

Article

# Functional Segments on Intrinsically Disordered Regions in Disease-Related Proteins

Hiroto Anbo, Masaya Sato, Atsushi Okoshi and Satoshi Fukuchi \*

Department of Life Science and Informatics, Faculty of Engineering, Maebashi Institute of Technology, 460-1, Kamisadori, Maebashi, Gunma 371-0816, Japan; koume8@icloud.com (H.A.); mail\_address107@icloud.com (M.S.); m1461008@maebashi-it.ac.jp (A.O.)

\* Correspondence: sfukuchi@maebashi-it.ac.jp

Received: 22 January 2019; Accepted: 25 February 2019; Published: 5 March 2019



**Abstract:** One of the unique characteristics of intrinsically disordered proteins (IPDs) is the existence of functional segments in intrinsically disordered regions (IDRs). A typical function of these segments is binding to partner molecules, such as proteins and DNAs. These segments play important roles in signaling pathways and transcriptional regulation. We conducted bioinformatics analysis to search these functional segments based on IDR predictions and database annotations. We found more than a thousand potential functional IDR segments in disease-related proteins. Large fractions of proteins related to cancers, congenital disorders, digestive system diseases, and reproductive system diseases have these functional IDRs. Some proteins in nervous system diseases have long functional segments in IDRs. The detailed analysis of some of these regions showed that the functional segments are located on experimentally verified IDRs. The proteins with functional IDR segments generally tend to come and go between the cytoplasm and the nucleus. Proteins involved in multiple diseases tend to have more protein-protein interactors, suggesting that hub proteins in the protein-protein interaction networks can have multiple impacts on human diseases.

**Keywords:** intrinsically disordered regions; functional segments; disease-related proteins; protein-protein interaction; subcellular location

## 1. Introduction

Intrinsically disordered proteins (IDPs) are proteins that do not adopt unique three-dimensional structures under physiological conditions [1–3]. They are fully or partially disordered and are abundant among eukaryotic proteins [4–6]. One of the unique features of IDPs is their ability to bind to binding partners. The regions performing such binding are generally short segments ranging from several residues to tens of residues and can adopt local two-dimensional structures in association with this binding. This has been referred to as the coupled folding and binding mechanism. These interactions are transient, specific, and low-affinity. Through this mechanism, intrinsically disordered regions (IDRs) play crucial roles in many biological processes, such as signal transduction and transcriptional regulation [1–3,7].

The importance of IDPs in human diseases has been reported [8,9]. Intrinsically disordered proteins are found in high concentrations in plaques and brain deposits in neurodegenerative patients, and mutations in IDRs can increase aggregation propensity. Intrinsically disordered proteins such as  $\alpha$ -synuclein, the amyloid  $\beta$  peptide, and huntingtin have been directly linked to diseases such as Alzheimer's, Parkinson's, and Huntington's diseases [10–16]. It has been shown that many IDPs participate in cell signaling and cancer-associated proteins [7]. Breast cancer type 1 susceptibility protein (BRCA1) is one of the most typical IDPs, with a long central region of 1480 residues shown to be disordered by nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy [17].

This long IDR has many binding segments for proteins such as p53, retinoblastoma protein, and the oncogenes c-Myc and JunB. p53 is a transcription factor that has IDRs in its N- and C-terminus. These IDRs have binding sites for many partner proteins. Among these, the interaction between p53 and E3 ubiquitin-protein ligase Mdm2 (MDM2) has been given much attention in cancer research, as p53 can induce apoptosis to suppress tumor progression [18,19]. Bioinformatics work has also shown that IDRs are rich in proteins involved in cancer, neurodegenerative diseases, cardiovascular diseases, and diabetes [9,20]. Intrinsically disordered proteins have gained attention as drug targets. Inhibitors targeting IDR and globular domain interactions have been developed for the interaction between Bcl-xL and BAK [21,22], MDM2 and p53 [23], interleukin (IL)-2 receptor  $\alpha$  and IL-2 [24,25], XIAP and Smac [26,27], and CBP and  $\beta$ -catenin [28].

As shown above, protein-protein interactions (PPIs) occurring on IDRs have high potential as drug targets. The IDP databases, IDEAL [29,30] and DisProt [31], have 913 and 803 proteins, and IDEAL has collected 559 protein-binding segments on IDRs called “protean segments (ProSs)” in the database. Protean segments are defined as sequences with experimental evidence of being both disordered in an isolated state and ordered in a binding state. In contrast, several tools for predicting such binding regions have been developed to suggest that there are more than 100,000 protein-binding segments in the human proteome [32]. Considering this prediction, our knowledge on IDR-mediated interactions is still limited because the number of ProSs with experimental evidence of ordered and disordered states is only about 600. However, we have a lot of PPI data accumulating and several computer programs to predict IDRs. The performance of IDR predictions has reached a standard for practical use, and PPI annotations found in predicted IDRs can be considered protein-binding segments in IDRs. In this study, we combined the annotations of the UniProt database and IDR predictions to find these possible protein-binding regions on IDRs and analyzed these regions in the context of human disease.

## 2. Materials and Methods

We selected human proteins from the Swiss-Prot section of the UniProt XML file [33] from UniProt release 2018\_07. We extracted the feature (FT) section information. A single FT section has a feature type, a description, and a location, and all of them were extracted. The IDEAL database provides binding segments in IDRs as “ProSs”. Protean segments are mostly short segments consisting of less than 30 amino acid residues to which more than 80% of ProSs belong. Thus, we selected for feature information shorter than 30 residues. Next, we picked binding-associated features from the selected features. By manually surveying feature descriptions, the features “region of interest” and “mutagenesis site” are found to contain binding-related features. Adding to these two features, “short sequence motif” also contains functional segments in IDRs. Out of the selected features of “region of interest”, the features having the terms “interact”, “bind”, or “motif” in their description were selected. Of the “mutagenesis site” features, those with “interact” or “bind” were selected. From the selected features of “mutagenesis site”, those having “no” or “not” were discarded. We wrote some in-house scripts to find the selected features located in IDRs. A feature region found in an IDR was defined as a possible ProS, hereafter referred to as a pProS.

We used three predictors for IDR prediction. MobiDB [31] provided predicted IDRs for several proteome datasets. We downloaded the human proteome dataset and used MobiDB-lite [34] predictions for predicted IDRs. MobiDB-lite uses eight different predictors, three variants of ESpritz (DisProt, NMR, X-ray), two variants of IUPred (long, short), two variants of DisEMBL (465, hot loops), and GlobPlot. MobiDB-lite uses the results of these predictors to combine them into a consensus result, where at least five out of eight methods must define a residue as disordered. Thus, the predicted IDRs reflect different types of predictions at one time. DISOPRED3 [35] is an extension of the previous program, DISOPRED2, to improve predictions of long IDRs. In order to achieve this goal, DISOPRED3 uses a neural network-based predictors trained on a dataset rich in long IDRs. The training was done on a position-specific scoring matrix (PSSM) generated by PSI-BLAST. This program is one of the

benchmark IDR predictors, which was one of the top ranked predictors in CASP10 [36]. DICHOT is the predictor combining the homology-based domain assignments and the support vector machine learning. First, it conducts a PSI-BLAST search against the Protein Data Bank (PDB) to mask structural regions, and then the unmasked regions are judged by using the support vector machine-based predictor trained on a multiple alignment of homologs. This predictor divides an entire amino acid chain into structural domains and IDRs, which is unique compared to other predictors. We selected regions where any of the two predictors predicted an IDR.

Human disease information was obtained from the KEGG database [37]. KEGG has a collection of disease entries called KEGG DISEASE. KEGG DISEASE provides Search Disease, which is a mapping tool against disease genes accumulated in KEGG DISEASE entries. Thus, the Search Disease tool provides information on which disease a protein is involved in. The mapped disease-related proteins were divided into pProS-containing proteins and non-pProS proteins, according to the existence or absence of pProSs.

We used the UniProt annotations of subcellular locations for protein localizations. We counted terms that appeared in the annotations by the disease-related proteins and all human proteins. We defined those proteins as only having the annotation of nucleus as nuclear protein, those proteins only having that of cytoplasm as cytoplasm protein, those only having that of membrane as membrane protein, and those having both of the annotations of cytoplasm and nucleus as cytoplasm and nuclear protein. The annotations of other locations were discarded because of shortages of appearance. The disease-related proteins were further divided into pProS-containing proteins and non-pProS proteins, and the ratios of each of the terms were obtained for the disease-related proteins and all human proteins. The logarithms of the ratios of two ratios were used in analysis.

The brief outline of the procedure can be found in Supplementary Figure S1.

### 3. Results

The UniProt database has 20,410 human proteins, and 3378 proteins (16.6%) were assigned human diseases by the Search Disease tool, as shown in Table 1. In the human UniProt and the disease-related proteins, 29,145 and 18,450 regions of the feature annotation shorter than 30 residues were found, respectively. From these regions, we selected the feature annotations of “region of interest”, “mutagenesis site”, and “short sequence motif” to pick up pProSs. Out of 3378 disease related proteins, 402 proteins (11.9%) had pProSs. Out of 18,450 feature annotations found in the disease related proteins, 8.3%, 3.4%, and 24.4% were found in the predicted IDRs for “region of interest”, “mutagenesis site”, and “short sequence motif”, respectively. Thus, the regions of these annotations are defined as pProSs in this study.

