically associated with myopathy (Vieira et al., 2014). In the future, it would be relevant to perform a transcriptome analysis of LGMD1G patients, and compare the obtained results with a recent transcriptome study of hnRNPD knockdown in HeLa cells (Li et al., 2019), to observe if the transcript profiles and splicing patterns coincide in both cellular contexts, thus supporting a loss of function mechanism for LGMD1G.

In their recent work, Batlle et al. (2020) not only show how disease-causing mutations impact protein aggregation, but they also focus on the effect of AS in the properties of prion-like proteins. It is becoming increasingly clear that AS plays a vital role in degenerative disorders. Many prion-like proteins experiment AS events at their PrLDs, generating isoforms with or without these domains, which critically influence their self-assembly properties. It has been reported that the presence of the PrLD in the spliced forms of ribonucleoproteins (hnRNPs) promotes the formation of assemblies that regulate the AS of other genes (Guerousov et al., 2017; Feng et al., 2019). Therefore, the upstream regulation of hnRNPs' AS resulting in isoforms with or without PrLDs, regulates at the same time their downstream splicing program. Misregulation of this AS events would lead to an improper intracellular balance of a variety of isoforms with different functionality and aggregation propensities, causing disease.

Three isoforms are produced by AS of the *HNRNPD* gene, named hnRNPD isoform 1, 2, and 3 (DL1, DL2, and DL3). The three isoforms consist of two RNA recognition motifs, but they differ in the number of IDRs present in their sequence (Figure 1B). DL2 is the predominant isoform, and it only contains a C-terminus Tyr-rich IDR, predicted to be a PrLD. DL1 is the longer isoform containing an additional N-terminus Arg-rich IDR, and DL3 does not contain any IDR. Batlle et al. showed that hnRNPD isoforms have different LLPS and aggregation behavior, that can be explained according to their IDRs composition (Batlle et al., 2020). The presence of Arg and Tyr residues at the two distal IDRs in DL1 mediates cation- π interactions promoting LLPS. Mutation of the Arg to Lys or Tyr to Phe decreases DL1 phase separation, indicating that long-distance and specific cation- π interactions govern LLPS in hnRNPD. Accordingly, DL2, missing the Arg-rich IDR, has lower LLPS propensity, but, in contrast, it is the isoform with higher potential to aggregate into amyloid fibrils. DL3, without IDRs, remains soluble in solution, indicating that the IDRs are the responsible regions for both LLPS and aggregation. LLPS has usually been suggested to be the first step towards amyloid aggregation in prion-like proteins associated with diseases. However, the authors show that this would not be the case for hnRNPD and LGMD1G. DL1 is the isoform with higher LLPS propensity, but it is instead DL2 the isoform with higher amyloid potential. This observation indicates that the N-terminus Arg-rich IDR in DL1 may act as a

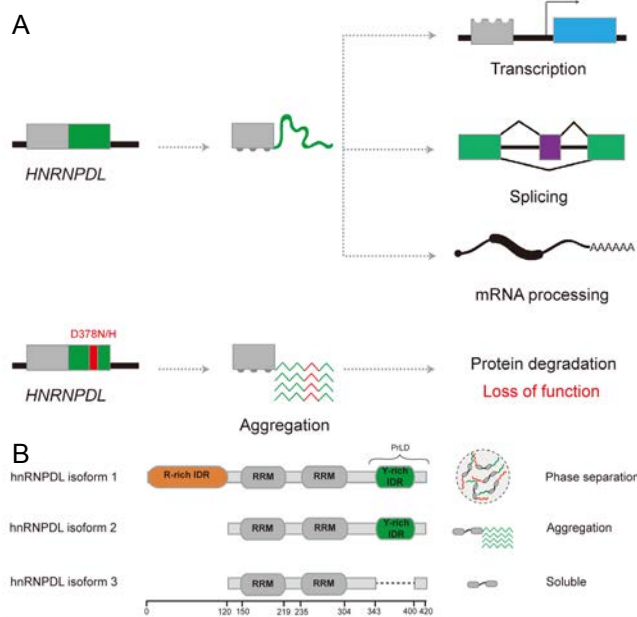


Figure 1 ????????

(A) Schematic representation of hnRNPD L loss of function by LGM-D1G-linked mutations. Genome sequencing of LGMD1G patients detected Asn (N) and His (H) substitutions in the Asp (D) residue at position 378 (red). This residue is located in the prion-like domain (PrLD) of hnRNPD L (green), and mutations enhance hnRNPD L aggregation propensity. hnRNPD L functions in transcription, splicing, and mRNA processing are lost because the mutated protein becomes insoluble and is eventually degraded by the protein quality control machinery. (B) Schematic diagram of hnRNPD L isoforms and their behavior. There exist three hnRNPD L isoforms produced by alternative splicing. hnRNPD L consists of two RNA recognition motifs (RRM, dark grey) and two, one or none predicted intrinsically disordered regions (IDRs). hnRNPD L isoform 1 contains an N-terminus Arg-rich IDR (orange) and a C-terminus Tyr-rich IDR (green), which is also predicted to be a PrLD. hnRNPD L isoform 2 only contains the C-terminus Tyr-rich IDR, and isoform 3 does not have any long IDR. The behavior of these hnRNPD L isoforms is connected to their composition: isoform 1 phase separates into liquid droplets, isoform 2 aggregates into amyloid fibrils, and isoform 3 remains soluble in solution.

protective mechanism against aggregation by promoting LLPS, something not previously described for other prion-like proteins.

The distinct self-assembly properties of hnRNPD L isoforms result in a differential behavior in mammalian cells, impacting their localization, dynamics, and interactions (Batlle et al., 2020). The PrLD in DL1 and DL2 isoforms determines their intranuclear compartmentalization, being both forms excluded from the nucleolus. Moreover, the presence of one or both IDRs in the protein sequence increases protein multivalency, favors protein-protein or protein-nucleic acids interactions, promotes the formation of more rigid and bigger complexes, and decreases protein mobility, being DL1 the less dynamic isoform. The formation of larger or more stable complexes in DL1, due to highly connected interaction networks, affects its nuclear-cytoplasmic shuttling, precluding its transport to the cytoplasm. Interestingly enough, other prion-like proteins, such as FUS, hnRNPD and TAF15 exhibit a similar dependence of their cellular properties on IDRs presence in their isoforms (Gueroussov et al., 2017; Batlle et al., 2020), indicating that AS is an essential and generic mechanism that regulates prion-like proteins' biological function in cells.

In conclusion, it is now clear that disease-causing mutations in prion-like proteins can promote both loss and gain of function phenotypes, AS controlling the self-assembly properties of these polypeptides, with significant downstream functional consequences. It is still not well understood the exact mechanism by which disease-linked protein aggregates appear and cause neurodegenerative

and muscular human disorders since LLPS seems to be able to both promote and prevent amyloid formation. Important open questions remain: Why among a large number of potential mutations in the PrLDs of proteins like hnRNPA1, hnRNPA2, and hnRNPD L, only specific Asp substitutions elicit disease onset? How and why PrLDs inclusion or exclusion is regulated by AS? The answer to these questions might allow developing generic therapeutic approaches for these diseases and provide new insights into human evolution complexity.

Cristina Batlle^{*}, **Salvador Ventura**^{*}

Departament de Bioquímica i Biologia Molecular and Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Spain

^{*}Correspondence to: *Cristina Batlle*, cristina.batlle@uab.cat;

Salvador Ventura, PhD, salvador.ventura@uab.cat.

orcid: 0000-0002-9652-6351 (*Salvador Ventura*)

Received: February 23, 2020

Peer review started: February 26, 2020

Accepted: March 19, 2020

Published online: 20???

doi: 10.

Copyright license agreement: The Copyright License Agreement has been signed by both authors before publication.

Plagiarism check: Checked twice by iThenticate.

Peer review: Externally peer reviewed.

Open access statement: This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

References

- Batlle C, Fernandez MR, Iglesias V, Ventura S (2017a) Perfecting prediction of mutational impact on the aggregation propensity of the ALS-associated hnRNPA2 prion-like protein. *FEBS Lett* 591:1966-1971.
- Batlle C, Iglesias V, Navarro S, Ventura S (2017b) Prion-like proteins and their computational identification in proteomes. *Expert Rev Proteomics* 14:335-350.
- Batlle C, Yang P, Coughlin M, Messing J, Pesarrodona M, Szulc E, Salvatella X, Kim HJ, Taylor JP, Ventura S (2020) hnRNPD L phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation. *Cell Rep* 30:1117-1128.
- Berardo A, Lornage X, Johari M, Evangelista T, Cejas C, Barroso F, Dubrovsky A, Bui MT, Brochier G, Saccoliti M, Bohm J, Udd B, Laporte J, Romero NB, Taratuto AL (2019) HNRNPDL-related muscular dystrophy: expanding the clinical, morphological and MRI phenotypes. *J Neurol* 266:2524-2534.
- Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, Tompa P, Fuxreiter M (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol* 28:420-435.
- Brundin P, Melki R, Kopito R (2010) Prion-like transmission of protein aggregates in neurodegenerative diseases. *Nat Rev Mol Cell Biol* 11:301-307.
- Feng H, Bao S, Rahman MA, Weyn-Vanhentenryck SM, Khan A, Wong J, Shah A, Flynn ED, Krainer AR, Zhang C (2019) Modeling RNA-binding protein specificity in vivo by precisely registering protein-RNA crosslink sites. *Mol Cell* 74:1189-1204.
- Gueroussov S, Weatheritt RJ, O'Hanlon D, Lin ZY, Narula A, Gingras AC, Blencowe BJ (2017) Regulatory expansion in mammals of multivalent hnRNP assemblies that globally control alternative splicing. *Cell* 170:324-339.
- Harrison AF, Shorter J (2017) RNA-binding proteins with prion-like domains in health and disease. *Biochem J* 474:1417-1438.
- Li RZ, Hou J, Wei Y, Luo X, Ye Y, Zhang Y (2019) hnRNPD L extensively regulates transcription and alternative splicing. *Gene* 687:125-134.
- Liewluck T, Milone M (2018) Untangling the complexity of limb-girdle muscular dystrophies. *Muscle Nerve* 58:167-177.
- Prakash T, Veerappa A, B Ramachandra N (2017) Complex interaction between HNRNPDL mutations and risk polymorphisms is associated with discordant Crohn's disease in monozygotic twins. *Autoimmunity* 50:275-276.
- Soto C, Pritzkow S (2018) Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat Neurosci* 21:1332-1340.
- Sun Y, Chen H, Lu Y, Duo J, Lei L, OuYang Y, Hao Y, Da Y, Shen XM (2019) Limb girdle muscular dystrophy D3 HNRNPDL related in a Chinese family with distal muscle weakness caused by a mutation in the prion-like domain. *J Neurol* 266:498-506.
- Vieira NM, Naslavsky MS, Licinio L, Kok F, Schlesinger D, Vainzof M, Sanchez N, Kitajima JB, Gal L, Cavacana N, Serafini PR, Chuartzman S, Vasquez C, Mimbacas A, Nigro V, Pavanello RC, Schuldiner M, Kunkel LM, Zatz M (2014) A defect in the RNA-processing protein HNRNPDL causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum Mol Genet* 23:4103-4110.

C-Editors: Zhao M, Li JY; T-Editor: Jia Y

6. Mediator complex subunit 15 (MED15)

The Q-rich prion-like domain of human MED15 forms a coiled-coil responsible for its conversion to amyloids and its propagation.

Battle C., Iglesias V., Calvo I., Lynch C., Serrano M. and Ventura S.

Future submission.

The Q-rich prion-like domain of human MED15 forms a coiled-coil responsible for its conversion to amyloids and its propagation

Cristina Batlle¹, Valentin Iglesias¹, Isabel Calvo², Cian Lynch², Manuel Serrano² and Salvador Ventura^{1*}

¹Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain

²Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain

*Correspondence: Salvador.ventura@uab.es

SUMMARY

Functional yeast prions are proteins with intrinsically disordered Q/N-rich prion domains (PrDs), which assemble into amyloids. A disordered to β -sheet transition was long thought to drive PrDs aggregation, similar to pathogenic amyloids. However, recent evidence indicates a critical role for coiled-coil (CC) regions within yeast PrDs in triggering amyloid formation. We show here that a significant number of human prion-like domains (PrLDs) contain CC regions, which overlap with polyQ tracts. The prion-like proteins bearing these domains are critical transcriptional coactivators, and they include the Mediator complex subunit 15 (MED15). We demonstrate that MED15-PrLD forms homodimers in solution sustained by CC interactions, and that it is this CC fold that mediates the transition towards a β -sheet amyloid state, its chemical or genetic disruption abolishing aggregation. As in functional yeast prions, globular domains adjacent to MED15-PrLD retain their structural integrity in the amyloid state. Expression of MED15-PrLD in human cells promotes the formation of cytoplasmic and perinuclear inclusions, kidnapping endogenous full-length MED15 to these aggregates in a prion-like manner. MED15 overexpression is associated with different cancers, especially in recurrent tumors possessing an aggressive phenotype. Increased protein concentration bursts MED15-PrLD multimerization and conformational conversion, suggesting that the prion-like properties of MED15 might be behind the association of this Mediator subunit and cancer.

INTRODUCTION

Prions were first described as protein-only infectious agents in the context of mammalian neurological disorders (Prusiner, 1982). Nevertheless, increasing evidence indicates that prion-like conformational conversion is not always deleterious, and instead, it can be exploited for beneficial purposes (Batlle et al., 2017a). The best characterized nonpathogenic prions are those of yeast, where they provide increased fitness in fluctuating environments (Shorter and Lindquist, 2005; Wickner et al., 2007), mainly by regulating transcription and translation (Liebman and Chernoff, 2012). Yeast prions are modular proteins, and their conformational promiscuity is encoded at their prion domains (PrDs), which are both sufficient and necessary for prion conversion (Alberti et al., 2009). PrDs are long and disordered sequences of low complexity, often enriched in glutamine and/or asparagine (Q/N) residues. Polypeptides containing similar Q/N-rich prion-like domains (PrLDs) have been identified in other organisms, including humans, and they are generically named prion-like proteins (Batlle et al., 2017a).

The aggregated state of yeast prions is macroscopically indistinguishable from that of pathogenic prions and amyloid proteins involved in neurodegenerative diseases (Jeon et al., 2013). Thus, it has been assumed that they all form amyloids through a common mechanism, that involves misfolding of the native structure, either folded or disordered, into β -sheet aggregation-prone conformations (Chiti and Dobson, 2017). However, this generic and uncontrolled misfolding mechanism was difficult to reconcile with the evidence that, in contrast to pathogenic amyloids, the structural transitions of functional prions should be regulated. Fiumara and co-workers demonstrated that yeast Q/N-rich PrDs have an intrinsic propensity to form coiled-coils (CC), and they proposed that CC-based structural changes might explain better the physiological conformational switches of functional Q/N-rich prions than stochastic misfolding (Fiumara et al., 2010). The propensity to adopt CC structures is also intrinsic to polyQ tracts (Schaefer et al., 2012). Several rare hereditary neurodegenerative diseases are originated when Q stretches exceed a critical length, and their severity increases with polyQ expansion (Orr and Zoghbi, 2007; Pearce and Kopito, 2018). These extensions would result in longer CC-prone helices, with a stronger polymerization propensity, which might ultimately cause their aggregation and the formation of the recurrent protein deposits observed in these diseases (Fiumara et al., 2010).

Although CC-mediated aggregation provides a plausible mechanism for functional transitions in yeast PrDs and for the pathogenesis of polyQ-expansion disorders, it remains to be demonstrated that the aggregation of human prion-like proteins also accommodates to the CC model. We, therefore, undertook a bioinformatic analysis of the human proteome, which revealed that a significant proportion of human PrLDs contain indeed sequences of high CC propensity, which, in most cases, overlap with Q-rich regions. This subgroup of prion-like proteins works in the regulation of transcription, and it comprises key transcription coactivators, including Mediator complex (Mediator) subunits 12 and 15 (MED12 and MED15).

Mediator is a large multi-protein complex that regulates most, if not all, gene transcription by RNA polymerase II (Pol II) (Kelleher et al., 1990; Malik and Roeder, 2010; Verger et al., 2019). Mediator is highly conserved among eukaryotes and consists of four modules: head, middle, tail, and the CDK/cyclin (Soutourina, 2018). The head and middle modules execute recruitment to the promoter regions, while the tail module mediates protein interactions with the transcription regulators, and the CDK/cyclin module has enzymatic activity (Soutourina, 2018). The tail

module consists of 7 subunits, one of them being MED15, which, as described above, we identified as a CC forming prion-like protein.

MED15 is located in the cell nucleus, and its knockdown causes slow growth and reduced transcriptional activation (Nakatsubo et al., 2014). MED15 is overexpressed in a wide range of human cancers: castration-resistant prostate cancer, head, and neck squamous cell carcinoma, hepatocellular carcinoma, breast cancer, renal cell carcinoma, and testicular germ cell tumors (Brown et al., 2003; Klümper et al., 2015; Ovchinnikov et al., 2014; Shaikhibrahim et al., 2014, 2015; Wang et al., 2018; Weiten et al., 2018; Zhao et al., 2013). Patients with MED15 overexpression in tumor tissues exhibit a more aggressive phenotype, associated with significantly shorter survival time (Wang et al., 2018). Consequently, MED15 may serve as a prognostic marker as well as a potential therapeutic target in cancer.

The MED15 subunit was first discovered in yeast as Gal11 (Fassler and Winston, 1989) and later on renamed as yMED15 (Bourbon et al., 2004). yMED15 is involved in a variety of biological processes, such as the expression of galactose-inducible genes, and it is an interaction hub of many transcription factors (TFs) (Cooper and Fassler, 2019). Interestingly, yMED15 contains a central Q-rich region reported to form amyloid-like intracellular inclusions spontaneously in yeast, whereas the deposition of the full-length yMED15 occurs under stress conditions (Zhu et al., 2015). The aggregation of yMED15 provokes the loss of the tail module, reducing transcription levels, which has been suggested to be an epigenetic mechanism for transcription regulation, similar to those of classical yeast prions (Zhu et al., 2015).

Human MED15 (MED15) has a Q-rich region, consisting of discontinuous polyQ tracts, at its N-terminus, which maps to its PrLD and is responsible for the predicted CC propensity of this domain (Fig. 1A). yMED15 and MED15 have low sequence identity and different distribution of their Q-rich regions; still, this kind of low complexity sequences have been reported to exert similar functions in orthologs without a strict positional requirement (Schaefer et al., 2012). Therefore, the MED15 prion-like domain (MED15-PrLD) might possess cryptic amyloid and prion-like features yet to be discovered.

Here, we provide experimental evidence that the MED15-PrLD forms a dimer with CC structure. This domain bears the ability to self-assemble into Congo Red positive, β -sheet enriched amyloid fibrils, displaying concentration-dependent aggregation kinetics, and self-seeding activity. Moreover, as in functional yeast prions, globular domains adjacent to the PrLD maintain their fold and activity in the amyloid state. We demonstrate that it is the CC structure that regulates MED15-PrLD amyloid formation, its chemical or genetic disruption abolishing aggregation. Expression of MED15-PrLD in human cells promotes the formation of cytoplasmic and perinuclear inclusions, kidnapping the endogenous full-length protein to these aggregates in a prion-like manner.

Overall, this study reveals that amyloid formation by human prion-like proteins can also follow the CC-mediated model and that the amyloid and prion-like properties of MED15 are conserved across species, which suggests novel mechanisms for MED15 function and malfunction.

RESULTS

Human PrLDs with predicted coiled-coil domains overlap with polyQ tracts

Coiled-coils (CC) are formed between proteins that contain repeats of seven amino acids (a-b-c-d-e-f-g) in which hydrophobic residues often occupy positions a/d to form a hydrophobic layer between the coiling helices; these heptad repeats are often discontinuous (Mason and Arndt, 2004; Parry et al., 2008a). CC regions are overrepresented in yeast prions, overlapping with their Q/N-rich prion domains, suggesting that the CC structure might represent a defining feature of these domains in yeast (Fiumara et al., 2010). This observation prompted us to assess whether human prion-like proteins might share the same architecture.

First, we performed a prion-like amino acid composition (PLAAC) analysis to identify human prion-like proteins and their respective PrLDs at the proteome level. PLAAC is an algorithm that allows the identification of polypeptide regions with a composition similar to that of a set of well-characterized yeast PrDs (Lancaster et al., 2014). We obtained a list of 193 reviewed human genes encoding for proteins with PrLDs (Table S1_1).

We analyzed which of the identified PrLDs might fold into a CC using COILS (Lupas Andrei et al., 1991), an algorithm that detects CC heptad repeats in primary sequences (Table S1_2). We found that 12% of human PrLDs ($n = 22$) were predicted to have CC regions. An 86% of these proteins ($n = 19$) is involved in transcriptional regulation (Table S1_3). Accordingly, a Gene Ontology (GO) analysis of the 22 candidates (Huang et al., 2009) indicated a high enrichment for the following ontologies: (i) transcription regulation in biological process, (ii) transcription coactivator and nucleic acid binding in molecular function, and (iii) nucleus and nucleoplasm in cellular component (Table S1_4). Importantly, 90% of the predicted CC regions ($n = 20$) correspond to or overlap with polyQ tracts in the PrLDs (Table S1_2), which is consistent with the observation that polyQ proteins are functionally biased towards transcriptional regulation (Gemayel et al., 2015; Schaefer et al., 2012).

These 20 prion-like candidates include proteins whose polyQ expansion is connected to autosomal dominant cerebellar ataxia, such as the TATA-box-binding protein (TBP) and ATXN1 (Orr and Zoghbi, 2007), as well as the CREB-binding protein (CBP), a transcriptional coactivator necessary for the survival of many neurons which is recruited into aggregates in polyQ diseases (Jiang et al., 2003).

Overall, our analysis indicates that, as it occurs in yeast prions, a significant fraction of human prion-like proteins might contain CC that overlap with polyQ tracts at their PrLDs. Proteins displaying these features include important transcriptional coactivators like CBP, TBP, KMT2D, FOXF2, NCOA2, NCOA3, MAML2, MED12 and MED15.

Apart from those involved in disease, the amyloid and prion-like properties of the PrLDs identified in these transcriptional regulators remain to be experimentally demonstrated. However, the Q-rich region of MED15 has been shown to facilitate ATX1 aggregation (Pettrakis et al., 2012), and the yMED15 polyQ can form amyloid-like aggregates on its own (Zhu et al., 2015). Despite human

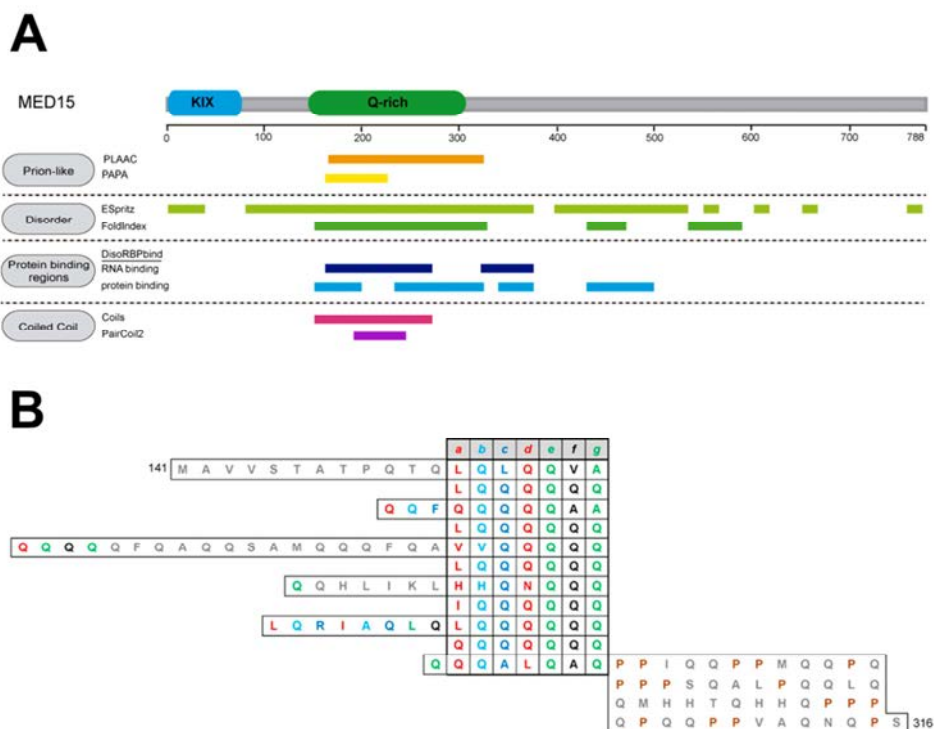


Figure 1. MED15 contains an N-terminal Q-rich coiled-coil PrLD. **A)** MED15 diagram showing the location of the KIX (blue) and Q-rich (green) regions. Prediction of prion-like domains, disorder regions, protein binding regions and coiled-coil regions are shown below the diagram. **B)** Schematic diagram of the heptad repeats (a-b-c-d-e-f-g) predicted by COILS (Lupas Andrei et al., 1991) (14-residue window) of MED15-PrLD (residues 141-316). Grey residues indicate COILS score < 0.2.

and yeast MED15 proteins display very low sequence identity (12%), we decided that it was worth to explore whether the identified human MED15-PrLD can access an amyloid state and propagate this conformation in a prion-like manner.

MED15 contains an N-terminal Q-rich coiled-coil PrLD

MED15 is highly enriched in Gln residues (Q, 20% of residues), especially at its N-terminus, with multiple short polyQ tracts (of 2-16 Qs) (**Fig. 1**). It is precisely this N-terminal Q-rich region that is identified as a PrLD by PLAAC (**Fig. 1A and Table S1_1**).

PrLDs are thought to be intrinsically disordered in their monomeric states, in such a way that their self-assembly can occur spontaneously, without an energy-dependent conformational change (Batlle et al., 2017a). The MED15-PrLD is consistently predicted to be disordered by FoldIndex and ESpritz predictors (**Fig. 1A**) (Prilusky et al., 2005; Walsh et al., 2012). Disordered regions can be cryptic protein binding sites that, upon interaction, undergo a disorder to order transition; DisoRDPbind is an algorithm aimed to predict such transitions, scoring putative associations between the disordered region of interest and RNA, DNA or protein (Peng and Kurgan, 2015). MED15-PrLD was predicted to interact with both RNA and protein (**Fig. 1A**), suggesting that it is susceptible to experiment conformational switches. Indeed, polyQ tracts are known to mediate PPI and multimerization, which can lead to the formation of homomeric or heteromeric CC structures (Schaefer et al., 2012). Accordingly, the MED15 N-terminal Q-rich region is predicted to adopt a CC

conformation (**Fig. 1A and Table S1_2**) and has been shown to exhibit α -helical structure when fused to a globular domain (Petrakis et al., 2012). This CC domain is discontinuous and consist of three consecutive CC regions (**Fig. S1**). As expected, in the MED15-PrLD CC model, almost all "a" and "d" positions in the helix are occupied by hydrophobic amino acids, Gln, which acts as an ambivalent hydrophobe (**Fig. 1B**) (Sodek et al., 1972). The N-terminal Q-rich segment is the less conserved region in MED15 vertebrate orthologs, with significant differences in the number, length, and location of polyQ tracts among species; however, in all cases, it is predicted to form CC (**Fig. S1**), suggesting that this region is structurally conserved, likely, because of its functional significance.

Overall, the bioinformatics analysis of MED15 suggests that we are in front of a prion-like protein with a disordered N-terminal Q-rich PrLD in its monomeric state, but with the ability to undergo a transition to a dimeric or higher-order CC structure.

MED15 interactors are predicted to have coiled-coil domains

Recently, a list of Mediator complex interactors was identified by performing FLAG-MED15 immunoprecipitation in mice (Quevedo et al., 2019). Therefore, at least a fraction of the identified binders may contact MED15 directly. All mouse MED15 potential interactors display >75% sequence identity with their human homologs (**Table S2_1**). In the light of the predicted propensity of MED15 to form CC, we examined the CC propensity of their putative interactors using

COILS (Lupas Andrei et al., 1991). Interestingly enough, 50% of these candidates resulted in positive predictions, indicating that CC containing proteins are overrepresented among MED15 binders (Table S2_2). Not surprisingly, a GO term enrichment analysis indicated that this subset of proteins is preferentially located at the nucleus/nucleoplasm, implicated in transcription regulation and with major involvement in Pol II transcription (Table S2_3). Importantly, among these polypeptides, we found four of the prion-like proteins identified in the previous section: NCOA2, MAML2, CBP, and MED12, which suggests that their PrLDs compositional bias and propensity to populate CC conformations might result in their eventual interaction. Indeed, an analysis with the STRING protein interactions database (Szklarczyk et al., 2017), indicates that all these five proteins are closely associated (Fig. S2) (PPI enrichment p-value: 7.5×10^{-08}).

Soluble MED15-PrLD characterization

Long polyQ stretches are commonly followed by polyproline (polyP) tracts at their C-terminus. These repeats stabilize and stop the CC region (Bhattacharyya et al., 2006; Schaefer et al., 2012), which suggests that the two tracts form a functional/structural unit and that the expression of the polyQ alone may not recapitulate the physiological context of the sequence. In MED15, a Pro-rich segment is also adjacent to the Q-rich region (Fig. 1B), and both stretches are predicted to be part of the PrLD. The presence of Pro residues flanking polyQ tracts is also frequent in MED15 orthologs (Fig. S3).

In functional proteins, polyQ tracts are adjacent to globular domains. Indeed, these low complexity regions are difficult to express recombinantly and to purify in the absence of a globular partner that modulates their inherent aggregation propensity (Adegbuyiro et al., 2017). The MED15-PrLD is flanked at its N-terminus by a globular KIX domain (Fig. 1A), a feature that is also conserved in MED15 orthologs (Fig. S3).

With above considerations in mind, we expressed the MED15-PrLD domain, containing both Q-rich and Pro-rich regions, fused at the C-terminus of GFP, which would act as a functionally traceable globular domain, hereinafter named as GFP-MED15CC. This construct should allow us to characterize the soluble, and eventually, the aggregated states of MED15-PrLD in a mimic of its sequential framework.

As intended, GFP-MED15CC was purified from the soluble cellular fraction. The GFP globular domain integrity was evaluated by monitoring the GFP absorption and fluorescence emission spectra (Fig. 2A-B). GFP-MED15CC spectral properties were identical to those of GFP alone, with an absorption maximum at 490 nm and an emission fluorescence maximum at 510 nm. In order to verify that GFP-MED15CC was not aggregated in the soluble state, we used synchronous light scattering (Fig. 2C). Neither GFP nor GFP-MED15CC showed a significant increase in light scattering signal, compared to the buffer alone.

The GFP-MED15CC secondary structure was assessed by far-UV circular dichroism (CD) (Fig. 2D). GFP alone showed a β -sheet spectrum with a characteristic minimum at 218 nm. In contrast, the GFP-MED15CC spectrum exhibited two minima at 208 and 222 nm, characteristic of α -helices. Deconvolution of the spectra using the K2D program (Whitmore and Wallace,

2008) indicated a predominant α -helix conformation for GFP-MED15CC (100%) and β -sheet for GFP (78%). Thus, the strong α -helix signal in GFP-MED15CC completely masks the signal of the folded and functional GFP, something that is characteristic of globular domains-CC fusions (Walavalkar et al., 2013).

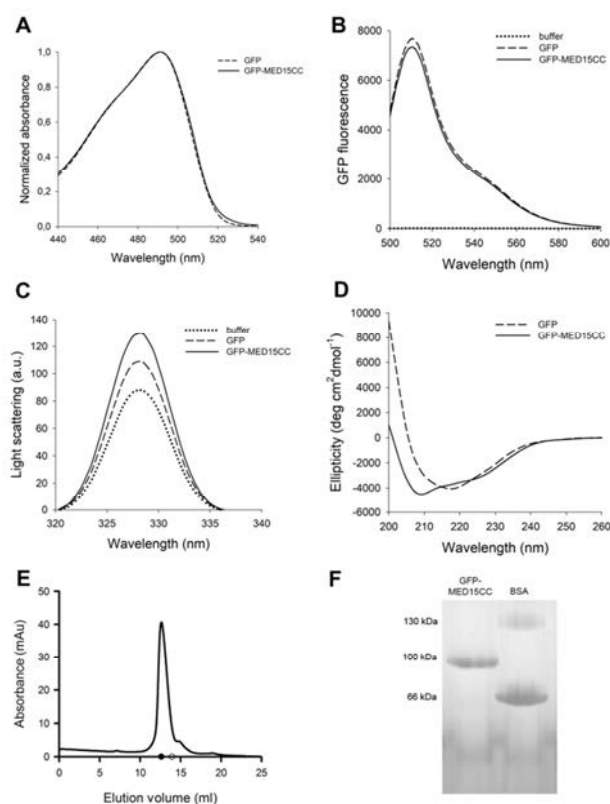


Figure 2. MED15-PrLD soluble characterization. **A)** Absorbance, **B)** fluorescence, **C)** light scattering and **D)** circular dichroism of 5 μ M GFP and GFP-MED15CC. **E)** GFP-MED15CC size exclusion chromatography elution profile. Filled circle and empty circle in the x axis indicates 100 and 50 kDa position, respectively. **F)** 12,5% blue native PAGE of GFP-MED15CC and bovine serum albumin (BSA) at 0,5 mg/ml. All assays are performed in 20 mM Tris pH 7.5 and 150 mM NaCl.

As described above, MED15-PrLD is predicted to be intrinsically disordered in its monomeric state. The fact that we observed a predominant α -helical secondary structure by far-UV CD is consistent with this region adopting a CC. This necessarily implies that, in solution, GFP-MED15CC acquires quaternary structure. GFP-MED15CC is 458 residues long, with an expected MW of 52 kDa. Size exclusion chromatography (SEC) and native PAGE were performed to evaluate the oligomerization state of GFP-MED15CC (Fig. 2E-F). Its SEC elution pattern corresponded to a protein size of \sim 100 kDa (elution at 12.65 ml), indicating that the protein was a dimer (Fig. 2E). In contrast, GFP (246 residues, MW 28 kDa) was eluted as a monomer (elution at 15 ml). In agreement with the SEC data, GFP-MED15CC migrates as a unique band of \sim 100 kDa in a native PAGE, located between the monomeric and dimeric forms of bovine serum albumin (BSA) used as control (Fig. 2F). In addition, dynamic light scattering (DLS) analysis revealed major populations with average diameters of 13 and 7 nm for GFP-MED15CC and GFP, respectively, further

supporting that the MED15 CC domain drives the formation of a dimeric structure.

In conclusion, in solution, GFP-MED15CC is a dimer sustained by a CC interaction, where the adjacent globular domain retains its integrity and function.

MED15-PrLD aggregates into β -sheet enriched amyloid-like fibrillar structures

Both PrLDs and polyQ tracts are well known for their ability to aggregate into amyloid β -sheet conformations (Adegbuyiro et al., 2017; Harrison and Shorter, 2017). Therefore, we assessed if GFP-MED15CC aggregates spontaneously using synchronous light scattering, GFP fluorescence, far-UV CD, and infrared spectroscopy (ATR-FTIR) (Fig. 3A-D). GFP was used as a control in all these assays.

The proteins were incubated at 5 μ M (0.25 mg/ml) for 24 hours at 37°C under quiescent conditions, and the light scattering of the correspondent solutions measured (Fig. 3A-B). In contrast to GFP, GFP-MED15CC exhibited a substantial increase in the light scattering signal upon incubation (Fig. 3A). GFP fluorescence was examined in the supernatant of the samples after their centrifugation (see Materials and Methods). In agreement with the light scattering data, the fluorescence of the GFP-MED15CC in the soluble fraction decreased significantly upon incubation, whereas that of GFP remained mostly the same (Fig. 3B). Together, these results demonstrate that GFP-MED15CC aggregates within 24 hours.

We monitored the secondary structure of incubated GFP-MED15CC solutions by far-UV CD and FTIR (Fig. 3C-D). GFP-MED15CC suffered a conformational change from the initial α -helical to a β -sheet structure after 24 hours of incubation, displaying a spectrum minimum at 218 nm, similar to that of GFP (Fig. 3C). Prolonged incubation for 4 days resulted in a weak β -sheet signal, indicating that a significant proportion of the protein was aggregated out of the solution, becoming undetectable by CD. The conformation of control GFP remained unaffected after 4 days of incubation. To further confirm the putative β -sheet secondary structure of incubated GFP-MED15CC, we analyzed its FTIR spectrum in the amide I region (1700-1600 cm^{-1}), which corresponds to the absorption of the main chain carbonyl group and is sensitive to the protein conformation. Deconvolution of the FTIR spectrum of incubated GFP-MED15CC resulted in a major peak at 1625 cm^{-1} , which is typically attributed to the presence of intermolecular β -sheet structure (Fig. 3D), in good agreement with the CD data.

Next, we evaluated whether GFP-MED15CC aggregates were arranged into an amyloid-like architecture. GFP-MED15CC emits green fluorescence; hence, we could not use Thioflavin-T (Biancalana and Koide, 2010). Instead, we used a Congo Red (CR) precipitation assay since this dye deposits on top of amyloid fibrils (Li et al., 2009). We incubated CR with buffer only, or with GFP, as negative controls, and with Sup35NM yeast prion amyloid fibrils as a positive control (Liebman and Chernoff, 2012). After 1 hour of incubation, only GFP-MED15CC and Sup35NM formed a red pellet, suggesting that MED15CC aggregates are amyloid-like (Fig. 3E). Next, we addressed the morphological features of GFP-MED15CC aggregates by transmission electron microscopy (TEM) to further confirm its amyloidogenic nature (Fig. 3F). Negative staining showed long, thin, and

unbranched amyloid-like fibrillar structures, without any significant accumulation of amorphous material. The fibrils exhibited a diameter of 19 ± 4 nm and a length of 2 ± 0.5 μ m.

In contrast to pathogenic proteins, a remarkable property of yeast PrDs is that their assembly into amyloids does not necessarily imply misfolding of the attached globular domains, which, at least in some cases, keep their native conformation in the aggregated state (Baxa et al., 2002). In the same manner, GFP-MED15CC fibrils were green fluorescent, indicating the maintenance of the native GFP fold within the fibril phase (Fig. S4).

Overall, these results indicate that the PrLD of the initially dimeric GFP-MED15CC protein promotes the spontaneous assembly into supramolecular β -sheet enriched fibrillar amyloid structures in which the adjacent globular domain keeps its functional conformation, all these features being consistent with a prion-like nature.

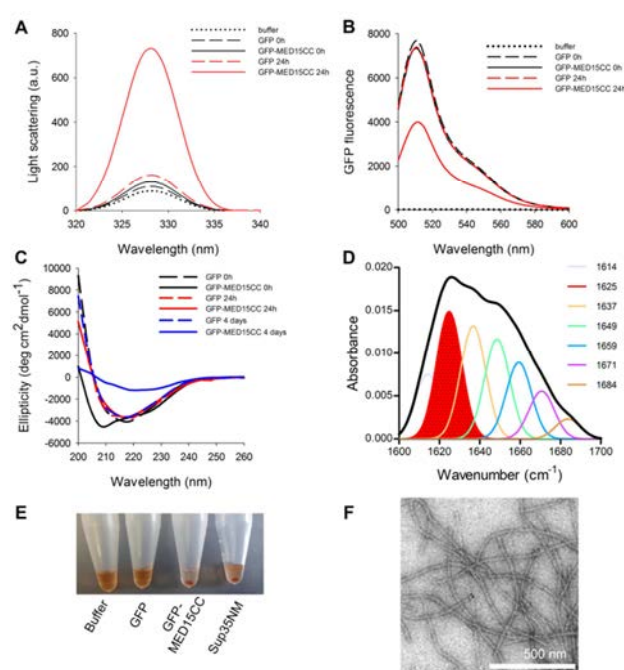


Figure 3. MED15-PrLD aggregates into β -sheet enriched amyloid-like fibrillar structures. **A)** Light scattering, **B)** GFP fluorescence (supernatant fraction) and **C)** circular dichroism of soluble ($t=0$ h) and aggregated ($t=24$ h or $t=4$ days) 5 μ M GFP and GFP-MED15CC. **D)** FTIR of aggregated ($t=24$ h) 5 μ M GFP-MED15CC. **E)** Congo Red precipitation of aggregated (4 days) 1 mg/ml GFP, GFP-MED15CC and Sup35NM. **F)** Transmission electron microscopy image of aggregated ($t=9$ h) 5 μ M GFP-MED15CC. All assays are performed in 20 mM Tris pH 7.5 and 150 mM NaCl.

MED15-PrLD aggregation kinetics

In the previous section, we showed that GFP-MED15CC aggregates into amyloid fibrils. To further characterize this reaction, we followed its kinetics, at 5 μ M protein concentration and 37°C without agitation. We monitored the reaction by synchronous light scattering, far-UV CD, and GFP fluorescence (Fig. 4A-D). Upon incubation, GFP-MED15CC showed a progressive increase in light scattering, indicative of protein aggregation (Fig. 4A). Far-UV CD analysis indicated that this is accompanied by a conformational change from an

α -helical to a β -sheet secondary structure (**Fig. 4B**). Moreover, GFP-MED15CC sample centrifugation at the same time points confirms a gradual decrease in the amount of soluble protein, with a concomitant decrease of GFP fluorescence in the supernatant (**Fig. 4C-D**). All the techniques converge to indicate that 9 hours of incubation are enough to reach the plateau phase of the aggregation reaction. In contrast, the parallel GFP control did not show signs of aggregation by any of the used techniques (**Fig. 3A-C and 4D**).

Aggregation into amyloids responds to a second or higher reaction and thus is, typically, very sensitive to the protein concentration. We evaluated if this is the case for GFP-MED15CC. We monitored GFP-MED15CC aggregation kinetics at 5 and 25 μ M by synchronous light scattering and observed a clear acceleration and signal increase at the higher concentration, as expected for a canonical amyloid protein (**Fig. 4E**).

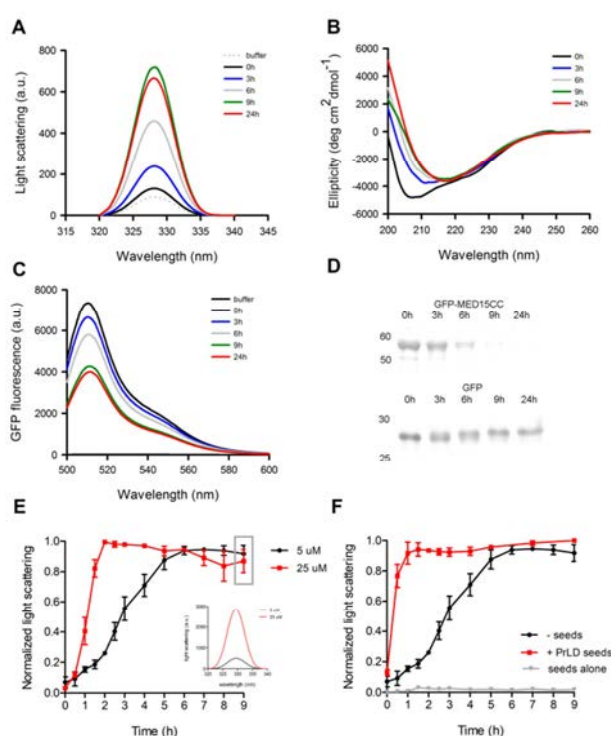


Figure 4. MED15-PrLD aggregation kinetics characterization. **A)** Light scattering, **B)** circular dichroism and **C)** GFP fluorescence (supernatant fraction) of 5 μ M GFP-MED15CC incubated at 37°C at the indicated time points. **D)** 12% SDS-PAGE of 5 μ M GFP and GFP-MED15CC supernatant fraction after centrifugation at the indicated time points of incubation. **E)** Light scattering kinetics of 5 and 25 μ M GFP-MED15CC at 37°C. The inset shows the real scattering intensities at the final time point (9h, gray rectangle). **F)** Light scattering kinetics of 5 μ M GFP-MED15CC incubated at 37°C in the absence or presence of 2% GFP-MED15CC seeds. Kinetics of seeds alone is shown as control. All assays are performed in 20 mM Tris pH 7.5 and 150 mM NaCl.

Many amyloids, and specially prion-like proteins, have the ability to self-propagate by seeding the aggregation of their soluble counterparts (Wickner et al., 2001). In this way, the presence of small amounts of preformed amyloid fibrils usually accelerates the protein aggregation kinetics, mostly by reducing or even abrogating the lag phase of

the reaction, which corresponds to the formation of the initial aggregation nuclei. In **Fig. 4F** we compare the aggregation kinetics of GFP-MED15CC at 5 μ M in the absence or presence of 2% (w/w) of its preformed fibrils. It can be observed how the presence of fibrils dramatically accelerates GFP-MED15CC aggregation. This result indicates that GFP-MED15CC amyloid formation follows a nucleation-polymerization mechanism and that, at least *in vitro*, the MED15-PrLD can propagate its amyloid conformation

MED15-PrLD aggregation is governed by its coiled-coil conformation

PolyQ tracts are generally thought to misfold spontaneously into aggregation-prone β -sheet conformations upon expansion. However, recent evidence suggest that their aggregation depends instead on the population of metastable CC structures (Fiumara et al., 2010). To test whether this mechanism drives GFP-MED15CC amyloid formation, we evaluated its aggregation in the presence of two reagents aimed to weaken or strengthen the CC domain selectively.

In the presence of 2 M urea, the GFP-MED15CC α -helical conformation was significantly disrupted, as assessed by far-UV CD, and the spectrum shifted towards the disordered region (**Fig. 5A**). Importantly, the conformation of the control GFP was not affected by this concentration of denaturant (**Fig. S5A**). We analyzed the GFP-MED15CC aggregation propensity in the presence or absence of 2 M urea by synchronous light scattering and GFP fluorescence (**Fig. 5B-C**). GFP-MED15CC did not show any signs of aggregation in the presence of the denaturing agent, indicating that CC disruption prevented protein aggregation. It is important to point out that, as expected, the integrity of the GFP moiety of GFP-MED15CC was not affected by urea because the protein remained fluorescent. The control GFP did not show any aggregation with, or without, urea (**Fig. S5B-C**).

TFE is a reagent used to enhance or stabilize α -helical conformations. However, the presence of > 20 % TFE might induce non-native α -helices, blurring any physiologically relevant interpretation of the data; to avoid this effect, we used an unusually low concentration of this cosolvent in our experiments. In the presence of 5% TFE, the CC of GFP-MED15CC was significantly enhanced as observed by far-UV CD (**Fig 5D**), whereas the structure of the control GFP was unaffected (**Fig. S5D**). Evaluation of GFP-MED15CC aggregation kinetics by synchronous light scattering and GFP fluorescence showed a significant acceleration of the process in the presence of TFE (**Fig. 5E-F**). Control GFP did not show signs of aggregation with, or without, TFE (**Fig. S5E-F**).

MED15-PrLD contains discontinuous CC segments, accounting a total of 11 heptad repeats, according to the COILS predictor (**Fig. 1B** and **Fig. S6**). To confirm the predominant role of MED15 CC on the PrLD aggregation, we used structure-guided mutagenesis to decrease the CC propensity by introducing Pro substitutions in the wild type protein, a strategy commonly used for α -helix disruption (Chang et al., 1999). We substituted five residues in positions "a" of the heptad repeats with a single Pro (**Fig. S6A**). These substitutions are predicted by COILS to disrupt the CC domain significantly (**Fig. S6B**). Hereinafter we named this mutant as GFP-MED15PP.

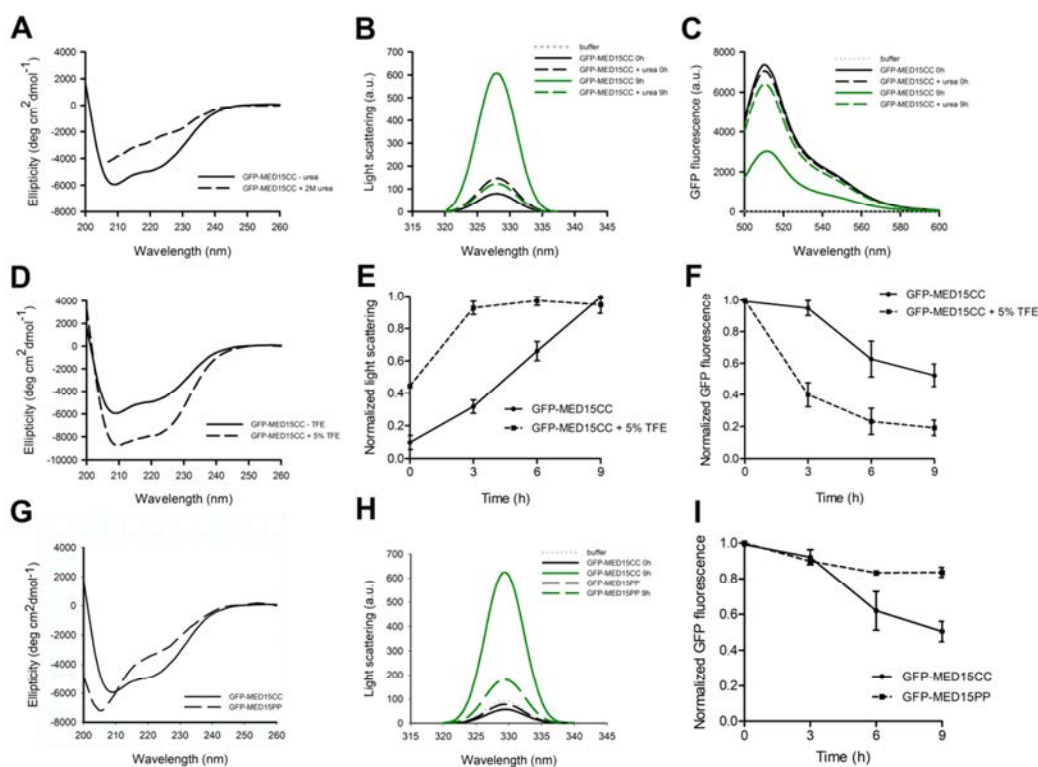


Figure 5. MED15-PrLD aggregation is governed by its coiled-coil conformation. **A)** Circular dichroism, **B)** light scattering and **C)** GFP fluorescence (supernatant fraction) of 5 μM GFP-MED15CC in the absence or presence of 2 M urea. **D)** Circular dichroism, **E)** light scattering kinetics and **F)** GFP fluorescence kinetics (supernatant fraction) of 5 μM GFP-MED15CC in the absence or presence of 5% TFE. **G)** Circular dichroism, **H)** light scattering and **I)** GFP fluorescence kinetics (supernatant fraction) of 5 μM GFP-MED15CC or GFP-MED15PP. In all cases, aggregation was performed for 9 hours at 37°C without agitation in 20 mM Tris pH 7.5 and 150 mM NaCl.

We first analyzed GFP-MED15PP secondary structure in solution by far-UV CD. The variant exhibited a minimum at 205 nm correspondent to a random coil conformation, although a small shoulder at 222 nm was still evident. In any case, the spectrum confirmed that the CC fold was significantly destabilized/disrupted by the introduced amino acid changes (Fig. 5G). Consistent with this observation, the GFP-MED15PP aggregation propensity was clearly reduced, relative to the GFP-MED15CC protein, as evidenced from the synchronous light scattering and GFP fluorescence signals after 9 hours of incubation at 37°C (Fig. 5H-I).

All in all, these results indicate that MED15 amyloid formation is mediated by its CC conformation and that disruption of this structure significantly reduces its aggregation potential. This aggregation mechanism is consistent with the one described for the functional switch of Q/N-rich yeast prions (Fiumara et al., 2010).

MED15-PrLD forms insoluble cytoplasmic inclusions in mammalian cells and kidnaps the endogenous protein

In the previous sections, we have shown that MED15-PrLD forms amyloids and that the CC is a central player in this reaction. To determine whether MED15-PrLD may aggregate in a cellular context, we stably transfected HEK293T cells with lentiviral vectors containing EGFP-tagged MED15-PrLD (EGFP-MED15CC) or EGFP, as a control, and expressed the two proteins using a doxycycline-inducible system.

MED15 is a component of the Mediator tail module, and, as expected, it displays a diffuse nuclear localization in non-induced cells (Fig. 6A). 72 hours after induction of protein expression, EGFP-MED15CC formed multiple cytoplasmic and perinuclear inclusions, whereas EGFP was diffusely distributed in the cell (Fig. 6A).

To demonstrate that the EGFP-MED15CC ability to form intracellular inclusions is cell and expression system independent and that this feature depends on the CC region of the PrLD, we transiently transfected EGFP, EGFP-MED15CC and the EGFP-MED15PP variant, containing the five disrupting Pro substitutions, in HeLa cells and expressed the three proteins for 24 hours. EGFP-MED15CC expression resulted again in the formation of cytoplasmic and perinuclear inclusions, which occur in ~50% of the transfected cells (Fig. 6B-C). In contrast, EGFP-MED15PP formed inclusions in only ~15% of the transfected cells (p-value = 0.0063), and they were small (Fig. 6B-C). Inclusion formation is a hallmark of protein aggregation, and aggregated amyloids exhibit protein detergent insolubility. In excellent agreement with their relative *in vitro* amyloid-forming and cellular inclusion-forming propensities, EGFP-MED15CC was localized mainly in the detergent-insoluble fraction upon cell fractionation, whereas EGFP-MED15PP remained mostly soluble (Fig. 7A).

Overall the cellular data converge to indicate that, as it happens *in vitro*, the CC conformation mediates MED15-PrLD aggregation and the formation of inclusions in human cells.

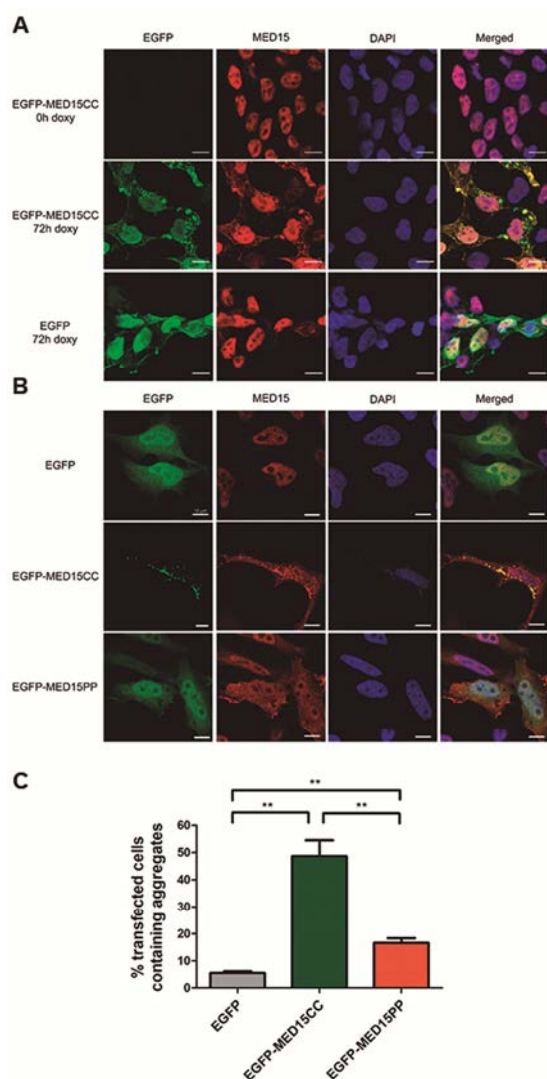


Figure 6. MED15-PrLD forms insoluble cytoplasmic inclusions in mammalian cells. A) Cellular localization by immunofluorescence of EGFP or EGFP-MED15CC in HEK293T cells after addition of doxycycline (doxy) for 0 or 72 h. **B)** Cellular localization by immunofluorescence of EGFP, EGFP-MED15CC and EGFP-MED15PP in HeLa cells after expression for 24 h. In both A and B, cells were stained with MED15 antibody (red) to evaluate the effect of transfection in the endogenous mediator subunit, and with DAPI (blue) as nuclear marker. **C)** Graph of % transfected cells containing cytoplasmic or perinuclear aggregates after EGFP, EGFP-MED15CC and EGFP-MED15PP expression in HeLa cells.

We have shown that, *in vitro*, the MED15-PrLD seeds the amyloid formation of its soluble counterpart efficiently. This activity immediately suggested that, when expressed in cells, it may also transmit its amyloid-like state and potentially seed the aggregation of the endogenous, soluble, and functional MED15 protein. Western blot and immunostaining of the endogenous full-length MED15 indicate that EGFP-MED15CC expression in HeLa cells results in a fraction of the otherwise soluble endogenous MED15 being recruited into the detergent-insoluble fraction (**Fig. 7B**). This pro-aggregation effect was reduced when we expressed the CC-disrupted EGFP-MED15PP variant instead (**Fig. 7B**). We speculated that sequestration of the endogenous protein by aggregated MED15-PrLD should increase its levels in the cytosol with

a concomitant decrease of the protein in the nucleus. HeLa cells fractionation into cytoplasmic and nuclear fractions, confirmed the expected redistribution of the endogenous protein upon EGFP-MED15CC expression (**Fig. 7C**). The mislocalization of endogenous MED15 in the cytosol might result from the retention of newly synthesized protein and/or by an alteration of the nucleo-cytoplasmic shuttling. In both cases, the interaction with aggregated cytosolic EGFP-MED15CC will make difficult its transit to the nucleus.

Overall, the data in this section converge to indicate that MED15-PrLD possesses intracellular prion-like activity, which is related to the CC propensity of its Q-rich segment.

DISCUSSION

We have identified the presence of CC domains in a significant fraction of human PrLDs. This result extends the previous observation that these α -helical domains were recurrent in yeast prions (Fiumara et al., 2010), indicating that CC constitute a generic feature of many functional prion and prion-like proteins.

PrLDs are considered and consistently predicted to be disordered in their soluble states. Accordingly, a disordered to β -sheet transition is seen as the generic mechanism behind their aggregation, similar to conventional amyloids (Shorter and Lindquist, 2005; Wickner et al., 2007). Instead, our data are consistent with CC domains playing a critical role in the structural dynamics of at least a fraction of human PrLDs. Accordingly, disrupting or destabilizing the MED15-PrLD CC, either genetically, by Pro mutations, or chemically, with urea, abolishes amyloid formation, whereas its stabilization with TFE accelerates aggregation.

To the best of our knowledge, this is the first time that the transition of a complete PrLD from an initial soluble CC conformation to an amyloid state has been characterized biophysically since previous reports involved relatively short peptides (17 residues) of the Ure2p PrD (Fiumara et al., 2010). These studies did not allow to discriminate if α -helical CC were self-sufficient mediators of the aggregation of functional prions or just intermediates or facilitators in the process of β -sheet formation. Our results support this second mechanism because the spontaneous aggregation of MED15-PrLD into amyloid fibrils involves a progressive conformational shift from an initial α -helical state to a β -sheet-rich structure, which occurs concomitantly with the aggregation of the domain.

The molecular mechanism by which a given CC converts into a β -sheet and forms amyloids is still unclear. In an elegant study, Kammerer et al. designed a short peptide (17 residues), referred to as cc β , that folds into a CC under ambient conditions but transforms into amyloid fibrils at elevated temperatures (Kammerer et al., 2004). This allowed the authors to trigger the CC to β -sheet transition in a controlled manner and to dissect the relative importance of the forces driving this conformational change. The study revealed that the nature of residues at position "f" of the heptad repeat, the most exposed region in a CC, was the most critical feature for the peptide transition. Hydrophobic residues at this position dramatically accelerated aggregation, whereas polar amino acids abolished it. These hydrophobic residues would facilitate multimerization, a reaction that depends on

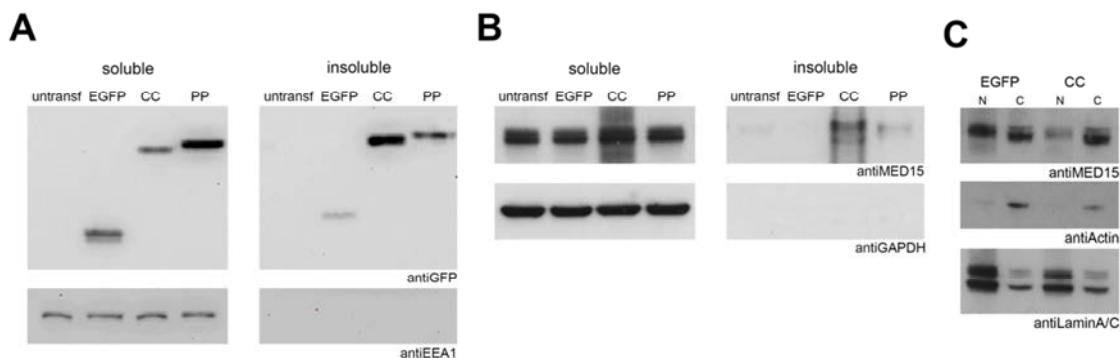


Figure 7. MED15-PrLD solubility and endogenous MED15 localization in HeLa cells. Cell extracts of HeLa were processed for soluble examination of **A)** EGFP, EGFP-MED15CC and EGFP-MED15PP, or **B)** endogenous MED15 by Western Blot in untransfected (untransf) cells or after expression of EGFP, EGFP-MED15CC (CC) or EGFP-MED15PP (PP) proteins. EEA1 and GAPDH were blotted as loading controls. **C)** Cell extracts of HeLa were processed for nuclear/cytoplasmic (N/C) examination of endogenous MED15 by Western Blot after expression of EGFP or EGFP-MED15CC (CC) proteins. Actin and LaminA/C were blotted as loading controls.

local contacts and thus would be favored at increased protein concentrations. Besides, a hydrophobic amino acid at position "f" would allow forming a continuous hydrophobic patch in a β -sheet, together with hydrophobic residues in positions "a" and "d", all lying in one side of the sequence in a fully extended conformation (Dong and Hartgerink, 2007). The same mechanism would likely apply for polyQ tracts, since Gln behaves as an ambivalent hydrophobe (Sodek et al., 1972), with contacts becoming more favorable as the polyQ, and therefore the CC, expands.

Although our model system is far more complex, it shares significant features with the behavior of cc β . In both cases, CD data indicated a transition from an initial α -helical to a β -sheet structure, the aggregated states corresponded to ordered amyloid fibrils, and conformational conversions were significantly accelerated by increased protein concentrations. Importantly, according to COILS, all residues at position "f" of MED15-PrLD CC domain are either hydrophobic or Gln (**Fig. 1B**). The need for specific spatial positioning of exposed hydrophobic residues to allow multimerization might explain the need to maintain the integrity of CC for MED15-PrLD aggregation. An essential difference between cc β and MED15-PrLD is that the structural transition of the former requires native state destabilization by temperature, whereas in the second case, it occurs spontaneously at room temperature. This indicates that the energy barrier between the CC and amyloid states of MED15-PrLD is significantly lower than that of cc β , something expected for a domain that should experiment conformational shifts under physiological conditions.

A CC-mediated model does not exclude the interplay of CCs with other PrLD elements in the aggregation process. The predicted CC regions in MED15-PrLD are discontinuous, with two short regions devoid of significant α -helical propensity linking them (**Fig. 1B**). We have shown that a majority of yeast PrDs contain sequence stretches able to assemble into highly ordered amyloid fibrils (Sant'Anna et al., 2016). These regions differ from the classical amyloid cores of pathogenic proteins in that they are slightly longer and more polar, in such a way that the amyloid potential is less concentrated, allowing the containing PrD to remain soluble under most physiological conditions, while still keeping a certain amyloid propensity

that might facilitate assembly in certain circumstances (Fernández et al., 2017; Sant'Anna et al., 2016). These soft amyloid cores (SAC) are necessary to sustain yeast prions conversion in human cells (Duernberger et al., 2018). We recently identified and characterized similar regions within several human PrLDs, including that of MED15 (Batlle et al., 2017b). The SAC of MED15-PrLD comprises residues 184-QQQQFQAQQSAMQQFQAVV-204 and a peptide correspondent to this stretch spontaneously forms highly ordered, β -sheet-rich, amyloid fibrils (Batlle et al., 2017b). This SAC corresponds almost precisely to the connecting region between the first and second CC in MED15-PrLD (**Fig. 1B** and **Fig. 8A**). This sequence is predicted to be devoid of any CC propensity by COILS and completely disordered by both FoldIndex and Espritz, and accordingly to be significantly exposed to the solvent. Thus, contacts between these accessible regions in different MED15-PrLD molecules are possible. Indeed, when we formed fibrils of the 21-residues peptide (**Fig. 8B**) and added them (2 %) to a solution of soluble GFP-MED15CC, they efficiently seeded its aggregation, indicating that this region was recognized in the soluble and dimeric protein (**Fig. 8C**). However, it is important to note that this amyloidogenic sequence alone is not sufficient to promote aggregation since this reaction does not occur in the absence of a previous CC conformation. Thus, either α -helical regions drive the initial MED15-PrLD multimerization, and intermolecular SAC contacts occur afterward, or both regions act in parallel to mediate aggregation. In both instances, the presence of the SAC would contribute to decrease the energy barrier for amyloid formation, respect that of a CC-only driven reaction. This leads us to propose a mechanism for MED15-PrLD amyloid formation in which both the misfolding of disordered regions to β -sheet conformations and CC multimerization cooperate to promote amyloid formation, unifying thus two apparently contradictory models for prion conversion. Such a cooperative model would explain why MED15-PrLD fibrillates spontaneously, even if its polyQ tracts are much shorter than those in triplet expansion genetic diseases (Bañez-Coronel et al., 2015).

We show that a dimeric CC structure is the thermodynamically dominant MED15-PrLD form in solution, immediately suggesting that MED15 might form

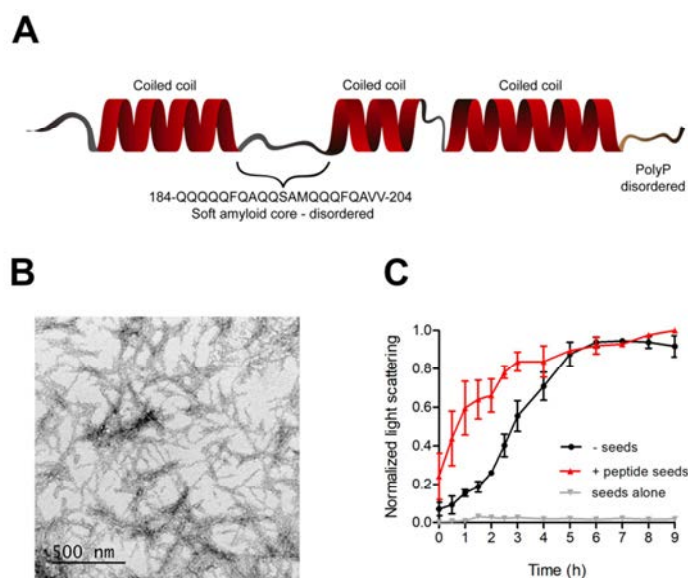


Figure 8. MED15 soft amyloid core effect in MED15-PrLD. **A)** Schematic diagram of MED15-PrLD structure. It consists of three discontinuous coiled-coil regions predicted by COILS (Lupas Andrei et al., 1991) (14-residue window) separated by disordered regions. The disorder region between the first and second coiled-coil corresponds to the soft amyloid core (SAC) predicted by pWaltz (Sabate et al., 2015). The final disordered region corresponds to the Pro residues flanking the polyQ tracts (polyP disordered). **B)** Transmission electron microscopy image of aggregated 100 μ M MED15CC SAC in 20 mM Tris pH 7.5 and 150 mM NaCl. **C)** Light scattering kinetics of 5 μ M GFP-MED15CC incubated at 37°C in the absence or presence of 2% MED15 SAC seeds. Kinetics of seeds alone is shown as control. Aggregation was performed in 20 mM Tris pH 7.5 and 150 mM NaCl.

homodimers spontaneously. No studies on the oligomeric state of full-length MED15 are available, but it has been demonstrated that yMED15 assembles into a dimer over a broad range of concentrations, suggesting that this state is also accessible to MED15 (Wands et al., 2011). The existence of functional MED15 as a dimeric protein stabilized by CC contacts is consistent with the observation that proteins containing CC are overrepresented among its interactors, since apart from homotypic interactions, CC also facilitate heterotypic PPI (Parry et al., 2008b). Human prion-like proteins containing CC at their PrLDs function in transcriptional regulation and comprise important coactivators, including NCOA2, MAML2, CBP, and MED12, all interconnected with MED15. As in humans, transcription regulators are prevailing in the yeast prion family (Alberti et al., 2009), and they include four *bona fide* prions: Ure2p, Swi1, Cyc8, Mot3. Not surprisingly, Fiumara et al. demonstrated that all these proteins contain CC at their PrDs (Fiumara et al., 2010). Thus, transcription appears to be an important cellular event that is influenced by prions and prion-like proteins across species, likely thanks to their ubiquitous CC domains. As it occurs in yeast, human prion-like transcription coactivators might allow generating phenotypic heterogeneity, allowing cells to adapt in front of intrinsic and extrinsic changes.

For most yeast prions, the new phenotype arising from their conversion to the amyloid state arises from a loss of function. In this way, Ure2p is a dimeric protein whose globular functional domain binds to the transcription factor Gln3p, preventing its migration to the nucleus (Thual et al., 2001). This interaction is lost in the Ure2p prion amyloid state, allowing Gln3p to activate the transcription of a series of genes that were previously silent. Importantly, the Ure2p globular domain is not misfolded in the process and maintains its original fold but becomes occluded from

interaction and loses its original function (Baxa et al., 2002). Similarly, when MED15-PrLD accesses a highly ordered amyloid state, the contiguous GFP moiety remains folded and fluorescent, suggesting that a mechanism equivalent to that of Ure2p might regulate MED15 activity.

In vitro, preformed fibrils of MED15-PrLD can seed the aggregation of their soluble counterpart, implying that they facilitate a conformational conversion from an initial α -helical to a β -sheet structure. The fact that when MED15-PrLD is expressed in human cells, a significant fraction of the endogenous MED15 protein localizes to the detergent-insoluble cytoplasmic fraction suggests that an analogous mechanism applies inside cells and thus that MED15 can potentially act in a prion-like manner in a physiological context. The ability to aggregate inside cells and to potentially transmit this conformation is, again, strongly dependent on the integrity of the N-terminal CC domain, as demonstrated by the reduced ability to form cellular inclusions and to interact with the endogenous protein of the MED15PP variant.

MED15 is part of the mediator tail module and is known to serve as a hub for signaling pathways. MED15 is overexpressed in different cancers and correlates with the clinical outcome and the recurrence of the disease (Shaikhibrahim et al., 2014, 2015; Syring et al., 2018; Wang et al., 2018; Zhao et al., 2013). The molecular mechanism by which increased levels of MED15 contributes to these malign phenotypes remains unknown, but it is assumed that they result from an increased and sustained transcriptional activation (Shaikhibrahim et al., 2014, 2015; Zhao et al., 2013). MED15-PrLD forms intracellular inclusions in different human cell lines, including laryngeal and oral squamous cell carcinoma (Fig. S8), two tumors with MED15 implication

(Shaikhbrahim et al., 2015). Under the assumption that, as it occurs for yMED15, the aggregation propensity of the PrLD is maintained in the context of the full-length protein, our results suggest alternative mechanisms to explain the association of this prion-like protein with cancer. Overexpression would increase the number of MED15 pre-associating N-terminal CC domains and thus their effective local concentration, facilitating both the establishment of CC homotypic and heterotypic interactions and PrLD conformational conversion.

MED15-PrLD mediated homotypic interactions would favor multimerization, and this might prevent the establishment of other relevant MED15 PPI, occluding the access to TFs interacting motifs, like those in the KIX domain, as it happens in Ure2p (Baxa et al., 2002). Alternatively, the self-association of MED15 might result in enhanced formation of transcriptional hubs by phase separation, as described for the ENL chromatin reader (Wan et al., 2020). Both situations are expected to result in widespread transcriptional changes. Because a significant fraction of MED15 interactors contains CC domains, increased heterotypic interactions and subsequent multimerization by side-to-side CC interactions, or from the swapping between protomers, are expected to exert a similar effect to homotypic contacts, either decreasing or exacerbating the interactors function. Another possibility is that the MED15-PrLD would establish promiscuous interactions with other Q-rich CC containing proteins, whose are not usual MED15 partners, sequestering them. Since many of these proteins are expected to be TFs, these degenerated CC interactions would impact transcription and signaling pathways significantly, as demonstrated for polyQ-expanded proteins in neurodegenerative diseases (Kwon et al., 2018; Orr and Zoghbi, 2007). In extreme cases, the MED15-PrLD might template and enhance the transition of other Q-rich proteins towards the formation of toxic amyloid aggregates, as previously shown for human ATXN1 (Petrakis et al., 2012). Deciphering if any of these mechanisms or a combination of them is behind MED15 association with different cancers might offer novel therapeutic avenues for their treatment.

MATERIALS AND METHODS

Bioinformatics analysis

Human reviewed proteins from the standard proteome (Proteome ID UP000005640, release 2019_06) were downloaded from UniProtKB/Swiss-Prot (Bateman et al., 2015). MED15 mouse interactors were obtained from (Quevedo et al., 2019). Their human homologous were obtained by blasting each sequence against their human UniProtKB/Swiss-Prot counterparts. For each protein, the highest sequential candidate (>75% homology) was obtained.

Human reviewed proteins were screened with PLAAC algorithm (Lancaster et al., 2014) using as background probability the precompiled frequencies of human proteome to obtain the list of putative proteins with PrLDs.

Coiled-coil prediction was performed using C version of COILS (Lupas Andrei et al., 1991) with a threshold of 0.2 and a standard 21 amino acids window.

Prion-like proteins and MED15 interactors datasets were analyzed enrichment with the Functional Annotation

Tool of DAVID 6.8 (Database for Annotation, Visualization and Integrated Discovery) (Huang et al., 2009) and ordered by p-value. Default "Direct" category was selected for each ontology as these are directly provided from the source annotation.

MED15 sequences

MED15 CC sequence (residues 141-316) was obtained from Erich E. Wanker research group and fused to GFP sequence in pET21b plasmid. The sequence is shown below with MED15 CC sequence underlined.

```
MASMTGGQQMGRDPNSSKGEELFTGVVPIVLVELDGDVNGH
KFSVSGEGEGDATYGKLTCLKFICTTGKLPVWPPTLVTTLYGV
QCFSTRYPDHMKRHHFFKSAMPEGYVQERTISFKDDGNYKTRAE
VKFEGDTLVNRIELKIDFKEDGNILGHKLEYNYNSHNVYITA
DKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITHGMDELYKKL
LEVLFGQPMVVSTATPQTQLQLQOVALQQQQQQQQFQQQQQA
ALQQQQQQQQQQQFQAQQSAMQQQFQAVVQQQQQLQQQQQQQQ
HLIKLHHQNNQQQIQQQQQQLQRIAQQLQQQQQQQQQQQQQQQ
QALQAQPPIQQPPMQPPQPPPSQALPQQLQOMHHTQHHPQPPQ
PQQPPVAQNQPSAAALEHHHHHHH
```

MED15 PP sequence was purchased from Genscript as GFP-MED15PP fusion in a pET21b plasmid. The sequence is the same as GFP-MED15CC except 5 Pro substitutions in MED15 CC region (**Fig. S6**).

Protein expression and purification

100 ml of Luria Broth (LB) with 100 µg/ml ampicillin (amp) and 34 µg/ml chloramphenicol (clm) were inoculated by single colony of BL21 Rosetta cells with pET21b-GFP-MED15CC/PP and incubated overnight at 37°C and 250 rpm. 25 ml of saturated overnight culture was transferred into 1 L LB-amp-clm and incubated at 37°C and 250 rpm. Protein expression was induced at OD₆₀₀ = 0.5 by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The induced culture was incubated O/N at 20°C and 250 rpm. Cells were then harvested by centrifugation for 15 min at 4,000 g and 4 °C (Beckman CoulterTM Avanti Centrifuge J-26XPI). The cell pellet was frosted at -80°C.

Pellets from 2 L cell culture were resuspended in 30 ml binding buffer (20 mM TrisHCl pH 8, 500 mM NaCl, 20 mM imidazole, 10% glycerol) supplemented with, 0.2 mg/ml lysozyme, 20 µg/ml DNase and 1 tablet of protease inhibitor cocktail EDTA free (Roche). The suspension was incubated for 20 min at 4°C with slow agitation and then lysed by 5 min sonication on ice (Branson Digital Sonifier). Lysate cells were centrifuged for 30 min at 30,000 g and 4°C. The supernatant was filtered through a 0.45 µm PVDF membrane and loaded onto a HisTrap FF Ni-column (GE Healthcare) at a flow rate of 4 ml/min. Protein was eluted by one-step procedure using elution buffer (20 mM TrisHCl pH 8, 500 mM NaCl, 250 mM imidazole, 10% glycerol). His purified protein was treated with 0.2 mg/ml RNase (Thermo Scientific) for 15 min at 37°C and 1 mM PMSF was added to avoid protein degradation. Native buffer (20 mM TrisHCl pH 8, 500 mM NaCl, 10% glycerol) was added to the protein sample to reduce imidazole to 20 mM and a second His trap step was performed as before to remove RNase protein. Then, protein was filtered using 0.22 µm PVDF membrane, concentrated using 10 K Amicon (Millipore) to 2.5 ml, filtered again and subjected to PD10 column equilibrated in 20 mM TrisHCl pH 7.5,

500 mM NaCl, 10% glycerol. Purified protein was concentrated to 0.5 ml, filtered and stored in small volume aliquots at -80°C. Protein concentration and absence of RNA was confirmed by 260nm and 280nm absorbance values using Specord® 200 Plus spectrophotometer (Analyticjena).

GFP was purified as described in (Gil-Garcia et al., 2018) and Sup35NM was purified as described in (Sant'Anna et al., 2016).

Size exclusion chromatography

2 mg/ml of purified protein was filtered with 0.22 µm PVDF membrane and loaded on a Superdex 200 HR 10/30 (GE Healthcare) equilibrated with 20 mM TrisHCl pH 7.4, 150 mM NaCl and 10% glycerol. The fractions were analyzed by SDS-PAGE 12%.

Native page

0,5 mg/ml of purified protein was loaded in a blue native polyacrylamide gel electrophoresis 12,5% for the analysis of protein oligomers.

Dynamic Light Scattering (DLS)

GFP and GFP-MED15CC protein size was determined using a Malvern Zetasizer Nano Series (Malvern instruments, UK) in 20 mM TrisHCl pH 7.5 and 150 mM NaCl at different protein concentrations per duplicate. Protein diameter was extracted from volume size.

GFP absorption

GFP absorbance was monitored from 400 to 600 nm using a Specord® 200 Plus spectrophotometer (Analyticjena).

Protein aggregation

Right before each experiment, the stock solutions were diluted to 5 µM in 20 mM TrisHCl pH 7.4 and 150 mM NaCl. For aggregation assays the samples were incubated at 37°C without agitation. In some experiments, protein aggregation was performed in the presence of 2 M urea or 5% 2,2,2-trifluoroethanol (TFE) (Fluka). All aggregation experiments were performed per triplicate.

GFP fluorescence

50 µl sample was centrifuged for 5 min at max speed and RT. Then, supernatant fraction was used for GFP fluorescence measurements unless indicated in the text. GFP fluorescence was monitored using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 485 nm, emission range from 500 to 600 nm, slit widths of 5 nm, 0.5 nm interval, 1 sec response and 1000 nm/min scan rate. The supernatant fraction was also analyzed by SDS-PAGE 12%.

Synchronous light scattering

Synchronous light scattering was monitored using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 330 nm, emission range from 320 to 340 nm, slit widths of 5 nm, 0.5 nm interval, 1 sec response and 1000 nm/min scan rate.

Congo Red (CR) precipitation

CR 200 µM in H₂O was centrifuged before the experiment to remove any possible precipitate. Buffer, GFP, GFP-MED15CC aggregates or Sup35NM aggregates were mixed 1:1 in volume with CR 200 µM (FC 100 µM). Samples were incubated for 1 hour at RT covered from light. Then, we centrifuged the samples for 5 min at max speed and RT. Appearance of a red pellet was indicative of the presence of amyloid fibrils.

Circular Dichroism (CD)

CD experiments were performed using a JASCO J-715 spectropolarimeter. Measurements of the far-UV CD spectra (260–200 nm) were made by the addition of 200 µl of 5 µM sample to a cuvette of 0.1 cm path-length. Spectra were recorded at room temperature, 1 nm band width and 100 nm/min scan rate. The resulting spectrum was the average of 5 scans. The contribution of the buffer was subtracted.

Fourier transform infrared (FTIR) spectroscopy

100 µl aggregated protein was centrifuged and washed one time with H₂O to remove the presence of salts. The final pellet was resuspended in 10 µl H₂O. FTIR experiments were performed using a Bruker Tensor 27 FT-IR spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Each spectrum consists of 32 independent scans, measured at a spectral resolution of 4 cm⁻¹ within 1800-1500 cm⁻¹ range. All spectral data were acquired and normalized using the OPUS MIR Tensor 27 software. Data was afterwards deconvoluted using the Peak Fit 4.12 program.

Transmission Electron Microscopy (TEM)

The morphology of the aggregated proteins was evaluated by negative staining and using a JEOL TEM-1400Plus Transmission Electron Microscope, 80 KV. 5 µM aggregated protein solution was diluted to 1 µM final concentration in H₂O. 5 µl of the diluted solution was placed on carbon-coated copper grids and incubated for 5 min. The grids were then washed and stained with 5 µl of 2 % w/v uranyl acetate for 1 min. Then, grids were washed again before analysis.

Mammalian molecular cloning

MED15 CC and PP sequence were amplified by PCR from pET21b-GFP-MED15CC/PP and assembled to pEGFP-C3 (Clontech, plasmid #6082-1) HindIII and BamHI digested plasmid using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). EGFP and EGFP-MED15CC were amplified by PCR from pEGFP-C3 and pEGFP-C3-MED15CC, respectively, and assembled to pTetO-FUW-OSKM (Addgene, plasmid #20321) EcoRI digested plasmid using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). The final vectors were transformed into XL1Blue chemically competent *E. coli* cells.

Mammalian cell culture

HeLa cells were grown and maintained in MEMα Glutamax medium (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific) at 37°C and 5% CO₂. HEK-293T were grown and maintained in DMEM medium supplemented with 10% FBS (Gibco) with antibiotics (penicillin/streptomycin 100 U/ml), on gelatin-coated plates

at 37°C and 5% CO₂. Cells were passaged and plated using 1x TrypLE Express (Thermo Fisher Scientific). Laryngeal HNSCC (UT-SCC-42B) and oral HNSCC (UT-SCC-2) were grown and maintained in DMEM medium (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific) at 37°C and 5% CO₂.

Production of lentivirus, and infection of recipient cells

Lentiviral supernatants were produced in HEK-293T cells (5x10⁶ cells per 100 mm diameter dish). Vector transfections were performed using Fugene-6 transfection reagent (Roche) according to the manufacturer's protocol.

For lentiviral production, per dish, 293T cells were transfected with 3 plasmids: (i) the ecotropic lentiviral envelope packaging plasmid pMD2.G (0.3 µg; Addgene, plasmid #12259; containing the VsVg gene); (ii) the lentiviral packaging plasmid pCMV-dR8.91 (3.0 µg); (from: Harvard Medical School, plasmid #516); (iii) plus one of the following lentiviral expression constructs (3.0 µg) expressing either the FUW-M2-rtTA vector (Addgene #20342), or, pTetO-FUW-EGFP-MED15CC or pTetO-FUW-EGFP.

Two days later, viral supernatants (10 ml) were collected serially during the subsequent 48 hours, at 12-hour intervals, each time adding fresh medium to the cells (10 ml). The recipient 293T cells were seeded the previous day (1.5x10⁵ cells per well in a 6-well plate) and each well received 1.0 ml of the corresponding lentiviral supernatants. This procedure was repeated every 12 hours for 2 days (a total of 3 additions). Recipient cells received both the FUW-M2-rtTA plus either pTetO-FUW-EGFP-MED15CC or pTetO-FUW-EGFP.

24h after lentiviral infection was completed; lentiviral expressing cells were used directly. Cell samples were tested for expression of EGFP or EGFP-MED15CC by Western blot 72h after addition of 1 µg/ml doxycycline.

Mammalian cells transfection

HeLa, laryngeal HNSCC (UT-SCC-42B) and oral HNSCC (UT-SCC-2) cells were transfected 24 h after seeding, on 8-well glass slides for immunofluorescence or on 10 cm dish for Western Blot, with 400 ng DNA using Lipofectamine2000 (Invitrogen). After 4 h of transfection, we changed cellular media to fresh one.

Immunofluorescence

Cells were fixed with 4% paraformaldehyde (141451, PanReacAppliChem) 24h post transfection, permeabilized with 0,2% TritonX100 and blocked with 1% bovine serum albumin (BSA). Primary antibody used was against MED15 (11566-1-AP, Proteintech). For visualization, the appropriate host-specific Alexa Fluor 555 (Thermo Fisher Scientific) secondary antibody was used. Slides were mounted using ProLong Gold antifade reagent with DAPI (Invitrogen). Images were captured using a Leica TCS SP5 confocal microscope (Leica Biosystems) with a 63x oil objective.

Soluble/insoluble fractionation

Transfected HeLa cells were trypsinized, washed with PBS1x and resuspended in 0.5 ml RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% TritonX100, 2 mM EDTA, 0.5% sodium deoxycholate and 0.1% SDS) supplemented with protease inhibitors. After 10 min incubation on ice,

cells were centrifuged for 10 min at max speed and 4°C. The supernatant corresponding to the soluble fraction was placed in a new eppendorf tube. The pellet corresponding to the insoluble fraction was washed with RIPA supplemented with protease inhibitors, resuspended in 0.5 ml urea buffer (30 mM Tris pH 8.5, 8 M urea, 2% CHAPS) and incubated for 20 min at RT and agitation. Protein concentration of the soluble fraction was measured by Bradford and 20 µg were loaded for Western Blot analysis. We loaded the same volume for the insoluble fraction. Primary antibody used was against MED15 (11566-1-AP, Proteintech), GFP (G6795, Sigma), GAPDH (AM4300, Invitrogen) and EEA1 (610457, BD Biosciences). The appropriate host-specific HRP conjugated secondary antibody (BioRad) was used. Western Blot was revealed using Immobilon® Forte Western HRP substrate (Millipore) in a VersaDoc (BioRad).

Nuclear/cytoplasmic fractionation

Nuclear and cytoplasmic fractionation was performed with the NE-PER kit (ThermoFisher #78833). 10 µg protein was loaded for Western Blot analyses. Primary antibody used was against MED15 (11566-1-AP, Proteintech), against β-actin (A5441, Sigma) and against Lamin A/C (Sc-6215 (N-18), Santa Cruz Biotechnology).

REFERENCES

- Adegbuyiro, A., Sedighi, F., Pilkington, A.W., Groover, S., et al. (2017). **Proteins Containing Expanded Polyglutamine Tracts and Neurodegenerative Disease**. *Biochemistry* 56, 1199–1217.
- Alberti, S., Halfmann, R., King, O., Kapila, A., et al. (2009). **A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins**. *Cell* 137, 146–158.
- Bañez-Coronel, M., Ayhan, F., Tarabochia, A.D., Zu, T., et al. (2015). **RAN Translation in Huntington Disease**. *Neuron* 88, 667–677.
- Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., et al. (2015). **UniProt: A hub for protein information**. *Nucleic Acids Res.* 43, D204–D212.
- Battle, C., Iglesias, V., Navarro, S., and Ventura, S. (2017a). **Prion-like proteins and their computational identification in proteomes**. *Expert Rev. Proteomics* 14, 335–350.
- Battle, C., De Groot, N.S., Iglesias, V., Navarro, S., et al. (2017b). **Characterization of Soft Amyloid Cores in Human Prion-Like Proteins**. *Sci. Rep.* 7, 1–16.
- Baxa, U., Speransky, V., Steven, A.C., and Wickner, R.B. (2002). **Mechanism of inactivation on prion conversion of the Saccharomyces cerevisiae Ure2 protein**. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5253–5260.
- Bhattacharyya, A., Thakur, A.K., Chellgren, V.M., Thiagarajan, G., et al. (2006). **Oligoproline effects on polyglutamine conformation and aggregation**. *J. Mol. Biol.* 355, 524–535.
- Biancalana, M., and Koide, S. (2010). **Molecular mechanism of Thioflavin-T binding to amyloid fibrils**. *Biochim. Biophys. Acta - Proteins Proteomics* 1804, 1405–1412.
- Bourbon, H.-M., Aguilera, A., Ansari, A.Z., Asturias, F.J., et al. (2004). **A Unified Nomenclature for Protein Subunits of Mediator Complexes Linking Transcriptional Regulators to RNA Polymerase II**. *Mol. Cell* 14, 553–557.
- Brown, J.J., Xu, H., Nishitani, J., Mohammed, H., et al. (2003). **Potential biomarkers for head and neck squamous cell carcinoma**. *Laryngoscope* 113, 393–400.
- Chang, D.-K., Cheng, S.-F., Trivedi, V.D., and Lin, K.-L. (1999). **Proline Affects Oligomerization of a Coiled Coil by Inducing a**

Kink in a Long Helix.

- Chiti, F., and Dobson, C.M. (2017). **Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade.** *Annu. Rev. Biochem.* *86*, 27–68.
- Cooper, D.G., and Fassler, J.S. (2019). **Med15: Glutamine-Rich Mediator Subunit with Potential for Plasticity.** *Trends Biochem. Sci.* 1–15.
- Dong, H., and Hartgerink, J.D. (2007). **Role of hydrophobic clusters in the stability of alpha-helical coiled coils and their conversion to amyloid-like beta-sheets.** *Biomacromolecules* *8*, 617–623.
- Duernberger, Y., Liu, S., Riemschoss, K., Paulsen, L., et al. (2018). **Prion Replication in the Mammalian Cytosol: Functional Regions within a Prion Domain Driving Induction, Propagation, and Inheritance.** *Mol. Cell. Biol.* *38*.
- Fassler, J.S., and Winston, F. (1989). **The *Saccharomyces cerevisiae* SPT131GALJJ Gene Has Both Positive and Negative Regulatory Roles in Transcription.**
- Fernández, M.R., Batlle, C., Gil-García, M., and Ventura, S. (2017). **Amyloid cores in prion domains: Key regulators for prion conformational conversion.** *Prion* *11*, 31–39.
- Fiumara, F., Fioriti, L., Kandel, E.R., and Hendrickson, W.A. (2010). **Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins.** *Cell* *143*, 1121–1135.
- Gemayel, R., Chavali, S., Pougach, K., Legendre, M., et al. (2015). **Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity.** *Mol. Cell* *59*, 615–627.
- Gil-García, M., Banó-Polo, M., Varejao, N., Jamroz, M., et al. (2018). **Combining Structural Aggregation Propensity and Stability Predictions to Redesign Protein Solubility.** *Mol. Pharm.* *15*, 3846–3859.
- Harrison, A.F., and Shorter, J. (2017). **RNA-binding proteins with prion-like domains in health and disease.** *Biochem. J.* *474*, 1417–1438.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res.* *37*, 1–13.
- Jeon, J., Lodge, M.S., Dawson, B.D., Ishigami, M., et al. (2013). **Superb resolution and contrast of transmission electron microscopy images of unstained biological samples on graphene-coated grids.** *Biochim. Biophys. Acta* *1830*, 3807–3815.
- Jiang, H., Nucifora, F.C., Ross, C.A., and DeFranco, D.B. (2003). **Cell death triggered by polyglutamine-expanded huntingtin in a neuronal cell line is associated with degradation of CREB-binding protein.** *Hum. Mol. Genet.* *12*, 1–12.
- Kammerer, R.A., Kostrewa, D., Zurdo, J., Detken, A., et al. (2004). **Exploring amyloid formation by a de novo design.** *Proc. Natl. Acad. Sci. U. S. A.* *101*, 4435–4440.
- Klümper, N., Syring, I., Offermann, A., Shaikhibrahim, Z., et al. (2015). **Differential expression of Mediator complex subunit MED15 in testicular germ cell tumors.** *Diagn. Pathol.* *10*, 1–6.
- Kwon, M.J., Han, M.H., Bagley, J.A., Hyeon, D.Y., et al. (2018). **Coiled-coil structure-dependent interactions between polyQ proteins and Foxo lead to dendrite pathology and behavioral defects.** *Proc. Natl. Acad. Sci. U. S. A.* *115*, E10748–E10757.
- Lancaster, A.K., Nutter-Upham, A., Lindquist, S., and King, O.D. (2014). **PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition.** *Bioinformatics* *30*, 2501–2502.
- Li, H., Rahimi, F., Sinha, S., Maiti, P., et al. (2009). **Amyloids and Protein Aggregation – Analytical Methods.** *Encycl. Anal. Chem.* 1–32.
- Liebman, S.W., and Chernoff, Y.O. (2012). **Prions in yeast.** *Genetics* *191*, 1041–1072.
- Lupas Andrei, Van Dyke Marc, and Stock Jeff (1991). **Predicting coiled coils from protein sequences.** *Science* (80-). *252*, 1162–1164.
- Madeira, F., Park, Y.M., Lee, J., Buso, N., et al. (2019). **The EMBL-EBI search and sequence analysis tools APIs in 2019.** *Nucleic Acids Res.* *47*, W636–W641.
- Mason, J.M., and Arndt, K.M. (2004). **Coiled coil domains: Stability, specificity, and biological implications.** *ChemBioChem* *5*, 170–176.
- McDonnell, A. V., Jiang, T., Keating, A.E., and Berger, B. (2006). **Paircoil2: improved prediction of coiled coils from sequence.** *Bioinformatics* *22*, 356–358.
- Nakatsubo, T., Nishitani, S., Kikuchi, Y., Iida, S., et al. (2014). **Human mediator subunit MED15 promotes transcriptional activation.** *Drug Discov. Ther.* *8*, 212–217.
- Orr, H.T., and Zoghbi, H.Y. (2007). **Trinucleotide Repeat Disorders.** *Annu. Rev. Neurosci.* *30*, 575–621.
- Ovchinnikov, D.A., Wan, Y., Coman, W.B., Pandit, P., et al. (2014). **DNA methylation at the novel CpG sites in the promoter of MED15/PCQAP gene as a biomarker for head and neck cancers.** *Biomark. Insights* *9*, 53–60.
- Parry, D.A.D., Bruce Fraser, R.D., and Squire, J.M. (2008a). **Fifty years of coiled-coils and alpha-helical bundles: A close relationship between sequence and structure.**
- Parry, D.A.D., Fraser, R.D.B., and Squire, J.M. (2008b). **Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure.** *J. Struct. Biol.* *163*, 258–269.
- Pearce, M.M.P., and Kopito, R.R. (2018). **Prion-Like Characteristics of Polyglutamine-Containing Proteins.** *Cold Spring Harb Perspect Med* *8*(2), a024257.
- Peng, Z., and Kurgan, L. (2015). **High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder.** *Nucleic Acids Res.* *43*, e121–e121.
- Petrakis, S., Raskó, T., Russ, J., Friedrich, R.P., et al. (2012). **Identification of Human Proteins That Modify Misfolding and Proteotoxicity of Pathogenic Ataxin-1.** *PLoS Genet.* *8*.
- Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., et al. (2005). **FoldIndex@: A simple tool to predict whether a given protein sequence is intrinsically unfolded.** *Bioinformatics* *21*, 3435–3438.
- Prusiner, S.B. (1982). **Novel proteinaceous infectious particles cause scrapie.** *Science* *216*, 136–144.
- Quevedo, M., Meert, L., Dekker, M.R., Dekkers, D.H.W., et al. (2019). **Mediator complex interaction partners organize the transcriptional network that defines neural stem cells.** *Nat. Commun.* *10*, 2669.
- Sabate, R., Rousseau, F., Schymkowitz, J., and Ventura, S. (2015). **What Makes a Protein Sequence a Prion?.** *PLoS Comput. Biol.* *11*, e1004013.
- Sant'Anna, R., Fernández, M.R., Batlle, C., Navarro, S., et al. (2016). **Characterization of Amyloid Cores in Prion Domains.** *Sci. Rep.* *6*, 34274.
- Schaefer, M.H., Wanker, E.E., and Andrade-Navarro, M.A. (2012). **Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks.** *Nucleic Acids Res.* *40*, 4273–4287.
- Shaikhibrahim, Z., Menon, R., Braun, M., Offermann, A., et al.