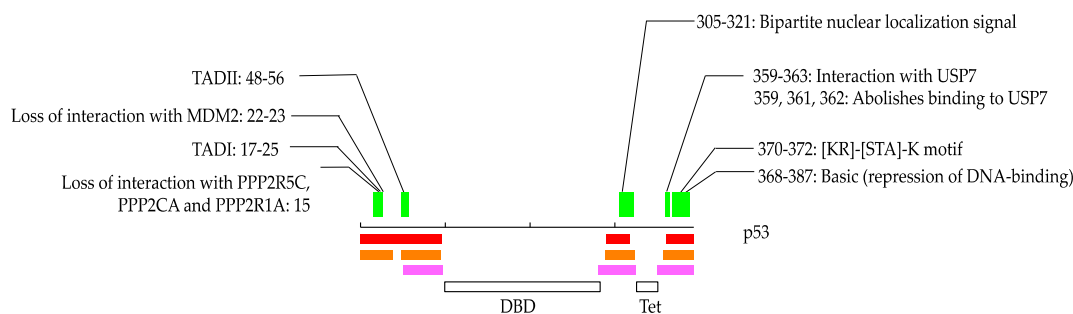
**Table 1.** Statistics of the UniProt annotations.

	All Proteins	Disease-Related	pProS	pProS (%)
No. proteins	20,410	3378	402	11.9
No. annotations shorter than 30 residues	29,145	18,450	1124	6.1
“Region of interest”	4646	2656	220	8.3
“Mutagenesis site”	21,269	14,056	479	3.4
“Short sequence motif”	3230	1740	425	24.4

pProS: Possible protean segment.

We illustrate how these annotations occur in IDRs in Figure 1. In this study, some of the proteins are stored in the IDEAL database [38]. In such cases, we denoted IDEAL identifiers for reference. p53 is one of the typical IDPs and has relatively short IDRs in the N- and C-terminus in addition to between the DNA-binding domain and the tetramerization domain (IDEAL: IID00015). p53 has three annotations from residues 15 to 25, one annotation from residues 48 to 56, one annotation from 305 to 321, four annotations from 359 to 363, and one annotation from 370 to 372 and 368 to 387. Among

them, “TAD I”, “TAD II”, “bipartite nuclear localization signal” and “[KR]-[STA]-K motif” are from “short sequence motif”, “interaction with USP7” and “basic” are from “region of interest”, and “loss of interactions with MDM2”, “loss of interaction with PPP2R5C, PPP2CA and PPP2R1A”, and “abolishes binding to USP7” are from “mutagenesis site”. As shown in Figure 1, some of these annotations overlap each other. We counted these overlapped annotations separately in the statistics in this study because they can be different annotations even if their regions are overlapped. For example, although “TAD I” overlaps “loss of interaction with MDM2” and “loss of interaction with PPP2R5C, PPP2CA and PPP2R1A”, they describe different phenomena. Thus, we would like readers to note that the statistics in this study contain such multiple counts in some cases.



**Figure 1.** An example of possible protean segment (pProS) definition, illustrated by p53. The black line in the middle represents the amino acid chain, and the intrinsically disordered regions (IDR) predictions are presented below. Pink, orange, and red represent the results by MobiDB-lite, DISOPRED3, and DICHOT, respectively. Regions where any of the two methods predict IDR are defined as IDRs. The green bars represent pProSs, and the annotations defining pProS are shown above with residue numbers of the annotations. DBD: DNA-binding domain; Tet: Tetramerization domain.

Most of the annotations we picked in this analysis occur at one time in the dataset. As seen in the example of p53, the annotations of interest in this study are in forms such as “interaction with protein A” or “loss of interaction with protein B”, etc., which can be assigned on a single or a small numbers of proteins binding onto a specific protein. However, some of the annotations appeared multiple times. These frequently appearing annotations in pProSs are listed in Supplementary Table S1. They contain targeting sequences, such as a nuclear localization signal (NLS). Retinoblastoma-associated protein has an NLS from residues 858 to 881. This region has been described as disordered in the isolated state [39], though the binding structure with importin has been solved (IDEAL: IID00017) [40]. Another class of frequently found annotations is segments binding upon promiscuously interacting domains, such as the SH3 domain and the PDZ domain. The SH3 domain mediates PPIs via a short ambiguous peptide motif to assemble cell regulatory systems [41] and is annotated as ProS in the IDEAL database (IDEAL: IIDE00256). The PDZ domain is a scaffold protein that forms protein complexes in signaling pathways or cell trafficking [42]. The binding segment is disordered in an isolated state (IDEAL: IID9005) [43]. We also frequently found the LXXLL motif, which exists in co-repressors or co-activators of nuclear receptors. The LXXLL motif in peroxisome proliferator-activated receptor gamma co-activator 1  $\alpha$  has been reported as disordered, and the binding structure with steroid hormone receptor has been elucidated (IDEAL: IID00103) [44,45].

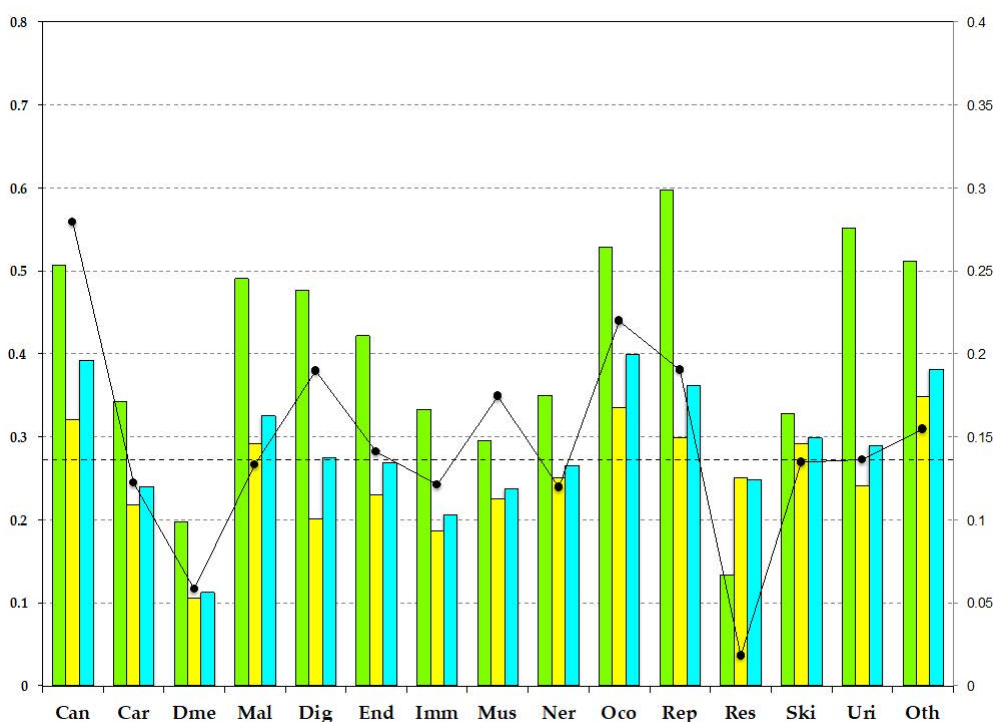
The disease-related proteins are classified into 15 categories by KEGG DISEASE, which are: cancers (Can); cardiovascular diseases (Car); congenital disorders of metabolism (Dme); congenital malformations (Mal); digestive system diseases (Dig); endocrine and metabolic diseases (End); immune system diseases (Imm); musculoskeletal diseases (Mus); nervous system diseases (Ner); other congenital disorders (Oco); reproductive system diseases (Rep); respiratory diseases (Res); skin diseases (Ski); urinary system diseases (Uri); and other diseases (Oth). These categories have subcategories into which these diseases are classified. The details of the disease classifications can be found on the web page for KEGG DISEASE [46].

We classified disease-related proteins into the disease categories and showed the statistics in Table 2. The rank of the numbers of unique pProSs is as follows, in descending order: “congenital malformations”; “nervous system diseases”; “cancers”; and “cardiovascular diseases”. The numbers depend on the number of proteins in each of the categories, as seen in Table 2. However, the category possessing many proteins does not necessarily contain many pProSs. For example, the “congenital disorders of metabolism” contains 687 proteins, though only 40 pProSs were identified. The values of protein coverage represent the ratio of the numbers of pProS-containing proteins to the numbers of proteins in a category. “cancers” shows the top coverage, followed by “congenital disorders,” “digestive system diseases,” and “reproductive system diseases.” The average values of annotations represent the values of the number of pProSs in a category divided by the number of pProS-containing proteins, wherein most of the values are greater than 2. As shown in Figure 1, some annotations are found in different regions, and some annotations overlap each other. The average annotation values here contain both cases. The annotations of “mutagenesis site” tend to overlap with other annotations, where 70% of them coincide with other annotations. This overlap of “mutagenesis site” may enlarge the average annotation values. The total amount of proteins with pProS over all disease categories differs from the number of pProS-containing proteins shown in Table 1. This is because some proteins were assigned to multiple diseases.