- (2014). **MED15, encoding a subunit of the mediator complex, is overexpressed at high frequency in castration-resistant prostate cancer.** *Int. J. Cancer* *135*, 19–26.
- Shaikhibrahim, Z., Offermann, A., Halbach, R., Vogel, W., et al. (2015). **Clinical and molecular implications of MED15 in head and neck squamous cell carcinoma.** *Am. J. Pathol.* *185*, 1114–1122.
- Shorter, J., and Lindquist, S. (2005). **Prions as adaptive conduits of memory and inheritance.** *Nat. Rev. Genet.* *6*, 435–450.
- Sodek, J., Hodges, R.S., Smillie, L.B., and Jurasek, L. (1972). **Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure.** *Proc. Natl. Acad. Sci. U. S. A.* *69*, 3800–3804.
- Syring, I., Weiten, R., Müller, T., Schmidt, D., et al. (2018). **The knockdown of the Mediator complex subunit MED15 restrains urothelial bladder cancer cells' malignancy.** *Oncol. Lett.* *16*, 3013–3021.
- Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., et al. (2017). **The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible.** *Nucleic Acids Res.* *45*, D362–D368.
- Thual, C., Bousset, L., Komar, A.A., Walter, S., et al. (2001). **Stability, folding, dimerization, and assembly properties of the yeast prion Ure2p.** *Biochemistry* *40*, 1764–1773.
- Walavalkar, N.M., Gordon, N., and Williams, D.C. (2013). **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/p66α).** *J. Biol. Chem.* *288*, 3419–3427.
- Walsh, I., Martin, A.J.M., Di Domenico, T., and Tosatto, S.C.E. (2012). **ESpritz: accurate and fast prediction of protein disorder.** *Bioinformatics* *28*, 503–509.
- Wan, L., Chong, S., Xuan, F., Liang, A., et al. (2020). **Impaired cell fate through gain-of-function mutations in a chromatin reader.** *Nature* *577*, 121–126.
- Wands, A.M., Wang, N., Lum, J.K., Hsieh, J., et al. (2011). **Transient-state kinetic analysis of transcriptional activator-DNA complexes interacting with a key coactivator.** *J. Biol. Chem.* *286*, 16238–16245.
- Wang, K., Duan, C., Zou, X., Song, Y., et al. (2018). **Increased mediator complex subunit 15 expression is associated with poor prognosis in hepatocellular carcinoma.** *Oncol. Lett.* *4303–4313*.
- Weiten, R., Müller, T., Schmidt, D., Steiner, S., et al. (2018). **The Mediator complex subunit MED15, a promoter of tumour progression and metastatic spread in renal cell carcinoma.** *Cancer Biomarkers* *21*, 839–847.
- Whitmore, L., and Wallace, B.A. (2008). **Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases.** *Biopolymers* *89*, 392–400.
- Wickner, R.B., Taylor, K.L., Edskes, H.K., Maddelein, M.L., et al. (2001). **Yeast prions act as genes composed of self-propagating protein amyloids.** *Adv. Protein Chem.*
- Wickner, R.B., Edskes, H.K., Shewmaker, F., and Nakayashiki, T. (2007). **Prions of fungi: Inherited structures and biological roles.** *Nat. Rev. Microbiol.* *5*, 611–618.
- Zhao, M., Yang, X., Fu, Y., Wang, H., et al. (2013). **Mediator MED15 modulates transforming growth factor beta (TGFβ)/Smad signaling and breast cancer cell metastasis.** *J. Mol. Cell Biol.* *5*, 57–60.
- Zhu, X., Chen, L., Carlsten, J.O.P., Liu, Q., et al. (2015). **Mediator tail subunits can form amyloid-like aggregates in vivo and affect stress response in yeast.** *Nucleic Acids Res.* *43*, 7306–7314.

7. GENERAL DISCUSSION

The term prion is generally associated with TSE and PrP. However, similar prionogenic behavior was observed in other proteins of eukaryotic and prokaryotic organisms with either pathogenic or functional consequences (*Publication I*). The best-characterized set of functional prions are yeast prions, which have disordered Q/N-rich PFDs responsible for their aggregation and propagation, behaving as epigenetic mechanisms of inheritance. The Lindquist group performed a genome-wide bioinformatics survey in order to identify novel yeast prions based on their compositional similarity to already known PFDs (Alberti et al., 2009). They further experimentally validated their predictions to identify a set of novel *bona-fide* yeast prions. These validations were later used as the core of several prion predictors relying on the so-called compositional model, which proposes that prion amyloid aggregation is driven by a large number of weak interactions along the PFD, for example, PAPA and PLAAC (Lancaster et al., 2014; Toombs et al., 2012). However, we proposed a new model, the short-stretch model, which proposes classical nucleation by short amyloidogenic stretches, whose amyloid propensity is strongly modulated by the structural context (Sabate et al., 2015; Zambrano et al., 2015). We termed this region as amyloid or nucleating core. These amyloid cores are longer than classical amyloid stretches (~21 residues long), enriched with a specific amino acid composition (Asn, Gln, Tyr and Ser), allowing the domain to remain soluble and disordered in physiological conditions but providing enough nucleation force for amyloid formation under defined conditions (*Publication II*). Consequently, the amyloid nucleating potency would be weaker and less concentrated than in pathogenic amyloids and these regions cannot be predicted by classical aggregation predictors such as TANGO or Aggrescan (Conchillo-Sole et al., 2007; Fernandez-Escamilla et al., 2004). According to this model, we developed the prion predictors pWaltz and PrionW (Sabate et al., 2015; Zambrano et al., 2015) and we experimentally validated it for four yeast prions (Sant'Anna et al., 2016) (*Publication III*). Importantly, the two models for yeast prion propagation are not in opposition. The compositional model would delineate the PFD and the short-stretch model would define the region with a higher probability to initiate the conformational conversion. Accordingly, the combination of both models allows better predictions for the mutational impact on hnRNPA2 PrLD (*Publication V*). This study led us to develop further AMYCO, a web server that allows a fast evaluation of the effect of mutations on the aggregation propensities of PrLDs, taking into account both composition and structural context (Iglesias et al., 2019).

Prion-like proteins are proteins from other proteomes, identified by some of the existing prion predictors, to have disordered domains similar in composition to those of yeast PFDs. Therefore, it could be hypothesized that these proteins might experiment a conformational conversion similar to the one of yeast prions, thus being also used for beneficial purposes. However, in humans, these proteins have been rapidly associated with many degenerative diseases (King et al., 2012). Prion-like proteins aggregation mediated by their PrLD has been suggested to be the process responsible for the disease onset, and disease-causing mutations

in these proteins have been reported to accelerate their aggregation, and thus the appearance of the disease. Interestingly, most, if not all, of the identified human prion-like proteins are regulatory nucleic acid-binding proteins, suggesting a common function for this kind of proteins. However, the reason why RNA and DNA-binding proteins are so frequently associated with PrLDs remains mostly unknown.

In this thesis, we have been interested in the study of human prion-like proteins, the characterization of the regions responsible for their aggregation, the effect of disease-causing mutations on their aggregation propensity and their mechanism of assembly. Combining the compositional and short-stretch model for prion-like proteins predictions, we identified a large number of putative prion-like proteins containing a PrLD in the human proteome. To further validate our short-stretch model, we selected seven prion-like candidates and synthesized their respective soft amyloid cores to evaluate their aggregation properties (*Publication IV*). We demonstrated that all of them can self-assemble into amyloid structures of low cytotoxicity. Moreover, they can seed the aggregation of their soluble counterparts according to their prionogenic property. These results converge to indicate that we identified the nucleating region responsible for aggregation in seven human prion-like proteins. It is important to note that the selected size of 21 residues for those amyloid cores is arbitrary and we do not propose that the identified cores are the exclusive players in the aggregation process. We selected 21 residues length because it was the length that best discriminated prion versus non-prion behavior in yeast Q/N-rich domains and also because it was the length of the minimal motif accounting for HET-s prion transmissibility (Wan and Stubbs, 2014). The selected nucleating cores were the highest scores in our predictions, but other short regions that collaborate either in a cooperative or additive manner to the process might exist.

It is interesting to notice that soft amyloid cores resemble IDPs' short linear motifs (SLiMs) properties. SLiMs are short stretches of the protein sequence that mediate protein-protein interactions (PPIs) (Ren et al., 2008). Similarly, soft amyloid cores could contribute to mediate PrLDs PPI and thus play a functional role. From an evolutionary point of view, the idea of human prion-like proteins being only involved in pathogenic processes makes no sense, as they have not been purged out by natural selection, as it should be expected in this case. The presence of these regions is indeed inherently risky since the described disease-causing mutations increase their amyloid potency generating a deleterious phenotype. Nevertheless, we know now that prion-like proteins are involved in the formation of dynamic and reversible MLOs where their PrLDs contribute to the intermolecular interactions sustaining a vast PPI network.

Many prion-like proteins reside in MLOs. Remarkably, there are no fusion events between distinct MLOs, although there can be distinct subcompartments and exist multiphase systems within them, such as in the nucleolus. Therefore, it is essential to understand what determines the specificity and ensures the integrity of these assemblies. For example, it has been suggested that the formation of multiphase MLOs could be mediated by differences in the surface tension of protein droplets (Feric et al., 2016). Besides physical properties, the integrity of MLOs formation may derive from the specificity of homotypic and heterotypic PPI. As

mentioned above, prion-like proteins are enriched in SLiMs, which can serve as primary protein-binding modules (Tompa et al., 2014). The different degrees of cooperativity among multiple SLiMs might determine the composition and material properties of the MLO. Multivalency, post-translational modifications (PTMs), RNA or changes in ionic strength can also contribute to the specificity of the droplets (Burke et al., 2015; Harmon et al., 2017; Nott et al., 2015; Wu and Fuxreiter, 2016). Furthermore, some prion-like proteins have dimerization or oligomerization domains through which they modulate their incorporation to MLOs, for example, G3BP or TDP43 (Conicella et al., 2016; Tourrière et al., 2003). Brangwynne lab demonstrated such a mechanism in “optodroplets”. A plant-derived light-inducible protein oligomerization domain provided an optogenetic switch to trigger PPI, and, when fused to PrLDs, to control LLPS (Shin et al., 2017). Accordingly, these “optodroplets” assembled only upon light stimulation. Finally, coiled-coils (CC) and labile β -zipper can also provide the interactions that drive LLPS, as in FUS protein (Hughes et al., 2018; Kato et al., 2012; Murray et al., 2017).

The formation of many MLOs relies on LLPS, as it occurs in P granules, nucleoli, stress granules, or in DNA damage regions. Novel prion-like proteins have been identified to be mapped into these MLOs, such as hnRNPA2, DDX4, FUS, TAF15 and TIA1. (Boeynaems et al., 2018). Intending to increase the knowledge on MLOs formation and the interactions driving LLPS, we selected a previously uncharacterized prion-like protein with high a score in our prion-like predictions: the heterogeneous ribonucleoprotein D-like (hnRNPD_L) (*Publication VI*). Three isoforms, DL1, DL2 and DL3, are produced by alternative splicing of *HNRNPDL* gene, containing two, one or none IDRs, respectively. We used these naturally existing variants to determine the regions responsible for LLPS and aggregation and studied their behavior in cells. Their self-assembly properties *in vitro* and cells are strikingly different and related to their domain composition. We identified that: i) the interactions responsible for LLPS in hnRNPD_L occur between the Arg residues of the N-terminal IDR and the Tyr residues at the C-terminal IDR; ii) the absence of the N-terminal domain in DL2 significantly enhances amyloid formation, and iii) the absence of IDRs keeps the protein soluble. The behavior of hnRNPD_L isoforms in cells suggests that they are involved in different PPI networks and the formation of complexes with different properties. Therefore, our results reinforce the idea that MLOs specificity is provided by the different PPI established according to the particular multivalency of the involved protein.

Prion-like proteins have evolved a high kinetic barrier to prevent amyloid formation. In contrast, this barrier is very low for MLOs formation as they can be readily broken by PTM or changes in salt, pH, temperature, which allows them to promote adaptive cellular responses. Changes in protein concentration, prolonged time of the stress, or mutations in the protein sequence, provoke a liquid-to-solid state transition of dynamic MLOs into static aggregates. Often mutations target the region of the soft amyloid core in the PrLD, making the protein more prone to aggregation, such as in hnRNPA1 and hnRNPA2 (Kim et al., 2013). This is also the case for hnRNPD_L. DL2, the predominant isoform with one IDR, aggregates into amyloid fibrils, and the described D378N/H disease-causing mutations responsible for LGMD1G accelerate

this process. A toxic gain-of-function phenotype is usually associated with these mutations (Martinez et al., 2016; Winklhofer et al., 2008), but in the hnRNPD case, we suggest that the disease responds to a loss-of-function mechanism, where hnRNPD mutated protein levels are decreased in the muscular tissue of patients (*Publication VII*). It is important to point out that in most cases PrLD aggregation is linked with pathogenesis, but, of course, there exist exceptions where the aggregation kinetic barrier is as low as for MLOs, for example the self-sustaining aggregates of Xvelo and Rim4 proteins (Berchowitz et al., 2015; Boke et al., 2016).

Many prion predictors have been used for the identification of prion-like proteins in proteomes. However, none of these predictions take into account protein diversity, such as alternative splicing isoforms or PTM. The different protein variants generated from a common gene might exhibit significant differences in their self-assembly and aggregation propensities. We observed this phenomenon with hnRNPD isoforms (*Publication VII*), and it was also reported for hnRNPD and FUS isoforms (Gueroussov et al., 2017). The presence of the PrLD in the sequence facilitates the incorporation of the protein to multivalent hnRNP assemblies that, in turn, control the AS of other genes. This situation has been recently appreciated by Ross lab, who developed a modified version of their PAPA predictor, taking into account this human variability (Cascarina and Ross, 2020). They observed that aggregation propensity scores differed considerably for alternative splicing isoforms. This study is important because in their predictions they identified prion-like proteins with isoforms that scored above the PAPA threshold, whereas other isoforms scored below it, for example hnRNPD, the prion-like protein we extensively studied in this thesis, and ILF3, for which we identified the presence of a soft amyloid core. Therefore, in addition to disease-causing mutations, the misregulation of alternative splicing events may also play an important role in disease, as it would lead to an improper cellular balance of a variety of aggregation-prone isoforms.

Maintaining proper PPI networks is fundamental for cell physiology. As described above, prion-like proteins PPIs can be regulated through SLiMs or by alternative splicing. Yeast PFDs have been reported to have an intrinsic propensity to form coiled-coils (CC) and a CC-model seemed to explain better their conformation switches that a disorder to β -sheet transition (Fiumara et al., 2010). CCs facilitate both homotypic and heterotypic PPIs and are enriched in transcription regulating proteins. Therefore, CC regions in prion-like proteins could be yet another mechanism to regulate their PPI network. We observed that a significant number of our human prion-like protein candidates were predicted to have CC regions in their PrLDs, which overlap with polyQ tracts (*MED15*). From this subset, we found a key transcription coactivator, the Mediator complex subunit 15 (MED15), for which we already identified the presence of a soft amyloid core (*Publication IV*). We demonstrated that MED15 contains an N-terminus Q-rich PrLD with discontinuous CC regions responsible for its homodimerization and aggregation into amyloid fibrils. MED15 aggregation involves a progressive conformational shift from an initial α -helical state to a β -sheet structure, which occurs concomitantly with the aggregation of the domain. Disruption of MED15-PrLD CC abolishes amyloid formation, whereas its stabilization accelerates aggregation. Between the first and second CC regions of MED15-PrLD, there is a

disordered connecting region that corresponds to the MED15 soft amyloid core. This leads us to propose that, in MED15-PrLD amyloid formation, it cooperates both the misfolding of disordered regions to β -sheet conformations and CC multimerization. MED15-PrLD aggregates not only *in vitro*, but also in human cells, forming insoluble cytoplasmic and perinuclear inclusions which possess the ability to kidnap the full-length endogenous MED15 in a prion-like manner. Consequently, accurate regulation of MED15, and prion-like proteins with similar behavior, is required to sustain a proper transcription function. Situations that alter homotypic or heterotypic interactions between this kind of prion-like proteins would impact transcription and downstream signaling pathways significantly, resulting in cell malfunction and, potentially, in diseases like cancer.

The identification of the region responsible for aggregation; the study and accurate prediction of disease-causing mutations effect on aggregation propensity; the understanding of the driving forces governing LLPS; the identification of the toxic step during aggregation processes; the characterization of CC mediated PPIs; and the global understanding of prion-like proteins behavior, could eventually provide insights to develop therapeutic strategies to treat the devastating diseases associated with these unique proteins.

8. CONCLUSIONS

Publication I. Prion-like proteins and their computational identification in proteomes:

- Different classifications and terminologies according to prionogenic properties have emerged.
- There exist different prion predictors to uncover proteins with domains similar to yeast PFDs.
- Many proteins with prion-like properties have been identified in other proteomes than yeast.

Publication II. Amyloids or prions? That is the question:

- Yeast prions and prion-like proteins might have an amyloid core responsible for their aggregation and propagation.
- The amyloid core should allow the domain to remain soluble and disordered, but also provide the nucleation force for amyloid formation.

Publication III. Amyloid cores in prion domains: key regulators for prion conformational conversion:

- Sup35, Ure2, Swi1 and Mot3 have an amyloid core in their PFD responsible for their aggregation and propagation.
- Amyloid cores could have different length and other short regions can contribute cooperatively or additively to the aggregation process.

Publication IV. Characterization of soft amyloid cores in human prion-like proteins:

- The experimental data validates the accuracy of the designed computational system to detect putative prion-like proteins.
- The seven soft amyloid cores selected from the human proteome experiment a conformational conversion from a predicted soluble and disordered state into an amyloid-like structure.
- The fibrils of the predicted soft amyloid cores are able to seed the aggregation of the soluble species.
- The human proteins displaying prionogenic features could be more abundant than previously assumed, and they could play important regulatory roles in health and disease, to which preferential PPI between this specific protein subclass might contribute significantly.

Publication V. Perfecting prediction of mutational impact on the aggregation propensity of the ALS-associated hnRNPA2 prion-like protein:

- The ability of hnRNPA2 PrLD to assemble into pathogenic macromolecular structures depends on both a special amino acid composition and on the presence of at least one specific soft amyloid core.

- The combination of the compositional and amyloid-stretch model provides a tool to identify the occurrence of disease-associated mutations in other prion-like proteins.

Publication VI. *hnRNPD*L phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation:

- hnRNPDL isoforms possess different phase separation propensity *in vitro*.
- The unique combination of IDRs in hnRNPDL isoforms is an important determinant of its self-assembly and localization in cells.
- hnRNPDL isoforms possess different aggregation propensity *in vitro*.
- Disease-causing mutations accelerate the aggregation process *in vitro*.
- Disease-causing mutations abolish hnRNPDL solubility in *Drosophila* muscle.

Publication VII. *Prion-like domain disease-causing mutations and misregulation of alternative splicing relevance in limb-girdle muscular dystrophy (LGMD) 1G:*

- Disease-causing mutations in prion-like proteins can promote both loss- and gain-of-function phenotypes.
- Alternative splicing is an essential and generic mechanism that regulates prion-like proteins' biological function in cells.

The Q-rich prion-like domain of human MED15 forms a coiled-coil responsible for its conversion to amyloids and its propagation:

- A significant number of human prion-like domains (PrLDs) are predicted to have CC regions which overlap with polyQ tracts, and they are functionally biased towards transcriptional regulation.
- MED15 N-terminus Q-rich PrLD forms CC homodimers in solution.
- MED15-PrLD aggregates into Congo Red positive, β -sheet enriched amyloid fibrils, displaying concentration-dependent aggregation kinetics and self-seeding activity.
- Chemical or genetic disruption of MED15-PrLD CC abolishes amyloid formation, whereas its stabilization accelerates it.
- Both the misfolding of disordered regions to β -sheet conformations and CC multimerization cooperate to promote MED15-PrLD amyloid formation.
- MED15-PrLD forms cytoplasmic and perinuclear inclusions in human cells, kidnapping the endogenous full-length protein to these aggregates in a prion-like manner.

9. REFERENCES

- Aguzzi, A., and Calella, A.M. (2009). **Prions: protein aggregation and infectious diseases.** *Physiol. Rev.* *89*, 1105–1152.
- Alberti, S., Halfmann, R., King, O., Kapila, A., et al. (2009). **A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins.** *Cell* *137*, 146–158.
- Anderson, P., and Kedersha, N. (2008). **Stress granules: the Tao of RNA triage.** *Trends Biochem. Sci.* *33*, 141–150.
- Battle, C., De Groot, N.S., Iglesias, V., Navarro, S., et al. (2017). **Characterization of Soft Amyloid Cores in Human Prion-Like Proteins.** *Sci. Rep.* *7*, 1–16.
- Berchowitz, L.E., Kabachinski, G., Walker, M.R., Carlile, T.M., et al. (2015). **Regulated Formation of an Amyloid-like Translational Repressor Governs Gametogenesis.** *Cell* *163*, 406–418.
- Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., et al. (2018). **Protein Phase Separation: A New Phase in Cell Biology.** *Trends Cell Biol.* *xx*, 1–16.
- Boke, E., Ruer, M., Wühr, M., Coughlin, M., et al. (2016). **Amyloid-like Self-Assembly of a Cellular Compartment.** *Cell* *166*, 637–650.
- Bolognesi, B., Gotor, N.L., Dhar, R., Cirillo, D., et al. (2016). **A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression.** *Cell Rep.* *16*, 222–232.
- Bolognesi, B., Faure, A.J., Seuma, M., Schmiedel, J.M., et al. (2019). **The mutational landscape of a prion-like domain.** *Nat. Commun.* *10*.
- Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., et al. (2009). **Germline P granules are liquid droplets that localize by controlled dissolution/condensation.** *Science* (80-.). *324*, 1729–1732.
- Brangwynne, C.P., Mitchison, T.J., and Hyman, A.A. (2011). **Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes.** *Proc. Natl. Acad. Sci. U. S. A.* *108*, 4334–4339.
- Burke, K.A., Janke, A.M., Rhine, C.L., and Fawzi, N.L. (2015). **Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II.** *Mol. Cell* *60*, 231–241.
- Cascarina, S.M., and Ross, E.D. (2020). **Natural and pathogenic protein sequence variation affecting prion-like domains within and across human proteomes.** *BMC Genomics* *21*, 23.
- Chakrabortee, S., Kayatekin, C., Newby, G.A., Mendillo, M.L., et al. (2016). **Luminidependens (LD) is an Arabidopsis protein with prion behavior.** *Proc. Natl. Acad. Sci. U. S. A.* *113*, 201604478.
- Chiti, F., and Dobson, C.M. (2006). **Protein misfolding, functional amyloid, and human disease.** *Annu. Rev. Biochem.* *75*, 333–366.
- Conchillo-Sole, O., de Groot, N.S., Avilés, F.X., Vendrell, J., et al. (2007). **AGGRESCAN: a**

- server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics* 8, 65.
- Conicella, A.E., Zerze, G.H., Mittal, J., and Fawzi, N.L. (2016). **ALS Mutations Disrupt Phase Separation Mediated by α -Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain.** *Structure* 24, 1537–1549.
 - Creighton, T.E. (1993). **Proteins : structures and molecular properties** (W.H. Freeman).
 - Dobson, C.M. (2003). **Protein folding and misfolding.** *Nature* 426, 884–890.
 - Elbaum-Garfinkle, S. (2019). **Matter over mind: Liquid phase separation and neurodegeneration.** *J. Biol. Chem.* jbc.REV118.001188.
 - Espinosa Angarica, V., Angulo, A., Giner, A., Losilla, G., et al. (2014). **PrionScan: an online database of predicted prion domains in complete proteomes.** *BMC Genomics* 15, 102.
 - Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., et al. (2016). **Coexisting Liquid Phases Underlie Nucleolar Subcompartments.** *Cell* 165, 1686–1697.
 - Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). **Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins.** *Nat. Biotechnol.* 22, 1302–1306.
 - Fiumara, F., Fioriti, L., Kandel, E.R., and Hendrickson, W.A. (2010). **Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins.** *Cell* 143, 1121–1135.
 - Franzmann, T., and Alberti, S. (2018). **Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior.** *J. Biol. Chem.* jbc.TM118.001190.
 - Franzmann, T.M., Jahnel, M., Pozniakovsky, A., Mahamid, J., et al. (2018). **Phase separation of a yeast prion protein promotes cellular fitness.** *Cell Biol.* 5654.
 - De Groot, N.S., Pallarés, I., Avilés, F.X., Vendrell, J., et al. (2005). **Prediction of “hot spots” of aggregation in disease-linked polypeptides.** *BMC Struct. Biol.* 5.
 - Gueroussov, S., Weatheritt, R.J., O’Hanlon, D., Lin, Z.Y., et al. (2017). **Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing.** *Cell* 170, 324-339.e23.
 - Halfmann, R., Alberti, S., and Lindquist, S. (2010). **Prions, protein homeostasis, and phenotypic diversity.** *Trends Cell Biol.* 20, 125–133.
 - Han, T.W., Kato, M., Xie, S., Wu, L.C., et al. (2012). **Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies.** *Cell* 149, 768–779.
 - Harbi, D., Parthiban, M., Gendoo, D.M.A., Ehsani, S., et al. (2012). **Prionhome: A database of prions and other sequences relevant to Prion phenomena.** *PLoS One* 7, 1–11.
 - Harmon, T.S., Holehouse, A.S., Rosen, M.K., and Pappu, R. V. (2017). **Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins.** *Elife* 6.
 - Harrison, A.F., and Shorter, J. (2017). **RNA-binding proteins with prion-like domains in health and disease.** *Biochem. J.* 474, 1417–1438.
 - Harrison, P.M., and Gerstein, M. (2003). **A method to assess compositional bias in**

biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. *Genome Biol.* 4, R40.

- Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). **Molecular chaperones in protein folding and proteostasis.** *Nature* 475, 324–332.

- Hofweber, M., and Dormann, D. (2018). **Friend or foe — post-translational modifications as regulators of phase separation and RNP granule dynamics.** *J. Biol. Chem.* jbc.TM118.001189.

- Hughes, M.P., Sawaya, M.R., Boyer, D.R., Goldschmidt, L., et al. (2018). **Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks.** *Science* (80-).

- Hutten, S., and Dormann, D. (2016). **hnRNPA2/B1 Function in Neurodegeneration: It's a Gain, Not a Loss.** *Neuron* 92, 672–674.

- Iglesias, V., Conchillo-Sole, O., Batlle, C., and Ventura, S. (2019). **AMYCO: Evaluation of mutational impact on prion-like proteins aggregation propensity.** *BMC Bioinformatics* 20.

- Invernizzi, G., Papaleo, E., Sabate, R., and Ventura, S. (2012). **Protein aggregation: Mechanisms and functional consequences.** *Int. J. Biochem. Cell Biol.* 44, 1541–1554.

- Ito, D., Hatano, M., and Suzuki, N. (2017). **RNA binding proteins and the pathological cascade in ALS / FTD neurodegeneration.** 1–11.

- Kato, M., Han, T.W., Xie, S., Shi, K., et al. (2012). **Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels.** *Cell* 149, 753–767.

- Kim, H.J., Kim, N.C., Wang, Y.-D., Scarborough, E.A., et al. (2013). **Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS.** *Nature* 495, 467–473.

- King, O.D., Gitler, A.D., and Shorter, J. (2012). **The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease.** *Brain Res.* 1462, 61–80.

- Kraus, A., Groveman, B.R., and Caughey, B. (2013). **Prions and the potential transmissibility of protein misfolding diseases.** *Annu. Rev. Microbiol.* 67, 543–564.

- Lancaster, A.K., Nutter-Upham, A., Lindquist, S., and King, O.D. (2014). **PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition.** *Bioinformatics* 30, 2501–2502.

- Li, Y.R., King, O.D., Shorter, J., and Gitler, A.D. (2013). **Stress granules as crucibles of ALS pathogenesis.** *J. Cell Biol.* 201, 361–372.

- Liebman, S.W., and Chernoff, Y.O. (2012). **Prions in yeast.** *Genetics* 191, 1041–1072.

- López De La Paz, M., and Serrano, L. (2004). **Sequence determinants of amyloid fibril formation.** *Proc. Natl. Acad. Sci. U. S. A.* 101, 87–92.

- Mackenzie, I.R., Nicholson, A.M., Sarkar, M., Messing, J., et al. (2017). **TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics.** *Neuron* 95, 808-816.e9.

- Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., et al. (2018). **RNA buffers the phase separation behavior of prion-like RNA binding proteins.** *Science* (80-). 7366, 1–10.

- Malinovska, L., and Alberti, S. (2015). **Protein misfolding in Dictyostelium: Using a freak of nature to gain insight into a universal problem.** *Prion* 9, 339–346.
- Malinovska, L., Palm, S., Gibson, K., Verbavatz, J.-M., et al. (2015). **Dictyostelium discoideum has a highly Q/N-rich proteome and shows an unusual resilience to protein aggregation.** *Proc. Natl. Acad. Sci.* 112, 201504459.
- Martinez, F.J., Pratt, G.A., Van Nostrand, E.L., Batra, R., et al. (2016). **Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System.** *Neuron* 0, 1077–1087.
- Michelitsch, M.D., and Weissman, J.S. (2000). **A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions.** *Proc. Natl. Acad. Sci. U. S. A.* 97, 11910–11915.
- Muralidharan, V., Oksman, A., Pal, P., Lindquist, S., et al. (2012). **Plasmodium falciparum heat shock protein 110 stabilizes the asparagine repeat-rich parasite proteome during malarial fevers.** *Nat. Commun.* 3.
- Murray, D.T., Kato, M., Lin, Y., Thurber, K.R., et al. (2017). **Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains.** *Cell* 171, 615-627.e16.
- Nedelsky, N.B., and Taylor, J.P. (2019). **Bridging biophysics and neurology: aberrant phase transitions in neurodegenerative disease.** *Nat. Rev. Neurol.*
- Newby, G.A., and Lindquist, S. (2013). **Blessings in disguise: Biological benefits of prion-like mechanisms.** *Trends Cell Biol.* 23, 251–259.
- Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., et al. (2015). **Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles.** *Mol. Cell* 57, 936–947.
- Owen, I., and Shewmaker, F. (2019). **The role of post-translational modifications in the phase transitions of intrinsically disordered proteins.** *Int. J. Mol. Sci.* 20, 5501.
- Pallarès, I., Iglesias, V., and Ventura, S. (2016). **The rho termination factor of Clostridium botulinum contains a prion-like domain with a highly amyloidogenic core.** *Front. Microbiol.* 6, 1–12.
- Patel, A., Lee, H.O., Jawerth, L., Maharana, S., et al. (2015). **A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation.** *Cell* 162, 1066–1077.
- Patel, A., Malinovska, L., Saha, S., Wang, J., et al. (2017). **ATP as a biological hydrotrope.** *Science* (80-.). 356, 753–756.
- Prusiner, S.B. (1982). **Novel proteinaceous infectious particles cause scrapie.** *Science* 216, 136–144.
- Ren, S., Uversky, V.N., Chen, Z., Dunker, A.K., et al. (2008). Short Linear Motifs recognized by SH2, SH3 and Ser/Thr Kinase domains are conserved in disordered protein regions. In *BMC Genomics*, p.
- Sabate, R., Rousseau, F., Schymkowitz, J., and Ventura, S. (2015). **What Makes a Protein Sequence a Prion?.** *PLoS Comput. Biol.* 11, e1004013.

- Sant'Anna, R., Fernández, M.R., Batlle, C., Navarro, S., et al. (2016). **Characterization of Amyloid Cores in Prion Domains**. *Sci. Rep.* *6*, 34274.
- Shin, Y., Berry, J., Pannucci, N., Haataja, M.P., et al. (2017). **Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets**. *Cell* *168*, 159-171.e14.
- Singh, G.P., Chandra, B.R., Bhattacharya, A., Akhouri, R.R., et al. (2004). **Hyper-expansion of asparagines correlates with an abundance of proteins with prion-like domains in Plasmodium falciparum**. *Mol. Biochem. Parasitol.* *137*, 307–319.
- Tompa, P., Davey, N.E., Gibson, T.J., and Babu, M.M. (2014). **A Million peptide motifs for the molecular biologist**. *Mol. Cell* *55*, 161–169.
- Toombs, J. a., Petri, M., Paul, K.R., Kan, G.Y., et al. (2012). **De novo design of synthetic prion domains**. *Proc. Natl. Acad. Sci.* *109*, 6519–6524.
- Tourrière, H., Chebli, K., Zekri, L., Courselaud, B., et al. (2003). **The RasGAP-associated endoribonuclease G3BP assembles stress granules**. *J. Cell Biol.* *160*, 823–831.
- Wan, W., and Stubbs, G. (2014). **Fungal prion HET-s as a model for structural complexity and self-propagation in prions**. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 5201–5206.
- Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., et al. (2018). **A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins**. *Cell* *0*, 1–12.
- Winklhofer, K.F., Tatzelt, J., and Haass, C. (2008). **The two faces of protein misfolding: Gain- and loss-of-function in neurodegenerative diseases**. *EMBO J.* *27*, 336–349.
- Woodruff, J.B., Hyman, A.A., and Boke, E. (2018). **Organization and Function of Non-dynamic Biomolecular Condensates**. *Trends Biochem. Sci.* *43*, 81–94.
- Wu, H., and Fuxreiter, M. (2016). **The Structure and Dynamics of Higher-Order Assemblies: Amyloids, Signalosomes, and Granules**. *Cell* *165*, 1055–1066.
- Yuan, A.H., and Hochschild, A. (2017). **A bacterial global regulator forms a prion**. *1*, 1–27.
- Zambrano, R., Conchillo-Sole, O., Iglesias, V., Illa, R., et al. (2015). **PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores**. *Nucleic Acids Res.* 1–7.

10.1. ANNEX. Supplementary material. PUBLICATION IV

Characterization of soft amyloid cores in human prion-like proteins.

Battle C., de Groot NS., Iglesias V., Navarro S. and Ventura S.

Sci Rep (2017)

DOI: [10.1038/s41598-017-09714-z](https://doi.org/10.1038/s41598-017-09714-z)

Supplementary information:

Characterization of Soft Amyloid Cores in Human Prion-Like Proteins

Cristina Batlle¹, Natalia S. de Groot^{2,3}, Valentin Iglesias¹, Susanna Navarro¹, Salvador Ventura¹

¹Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular,
Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain

²Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader
88, 08003 Barcelona, Spain.

³Universitat Pompeu Fabra (UPF), Barcelona, Spain.

SUPPLEMENTARY FIGURES

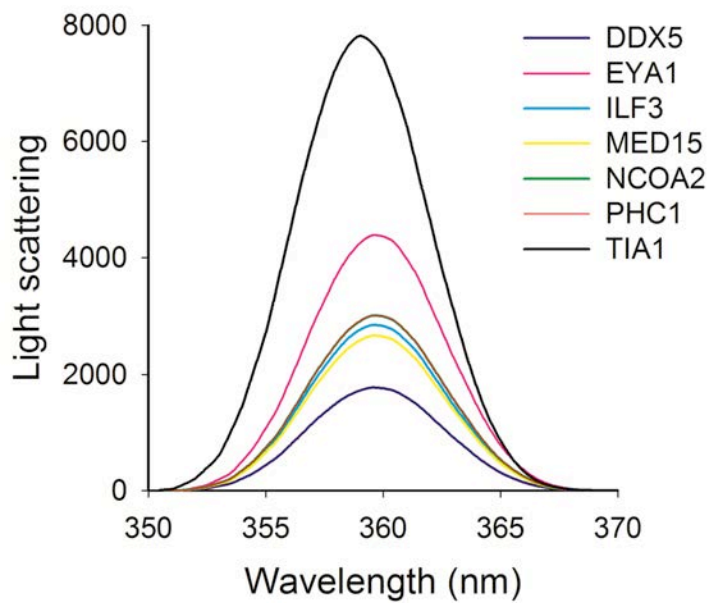


Figure S1. Aggregation of human PrLD amyloid cores. The ability to form macromolecular structures was measured by synchronous light scattering. All selected peptides exhibited a significant scattering signal at 100 μ M in 5 mM potassium phosphate buffer at pH 7.4 after incubation at 37 $^{\circ}$ C for 2 days.

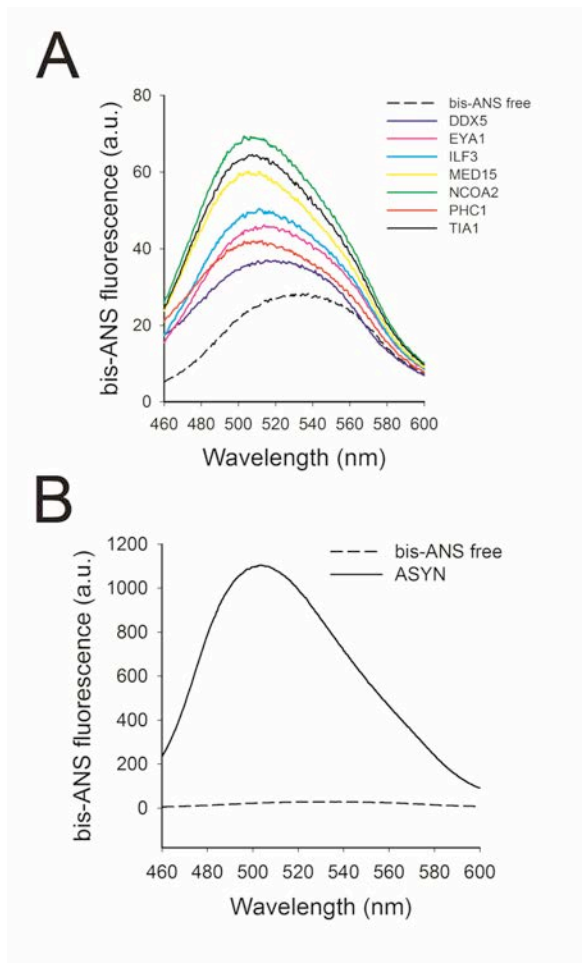


Figure S2. Binding of bis-ANS to aggregated human PrLD amyloid cores. A) Spectra of bis-ANS in the presence of the seven peptides after their incubation at 100 μ M in 5 mM potassium phosphate buffer at pH 7.4 and 37 $^{\circ}$ C for 2 days (color solid lines). The spectrum of free bis-ANS is shown as a dashed line. B) Bis-ANS binding for α -synuclein (ASYN) at 100 μ M in 5 mM potassium phosphate buffer at pH 7.4 incubated at 37 $^{\circ}$ C for 7 days.

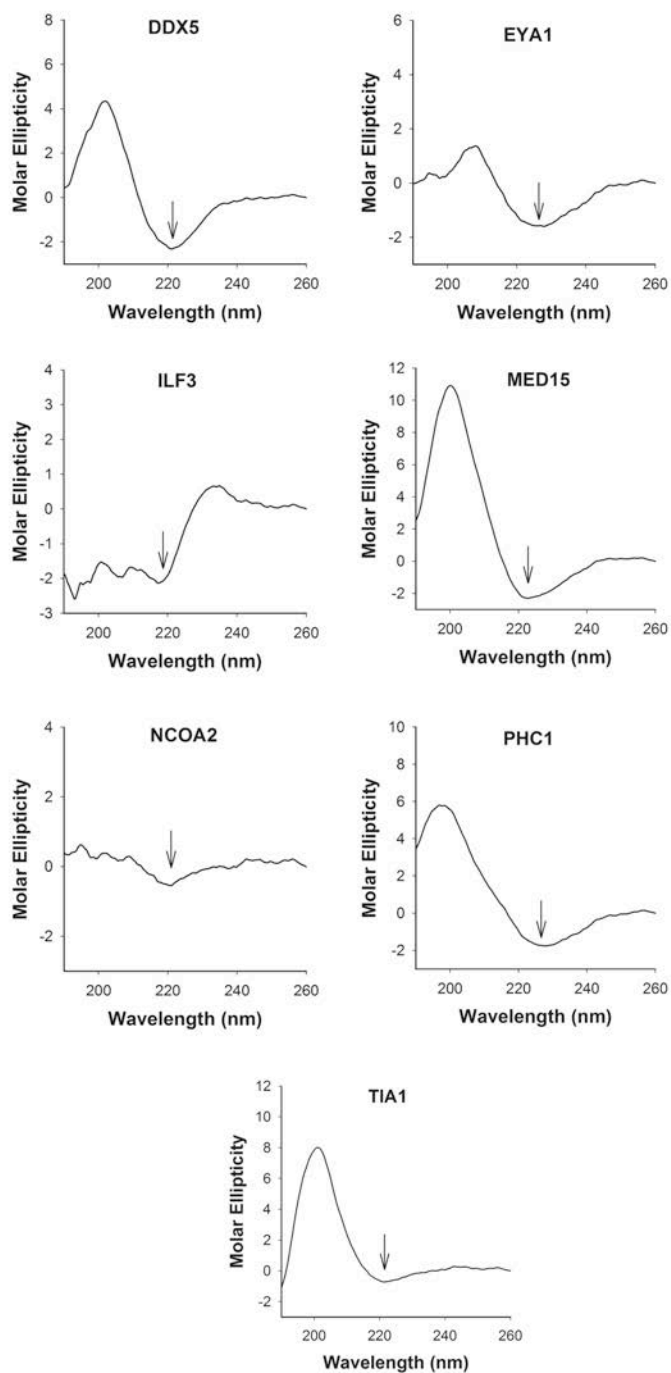


Figure S3. CD spectra of aggregated human PrLD amyloid cores. CD spectra in the far-UV region for peptides incubated at 100 μ M in 5 mM potassium phosphate buffer at pH 7.4 and 37 $^{\circ}$ C for 2 days. The arrow indicates the spectra minima correspondent to β -sheet secondary structure.

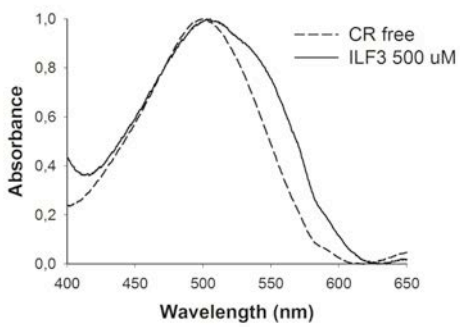
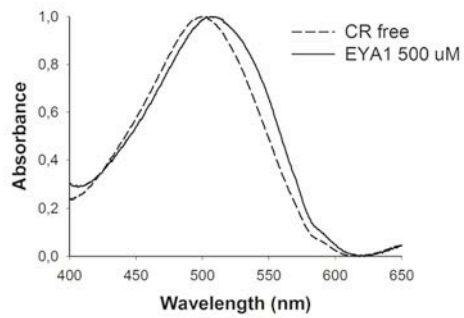
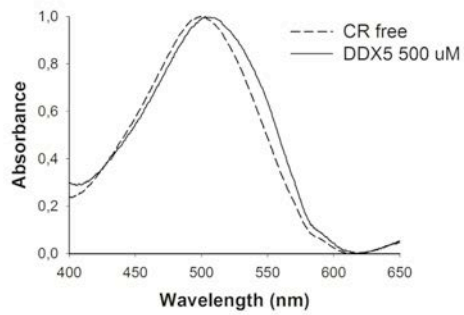


Figure S4. CR binding to DDX5, EYA1 and ILF3 amyloid cores. CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of DDX5, EYA1 and ILF3 peptides incubated at 500 μM in 5 mM potassium phosphate buffer at pH 7.4 and 37 $^{\circ}\text{C}$ for 2 days.

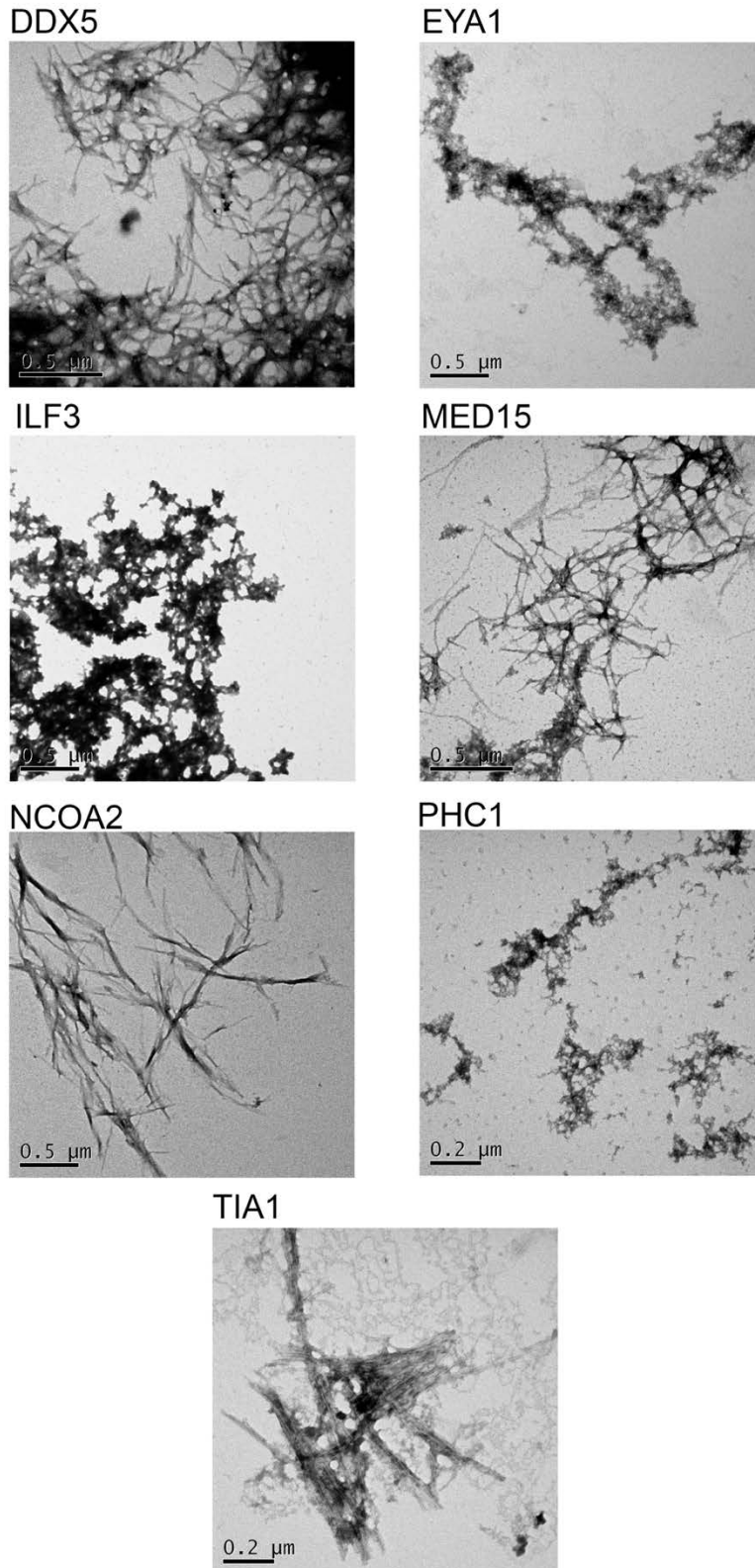


Figure S5. Transmission electron micrographs of peptides incubated at 10 μM. The images were acquired upon incubation of the peptides for 5 days in 5 mM potassium phosphate buffer pH 7.4 at 37 °C.

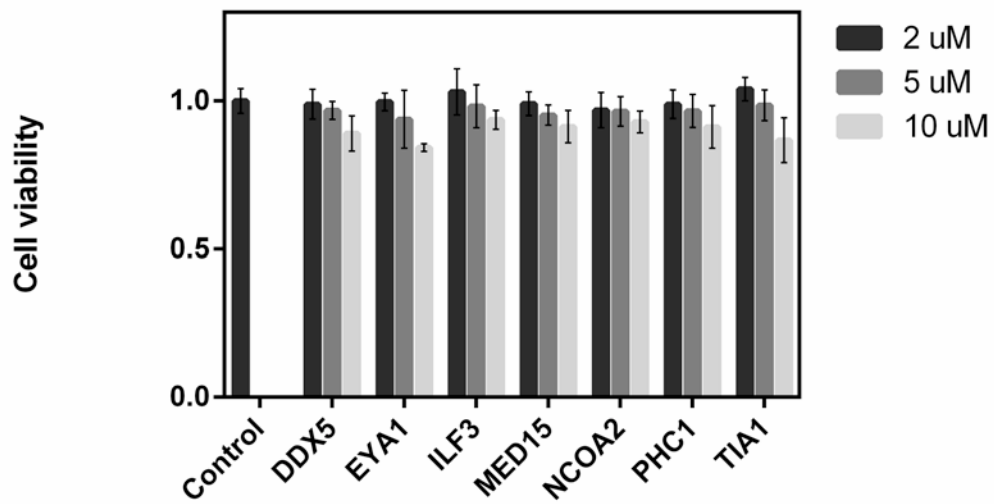


Figure S6. Cytotoxicity of aggregated human PrLD amyloid cores. Cell viability was calculated relative to cells treated with PBS1x buffer (100% viability). The assays were done in five wells for each concentration and the experiment was repeated twice. The data show the averaged values of two independent experiments, and the bars represent the standard error of the mean. Peptides were added at 2, 5 or 10 μ M final concentration to SH-SHY cells.

SUPPLEMENTARY TABLES

Table S1. Human PrLD-containing proteins. Upon an orthogonal screening of the human proteome with PAPA¹ and pWALTZ² we detected 535 human proteins bearing putative PrLDs. For each candidate it is shown its PAPA score, its amyloid core sequence, the pWALTZ score for this protein region and its Ensembl ID according to Uniprot³.

HUMAN PrLD-CONTAINING PROTEINS				
Uniprot ID	PAPA Score	pWALTZ Amyloid Core	pWALTZ Score	Ensembl ID
Q9UN88	0,0866361	CLFFVFLSLLEYVYINLYFYS	89,8709	ENSG00000268089
O43300	0,07596259	ITGTMALLFSFFFIIFIVFIS	89,6002	ENSG00000146006
Q9P035	0,06972523	VRFSFFLQIYLIMIFLGLYIN	89,425	ENSG00000074696
H3BRL8	0,06972523	VRFSFFLQIYLIMIFLGLYIN	89,425	ENSG00000074696
H3BPZ1	0,06972523	VRFSFFLQIYLIMIFLGLYIN	89,425	ENSG00000074696
H3BS72	0,06972523	VRFSFFLQIYLIMIFLGLYIN	89,425	ENSG00000074696
Q13936	0,05912466	KSNVFWLVIFLVFLNLTIA	88,1191	ENSG00000151067
F5H638	0,05912466	KSNVFWLVIFLVFLNLTIA	88,1191	ENSG00000151067
A0A0A0MSA1	0,05912466	KSNVFWLVIFLVFLNLTIA	88,1191	ENSG00000151067
A0A0A0MR67	0,05912466	KSNVFWLVIFLVFLNLTIA	88,1191	ENSG00000151067
F5H522	0,05912466	KSNVFWLVIFLVFLNLTIA	88,1191	ENSG00000151067
F5GY28	0,05912466	KSNVFWLVIFLVFLNLTIA	88,1191	ENSG00000151067
E9PDI6	0,05912466	KSNVFWLVIFLVFLNLTIA	88,1191	ENSG00000151067
H7BYL8	0,05387899	GYVFFEYQYVDNNIFFEFFIQ	87,4343	ENSG00000164659
A8MWY0	0,05021104	GYVFFEYQYVDNNIFFEFFIQ	87,4343	ENSG00000164659
H7C2N5	0,05021104	GYVFFEYQYVDNNIFFEFFIQ	87,4343	ENSG00000164659
C9JA41	0,05021104	GYVFFEYQYVDNNIFFEFFIQ	87,4343	ENSG00000164659
Q9H7F0	0,05578504	NYENTTVFFISSFQYLIVAIA	86,8609	ENSG00000133657
H7C4S8	0,07745829	KSVTFYWLIVLVFLNLTIS	86,4309	ENSG00000157388
Q01668	0,06549277	KSVTFYWLIVLVFLNLTIS	86,4309	ENSG00000157388
A0A1BOGUB6	0,06549277	KSVTFYWLIVLVFLNLTIS	86,4309	ENSG00000157388
A0A1BOGTN0	0,06549277	KSVTFYWLIVLVFLNLTIS	86,4309	ENSG00000157388
A0A1BOGUN6	0,06549277	KSVTFYWLIVLVFLNLTIS	86,4309	ENSG00000157388
Q59GD8	0,06429075	KSVTFYWLIVLVFLNLTIS	86,4309	ENSG00000157388
A0A1BOGWE1	0,06429075	KSVTFYWLIVLVFLNLTIS	86,4309	ENSG00000157388
Q9NX07	0,1563495	YSQMYSYSYNQYQQYQNYA	85,2683	ENSG00000180098
Q13324	0,06239437	LSQIMFIYFNSFLQSFQGFV	84,9976	ENSG00000106113
Q6PJF5	0,06162701	FTYWLTFVHVLIITLLVICTYG	84,8224	ENSG00000129667
P34998	0,0566561	FIYFNSFLESFQGFVSVFYC	84,4879	ENSG00000120088;ENSG00000278232;ENSG00000276191
B3TIK8	0,0566561	FIYFNSFLESFQGFVSVFYC	84,4879	ENSG00000120088;ENSG00000278232;ENSG00000276191
Q9NR82	0,0714548	RGWAFIYHAFVLLVFGCLIL	84,0101	ENSG00000185760
Q96G97	0,07405873	ASNFTFLSVIVLFSYMQVWVG	83,7553	ENSG00000168000
J3KQ12	0,05574767	ASNFTFLSVIVLFSYMQVWVG	83,7553	ENSG00000168000
P51861	0,07019652	GYWKTWIFWKTWIFWKTWIFR	83,3572	ENSG00000184258
Q9Y4I1	0,05922493	DIYGFETFEINSFEQFCINYA	83,2138	ENSG00000197535
F8W6H6	0,05922493	DIYGFETFEINSFEQFCINYA	83,2138	ENSG00000197535
G3V394	0,05922493	DIYGFETFEINSFEQFCINYA	83,2138	ENSG00000197535
F8WE88	0,05922493	DIYGFETFEINSFEQFCINYA	83,2138	ENSG00000197535
A0A087WY00	0,05922493	DIYGFETFEINSFEQFCINYA	83,2138	ENSG00000197535
Q9ULS6	0,05845085	LCVFSFSQEIEYWFGINEFFID	83,0387	ENSG00000156486
Q6V1P9	0,06386404	LVTFSNIDHDWTRENTYVEYS	82,6405	ENSG00000197410
A0A096LNH0	0,06386404	LVTFSNIDHDWTRENTYVEYS	82,6405	ENSG00000197410
Q96G30	0,05753923	AHKYSIVIGFVWGLAVFVIFM	82,5449	ENSG00000135324
H7BZA3	0,11752418	GQQWIWLQTHYYITYHQWNSK	82,4335	ENSG00000170485
O14979	0,11734311	QQQWNNQGFNNYYDQGYGNYN	82,2742	ENSG00000152795
P35749	0,05757759	DIAGFEIFEVNSFEQLCINYT	82,1149	ENSG00000133392;ENSG00000276480
Q6EIG7	0,05809473	NTEAEQNFIYQQLNESFSYFL	81,892	ENSG00000205846
Q14330	0,11151439	AFTTFLMNLSTCLDVILYYIV	81,876	ENSG00000125245

Q8NEV4	0,07197329	NEQIQYYYNQHVFAWEQNEYL	81,8442	ENSG00000095777
O15516	0,09271811	GQQWIWLQTHYYITYHQWNSR	81,7646	ENSG00000134852
Q9UKN7	0,06258927	FEQLCINYANENLQYLFNKIV	81,7486	ENSG00000091536
A0A087WYA1	0,06258927	FEQLCINYANENLQYLFNKIV	81,7486	ENSG00000091536
Q9UM54	0,08939663	DIAGFEYFEHNSFEQFCINYC	81,7009	ENSG00000196586
E7EW20	0,08939663	DIAGFEYFEHNSFEQFCINYC	81,7009	ENSG00000196586
A0A0A0MRM8	0,08939663	DIAGFEYFEHNSFEQFCINYC	81,7009	ENSG00000196586
A0A0D9SGC1	0,08939663	DIAGFEYFEHNSFEQFCINYC	81,7009	ENSG00000196586
O95178	0,06500774	QVFQSEFFSGLMWFILWRFW	81,6212	ENSG00000090266
H7C5B8	0,06500774	QVFQSEFFSGLMWFILWRFW	81,6212	ENSG00000090266
C9IZW8	0,05916781	QVFQSEFFSGLMWFILWRFW	81,6212	ENSG00000090266
C9JXM4	0,05857762	QVFQSEFFSGLMWFILWRFW	81,6212	ENSG00000090266
Q13459	0,10386017	RNSFEQFCINYANEQLQYYFN	81,4779	ENSG00000099331
M0R0P8	0,10386017	RNSFEQFCINYANEQLQYYFN	81,4779	ENSG00000099331
M0R300	0,10386017	RNSFEQFCINYANEQLQYYFN	81,4779	ENSG00000099331
Q8N119	0,09105683	GEVMVRFSTYFFRNSWYWLVE	81,4142	ENSG00000154485
Q8NHU2	0,0599964	NEMDIEYIRSHYNIEDFIYFS	81,3983	ENSG00000089101
P42261	0,06674154	AYEIWMCIVFAYIGVSVLFL	81,1594	ENSG00000155511
P42262	0,06542405	AYEIWMCIVFAYIGVSVLFL	81,1594	ENSG00000120251
F8W7L6	0,06542405	AYEIWMCIVFAYIGVSVLFL	81,1594	ENSG00000120251
E7EWC7	0,16652972	AEASALQQQQYYQWYQQYNYA	81,1434	ENSG00000167615
A0A0G2JPW6	0,16652972	AEASALQQQQYYQWYQQYNYA	81,1434	ENSG00000276681
A0A087WUE4	0,16652972	AEASALQQQQYYQWYQQYNYA	81,1434	ENSG00000167615
Q96PV6	0,12227261	AEASALQQQQYYQWYQQYNYA	81,1434	ENSG00000167615;ENSG00000276681;ENSG000000276458;ENSG00000274305
A0A087WTE7	0,12227261	AEASALQQQQYYQWYQQYNYA	81,1434	ENSG00000274305;ENSG00000276681;ENSG00000167615;ENSG00000276458
C9JMY0	0,12227261	AEASALQQQQYYQWYQQYNYA	81,1434	ENSG00000167615
A0A0G2JN70	0,12227261	AEASALQQQQYYQWYQQYNYA	81,1434	ENSG00000276681
Q9UKF5	0,0936033	WWIHFRIVEIVVVIDNLYLIR	81,1116	ENSG00000168594
Q9Y6U3	0,07828965	VDQNSYGEFYGGDCYIILYTY	81,0957	ENSG00000006747
Q9UL59	0,05085828	QEEKFRYLEYENFSYWQGWVN	81,0638	ENSG00000149050
E3W988	0,06579112	YTVHLKQRYFLADNFMIIYLYN	80,809	ENSG00000197140;ENSG00000275594
Q8WXR4	0,13605077	RNSFEQLCINIANEQIQYYFN	80,7134	ENSG00000071909
F5H2J1	0,13605077	RNSFEQLCINIANEQIQYYFN	80,7134	ENSG00000071909
A0A1B0GUS7	0,06123439	DGQYIYSLTDTSTGQYAYLFI	80,602	ENSG00000198722
A4D1P6	0,05156147	YDENTVYSIGEDGKFIQWNIH	80,5701	ENSG00000105875
C9J1X0	0,05156147	YDENTVYSIGEDGKFIQWNIH	80,5701	ENSG00000105875
Q6PIF6	0,09351191	DIFGFENFENNSFEQLCINF	80,5223	ENSG00000169994
Q9BVK6	0,05865928	QRVLWWSIQLTLILVAIGVWQ	80,5223	ENSG00000184840
Q9BQ31	0,06148551	LCVFSFCQIEYWGINELFID	80,4268	ENSG00000170745
Q9H114	0,0754019	STLNFFIQSYNNASNDTYLYR	80,3949	ENSG00000125823
P10643	0,05645944	FQVKINNDFNYEFYNSTWSYV	80,379	ENSG00000112936
Q8WW62	0,0607709	IQSNYNYVNWWSAQSLVIL	80,2675	ENSG00000157315
Q16099	0,10851645	ARVLNSNYAFLESTMNEYR	80,156	ENSG00000149403
Q8IXK2	0,05518623	GHQVILYLCHGMGQNFQFEYT	80,1242	ENSG00000119514
Q02817	0,06500462	FDGLYYSYQGNCTYVVLVEEIS	80,0605	#N/A
A0A0G2JR65	0,06500462	FDGLYYSYQGNCTYVVLVEEIS	80,0605	ENSG00000278466
Q86WI1	0,05742728	TNISYTSTFYGFKEEDYVVIS	79,9012	ENSG00000205038
Q8NDX9	0,12844553	QYISYRCQEKRNTYFAEYWYQ	79,8375	ENSG00000240053;ENSG00000239285;ENSG000000241132;ENSG00000240433;ENSG00000241713;ENSG00000244672;ENSG00000239497
H0Y6P8	0,12844553	QYISYRCQEKRNTYFAEYWYQ	79,8375	ENSG00000263020;ENSG00000258589;ENSG00000258543
A0A0G2JM12	0,12844553	QYISYRCQEKRNTYFAEYWYQ	79,8375	ENSG00000224774;ENSG00000206406
N0E472	0,12844553	QYISYRCQEKRNTYFAEYWYQ	79,8375	ENSG00000258589;ENSG00000263020;ENSG00000228875;ENSG00000224398;EN

				SG00000258543
A0A140T9M8	0,12844553	QYISYRCQEKRNTYFAEYWYQ	79,8375	ENSG00000241713
A0A140T9C1	0,12844553	QYISYRCQEKRNTYFAEYWYQ	79,8375	ENSG00000241132
A0A140T8X3	0,12844553	QYISYRCQEKRNTYFAEYWYQ	79,8375	ENSG00000239285
A0A140T990	0,12844553	QYISYRCQEKRNTYFAEYWYQ	79,8375	ENSG00000239497
P35499	0,05836608	VMILTVFCLSVFALVGLQLFM	79,7738	ENSG00000007314
J3QQZ1	0,05836608	VMILTVFCLSVFALVGLQLFM	79,7738	ENSG00000007314
H0YBY8	0,05675816	QNSEFYGFSEFYCTEDVLRM	79,6305	ENSG00000197217
Q8N660	0,05879675	QHYSRVFYSFEEHISFALYV	79,5668	ENSG00000266338
Q9UHI6	0,08907492	TYQDYEEYWRAYYRAWQEYYA	79,5349	ENSG00000064703
E9PJ60	0,08907492	TYQDYEEYWRAYYRAWQEYYA	79,5349	ENSG00000064703
Q8N699	0,07278373	FTVSMALGLVLGGFIWAVFIC	79,4871	ENSG00000120279
H0YDV5	0,07278373	FTVSMALGLVLGGFIWAVFIC	79,4871	ENSG00000120279
Q8N9R8	0,05212206	YHYLRTSETSFLNEAFSFYS	79,4553	ENSG00000173611
Q3SXZ0	0,05212206	YHYLRTSETSFLNEAFSFYS	79,4553	ENSG00000173611
H3BTA9	0,06513198	MLFQLMVEHDHETFWLFQFFL	79,4075	ENSG00000167139
Q8N7C4	0,11650313	RTVMHCFWMMFFVINYAHITYK	79,3916	ENSG00000172738
P27487	0,0553414	QENNILVFNAEYGNSSVFLEN	79,3916	ENSG00000197635
F8WE17	0,0553414	QENNILVFNAEYGNSSVFLEN	79,3916	ENSG00000197635
B2RTY4	0,05122702	DIFGFEDYENNSFEQFCINF	79,3119	ENSG00000066933
H3BMM1	0,05122702	DIFGFEDYENNSFEQFCINF	79,3119	ENSG00000066933
H3BRD5	0,05122702	DIFGFEDYENNSFEQFCINF	79,3119	ENSG00000066933
Q5QGZ9	0,06335282	YCGYINRLYVQYYHCTYKKRM	79,0253	ENSG00000172322
Q96MH7	0,05231866	GSVFKSEGAYFGNYFTYYSIQ	78,9297	ENSG00000172244
Q6ZMG9	0,07885875	CESMWRFSFYLYVFTYGVVFL	78,8819	ENSG00000172292
P12259	0,08243021	CYTTEFYVAYSSNQINWQIFK	78,8342	ENSG00000198734
A0A0AOMRJ7	0,08243021	CYTTEFYVAYSSNQINWQIFK	78,8342	ENSG00000198734
Q9BXY4	0,08910904	MHLRLISWLFIIILNFMEYIGS	78,7227	ENSG00000146374
O43823	0,06285794	YENNYGQAQNTSVTTGATYS	78,6908	ENSG00000105127
O00219	0,05941801	YFREWLNSLWFHKKHLLWMTY	78,3086	ENSG00000103044
Q8J025	0,08605179	ITRSYRFYHNNTFKAYQFYYG	78,2927	ENSG00000154856
J3KTQ6	0,08605179	ITRSYRFYHNNTFKAYQFYYG	78,2927	ENSG00000154856
O75907	0,10169507	LIWLIFFYWLNFHSCLNVAEAL	78,0856	ENSG00000185000
Q16832	0,05704869	HCNNMFAKGVKIFKEVQCYFR	78,0538	ENSG00000162733
Q9Y2K6	0,11023015	CQNVINGQWYEFDDQYVTEVH	78,0219	ENSG00000136878
Q9NPL8	0,07422086	RTAVFVTIFNTVNTSLNVYRN	77,9901	ENSG00000113845
G3XA94	0,07422086	RTAVFVTIFNTVNTSLNVYRN	77,9901	ENSG00000113845
P30988	0,07031394	YVMHSLIHFQGGFFVATIYCFC	77,9741	ENSG00000004948
A0A0A0MSQ7	0,07031394	YVMHSLIHFQGGFFVATIYCFC	77,9741	ENSG00000004948
A0A0A0MRG0	0,07031394	YVMHSLIHFQGGFFVATIYCFC	77,9741	ENSG00000004948
A0A0C4DG16	0,07031394	YVMHSLIHFQGGFFVATIYCFC	77,9741	ENSG00000004948
Q9UH17	0,05251703	ARVTIMDYEEFAYCWENFVYN	77,9741	ENSG00000179750
B0QYD3	0,05251703	ARVTIMDYEEFAYCWENFVYN	77,9741	ENSG00000179750
Q9NP73	0,09341816	VLQYYFNLGLQCYHSHYWHSM	77,8308	ENSG00000101901
Q9NZP6	0,05469753	TERKFYTSSTHYGQETVYRR	77,783	ENSG00000185823
Q6ZS30	0,05927247	SSFFEDFQYECNSNEWQVYIE	77,7034	ENSG00000144426
Q68BL8	0,08969852	VVYNGAFYFYNRAFRNIIKYD	77,5123	ENSG00000162745
F2Z3N3	0,08969852	VVYNGAFYFYNRAFRNIIKYD	77,5123	ENSG00000162745
Q5JY77	0,06392352	EEVNQEAEEETIFGSWFWVID	77,5123	ENSG00000198932
Q5VYJ5	0,0822926	CTFRFYHMFQKRIYRLAIYQ	77,4964	ENSG00000204740
U5GXS0	0,0822926	CTFRFYHMFQKRIYRLAIYQ	77,4964	ENSG00000204740
H0Y8G5	0,17667036	QNWNOGYSNYWNOGYGNYGYN	77,4327	ENSG00000138668
Q14103	0,16396614	QNWNOGYSNYWNOGYGNYGYN	77,4327	ENSG00000138668
Q2NKX8	0,06241224	RNGVIITTYQMLINWQQLSS	77,4167	ENSG00000186871
B5MDQ0	0,06241224	RNGVIITTYQMLINWQQLSS	77,4167	ENSG00000186871
Q5TCS8	0,10725238	QHQNWYVIDGFHSHKWWVWNEV	77,4008	ENSG00000155085
H7C517	0,10725238	QHQNWYVIDGFHSHKWWVWNEV	77,4008	ENSG00000155085
Q8N3X1	0,06385026	MGDWQEVNDENTGCYYWNTQ	77,3371	ENSG00000109920
Q9Y6F1	0,05656975	TLNQTNIENNNTKFYIIQLLQ	77,3371	ENSG00000041880
C9J9C7	0,05475247	TLNQTNIENNNTKFYIIQLLQ	77,3371	ENSG00000041880
A0A024R2X5	0,05475247	TLNQTNIENNNTKFYIIQLLQ	77,3371	ENSG00000041880
Q9UBI9	0,05394886	MHLQCFYEWESSILVQFNCIG	77,2415	ENSG00000112406
Q8N987	0,08028421	MIYEFWENSSVWNSHLQTNYS	77,1301	ENSG00000123119
Q12809	0,06760619	GSLMYASIFGNVSAIIQRLYS	77,0823	ENSG00000055118
Q86U57	0,06760619	GSLMYASIFGNVSAIIQRLYS	77,0823	ENSG00000055118

Q9UKU9	0,07026607	NGVWYRGGHYRSRYQDGVYWA	77,0345	ENSG00000136859
Q99784	0,0560049	GQVVYNGSIYFNKFSHIIR	76,8593	ENSG00000130558
O60469	0,05253284	GYQIGYREYSTGGNFQFNIIS	76,7319	ENSG00000171587
Q8WY19	0,05253284	GYQIGYREYSTGGNFQFNIIS	76,7319	ENSG00000171587
A0A087WUI7	0,05253284	GYQIGYREYSTGGNFQFNIIS	76,7319	ENSG00000171587
Q9BXT5	0,08585683	SAWCVYQYSNSNGNAITQTYQ	76,6364	ENSG00000133863
Q68BL7	0,09373137	YVTNYYYGNSLVEFRNLENFK	76,6204	ENSG00000185585
Q9C0C4	0,05184922	FFGVFQAQWQDMYLSAICEYQ	76,6045	ENSG00000168758
O95677	0,08962113	TAFGQNQYAYQYYSASTYGAYM	76,3656	ENSG00000112319
F2Z2Y1	0,08962113	TAFGQNQYAYQYYSASTYGAYM	76,3656	ENSG00000112319
E7ESD5	0,08962113	TAFGQNQYAYQYYSASTYGAYM	76,3656	ENSG00000112319
E9PLN6	0,08962113	TAFGQNQYAYQYYSASTYGAYM	76,3656	ENSG00000112319
Q6ZMW3	0,05828773	CRNNLYTAGKEVVYFVAGVG	76,2382	ENSG00000214595
K7END1	0,06084251	NGINLALAWSQOEAIWQVLYL	76,2223	ENSG00000108784
O15372	0,05733634	ALLDSQFSYQHAIEESVLLIY	76,1426	ENSG00000147677
B3KS98	0,05733634	ALLDSQFSYQHAIEESVLLIY	76,1426	ENSG00000147677
A0A087WZK9	0,05733634	ALLDSQFSYQHAIEESVLLIY	76,1426	ENSG00000147677
E5RGU4	0,05733634	ALLDSQFSYQHAIEESVLLIY	76,1426	ENSG00000147677
Q6IE37	0,06210967	IAQVQTNLDIFTLTLCSSFLTIV	76,0471	#N/A
P14735	0,09942826	GWVFYQQRNEVHNNCGIETIYY	76,0152	ENSG00000119912
D6RBZ0	0,12263152	GQSQSWNQGYGNYWVQYGYQ	75,9993	ENSG00000197451
Q99729	0,11929034	GQSQSWNQGYGNYWVQYGYQ	75,9993	ENSG00000197451
E9PS35	0,05230367	AIQQYGSSETGVVITFKNYL	75,9197	ENSG00000166938
Q99590	0,08203334	GESSFTYRAYCTEFIEASEIS	75,8719	ENSG00000139218
A0A0A0MTP7	0,08203334	GESSFTYRAYCTEFIEASEIS	75,8719	ENSG00000139218
P27540	0,05041252	KGQVLSVMFRFRSKNQEWLWM	75,856	ENSG00000143437
B8ZZ71	0,10982166	CWFNGIVEENDSNIWKFWYTN	75,8401	ENSG00000170417
C9IYX5	0,10982166	CWFNGIVEENDSNIWKFWYTN	75,8401	ENSG00000170417
Q969J5	0,07793298	NSSVYFVQYKIMFSCSMKSSH	75,8401	ENSG00000164485
Q6ZP80	0,07502435	CWFNGIVEENDSNIWKFWYTN	75,8401	ENSG00000170417
Q12768	0,08698796	FHRSFYIQQYVNIYGLKIWQ	75,8082	ENSG00000164961
E7EQI7	0,08698796	FHRSFYIQQYVNIYGLKIWQ	75,8082	ENSG00000164961
Q6R2W3	0,05627706	IFSWMQTNNSSHWTFLWFIQ	75,8082	ENSG00000232040;ENSG00000248496
A0A140T9Y6	0,05627706	IFSWMQTNNSSHWTFLWFIQ	75,8082	ENSG00000248496
Q6ZW05	0,06020776	HHFIQHFRLREHYNEWITNIYV	75,7604	ENSG00000244694
Q5GH73	0,14903539	SVWIWQSVIHLQMQGVWRYI	75,6967	ENSG00000171044
B1AJW0	0,05669377	QTKLYLAMNSEGYLYTSELFT	75,6489	ENSG00000129682
Q92913	0,055885	QTKLYLAMNSEGYLYTSELFT	75,6489	ENSG00000129682
H0YCE8	0,05602707	VQEDYDQAFQYQQATQFASS	75,633	ENSG00000198730
Q6PD62	0,05176503	VQEDYDQAFQYQQATQFASS	75,633	ENSG00000198730
Q86UK0	0,06020608	IYNLTGQRVENYLISTANEFV	75,3463	ENSG00000144452
Q9BZC7	0,06005375	GRFYFLYGFVWIQDMMERAI	75,3304	ENSG00000107331
E9PGB2	0,06005375	GRFYFLYGFVWIQDMMERAI	75,3304	ENSG00000107331
J3QSS3	0,06005375	GRFYFLYGFVWIQDMMERAI	75,3304	ENSG00000107331
A0A087WVK5	0,06005375	GRFYFLYGFVWIQDMMERAI	75,3304	ENSG00000107331
H0YD08	0,05357163	SSGHYIAYCRNNLNLWYEFD	75,3145	ENSG00000077254
A0A0A6YYK7	0,05363945	TSSFNFITITASQVVDASVYFC	75,2986	ENSG00000211799
P04436	0,05240038	TSSFNFITITASQVVDASVYFC	75,2986	#N/A
E5RJT0	0,06936255	QSTYYGSFVTRALLDSQFSYQ	75,2508	ENSG00000147677
Q6UXN8	0,08785334	WIQNRESCYYVSEIWSIWHTS	75,2189	ENSG00000197992
Q96L03	0,17039236	RKQYQLTVQVAYYTMNNLYN	75,1552	ENSG00000162814
Q86VZ5	0,08494212	NQQVLKEASQMNLLARVWYR	75,1552	#N/A
D3DWC4	0,08494212	NQQVLKEASQMNLLARVWYR	75,1552	ENSG00000198964
Q92802	0,09794422	SNIFQAQDDSQIQNGYVNNC	75,0915	ENSG00000244754
D6R968	0,09794422	SNIFQAQDDSQIQNGYVNNC	75,0915	ENSG00000244754
A0A0B4J276	0,05748387	GEDFTTYCNSSTLSNIQWYK	75,0756	ENSG00000211806
Q9H159	0,06574854	ISAWYNLSITATEKYNIEQIS	74,996	ENSG00000071991
J3KTP3	0,06574854	ISAWYNLSITATEKYNIEQIS	74,996	ENSG00000071991
Q8IZT6	0,07521763	IRMIIAVTSYKRYLWATVTIQ	74,9163	ENSG00000066279
P0C881	0,06174665	SIYYNQEGTCWYEGDWVQNIK	74,9004	ENSG00000155026
B2RC85	0,06174665	SIYYNQEGTCWYEGDWVQNIK	74,9004	ENSG00000169402
Q05BV3	0,06132981	LFYQTQIGEIVYHVAAVGVIYN	74,8367	ENSG00000165521
Q99502	0,10337948	MQGSSFTTSSGIYTGNNSLTN	74,6934	ENSG00000104313
A6NCB9	0,10337948	MQGSSFTTSSGIYTGNNSLTN	74,6934	ENSG00000104313

E7EQM5	0,10337948	MQGSSFTTSSGIYTGNNSLTN	74,6934	ENSG00000104313
F8WB53	0,10337948	MQGSSFTTSSGIYTGNNSLTN	74,6934	ENSG00000104313
P11230	0,0818329	GHQEIHIHEGTFIENGQWEII	74,6297	ENSG00000170175
I3L1T7	0,0818329	GHQEIHIHEGTFIENGQWEII	74,6297	ENSG00000170175
Q9UKJ0	0,05310199	YVNRLFNLNWTGQESGFLRIS	74,5978	ENSG00000121716
C9J8P3	0,05310199	YVNRLFNLNWTGQESGFLRIS	74,5978	ENSG00000121716
V9GYC2	0,05310199	YVNRLFNLNWTGQESGFLRIS	74,5978	ENSG00000121716
C9JNA4	0,05310199	YVNRLFNLNWTGQESGFLRIS	74,5978	ENSG00000121716
F78364	0,07126115	QQQQIHLQKQVVIQQQIAIH	74,5819	ENSG00000111752
J3KQH6	0,07126115	QQQQIHLQKQVVIQQQIAIH	74,5819	ENSG00000111752
Q9H2Y7	0,05412621	NHSNSGGWLSNSGAVDWNHN	74,4863	ENSG00000103994
H3BSS6	0,05412621	NHSNSGGWLSNSGAVDWNHN	74,4863	ENSG00000103994
Q8N7L0	0,06546473	LLKALNQQQRYFYFSIMRIYN	74,4226	ENSG00000179813
Q8TDW7	0,05700029	ASIVTVIQLVNNVVDTIENEV	74,3908	ENSG00000165323;ENSG0000282908
E9PQ73	0,05700029	ASIVTVIQLVNNVVDTIENEV	74,3908	ENSG00000165323;ENSG0000282908
E9PJL8	0,06251799	RTFAVYLNNTGYRTAFFGKYL	74,3749	ENSG00000137573
Q8IWU6	0,05373492	RTFAVYLNNTGYRTAFFGKYL	74,3749	ENSG00000137573
Q9NRD1	0,07067485	VRYLFLQHGGRDTQYWAGWYG	74,3271	ENSG00000116663
J3KQ72	0,07067485	VRYLFLQHGGRDTQYWAGWYG	74,3271	ENSG00000116663
H0YNJ6	0,0883015	GMSSEMAMKKYAGGVAEYRYV	74,2634	ENSG00000100938
E5RI03	0,05168671	MRLRFCISSQEYNINNAESFS	74,1997	ENSG00000169398
P04745	0,08106374	VAFGRGNRGFIVFNDDWTFS	74,1519	ENSG00000174876;ENSG0000187733;ENSG00000237763
P19961	0,08106374	VAFGRGNRGFIVFNDDWTFS	74,1519	ENSG00000240038
Q9H9S0	0,13721043	QTWNNSTWSNQTONIQSWSNH	74,0882	ENSG00000111704
Q6NSW7	0,13638297	QTWNNSTWSNQTONIQSWSNH	74,0882	#N/A
J7H3Z5	0,13638297	QTWNNSTWSNQTONIQSWSNH	74,0882	ENSG00000255192
Q8N7R0	0,13590225	QTWNNSTWSNQTONIQSWSNH	74,0882	#N/A
A0A0D9SG05	0,13486792	QTWNNSTWSNQTONIQSWSNH	74,0882	ENSG00000255192
Q16478	0,0780959	IARVLNSRYAFLLESTMNEYH	74,0882	ENSG00000105737
Q15436	0,05028237	IRVTTIARNWADAQTQIQNIA	74,0086	ENSG00000100934
F5H365	0,05028237	IRVTTIARNWADAQTQIQNIA	74,0086	ENSG00000100934
Q96DX8	0,11143341	AQVQILCHTYWEHWTSSQGQVR	73,9448	ENSG00000136514
Q96NE9	0,05192421	RAARYYYYWHLRQVLHSQCV	73,9448	ENSG00000139926
G3V3V8	0,05192421	RAARYYYYWHLRQVLHSQCV	73,9448	ENSG00000139926
Q96Q05	0,07622392	SEENFRIYKRICSVSISVR	73,8811	ENSG00000167632
P31321	0,05640552	GEWVTNISEGGSFGELALIYG	73,8493	ENSG00000188191
C9JSK5	0,05640552	GEWVTNISEGGSFGELALIYG	73,8493	ENSG00000188191
C9JR00	0,05640552	GEWVTNISEGGSFGELALIYG	73,8493	ENSG00000188191
P13727	0,15379692	LVSIHNFNINRYIQCSVSALN	73,7537	ENSG00000186652
Q9H9V9	0,06168903	TISINHNWVNGFNLANMWRF	73,7378	ENSG00000081692
A0A087WT84	0,06168903	TISINHNWVNGFNLANMWRF	73,7378	ENSG00000081692
F8VU51	0,10657491	AAAHWQQHQHRVGFQYQGIM	73,5785	ENSG00000119596
O94929	0,06945916	SSNVIQCYRCGDTCKGEVVRV	73,2441	ENSG00000173210
D6RHE7	0,06945916	SSNVIQCYRCGDTCKGEVVRV	73,2441	ENSG00000173210
A0A0C4DGA7	0,06945916	SSNVIQCYRCGDTCKGEVVRV	73,2441	ENSG00000173210
Q13117	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205944
Q86SG3	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205916
Q9NQZ3	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000188120
E7EU39	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205944;ENSG0000205916
E9PBY2	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205944
E7EU38	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205916
E7ENA5	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205916
A0A0A0MSR9	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000188120
A0A0A0MSS9	0,07849245	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205916
Q9NR90	0,07847705	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000187191
E7ERQ6	0,07847705	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000187191
E7ETR3	0,07847705	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000187191
I3LOB6	0,07847705	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205944
E7ENA6	0,07847705	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205944
A0A140T8Y1	0,07847705	RRNLWTEAYKWWYLVCLIQRR	73,1804	ENSG00000205916

Q92839	0,07410492	QQTRWSKSYFREWLYNALWWH	73,1645	ENSG00000105509
G3VLS7	0,07410492	QQTRWSKSYFREWLYNALWWH	73,1645	ENSG00000105509
M0R2V0	0,07410492	QQTRWSKSYFREWLYNALWWH	73,1645	ENSG00000105509
Q9H3P7	0,08617874	QKQOIMAAALNSQTAVQFQOYA	73,0689	ENSG00000182827
P28827	0,06937638	YNVTRCHSYNLTVHYCYQVGG	73,0689	ENSG00000173482
E7EPS8	0,06937638	YNVTRCHSYNLTVHYCYQVGG	73,0689	ENSG00000173482
P38435	0,06425701	HFLTQGYNNWTNGLYGYSWDM	73,053	ENSG00000115486
P31943	0,07854857	SMSGYDQVLQENSSDFQSNIA	72,9574	ENSG00000169045
H0YBK1	0,07854857	SMSGYDQVLQENSSDFQSNIA	72,9574	ENSG00000169045
E9PCY7	0,07292214	SMSGYDQVLQENSSDFQSNIA	72,9574	ENSG00000169045
H0YBG7	0,07292214	SMSGYDQVLQENSSDFQSNIA	72,9574	ENSG00000169045
P07510	0,0519755	FDWQNCSLIFQSQTYSTNEID	72,9415	ENSG00000196811
C9JGE3	0,07807744	YGQOIYFVGSGRTKNGFETRA	72,8778	ENSG00000182944
Q86Y38	0,05531318	RQLQRMFKAIYHKDHFYIYHV	72,7822	ENSG00000103489
Q9NZM6	0,05693646	DLSNFGLQINTEWRYSTSNTN	72,7504	ENSG00000078795
D6RF71	0,05693646	DLSNFGLQINTEWRYSTSNTN	72,7504	ENSG00000078795
Q86X55	0,09376256	VQYFQFYGYLSQQQNMMDYV	72,7185	ENSG00000142453
K7EQA8	0,09376256	VQYFQFYGYLSQQQNMMDYV	72,7185	ENSG00000142453
Q12805	0,05409457	ECDASNQCAQQCYNILGSFIC	72,7026	ENSG00000115380
J3QR85	0,07518691	RNGQYVACCYFSDLQSYRNR	72,5752	ENSG00000011260
Q8N831	0,06318525	IVKVYEVRSFGQVVSFSTLIM	72,5115	ENSG00000178021
P23471	0,06519049	GYVMLMDYLQNFREQYKFS	72,4478	ENSG00000106278
Q8NB12	0,05293006	FSMQYISHIFGVINCNGFTLS	72,4478	ENSG00000115593
E9PHG3	0,05293006	FSMQYISHIFGVINCNGFTLS	72,4478	ENSG00000115593
Q9H0B3	0,05081533	SMQAAEEIRILAVITIQAGVR	72,4	ENSG00000130518
A0A087WXN0	0,05081533	SMQAAEEIRILAVITIQAGVR	72,4	ENSG00000130518
Q9UK61	0,10196795	EGENSNSTEQDSYSNFQVYHS	72,2089	ENSG00000163946
A0A087X0F1	0,10196795	EGENSNSTEQDSYSNFQVYHS	72,2089	ENSG00000163946
Q96E52	0,08559848	NHVFFRFNSLSNWRKCNTLAS	72,1611	ENSG00000162600
S4R3A3	0,08559848	NHVFFRFNSLSNWRKCNTLAS	72,1611	ENSG00000162600
X6RDQ1	0,08559848	NHVFFRFNSLSNWRKCNTLAS	72,1611	ENSG00000162600
X6RIG5	0,08559848	NHVFFRFNSLSNWRKCNTLAS	72,1611	ENSG00000162600
X6RL62	0,08559848	NHVFFRFNSLSNWRKCNTLAS	72,1611	ENSG00000162600
X6RD49	0,08559848	NHVFFRFNSLSNWRKCNTLAS	72,1611	ENSG00000162600
Q9Y6X6	0,05021586	QLCVNMTNEKMHYINEVLFL	72,0496	ENSG00000041515;ENSG00000282848
F8W883	0,05021586	QLCVNMTNEKMHYINEVLFL	72,0496	ENSG00000041515;ENSG00000282848
Q9NZR2	0,07833707	HQQISHIEHNSRITGMVYYQ	72,0337	ENSG00000168702
H0Y7T7	0,07833707	HQQISHIEHNSRITGMVYYQ	72,0337	ENSG00000168702
P35555	0,05446294	GYLQHYQWNQCVDENECLSAH	72,0178	ENSG00000166147
H0YEX9	0,06493815	VERRWMTNYLRLWQLGVEKIY	71,9541	ENSG00000154721
P37088	0,05638715	VNNKRNGVAKVNIFFKELNYK	71,9382	ENSG00000111319
F5GXE6	0,05638715	VNNKRNGVAKVNIFFKELNYK	71,9382	ENSG00000111319
J3QQR9	0,05529633	KILTRNQIETVLSTRIQVMIS	71,8267	ENSG00000213424
M0R010	0,06052653	AGYEGYGYGYGQDNTTNYG	71,6356	ENSG00000011243
P17844	0,07038455	TQNGVYSAANYTNGSFGSNFV	71,6196	ENSG00000108654
J3KTA4	0,07038455	TQNGVYSAANYTNGSFGSNFV	71,6196	ENSG00000108654
A0A075B7F2	0,1271674	TDSSYGQNYSGYSSYGQSYSQ	71,5878	ENSG00000270647;ENSG00000276833
P35557	0,0747984	AMVNDTVATMISCYEDHQCE	71,5878	ENSG00000106633
C9JQD1	0,0747984	AMVNDTVATMISCYEDHQCE	71,5878	ENSG00000106633
P10153	0,105875	QHINMTSQCTNAMQVINNYQ	71,54	ENSG00000169385
P12724	0,07330481	TIAMRAINNYRWRCKNQNTFL	71,54	ENSG00000169397
Q14435	0,05442499	SISKEYFEYIGSYDEEMEIWG	71,5082	ENSG00000115339
C9J2C3	0,05442499	SISKEYFEYIGSYDEEMEIWG	71,5082	ENSG00000115339
E7EUL0	0,05442499	SISKEYFEYIGSYDEEMEIWG	71,5082	ENSG00000115339
O60941	0,09797935	CQQCHNYQLCQNCFWRGHAGG	71,3967	ENSG00000138101
E9PEY4	0,09797935	CQQCHNYQLCQNCFWRGHAGG	71,3967	ENSG00000138101
E7EVB6	0,09797935	CQQCHNYQLCQNCFWRGHAGG	71,3967	ENSG00000138101
R4GN71	0,05882989	IQSWFRGCQVRAIRHLNRIV	71,3648	ENSG00000162814
Q8TE57	0,11287522	GARSIRIYEMNVSTSYISVRN	71,333	ENSG00000145536
Q09MP3	0,07179926	QQVVNVENWAHYNSSSTVKAHG	71,2533	ENSG00000214842
H0YEG2	0,05339526	SATYGEHWFETNVSGDFCYV	71,2374	ENSG00000149091
Q15678	0,05657175	FHGNEEALYCNSHNSLDLNYL	71,2215	ENSG00000152104