**Table 2.** Statistics of pProS by the disease category.

Category	No. Unique pProSs	No. Proteins with pProS	No. Proteins	Protein Coverage (%)	Average Annotations
Cancers	147	57	204	27.9	2.6
Cardiovascular diseases	93	41	335	12.2	2.3
Congenital disorders of metabolism	53	40	687	5.8	1.3
Congenital malformations	242	111	832	13.3	2.2
Digestive system diseases	32	15	79	19.0	2.1
Endocrine and metabolic diseases	63	30	213	14.1	2.1
Immune system diseases	56	31	256	12.1	1.8
Musculoskeletal diseases	69	26	149	17.4	2.7
Nervous system diseases	199	95	795	11.9	2.1
Other congenital disorders	39	20	91	22.0	2.0
Reproductive system diseases	21	12	63	19.0	1.8
Respiratory diseases	1	1	55	1.8	1.0
Skin diseases	22	14	104	13.5	1.6
Urinary system diseases	33	9	66	13.6	3.7
Other diseases	68	30	194	15.5	2.3

Figure 2 shows the IDR ratios by the disease categories. The IDR ratio of all disease proteins (27.0%) is similar to that of the UniProt human proteins (28.7%, the dashed line in Figure 2). However, the IDR fractions of each category differ, where “cancers,” “congenital malformations,” “other congenital disorders,” “reproductive system diseases,” and “other diseases” are over-represented. These IDR-rich categories also have high fractions of IDRs in the pProS-containing proteins. Although the IDR fractions of pProS-containing proteins correlate with the protein coverage (Table 2), a high fraction of IDRs does not necessarily mean high protein coverage. For example, although “urinary diseases” shows a very high IDR fraction in their pProS-containing proteins, the protein coverage is not high. This means that the pProS-containing proteins in this category contain high fractions of IDR, but a number of non-pProS-containing proteins also belong to this category. On the other hand, “musculoskeletal diseases” has low IDR fractions in both pProS-containing proteins and all proteins. However, the protein coverage is relatively high, suggesting that many proteins with relatively short IDR fractions have functional regions in such IDRs. Iakoucheva et al. [7] showed that cancer related proteins have high fractions of IDRs, and the result of the present analysis confirms this trend. Moreover, the current results suggest that cancer-related proteins have a lot of pProSs compared to the proteins in the other categories.



**Figure 2.** The IDR ratios by disease category. The green, yellow, and blue bars represent the IDR fractions of the pProS-containing proteins, the non-pProS proteins, and the total proteins in each of the disease categories, respectively. The measure on the left axis represents the IDR fractions. The black line with dots represents the protein coverage found in Table 2. The measure on the right axis represents the protein coverage. The dashed line represents the IDR ratio of the human proteome. Can: Cancers; Car: Cardiovascular diseases; Dme: Congenital disorders of metabolism; Mal: Congenital malformations; Dig: Digestive system diseases; End: Endocrine and metabolic diseases; Imm: Immune system diseases; Mus: Musculoskeletal diseases; Ner: Nervous system diseases; Oco: Other congenital disorders; Rep: Reproductive system diseases; Res: Respiratory diseases; Ski: Skin diseases; Uri: Urinary system diseases; Oth: Other diseases.

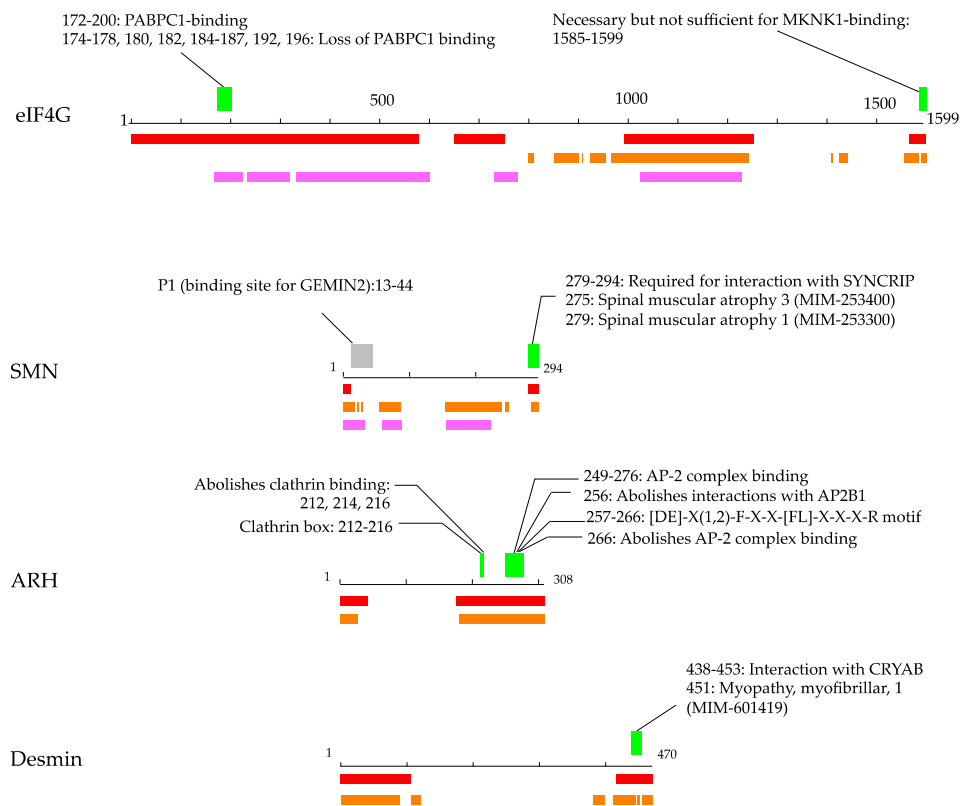
Table 3 shows the top-ranked proteins in terms of the numbers of pProS residues. Although we counted overlapped annotations separately in Table 2, we counted “pProS residues” without redundancy. For example, when there are annotations of “from 360 to 365” and “from 360 to 362”, we counted six for this region. “Cancers” and p53 appear again, and the proteins belonging to “nervous system diseases” are listed frequently. As shown in Figure 2, “cancers” has the largest IDR fraction and the top ranked protein coverage. Thus, it is natural for the proteins in “cancers” to contain proteins with a large fraction of pProS residues. On the other hand, “nervous system diseases” has an IDR fraction smaller than that of the UniProt human proteins, and the protein coverage is low. Therefore, the proteins in “nervous system diseases” do not generally contain a large fraction of IDR and pProS. However, some of them have long pProSs, even in relatively short IDRs. On the other hand, “nervous system diseases” has an IDR fraction smaller than that of the UniProt human proteins. Raychaudhuri et al. reported similar results, where in the nervous system disease related proteins, Huntington’s and Alzheimer’s disease proteins have large fractions of IDRs, and Parkinson’s disease proteins do not [20], suggesting the similarity of the average IDR fraction of these nervous disease proteins to that of UniProt human proteins. Therefore, the proteins in “nervous system diseases” do not generally contain a large fraction of IDR and pProS. However, some of them have long pProSs, even in relatively short IDRs. The example of “nervous system diseases” suggests that even if the IDR fraction and protein coverage of pProS-containing proteins are low, some proteins in a disease category can have considerable lengths of pProSs.

**Table 3.** The list of the proteins with long pProSs.

Protein Name	UniProt Accession	pProS Residues	No. Disease	Disease Category	Disease
DNA excision repair protein ERCC-6	Q03468	63	4	Ner Ner Mal Ski	Age-related macular degeneration Cockayne syndrome Disorders of nucleotide excision repair Ultra violet-sensitive syndrome
Cellular tumor antigen p53	P04637	61	46		*
E3 ubiquitin-protein ligase RNF168	Q8IYW5	55	1	Imm	RIDDLE syndrome
CD2-associated protein	Q9Y5K6	50	1	Uri	Focal segmental glomerulosclerosis
Synaptic functional regulator FMR1	Q06787	47	3	Rep	Premature ovarian failure
Low-density lipoprotein receptor-related protein 2	P98164	44	1	Mal	Donnai-Barrow syndrome
Eukaryotic translation initiation factor 4 gamma 1	Q04637	44	1	Ner	Parkinson disease
DNA (cytosine-5)-methyltransferase 1	P26358	41	1	Ner	Hereditary sensory and autonomic neuropathy
Period circadian protein homolog 2	O15055	40	1	Ner	Familial advanced sleep phase syndrome
Latent-transforming growth factor $\beta$ -binding protein 2	Q14767	40	1	Ner	Primary congenital glaucoma
Low-density lipoprotein receptor-related protein 6	O75581	39	3	Can Car Dig	Breast cancer Coronary artery disease Tooth agenesis
DNA damage-inducible transcript 3 protein	P35638	39	1	Can	Myxoid liposarcoma
KN motif and ankyrin repeat domain-containing protein 1	Q14678	39	1	Ner	Spastic quadriplegic cerebral palsy
Retinoic acid-induced protein 1	Q7Z5J4	36	1	Oco	Smith-Magenis syndrome
Histone-lysine N-methyltransferase 2D	O14686	35	2	Can Mal	Follicular lymphoma Kabuki syndrome
FYN-binding protein 1	O15117	35	1	Car	Thrombocytopenia
Low-density lipoprotein receptor adapter protein 1	Q5SW96	33	1	Dme	Familial autosomal recessive hypercholesterolemia
Catenin $\beta$ -1	P35222	32	8	Can Can Can Can Can Can Oth Ski	Thyroid cancer Medulloblastoma Endometrial cancer Colorectal cancer Gastric cancer Hepatocellular carcinoma Autosomal dominant mental retardation Pilomatricoma
Sp110 nuclear body protein	Q9HB58	31	1	Dig	Hepatic veno-occlusive disease with immunodeficiency
Low-density lipoprotein receptor-related protein 5	O75197	30	6	Mal Mal Mal Mus Mus Ner	Osteopetrosis Worth type autosomal dominant osteosclerosis Osteoporosis-pseudoglioma syndrome Hyperostosis corticalis generalisata Osteoporosis Familial exudative vitreoretinopathy
LEM domain-containing protein 2	Q8NC56	30	1	Ner	Cataract
Single-stranded DNA cytosine deaminase	Q9GZX7	30	1	Imm	Hyper IgM syndromes, autosomal recessive type

\* The list of diseases involving p53 is found in Supplementary Table S2. Can: Cancers; Car: Cardiovascular diseases; Dme: Congenital disorders of metabolism; Mal: Congenital malformations; Dig: Digestive system diseases; End: Endocrine and metabolic diseases; Imm: Immune system diseases; Mus: Musculoskeletal diseases; Ner: Nervous system diseases; Oco: Other congenital disorders; Rep: Reproductive system diseases; Res: Respiratory diseases; Ski: Skin diseases; Uri: Urinary system diseases; Oth: Other diseases.