O60674	0,06695583	HVFHIDESTRHNVLRYRIRFYF	71,1896	ENSG00000096968
A0A1BOGTR9	0,06695583	HVFHIDESTRHNVLRYRIRFYF	71,1896	ENSG00000096968
Q9ULJ6	0,05014214	EQFNGQNNTFSGSSYSNYSQG	71,1419	ENSG00000108175
F6WR09	0,05014214	EQFNGQNNTFSGSSYSNYSQG	71,1419	ENSG00000108175
P29323	0,06351217	MNTIRTYQVCNVFESSQNNWL	71,11	ENSG00000133216
Q6NVW1	0,06351217	MNTIRTYQVCNVFESSQNNWL	71,11	ENSG00000133216
B1AKC9	0,06351217	MNTIRTYQVCNVFESSQNNWL	71,11	ENSG00000133216
Q6ZR05	0,06218373	GSSAWVAVCKQVCTRVGTYA	71,0304	#N/A
Q12906	0,0917831	YGSYGYGNSATAGYSQFYSN	71,0144	ENSG00000129351
V9GYX2	0,06612757	RVEYQCQSYELQGSNYVTCS	71,0144	ENSG00000134365
Q92496	0,06551282	RVEYQCQSYELQGSNYVTCS	71,0144	ENSG00000134365
A0A0C4DH21	0,06551282	RVEYQCQSYELQGSNYVTCS	71,0144	ENSG00000134365
H7BZF3	0,06000693	KARYLYNLMFQTKWTVRQQ	70,9985	ENSG00000198089
Q8WYK1	0,07709834	YEQSCVYRHOQNTAGFFYID	70,9507	ENSG00000155052
H7BXM1	0,06487566	QDLGMFLVTISCYTRGGRIIS	70,8711	ENSG00000168000
Q8N6M8	0,0852858	ARMWRIRRRYQCVLNAVRIIQ	70,8233	ENSG00000173389
Q8IZL2	0,0902506	QQQQQSSISAQQQQQQSSIS	70,7915	ENSG00000184384
A0A087X0G5	0,0902506	QQQQQSSISAQQQQQQSSIS	70,7915	ENSG00000184384
Q9HCJ0	0,06707012	MTMLNQLYQLQLAYQLRQIQQ	70,7915	ENSG00000078687
A0A1BOGU24	0,06707012	MTMLNQLYQLQLAYQLRQIQQ	70,7915	ENSG00000078687
Q5T4I8	0,0556674	RYGNWYARQHGSYLLSGYSYG	70,7596	ENSG00000137434
Q9H4W6	0,05422166	SSQLAVNVSETSQANDQVGYG	70,7278	ENSG00000108001
Q07507	0,07795041	VAGFQSYRHFESVLDREWQFYC	70,6959	ENSG00000143196
Q15437	0,06101058	QFVTHYQHSSTQRRIRVTTIA	70,6959	ENSG00000101310
Q86TS7	0,1051753	MVQECCSQSLYEEELHSYHIV	70,68	#N/A
Q9H0C5	0,07948308	SDRIRFTVNRRISIVGFLYLG	70,6641	ENSG00000064726
A0A0U1RQI7	0,09235625	GNQALYGGQMMTSTGNQTLYW	70,6163	ENSG00000283039
C9JCQ3	0,10682809	QQQQQFQAQQSAMQQQFQAVV	70,5844	ENSG00000099917
Q96RN5	0,09282619	QQQQQFQAQQSAMQQQFQAVV	70,5844	ENSG00000099917
G3V1P5	0,09282619	QQQQQFQAQQSAMQQQFQAVV	70,5844	ENSG00000099917
H7C308	0,09282619	QQQQQFQAQQSAMQQQFQAVV	70,5844	ENSG00000099917
Q9P267	0,05474238	MSSINNTLSNHQLTHLQSLLN	70,5526	ENSG00000204406
E9PHH0	0,05474238	MSSINNTLSNHQLTHLQSLLN	70,5526	ENSG00000204406
A0A0D9SF16	0,05474238	MSSINNTLSNHQLTHLQSLLN	70,5526	ENSG00000204406
A0A1BOGW10	0,05474238	MSSINNTLSNHQLTHLQSLLN	70,5526	ENSG00000204406
A0A0D9SG23	0,05474238	MSSINNTLSNHQLTHLQSLLN	70,5526	ENSG00000204406
P05813	0,06088029	GWFNNEVGSQKIQSGAWVCYQ	70,473	ENSG00000108255
P0C7M6	0,10372894	LVQRRIRQRQALLRVYVIQE	70,4411	ENSG00000229972
F8VUB4	0,07996259	FGMNRNQAFGMNNSLSSNIFN	70,4252	ENSG00000111596
F8VQD8	0,07996259	FGMNRNQAFGMNNSLSSNIFN	70,4252	ENSG00000111596
Q15596	0,12863835	HFGQQANTSMYSNNMNINVSM	70,3933	ENSG00000140396
H0YBB6	0,12863835	HFGQQANTSMYSNNMNINVSM	70,3933	ENSG00000140396
Q8NFP9	0,0969324	EIRCYVNGQLVSYGDMAWHVN	70,3455	ENSG00000172915
Q5T321	0,0969324	EIRCYVNGQLVSYGDMAWHVN	70,3455	ENSG00000172915
F5GXV7	0,0969324	EIRCYVNGQLVSYGDMAWHVN	70,3455	ENSG00000172915
A0A0D9SF28	0,0969324	EIRCYVNGQLVSYGDMAWHVN	70,3455	ENSG00000172915
Q9HCT0	0,07789426	VHVGVVV1KAVSSGFYVAMNR	70,3137	ENSG00000070388
Q92968	0,10647873	AYSSFSSGYGAYGNSFYGGYS	70,2818	ENSG00000162928
Q9HAK2	0,10014065	GSQLGVSISESTQGNQGYIR	70,2659	ENSG00000221818
B7Z934	0,10014065	GSQLGVSISESTQGNQGYIR	70,2659	ENSG00000221818
H0Y7Z8	0,08254524	STGVLGRGLDLISVRLVNFDF	70,2341	ENSG00000118197
K7ELV3	0,08316219	GAGYGSYGYGNSATAGYNYS	70,1863	ENSG00000129351
Q01085	0,10977826	MTKNFQQVDYSQWGQWSQVYG	70,1544	ENSG00000151923
Q9NUM4	0,1158373	SIKVHNIVLMMQVTVTTTTYFG	70,1226	ENSG00000106460
Q969I3	0,05209692	RFVGFQGFEEASCEWHQWTCY	70,0748	ENSG00000166840
P31942	0,06838781	GLGGYGRGGGGSGGYYQGGGM	70,0589	ENSG00000096746
Q6P3S6	0,05580397	TQRFSSHACYYDANQSMYVFG	70,0589	ENSG00000037637
P55317	0,08241898	MEGHETSQDWSYADTQEQAYS	69,9155	ENSG00000129514
C9JLN7	0,06650375	QQQQQQQFQAQQSAMQQQFQ	69,9155	ENSG00000099917
Q3L8U1	0,0870187	MTSCSVNSQGFSSHSYFSSN	69,8678	ENSG00000177200
H3BSP3	0,0870187	MTSCSVNSQGFSSHSYFSSN	69,8678	ENSG00000177200
P40426	0,0533757	RHVINQTTGGYSDGLGGNSLYS	69,8359	ENSG00000167081
Q5JS98	0,0533757	RHVINQTTGGYSDGLGGNSLYS	69,8359	ENSG00000167081
Q92804	0,12636454	ASQSYSGYQTTDSSYQGNYS	69,82	ENSG00000270647;ENSG0000276833

Q495T6	0,05182705	QYGNYSWDLADEQNVNGFNTL	69,82	ENSG00000142606;ENSG0000277131
P55795	0,06941147	SMSGYDQVLQENSSDYQSNLA	69,8041	ENSG00000126945
B7ZKW0	0,05699281	HVVISLNGFLQGYNDLSQEEM	69,7563	ENSG00000101892
Q01432	0,05010183	MYANIMVLNNLRRERGLSTFL	69,6448	ENSG00000133805
E9PKC5	0,05010183	MYANIMVLNNLRRERGLSTFL	69,6448	ENSG00000133805
A0A087WXN3	0,06733346	SDRIRFSVNKRIFVVGFGLYG	69,5492	ENSG00000133243
Q9BX70	0,06502122	SDRIRFSVNKRIFVVGFGLYG	69,5492	ENSG00000133243
Q9BYJ4	0,05452387	VDVSKKTAWILGVYCRYSRH	69,5333	ENSG00000258659
B2RNG4	0,05452387	VDVSKKTAWILGVYCRYSRH	69,5333	ENSG00000258588
P04746	0,07041337	YDNGSNQVAFGRGNRGFIVFN	69,4378	ENSG00000243480
Q9Y485	0,06907937	RFNYQGNKFGIVDADGYLSLY	69,39	ENSG00000172869
F5H269	0,06907937	RFNYQGNKFGIVDADGYLSLY	69,39	ENSG00000172869
J3QKV3	0,07461496	RVQNEGSWNSYVDYKIFLHME	69,3741	ENSG00000002919
Q7Z353	0,06989489	LRSVFTVEQQRILQRYYENG	69,3741	ENSG00000165259
Q9NSY1	0,07259401	QHATQQQQMLQQQFLMHSVYQ	69,3581	ENSG00000138756
H0Y9P1	0,07259401	QHATQQQQMLQQQFLMHSVYQ	69,3581	ENSG00000138756
Q2KHR3	0,07810524	SQVLSVSVLSSESYASGESLTL	69,3263	ENSG00000060749
Q9NUQ7	0,07140235	DDNGWGCAYSRLQITICSWFKH	69,3103	ENSG00000109775
D6RGX2	0,07140235	DDNGWGCAYSRLQITICSWFKH	69,3103	ENSG00000109775
Q9BY15	0,06976103	YSVYCGFNAVCYNVEGSFYCQ	69,2148	ENSG00000131355
P29074	0,05938454	SYGCVQVCHSEEGNTAYIFR	68,96	ENSG00000088179
J3KQD3	0,05938454	SYGCVQVCHSEEGNTAYIFR	68,96	ENSG00000088179
Q9NQV6	0,07371252	GQALQQQQQQQQNSVQHTYL	68,8963	ENSG00000170325
E9PLV1	0,07371252	GQALQQQQQQQQNSVQHTYL	68,8963	ENSG00000170325
E9PRS0	0,07371252	GQALQQQQQQQQNSVQHTYL	68,8963	ENSG00000170325
Q76N89	0,0509086	SSSCYSTSCYSSSCYSASCYS	68,7689	ENSG00000002746
Q8NCA5	0,06147178	QGGQFEQHFGHGGYQYNHSGF	68,7529	ENSG00000119812
Q8WYB5	0,14643458	AYNVNSVMNMTLNAMNGYS	68,6255	ENSG00000156650;ENSG0000281813
Q6ZMW2	0,06456858	TFQASVSFDVTVFEFSQEEWQ	68,5937	ENSG00000196597
C9J9Y8	0,06456858	TFQASVSFDVTVFEFSQEEWQ	68,5937	ENSG00000196597
P05787	0,06223608	YGGASGMGGITAVTVNQSLLS	68,3866	ENSG00000170421
F8VUG2	0,06223608	YGGASGMGGITAVTVNQSLLS	68,3866	ENSG00000170421
F8VQY3	0,05831083	YGGASGMGGITAVTVNQSLLS	68,3866	ENSG00000170421
F8W1U3	0,05831083	YGGASGMGGITAVTVNQSLLS	68,3866	ENSG00000170421
Q92794	0,12711816	AYNVNSMNMNTLNMNSYRMT	68,3548	ENSG00000083168
Q9Y4J8	0,08418553	QCHNYQLCQDCFWRGHAGGSH	68,2114	ENSG00000134769
Q13946	0,05266863	EIEVSVSARNIRLLSFQRYL	68,2114	ENSG00000205268
M0QYH6	0,05018897	KDLETLKLSCRIMDNGFNGFV	68,1796	ENSG00000104960
O94916	0,09937369	MSLQSGNFRLQQSSHSQAQLFH	68,1637	ENSG00000102908
P40189	0,05250454	NESSQNTSSTVQYSTVHVHSGY	68,1159	ENSG00000134352
Q09327	0,05199978	MLQAVYGLDGI RLRRRQYYTM	68,1	ENSG00000128268
A0A0U1RQR8	0,06695533	QQLQALLQQQAVMLQQLLQQ	67,9885	ENSG00000128573
A0A087WZF3	0,05417407	QSQGYNQWQQGSVHVNVLCGR	67,877	ENSG00000153187
P31483	0,13060778	AYGMYGQAWNQGFNQTSASA	67,8133	ENSG00000116001
F8W8I6	0,12931171	AYGMYGQAWNQGFNQTSASA	67,8133	ENSG00000116001
Q9UMD9	0,10255704	GQEIQQYISEYMQSDSIRSYL	67,8133	ENSG00000065618
C9JWK5	0,06584692	SAQLQLQQVALQQQQQQQFQ	67,7974	ENSG00000099917
Q96PZ7	0,07132365	SDHSQNRQGFKLAYQAYELQN	67,7337	#N/A
F8W9C3	0,07132365	SDHSQNRQGFKLAYQAYELQN	67,7337	ENSG00000183117
E5RIG2	0,07132365	SDHSQNRQGFKLAYQAYELQN	67,7337	ENSG00000183117
F5GZ18	0,07132365	SDHSQNRQGFKLAYQAYELQN	67,7337	ENSG00000183117
A0A0U1RQY1	0,07132365	SDHSQNRQGFKLAYQAYELQN	67,7337	ENSG00000183117
H7BXU2	0,07132365	SDHSQNRQGFKLAYQAYELQN	67,7337	ENSG00000183117
O15405	0,11558891	QLQQHQMHHQQIQQQMQQQHFQ	67,5107	ENSG00000103460
Q9Y2C2	0,05414448	RFLSNYFFRRFGDWRGEQNHM	67,5107	ENSG00000111962
Q86YT9	0,07782091	AKEEIVFRYHKLKLRMSVEYSQ	67,4948	ENSG00000160593
E9PKK2	0,07782091	AKEEIVFRYHKLKLRMSVEYSQ	67,4948	ENSG00000160593
C9JJ12	0,05601152	QTQLQLQQVALQQQQQQQFQ	67,4948	ENSG00000099917
A0A087X1A2	0,1757121	YYGGLGYGCGGFGGLGYGYSC	67,3355	ENSG00000244362
P41221	0,05024537	GSRETAFTYAVSAAGVVNAMS	67,0807	ENSG00000114251
C9J8I8	0,05024537	GSRETAFTYAVSAAGVVNAMS	67,0807	ENSG00000114251
O00755	0,05094925	NTHQYARVWQCNCCKFWCCYV	67,0488	ENSG00000154764
Q9UGU0	0,08762422	QYEGHNVGSNAQAYGTQSNYS	67,0329	ENSG00000100207;ENSG00

				000262024;ENSG00000276461;ENSG00000280467;ENSG00000281897;ENSG00000282892;ENSG00000283026;ENSG00000283681
O15523	0,11198352	GGGGYGNRGRFGGGGGYGGFYN	66,9533	ENSG00000067048
O75570	0,05585811	RLQVFRQNRNCILHLLSKNWS	66,9214	ENSG00000120662
X6RFD4	0,05585811	RLQVFRQNRNCILHLLSKNWS	66,9214	ENSG00000120662
A0A087X1S1	0,05585811	RLQVFRQNRNCILHLLSKNWS	66,9214	ENSG00000120662
Q8N6Z2	0,05585811	RLQVFRQNRNCILHLLSKNWS	66,9214	ENSG00000120662
Q8WXH2	0,05730998	LQDGYGTETYSDDGGTYQGQWV	66,8737	ENSG00000154118
X6RCC3	0,05349061	HRNLGVHISRKSVNLDQWTQ	66,8418	ENSG00000084070
Q15714	0,05430174	ERESTSGSSVSSSVSTLSHYT	66,81	ENSG00000102804
Q9NZC4	0,05021764	TDSYSTCNVSSGFFGGQWHEI	66,7781	ENSG00000135373
E9PQX0	0,05021764	TDSYSTCNVSSGFFGGQWHEI	66,7781	ENSG00000135373
E9PN75	0,05021764	TDSYSTCNVSSGFFGGQWHEI	66,7781	ENSG00000135373
E9PQR6	0,05021764	TDSYSTCNVSSGFFGGQWHEI	66,7781	ENSG00000135373
E9PPS9	0,05021764	TDSYSTCNVSSGFFGGQWHEI	66,7781	ENSG00000135373
Q00839	0,05540415	NRGYKNQSQGYNQWQQGQFWG	66,6507	ENSG00000153187
P57052	0,05216021	SFESCVKINSHNYRNEEMLVG	66,6507	ENSG00000185272
Q93074	0,07707357	QQQQQQQQYHIRQQQQQQIL	66,587	ENSG00000184634
Q7Z3Z5	0,07707357	QQQQQQQQYHIRQQQQQQIL	66,587	ENSG00000184634
Q13342	0,05152808	QQQMASGSDSNLFRMVAETQ	66,587	ENSG00000079263
P07196	0,06787228	VHISSVRSYGSTARSAYSSYS	66,5711	ENSG00000277586
A0A087X0W2	0,06787228	VHISSVRSYGSTARSAYSSYS	66,5711	ENSG00000277586
O75038	0,05905937	QMVALNYQSEGRMLQLNRAKF	66,5711	ENSG00000149527;ENSG0000276429
D6RF34	0,05124648	CRLNEYNQQLQAAHAQEQR	66,5711	ENSG00000126545
E9PDQ1	0,05124648	CRLNEYNQQLQAAHAQEQR	66,5711	ENSG00000126545
Q9HBZ2	0,0675228	TGQNMSQISRQLNQSQVAWTG	66,5392	ENSG00000172379
A0A087WVE9	0,0675228	TGQNMSQISRQLNQSQVAWTG	66,5392	ENSG00000172379
Q9H4Z2	0,06116659	TVQHLVTDNDQVQYIISQDGV	66,4596	ENSG00000198026
P46100	0,06900578	RLQQQYNQQQQQMTYQQATL	66,4437	ENSG00000085224
Q9BUJ2	0,1338356	SYNQYQQYQAQQWNOYYQNCQQ	66,4118	ENSG00000105323
B7Z4B8	0,1338356	SYNQYQQYQAQQWNOYYQNCQQ	66,4118	ENSG00000105323
A0A0A0MRA5	0,1338356	SYNQYQQYQAQQWNOYYQNCQQ	66,4118	ENSG00000105323
P35637	0,10053937	SGYSQSTDTSGYGQSSYSSYG	66,4118	ENSG00000089280
H3BNZ4	0,10053937	SGYSQSTDTSGYGQSSYSSYG	66,4118	ENSG00000089280
H3BPE7	0,10053937	SGYSQSTDTSGYGQSSYSSYG	66,4118	ENSG00000089280
Q9P2D1	0,05800311	SNLNQGLVNTGMNQNLGLTN	66,4118	ENSG00000171316
P11844	0,0576794	LRVYFSRCNSIRVDSGCWMLY	66,0774	ENSG00000168582
Q7Z794	0,09137154	ISVQNSQVSVNGGAGGGGSYG	65,9977	ENSG00000189182
P04264	0,09085132	RSGYRSGGGFSSGSAGIINYQ	65,727	ENSG00000167768
H7C557	0,0613007	WCVGHERRTVYMGYRQVYTT	65,6633	ENSG00000162591
Q5T6S3	0,05092728	MLQCYRCRQWFHEACTQCLNE	65,6473	ENSG00000119403
A0A087X169	0,05092728	MLQCYRCRQWFHEACTQCLNE	65,6473	ENSG00000119403
G3V5F8	0,08867231	QCQVCTCVVHKRCHHLIVTAC	65,5677	ENSG00000027075
G3V5U5	0,08867231	QCQVCTCVVHKRCHHLIVTAC	65,5677	ENSG00000027075
G3V304	0,08867231	QCQVCTCVVHKRCHHLIVTAC	65,5677	ENSG00000027075
G3V4Q9	0,08867231	QCQVCTCVVHKRCHHLIVTAC	65,5677	ENSG00000027075
G3V4Q6	0,08867231	QCQVCTCVVHKRCHHLIVTAC	65,5677	ENSG00000027075
G3V4X4	0,08867231	QCQVCTCVVHKRCHHLIVTAC	65,5677	ENSG00000027075
G3V3E7	0,08867231	QCQVCTCVVHKRCHHLIVTAC	65,5677	ENSG00000027075
P24723	0,08063336	QCQVCTCVVHKRCHHLIVTAC	65,5677	ENSG00000027075
O15014	0,05613261	DQSYHTHLLSTNTAYRQQYEE	65,5518	ENSG00000180357
O00571	0,09440444	FYNSDGYGGNYNSQGVDDWGN	65,4722	ENSG00000215301
A0A0C4DH59	0,05406683	GQQVTLRCSQSGHNTVSWYQ	65,3925	ENSG00000230099
Q9Y2E5	0,09375745	NREAVLRTSTNLNSQQVIYSD	65,3288	ENSG00000013288
E9PCD7	0,09375745	NREAVLRTSTNLNSQQVIYSD	65,3288	ENSG00000013288
H0YA68	0,09375745	NREAVLRTSTNLNSQQVIYSD	65,3288	ENSG00000013288
Q9HBW9	0,05834635	GTVCIENTVNANCHLDNVCIAA	65,297	ENSG00000162618
P51991	0,05665309	DGYNEGGNFGGGNYGGGGNYN	65,281	ENSG00000170144
Q92841	0,05329358	AYGTSSYTAQEYAGATYGASS	65,1377	ENSG00000100201
H3BLZ8	0,05329358	AYGTSSYTAQEYAGATYGASS	65,1377	ENSG00000100201
P12035	0,10457784	YGVSGGGFSSASNRGGSIKFS	65,0262	ENSG00000186442

Table S2. Selected human PrLD amyloid cores predicted aggregation propensity. AGGRESCAN ⁴, TANGO ⁵ and Zyggregator ⁶ predictions for the human PrLD amyloid cores candidates. Aggregation propensities values are shown for each peptide according to the corresponding predictors using the default setting and values above the aggregation threshold shown in bold.

<i>PROTEIN</i>	<i>PrLD AMYLOID CORES</i>	<i>AGGRESCAN</i>	<i>TANGO</i>	<i>ZYGGREGATOR</i>
DDX5	TQNGVYSAANYTNGSFSGSNFV	-10.6	1.28	-3.66
EYA1	MQGSSFTTSSGIYTGNNSLTN	-12.3	0	-3.65
ILF3	YGSYGYGGNSATAGYSQFYSN	-3.1	0	-3.72
MED15	QQQQQFQAQQSAMQQQFQAVV	-44.6	0	-4.63
NCOA2	HFGQQANTSMYSNNMNINVSM	-20.9	0	-1.17
PHC1	QQQQIHLQKQVVIQQQIAIH	-17.5	57.17	-2.10
TIA1	AYGMYGQAWNQQGFNQTQSSA	-30.9	0	-4.09

Table S3. Selected human PrLD amyloid cores hydrophaticity. For each candidate, the grand average of hydrophathy (GRAVY) value was evaluated using the EXPASY ProtParam tool ⁷. Positive values corresponding to hydrophobic sequences are shown in bold. α -synuclein (ASYN) and A β 42 short amyloidogenic stretches were predicted with AmylPred2 ⁸.

<i>PROTEIN</i>	<i>PrLD AMYLOID CORES</i>	<i>GRAVY SCORE</i>
DDX5	TQNGVYSAANYTNGSFGSNFV	-0.357
EYA1	MQGSSFTTSSGIYTGNNSLTN	-0.49
ILF3	YGSYGYGGNSATAGYSQFYSN	-0.786
MED15	QQQQQFQAQQSAMQQQFQAVV	-1.024
NCOA2	HFGQQANTSMYSNNMNINVSM	-0.643
PHC1	QQQQIHLQQKQVVIQQQIAIH	-0.633
TIA1	AYGMYGQAWNQQGFNQTQSSA	-1.019
ASYN	GVLYVG	1.683
	GGAVVTGVTAVAQ	1.238
Aβ42	GAIIGLMVGGVVI	2.462
	QKLVFFAE	0.562

Table S4. Disorder context of PrLD amyloid cores. FoldIndex⁹, IUPRED¹⁰, PONDR-FIT¹¹ and RONN¹² algorithms were used for disorder prediction. Disorder was analyzed for the 21 residues-long peptides and 20 flanking residues at each end and expressed as the percentage of disordered residues in these 61 residues-long segments. Average disorder accounts for the mean of all disorder predictions for a given segment.

DISORDER PREDICTOR	DDX5	EYA1	ILF3	MED15	NCOA2	PHC1	TIA1
FOLDINDEX	91	61	74	100	25	62	100
IUPRED	37	75	64	31	100	69	38
PONDR-FIT	51	100	100	100	100	100	56
RONN	41	100	100	100	100	85	51
AVERAGE	55	84	85	83	82	79	61

Table S5. Selected human PrLD amyloid cores secondary structure assignment. Assignment and area of the secondary structure components of aggregated PrLD amyloid cores in the amide I region of the FTIR spectra.

ASSIGNMENTS	DDX5	EYA1	ILF3	MED15	NCOA2	PHC1	TIA1
Inter β-sheet (1623-1641 cm^{-1})	37%	67.5%	46%	54%	52%	66.5%	62%
Disordered/Loops/Turns (1658-1665 cm^{-1})	63%	32.5%	34%	35%	24%	33.5%	31.5%
β-sheet (1674-1695 cm^{-1})	-	-	20%	11%	24%	-	6.5%

Table S6. Protein-protein interactions (PPI) statistics for human PrLD-containing proteins. The PPI enrichment p-value was obtained from the STRING database ¹³ and reflects the relationship between the number of interactions established by the different proteins and those randomly expected.

<i>PROTEIN</i>	<i>PPI enrichment p-value</i>
DDX5	7.76e-06
EYA1	1.12e-10
ILF3	0.00294
MED15	1.87e-10
NCOA2	2.03e-12
PHC1	0
TIA1	0.45

References:

- 1 Toombs, J. a. *et al.* De novo design of synthetic prion domains. *Proceedings of the National Academy of Sciences* **109**, 6519-6524, (2012).
- 2 Sabate, R., Rousseau, F., Schymkowitz, J. & Ventura, S. What Makes a Protein Sequence a Prion? *PLoS Computational Biology* **11**, e1004013, (2015).
- 3 UniProt Consortium, T. U. UniProt: a hub for protein information. *Nucleic acids research* **43**, D204-212, (2015).
- 4 Conchillo-Sole, O. *et al.* AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC bioinformatics* **8**, 65, (2007).
- 5 Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature biotechnology* **22**, 1302-1306, (2004).
- 6 Tartaglia, G. G. & Vendruscolo, M. The Zyggregator method for predicting protein aggregation propensities. *Chemical Society reviews* **37**, 1395-1401, (2008).
- 7 Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* **40**, W597-603, (2012).
- 8 Tsolis, A. C., Papandreou, N. C., Iconomidou, V. A. & Hamodrakas, S. J. A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS one* **8**, e54175, (2013).
- 9 Prilusky, J. *et al.* FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435-3438, (2005).
- 10 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* **21**, 3433-3434, (2005).
- 11 Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochimica et biophysica acta* **1804**, 996-1010, (2010).
- 12 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* **21**, 3369-3376, (2005).
- 13 Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447-452, (2015).

10.2. ANNEX. Supplementary material. PUBLICATION VI

hnRNPD L phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation.

Battle C., Yang P., Coughlin M., Messing J., Pesarro dona M., Szulc E., Salvatella X., Kim HJ., Taylor JP. and Ventura S.

Cell Rep. (2020).

DOI: 10.1016/j.celrep.2019.12.080

Cell Reports, Volume 30

Supplemental Information

hnRNPD L Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation

Cristina Batlle, Peiguo Yang, Maura Coughlin, James Messing, Mireia Pesarrodonà, Elzbieta Szulc, Xavier Salvatella, Hong Joo Kim, J. Paul Taylor, and Salvador Ventura

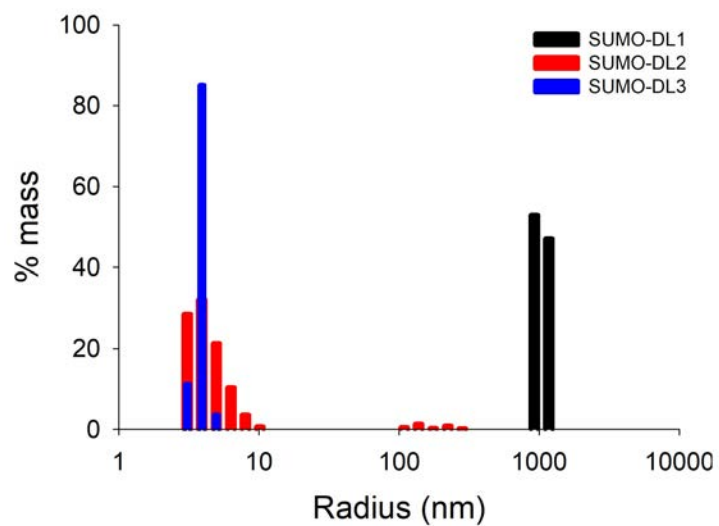


Figure S1. Dynamic Light Scattering of hnRNPD L isoforms. Related to Figure 1.

Dynamic Light Scattering (DLS) radius (nm) versus % of mass of hnRNPD L isoforms 1, 2 and 3 (SUMO-DL1, DL2 and DL3 fusions) at 50 μ M in 50 mM HEPES pH 7.5 and 150 mM NaCl, in the absence of crowding agents.

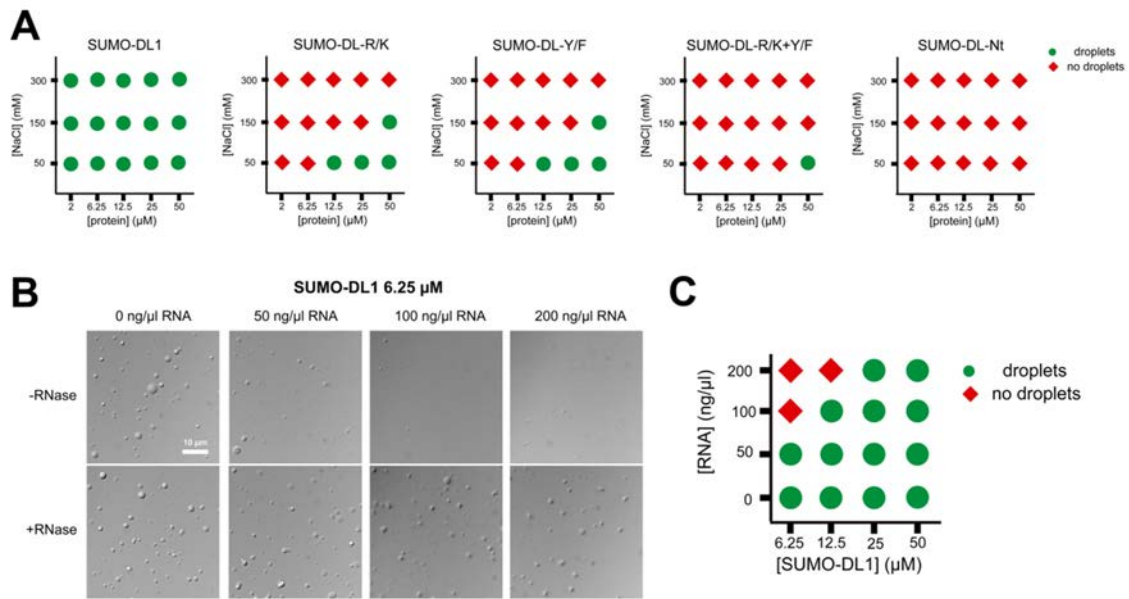


Figure S2. Liquid-liquid phase separation diagram of hnRNPDL isoform 1 after mutation or RNA addition. Related to Figure 2 and Figure 3.

A) LLPS diagram of hnRNPDL isoform 1 (SUMO-DL1) and the four hnRNPDL variants: SUMO-DL-R/K, SUMO-DL-Y/F, SUMO-DL-R/K+Y/F and SUMO-DL-Nt, in the absence of crowding agents.

Green circles indicate positive and red diamonds indicate negative for the appearance of droplets at the indicated NaCl/protein concentration combinations.

B) SUMO-hnRNPDL isoform 1 (DL1) LLPS at 6.25 μM (366 ng/μl) in the presence of different concentrations of total RNA with or without 5 ng/μl RNase in 50 mM HEPES pH 7.5 and 150 mM NaCl.

C) LLPS diagram of DL1 as a function of protein and RNA concentration in 50 mM HEPES pH 7.5 and 150 mM NaCl. Green circles indicate positive and red diamonds indicate negative for the appearance of droplets at the indicated RNA/protein concentration combinations.

Green circles indicate positive and red diamonds indicate negative for the appearance of droplets at the indicated RNA/protein concentration combinations.

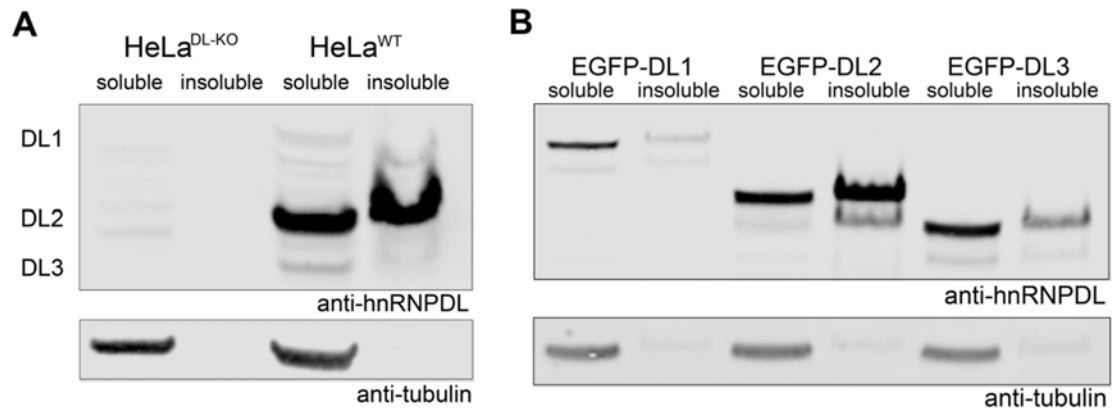


Figure S3. Endogenous or individual isoforms hnRNPDL solubility analysis after expression in HeLa WT or hnRNPDL KO cells. Related to Figure 3 and Figure 5.

A) Cell extracts of HeLa hnRNPDL KO (HeLa^{DL-KO}) and HeLa WT were processed for soluble examination by Western Blot using an antibody against hnRNPDL protein. Tubulin was blotted as a loading control. B) Cell extracts of HeLa^{DL-KO} after EGFP-tagged hnRNPDL isoform 1, 2 or 3 (EGFP-DL1, DL2 and DL3) expression were fractionated and the soluble and insoluble fractions analyzed by Western Blot using an antibody against hnRNPDL protein. Tubulin was blotted as a loading control.

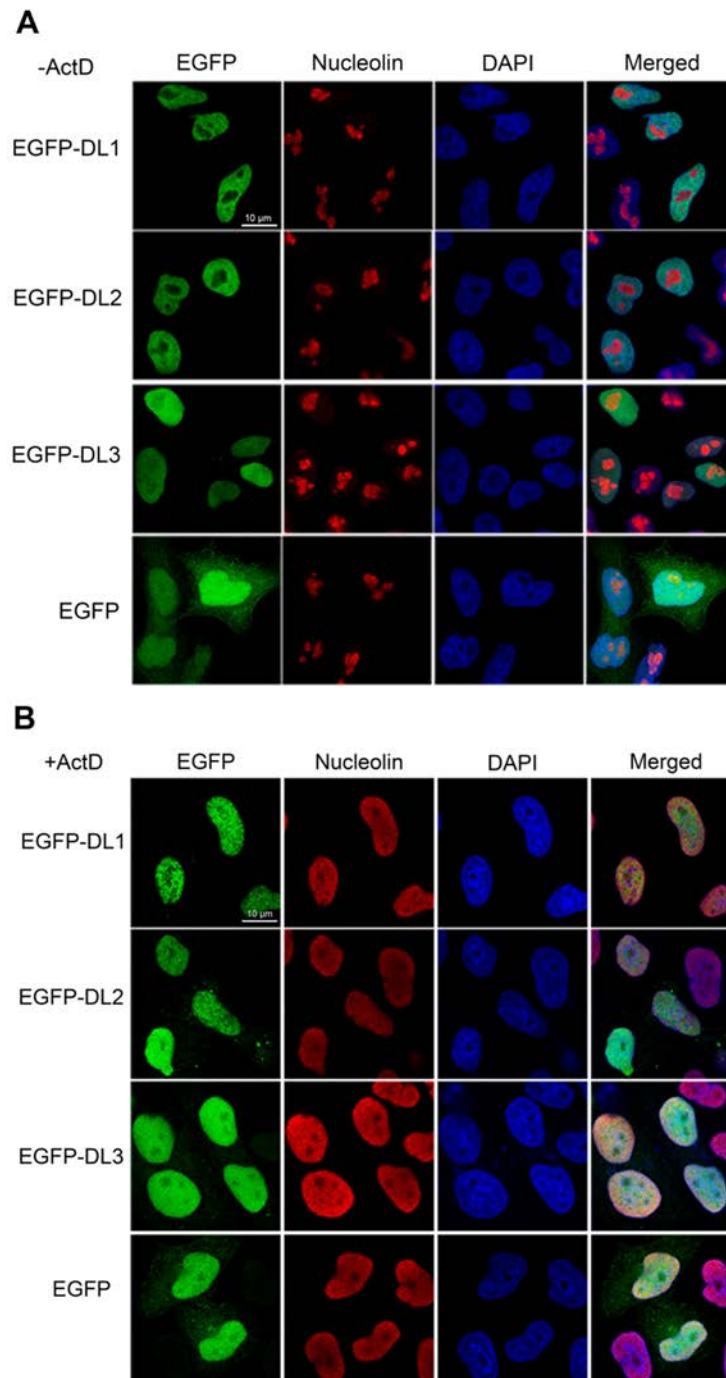


Figure S4. HnRNPD^L isoforms localization in HeLa^{DL-KO} cells. Related to Figure 3 and Figure 4. Cellular localization by immunofluorescence of EGFP-hnRNPD^L isoform 1, 2 and 3 (DL1, DL2, DL3) and unfused EGFP after expression in HeLa^{DL-KO} cells in the absence (A) or the presence (B) of actinomycin D. Cells were stained with nucleolin antibody (red) as nucleolus marker and DAPI (blue) as nuclear marker.

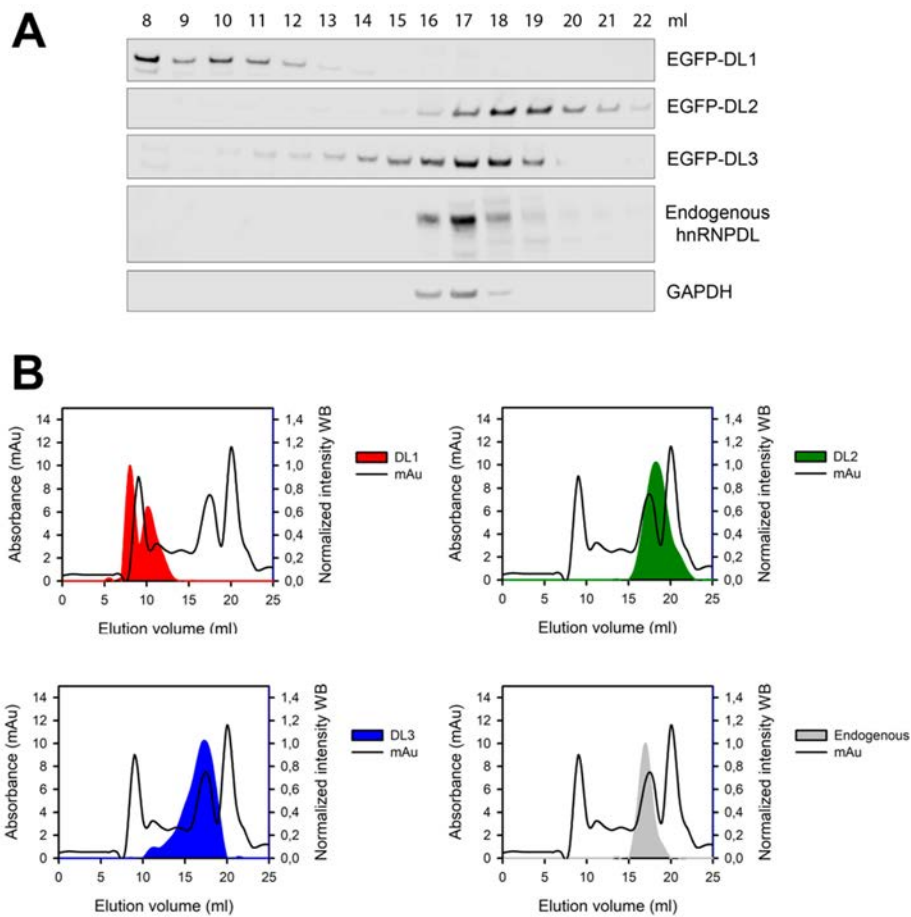


Figure S5. Elution pattern of cellular hnRNPDL isoforms. Related to Figure 3.

A) Cell extracts of HeLa^{DL-KO} cells after EGFP-tagged hnRNPDL isoform 1, 2 or 3 (EGFP-DL1, DL2 and DL3) expression as well as HeLa WT cells were fractionated by size exclusion chromatography (SEC) and fractions at different elution volumes (ml) were analyzed by Western Blot (WB) using an antibody against hnRNPDL protein. GAPDH was blotted as a loading and molecular weight control. B) WB intensities of EGFP-DL1, DL2 and DL3 and endogenous hnRNPDL were plotted on top of a representative SEC graph of HeLa cells extracts.

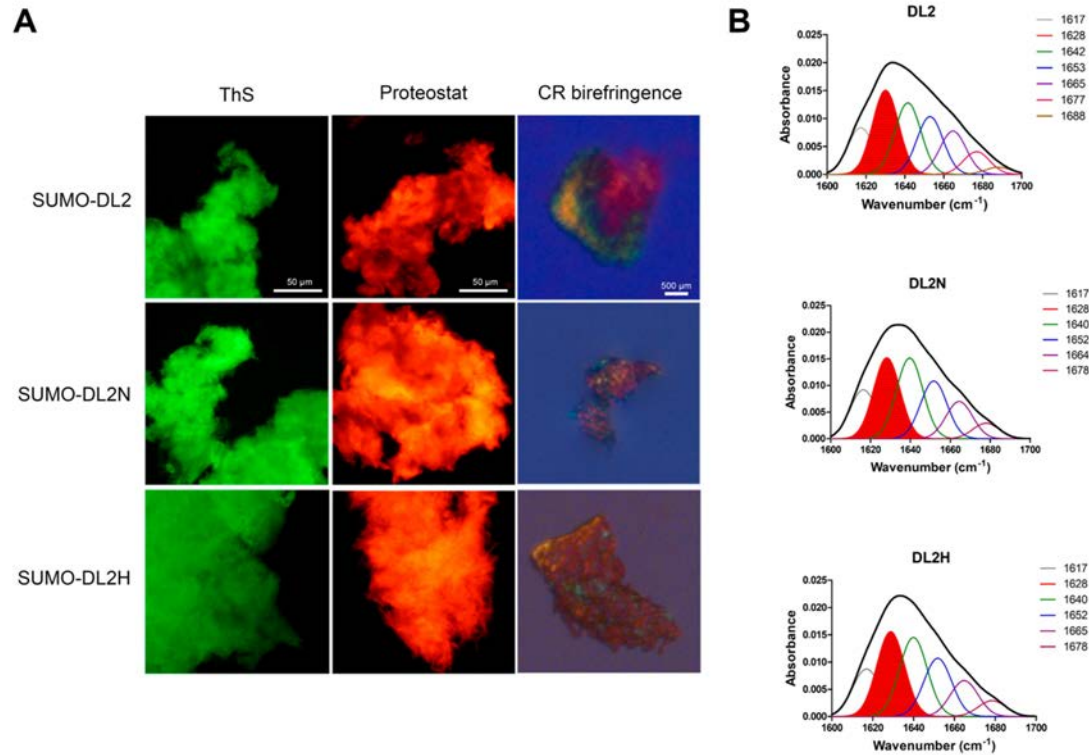


Figure S6. Amyloid properties of hnRNPD L isoform 2 and disease-causing mutations. Related to Figure 5 and 6.

A) Thioflavin-S (ThS) staining, Proteostat® staining and Congo Red (CR) birefringence of 4 days incubated 50 μ M SUMO-hnRNPD L isoform 2 (DL2) and the disease-causing mutations D259N and D259H (DL2N and DL2H) in 50 mM HEPES pH 7.5 and 150 mM NaCl. B) DL2, DL2N and DL2H FTIR absorbance spectrum in the amide I region after aggregation at 50 μ M in 50 mM HEPES pH 7.5 and 150 mM NaCl. The black line corresponds to the original absorbance spectrum and the red dotted area indicates the contribution of the intermolecular β -sheet signal to the total area upon Gaussian deconvolution. Aggregation was conducted at 37°C and 600 rpm in both A and B.

Human	YGNYN SAYGG-DQNY S-GYGGY D YTG YNYGNYGYGQGYADYSG 41
Mouse	YGNYN SAYGG-DQNY S-GYGGY D YTG YNYGNYGYGQGYAD--- 38
Rat	YGNYN SAYGD-ES--YSGYGGY D YTG YNYGSYGYGQGYTD--- 37
Chicken	YGNYN SAYSD-QS--YSGYGGY D YSGYNYPNYGYGPGYTD--- 37
Xenopus	YGSYGNNGSYADQGYNNSYSGY D YSGYNYGSYGYNQGYTD--- 40
Zebrafish	GQNYGGGYGNGYNQGYNGYSGY D YSGYNYQNYGYGQGYDD--- 40
	.*. *.*****;**** .****. ** *

Figure S7. HnRNPD L exon 6 alignment in vertebrates. Related to Figure 6.

hnRNPD L exon 6 alignment between human, mouse, rat, chicken, xenopus and zebrafish organisms using Clustal Omega. Asp disease-causing mutation is in red.

Rosetta energy (kcal/mol)	DL2	DL2N	DL2H
Average six hexapeptides	-19.4	-20.7	-21.2
Maximum scored hexapeptide GGYDYT	-20.7	-22.4	-22.2

Table S1. ZipperDB analysis of hnRNPD L isoform 2 and disease-causing mutations. Related to Figure 6.

ZipperDB analysis of hnRNPD L isoform 2 and the disease-causing mutations D259N and D259H (position 378 in hnRNPD L isoform 1) (DL2N and DL2H). The average of the Rosetta energy for the 6 possible hexapeptides containing the Asp mutated residue and the hexapeptide with the highest score are presented in the table.

SUMO-DL2		SUMO-DL2N		SUMO-DL2H	
Peak	% area	Peak	% area	Peak	% area
1617	13.99	1617	15.24	1617	14.78
1628	25.23	1628	25.23	1628	26.39
1642	21.45	1640	25.12	1640	24.59
1653	17.36	1652	17.98	1652	18.13
1665	13.00	1664	11.60	1665	11.19
1677	6.82	1678	4.83	1678	4.93
1688	2.15				

Table S2. Secondary structure content of hnRNPD L isoform 2 and disease-causing mutations aggregates. Related to Figure 6 and Figure S6.

Position and relative area of spectral components in the amide I region of the FTIR absorbance spectrum for the aggregated hnRNPD L isoform 2 and the disease-causing mutations D259N and D259H (DL2N and DL2H).

Bacteria primers	Primers 5' → 3'
SUMO-DL1_F	CGCGAACAGATTGGAGGTGAAGTCCC GCCCGTCTG
SUMO-DL1_R	GTGGCGGCCGCTCTATTAGTACGGTTGATAATTGTT
SUMO-DL2_F	GAAGACATGAACGAATACAGC
SUMO-DL2_R	ACCTCCAATCTGTTTCGCGGTG
SUMO-DL3_F	CAAAGCACGTACGGTAAAGCAAG
SUMO-DL3_R	CTGACCACGGCCGCGACCACG
SUMO-DL2N_F	AACTACACCGGCTATAACTAC
SUMO-DL2N_R	ATAACCGCCGTAACCGCTATAG
SUMO-DL2H_F	CACTACACCGGCTATAACTAC
SUMO-DL2H_R	ATAACCGCCGTAACCGCTATAG
SUMO-DL1R/K_F	GAAGACATGAACGAATACAGC
SUMO-DL1R/K_R	ACCTCCAATCTGTTTCGCGGTG
SUMO-DL1-Y/F_F	CAAAGCACGTACGGTAAAGC
SUMO-DL1-Y/F_R	CTGACCACGGCCGCGACCACGGGT
SUMO-DL-Nt_F	TAATAGAGCGGCCGCCACCGCT
SUMO-DL-Nt_R	CATCGTGACGCTCGAATCTG
Mammalian primers	Primers 5' → 3'
EGFP-C3-DL1_F	GTA CT CAGATCTCGAGCTCAAGCTTATGGAGGTCCC GCCCAGGCTTTC
EGFP-C3-DL1_R	CAGTTATCTAGATCCGGTGGATCCTTAGTATGGCTGGTAATTGTTT
EGFP-C3-DL2_F	GAGGATATGAACGAGTACAGC
EGFP-C3-DL2_R	AAGCTTGAGCTCGAGATCTGAG
EGFP-C3-DL3_F	CAGAGCACTTATGGCAAGGCATC
EGFP-C3-DL3_R	CTGACCTCGGCCACGACCCCTC
EGFP-C3-DL2N_F	AATTATACTGGGTATAACTATG
EGFP-C3-DL2N_R	ATATCCGCCATAGCCACTATAG
EGFP-C3-DL2H_F	CATTATACTGGGTATAACTATG
EGFP-C3-DL2H_R	ATATCCGCCATAGCCACTATAG
DNA fragments	Primers 5' → 3'
DNA fragment R/K	CACCGCGAACAGATTGGAGGTGAAGTCCC GCCGAAACTGAGTCATGTCCC GC CGCCGCTGTTCCC GAGCGCACCCGGCAACCCTGGCAAGCAAGAGCCTGTCGCA CTGGAAGCCGAAACCGCCGAAACAGCTGGCACCGCTGCTGCCGTCCCTGGCC CCGAGCTCTGCAAAGCAGGGCGCTAAGAAAGCGCAAAAGCATGTTACCGCAC AGCAACCGAGTAAACTGGCAGGCGGTGCGGCCATTAAAGGCGGTAAGAAGAA GAAACCGGACCTGTTTAAGAAACATTTCAAAGTTCTCAATCCAGAAGAGC GCAGCTGCGGCCGAGCTACCAAGACGGCTAAACAGCACCCGCCGCGCAGATT CGAGCGTACGATGGAAGACATGAACGAATACAGC
DNA fragment Y/F	ACCCGTGGTTCGCGGCCGTGGTCAGGGCCAAA ACTGGAACCAGGGTTTCAACA ACTTCTTCGATCAAGGTTTTCGGCAACTTCAATTCGGCGTTTTGGCGGTGATCA GAAC TTTAGCGGTTTTCGGCGGTTTTGACTTCACCGGCTTTAACTTCGGTAAT TTTGGTTTTCGGCCAGGGTTTTGCCGATTTCTCGGGCCAGCAAAGCACGTACG GTAAAGC

Table S3. List of the primers used in this study. Related to STAR Methods.

The source of all the primers is from this study and there is no identifier.

10.3. ANNEX. Supplementary material. MED15

The Q-rich prion-like domain of human MED15 forms a coiled-coil responsible for its conversion to amyloids and its propagation.

Batlle C., Iglesias V., Calvo I., Lynch C., Serrano M. and Ventura S.

Future submission.

SUPPLEMENTARY INFORMATION:

The Q-rich prion-like domain of human MED15 forms a coiled-coil responsible for its conversion to amyloids and its propagation

Cristina Batlle¹, Valentin Iglesias¹, Isabel Calvo², Cian Lynch², Manuel Serrano² and Salvador Ventura^{1*}

¹Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain

²Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldri Reixac 10, 08028 Barcelona, Spain

*Correspondence: Salvador.ventura@uab.es

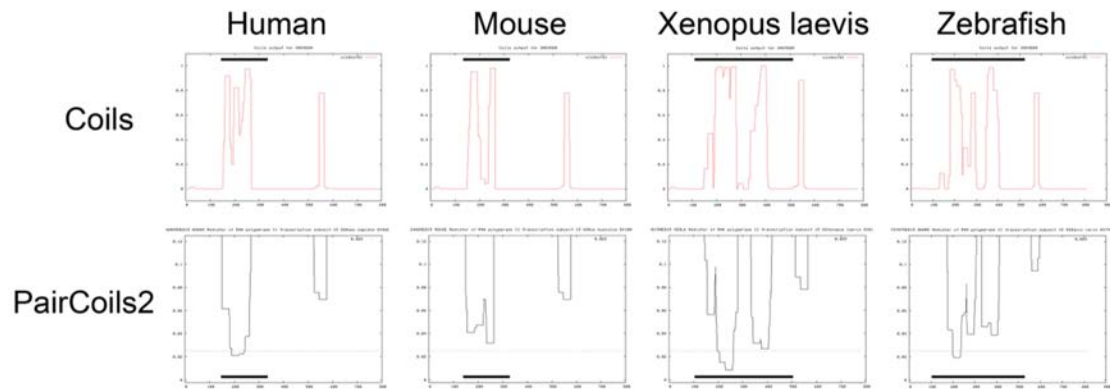


Figure S1. MED15 orthologs coiled-coil prediction.

Coils (Lupas Andrei et al., 1991) and PairCoils2 (McDonnell et al., 2006) predictions (21-residue window) of MED15 orthologs: human, mouse, xenopus laevis and zebrafish. Black rectangle indicates approximate prion-like domain position predicted by PLAAC (Lancaster et al., 2014).

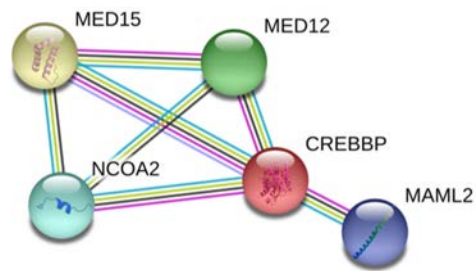


Figure S2. MED15 interaction network.

STRING (Szklarczyk et al., 2017) prediction of MED15, MED12, NCOA2, CREBBP and MAML2 interactions.

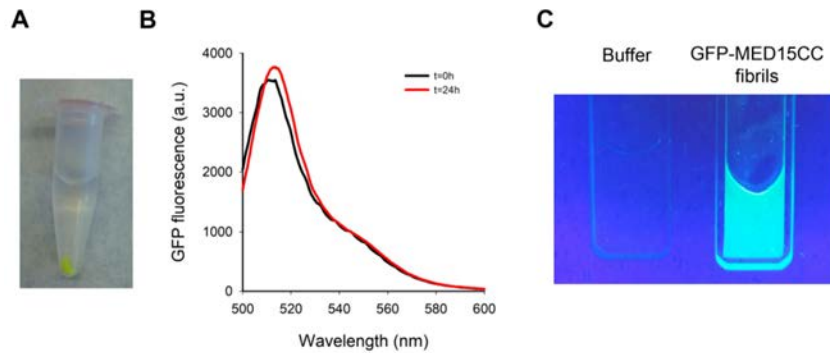


Figure S4. GFP-MED15CC keeps the native GFP fold within the fibril phase.

A) Visualization of aggregated GFP-MED15CC pellet after centrifugation. **B)** GFP fluorescence of GFP-MED15CC at initial time point ($t=0h$), and final time point of the pellet fraction ($t=24h$). **C)** Image under UV of aggregated GFP-MED15CC pellet after centrifugation and resuspended in buffer. Buffer alone is used as control.

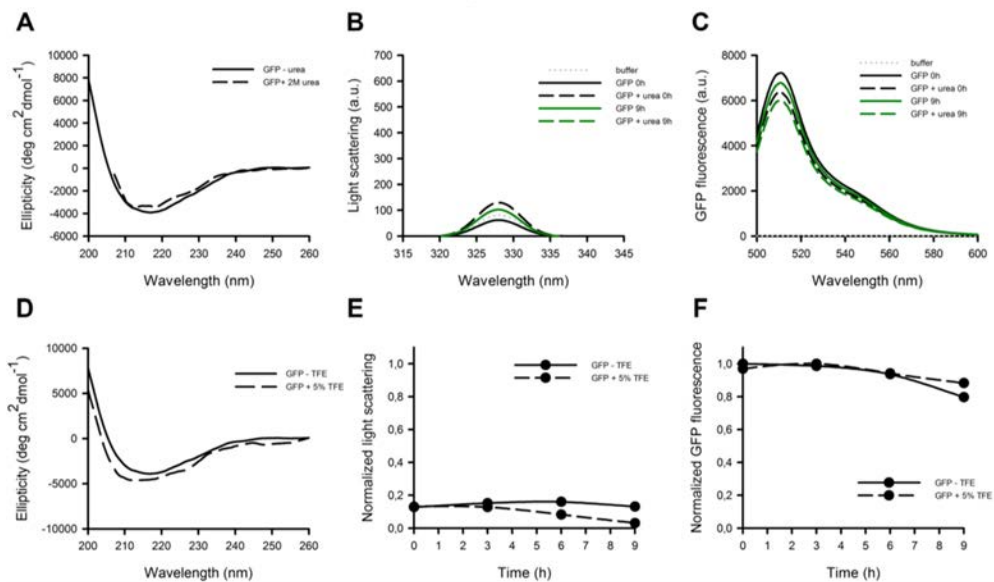


Figure S5. GFP aggregation in the presence of urea or TFE. A) Far-UV CD of $5 \mu\text{M}$ GFP in the presence or absence of 2 M urea. **B)** Synchronous light scattering and **C)** GFP fluorescence of $5 \mu\text{M}$ GFP incubated in 20 mM Tris pH 7.5 and 150 mM NaCl in the presence or absence of 2 M urea at initial time point (0 hours) or final incubation time point (9 hours). **D)** Far-UV CD of $5 \mu\text{M}$ GFP in the presence or absence of 5% TFE. **E)** Synchronous light scattering and **F)** GFP fluorescence aggregation kinetics of $5 \mu\text{M}$ GFP incubated in 20 mM Tris pH 7.5 and 150 mM NaCl in the presence or absence of 5% TFE.

TableS1_1
PLAAC positive reviewed genes

Uniprot ID	Gene names
A1KXE4	FAM168B KIAA0280L MANI
O00167	EYA2 EAB1
O00267	SUPT5H SPT5 SPT5H
O14497	ARID1A BAF250 BAF250A C1orf4 OSA1 SMARCF1
O14654	IRS4
O14686	KMT2D ALR MLL2 MLL4
O14776	TCERG1 CA150 TAF2S
O14964	HGS HRS
O14979	HNRNPD L HNRPDL JKTBP
O14994	SYN3
O15162	PLSCR1
O15405	TOX3 CAGF9 TNRC9
O15409	FOXP2 CAGH44 TNRC10
O15516	CLOCK BHLHE8 KIAA0334
O43765	SGTA SGT SGT1
O60885	BRD4 HUNK1
O75177	SS18L1 CREST KIAA0693
O75886	STAM2 HBP
O75909	CCNK CPR4
O94913	PCF11 KIAA0824
O94916	NFAT5 KIAA0827 TONEBP
O94979	SEC31A KIAA0905 SEC31L1 HSPC275 HSPC334
O95486	SEC24A
O95677	EYA4
O95835	LATS1 WARTS
P02810	PRH1; PRH2
P02812	PRB2
P04264	KRT1 KRTA
P04280	PRB1
P08047	SP1 TSFP1
P08247	SYP
P09651	HNRNPA1 HNRPA1
P0CG23	ZNF853
P13645	KRT10 KPP
P17844	DDX5 G17P1 HELR HLR1
P20073	ANXA7 ANX7 SNX OK/SW-cl.95
P20226	TBP GTF2D1 TF2D TFIID
P22626	HNRNPA2B1 HNRPA2B1
P23511	NFYA
P24928	POLR2A POLR2
P31483	TIA1
P31942	HNRNPH3 HNRPH3
P31943	HNRNPH1 HNRPH HNRPH1
P35527	KRT9
P35637	FUS TLS
P35908	KRT2 KRT2A KRT2E
P40426	PBX3
P46100	ATRX RAD54L XH2
P48431	SOX2
P48436	SOX9
P49750	YLPM1 C14orf170 ZAP3
P49790	NUP153
P50616	TOB1 TOB TROB1
P50995	ANXA11 ANX11
P51991	HNRNPA3 HNRPA3
P52594	AGFG1 HRB RAB RIP

P52948	NUP98 ADAR2
P53992	SEC24C KIAA0079
P54253	ATXN1 ATX1 SCA1
P55795	HNRNPH2 FTP3 HNRPH2
P62683	ERVK-21
P62684	HERVK_113
P63126	ERVK-9
P63130	ERVK-7
P63145	ERVK-24
P78364	PHC1 EDR1 PH1
P78424	POU6F2 RPF1
P81877	SSBP2 SSDP2
P87889	ERVK-10
Q00839	HNRNPU C1orf199 HNRPU SAFA U21.1
Q01085	TIAL1
Q01844	EWSR1 EWS
Q02446	SP4
Q02447	SP3
Q07687	DLX2
Q08211	DHX9 DDX9 LKP NDH2
Q09472	EP300 P300
Q12778	FOXO1 FKHR FOXO1A
Q12791	KCNMA1 KCNMA SLO
Q12830	BPTF FAC1 FALZ
Q12906	ILF3 DRBF MPHOSPH4 NF90
Q13117	DAZ2
Q13148	TARDBP TDP43
Q13151	HNRNPA0 HNRPA0
Q13952	NFYC
Q14103	HNRNPD AUF1 HNRPD
Q14444	CAPRIN1 GPIAP1 GPIP137 M11S1 RNG105
Q14677	CLINT1 ENTH EPN4 EPNR KIAA0171
Q14686	NCOA6 AIB3 KIAA0181 RAP250 TRBP
Q15032	R3HDM1 KIAA0029 R3HDM
Q15517	CDSN
Q15532	SS18 SSXT SYT
Q15596	NCOA2 BHLHE75 SRC2 TIF2
Q156A1	ATXN8
Q15788	NCOA1 BHLHE74 SRC1
Q17RH7	TPRXL
Q1KMD3	HNRNPUL2 HNRPUL2
Q2M218	AAK1 KIAA1048
Q2TAL8	QRICH1
Q32P51	HNRNPA1L2 HNRNPA1L
Q3L8U1	CHD9 KIAA0308 KISH2 PRIC320 AD-013 x0008
Q5D862	FLG2 IFPS
Q5TAX3	TUT4 KIAA0191 ZCCHC11
Q68CP9	ARID2 BAF200 KIAA1557
Q68DE3	USF3 KIAA2018
Q6AI39	BICRAL GLTSCR1L KIAA0240
Q6E0U4	DMKN UNQ729/PRO1411
Q6KC79	NIPBL IDN3 SCC2
Q6N021	TET2 KIAA1546 Nbla00191
Q6P1W5	C1orf94
Q6P3W7	SCYL2 CVAK104 KIAA1360
Q6XPR3	RPTN
Q6ZW49	PAXIP1 PAXIP1L PTIP CAGF28
Q7KZ85	SUPT6H KIAA0162 SPT6H
Q7LDI9	ERVK-6 ERVK6

Q7Z353	HDX CXorf43
Q7Z429	GRINA LFG1 NMDARA1 TMBIM3
Q86SG3	DAZ4
Q86US8	SMG6 C17orf31 EST1A KIAA0732
Q86VE3	SATL1
Q86YW9	MED12L KIAA1635 TNRC11L TRALP TRALPUSH PRO0314
Q86YZ3	HRNR S100A18
Q8HWS3	RFX6 RFXDC1
Q8IYB5	SMAP1
Q8IZD2	KMT2E MLL5
Q8IZL2	MAML2 KIAA1819
Q8N5C8	TAB3 MAP3K7IP3
Q8N8F6	YIPF7 FINGER9 YIP1B
Q8NCA5	FAM98A
Q8NDV7	TNRC6A CAGH26 KIAA1460 TNRC6
Q8NFD5	ARID1B BAF250B DAN15 KIAA1235 OSA2
Q8TF68	ZNF384 CAGH1 CIZ NMP4 TNRC1
Q8WU79	SMAP2 SMAP1L
Q8WUM4	PDCD6IP AIP1 ALIX KIAA1375
Q8WYB5	KAT6B KIAA0383 MORF MOZ2 MYST4
Q92540	SMG7 C1orf16 EST1C KIAA0250
Q92585	MAML1 KIAA0200
Q92734	TFG
Q92783	STAM STAM1
Q92793	CREBBP CBP
Q92794	KAT6A MOZ MYST3 RUNXBP2 ZNF220
Q92804	TAF15 RBP56 TAF2N
Q92841	DDX17
Q93074	MED12 ARC240 CAGH45 HOPA KIAA0192 TNRC11 TRAP230
Q969M3	YIPF5 FINGER5 YIP1A PP12723 SB140 UNQ3123/PRO10275
Q96JK9	MAML3 KIAA1816
Q96KN3	PKNOX2 PREP2
Q96MM7	HS6ST2 PSEC0092
Q96PK6	RBM14 SIP
Q96PN7	TRERF1 BCAR2 RAPA TREP132
Q96PV6	LENG8 KIAA1932
Q96RN5	MED15 ARC105 CTG7A PCQAP TIG1 TNRC7
Q99502	EYA1
Q99697	PITX2 ARP1 RGS RIEG RIEG1
Q99729	HNRNPAB ABBP1 HNRPAB
Q99967	CITED2 MRG1
Q9BUJ2	HNRNPUL1 E1BAP5 HNRPUL1
Q9BWU1	CDK19 CDC2L6 CDK11 KIAA1028
Q9BWW4	SSBP3 SSDP SSDP1
Q9BYJ9	YTHDF1 C20orf21
Q9BYW2	SETD2 HIF1 HYPB KIAA1732 KMT3A SET2 HSPC069
Q9BZC1	CELF4 BRUNOL4
Q9C0B9	ZCCHC2 C18orf49 KIAA1744
Q9C0J8	WDR33 WDC146
Q9GZV5	WWTR1 TAZ
Q9H0L4	CSTF2T KIAA0689
Q9H3P7	ACBD3 GCP60 GOCAP1 GOLPH1
Q9H4A3	WNK1 HSN2 KDP KIAA0344 PRKWNK1
Q9H4W6	EBF3 COE3
Q9HCJ0	TNRC6C KIAA1582
Q9NQV6	PRDM10 KIAA1231 PFM7 TRIS
Q9NQZ3	DAZ1 DAZ SPGY
Q9NR90	DAZ3

Q9NRM1	ENAM
Q9NSY1	BMP2K BIKE HRIHFB2017
Q9NYJ8	TAB2 KIAA0733 MAP3K7IP2
Q9NZN8	CNOT2 CDC36 NOT2 HSPC131 MSTP046
Q9NZW4	DSPP
Q9P267	MBD5 KIAA1461
Q9P2D1	CHD7 KIAA1416
Q9UBL0	ARPP21 TARPP
Q9UBV8	PEF1 ABP32 UNQ1845/PRO3573
Q9UGU0	TCF20 KIAA0292 SPBP
Q9UH73	EBF1 COE1 EBF
Q9UI36	DACH1 DACH
Q9UKF5	ADAM29
Q9ULJ6	ZMIZ1 KIAA1224 RAI17 ZIMP10
Q9UMX0	UBQLN1 DA41 PLIC1
Q9UPA5	BSN KIAA0434 ZNF231
Q9Y2K5	R3HDM2 KIAA1002
Q9Y3Y4	PYGO1
Q9Y6Q9	NCOA3 AIB1 BHLHE42 RAC3 TRAM1
Q9YNA8	ERVK-19

**Table S1_2. Human PrLDs with CC regions.
Coils positive, score > 0.2**

Uniprot ID	Gene names	CC domain
O14686	KMT2D ALR MLL2 MLL4	LQQLQQQQQLQQQQQLQQQQQQQLQQQQQLQQQQQLQQQQQQQQQLQQQQ QQQLQQQQQQQLQQQQQQQQQQFQQQQQQQQ
O15405	TOX3 CAGF9 TNRC9	TQQHQMQLQQQQQQQQQQMQQMQQQQQLQQHQMHQQIQQQMQQQHFQH HMQQHLQQQQQHLLQQQINQQQLQQQLQQRLQLQQQLQHMQHQSQ
O15409	FOXP2 CAGH44 TNRC10	KKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ QQQQQHPGKQAKEQQQQQQQQQQQLAAQQLVFQQQLLQMQLQQQ
O94916	NFAT5 KIAA0827 TONEBP	SQEQQQQQQQQQQQQQQQQQS
P0CG23	ZNF853	QVQEQQRLQQQQEQLQTQQAQEQQVLLQQQEQLQQQVQEQQLLQQQQEQL QQQQLLQQQEQLQQQ
P20226	TBP GTF2D1 TF2D TFIID	LSILEEQRRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQA
P54253	ATXN1 ATX1 SCA1	QQQQQQQQQQQQHQHQQQQQQQQQQQQQ
P78424	POU6F2 RPF1	TSSLNSQLQLQLQLQQQQQQQQQ
Q15596	NCOA2 BHLHE75 SRC2 TIF2	AQRQREILNQHLRQRQMHQQQ
Q156A1	ATXN8	MQQ QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
Q2M218	AAK1 KIAA1048	QQQQQQQQQQQQQQQLATALHQ
Q6ZW49	PAXIP1 PAXIP1L PTIP CAGF28	HQLQQQQLAQLQQQHSLLLQQQQQQQIQQQQLQRMHQQQQQQQMQSQ
Q8IZL2	MAML2 KIAA1819	LLHYTQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQSSISAQQQQ QQSSISAQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
Q8TF68	ZNF384 CAGH1 CIZ NMP4 TNRC1	AQAQAQAQAQAQAQAQAQAQASQASQQQQQQQQQQQQQQQQ
Q92793	CREBBP CBP	REMLRRQLLQQQQQQQQQQQQQQQQQGS
Q93074	MED12 ARC240 CAGH45 HOPA KIAA0192 TNRC11 TRAP230	QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQYHIRQQQQQQILRQQQQQQQQ QQQQQQQQQQQQQQQQQQHQHQQQQQ
Q96JK9	MAML3 KIAA1816	MAAQQQRAKLMQQKQQQQQQQQQQQQQQQQQQQQQQQQQQQQHSNQ
Q96MM7	HS6ST2 PSEC0092	QNLTQNLMQNLTQSLSQKENR
Q96RN5	MED15 ARC105 CTG7A	QQQQQQQQFQQQQQAALQQQQQQQQQQQFQAQQSAMQQQFQAVVQQQQ QLQQQQQQQHLLIKLHHQNQQIQQQQQQLQRIQLQLQQQQQQQQQQQQ QQQQALQAQ

	PCQAP TIG1 TNRC7	
Q9NSY1	BMP2K BIKE HRIHFB20 17	QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQQDAYMQQYQHATQQQQMLQ QQ
Q9Y3Y4	PYGO1	VNQSNIELKNVNRNNAVNQEN
Q9Y6Q9	NCOA3 AIB1 BHLHE42 RAC3 TRAM1	MMMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQTQ

Table S1_3. Function of human PrLD with CC regions.

Uniprot ID	Function
O14686	Histone methyltransferase. Methylates 'Lys-4' of histone H3 (H3K4me). H3K4me represents a specific tag for epigenetic transcriptional activation. Acts as a coactivator for estrogen receptor by being recruited by ESR1, thereby activating transcription. {ECO:0000269 PubMed:16603732, ECO:0000269 PubMed:17500065, ECO:0000269 PubMed:17851529}.
O15405	Transcriptional coactivator of the p300/CBP-mediated transcription complex. Activates transactivation through cAMP response element (CRE) sites. Protects against cell death by inducing antiapoptotic and repressing pro-apoptotic transcripts. Stimulates transcription from the estrogen-responsive or BCL-2 promoters. Required for depolarization-induced transcription activation of the C-FOS promoter in neurons. Associates with chromatin to the estrogen-responsive C3 promoter region. {ECO:0000269 PubMed:21172805}.
O15409	Transcriptional repressor that may play a role in the specification and differentiation of lung epithelium. May also play a role in developing neural, gastrointestinal and cardiovascular tissues. Can act with CTBP1 to synergistically repress transcription but CTPBP1 is not essential. Plays a role in synapse formation by regulating SRPX2 levels. Involved in neural mechanisms mediating the development of speech and language.
O94916	Transcription factor involved, among others, in the transcriptional regulation of osmoprotective and inflammatory genes. Mediates the transcriptional response to hypertonicity (PubMed:10051678). Positively regulates the transcription of LCN2 and S100A4 genes; optimal transactivation of these genes requires the presence of DDX5/DDX17 (PubMed:22266867). Binds the DNA consensus sequence 5'-[ACT][AG]TGGAAA[CAT]A[TA][ATC][CA][ATG][GT][GAC][CG][CT]-3' (PubMed:10377394). {ECO:0000269 PubMed:10051678, ECO:0000269 PubMed:10377394, ECO:0000269 PubMed:22266867}.
P0CG23	
P20226	General transcription factor that functions at the core of the DNA-binding multiprotein factor TFIID (PubMed:2374612, PubMed:2363050, PubMed:2194289, PubMed:9836642, PubMed:27193682). Binding of TFIID to the TATA box is the initial transcriptional step of the pre-initiation complex (PIC), playing a role in the activation of eukaryotic genes transcribed by RNA polymerase II (PubMed:2374612, PubMed:2363050, PubMed:2194289, PubMed:9836642, PubMed:27193682). Component of a BRF2-containing transcription factor complex that regulates transcription mediated by RNA polymerase III (PubMed:26638071). Component of the transcription factor SL1/TIF-IB complex, which is involved in the assembly of the PIC (pre-initiation complex) during RNA polymerase I-dependent transcription (PubMed:15970593). The rate of PIC formation probably is primarily dependent on the rate of association of SL1 with the rDNA promoter. SL1 is involved in stabilization of nucleolar transcription factor 1/UBTF on rDNA. {ECO:0000269 PubMed:15970593, ECO:0000269 PubMed:2194289, ECO:0000269 PubMed:2363050, ECO:0000269 PubMed:2374612, ECO:0000269 PubMed:26638071, ECO:0000269 PubMed:27193682, ECO:0000269 PubMed:9836642, ECO:0000305}.

P54253	Chromatin-binding factor that repress Notch signaling in the absence of Notch intracellular domain by acting as a CBF1 corepressor. Binds to the HEY promoter and might assist, along with NCOR2, RBPJ-mediated repression. Binds RNA in vitro. May be involved in RNA metabolism (PubMed:21475249). In concert with CIC and ATXN1L, involved in brain development (By similarity). {ECO:0000250 UniProtKB:P54254, ECO:0000269 PubMed:21475249}.
P78424	Probable transcription factor likely to be involved in early steps in the differentiation of amacrine and ganglion cells. Recognizes and binds to the DNA sequence 5'-ATGCAAAT-3'. Isoform 1 does not bind DNA.
Q15596	Transcriptional coactivator for steroid receptors and nuclear receptors. Coactivator of the steroid binding domain (AF-2) but not of the modulating N-terminal domain (AF-1). Required with NCOA1 to control energy balance between white and brown adipose tissues. Critical regulator of glucose metabolism regulation, acts as RORA coactivator to specifically modulate G6PC expression. Involved in the positive regulation of the transcriptional activity of the glucocorticoid receptor NR3C1 by sumoylation enhancer RWDD3. Positively regulates the circadian clock by acting as a transcriptional coactivator for the CLOCK-ARNTL/BMAL1 heterodimer (By similarity). {ECO:0000250 UniProtKB:Q61026, ECO:0000269 PubMed:23508108, ECO:0000269 PubMed:9430642}.
Q156A1	
Q2M2I8	Regulates clathrin-mediated endocytosis by phosphorylating the AP2M1/mu2 subunit of the adaptor protein complex 2 (AP-2) which ensures high affinity binding of AP-2 to cargo membrane proteins during the initial stages of endocytosis. Isoform 1 and isoform 2 display similar levels of kinase activity towards AP2M1. Regulates phosphorylation of other AP-2 subunits as well as AP-2 localization and AP-2-mediated internalization of ligand complexes. Phosphorylates NUMB and regulates its cellular localization, promoting NUMB localization to endosomes. Binds to and stabilizes the activated form of NOTCH1, increases its localization in endosomes and regulates its transcriptional activity. {ECO:0000269 PubMed:12952931, ECO:0000269 PubMed:17494869, ECO:0000269 PubMed:18657069, ECO:0000269 PubMed:21464124}.

Q6ZW49	<p>Involved in DNA damage response and in transcriptional regulation through histone methyltransferase (HMT) complexes. Plays a role in early development. In DNA damage response is required for cell survival after ionizing radiation. In vitro shown to be involved in the homologous recombination mechanism for the repair of double-strand breaks (DSBs). Its localization to DNA damage foci requires RNF8 and UBE2N. Recruits TP53BP1 to DNA damage foci and, at least in particular repair processes, effective DNA damage response appears to require the association with TP53BP1 phosphorylated by ATM at 'Ser-25'. Together with TP53BP1 regulates ATM association. Proposed to recruit PAGR1 to sites of DNA damage and the PAGR1:PAXIP1 complex is required for cell survival in response to DNA damage; the function is probably independent of MLL-containing histone methyltransferase (HMT) complexes. However, this function has been questioned (By similarity). Promotes ubiquitination of PCNA following UV irradiation and may regulate recruitment of polymerase eta and RAD51 to chromatin after DNA damage. Proposed to be involved in transcriptional regulation by linking MLL-containing histone methyltransferase (HMT) complexes to gene promoters by interacting with promoter-bound transcription factors such as PAX2. Associates with gene promoters that are known to be regulated by KMT2D/MLL2. During immunoglobulin class switching in activated B-cells is involved in trimethylation of histone H3 at 'Lys-4' and in transcription initiation of downstream switch regions at the immunoglobulin heavy-chain (Igh) locus; this function appears to involve the recruitment of MLL-containing HMT complexes. Conflictingly, its function in transcriptional regulation during immunoglobulin class switching is reported to be independent of the MLL2/MLL3 complex (By similarity). {ECO:0000250 UniProtKB:Q6NZQ4, ECO:0000269 PubMed:14576432, ECO:0000269 PubMed:15456759, ECO:0000269 PubMed:17690115, ECO:0000269 PubMed:17925232, ECO:0000269 PubMed:18353733, ECO:0000269 PubMed:20088963, ECO:0000269 PubMed:23727112}.</p>
Q8IZL2	<p>Acts as a transcriptional coactivator for NOTCH proteins. Has been shown to amplify NOTCH-induced transcription of HES1. Potentiates activation by NOTCH3 and NOTCH4 more efficiently than MAML1 or MAML3. {ECO:0000269 PubMed:12370315, ECO:0000269 PubMed:12386158, ECO:0000269 PubMed:12539049}.</p>
Q8TF68	<p>Transcription factor that binds the consensus DNA sequence [GC]AAAAA. Seems to bind and regulate the promoters of MMP1, MMP3, MMP7 and COL1A1 (By similarity). {ECO:0000250}.</p>
Q92793	<p>Acetylates histones, giving a specific tag for transcriptional activation. Also acetylates non-histone proteins, like NCOA3 and FOXO1. Binds specifically to phosphorylated CREB and enhances its transcriptional activity toward cAMP-responsive genes. Acts as a coactivator of ALX1. Acts as a circadian transcriptional coactivator which enhances the activity of the circadian transcriptional activators: NPAS2-ARNTL/BMAL1 and CLOCK-ARNTL/BMAL1 heterodimers. Acetylates PCNA; acetylation promotes removal of chromatin-bound PCNA and its degradation during nucleotide excision repair (NER) (PubMed:24939902). Functions as a transcriptional coactivator for SMAD4 in the TGF-beta signaling pathway (PubMed:25514493). {ECO:0000269 PubMed:11154691, ECO:0000269 PubMed:12738767, ECO:0000269 PubMed:12929931, ECO:0000269 PubMed:14645221, ECO:0000269 PubMed:24939902, ECO:0000269 PubMed:25514493, ECO:0000269 PubMed:9707565}.</p>

Q93074	<p>Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. This subunit may specifically regulate transcription of targets of the Wnt signaling pathway and SHH signaling pathway. {ECO:0000269 PubMed:16565090, ECO:0000269 PubMed:16595664, ECO:0000269 PubMed:17000779}.</p>
Q96JK9	<p>Acts as a transcriptional coactivator for NOTCH proteins. Has been shown to amplify NOTCH-induced transcription of HES1. {ECO:0000269 PubMed:12370315, ECO:0000269 PubMed:12386158}.</p>
Q96MM7	<p>6-O-sulfation enzyme which catalyzes the transfer of sulfate from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to position 6 of the N-sulfoglucosamine residue (GlcNS) of heparan sulfate.</p>
Q96RN5	<p>Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. Required for cholesterol-dependent gene regulation. Positively regulates the Nodal signaling pathway. {ECO:0000269 PubMed:12167862, ECO:0000269 PubMed:16630888, ECO:0000269 PubMed:16799563}.</p>
Q9NSY1	<p>May be involved in osteoblast differentiation.</p>
Q9Y3Y4	<p>Involved in signal transduction through the Wnt pathway.</p>
Q9Y6Q9	<p>Nuclear receptor coactivator that directly binds nuclear receptors and stimulates the transcriptional activities in a hormone-dependent fashion. Plays a central role in creating a multisubunit coactivator complex, which probably acts via remodeling of chromatin. Involved in the coactivation of different nuclear receptors, such as for steroids (GR and ER), retinoids (RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR) and prostanoids (PPARs). Displays histone acetyltransferase activity. Also involved in the coactivation of the NF-kappa-B pathway via its interaction with the NFKB1 subunit.</p>

Table S1_4. GO terms of the PrLDs with CC regions.

Category	Term	Count	%	PValue
GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	47	24.4791667	9.77E-19
GOTERM_BP_DIRECT	GO:0045893~positive regulation of transcription, DNA-templated	27	14.0625	1.94E-11
GOTERM_BP_DIRECT	GO:0006355~regulation of transcription, DNA-templated	44	22.9166667	4.44E-10
GOTERM_BP_DIRECT	GO:0006366~transcription from RNA polymerase II promoter	23	11.9791667	1.57E-08
GOTERM_BP_DIRECT	GO:0006351~transcription, DNA-templated	48	25	1.59E-08
GOTERM_BP_DIRECT	GO:0010467~gene expression	9	4.6875	2.87E-08
GOTERM_BP_DIRECT	GO:0000398~mRNA splicing, via spliceosome	15	7.8125	7.45E-08
GOTERM_BP_DIRECT	GO:0016032~viral process	17	8.85416667	8.34E-08
GOTERM_BP_DIRECT	GO:0006338~chromatin remodeling	9	4.6875	2.89E-06
GOTERM_BP_DIRECT	GO:0006396~RNA processing	9	4.6875	7.15E-06
GOTERM_BP_DIRECT	GO:0016569~covalent chromatin modification	9	4.6875	2.20E-05
GOTERM_BP_DIRECT	GO:0016573~histone acetylation	6	3.125	2.39E-05
GOTERM_BP_DIRECT	GO:0006367~transcription initiation from RNA polymerase II promoter	10	5.20833333	2.82E-05
GOTERM_BP_DIRECT	GO:0006473~protein acetylation	4	2.08333333	8.63E-05
GOTERM_BP_DIRECT	GO:0051028~mRNA transport	6	3.125	1.18E-04
GOTERM_BP_DIRECT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	20	10.4166667	1.62E-04
GOTERM_BP_DIRECT	GO:0006310~DNA recombination	7	3.64583333	2.15E-04
GOTERM_BP_DIRECT	GO:0017148~negative regulation of translation	6	3.125	3.23E-04
GOTERM_BP_DIRECT	GO:0051592~response to calcium ion	6	3.125	3.23E-04
GOTERM_BP_DIRECT	GO:0070935~3'-UTR-mediated mRNA stabilization	4	2.08333333	6.57E-04
GOTERM_BP_DIRECT	GO:0044255~cellular lipid metabolic process	5	2.60416667	8.18E-04
GOTERM_BP_DIRECT	GO:0007275~multicellular organism development	15	7.8125	0.0010193
GOTERM_BP_DIRECT	GO:0016576~histone dephosphorylation	3	1.5625	0.00154052
GOTERM_BP_DIRECT	GO:0006357~regulation of transcription from RNA polymerase II promoter	13	6.77083333	0.00209363
GOTERM_BP_DIRECT	GO:0007221~positive regulation of transcription of Notch receptor target	3	1.5625	0.00214213
GOTERM_BP_DIRECT	GO:0048511~rhythmic process	5	2.60416667	0.00230413
GOTERM_BP_DIRECT	GO:0042921~glucocorticoid receptor signaling pathway	3	1.5625	0.00362274
GOTERM_BP_DIRECT	GO:0006406~mRNA export from nucleus	6	3.125	0.00375129
GOTERM_BP_DIRECT	GO:0036258~multivesicular body assembly	4	2.08333333	0.00390807
GOTERM_BP_DIRECT	GO:0033148~positive regulation of intracellular estrogen receptor signaling pathway	3	1.5625	0.00449785
GOTERM_BP_DIRECT	GO:0006278~RNA-dependent DNA biosynthetic process	3	1.5625	0.00650819

GOTERM_BP_DIRECT	GO:0007283~spermatogenesis	11	5.72916667	0.00667453
GOTERM_BP_DIRECT	GO:0007219~Notch signaling pathway	6	3.125	0.00676297
GOTERM_BP_DIRECT	GO:0001829~trophoblast cell differentiation	3	1.5625	0.0076397
GOTERM_BP_DIRECT	GO:0030521~androgen receptor signaling pathway	4	2.08333333	0.00859803
GOTERM_BP_DIRECT	GO:1904837~beta-catenin-TCF complex assembly	4	2.08333333	0.00980604
GOTERM_BP_DIRECT	GO:0045948~positive regulation of translational initiation	3	1.5625	0.01014633
GOTERM_BP_DIRECT	GO:1903543~positive regulation of exosomal secretion	3	1.5625	0.01014633
GOTERM_BP_DIRECT	GO:0008543~fibroblast growth factor receptor signaling pathway	5	2.60416667	0.010207
GOTERM_BP_DIRECT	GO:0006397~mRNA processing	7	3.64583333	0.0106668
GOTERM_BP_DIRECT	GO:0015074~DNA integration	3	1.5625	0.01151791
GOTERM_BP_DIRECT	GO:0045109~intermediate filament organization	3	1.5625	0.01151791
GOTERM_BP_DIRECT	GO:0030900~forebrain development	4	2.08333333	0.01179451
GOTERM_BP_DIRECT	GO:0006368~transcription elongation from RNA polymerase II promoter	5	2.60416667	0.01200783
GOTERM_BP_DIRECT	GO:0045892~negative regulation of transcription, DNA-templated	12	6.25	0.01446077
GOTERM_BP_DIRECT	GO:0061436~establishment of skin barrier	3	1.5625	0.01448895
GOTERM_BP_DIRECT	GO:0051568~histone H3-K4 methylation	3	1.5625	0.01775259
GOTERM_BP_DIRECT	GO:0001892~embryonic placenta development	3	1.5625	0.01775259
GOTERM_BP_DIRECT	GO:0090502~RNA phosphodiester bond hydrolysis, endonucleolytic	4	2.08333333	0.01906229
GOTERM_BP_DIRECT	GO:0097150~neuronal stem cell population maintenance	3	1.5625	0.01949003
GOTERM_BP_DIRECT	GO:0030520~intracellular estrogen receptor signaling pathway	3	1.5625	0.01949003
GOTERM_BP_DIRECT	GO:0001570~vasculogenesis	4	2.08333333	0.0199922
GOTERM_BP_DIRECT	GO:1904017~cellular response to Thyroglobulin triiodothyronine	2	1.04166667	0.02049952
GOTERM_BP_DIRECT	GO:0034728~nucleosome organization	2	1.04166667	0.02049952
GOTERM_BP_DIRECT	GO:0048208~COPII vesicle coating	4	2.08333333	0.02501168
GOTERM_BP_DIRECT	GO:0030099~myeloid cell differentiation	3	1.5625	0.0251057
GOTERM_BP_DIRECT	GO:0090090~negative regulation of canonical Wnt signaling pathway	6	3.125	0.02688924
GOTERM_BP_DIRECT	GO:0031047~gene silencing by RNA	5	2.60416667	0.02783713
GOTERM_BP_DIRECT	GO:0031018~endocrine pancreas development	3	1.5625	0.02917018
GOTERM_BP_DIRECT	GO:0035303~regulation of dephosphorylation	2	1.04166667	0.03059205
GOTERM_BP_DIRECT	GO:1903551~regulation of extracellular exosome assembly	2	1.04166667	0.03059205
GOTERM_BP_DIRECT	GO:0018076~N-terminal peptidyl-lysine acetylation	2	1.04166667	0.03059205

GOTERM_BP_DIRECT	GO:0016925~protein sumoylation	5	2.60416667	0.03288452
GOTERM_BP_DIRECT	GO:0042795~snRNA transcription from RNA polymerase II promoter	4	2.08333333	0.0355945
GOTERM_BP_DIRECT	GO:0002474~antigen processing and presentation of peptide antigen via MHC class I	3	1.5625	0.03801543
GOTERM_BP_DIRECT	GO:0045739~positive regulation of DNA repair	3	1.5625	0.03801543
GOTERM_BP_DIRECT	GO:0043487~regulation of RNA stability	2	1.04166667	0.04058118
GOTERM_BP_DIRECT	GO:0014706~striated muscle tissue development	2	1.04166667	0.04058118
GOTERM_BP_DIRECT	GO:0017145~stem cell division	2	1.04166667	0.04058118
GOTERM_BP_DIRECT	GO:2000373~positive regulation of DNA topoisomerase (ATP-hydrolyzing) activity	2	1.04166667	0.04058118
GOTERM_BP_DIRECT	GO:1900034~regulation of cellular response to heat	4	2.08333333	0.0423212
GOTERM_BP_DIRECT	GO:0003007~heart morphogenesis	3	1.5625	0.04277351
GOTERM_BP_DIRECT	GO:0001843~neural tube closure	4	2.08333333	0.04517791
GOTERM_BP_DIRECT	GO:0006914~autophagy	5	2.60416667	0.04880288
GOTERM_BP_DIRECT	GO:0042634~regulation of hair cycle	2	1.04166667	0.05046797
GOTERM_BP_DIRECT	GO:0048752~semicircular canal morphogenesis	2	1.04166667	0.05046797
GOTERM_BP_DIRECT	GO:0070934~CRD-mediated mRNA stabilization	2	1.04166667	0.05046797
GOTERM_BP_DIRECT	GO:0007501~mesodermal cell fate specification	2	1.04166667	0.05046797
GOTERM_BP_DIRECT	GO:0042471~ear morphogenesis	2	1.04166667	0.05046797
GOTERM_BP_DIRECT	GO:0042059~negative regulation of epidermal growth factor receptor signaling pathway	3	1.5625	0.05290649
GOTERM_BP_DIRECT	GO:0048536~spleen development	3	1.5625	0.05290649
GOTERM_BP_DIRECT	GO:0021549~cerebellum development	3	1.5625	0.05556071
GOTERM_BP_DIRECT	GO:2001240~negative regulation of extrinsic apoptotic signaling pathway in absence of ligand	3	1.5625	0.05556071
GOTERM_BP_DIRECT	GO:0030522~intracellular receptor signaling pathway	3	1.5625	0.05826081
GOTERM_BP_DIRECT	GO:0010793~regulation of mRNA export from nucleus	2	1.04166667	0.06025346
GOTERM_BP_DIRECT	GO:0060136~embryonic process involved in female pregnancy	2	1.04166667	0.06025346
GOTERM_BP_DIRECT	GO:0009725~response to hormone	3	1.5625	0.06949609
GOTERM_BP_DIRECT	GO:2001014~regulation of skeletal muscle cell differentiation	2	1.04166667	0.06993869
GOTERM_BP_DIRECT	GO:1903608~protein localization to cytoplasmic stress granule	2	1.04166667	0.06993869
GOTERM_BP_DIRECT	GO:0035194~posttranscriptional gene silencing by RNA	2	1.04166667	0.06993869
GOTERM_BP_DIRECT	GO:0008584~male gonad development	4	2.08333333	0.07311489
GOTERM_BP_DIRECT	GO:1903071~positive regulation of ER-associated ubiquitin-dependent protein catabolic process	2	1.04166667	0.07952466
GOTERM_BP_DIRECT	GO:2000020~positive regulation of male gonad	2	1.04166667	0.07952466

	development			
	GO:2000369~regulation of clathrin-mediated endocytosis	2	1.04166667	0.07952466
GOTERM_BP_DIRECT	GO:0003151~outflow tract morphogenesis	3	1.5625	0.0813719
GOTERM_BP_DIRECT	GO:0007289~spermatid nucleus differentiation	2	1.04166667	0.0890124
	GO:0048096~chromatin-mediated maintenance of transcription	2	1.04166667	0.0890124
GOTERM_BP_DIRECT	GO:0014070~response to organic cyclic compound	3	1.5625	0.09065693
GOTERM_BP_DIRECT	GO:0001649~osteoblast differentiation	4	2.08333333	0.09238192
GOTERM_BP_DIRECT	GO:0019827~stem cell population maintenance	3	1.5625	0.09381843
	GO:0002223~stimulatory C-type lectin receptor signaling pathway	4	2.08333333	0.09441417
GOTERM_BP_DIRECT	GO:0032481~positive regulation of type I interferon production	3	1.5625	0.09701149
GOTERM_BP_DIRECT	GO:0060009~Sertoli cell development	2	1.04166667	0.09840291
GOTERM_BP_DIRECT	GO:0060487~lung epithelial cell differentiation	2	1.04166667	0.09840291
GOTERM_BP_DIRECT	GO:0006468~protein phosphorylation	9	4.6875	0.09978548

Category	Term	Count	%	PValue
GOTERM_MF_DIRECT	GO:0003713~transcription coactivator activity	23	11.9791667	3.65E-14
GOTERM_MF_DIRECT	GO:0000166~nucleotide binding	26	13.5416667	7.56E-14
GOTERM_MF_DIRECT	GO:0003676~nucleic acid binding	40	20.8333333	1.40E-12
GOTERM_MF_DIRECT	GO:0003723~RNA binding	30	15.625	1.50E-12
GOTERM_MF_DIRECT	GO:0044822~poly(A) RNA binding	42	21.875	5.33E-12
GOTERM_MF_DIRECT	GO:0003677~DNA binding	49	25.5208333	2.38E-10
GOTERM_MF_DIRECT	GO:0003682~chromatin binding	20	10.4166667	5.67E-08
GOTERM_MF_DIRECT	GO:0004402~histone acetyltransferase activity	8	4.16666667	8.23E-07
GOTERM_MF_DIRECT	GO:0030374~ligand-dependent nuclear receptor transcription coactivator activity	8	4.16666667	1.26E-06
GOTERM_MF_DIRECT	GO:0004190~aspartic-type endopeptidase activity	7	3.64583333	1.34E-06
GOTERM_MF_DIRECT	GO:0001105~RNA polymerase II transcription coactivator activity	7	3.64583333	2.71E-06
GOTERM_MF_DIRECT	GO:0005515~protein binding	127	66.1458333	3.90E-06
GOTERM_MF_DIRECT	GO:0001104~RNA polymerase II transcription cofactor activity	6	3.125	4.18E-05
GOTERM_MF_DIRECT	GO:0008270~zinc ion binding	29	15.1041667	6.75E-05
GOTERM_MF_DIRECT	GO:0008134~transcription factor binding	12	6.25	2.91E-04
GOTERM_MF_DIRECT	GO:0016407~acetyltransferase activity	4	2.08333333	3.36E-04
GOTERM_MF_DIRECT	GO:0030331~estrogen receptor binding	5	2.60416667	6.81E-04
GOTERM_MF_DIRECT	GO:0001077~transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	10	5.20833333	0.00115259
GOTERM_MF_DIRECT	GO:0016922~ligand-dependent nuclear receptor binding	4	2.08333333	0.00146627
GOTERM_MF_DIRECT	GO:0005198~structural molecule activity	10	5.20833333	0.00157978
GOTERM_MF_DIRECT	GO:0035257~nuclear hormone receptor binding	4	2.08333333	0.00168426
GOTERM_MF_DIRECT	GO:0000978~RNA polymerase II core promoter proximal region sequence-specific DNA binding	12	6.25	0.00180777
GOTERM_MF_DIRECT	GO:0044212~transcription regulatory region DNA binding	9	4.6875	0.00236871
GOTERM_MF_DIRECT	GO:0034046~poly(G) binding	3	1.5625	0.00239375
GOTERM_MF_DIRECT	GO:0046966~thyroid hormone receptor binding	4	2.08333333	0.00307372
GOTERM_MF_DIRECT	GO:0097157~pre-mRNA intronic binding	3	1.5625	0.00316882
GOTERM_MF_DIRECT	GO:0003964~RNA-directed DNA polymerase activity	3	1.5625	0.00316882
GOTERM_MF_DIRECT	GO:0048306~calcium-dependent protein binding	5	2.60416667	0.00366429