In Figure 3 and the following sections, we illustrate how annotations can be found in IDRs of the disease-related proteins.



**Figure 3.** Examples of proteins with pProS. The black line in the middle represents the amino acid chain, and the IDR predictions are presented below. Pink, orange, and red represent the results by MobiDB-lite, DISOPRED3, and DICHOT, respectively. The green bars represent pProSs, and the annotations defining pProS are shown above with the residue numbers of the annotations. The gray bar in the example of survival of motor neuron (SMN) represents the regions of a pseudo-pProS, which was not taken as pProS because the region of the annotation is longer than 30. In the case of low-density lipoprotein receptor adaptor protein 1 (ARH) and desmin, MobiDB-lite does not predict any IDRs. The scale of eIF4G1 (eukaryotic translation initiation factor 4 gamma 1) differs from other three.

### 3.1. Eukaryotic Translation Initiation Factor 4 Gamma 1

Parkinson's disease is a progressive neurodegenerative movement disorder caused by the death of dopaminergic neurons in the substantia nigra pars compacta (KEGG: H00057). Although deleterious mutations in  $\alpha$ -synuclein (OMIM-163890) [47], leucine-rich repeat kinase 2 (OMIM-609007) [48], vesicular protein sorting 35 (OMIM-601501) [49], parkin (OMIM-602544) [50], PTEN induced putative kinase 1 (OMIM-608309) [51], and DJ-1 (OMIM-602533) [52] have been found in multi-incident families with parkinsonism, mutations in the translation initiator, eukaryotic translation initiation factor 4 gamma 1 (eIF4G1, UniProt: Q04637), have also been reported [53]. The eIF4G1 is a component of the protein complex eIF4F, which is involved in the recognition of the mRNA cap, ATP-dependent unwinding of the 5'-terminal secondary structure and recruitment of mRNA to the ribosome. eIF4G1 has long IDRs in its 1599 amino acid residues. In the IDRs, the region from residues 172 to 200 possesses polyadenylate-binding protein 1 (PABPC) binding ability [54], and the region from residues 1585 to 1599 has an annotation of "necessary for binding of MAP kinase-interacting serine/threonine-protein kinase 1 (MKNK1)". Although IDRs in this study were defined by computer predictions, the region of PABPC binding is reported to be unfolded in the isolated state [54]. The eIF4G1 has been found to have five Parkinson's disease-associated mutations: Ala502Val; Gly686Cys; Ser1164Arg; Arg1197Trp;



and Arg1205His. Out of five mutations, Ala502Val and Arg1205His appear to disrupt eIF4E or eIF3E binding and share haplotypes consistent with ancestral founders [53]. These two mutations reside on IDRs, suggesting an association of binding sites on IDRs with Parkinson's disease.

### 3.2. Survival of Motor Neuron Protein

Spinal muscular atrophy (SMA) is a neuromuscular disease characterized by degeneration of motor neurons, resulting in progressive muscle atrophy and paralysis. The most common form of SMA is caused by mutation of the survival of motor neuron (SMN, UniProt: Q16637) protein (KEGG: H00455). Survival of motor neuron (SMN) protein is in a complex with several proteins, including Gemin2, Gemin3, and Gemin4, and plays important roles in small nuclear ribonucleoprotein (snRNP) biogenesis and pre-mRNA splicing. The SMN protein has two highly homologous genes, SMN1 and SMN2, which lie within the telomeric and centromeric halves of a large inverted repeat on chromosome 5q. The coding sequence of SMN2 differs from that of SMN1 by a single nucleotide in exon 7 (840C-T) [55–57]. Thirty-eight patients with SMA have a homozygous deletion of exon 7 of the SMN1 gene, and the deletion is associated with homozygous deletion of exon 8 in 31 of 34 patients [58]. Exon 7 and 8 cover the region from residues 242 to 294, which is predicted to be an IDR. The region from residues 252 to 281 contains the YG-box for forming helical oligomers (PDB: 4gli) [59]. In the oligomers, the C-terminal region from residues 281 to 297 is disordered [59]. This C-terminal region contains a binding region for heterogeneous nuclear ribonucleoprotein (hnRNP) Q, and the most common SMN mutant found in SMA patients is defective in its interactions with snRNPs [60]. Thus, SMN protein provides an example that a defect of IDR binding ability causes a serious disorder. SMN has another IDR in its N-terminal; this region also contains binding sites for GEMIN2, though the length of the annotation and IDR coverage do not meet the threshold of pProS in this study.

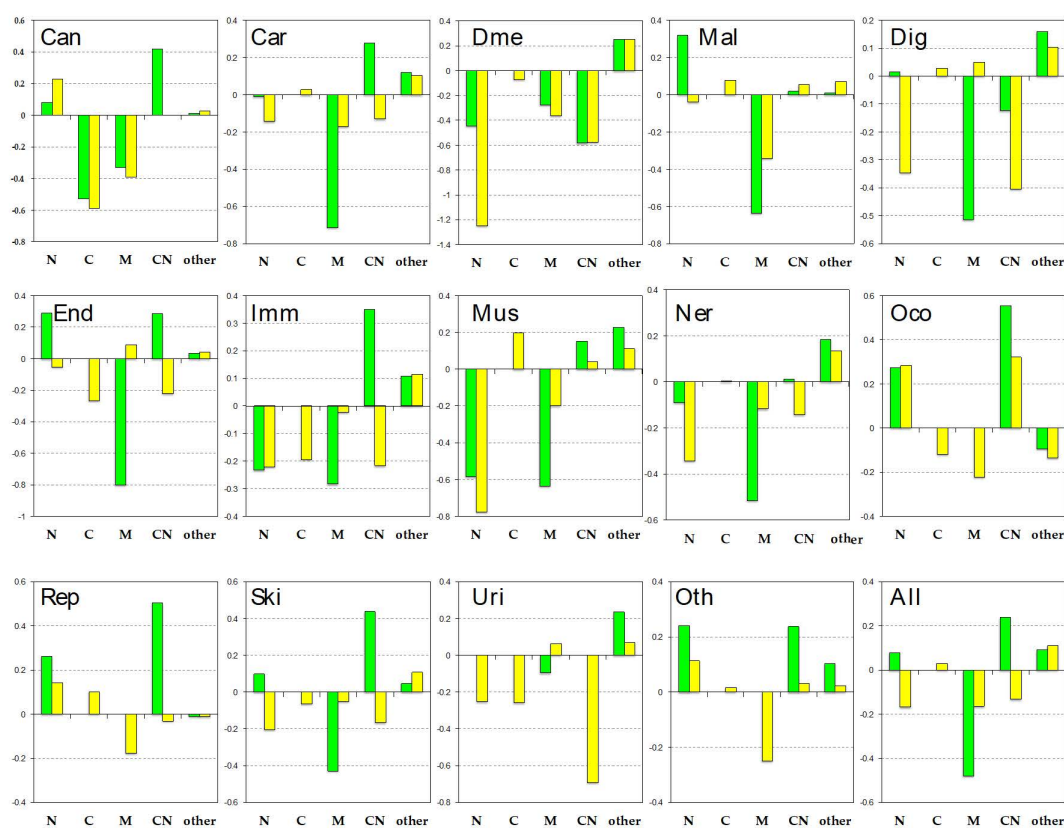
### 3.3. Low-Density Lipoprotein Receptor Adaptor Protein 1

A known cause of hypercholesterolemia is deficiency of low-density lipoprotein receptors (LDLR) or apolipoprotein B [61]. Recently, mutations in the low-density lipoprotein receptor adaptor protein 1 (ARH, UniProt: Q5SW96) were found to cause the autosomal recessive form of hypercholesterolemia [62]. Low-density lipoprotein receptors-mediated endocytosis in the liver is the primary pathway for clearance of circulating LDL, to prevent LDL accumulation. The ARH protein is an adaptor protein required for efficient endocytosis of the LDL receptor. The ARH protein can interact with the internalization sequence in the cytoplasmic tail of LDLR, and the N-terminal region of the clathrin heavy chain, a component of a polyhedral lattice on the transport vesicles in the clathrin-mediated membrane traffic. It also binds upon the beta subunit of adaptor protein complex 2 (AP-2), which is a vesicle coat component involved in cargo selection and vesicle formation. The ARH protein is predicted to have two IDRs in its N- and C-terminus. The IDRs in the C-terminal end have two functional regions: clathrin binding and AP-2 complex binding, which are crucial interactions for LDLR-mediated endocytosis. Although the IDRs are defined only by the computer predictions, the clathrin-binding region of auxilin has been found to be disordered [63], and the tertiary structure of the AP-2 complex binding region of ARH ([DE]-X(1,2)-F-X-X-[FL]-X-X-X-R motif) in the complex with AP-2  $\beta$  subunit has been solved (PDB: 2g30) [64]. Supplementary Figure S2 shows the structure of this segment binding upon AP-2  $\beta$  subunit, together with a typical ProS structure of nuclear receptor co-activator 1 (IDEAL: IID50084) binding upon nuclear receptor subfamily 1 group I. Nuclear receptor co-activators or co-repressors have the LXXLL motif in their binding site for nuclear receptors. This motif is one of the annotations frequently found to be a pProS in this study (Supplementary Table S1) and has been shown to be disordered in the isolated state, as mentioned above. The AP-2 binding region shows a similar  $\alpha$ -helix structure attached to the globular structure of the partner protein.