GOTERM_MF_DIRECT	GO:0003700~transcription factor activity, sequence-specific DNA binding	21	10.9375	0.00414868
GOTERM_MF_DIRECT	GO:0043021~ribonucleoprotein complex binding	4	2.08333333	0.00416559
GOTERM_MF_DIRECT	GO:0001078~transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding	6	3.125	0.0073691
GOTERM_MF_DIRECT	GO:0003724~RNA helicase activity	3	1.5625	0.00851682
GOTERM_MF_DIRECT	GO:0050681~androgen receptor binding	4	2.08333333	0.00936272
GOTERM_MF_DIRECT	GO:0004523~RNA-DNA hybrid ribonuclease activity	3	1.5625	0.01130235
GOTERM_MF_DIRECT	GO:0046965~retinoid X receptor binding	3	1.5625	0.01130235
GOTERM_MF_DIRECT	GO:0001046~core promoter sequence-specific DNA binding	4	2.08333333	0.01142036
GOTERM_MF_DIRECT	GO:0003729~mRNA binding	6	3.125	0.01194807
GOTERM_MF_DIRECT	GO:0008013~beta-catenin binding	5	2.60416667	0.01235383
GOTERM_MF_DIRECT	GO:0070034~telomerase RNA binding	3	1.5625	0.01282517
GOTERM_MF_DIRECT	GO:0004386~helicase activity	5	2.60416667	0.013948
GOTERM_MF_DIRECT	GO:0017091~AU-rich element binding	3	1.5625	0.01443198
GOTERM_MF_DIRECT	GO:0003730~mRNA 3'-UTR binding	4	2.08333333	0.0162605
GOTERM_MF_DIRECT	GO:0034212~peptide N-acetyltransferase activity	2	1.04166667	0.02168149
GOTERM_MF_DIRECT	GO:0046983~protein dimerization activity	6	3.125	0.02431311
GOTERM_MF_DIRECT	GO:0042162~telomeric DNA binding	3	1.5625	0.02786828
GOTERM_MF_DIRECT	GO:0019899~enzyme binding	9	4.6875	0.02987181
GOTERM_MF_DIRECT	GO:0001047~core promoter binding	4	2.08333333	0.03268624
GOTERM_MF_DIRECT	GO:0001103~RNA polymerase II repressing transcription factor binding	3	1.5625	0.03469756
GOTERM_MF_DIRECT	GO:0000981~RNA polymerase II transcription factor activity, sequence-specific DNA binding	6	3.125	0.03949458
GOTERM_MF_DIRECT	GO:0003712~transcription cofactor activity	4	2.08333333	0.04245618
GOTERM_MF_DIRECT	GO:0070087~chromo shadow domain binding	2	1.04166667	0.06365173
GOTERM_MF_DIRECT	GO:0016817~hydrolase activity, acting on acid anhydrides	2	1.04166667	0.06365173
GOTERM_MF_DIRECT	GO:0001102~RNA polymerase II activating transcription factor binding	3	1.5625	0.06432763
GOTERM_MF_DIRECT	GO:0018024~histone-lysine N-methyltransferase activity	3	1.5625	0.06733293
GOTERM_MF_DIRECT	GO:1990247~N6-methyladenosine-containing RNA binding	2	1.04166667	0.0738614
GOTERM_MF_DIRECT	GO:0003697~single-stranded DNA binding	4	2.08333333	0.08131122

GOTERM_MF_DIRECT	GO:0001085~RNA polymerase II transcription factor binding	3	1.5625	0.09291282
GOTERM_MF_DIRECT	GO:0008494~translation activator activity	2	1.0416667	0.09394975

Category	Term	Count	%	PValue
GOTERM_CC_DIRECT	GO:0005634~nucleus	122	63.5416667	2.97E-24
GOTERM_CC_DIRECT	GO:0005654~nucleoplasm	83	43.2291667	1.47E-21
	GO:0030529~intracellular ribonucleoprotein			
GOTERM_CC_DIRECT	complex	16	8.33333333	7.75E-12
GOTERM_CC_DIRECT	GO:0019028~viral capsid	7	3.64583333	4.53E-09
	GO:0019013~viral			
GOTERM_CC_DIRECT	nucleocapsid	7	3.64583333	3.06E-07
GOTERM_CC_DIRECT	GO:0005737~cytoplasm	84	43.75	1.38E-06
	GO:0019031~viral			
GOTERM_CC_DIRECT	envelope	5	2.60416667	4.29E-05
	GO:0005667~transcription factor complex			
GOTERM_CC_DIRECT	GO:0033565~ESCRT-0	10	5.20833333	1.55E-04
GOTERM_CC_DIRECT	complex	3	1.5625	3.02E-04
	GO:0090544~BAF-type			
GOTERM_CC_DIRECT	complex	3	1.5625	0.00148083
	GO:0032993~protein-DNA			
GOTERM_CC_DIRECT	complex	4	2.08333333	0.00197746
	GO:0071013~catalytic			
GOTERM_CC_DIRECT	step 2 spliceosome	6	3.125	0.00240524
GOTERM_CC_DIRECT	GO:0005730~nucleolus	19	9.89583333	0.00267632
	GO:0030127~COPII vesicle			
GOTERM_CC_DIRECT	coat	3	1.5625	0.00432585
	GO:0016592~mediator			
GOTERM_CC_DIRECT	complex	4	2.08333333	0.00522658
	GO:0070971~endoplasmic			
GOTERM_CC_DIRECT	reticulum exit site	3	1.5625	0.00735042
	GO:0042405~nuclear			
GOTERM_CC_DIRECT	inclusion body	3	1.5625	0.00735042
	GO:0071565~nBAF			
GOTERM_CC_DIRECT	complex	3	1.5625	0.00851889
	GO:0000790~nuclear			
GOTERM_CC_DIRECT	chromatin	7	3.64583333	0.01374281
	GO:0005681~spliceosomal			
GOTERM_CC_DIRECT	complex	5	2.60416667	0.0151672
	GO:0016607~nuclear			
GOTERM_CC_DIRECT	speck	7	3.64583333	0.01647892
	GO:0005697~telomerase			
GOTERM_CC_DIRECT	holoenzyme complex	3	1.5625	0.01877141
	GO:0035097~histone methyltransferase			
GOTERM_CC_DIRECT	complex	3	1.5625	0.02231963
	GO:0043231~intracellular membrane-bounded			
GOTERM_CC_DIRECT	organelle	12	6.25	0.02649128

GOTERM_CC_DIRECT	GO:0016602~CCAAT-binding factor complex	2	1.04166667	0.03978199
GOTERM_CC_DIRECT	GO:0097165~nuclear stress granule	2	1.04166667	0.04947903
GOTERM_CC_DIRECT	GO:0010494~cytoplasmic stress granule	3	1.5625	0.05105274
GOTERM_CC_DIRECT	GO:0070937~CRD-mediated mRNA stability complex	2	1.04166667	0.05907867
GOTERM_CC_DIRECT	GO:0070776~MOZ/MORF histone acetyltransferase complex	2	1.04166667	0.06858189
GOTERM_CC_DIRECT	GO:0001533~cornified envelope	3	1.5625	0.07861826
GOTERM_CC_DIRECT	GO:0044666~MLL3/4 complex	2	1.04166667	0.08730287
GOTERM_CC_DIRECT	GO:0012507~ER to Golgi transport vesicle membrane	3	1.5625	0.09691428

Table S2_1. Mouse MED15 interactors homology with humans.**List of MED15 interactors extracted from (Quevedo et al, 2019)**

Mouse input	Human best scoring homologue ID	% Identity BLAST
sp Q924H2 MED15_MOUSE	Q96RN5.2	90.397
sp Q9CQA5 MED4_MOUSE	Q9NPJ6.1	94.574
sp Q9DB40 MED27_MOUSE	Q6P2C8.1	98.071
sp Q9R0X0 MED20_MOUSE	Q9H944.1	97.17
tr Q9DAY7 Q9DAY7_MOUSE	Q96G25.2	75
sp Q921D4 MED6_MOUSE	O75586.2	95.122
sp A2ABV5 MED14_MOUSE	O60244.2	96.438
sp Q920D3 MED28_MOUSE	Q9H204.1	94.382
sp Q9CXU1 MED31_MOUSE	Q9Y3C7.1	97.71
sp Q9CXU0 MED10_MOUSE	Q9BTT4.1	99.259
sp Q62276 MED22_MOUSE	Q15528.2	95.477
sp Q8VCD5 MED17_MOUSE	Q9NVC6.2	95.699
sp Q9CZ82 MED18_MOUSE	Q9BUE0.1	98.558
tr A2AGH9 A2AGH9_MOUSE	Q93074.4	96.703
sp Q99K74 MED24_MOUSE	O75448.1	94.237
tr G3UW74 G3UW74_MOUSE	Q9Y2X0.2	92.692
tr E9QNV2 E9QNV2_MOUSE	Q9ULK4.2	95.852
tr Q3UXL9 Q3UXL9_MOUSE	P24863.2	99.248
sp Q9DB91 MED29_MOUSE	Q9NX70.1	93.514
sp Q9CQI9 MED30_MOUSE	Q96HR3.1	94.944
sp Q8BWD8 CDK19_MOUSE	Q9BWU1.1	96.414
sp Q5SWW4 MED13_MOUSE	Q9UHV7.3	93.468
sp Q925J9 MED1_MOUSE	Q15648.4	93.493
sp Q8C1S0 MED19_MOUSE	A0JLT2.2	94.262
sp Q8VCB2 MED25_MOUSE	Q71SY5.2	91.299
sp Q7TN02 MED26_MOUSE	O95402.2	84.667
tr E9QLJ3 E9QLJ3_MOUSE	Q71F56.1	91.964
tr F8VPR5 F8VPR5_MOUSE	Q92793.3	95.624
sp B2RWS6 EP300_MOUSE	Q09472.2	92.81
sp Q9Z2D8 MBD3_MOUSE	O95983.1	97.753
tr E9QAS5 E9QAS5_MOUSE	Q14839.2	98.127
sp Q9R190 MTA2_MOUSE	O94776.1	98.503
tr E9QMN5 E9QMN5_MOUSE	Q86YP4.1	87.912
sp Q8VHR5 P66B_MOUSE	Q8WXI9.1	98.316
tr F8WHY8 F8WHY8_MOUSE	Q13330.2	94.825
sp Q80UW8 RPAB1_MOUSE	P19388.4	98.571
tr A0A087WR08 A0A087WR08_MOUSE	Q16594.1	96.774
sp P08775 RPB1_MOUSE	P24928.2	99.898

sp Q99J95 CDK9_MOUSE	P50750.3	98.656
sp Q8BFQ4 WDR82_MOUSE	Q6UXN9.1	100
sp Q8BX09 RBBP5_MOUSE	Q15291.2	99.071
tr D3YYA0 D3YYA0_MOUSE	Q9UBL3.1	93.258
sp Q61466 SMRD1_MOUSE	Q96GM5.2	99.417
sp O54941 SMCE1_MOUSE	Q969G3.2	97.324
tr Q3UID0 Q3UID0_MOUSE	Q8TAQ2.1	93.937
sp Q8BHJ5 TBL1R_MOUSE	Q9BZK7.1	99.027
sp Q9QXE7 TBL1X_MOUSE	O60907.3	94.118
tr F8VQL9 F8VQL9_MOUSE	Q9Y618.3	86.046
tr Q5RIM6 Q5RIM6_MOUSE	O75376.2	91.13
sp Q6ZQ88 KDM1A_MOUSE	O60341.2	98.124
sp Q8C796 RCOR2_MOUSE	Q8IZ40.2	97.514
tr Z4YJZ7 Z4YJZ7_MOUSE	Q9H9B1.4	88.13
sp Q9WVG6 CARM1_MOUSE	Q86X55.3	98.023
sp Q61026 NCOA2_MOUSE	Q15596.2	94.057
sp Q9CXY6 ILF2_MOUSE	Q12905.2	100
sp Q9R059 FHL3_MOUSE	Q13643.4	94.464
tr G3X8R8 G3X8R8_MOUSE	Q9ULH7.3	82.386
sp Q9WUP7 UCHL5_MOUSE	Q9Y5K5.3	96.96
sp P70365 NCOA1_MOUSE	Q15788.3	91.975
sp Q9DBR0 AKAP8_MOUSE	O43823.1	79.076
sp A2AJK6 CHD7_MOUSE	Q9P2D1.3	94.663
tr E9Q8Z6 E9Q8Z6_MOUSE	O60716.1	96.798
tr Q45VK5 Q45VK5_MOUSE	Q12906.3	94.949
tr D3YUG5 D3YUG5_MOUSE	Q969V6.1	82.56
tr F6U238 F6U238_MOUSE	Q8IZL2.2	84.475
sp Q9QYH6 MAGD1_MOUSE	Q9Y5V3.3	86.752
sp Q91ZW3 SMCA5_MOUSE	O60264.1	97.624
sp Q8CH18 CCAR1_MOUSE	Q8IX12.2	94.087
tr A0A0A0MQ98 A0A0A0MQ98_MOUSE	Q15652.2	86.04
tr A2A655 A2A655_MOUSE	Q12830.3	78.196
tr E9PZA7 E9PZA7_MOUSE	Q9Y4A5.3	98.161
sp Q8CHI8 EP400_MOUSE	Q96L91.4	85.319
sp Q9CU62 SMC1A_MOUSE	Q14683.2	99.838
sp Q8BHG9 CGBP1_MOUSE	Q9UFW8.2	98.204
sp Q00899 TTY1_MOUSE	P25490.2	98.558
tr B1AUC0 B1AUC0_MOUSE	Q12857.2	99.154
tr Q60I23 Q60I23_MOUSE	P48431.1	97.806
tr A2BG76 A2BG76_MOUSE	O00712.2	99.275
sp Q62255 SALL3_MOUSE	Q9BXA9.2	81.619
sp Q9CU65 ZMYM2_MOUSE	Q9UBW7.1	96.078
sp Q61286 HTF4_MOUSE	Q99081.1	91.926

sp Q91VN1 ZNF24_MOUSE	P17028.4	95.109
sp Q04207 TF65_MOUSE	Q04206.2	88.246
sp Q61624 ZN148_MOUSE	Q9UQR1.2	97.607
sp Q925H1 TRPS1_MOUSE	Q9UHF7.2	93.13
sp Q99LI5 ZN281_MOUSE	Q9Y2X9.1	95.576
sp B1AWL2 ZN462_MOUSE	Q96JM2.3	93.678
tr E9Q8G4 E9Q8G4_MOUSE	P15884.3	94.228
sp Q99PQ2 TRI11_MOUSE	Q96F44.2	89.53
tr A2AMY5 A2AMY5_MOUSE	Q5T6F2.1	79.332
tr E9Q2H1 E9Q2H1_MOUSE	O95071.2	97.999
sp Q8CGY8 OGT1_MOUSE	O15294.3	99.618
sp Q8N7N5 DCAF8_MOUSE	Q5TAQ9.1	95.485
tr E9QME5 E9QME5_MOUSE	Q9UPN9.3	92.989
sp G5E870 TRIPC_MOUSE	Q14669.1	96.691
sp O88196 TTC3_MOUSE	P53804.2	76.01
sp Q9Z2X1 HNRPF_MOUSE	P52597.3	98.072
sp P97315 CSR1_MOUSE	P21291.3	99.482
sp Q8VDP4 CCAR2_MOUSE	Q8N163.2	91.441
tr A0A0A6YWB0 A0A0A6YWB0_MOUSE	Q9NR56.2	99.408
sp Q91YT7 YTHD2_MOUSE	Q9Y5A9.2	99.655
sp Q9EPU0 RENT1_MOUSE	Q92900.2	98.494
sp Q8C2Q3 RBM14_MOUSE	Q96PK6.2	98.356
sp O88532 ZFR_MOUSE	Q96KR1.2	97.765
sp Q9QYS9 QKI_MOUSE	Q96PU8.1	100
tr B7ZWL1 B7ZWL1_MOUSE	A5YKK6.2	97.264
sp O35218 CPSF2_MOUSE	Q9P210.2	97.954
sp Q9EPU4 CPSF1_MOUSE	Q10570.2	96.535
sp Q8K2A7 INT10_MOUSE	Q9NVR2.2	95.775
sp Q6P4S8 INT1_MOUSE	Q8N201.2	90.114
sp P23198 CBX3_MOUSE	Q13185.4	99.454
sp Q7TSZ8 NACC1_MOUSE	Q96RE7.1	86.415
sp Q80X50 UBP2L_MOUSE	Q14157.2	96.387
sp Q9CQJ4 RING2_MOUSE	Q99496.1	99.702
sp Q60520 SIN3A_MOUSE	Q96ST3.2	98.038
sp Q9Z103 ADNP_MOUSE	Q9H2P0.1	93.604
tr A2BIE1 A2BIE1_MOUSE	Q2KHR3.3	81.312
sp Q3UA37 QRIC1_MOUSE	Q2TAL8.1	99.099
sp Q9DCT8 CRIP2_MOUSE	P52943.1	92.788
sp P70168 IMB1_MOUSE	Q14974.2	99.201
sp Q3U1J4 DDB1_MOUSE	Q16531.1	99.649
sp Q9Z191 EYA4_MOUSE	O95677.2	91.236
TrEMBL F7ARK3 Release	Q9Y467.4	87.778

Table S2_2. MED15 interactors with CC regions and their function.

Coils predictions with positive score > 0.2

Uniprot ID	Gene names	Function [CC]
O60264	SMARCA5 SNF2H WCRF135	Helicase that possesses intrinsic ATP-dependent nucleosome-remodeling activity. Complexes containing SMARCA5 are capable of forming ordered nucleosome arrays on chromatin; this may require intact histone H4 tails. Also required for replication of pericentric heterochromatin in S-phase specifically in conjunction with BAZ1A. Probably plays a role in repression of polI dependent transcription of the rDNA locus, through the recruitment of the SIN3/HDAC1 corepressor complex to the rDNA promoter. Essential component of the WICH complex, a chromatin remodeling complex that mobilizes nucleosomes and reconfigures irregular chromatin to a regular nucleosomal array structure. The WICH complex regulates the transcription of various genes, has a role in RNA polymerase I and RNA polymerase III transcription, mediates the histone H2AX phosphorylation at 'Tyr-142', and is involved in the maintenance of chromatin structures during DNA replication processes. Essential component of the NoRC (nucleolar remodeling complex) complex, a complex that mediates silencing of a fraction of rDNA by recruiting histone-modifying enzymes and DNA methyltransferases, leading to heterochromatin formation and transcriptional silencing. {ECO:0000269 PubMed:10880450, ECO:0000269 PubMed:11980720, ECO:0000269 PubMed:12198550, ECO:0000269 PubMed:12434153, ECO:0000269 PubMed:12972596, ECO:0000269 PubMed:15543136, ECO:0000269 PubMed:16603771}.
O60341	KDM1A AOF2 KDM1 KIAA0601 LSD1	Histone demethylase that can demethylate both 'Lys-4' (H3K4me) and 'Lys-9' (H3K9me) of histone H3, thereby acting as a coactivator or a corepressor, depending on the context (PubMed:15620353, PubMed:15811342, PubMed:16140033, PubMed:16079794, PubMed:16079795, PubMed:16223729). Acts by oxidizing the substrate by FAD to generate the corresponding imine that is subsequently hydrolyzed (PubMed:15620353, PubMed:15811342, PubMed:16079794, PubMed:21300290). Acts as a corepressor by mediating demethylation of H3K4me, a specific tag for epigenetic transcriptional activation. Demethylates both mono- (H3K4me1) and di-methylated (H3K4me2) H3K4me (PubMed:15620353, PubMed:20389281, PubMed:21300290, PubMed:23721412). May play a role in the repression of neuronal genes. Alone, it is unable to demethylate H3K4me on nucleosomes and requires the presence of RCOR1/CoREST to achieve such activity (PubMed:16140033, PubMed:16079794, PubMed:16885027, PubMed:21300290, PubMed:23721412). Also acts as a coactivator of androgen receptor (ANDR)-dependent transcription, by being recruited to ANDR target genes and mediating demethylation of H3K9me, a specific tag for epigenetic transcriptional repression. The presence of PRKCB in ANDR-containing complexes, which mediates phosphorylation of 'Thr-6' of histone H3 (H3T6ph), a specific tag that prevents demethylation H3K4me, prevents H3K4me demethylase activity of KDM1A (PubMed:16079795). Demethylates di-methylated 'Lys-370' of p53/TP53 which prevents interaction of p53/TP53 with TP53BP1 and represses p53/TP53-mediated transcriptional activation. Demethylates and stabilizes the DNA methylase DNMT1. Required for gastrulation during embryogenesis. Component of a RCOR/GFI/KDM1A/HDAC complex that suppresses, via histone deacetylase (HDAC) recruitment, a number of genes implicated in multilineage blood cell development. Effector of SNAI1-mediated transcription repression of E-cadherin/CDH1, CDN7 and KRT8. Required for the maintenance of the silenced state of the SNAI1 target genes E-cadherin/CDH1 and CDN7 (PubMed:20389281). {ECO:0000269 PubMed:12032298, ECO:0000269 PubMed:15620353, ECO:0000269 PubMed:15811342, ECO:0000269 PubMed:16079794, ECO:0000269 PubMed:16079795, ECO:0000269 PubMed:16140033, ECO:0000269 PubMed:16223729, ECO:0000269 PubMed:16885027, ECO:0000269 PubMed:16956976, ECO:0000269 PubMed:17805299, ECO:0000269 PubMed:20228790, ECO:0000269 PubMed:20389281, ECO:0000269 PubMed:20562920, ECO:0000269 PubMed:21300290, ECO:0000269 PubMed:23721412}.
O60716	CTNND1 KIAA0384	Binds to and inhibits the transcriptional repressor ZBTB33, which may lead to activation of target genes of the Wnt signaling pathway (By similarity). Associates with and regulates the cell adhesion properties of both C-, E- and N-cadherins, being critical for their surface stability. Implicated both in cell transformation by SRC and in ligand-induced receptor

		signaling through the EGF, PDGF, CSF-1 and ERBB2 receptors. Promotes GLIS2 C-terminal cleavage. {ECO:0000250, ECO:0000269 PubMed:17344476, ECO:0000269 PubMed:20371349}.
O75376	NCOR1 KIAA1047	Mediates transcriptional repression by certain nuclear receptors (PubMed:20812024). Part of a complex which promotes histone deacetylation and the formation of repressive chromatin structures which may impede the access of basal transcription factors. Participates in the transcriptional repressor activity produced by BCL6. Recruited by ZBTB7A to the androgen response elements/ARE on target genes, negatively regulates androgen receptor signaling and androgen-induced cell proliferation (PubMed:20812024). Mediates the NR1D1-dependent repression and circadian regulation of TSHB expression (By similarity). The NCOR1-HDAC3 complex regulates the circadian expression of the core clock gene ARTNL/BMAL1 and the genes involved in lipid metabolism in the liver (By similarity). {ECO:0000250 UniProtKB:Q60974, ECO:0000269 PubMed:14527417, ECO:0000269 PubMed:20812024}.
O95071	UBR5 EDD EDD1 HYD KIAA0896	E3 ubiquitin-protein ligase which is a component of the N-end rule pathway. Recognizes and binds to proteins bearing specific N-terminal residues that are destabilizing according to the N-end rule, leading to their ubiquitination and subsequent degradation (By similarity). Involved in maturation and/or transcriptional regulation of mRNA by activating CDK9 by polyubiquitination. May play a role in control of cell cycle progression. May have tumor suppressor function. Regulates DNA topoisomerase II binding protein (TopBP1) in the DNA damage response. Plays an essential role in extraembryonic development. Ubiquitinates acetylated PCK1. Also acts as a regulator of DNA damage response by acting as a suppressor of RNF168, an E3 ubiquitin-protein ligase that promotes accumulation of 'Lys-63'-linked histone H2A and H2AX at DNA damage sites, thereby acting as a guard against excessive spreading of ubiquitinated chromatin at damaged chromosomes. {ECO:0000250, ECO:0000269 PubMed:21127351, ECO:0000269 PubMed:21726808, ECO:0000269 PubMed:22884692}.
O95983	MBD3	Acts as transcriptional repressor and plays a role in gene silencing. Does not bind to DNA by itself (PubMed:12124384). Binds to DNA with a preference for sites containing methylated CpG dinucleotides (in vitro). Binds to a lesser degree DNA containing unmethylated CpG dinucleotides (PubMed:24307175). Recruits histone deacetylases and DNA methyltransferases. {ECO:0000269 PubMed:10947852, ECO:0000269 PubMed:12124384, ECO:0000269 PubMed:18644863, ECO:0000269 PubMed:23361464, ECO:0000269 PubMed:24307175, ECO:0000269 PubMed:9774669}.
P24928	POLR2A POLR2	DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates. Largest and catalytic component of RNA polymerase II which synthesizes mRNA precursors and many functional non-coding RNAs. Forms the polymerase active center together with the second largest subunit. Pol II is the central component of the basal RNA polymerase II transcription machinery. It is composed of mobile elements that move relative to each other. RPB1 is part of the core element with the central large cleft, the clamp element that moves to open and close the cleft and the jaws that are thought to grab the incoming DNA template. At the start of transcription, a single-stranded DNA template strand of the promoter is positioned within the central active site cleft of Pol II. A bridging helix emanates from RPB1 and crosses the cleft near the catalytic site and is thought to promote translocation of Pol II by acting as a ratchet that moves the RNA-DNA hybrid through the active site by switching from straight to bent conformations at each step of nucleotide addition. During transcription elongation, Pol II moves on the template as the transcript elongates. Elongation is influenced by the phosphorylation status of the C-terminal domain (CTD) of Pol II largest subunit (RPB1), which serves as a platform for assembly of factors that regulate transcription initiation, elongation, termination and mRNA processing. Regulation of gene expression levels depends on the balance between methylation and acetylation levels of the CTD-lysines (By similarity). Initiation or early elongation steps of transcription of growth-factors-induced immediate early genes are regulated by the acetylation status of the CTD (PubMed:24207025). Methylation and dimethylation have a repressive effect on target genes expression (By similarity). {ECO:0000250 UniProtKB:P08775, ECO:0000269 PubMed:20231364, ECO:0000269 PubMed:23748380, ECO:0000269 PubMed:24207025, ECO:0000269 PubMed:26124092, ECO:0000269 PubMed:9852112};. FUNCTION: (Microbial infection) Acts as an RNA-dependent RNA polymerase when associated with small delta antigen of Hepatitis delta virus, acting both as a replicate and transcriptase for the viral RNA circular genome.

		{ECO:0000269 PubMed:18032511}.
P53804	TTC3 DCRR1 RNF105 TPRD	E3 ubiquitin-protein ligase that mediates the ubiquitination and subsequent degradation of phosphorylated Akt (AKT1, AKT2 and AKT3) in the nucleus. Acts as a terminal regulator of Akt signaling after activation; its phosphorylation by Akt, which is a prerequisite for ubiquitin ligase activity, suggests the existence of a regulation mechanism required to control Akt levels after activation. Catalyzes the formation of 'Lys-48'-polyubiquitin chains. May play a role in neuronal differentiation inhibition via its interaction with CIT. {ECO:0000269 PubMed:17488780, ECO:0000269 PubMed:20059950}.
Q09472	EP300 P300	Functions as histone acetyltransferase and regulates transcription via chromatin remodeling (PubMed:23415232, PubMed:23934153, PubMed:8945521). Acetylates all four core histones in nucleosomes. Histone acetylation gives an epigenetic tag for transcriptional activation (PubMed:23415232, PubMed:23934153, PubMed:8945521). Mediates cAMP-gene regulation by binding specifically to phosphorylated CREB protein. Mediates acetylation of histone H3 at 'Lys-122' (H3K122ac), a modification that localizes at the surface of the histone octamer and stimulates transcription, possibly by promoting nucleosome instability. Mediates acetylation of histone H3 at 'Lys-27' (H3K27ac) (PubMed:23911289). Also functions as acetyltransferase for non-histone targets, such as ALX1, HDAC1, PRMT1 or SIRT2 (PubMed:12929931, PubMed:16762839, PubMed:18722353). Acetylates 'Lys-131' of ALX1 and acts as its coactivator (PubMed:12929931). Acetylates SIRT2 and is proposed to indirectly increase the transcriptional activity of TP53 through acetylation and subsequent attenuation of SIRT2 deacetylase function (PubMed:18722353). Acetylates HDAC1 leading to its inactivation and modulation of transcription (PubMed:16762839). Acts as a TFAP2A-mediated transcriptional coactivator in presence of CITED2 (PubMed:12586840). Plays a role as a coactivator of NEUROD1-dependent transcription of the secretin and p21 genes and controls terminal differentiation of cells in the intestinal epithelium. Promotes cardiac myocyte enlargement. Can also mediate transcriptional repression. Acetylates FOXO1 and enhances its transcriptional activity (PubMed:15890677). Acetylates BCL6 which disrupts its ability to recruit histone deacetylases and hinders its transcriptional repressor activity (PubMed:12402037). Participates in CLOCK or NPAS2-regulated rhythmic gene transcription; exhibits a circadian association with CLOCK or NPAS2, correlating with increase in PER1/2 mRNA and histone H3 acetylation on the PER1/2 promoter (PubMed:14645221). Acetylates MTA1 at 'Lys-626' which is essential for its transcriptional coactivator activity (PubMed:16617102). Acetylates XBP1 isoform 2; acetylation increases protein stability of XBP1 isoform 2 and enhances its transcriptional activity (PubMed:20955178). Acetylates PCNA; acetylation promotes removal of chromatin-bound PCNA and its degradation during nucleotide excision repair (NER) (PubMed:24939902). Acetylates MEF2D (PubMed:21030595). Acetylates and stabilizes ZBTB7B protein by antagonizing ubiquitin conjugation and degradation, this mechanism may be involved in CD4/CD8 lineage differentiation (PubMed:20810990). In addition to protein acetyltransferase, can use different acyl-CoA substrates, such as (2E)-butenoyl-CoA (crotonyl-CoA), butanoyl-CoA (butyryl-CoA) or propanoyl-CoA (propionyl-CoA), and is able to mediate protein crotonylation, butyrylation or propionylation, respectively (PubMed:25818647, PubMed:17267393). Acts as a histone crotonyltransferase; crotonylation marks active promoters and enhancers and confers resistance to transcriptional repressors (PubMed:25818647). Histone crotonyltransferase activity is dependent on the concentration of (2E)-butenoyl-CoA (crotonyl-CoA) substrate and such activity is weak when (E)-but-2-enoyl-CoA (crotonyl-CoA) concentration is low (PubMed:25818647). Also acts as a histone butyryltransferase; butyrylation marks active promoters (PubMed:17267393). Functions as a transcriptional coactivator for SMAD4 in the TGF-beta signaling pathway (PubMed:25514493). Acetylates PCK1 and promotes PCK1 anaplerotic activity (PubMed:30193097). {ECO:0000250 UniProtKB:B2RWS6, ECO:0000269 PubMed:10733570, ECO:0000269 PubMed:11430825, ECO:0000269 PubMed:11701890, ECO:0000269 PubMed:12402037, ECO:0000269 PubMed:12586840, ECO:0000269 PubMed:12929931, ECO:0000269 PubMed:14645221, ECO:0000269 PubMed:15186775, ECO:0000269 PubMed:15890677, ECO:0000269 PubMed:16617102, ECO:0000269 PubMed:16762839, ECO:0000269 PubMed:17267393, ECO:0000269 PubMed:18722353, ECO:0000269 PubMed:18995842, ECO:0000269 PubMed:20810990, ECO:0000269 PubMed:21030595, ECO:0000269 PubMed:23415232, ECO:0000269 PubMed:23911289, ECO:0000269 PubMed:23934153, ECO:0000269 PubMed:24939902,

		ECO:0000269 PubMed:25514493, ECO:0000269 PubMed:25818647, ECO:0000269 PubMed:30193097, ECO:0000269 PubMed:8945521, ECO:0000305 PubMed:20955178}.; (Microbial infection) In case of HIV-1 infection, it is recruited by the viral protein Tat. Regulates Tat's transactivating activity and may help inducing chromatin remodeling of proviral genes. Binds to and may be involved in the transforming capacity of the adenovirus E1A protein. {ECO:0000269 PubMed:10545121, ECO:0000269 PubMed:11080476}.
Q12830	BPTF FAC1 FALZ	Histone-binding component of NURF (nucleosome-remodeling factor), a complex which catalyzes ATP-dependent nucleosome sliding and facilitates transcription of chromatin. Specifically recognizes H3 tails trimethylated on 'Lys-4' (H3K4me3), which mark transcription start sites of virtually all active genes. May also regulate transcription through direct binding to DNA or transcription factors.
Q14669	TRIP12 KIAA0045 ULF	E3 ubiquitin-protein ligase involved in ubiquitin fusion degradation (UFD) pathway and regulation of DNA repair. Part of the ubiquitin fusion degradation (UFD) pathway, a process that mediates ubiquitination of protein at their N-terminus, regardless of the presence of lysine residues in target proteins. In normal cells, mediates ubiquitination and degradation of isoform p19ARF/ARF of CDKN2A, a lysine-less tumor suppressor required for p53/TP53 activation under oncogenic stress. In cancer cells, however, isoform p19ARF/ARF and TRIP12 are located in different cell compartments, preventing isoform p19ARF/ARF ubiquitination and degradation. Does not mediate ubiquitination of isoform p16-INK4a of CDKN2A. Also catalyzes ubiquitination of NAE1 and SMARCE1, leading to their degradation. Ubiquitination and degradation of target proteins is regulated by interaction with proteins such as MYC, TRADD or SMARCC1, which disrupt the interaction between TRIP12 and target proteins. Acts as a key regulator of DNA damage response by acting as a suppressor of RNF168, an E3 ubiquitin-protein ligase that promotes accumulation of 'Lys-63'-linked histone H2A and H2AX at DNA damage sites, thereby acting as a guard against excessive spreading of ubiquitinated chromatin at damaged chromosomes. {ECO:0000269 PubMed:18627766, ECO:0000269 PubMed:19028681, ECO:0000269 PubMed:20208519, ECO:0000269 PubMed:20829358, ECO:0000269 PubMed:22884692}.
Q14683	SMC1A DXS423E KIAA0178 SB1.8 SMC1 SMC1L1	Involved in chromosome cohesion during cell cycle and in DNA repair. Central component of cohesin complex. The cohesin complex is required for the cohesion of sister chromatids after DNA replication. The cohesin complex apparently forms a large proteinaceous ring within which sister chromatids can be trapped. At anaphase, the complex is cleaved and dissociates from chromatin, allowing sister chromatids to segregate. The cohesin complex may also play a role in spindle pole assembly during mitosis. Involved in DNA repair via its interaction with BRCA1 and its related phosphorylation by ATM, or via its phosphorylation by ATR. Works as a downstream effector both in the ATM/NBS1 branch and in the ATR/MSH2 branch of S-phase checkpoint. {ECO:0000269 PubMed:11877377}.
Q14839	CHD4	Component of the histone deacetylase NuRD complex which participates in the remodeling of chromatin by deacetylating histones. {ECO:0000269 PubMed:17626165, ECO:0000269 PubMed:9804427}.
Q15596	NCOA2 BHLHE75 SRC2 TIF2	Transcriptional coactivator for steroid receptors and nuclear receptors. Coactivator of the steroid binding domain (AF-2) but not of the modulating N-terminal domain (AF-1). Required with NCOA1 to control energy balance between white and brown adipose tissues. Critical regulator of glucose metabolism regulation, acts as RORA coactivator to specifically modulate G6PC expression. Involved in the positive regulation of the transcriptional activity of the glucocorticoid receptor NR3C1 by sumoylation enhancer RWDD3. Positively regulates the circadian clock by acting as a transcriptional coactivator for the CLOCK-ARNTL/BMAL1 heterodimer (By similarity). {ECO:0000250 UniProtKB:Q61026, ECO:0000269 PubMed:23508108, ECO:0000269 PubMed:9430642}.
Q15652	JMJD1C JHDM2C KIAA1380 TRIP8	Probable histone demethylase that specifically demethylates 'Lys-9' of histone H3, thereby playing a central role in histone code. Demethylation of Lys residue generates formaldehyde and succinate. May be involved in hormone-dependent transcriptional activation, by participating in recruitment to androgen-receptor target genes (By similarity). {ECO:0000250}.
Q5T6F2	UBAP2 KIAA1491	
Q5TAQ9	DCAF8 H326 WDR42A	May function as a substrate receptor for CUL4-DDB1 E3 ubiquitin-protein ligase complex. {ECO:0000269 PubMed:16949367, ECO:0000269 PubMed:16964240}.
Q6P2C8	MED27 CRSP34	Component of the Mediator complex, a coactivator involved in the regulated transcription of

	CRSP8	nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. {ECO:0000269 PubMed:10882111, ECO:0000269 PubMed:9989412}.
Q86YP4	GATAD2A	Transcriptional repressor. Enhances MBD2-mediated repression. Efficient repression requires the presence of GATAD2B. {ECO:0000269 PubMed:12183469, ECO:0000269 PubMed:16415179}.
Q8IX12	CCAR1 CARP1 DIS	Associates with components of the Mediator and p160 coactivator complexes that play a role as intermediaries transducing regulatory signals from upstream transcriptional activator proteins to basal transcription machinery at the core promoter. Recruited to endogenous nuclear receptor target genes in response to the appropriate hormone. Also functions as a p53 coactivator. May thus play an important role in transcriptional regulation (By similarity). May be involved in apoptosis signaling in the presence of the reinoid CD437. Apoptosis induction involves sequestration of 14-3-3 protein(s) and mediated altered expression of multiple cell cycle regulatory genes including MYC, CCNB1 and CDKN1A. Plays a role in cell cycle progression and/or cell proliferation (PubMed:12816952). In association with CALCOCO1 enhances GATA1- and MED1-mediated transcriptional activation from the gamma-globin promoter during erythroid differentiation of K562 erythroleukemia cells (PubMed:24245781). Can act as a both a coactivator and corepressor of AR-mediated transcription. Contributes to chromatin looping and AR transcription complex assembly by stabilizing AR-GATA2 association on chromatin and facilitating MED1 and RNA polymerase II recruitment to AR-binding sites. May play an important role in the growth and tumorigenesis of prostate cancer cells (PubMed:23887938). {ECO:0000250 UniProtKB:Q8CH18, ECO:0000269 PubMed:12816952, ECO:0000269 PubMed:23887938, ECO:0000269 PubMed:24245781}.
Q8IZ40	RCOR2	May act as a component of a corepressor complex that represses transcription. {ECO:0000305}.
Q8IZL2	MAML2 KIAA1819	Acts as a transcriptional coactivator for NOTCH proteins. Has been shown to amplify NOTCH-induced transcription of HES1. Potentiates activation by NOTCH3 and NOTCH4 more efficiently than MAML1 or MAML3. {ECO:0000269 PubMed:12370315, ECO:0000269 PubMed:12386158, ECO:0000269 PubMed:12539049}.
Q8N163	CCAR2 DBC1 KIAA1967	Core component of the DBIRD complex, a multiprotein complex that acts at the interface between core mRNP particles and RNA polymerase II (RNAPII) and integrates transcript elongation with the regulation of alternative splicing: the DBIRD complex affects local transcript elongation rates and alternative splicing of a large set of exons embedded in (A + T)-rich DNA regions. Inhibits SIRT1 deacetylase activity leading to increasing levels of p53/TP53 acetylation and p53-mediated apoptosis. Inhibits SUV39H1 methyltransferase activity. As part of a histone H3-specific methyltransferase complex may mediate ligand-dependent transcriptional activation by nuclear hormone receptors. Plays a critical role in maintaining genomic stability and cellular integrity following UV-induced genotoxic stress. Regulates the circadian expression of the core clock components NR1D1 and ARNTL/BMAL1. Enhances the transcriptional repressor activity of NR1D1 through stabilization of NR1D1 protein levels by preventing its ubiquitination and subsequent degradation (PubMed:18235501, PubMed:18235502, PubMed:19131338, PubMed:19218236, PubMed:22446626, PubMed:23352644, PubMed:23398316). Represses the ligand-dependent transcriptional activation function of ESR2 (PubMed:20074560). Acts as a regulator of PCK1 expression and gluconeogenesis by a mechanism that involves, at least in part, both NR1D1 and SIRT1 (PubMed:24415752). Negatively regulates the deacetylase activity of HDAC3 and can alter its subcellular localization (PubMed:21030595). Positively regulates the beta-catenin pathway (canonical Wnt signaling pathway) and is required for MCC-mediated repression of the beta-catenin pathway (PubMed:24824780). Represses ligand-dependent transcriptional activation function of NR1H2 and NR1H3 and inhibits the interaction of SIRT1 with NR1H3 (PubMed:25661920). Plays an important role in tumor suppression through p53/TP53 regulation; stabilizes p53/TP53 by affecting its interaction with ubiquitin ligase MDM2 (PubMed:25732823). Represses the transcriptional activator activity of BRCA1 (PubMed:20160719). Inhibits SIRT1 in a CHEK2 and PSEM3-dependent manner and inhibits the activity of CHEK2 in vitro (PubMed:25361978). {ECO:0000269 PubMed:18235501, ECO:0000269 PubMed:18235502,

		ECO:0000269 PubMed:19131338, ECO:0000269 PubMed:19218236, ECO:0000269 PubMed:20074560, ECO:0000269 PubMed:20160719, ECO:0000269 PubMed:21030595, ECO:0000269 PubMed:22446626, ECO:0000269 PubMed:23352644, ECO:0000269 PubMed:23398316, ECO:0000269 PubMed:24415752, ECO:0000269 PubMed:24824780, ECO:0000269 PubMed:25361978, ECO:0000269 PubMed:25661920, ECO:0000269 PubMed:25732823}.
Q8N201	INTS1 KIAA1440 UNQ1821/PRO3434	Component of the Integrator (INT) complex, a complex involved in the small nuclear RNAs (snRNA) U1 and U2 transcription and in their 3'-box-dependent processing. The Integrator complex is associated with the C-terminal domain (CTD) of RNA polymerase II largest subunit (POLR2A) and is recruited to the U1 and U2 snRNAs genes (Probable). Mediates recruitment of cytoplasmic dynein to the nuclear envelope, probably as component of the INT complex (PubMed:23904267). {ECO:0000269 PubMed:23904267, ECO:0000305 PubMed:16239144}.
Q8WXI9	GATAD2B KIAA1150	Transcriptional repressor. Enhances MBD2-mediated repression. Efficient repression requires the presence of GATAD2A. Targets MBD3 to discrete loci in the nucleus. May play a role in synapse development. {ECO:0000269 PubMed:12183469, ECO:0000269 PubMed:16415179}.
Q92793	CREBBP CBP	Acetylates histones, giving a specific tag for transcriptional activation. Also acetylates non-histone proteins, like NCOA3 and FOXO1. Binds specifically to phosphorylated CREB and enhances its transcriptional activity toward cAMP-responsive genes. Acts as a coactivator of ALX1. Acts as a circadian transcriptional coactivator which enhances the activity of the circadian transcriptional activators: NPAS2-ARNTL/BMAL1 and CLOCK-ARNTL/BMAL1 heterodimers. Acetylates PCNA; acetylation promotes removal of chromatin-bound PCNA and its degradation during nucleotide excision repair (NER) (PubMed:24939902). Functions as a transcriptional coactivator for SMAD4 in the TGF-beta signaling pathway (PubMed:25514493). {ECO:0000269 PubMed:11154691, ECO:0000269 PubMed:12738767, ECO:0000269 PubMed:12929931, ECO:0000269 PubMed:14645221, ECO:0000269 PubMed:24939902, ECO:0000269 PubMed:25514493, ECO:0000269 PubMed:9707565}.
Q93074	MED12 ARC240 CAGH45 HOPA KIAA0192 TNRC11 TRAP230	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. This subunit may specifically regulate transcription of targets of the Wnt signaling pathway and SHH signaling pathway. {ECO:0000269 PubMed:16565090, ECO:0000269 PubMed:16595664, ECO:0000269 PubMed:17000779}.
Q969G3	SMARCE1 BAF57	Involved in transcriptional activation and repression of select genes by chromatin remodeling (alteration of DNA-nucleosome topology). Component of SWI/SNF chromatin remodeling complexes that carry out key enzymatic activities, changing chromatin structure by altering DNA-histone contacts within a nucleosome in an ATP-dependent manner. Belongs to the neural progenitors-specific chromatin remodeling complex (npBAF complex) and the neuron-specific chromatin remodeling complex (nBAF complex). During neural development a switch from a stem/progenitor to a postmitotic chromatin remodeling mechanism occurs as neurons exit the cell cycle and become committed to their adult state. The transition from proliferating neural stem/progenitor cells to postmitotic neurons requires a switch in subunit composition of the npBAF and nBAF complexes. As neural progenitors exit mitosis and differentiate into neurons, npBAF complexes which contain ACTL6A/BAF53A and PHF10/BAF45A, are exchanged for homologous alternative ACTL6B/BAF53B and DPF1/BAF45B or DPF3/BAF45C subunits in neuron-specific complexes (nBAF). The npBAF complex is essential for the self-renewal/proliferative capacity of the multipotent neural stem cells. The nBAF complex along with CREST plays a role regulating the activity of genes essential for dendrite growth (By similarity). Required for the coactivation of estrogen responsive promoters by SWI/SNF complexes and the SRC/p160 family of histone acetyltransferases (HATs). Also specifically interacts with the CoREST corepressor resulting in repression of neuronal specific gene promoters in non-neuronal cells. {ECO:0000250 UniProtKB:O54941, ECO:0000303 PubMed:12672490, ECO:0000303 PubMed:22952240, ECO:0000303 PubMed:26601204}.
Q969V6	MRTFA KIAA1438 MAL MKL1	Transcription coactivator that associates with the serum response factor (SRF) transcription factor to control expression of genes regulating the cytoskeleton during

		development, morphogenesis and cell migration. The SRF-MRTFA complex activity responds to Rho GTPase-induced changes in cellular globular actin (G-actin) concentration, thereby coupling cytoskeletal gene expression to cytoskeletal dynamics. MRTFA binds G-actin via its RPEL repeats, regulating activity of the MRTFA-SRF complex. Activity is also regulated by filamentous actin (F-actin) in the nucleus. {ECO:0000250 UniProtKB:Q8K4J6}.
Q96F44	TRIM11 RNF92	E3 ubiquitin-protein ligase that promotes the degradation of insoluble ubiquitinated proteins, including insoluble PAX6, poly-Gln repeat expanded HTT and poly-Ala repeat expanded ARX. Mediates PAX6 ubiquitination leading to proteasomal degradation, thereby modulating cortical neurogenesis. May also inhibit PAX6 transcriptional activity, possibly in part by preventing the binding of PAX6 to its consensus sequences. May contribute to the regulation of the intracellular level of HN (humanin) or HN-containing proteins through the proteasomal degradation pathway. Mediates MED15 ubiquitination leading to proteasomal degradation. May contribute to the innate restriction of retroviruses. Upon overexpression, reduces HIV-1 and murine leukemia virus infectivity, by suppressing viral gene expression. Antiviral activity depends on a functional E3 ubiquitin-protein ligase domain. May regulate TRIM5 turnover via the proteasome pathway, thus counteracting the TRIM5-mediated cross-species restriction of retroviral infection at early stages of the retroviral life cycle. {ECO:0000269 PubMed:18248090}.
Q96G25	MED8	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. May play a role as a target recruitment subunit in E3 ubiquitin-protein ligase complexes and thus in ubiquitination and subsequent proteasomal degradation of target proteins.
Q96GM5	SMARCD1 BAF60A	Involved in transcriptional activation and repression of select genes by chromatin remodeling (alteration of DNA-nucleosome topology). Component of SWI/SNF chromatin remodeling complexes that carry out key enzymatic activities, changing chromatin structure by altering DNA-histone contacts within a nucleosome in an ATP-dependent manner (PubMed:8804307, PubMed:29374058). Belongs to the neural progenitors-specific chromatin remodeling complex (npBAF complex) and the neuron-specific chromatin remodeling complex (nBAF complex). During neural development a switch from a stem/progenitor to a postmitotic chromatin remodeling mechanism occurs as neurons exit the cell cycle and become committed to their adult state. The transition from proliferating neural stem/progenitor cells to postmitotic neurons requires a switch in subunit composition of the npBAF and nBAF complexes. As neural progenitors exit mitosis and differentiate into neurons, npBAF complexes which contain ACTL6A/BAF53A and PHF10/BAF45A, are exchanged for homologous alternative ACTL6B/BAF53B and DPF1/BAF45B or DPF3/BAF45C subunits in neuron-specific complexes (nBAF). The npBAF complex is essential for the self-renewal/proliferative capacity of the multipotent neural stem cells. The nBAF complex along with CREST plays a role regulating the activity of genes essential for dendrite growth (By similarity). Has a strong influence on vitamin D-mediated transcriptional activity from an enhancer vitamin D receptor element (VDRE). May be a link between mammalian SWI-SNF-like chromatin remodeling complexes and the vitamin D receptor (VDR) heterodimer (PubMed:14698202). Mediates critical interactions between nuclear receptors and the BRG1/SMARCA4 chromatin-remodeling complex for transactivation (PubMed:12917342). {ECO:0000250 UniProtKB:Q61466, ECO:0000269 PubMed:12917342, ECO:0000269 PubMed:14698202, ECO:0000269 PubMed:29374058, ECO:0000269 PubMed:8804307, ECO:0000303 PubMed:22952240, ECO:0000303 PubMed:26601204}.
Q96HR3	MED30 THRAP6 TRAP25	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. {ECO:0000269 PubMed:11909976, ECO:0000269 PubMed:16595664}.
Q96JM2	ZNF462 KIAA1803	Zinc finger nuclear factor involved in transcription by regulating chromatin structure and organization (PubMed:20219459, PubMed:21570965). Involved in the pluripotency and differentiation of embryonic stem cells by regulating SOX2, POU5F1/OCT4, and NANOG