### 3.4. Desmin

Desminopathy belongs to a genetically heterogeneous group of disorders named myofibrillar myopathy, caused by mutations in desmin,  $\alpha$ B-crystallin, myotilin, Z-band alternatively spliced PDZ-containing protein, filamin, or Bcl-2-associated athanogene. Desmin (Uniprot: P17661) is the main intermediate filament (IF) protein. It interacts with other proteins to support myofibrils at the level of the Z-disc and forms a continuous cytoskeletal IF network [65]. Desmin has two predicted IDRs in its N- and C-terminus, and the binding region for  $\alpha$ B-crystallin (CRYAB) is located in the C-terminal IDR [66]. Several mutations in this region have been reported to cause severe disturbance of filament formation. The desmin mutations Thr442Ile, Arg454Trp, and Ser460Ile reveal a severe disturbance of filament formation competence and filament interactions [66–68], and Thr453Ile exhibits significantly delayed filament assembly kinetics [69,70]. These sites locate on one of the predicted IDRs.

We analyzed the disease-related proteins from subcellular localizations in Figure 4. The location distributions of all disease-related proteins (“all” shown in the right bottom panel) showed that pProS-containing proteins are rich in cytoplasm and nuclear (CN) proteins and nucleus (N) proteins. The distribution of the non-pProS proteins shows a shortage of CN and N proteins. Ota et al. [71] pointed out the IDR richness in the mobile proteins shuttling from cytoplasm to nucleus and vice versa. Cytoplasm and nuclear proteins localize both in the cytoplasm and nucleus, and thus, the results of this study also showed that these CN proteins have functional regions on their IDRs. Although most of the localizations by the disease category showed this trend, some of them show some divergence. “cancers”, “cardiovascular diseases”, “endocrine and metabolic diseases”, “other congenital disorders”, “reproductive system diseases”, “skin diseases”, and “other diseases” showed similarity to that of all disease-related proteins. On the other hand, “congenital disorders of metabolism”, “digestive system diseases”, “nervous system diseases”, and “urinary system diseases” showed another trend. The pProS-containing proteins in these categories do not have many CN proteins and N proteins but have proteins in other locations. It has been indicated that IDRs are found in large numbers in nuclear proteins, such as transcription factors and proteins in the signaling pathways [4,6]. The CN proteins and N proteins are those proteins under such categories because stimulus from outside of the cell is received by receptors and must be transmitted to the nucleus. The system therefore needs to have mobile proteins from cytoplasm to the nucleus. Additionally, in order to regulate such systems, information flow from the nucleus to the cytoplasm is required. Then, the pProS-containing proteins under the first trend can be the typical IDPs previously reported. The other locations contain “endoplasmic reticulum”, “cell projection”, and “cell junction”, etc. The proteins under the second trend suggest that IDPs not belonging to typical IDPs, such as transcription factors, can have functional sites in their IDRs and are involved in human diseases. For example, ARH as shown in Figure 3 has functions in the endocytosis in cytoplasm, and its IDRs have the key regions associated with hypercholesterolemia.



**Figure 4.** Subcellular localizations by disease category. The bars represent the degree of over-representation in each of the location categories, where green represents pProS-containing proteins, and yellow represents non-pProS proteins (see also Materials and Methods). N: Nuclear; C: Cytoplasm; M: Membrane; CN: Cytoplasm and nuclear; Can: Cancers; Car: Cardiovascular diseases; Dme: Congenital disorders of metabolism; Mal: Congenital malformations; Dig: Digestive system diseases; End: Endocrine and metabolic diseases; Imm: Immune system diseases; Mus: Musculoskeletal diseases; Ner: Nervous system diseases; Oco: Other congenital disorders; Rep: Reproductive system diseases; Ski: Skin diseases; Uri: Urinary system diseases; Oth: Other diseases; All: All disease-related proteins.

#### 4. Discussion

This study is based on IDR predictions and database annotations. We employed consensus methods to define IDRs using three predictors. Because the MobiDB-lite program contains eight different prediction models, we substantially used ten predictors to make consensus IDRs. We found experimental evidence for predicted IDRs in the detailed analysis of some proteins. For example, the annotations frequently found in the IDRs (Supplementary Table S1) were found in the regions experimentally verified as IDRs, and the examples of eIF4G1, SMN protein, and ARH also showed that some parts of the predicted IDRs have been experimentally verified to be IDRs (Figure 3). Due to these examples, we were convinced that the IDRs in this study were promising. Even with the conservative IDR definition, we found more than 1000 pProSs. If the condition to select IDRs was relaxed, more candidates for functional regions in IDRs could be obtained.

Functional segments on IDRs have been collected as the databases such as SLiM/ELM [72]. A proteome-wide analysis combined with high-throughput sequencing data showed that disease-related mutations were enriched in SLiMs on IDRs and occurred more frequently at functionally important residues in SLiMs [73]. Most of these motifs are annotated as “short sequence motif” in UniProt. In this sense, the analysis in this study converts not only these motifs but also potential functional segments annotated as “mutagenesis site” and “region of interest”. The ratio of the numbers of “short sequence motif” annotations to those of the other two is one-eighth (Table 1). This result

suggests that knowledge of functional segments on IDRs can exist other than the information stored in the motif databases.

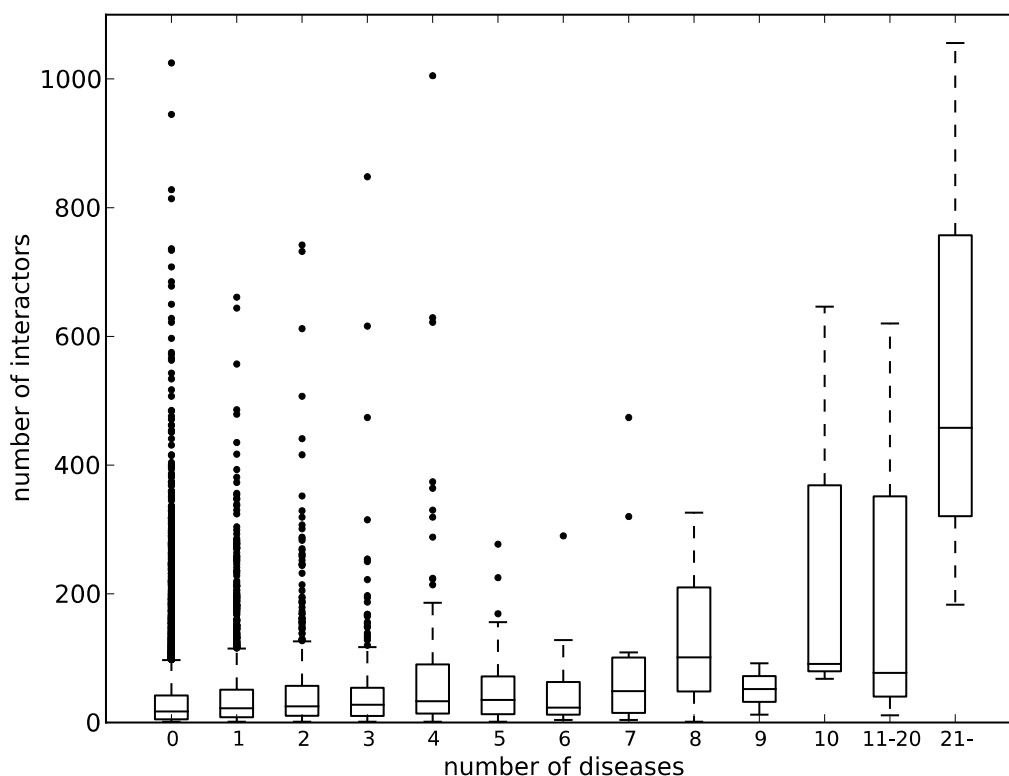
We searched potential binding segments in IDRs by referring to the UniProt annotations. The UniProt annotations are human-curated and highly reliable. Some of the information on a protein, however, does not appear in the FT section of the UniProt annotation. For example, the N-terminal IDR of p53 has 11 binding partners, and the C-terminal IDR has 15 partners solved in PDB structures (IDEAL: IID00015). The UniProt annotations in these IDRs are concise, as shown in Figure 2, although the links to the PDB entries are described in the “cross reference” section in UniProt. We picked eIF4G1 as an example in Figure 3. In this example, we found the pProS of the PABPC1-binding region in the UniProt annotation. However, this region also binds upon rotavirus nonstructural protein 3 [74], and this information does not appear in the feature table. In this sense, this study does not cover all of the knowledge of potential functional regions in IDRs. Thus, detailed analysis of each of the proteins would provide more information about functional segments on IDRs in the disease-related proteins.

In fact, we found potential functional sites from another information in UniProt. As shown in Figure 3, some of pProSs coincide with mutation sites associated with human disease. UniProt describes these associations as the link of mutation sites to the OMIM database, which are not directly linked with the feature annotations used in this study. Although we did not use such mutation site information to define pProS, we found considerable numbers of mutation sites in the predicted IDRs. We found 1611 mutation sites in the predicted IDRs of 572 proteins. Among them, UniProt describes links to the OMIM database for 919 mutation sites in 359 proteins. The list of these mutation sites can be found in Supplementary Table S3. Although it is not clear whether these mutation sites are binding sites for other molecules, these sites may possibly be regarded as functional regions in IDRs. Some of the readers may be interested in other model organisms. We briefly surveyed some of the represented model organisms, as shown in Supplementary Table S4. Although the results are preliminary, considerable numbers of annotations were found in the predicted IDRs for the mouse, rat, *Arabidopsis thaliana*, and yeast. These organisms have been long used for model organisms, and knowledge verified by experiments has been accumulated. Thus, for such model organisms, the strategy of the present study can be applied.

When we simply count the numbers of pProSs in the disease categories, the statistics were different from Table 2. The statistics are shown in Supplementary Table S5. The “pProS counts” represent how many times pProSs occur in each of the categories. “Cancers” were top ranked, followed by “congenital malformation” and “nervous system diseases”. The large numbers of proteins can account for the large numbers of “pProS counts” of “nervous system diseases” and “congenital malformation”. “Cancers”, however, does not have as many proteins as these two categories, in spite of many “pProS counts.” This is due to the multiple involvement of a protein in different diseases. When a protein having a pProS and is assigned to two diseases, we counted two for the pProS count. Thus, when a protein is assigned to multiple diseases in a disease category, redundant counts in the disease category occur. Supplementary Table S6 lists the top 10 redundant proteins in terms of multiple disease annotations. p53 Has the maximum multiple disease annotations, and most of these diseases are in the category “cancers”. The multiple annotations on the other proteins tend to be also redundant in “cancers”. The pProS redundancy in Supplementary Table S5, which is the ratio of the numbers of unique pProS to the numbers of pProS counts, shows the trend of this multiple disease association. “cancers” has a remarkably high value of pProS redundancy, suggesting that pProS-containing proteins in “cancers” tend to associate with several kinds of cancers.

p53 is involved in an extremely large number of diseases and has been also known as a hub protein that has great numbers of binding partners in PPI networks. In fact, the BioGRID database [75] lists 1056 interactors for p53, which is ranked 11th in terms of the number of PPIs in the human proteome. Then, we looked at the relation between the number of diseases assigned by KEGG DISEASE and the number of protein interactors. Figure 5 shows the relationship between the two. The correlation coefficients between the mean and median values of the numbers of interactors to the numbers of

assigned diseases were 0.86 and 0.87. The distribution of the numbers of assigned diseases likely follows the power law. Therefore, most of the samples are found in the bins of the left side of the chart, and only small numbers of samples are found in the right side of the chart, as suggested by the long boxes. Even within the data of less than 10 assigned diseases, the correlation coefficient of the mean value is 0.73. It can therefore be said that there is generally a correlation between the numbers of interactors and the numbers of diseases involved. It has been pointed out that hub proteins in PPI networks are rich in IDRs [76,77]. If we accept the relationship between the number of protein interactors and the number of diseases involved, these hub proteins, namely IDPs, must be associated with many human diseases.



**Figure 5.** The correlation between the number of protein–protein interactions and the number of diseases involved. The horizontal axis represents the number of diseases, and the vertical one represents the number of interactors. A box and a pair of whiskers represent quartiles, and the line in the middle of the box represents the median. The dots represent outliers.

Post-translational modifications (PTM) are an important modification on proteins to regulate their function. In particular, phosphorylation has been known to occur preferentially on IDRs [78] to regulate signaling pathways and other many biological processes. We surveyed phosphorylation sites of the pProS-containing proteins by using the UniProt annotations and found that 876 phosphorylation sites coincide with pProS regions in 290 proteins. Most of these phosphorylation related pProSs have the annotations relating to protein binding, suggesting that the PPI via these pProSs may be regulated by phosphorylation. We also checked links of these regions to the OMIM database and found that about 30 phosphorylation sites have the links to the OMIM database. Although the number of the direct link to OMIM is short, a lot of phosphorylation sites can regulate PPIs, and the defect of these phosphorylation related pProSs may make an impact on PPI networks.

We analyzed localizations of disease-related proteins to find the richness of the mobile proteins coming and going between the cytoplasm and nucleus in pProS-containing proteins. Recently, proteins and protein–nucleic acid mixtures can undergo liquid–liquid phase separation (LLPS) to form non-membrane organelles or lipid droplets, and have several important biological functions [79–82].

Intrinsically disordered regions have been shown to play important roles in LLPS [83–85]. Some of the non-membrane organelles appeared in the UniProt annotations of subcellular localization. Supplementary Figure S3 shows the distributions of these non-membrane organelles found in disease-related proteins. It is clear that the pProS-containing proteins are over-represented, while the non-pProS proteins are not, with the exception of “lipid droplet” and “P-body”. Although the numbers of annotations for these non-membrane organelles are small, these results suggest that IDPs play an important role in forming such organelles and have functional segments on their IDRs.

## 5. Conclusions

We conducted bioinformatics analyses to survey our knowledge on functional segments in IDRs from the perspective of disease-related proteins. We found more than a thousand annotations in the predicted IDRs, and considerable fractions of the disease-related proteins contained functional segments on IDRs. The detailed analysis on some of the examples showed that the pProSs found in this study were located in the experimentally verified IDRs and could directly associate with the diseases. Hub proteins in the PPI network tend to be involved in many human diseases, and some of the pProS-containing proteins are embedded in non-membrane organelles. We should note that the statistics in this study convey only current research, and more than 100,000 functional segments are expected to exist in IDRs. However, even with limited information, this study showed the power of database search for retrieving knowledge of functional segments in IDRs. The complete lists of pProSs can be found in Supplementary Table S7.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2218-273X/9/3/88/s1>, Figure S1: The brief outline of the procedure to obtain pProSs; Figure S2: A structural example of a functional fragment found in an experimentally verified IDR; Figure S3: The non-membrane organelles found in the annotations of the disease-related proteins; Table S1: pProSs frequently found in the disease-related proteins; Table S2: The list of disease involving p53; Table S3: The disease associated mutation sites found in the predicted IDRs; Table S4: The numbers of annotations found on IDRs in the other model organisms; Table S5: The redundancy of pProS by the disease categories; Table S6: Proteins involved in multiple diseases; Table S7: The list of pProSs found in this study.

**Author Contributions:** Investigation, H.A., M.S., A.O. and S.F.; writing—original draft preparation, H.A. and S.F.; supervision, S.F.

**Funding:** This research received no external funding.

**Acknowledgments:** We are grateful to the IDEAL development team for their effort to develop and maintain the IDEAL database.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dunker, A.K.; Brown, C.J.; Lawson, J.D.; Iakoucheva, L.M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573–6582. [[CrossRef](#)] [[PubMed](#)]
2. Dyson, H.J.; Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208. [[CrossRef](#)] [[PubMed](#)]
3. Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **2005**, *579*, 3346–3354. [[CrossRef](#)] [[PubMed](#)]
4. Fukuchi, S.; Hosoda, K.; Homma, K.; Gojobori, T.; Nishikawa, K. Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC Struct. Biol.* **2011**, *11*, 29. [[CrossRef](#)] [[PubMed](#)]
5. Minezaki, Y.; Homma, K.; Kinjo, A.R.; Nishikawa, K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* **2006**, *359*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
6. Ward, J.J.; Sodhi, J.S.; McGuffin, L.J.; Buxton, B.F.; Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004**, *337*, 635–645. [[CrossRef](#)] [[PubMed](#)]
7. Iakoucheva, L.M.; Brown, C.J.; Lawson, J.D.; Obradovic, Z.; Dunker, A.K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573–584. [[CrossRef](#)]