		(PubMed:21570965). By binding PBX1, prevents the heterodimerization of PBX1 and HOXA9 and their binding to DNA (By similarity). Regulates neuronal development and neural cell differentiation (PubMed:21570965). {ECO:0000250 UniProtKB:B1AWL2, ECO:0000269 PubMed:20219459, ECO:0000269 PubMed:21570965}.
Q96L91	EP400 CAGH32 KIAA1498 KIAA1818 TNRC12	Component of the NuA4 histone acetyltransferase complex which is involved in transcriptional activation of select genes principally by acetylation of nucleosomal histones H4 and H2A. This modification may both alter nucleosome - DNA interactions and promote interaction of the modified histones with other proteins which positively regulate transcription. May be required for transcriptional activation of E2F1 and MYC target genes during cellular proliferation. The NuA4 complex ATPase and helicase activities seem to be, at least in part, contributed by the association of RUVBL1 and RUVBL2 with EP400. May regulate ZNF42 transcription activity. Component of a SWR1-like complex that specifically mediates the removal of histone H2A.Z/H2AFZ from the nucleosome. {ECO:0000269 PubMed:14966270, ECO:0000269 PubMed:24463511}.
Q96RN5	MED15 ARC105 CTG7A PCQAP TIG1 TNRC7	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. Required for cholesterol-dependent gene regulation. Positively regulates the Nodal signaling pathway. {ECO:0000269 PubMed:12167862, ECO:0000269 PubMed:16630888, ECO:0000269 PubMed:16799563}.
Q96ST3	SIN3A	Acts as a transcriptional repressor. Corepressor for REST. Interacts with MXI1 to repress MYC responsive genes and antagonize MYC oncogenic activities. Also interacts with MXD1-MAX heterodimers to repress transcription by tethering SIN3A to DNA. Acts cooperatively with OGT to repress transcription in parallel with histone deacetylation. Involved in the control of the circadian rhythms. Required for the transcriptional repression of circadian target genes, such as PER1, mediated by the large PER complex through histone deacetylation. Cooperates with FOXK1 to regulate cell cycle progression probably by repressing cell cycle inhibitor genes expression (By similarity). Required for cortical neuron differentiation and callosal axon elongation (By similarity). {ECO:0000250 UniProtKB:Q60520, ECO:0000269 PubMed:12150998}.
Q99496	RNF2 BAP1 DING HIPI3 RING1B	E3 ubiquitin-protein ligase that mediates monoubiquitination of 'Lys-119' of histone H2A (H2AK119Ub), thereby playing a central role in histone code and gene regulation (PubMed:15386022, PubMed:16359901, PubMed:25519132, PubMed:21772249, PubMed:25355358, PubMed:26151332). H2AK119Ub gives a specific tag for epigenetic transcriptional repression and participates in X chromosome inactivation of female mammals. May be involved in the initiation of both imprinted and random X inactivation (By similarity). Essential component of a Polycomb group (PcG) multiprotein PRC1-like complex, a complex class required to maintain the transcriptionally repressive state of many genes, including Hox genes, throughout development (PubMed:16359901, PubMed:26151332). PcG PRC1 complex acts via chromatin remodeling and modification of histones, rendering chromatin heritably changed in its expressibility (PubMed:26151332). E3 ubiquitin-protein ligase activity is enhanced by BMI1/PCGF4 (PubMed:21772249). Acts as the main E3 ubiquitin ligase on histone H2A of the PRC1 complex, while RING1 may rather act as a modulator of RNF2/RING2 activity (Probable). Association with the chromosomal DNA is cell-cycle dependent. In resting B- and T-lymphocytes, interaction with AURKB leads to block its activity, thereby maintaining transcription in resting lymphocytes (By similarity). {ECO:0000250 UniProtKB:Q9CQJ4, ECO:0000269 PubMed:11513855, ECO:0000269 PubMed:15386022, ECO:0000269 PubMed:16359901, ECO:0000269 PubMed:16714294, ECO:0000269 PubMed:20696397, ECO:0000269 PubMed:21772249, ECO:0000269 PubMed:25355358, ECO:0000269 PubMed:25519132, ECO:0000269 PubMed:26151332, ECO:0000305}.
Q9NPJ6	MED4 ARC36 DRIP36 VDRIP HSPC126	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors.
Q9P210	CPSF2 CPSF100	Component of the cleavage and polyadenylation specificity factor (CPSF) complex that

	KIAA1367	play a key role in pre-mRNA 3'-end formation, recognizing the AAUAAA signal sequence and interacting with poly(A) polymerase and other factors to bring about cleavage and poly(A) addition. Involved in the histone 3' end pre-mRNA processing. {ECO:0000269 PubMed:14749727, ECO:0000269 PubMed:18688255}.
Q9UHV7	MED13 ARC250 KIAA0593 THRAP1 TRAP240	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. {ECO:0000269 PubMed:16595664}.
Q9ULH7	MRTFB KIAA1243 MKL2	Acts as a transcriptional coactivator of serum response factor (SRF). Required for skeletal myogenic differentiation. {ECO:0000269 PubMed:14565952}.
Q9UPN9	TRIM33 KIAA1113 RFG7 TIF1G	Acts as an E3 ubiquitin-protein ligase. Promotes SMAD4 ubiquitination, nuclear exclusion and degradation via the ubiquitin proteasome pathway. According to PubMed:16751102, does not promote a decrease in the level of endogenous SMAD4. May act as a transcriptional repressor. Inhibits the transcriptional response to TGF-beta/BMP signaling cascade. Plays a role in the control of cell proliferation. Its association with SMAD2 and SMAD3 stimulates erythroid differentiation of hematopoietic stem/progenitor (By similarity). Monoubiquitinates SMAD4 and acts as an inhibitor of SMAD4-dependent TGF-beta/BMP signaling cascade (Monoubiquitination of SMAD4 hampers its ability to form a stable complex with activated SMAD2/3 resulting in inhibition of TGF-beta/BMP signaling cascade). {ECO:0000250, ECO:0000269 PubMed:10022127, ECO:0000269 PubMed:15820681, ECO:0000269 PubMed:16751102, ECO:0000269 PubMed:19135894}.
Q9UQR1	ZNF148 ZBP89	Involved in transcriptional regulation. Represses the transcription of a number of genes including gastrin, stromelysin and enolase. Binds to the G-rich box in the enhancer region of these genes.
Q9Y4A5	TRRAP PAF400	Adapter protein, which is found in various multiprotein chromatin complexes with histone acetyltransferase activity (HAT), which gives a specific tag for epigenetic transcription activation. Component of the NuA4 histone acetyltransferase complex which is responsible for acetylation of nucleosomal histones H4 and H2A. Plays a central role in MYC transcription activation, and also participates in cell transformation by MYC. Required for p53/TP53-, E2F1- and E2F4-mediated transcription activation. Also involved in transcription activation mediated by the adenovirus E1A, a viral oncoprotein that deregulates transcription of key genes. Probably acts by linking transcription factors such as E1A, MYC or E2F1 to HAT complexes such as STAGA thereby allowing transcription activation. Probably not required in the steps following histone acetylation in processes of transcription activation. May be required for the mitotic checkpoint and normal cell cycle progression. Component of a SWR1-like complex that specifically mediates the removal of histone H2A.Z/H2AFZ from the nucleosome. {ECO:0000269 PubMed:11418595, ECO:0000269 PubMed:12138177, ECO:0000269 PubMed:12660246, ECO:0000269 PubMed:12743606, ECO:0000269 PubMed:14966270, ECO:0000269 PubMed:17967892, ECO:0000269 PubMed:24463511, ECO:0000269 PubMed:9708738}.
Q9Y5K5	UCHL5 UCH37 AD-019 CGI-70	Protease that specifically cleaves 'Lys-48'-linked polyubiquitin chains. Deubiquitinating enzyme associated with the 19S regulatory subunit of the 26S proteasome. Putative regulatory component of the INO80 complex; however is inactive in the INO80 complex and is activated by a transient interaction of the INO80 complex with the proteasome via ADRM1. {ECO:0000269 PubMed:16906146, ECO:0000269 PubMed:18922472}.
Q9Y5V3	MAGED1 NRAGE PP2250 PRO2292	Involved in the apoptotic response after nerve growth factor (NGF) binding in neuronal cells. Inhibits cell cycle progression, and facilitates NGFR-mediated apoptosis. May act as a regulator of the function of DLX family members. May enhance ubiquitin ligase activity of RING-type zinc finger-containing E3 ubiquitin-protein ligases. Proposed to act through recruitment and/or stabilization of the Ubl-conjugating enzyme (E2) at the E3:substrate complex. Plays a role in the circadian rhythm regulation. May act as RORA co-regulator, modulating the expression of core clock genes such as ARNTL/BMAL1 and NFIL3, induced, or NR1D1, repressed. {ECO:0000269 PubMed:20864041}.
Q9Y618	NCOR2 CTG26	Transcriptional corepressor (PubMed:20812024). Mediates the transcriptional repression activity of some nuclear receptors by promoting chromatin condensation, thus preventing access of the basal transcription. Isoform 1 and isoform 4 have different affinities for different nuclear receptors. Involved in the regulation BCL6-dependent of the germinal

		center (GC) reactions, mainly through the control of the GC B-cells proliferation and survival. Recruited by ZBTB7A to the androgen response elements/ARE on target genes, negatively regulates androgen receptor signaling and androgen-induced cell proliferation (PubMed:20812024). {ECO:0000269 PubMed:18212045, ECO:0000269 PubMed:20812024, ECO:0000269 PubMed:23911289}.
O15294	OGT	Catalyzes the transfer of a single N-acetylglucosamine from UDP-GlcNAc to a serine or threonine residue in cytoplasmic and nuclear proteins resulting in their modification with a beta-linked N-acetylglucosamine (O-GlcNAc) (PubMed:26678539, PubMed:23103939, PubMed:21240259, PubMed:21285374, PubMed:15361863). Glycosylates a large and diverse number of proteins including histone H2B, AKT1, EZH2, PFKL, KMT2E/MLL5, MAPT/TAU and HCFC1. Can regulate their cellular processes via cross-talk between glycosylation and phosphorylation or by affecting proteolytic processing (PubMed:21285374). Probably by glycosylating KMT2E/MLL5, stabilizes KMT2E/MLL5 by preventing its ubiquitination (PubMed:26678539). Involved in insulin resistance in muscle and adipocyte cells via glycosylating insulin signaling components and inhibiting the 'Thr-308' phosphorylation of AKT1, enhancing IRS1 phosphorylation and attenuating insulin signaling (By similarity). Involved in glycolysis regulation by mediating glycosylation of 6-phosphofructokinase PFKL, inhibiting its activity (PubMed:22923583). Component of a THAP1/THAP3-HCFC1-OGT complex that is required for the regulation of the transcriptional activity of RRM1. Plays a key role in chromatin structure by mediating O-GlcNAcylation of 'Ser-112' of histone H2B: recruited to CpG-rich transcription start sites of active genes via its interaction with TET proteins (TET1, TET2 or TET3) (PubMed:22121020, PubMed:23353889). As part of the NSL complex indirectly involved in acetylation of nucleosomal histone H4 on several lysine residues (PubMed:20018852). O-GlcNAcylation of 'Ser-75' of EZH2 increases its stability, and facilitating the formation of H3K27me3 by the PRC2/EED-EZH2 complex (PubMed:24474760). Regulates circadian oscillation of the clock genes and glucose homeostasis in the liver. Stabilizes clock proteins ARNTL/BMAL1 and CLOCK through O-glycosylation, which prevents their ubiquitination and subsequent degradation. Promotes the CLOCK-ARNTL/BMAL1-mediated transcription of genes in the negative loop of the circadian clock such as PER1/2 and CRY1/2 (PubMed:12150998, PubMed:19451179, PubMed:20018868, PubMed:20200153, PubMed:21285374, PubMed:15361863). O-glycosylates HCFC1 and regulates its proteolytic processing and transcriptional activity (PubMed:21285374, PubMed:28584052, PubMed:28302723). Regulates mitochondrial motility in neurons by mediating glycosylation of TRAK1 (By similarity). Glycosylates HOXA1 (By similarity). {ECO:0000250 UniProtKB:P56558, ECO:0000250 UniProtKB:Q8CGY8, ECO:0000269 PubMed:12150998, ECO:0000269 PubMed:15361863, ECO:0000269 PubMed:19451179, ECO:0000269 PubMed:20018852, ECO:0000269 PubMed:20018868, ECO:0000269 PubMed:20200153, ECO:0000269 PubMed:21240259, ECO:0000269 PubMed:21285374, ECO:0000269 PubMed:22121020, ECO:0000269 PubMed:22923583, ECO:0000269 PubMed:23103939, ECO:0000269 PubMed:23353889, ECO:0000269 PubMed:24474760, ECO:0000269 PubMed:26678539, ECO:0000269 PubMed:28302723, ECO:0000269 PubMed:28584052}; Isoform 2: the mitochondrial isoform (mOGT) is cytotoxic and triggers apoptosis in several cell types including INS1, an insulinoma cell line. {ECO:0000269 PubMed:20824293}.
O60244	MED14 ARC150 CRSP2 CXorf4 DRIP150 EXLM1 RGR1 TRAP170	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. {ECO:0000269 PubMed:15340088, ECO:0000269 PubMed:15625066, ECO:0000269 PubMed:16595664}.
P15884	TCF4 BHLHB19 ITF2 SEF2	Transcription factor that binds to the immunoglobulin enhancer Mu-E5/KE5-motif. Involved in the initiation of neuronal differentiation. Activates transcription by binding to the E box (5'-CANNTG-3'). Binds to the E-box present in the somatostatin receptor 2 initiator element (SSTR2-INR) to activate transcription (By similarity). Preferentially binds to either 5'-ACANNTGT-3' or 5'-CCANNTGG-3'. {ECO:0000250}.
P24863	CCNC	Component of the Mediator complex, a coactivator involved in regulated gene transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and

		<p>serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. Binds to and activates cyclin-dependent kinase CDK8 that phosphorylates the CTD (C-terminal domain) of the large subunit of RNA polymerase II (RNAP II), which may inhibit the formation of a transcription initiation complex. {ECO:0000269 PubMed:16595664, ECO:0000269 PubMed:8700522}.</p>
P50750	CDK9 CDC2L4 TAK	<p>Protein kinase involved in the regulation of transcription. Member of the cyclin-dependent kinase pair (CDK9/cyclin-T) complex, also called positive transcription elongation factor b (P-TEFb), which facilitates the transition from abortive to productive elongation by phosphorylating the CTD (C-terminal domain) of the large subunit of RNA polymerase II (RNAP II) POLR2A, SUPT5H and RDBP. This complex is inactive when in the 7SK snRNP complex form. Phosphorylates EP300, MYOD1, RPB1/POLR2A and AR and the negative elongation factors DSIF and NELF. Regulates cytokine inducible transcription networks by facilitating promoter recognition of target transcription factors (e.g. TNF-inducible RELA/p65 activation and IL-6-inducible STAT3 signaling). Promotes RNA synthesis in genetic programs for cell growth, differentiation and viral pathogenesis. P-TEFb is also involved in cotranscriptional histone modification, mRNA processing and mRNA export. Modulates a complex network of chromatin modifications including histone H2B monoubiquitination (H2Bub1), H3 lysine 4 trimethylation (H3K4me3) and H3K36me3; integrates phosphorylation during transcription with chromatin modifications to control co-transcriptional histone mRNA processing. The CDK9/cyclin-K complex has also a kinase activity towards CTD of RNAP II and can substitute for CDK9/cyclin-T P-TEFb in vitro. Replication stress response protein; the CDK9/cyclin-K complex is required for genome integrity maintenance, by promoting cell cycle recovery from replication arrest and limiting single-stranded DNA amount in response to replication stress, thus reducing the breakdown of stalled replication forks and avoiding DNA damage. In addition, probable function in DNA repair of isoform 2 via interaction with KU70/XRCC6. Promotes cardiac myocyte enlargement. RPB1/POLR2A phosphorylation on 'Ser-2' in CTD activates transcription. AR phosphorylation modulates AR transcription factor promoter selectivity and cell growth. DSIF and NELF phosphorylation promotes transcription by inhibiting their negative effect. The phosphorylation of MYOD1 enhances its transcriptional activity and thus promotes muscle differentiation. {ECO:0000269 PubMed:10393184, ECO:0000269 PubMed:10574912, ECO:0000269 PubMed:10757782, ECO:0000269 PubMed:10912001, ECO:0000269 PubMed:11112772, ECO:0000269 PubMed:11145967, ECO:0000269 PubMed:11575923, ECO:0000269 PubMed:11809800, ECO:0000269 PubMed:11884399, ECO:0000269 PubMed:12037670, ECO:0000269 PubMed:14701750, ECO:0000269 PubMed:15564463, ECO:0000269 PubMed:16109376, ECO:0000269 PubMed:16109377, ECO:0000269 PubMed:17956865, ECO:0000269 PubMed:18362169, ECO:0000269 PubMed:19575011, ECO:0000269 PubMed:19844166, ECO:0000269 PubMed:20081228, ECO:0000269 PubMed:20493174, ECO:0000269 PubMed:20930849, ECO:0000269 PubMed:20980437, ECO:0000269 PubMed:21127351, ECO:0000269 PubMed:9857195}.</p>
Q12857	NFIA KIAA1439	<p>Recognizes and binds the palindromic sequence 5'-TTGCNNNNNGCCAA-3' present in viral and cellular promoters and in the origin of replication of adenovirus type 2. These proteins are individually capable of activating transcription and replication.</p>
Q15528	MED22 SURF5	<p>Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors.</p>
Q15788	NCOA1 BHLHE74 SRC1	<p>Nuclear receptor coactivator that directly binds nuclear receptors and stimulates the transcriptional activities in a hormone-dependent fashion. Involved in the coactivation of different nuclear receptors, such as for steroids (PGR, GR and ER), retinoids (RXRs), thyroid hormone (TRs) and prostanoids (PPARs). Also involved in coactivation mediated by STAT3, STAT5A, STAT5B and STAT6 transcription factors. Displays histone acetyltransferase activity toward H3 and H4; the relevance of such activity remains however unclear. Plays a central role in creating multisubunit coactivator complexes that act via remodeling of chromatin, and possibly acts by participating in both chromatin remodeling and recruitment of general transcription factors. Required with NCOA2 to control energy balance between white and brown adipose tissues. Required for mediating</p>

		steroid hormone response. Isoform 2 has a higher thyroid hormone-dependent transactivation activity than isoform 1 and isoform 3. {ECO:0000269 PubMed:10449719, ECO:0000269 PubMed:12954634, ECO:0000269 PubMed:7481822, ECO:0000269 PubMed:9223281, ECO:0000269 PubMed:9223431, ECO:0000269 PubMed:9296499, ECO:0000269 PubMed:9427757}.
Q2KHR3	QSER1	
Q71SY5	MED25 ACID1 ARC92 PTOV2 TCBAPO758	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. Required for RARA/RXRA-mediated transcription. {ECO:0000269 PubMed:14657022, ECO:0000269 PubMed:14983011, ECO:0000269 PubMed:17641689}.
Q8TAQ2	SMARCC2 BAF170	Involved in transcriptional activation and repression of select genes by chromatin remodeling (alteration of DNA-nucleosome topology). Component of SWI/SNF chromatin remodeling complexes that carry out key enzymatic activities, changing chromatin structure by altering DNA-histone contacts within a nucleosome in an ATP-dependent manner (PubMed:11018012). Can stimulate the ATPase activity of the catalytic subunit of these complexes (PubMed:10078207). May be required for CoREST dependent repression of neuronal specific gene promoters in non-neuronal cells (PubMed:12192000). Belongs to the neural progenitors-specific chromatin remodeling complex (npBAF complex) and the neuron-specific chromatin remodeling complex (nBAF complex). During neural development a switch from a stem/progenitor to a postmitotic chromatin remodeling mechanism occurs as neurons exit the cell cycle and become committed to their adult state. The transition from proliferating neural stem/progenitor cells to postmitotic neurons requires a switch in subunit composition of the npBAF and nBAF complexes. As neural progenitors exit mitosis and differentiate into neurons, npBAF complexes which contain ACTL6A/BAF53A and PHF10/BAF45A, are exchanged for homologous alternative ACTL6B/BAF53B and DPF1/BAF45B or DPF3/BAF45C subunits in neuron-specific complexes (nBAF). The npBAF complex is essential for the self-renewal/proliferative capacity of the multipotent neural stem cells. The nBAF complex along with CREST plays a role regulating the activity of genes essential for dendrite growth (By similarity). Critical regulator of myeloid differentiation, controlling granulocytopenesis and the expression of genes involved in neutrophil granule formation (By similarity). {ECO:0000250 UniProtKB:Q6PDG5, ECO:0000269 PubMed:10078207, ECO:0000269 PubMed:11018012, ECO:0000269 PubMed:12192000, ECO:0000303 PubMed:22952240, ECO:0000303 PubMed:26601204}.
Q96PU8	QKI HKQ	RNA-binding protein that plays a central role in myelination (PubMed:16641098). Binds to the 5'-NACUAAAY-N(1,20)-UAAAY-3' RNA core sequence. Regulates target mRNA stability (PubMed:23630077). In addition, acts by regulating pre-mRNA splicing, mRNA export and protein translation. Required to protect and promote stability of mRNAs such as MBP and CDKN1B. Regulator of oligodendrocyte differentiation and maturation in the brain that may play a role in myelin and oligodendrocyte dysfunction in schizophrenia (PubMed:16641098). Participates in mRNA transport by regulating the nuclear export of MBP mRNA. Also involved in regulation of mRNA splicing of MAG pre-mRNA. Acts as a translational repressor (By similarity). {ECO:0000250 UniProtKB:Q9QYS9, ECO:0000269 PubMed:16641098, ECO:0000269 PubMed:23630077}.
Q99081	TCF12 BHLHB20 HEB HTF4	Transcriptional regulator. Involved in the initiation of neuronal differentiation. Activates transcription by binding to the E box (5'-CANNTG-3').
Q9H204	MED28 EG1 FKSG20	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. May be part of a complex containing NF2/merlin that participates in cellular signaling to the actin cytoskeleton downstream of tyrosine kinase signaling pathways. {ECO:0000269 PubMed:15467741}.
Q9H9B1	EHMT1 EUHMTASE1 GLP KIAA1876 KMT1D	Histone methyltransferase that specifically mono- and dimethylates 'Lys-9' of histone H3 (H3K9me1 and H3K9me2, respectively) in euchromatin. H3K9me represents a specific tag for

		epigenetic transcriptional repression by recruiting HP1 proteins to methylated histones. Also weakly methylates 'Lys-27' of histone H3 (H3K27me). Also required for DNA methylation, the histone methyltransferase activity is not required for DNA methylation, suggesting that these 2 activities function independently. Probably targeted to histone H3 by different DNA-binding proteins like E2F6, MGA, MAX and/or DP1. During G0 phase, it probably contributes to silencing of MYC- and E2F-responsive genes, suggesting a role in G0/G1 transition in cell cycle. In addition to the histone methyltransferase activity, also methylates non-histone proteins: mediates dimethylation of 'Lys-373' of p53/TP53. {ECO:0000269 PubMed:12004135, ECO:0000269 PubMed:20118233}.
Q9P2D1	CHD7 KIAA1416	Probable transcription regulator. Maybe involved in the in 45S precursor rRNA production. {ECO:0000269 PubMed:22646239}.
Q9Y3C7	MED31 SOH1 CGI-125	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors.
Q9Y467	SALL2 KIAA0360 SAL2 ZNF795	Probable transcription factor that plays a role in eye development before, during, and after optic fissure closure. {ECO:0000269 PubMed:24412933}.

Table S2_3. GO terms of MED15 interactors with CC regions.

Category	Term	Count	%	PValue
GOTERM_BP_DIRECT	GO:0006367~transcription initiation from RNA polymerase II promoter	27	21.9512195	7.49E-29
GOTERM_BP_DIRECT	GO:0006351~transcription, DNA-templated	55	44.7154472	5.63E-20
GOTERM_BP_DIRECT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	34	27.6422764	4.48E-18
GOTERM_BP_DIRECT	GO:0006357~regulation of transcription from RNA polymerase II promoter	28	22.7642276	6.16E-18
GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	38	30.8943089	1.62E-17
GOTERM_BP_DIRECT	GO:0030518~intracellular steroid hormone receptor signaling pathway	9	7.31707317	7.05E-14
GOTERM_BP_DIRECT	GO:0044255~cellular lipid metabolic process	10	8.1300813	1.12E-11
GOTERM_BP_DIRECT	GO:0030521~androgen receptor signaling pathway	10	8.1300813	1.12E-11
GOTERM_BP_DIRECT	GO:0045893~positive regulation of transcription, DNA-templated	23	18.699187	1.50E-11
GOTERM_BP_DIRECT	GO:0016575~histone deacetylation	9	7.31707317	1.19E-09
GOTERM_BP_DIRECT	GO:0019827~stem cell population maintenance	9	7.31707317	2.40E-09
GOTERM_BP_DIRECT	GO:0043044~ATP-dependent chromatin remodeling	7	5.69105691	1.13E-08
GOTERM_BP_DIRECT	GO:0006366~transcription from RNA polymerase II promoter	19	15.4471545	2.77E-08
GOTERM_BP_DIRECT	GO:0045892~negative regulation of transcription, DNA-templated	17	13.8211382	5.89E-07
GOTERM_BP_DIRECT	GO:0016567~protein ubiquitination	14	11.3821138	1.92E-06
GOTERM_BP_DIRECT	GO:0006355~regulation of transcription, DNA-templated	29	23.5772358	2.22E-06
GOTERM_BP_DIRECT	GO:0016569~covalent chromatin modification	8	6.50406504	1.74E-05
GOTERM_BP_DIRECT	GO:0006338~chromatin remodeling	7	5.69105691	3.63E-05
GOTERM_BP_DIRECT	GO:0006281~DNA repair	10	8.1300813	5.13E-05
GOTERM_BP_DIRECT	GO:0016573~histone acetylation	5	4.06504065	1.01E-04
GOTERM_BP_DIRECT	GO:0042795~snRNA transcription from RNA polymerase II promoter	6	4.87804878	1.49E-04
GOTERM_BP_DIRECT	GO:1904837~beta-catenin-TCF complex assembly	5	4.06504065	2.55E-04
GOTERM_BP_DIRECT	GO:0006306~DNA methylation	4	3.25203252	7.48E-04
GOTERM_BP_DIRECT	GO:0043967~histone H4 acetylation	4	3.25203252	0.00141584
GOTERM_BP_DIRECT	GO:1901796~regulation of signal transduction by p53 class mediator	6	4.87804878	0.00203481
GOTERM_BP_DIRECT	GO:0006325~chromatin organization	4	3.25203252	0.00365073
GOTERM_BP_DIRECT	GO:2000273~positive regulation of receptor activity	3	2.43902439	0.00381345
GOTERM_BP_DIRECT	GO:0034644~cellular response to UV	4	3.25203252	0.00389747
GOTERM_BP_DIRECT	GO:0006974~cellular response to DNA damage stimulus	7	5.69105691	0.00397089
GOTERM_BP_DIRECT	GO:0032481~positive regulation of type I interferon production	4	3.25203252	0.00590888
GOTERM_BP_DIRECT	GO:0006337~nucleosome disassembly	3	2.43902439	0.00652518
GOTERM_BP_DIRECT	GO:0044849~estrous cycle	3	2.43902439	0.00652518
GOTERM_BP_DIRECT	GO:0048511~rhythmic process	4	3.25203252	0.00692888
GOTERM_BP_DIRECT	GO:0032922~circadian regulation of gene expression	4	3.25203252	0.00804719
GOTERM_BP_DIRECT	GO:0051568~histone H3-K4 methylation	3	2.43902439	0.00898858
GOTERM_BP_DIRECT	GO:1904017~cellular response to Thyroglobulin triiodothyronine	2	1.62601626	0.01436013
GOTERM_BP_DIRECT	GO:0006283~transcription-coupled nucleotide-excision	4	3.25203252	0.01632538

repair

GOTERM_BP_DIRECT	GO:0043392~negative regulation of DNA binding	3	2.43902439	0.01722546
GOTERM_BP_DIRECT	GO:0050434~positive regulation of viral transcription	3	2.43902439	0.01841527
GOTERM_BP_DIRECT	GO:0018076~N-terminal peptidyl-lysine acetylation	2	1.62601626	0.02146331
GOTERM_BP_DIRECT	GO:1901315~negative regulation of histone H2A K63-linked ubiquitination	2	1.62601626	0.02146331
GOTERM_BP_DIRECT	GO:0000398~mRNA splicing, via spliceosome	6	4.87804878	0.0223228
GOTERM_BP_DIRECT	GO:0006368~transcription elongation from RNA polymerase II promoter	4	3.25203252	0.02422782
GOTERM_BP_DIRECT	GO:0006260~DNA replication	5	4.06504065	0.02583551
GOTERM_BP_DIRECT	GO:0006352~DNA-templated transcription, initiation	3	2.43902439	0.02765898
GOTERM_BP_DIRECT	GO:2000780~negative regulation of double-strand break repair	2	1.62601626	0.03551772
GOTERM_BP_DIRECT	GO:0007595~lactation	3	2.43902439	0.03676295
GOTERM_BP_DIRECT	GO:1903799~negative regulation of production of miRNAs involved in gene silencing by miRNA	2	1.62601626	0.04246967
GOTERM_BP_DIRECT	GO:0001701~in utero embryonic development	5	4.06504065	0.04637029
GOTERM_BP_DIRECT	GO:0010467~gene expression	3	2.43902439	0.04684581
GOTERM_BP_DIRECT	GO:0071222~cellular response to lipopolysaccharide	4	3.25203252	0.0482757
GOTERM_BP_DIRECT	GO:0060992~response to fungicide	2	1.62601626	0.04937192
GOTERM_BP_DIRECT	GO:0035563~positive regulation of chromatin binding	2	1.62601626	0.06302873
GOTERM_BP_DIRECT	GO:0006473~protein acetylation	2	1.62601626	0.06302873
GOTERM_BP_DIRECT	GO:0032091~negative regulation of protein binding	3	2.43902439	0.06359261
GOTERM_BP_DIRECT	GO:0045475~locomotor rhythm	2	1.62601626	0.06978399
GOTERM_BP_DIRECT	GO:0033160~positive regulation of protein import into nucleus, translocation	2	1.62601626	0.06978399
GOTERM_BP_DIRECT	GO:0045815~positive regulation of gene expression, epigenetic	3	2.43902439	0.07364007
GOTERM_BP_DIRECT	GO:0006977~DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	3	2.43902439	0.07364007
GOTERM_BP_DIRECT	GO:0033169~histone H3-K9 demethylation	2	1.62601626	0.07649095
GOTERM_BP_DIRECT	GO:0010629~negative regulation of gene expression	4	3.25203252	0.07653644
GOTERM_BP_DIRECT	GO:0006369~termination of RNA polymerase II transcription	3	2.43902439	0.07779312
GOTERM_BP_DIRECT	GO:0035019~somatic stem cell population maintenance	3	2.43902439	0.07989688
GOTERM_BP_DIRECT	GO:0043627~response to estrogen	3	2.43902439	0.07989688
GOTERM_BP_DIRECT	GO:0061158~3'-UTR-mediated mRNA destabilization	2	1.62601626	0.08314994
GOTERM_BP_DIRECT	GO:0006379~mRNA cleavage	2	1.62601626	0.08314994
GOTERM_BP_DIRECT	GO:0061014~positive regulation of mRNA catabolic process	2	1.62601626	0.08314994
GOTERM_BP_DIRECT	GO:0035518~histone H2A monoubiquitination	2	1.62601626	0.08314994
GOTERM_BP_DIRECT	GO:1902166~negative regulation of intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator	2	1.62601626	0.08976132
GOTERM_BP_DIRECT	GO:0016180~snRNA processing	2	1.62601626	0.08976132
GOTERM_BP_DIRECT	GO:0051145~smooth muscle cell differentiation	2	1.62601626	0.08976132
GOTERM_BP_DIRECT	GO:0071347~cellular response to interleukin-1	3	2.43902439	0.09287811
GOTERM_BP_DIRECT	GO:0035729~cellular response to hepatocyte growth factor stimulus	2	1.62601626	0.09632541

Category	Term	Count	%	PValue
GOTERM_MF_DIRECT	GO:0001104~RNA polymerase II transcription cofactor activity	20	16.2601626	2.24E-32
GOTERM_MF_DIRECT	GO:0003713~transcription coactivator activity	28	22.7642276	9.28E-25
GOTERM_MF_DIRECT	GO:0003682~chromatin binding	27	21.9512195	2.06E-18
GOTERM_MF_DIRECT	GO:0005515~protein binding	107	86.9918699	3.78E-18
GOTERM_MF_DIRECT	GO:0030374~ligand-dependent nuclear receptor transcription coactivator activity	14	11.3821138	2.03E-17
GOTERM_MF_DIRECT	GO:0046966~thyroid hormone receptor binding	11	8.94308943	1.56E-15
GOTERM_MF_DIRECT	GO:0042809~vitamin D receptor binding	9	7.31707317	2.96E-14
GOTERM_MF_DIRECT	GO:0044212~transcription regulatory region DNA binding	17	13.8211382	2.01E-12
GOTERM_MF_DIRECT	GO:0003712~transcription cofactor activity	12	9.75609756	2.39E-12
GOTERM_MF_DIRECT	GO:0008134~transcription factor binding	16	13.0081301	1.46E-09
GOTERM_MF_DIRECT	GO:0061630~ubiquitin protein ligase activity	13	10.5691057	8.10E-09
GOTERM_MF_DIRECT	GO:0000978~RNA polymerase II core promoter proximal region sequence-specific DNA binding	16	13.0081301	2.96E-08
GOTERM_MF_DIRECT	GO:0003677~DNA binding	34	27.6422764	3.01E-08
GOTERM_MF_DIRECT	GO:0016922~ligand-dependent nuclear receptor binding	6	4.87804878	2.97E-07
GOTERM_MF_DIRECT	GO:0000980~RNA polymerase II distal enhancer sequence-specific DNA binding	8	6.50406504	3.59E-07
GOTERM_MF_DIRECT	GO:0008270~zinc ion binding	25	20.3252033	1.74E-06
GOTERM_MF_DIRECT	GO:0004407~histone deacetylase activity	6	4.87804878	7.60E-06
GOTERM_MF_DIRECT	GO:0003700~transcription factor activity, sequence-specific DNA binding	21	17.0731707	1.20E-05
GOTERM_MF_DIRECT	GO:0003714~transcription corepressor activity	10	8.1300813	1.36E-05
GOTERM_MF_DIRECT	GO:0035257~nuclear hormone receptor binding	5	4.06504065	1.56E-05
GOTERM_MF_DIRECT	GO:0031492~nucleosomal DNA binding	6	4.87804878	1.74E-05
GOTERM_MF_DIRECT	GO:0004402~histone acetyltransferase activity	6	4.87804878	2.15E-05
GOTERM_MF_DIRECT	GO:0001047~core promoter binding	6	4.87804878	8.75E-05
GOTERM_MF_DIRECT	GO:0001105~RNA polymerase II transcription coactivator activity	5	4.06504065	1.30E-04
GOTERM_MF_DIRECT	GO:0004872~receptor activity	9	7.31707317	1.51E-04
GOTERM_MF_DIRECT	GO:0008013~beta-catenin binding	6	4.87804878	2.83E-04
GOTERM_MF_DIRECT	GO:0030375~thyroid hormone receptor coactivator activity	3	2.43902439	2.93E-04
GOTERM_MF_DIRECT	GO:0044822~poly(A) RNA binding	20	16.2601626	3.37E-04
GOTERM_MF_DIRECT	GO:0003723~RNA binding	13	10.5691057	4.83E-04
GOTERM_MF_DIRECT	GO:0001103~RNA polymerase II repressing transcription factor binding	4	3.25203252	8.83E-04
GOTERM_MF_DIRECT	GO:0043565~sequence-specific DNA binding	12	9.75609756	0.0010873
GOTERM_MF_DIRECT	GO:0002039~p53 binding	5	4.06504065	0.00127774
GOTERM_MF_DIRECT	GO:0001102~RNA polymerase II activating transcription factor binding	4	3.25203252	0.00240705
GOTERM_MF_DIRECT	GO:0042974~retinoic acid receptor binding	3	2.43902439	0.00310553
GOTERM_MF_DIRECT	GO:0001085~RNA polymerase II transcription factor binding	4	3.25203252	0.00441884
GOTERM_MF_DIRECT	GO:0047485~protein N-terminus binding	5	4.06504065	0.00491439
GOTERM_MF_DIRECT	GO:0042826~histone deacetylase binding	5	4.06504065	0.00586789

GOTERM_MF_DIRECT	GO:0042800~histone methyltransferase activity (H3-K4 specific)	3	2.43902439	0.00700317
GOTERM_MF_DIRECT	GO:0001078~transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding	5	4.06504065	0.0078805
GOTERM_MF_DIRECT	GO:0031490~chromatin DNA binding	4	3.25203252	0.00795094
GOTERM_MF_DIRECT	GO:0000976~transcription regulatory region sequence-specific DNA binding	4	3.25203252	0.00795094
GOTERM_MF_DIRECT	GO:0033613~activating transcription factor binding	3	2.43902439	0.01038115
GOTERM_MF_DIRECT	GO:0034212~peptide N-acetyltransferase activity	2	1.62601626	0.01404941
GOTERM_MF_DIRECT	GO:0019904~protein domain specific binding	6	4.87804878	0.01595281
GOTERM_MF_DIRECT	GO:0004386~helicase activity	4	3.25203252	0.02220255
GOTERM_MF_DIRECT	GO:0001077~transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	6	4.87804878	0.02588126
GOTERM_MF_DIRECT	GO:0036033~mediator complex binding	2	1.62601626	0.02790307
GOTERM_MF_DIRECT	GO:0030331~estrogen receptor binding	3	2.43902439	0.02794981
GOTERM_MF_DIRECT	GO:0003899~DNA-directed RNA polymerase activity	3	2.43902439	0.02936885
GOTERM_MF_DIRECT	GO:0019899~enzyme binding	7	5.69105691	0.03058652
GOTERM_MF_DIRECT	GO:0018024~histone-lysine N-methyltransferase activity	3	2.43902439	0.03081612
GOTERM_MF_DIRECT	GO:0050681~androgen receptor binding	3	2.43902439	0.03229116
GOTERM_MF_DIRECT	GO:0043426~MRF binding	2	1.62601626	0.03475735
GOTERM_MF_DIRECT	GO:0001012~RNA polymerase II regulatory region DNA binding	2	1.62601626	0.03475735
GOTERM_MF_DIRECT	GO:0016874~ligase activity	6	4.87804878	0.04237885
GOTERM_MF_DIRECT	GO:0001055~RNA polymerase II activity	2	1.62601626	0.06831642
GOTERM_MF_DIRECT	GO:0003725~double-stranded RNA binding	3	2.43902439	0.06891556
GOTERM_MF_DIRECT	GO:0003684~damaged DNA binding	3	2.43902439	0.07289174
GOTERM_MF_DIRECT	GO:0032454~histone demethylase activity (H3-K9 specific)	2	1.62601626	0.07488807
GOTERM_MF_DIRECT	GO:0016407~acetyltransferase activity	2	1.62601626	0.08789379

Category	Term	Count	%	PValue
GOTERM_CC_DIRECT	GO:0016592~mediator complex	25	20.3252033	1.87E-45
GOTERM_CC_DIRECT	GO:0005654~nucleoplasm	84	68.2926829	5.61E-40
GOTERM_CC_DIRECT	GO:0005634~nucleus	93	75.6097561	5.75E-26
GOTERM_CC_DIRECT	GO:0000790~nuclear chromatin	16	13.0081301	2.77E-12
GOTERM_CC_DIRECT	GO:0005667~transcription factor complex	15	12.195122	3.91E-11
GOTERM_CC_DIRECT	GO:0070847~core mediator complex	6	4.87804878	7.09E-11
GOTERM_CC_DIRECT	GO:0000151~ubiquitin ligase complex	12	9.75609756	1.40E-10
GOTERM_CC_DIRECT	GO:0017053~transcriptional repressor complex	8	6.50406504	7.44E-08
GOTERM_CC_DIRECT	GO:0016581~NuRD complex	6	4.87804878	9.49E-08
GOTERM_CC_DIRECT	GO:0000785~chromatin	7	5.69105691	2.80E-05
GOTERM_CC_DIRECT	GO:0000118~histone deacetylase complex	5	4.06504065	9.24E-05
GOTERM_CC_DIRECT	GO:0000123~histone acetyltransferase complex	4	3.25203252	4.59E-04
GOTERM_CC_DIRECT	GO:0071339~MLL1 complex	4	3.25203252	9.20E-04
GOTERM_CC_DIRECT	GO:0016604~nuclear body	4	3.25203252	0.00134696
GOTERM_CC_DIRECT	GO:0048188~Set1C/COMPASS complex	3	2.43902439	0.00231231
GOTERM_CC_DIRECT	GO:0071564~npBAF complex	3	2.43902439	0.00276275
GOTERM_CC_DIRECT	GO:0071565~nBAF complex	3	2.43902439	0.00377632
GOTERM_CC_DIRECT	GO:0016514~SWI/SNF complex	3	2.43902439	0.00433844
GOTERM_CC_DIRECT	GO:0005730~nucleolus	14	11.3821138	0.00454333
GOTERM_CC_DIRECT	GO:0035097~histone methyltransferase complex	3	2.43902439	0.01009795
GOTERM_CC_DIRECT	GO:0000784~nuclear chromosome, telomeric region	5	4.06504065	0.01105775
GOTERM_CC_DIRECT	GO:0005719~nuclear euchromatin	3	2.43902439	0.0137698
GOTERM_CC_DIRECT	GO:0031519~PcG protein complex	3	2.43902439	0.0137698
GOTERM_CC_DIRECT	GO:0005737~cytoplasm	47	38.2113821	0.01617858
GOTERM_CC_DIRECT	GO:0016020~membrane	24	19.5121951	0.01776757
GOTERM_CC_DIRECT	GO:0043234~protein complex	8	6.50406504	0.02009723
GOTERM_CC_DIRECT	GO:0005876~spindle microtubule	3	2.43902439	0.03450218
GOTERM_CC_DIRECT	GO:0000125~PCAF complex	2	1.62601626	0.03918743
GOTERM_CC_DIRECT	GO:0016589~NURF complex	2	1.62601626	0.04556893
GOTERM_CC_DIRECT	GO:0000812~Swr1 complex	2	1.62601626	0.05820612
GOTERM_CC_DIRECT	GO:0044666~MLL3/4 complex	2	1.62601626	0.05820612
GOTERM_CC_DIRECT	GO:0030529~intracellular ribonucleoprotein complex	4	3.25203252	0.06206279
GOTERM_CC_DIRECT	GO:0016580~Sin3 complex	2	1.62601626	0.08298477
GOTERM_CC_DIRECT	GO:0032039~integrator complex	2	1.62601626	0.08298477
GOTERM_CC_DIRECT	GO:0030914~STAGA complex	2	1.62601626	0.08907772
GOTERM_CC_DIRECT	GO:0005847~mRNA cleavage and polyadenylation specificity factor complex	2	1.62601626	0.08907772
GOTERM_CC_DIRECT	GO:0033276~transcription factor TFTC complex	2	1.62601626	0.08907772

GOTERM_CC_DIRECT	GO:0000932~cytoplasmic mRNA processing body	3	2.43902439	0.09485405
GOTERM_CC_DIRECT	GO:0031011~Ino80 complex	2	1.62601626	0.09513053