8. Babu, M.M.; van der Lee, R.; de Groot, N.S.; Gsponer, J. Intrinsically disordered proteins: Regulation and disease. *Curr. Opin. Struct. Biol.* **2011**, *21*, 432–440. [[CrossRef](#)] [[PubMed](#)]
9. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Ann. Rev. Biophys.* **2008**, *37*, 215–246. [[CrossRef](#)] [[PubMed](#)]
10. Wu, H.; Fuxreiter, M. The structure and dynamics of higher-order assemblies: Amyloids, signalosomes, and granules. *Cell* **2016**, *165*, 1055–1066. [[CrossRef](#)] [[PubMed](#)]
11. Toretsky, J.A.; Wright, P.E. Assemblages: Functional units formed by cellular phase separation. *J. Cell Biol.* **2014**, *206*, 579–588. [[CrossRef](#)] [[PubMed](#)]
12. Calabretta, S.; Richard, S. Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem. Sci.* **2015**, *40*, 662–672. [[CrossRef](#)] [[PubMed](#)]
13. Aguzzi, A.; Altmeyer, M. Phase separation: Linking cellular compartmentalization to disease. *Trends Cell Biol.* **2016**, *26*, 547–558. [[CrossRef](#)] [[PubMed](#)]
14. Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A.P.; Kim, H.J.; Mittag, T.; Taylor, J.P. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **2015**, *163*, 123–133. [[CrossRef](#)] [[PubMed](#)]
15. Knowles, T.P.; Vendruscolo, M.; Dobson, C.M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 384–396. [[CrossRef](#)] [[PubMed](#)]
16. Chiti, F.; Dobson, C.M. Protein misfolding, functional amyloid, and human disease. *Ann. Rev. Biochem.* **2006**, *75*, 333–366. [[CrossRef](#)] [[PubMed](#)]
17. Mark, W.Y.; Liao, J.C.; Lu, Y.; Ayed, A.; Laister, R.; Szymczyna, B.; Chakrabarty, A.; Arrowsmith, C.H. Characterization of segments from the central region of BRCA1: An intrinsically disordered scaffold for multiple protein–protein and protein–DNA interactions? *J. Mol. Biol.* **2005**, *345*, 275–287. [[CrossRef](#)] [[PubMed](#)]
18. Chene, P. The role of tetramerization in p53 function. *Oncogene* **2001**, *20*, 2611–2617. [[CrossRef](#)] [[PubMed](#)]
19. Oliner, J.D.; Pietenpol, J.A.; Thiagalingam, S.; Gyuris, J.; Kinzler, K.W.; Vogelstein, B. Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. *Nature* **1993**, *362*, 857–860. [[CrossRef](#)] [[PubMed](#)]
20. Raychaudhuri, S.; Dey, S.; Bhattacharyya, N.P.; Mukhopadhyay, D. The role of intrinsically unstructured proteins in neurodegenerative diseases. *PLoS ONE* **2009**, *4*, e5566. [[CrossRef](#)] [[PubMed](#)]
21. Qian, J.; Voorbach, M.J.; Huth, J.R.; Coen, M.L.; Zhang, H.; Ng, S.C.; Comess, K.M.; Petros, A.M.; Rosenberg, S.H.; Warrior, U.; et al. Discovery of novel inhibitors of Bcl-xL using multiple high-throughput screening platforms. *Anal. Biochem.* **2004**, *328*, 131–138. [[CrossRef](#)] [[PubMed](#)]
22. Real, P.J.; Cao, Y.; Wang, R.; Nikolovska-Coleska, Z.; Sanz-Ortiz, J.; Wang, S.; Fernandez-Luna, J.L. Breast cancer cells can evade apoptosis-mediated selective killing by a novel small molecule inhibitor of Bcl-2. *Cancer Res.* **2004**, *64*, 7947–7953. [[CrossRef](#)] [[PubMed](#)]
23. Vassilev, L.T.; Vu, B.T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **2004**, *303*, 844–848. [[CrossRef](#)] [[PubMed](#)]
24. Braisted, A.C.; Oslob, J.D.; Delano, W.L.; Hyde, J.; McDowell, R.S.; Waal, N.; Yu, C.; Arkin, M.R.; Raimundo, B.C. Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J. Am. Chem. Soc.* **2003**, *125*, 3714–3715. [[CrossRef](#)] [[PubMed](#)]
25. Emerson, S.D.; Palermo, R.; Liu, C.M.; Tilley, J.W.; Chen, L.; Danho, W.; Madison, V.S.; Greeley, D.N.; Ju, G.; Fry, D.C. NMR characterization of interleukin-2 in complexes with the IL-2R $\alpha$  receptor component, and with low molecular weight compounds that inhibit the IL-2/IL-R $\alpha$  interaction. *Protein Sci.* **2003**, *12*, 811–822. [[CrossRef](#)] [[PubMed](#)]
26. Li, L.; Thomas, R.M.; Suzuki, H.; De Brabander, J.K.; Wang, X.; Harran, P.G. A small molecule Smac mimic potentiates TRAIL- and TNF $\alpha$ -mediated cell death. *Science* **2004**, *305*, 1471–1474. [[CrossRef](#)] [[PubMed](#)]
27. Oost, T.K.; Sun, C.; Armstrong, R.C.; Al-Assaad, A.S.; Betz, S.F.; Deckwerth, T.L.; Ding, H.; Elmore, S.W.; Meadows, R.P.; Olejniczak, E.T.; et al. Discovery of potent antagonists of the antiapoptotic protein XIAP for the treatment of cancer. *J. Med. Chem.* **2004**, *47*, 4417–4426. [[CrossRef](#)] [[PubMed](#)]
28. Ko, A.H.; Chiorean, E.G.; Kwak, E.L.; Lenz, H.-J.; Nadler, P.I.; Wood, D.L.; Fujimori, M.; Inada, T.; Kouji, H.; McWilliams, R.R. Final results of a phase Ib dose-escalation study of PRI-724, a CBP/ $\beta$ -catenin modulator, plus gemcitabine (GEM) in patients with advanced pancreatic adenocarcinoma (APC) as second-line therapy after FOLFIRINOX or FOLFOX. *J. Clin. Oncol.* **2016**, *34*, e15721. [[CrossRef](#)]

29. Fukuchi, S.; Amemiya, T.; Sakamoto, S.; Nobe, Y.; Hosoda, K.; Kado, Y.; Murakami, S.D.; Koike, R.; Hiroaki, H.; Ota, M. IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. *Nucleic Acids Res.* **2014**, *42*, D320–D325. [[CrossRef](#)] [[PubMed](#)]
30. Fukuchi, S.; Sakamoto, S.; Nobe, Y.; Murakami, S.D.; Amemiya, T.; Hosoda, K.; Koike, R.; Hiroaki, H.; Ota, M. IDEAL: Intrinsically disordered proteins with extensive annotations and literature. *Nucleic Acids Res.* **2012**, *40*, D507–D511. [[CrossRef](#)] [[PubMed](#)]
31. Piovesan, D.; Tabaro, F.; Micetic, I.; Necci, M.; Quaglia, F.; Oldfield, C.J.; Aspromonte, M.C.; Davey, N.E.; Davidovic, R.; Dosztanyi, Z.; et al. DisProt 7.0: A major update of the database of disordered proteins. *Nucleic Acids Res.* **2017**, *45*, D1123–D1124. [[CrossRef](#)] [[PubMed](#)]
32. Tompa, P.; Davey, N.E.; Gibson, T.J.; Babu, M.M. A million peptide motifs for the molecular biologist. *Mol. Cell* **2014**, *55*, 161–169. [[CrossRef](#)] [[PubMed](#)]
33. The UniProt Consortium. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169. [[CrossRef](#)] [[PubMed](#)]
34. Necci, M.; Piovesan, D.; Dosztanyi, Z.; Tosatto, S.C.E. MobiDB-lite: Fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* **2017**, *33*, 1402–1404. [[CrossRef](#)] [[PubMed](#)]
35. Jones, D.T.; Cozzetto, D. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **2015**, *31*, 857–863. [[CrossRef](#)] [[PubMed](#)]
36. Monastyrskyy, B.; Kryshtafovych, A.; Moulton, J.; Tramontano, A.; Fidelis, K. Assessment of protein disorder region predictions in CASP10. *Proteins* **2014**, *82* (Suppl. S2), 127–137. [[CrossRef](#)] [[PubMed](#)]
37. Kanehisa, M.; Sato, Y.; Furumichi, M.; Morishima, K.; Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **2018**. [[CrossRef](#)] [[PubMed](#)]
38. The IDEAL database. Available online: <http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/> (accessed on 22 January 2019).
39. Rubin, S.M.; Gall, A.L.; Zheng, N.; Pavletich, N.P. Structure of the Rb C-terminal domain bound to E2F1-DP1: A mechanism for phosphorylation-induced E2F release. *Cell* **2005**, *123*, 1093–1106. [[CrossRef](#)] [[PubMed](#)]
40. Fontes, M.R.; Teh, T.; Jans, D.; Brinkworth, R.I.; Kobe, B. Structural basis for the specificity of bipartite nuclear localization sequence binding by importin- $\alpha$ . *J. Biol. Chem.* **2003**, *278*, 27981–27987. [[CrossRef](#)] [[PubMed](#)]
41. Pawson, T.; Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **2003**, *300*, 445–452. [[CrossRef](#)] [[PubMed](#)]
42. Lee, H.J.; Zheng, J.J. PDZ domains and their binding partners: Structure, specificity, and modification. *Cell Commun. Signal.* **2010**, *8*, 8. [[CrossRef](#)] [[PubMed](#)]
43. Nomine, Y.; Masson, M.; Charbonnier, S.; Zanier, K.; Ristriani, T.; Deryckere, F.; Sibler, A.P.; Desplancq, D.; Atkinson, R.A.; Weiss, E.; et al. Structural and functional analysis of E6 oncoprotein: Insights in the molecular pathways of human papillomavirus-mediated pathogenesis. *Mol. Cell* **2006**, *21*, 665–678. [[CrossRef](#)] [[PubMed](#)]
44. Greschik, H.; Althage, M.; Flaig, R.; Sato, Y.; Chavant, V.; Peluso-Iltis, C.; Choulier, L.; Cronet, P.; Rochel, N.; Schule, R.; et al. Communication between the ERR $\alpha$  homodimer interface and the PGC-1 $\alpha$  binding surface via the helix 8-9 loop. *J. Biol. Chem.* **2008**, *283*, 20220–20230. [[CrossRef](#)] [[PubMed](#)]
45. Kallen, J.; Schlaeppli, J.M.; Bitsch, F.; Filipuzzi, I.; Schilb, A.; Riou, V.; Graham, A.; Strauss, A.; Geiser, M.; Fournier, B. Evidence for ligand-independent transcriptional activation of the human estrogen-related receptor alpha (ERR $\alpha$ ): Crystal structure of ERR $\alpha$  ligand binding domain in complex with peroxisome proliferator-activated receptor coactivator-1 $\alpha$ . *J. Biol. Chem.* **2004**, *279*, 49330–49337. [[CrossRef](#)] [[PubMed](#)]
46. KEGG DISEASE Database. Available online: [https://www.genome.jp/kegg-bin/get\\_htext?br08402.keg](https://www.genome.jp/kegg-bin/get_htext?br08402.keg) (accessed on 22 January 2019).
47. Giasson, B.I.; Duda, J.E.; Murray, I.V.; Chen, Q.; Souza, J.M.; Hurtig, H.I.; Ischiropoulos, H.; Trojanowski, J.Q.; Lee, V.M. Oxidative damage linked to neurodegeneration by selective  $\alpha$ -synuclein nitration in synucleinopathy lesions. *Science* **2000**, *290*, 985–989. [[CrossRef](#)] [[PubMed](#)]
48. Gandhi, P.N.; Wang, X.; Zhu, X.; Chen, S.G.; Wilson-Delfosse, A.L. The Roc domain of leucine-rich repeat kinase 2 is sufficient for interaction with microtubules. *J. Neurosci. Res.* **2008**, *86*, 1711–1720. [[CrossRef](#)] [[PubMed](#)]
49. Vilarino-Guell, C.; Wider, C.; Ross, O.A.; Dachsel, J.C.; Kachergus, J.M.; Lincoln, S.J.; Soto-Ortolaza, A.I.; Cobb, S.A.; Wilhoite, G.J.; Bacon, J.A.; et al. VPS35 mutations in Parkinson disease. *Am. J. Hum. Genet.* **2011**, *89*, 162–167. [[CrossRef](#)] [[PubMed](#)]



50. Yoshii, S.R.; Kishi, C.; Ishihara, N.; Mizushima, N. Parkin mediates proteasome-dependent protein degradation and rupture of the outer mitochondrial membrane. *J. Biol. Chem.* **2011**, *286*, 19630–19640. [[CrossRef](#)] [[PubMed](#)]
51. Poole, A.C.; Thomas, R.E.; Andrews, L.A.; McBride, H.M.; Whitworth, A.J.; Pallanck, L.J. The PINK1/Parkin pathway regulates mitochondrial morphology. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1638–1643. [[CrossRef](#)] [[PubMed](#)]
52. Ottolini, D.; Cali, T.; Negro, A.; Brini, M. The Parkinson disease-related protein DJ-1 counteracts mitochondrial impairment induced by the tumour suppressor protein p53 by enhancing endoplasmic reticulum-mitochondria tethering. *Hum. Mol. Genet.* **2013**, *22*, 2152–2168. [[CrossRef](#)] [[PubMed](#)]
53. Chartier-Harlin, M.C.; Dachsel, J.C.; Vilarino-Guell, C.; Lincoln, S.J.; LePrete, F.; Hulihan, M.M.; Kachergus, J.; Milnerwood, A.J.; Tapia, L.; Song, M.S.; et al. Translation initiator EIF4G1 mutations in familial Parkinson disease. *Am. J. Hum. Genet.* **2011**, *89*, 398–406. [[CrossRef](#)] [[PubMed](#)]
54. Safaee, N.; Kozlov, G.; Noronha, A.M.; Xie, J.; Wilds, C.J.; Gehring, K. Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol. Cell* **2012**, *48*, 375–386. [[CrossRef](#)] [[PubMed](#)]
55. Biros, I.; Forrest, S. Spinal muscular atrophy: Untangling the knot? *J. Med. Genet.* **1999**, *36*, 1–8. [[PubMed](#)]
56. Wirth, B. An update of the mutation spectrum of the survival motor neuron gene (*SMN1*) in autosomal recessive spinal muscular atrophy (SMA). *Hum. Mutat.* **2000**, *15*, 228–237. [[CrossRef](#)]
57. Ogino, S.; Wilson, R.B. Spinal muscular atrophy: Molecular genetics and diagnostics. *Expert Rev. Mol. Diagn.* **2004**, *4*, 15–29. [[CrossRef](#)] [[PubMed](#)]
58. Matthijs, G.; Schollen, E.; Legius, E.; Devriendt, K.; Goemans, N.; Kayserili, H.; Apak, M.Y.; Cassiman, J.J. Unusual molecular findings in autosomal recessive spinal muscular atrophy. *J. Med. Genet.* **1996**, *33*, 469–474. [[CrossRef](#)] [[PubMed](#)]
59. Martin, R.; Gupta, K.; Ninan, N.S.; Perry, K.; Van Duyne, G.D. The survival motor neuron protein forms soluble glycine zipper oligomers. *Structure* **2012**, *20*, 1929–1939. [[CrossRef](#)] [[PubMed](#)]
60. Mourelatos, Z.; Abel, L.; Yong, J.; Kataoka, N.; Dreyfuss, G. SMN interacts with a novel family of hnRNP and spliceosomal proteins. *EMBO J.* **2001**, *20*, 5443–5452. [[CrossRef](#)] [[PubMed](#)]
61. Zuliani, G.; Arca, M.; Signore, A.; Bader, G.; Fazio, S.; Chianelli, M.; Bellosta, S.; Campagna, F.; Montali, A.; Maioli, M.; et al. Characterization of a new form of inherited hypercholesterolemia: Familial recessive hypercholesterolemia. *Arterioscler. Thromb. Vasc. Biol.* **1999**, *19*, 802–809. [[CrossRef](#)] [[PubMed](#)]
62. Garcia, C.K.; Wilund, K.; Arca, M.; Zuliani, G.; Fellin, R.; Maioli, M.; Calandra, S.; Bertolini, S.; Cossu, F.; Grishin, N.; et al. Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science* **2001**, *292*, 1394–1398. [[CrossRef](#)] [[PubMed](#)]
63. Scheele, U.; Alves, J.; Frank, R.; Duwel, M.; Kalthoff, C.; Ungewickell, E. Molecular and functional characterization of clathrin- and AP-2-binding determinants within a disordered domain of auxilin. *J. Biol. Chem.* **2003**, *278*, 25357–25368. [[CrossRef](#)] [[PubMed](#)]
64. Edeling, M.A.; Mishra, S.K.; Keyel, P.A.; Steinhäuser, A.L.; Collins, B.M.; Roth, R.; Heuser, J.E.; Owen, D.J.; Traub, L.M. Molecular switches involving the AP-2  $\beta 2$  appendage regulate endocytic cargo selection and clathrin coat assembly. *Dev. Cell* **2006**, *10*, 329–342. [[CrossRef](#)] [[PubMed](#)]
65. Goldfarb, L.G.; Dalakas, M.C. Tragedy in a heartbeat: Malfunctioning desmin causes skeletal and cardiac muscle disease. *J. Clin. Investig.* **2009**, *119*, 1806–1813. [[CrossRef](#)] [[PubMed](#)]
66. Sharma, S.; Conover, G.M.; Elliott, J.L.; Der Perng, M.; Herrmann, H.; Quinlan, R.A.  $\alpha$ B-crystallin is a sensor for assembly intermediates and for the subunit topology of desmin intermediate filaments. *Cell Stress Chaperones* **2017**, *22*, 613–626. [[CrossRef](#)] [[PubMed](#)]
67. Bar, H.; Goudeau, B.; Walde, S.; Casteras-Simon, M.; Mucke, N.; Shatunov, A.; Goldberg, Y.P.; Clarke, C.; Holton, J.L.; Eymard, B.; et al. Conspicuous involvement of desmin tail mutations in diverse cardiac and skeletal myopathies. *Hum. Mutat.* **2007**, *28*, 374–386. [[CrossRef](#)] [[PubMed](#)]
68. Vattemi, G.; Neri, M.; Piffer, S.; Vicart, P.; Gualandi, F.; Marini, M.; Guglielmi, V.; Filosto, M.; Tonin, P.; Ferlini, A.; et al. Clinical, morphological and genetic studies in a cohort of 21 patients with myofibrillar myopathy. *Acta Myol.* **2011**, *30*, 121–126. [[PubMed](#)]
69. Arbustini, E.; Pasotti, M.; Pilotto, A.; Pellegrini, C.; Grasso, M.; Previtali, S.; Repetto, A.; Bellini, O.; Azan, G.; Scaffino, M.; et al. Desmin accumulation restrictive cardiomyopathy and atrioventricular block associated with desmin gene defects. *Eur. J. Heart Fail.* **2006**, *8*, 477–483. [[CrossRef](#)] [[PubMed](#)]

70. Baker, L.K.; Gillis, D.C.; Sharma, S.; Ambrus, A.; Herrmann, H.; Conover, G.M. Nebulin binding impedes mutant desmin filament assembly. *Mol. Biol. Cell* **2013**, *24*, 1918–1932. [[CrossRef](#)] [[PubMed](#)]
71. Ota, M.; Gonja, H.; Koike, R.; Fukuchi, S. Multiple-localization and Hub proteins. *PLoS ONE* **2016**, *11*, e0156455. [[CrossRef](#)] [[PubMed](#)]
72. Dinkel, H.; Van Roey, K.; Michael, S.; Kumar, M.; Uyar, B.; Altenberg, B.; Milchevskaya, V.; Schneider, M.; Kuhn, H.; Behrendt, A.; et al. ELM 2016-data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.* **2016**, *44*, D294–300. [[CrossRef](#)] [[PubMed](#)]
73. Uyar, B.; Weatheritt, R.J.; Dinkel, H.; Davey, N.E.; Gibson, T.J. Proteome-wide analysis of human disease mutations in short linear motifs: Neglected players in cancer? *Mol. Biosyst.* **2014**, *10*, 2626–2642. [[CrossRef](#)] [[PubMed](#)]
74. Groft, C.M.; Burley, S.K. Recognition of eIF4G by rotavirus NSP3 reveals a basis for mRNA circularization. *Mol. Cell* **2002**, *9*, 1273–1283. [[CrossRef](#)]
75. Chatr-Aryamontri, A.; Oughtred, R.; Boucher, L.; Rust, J.; Chang, C.; Kolas, N.K.; O'Donnell, L.; Oster, S.; Theesfeld, C.; Sellam, A.; et al. The BioGRID interaction database: 2017 Update. *Nucleic Acids Res.* **2017**, *45*, D369–D379. [[CrossRef](#)] [[PubMed](#)]
76. Haynes, C.; Oldfield, C.J.; Ji, F.; Klitgord, N.; Cusick, M.E.; Radivojac, P.; Uversky, V.N.; Vidal, M.; Iakoucheva, L.M. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput. Biol.* **2006**, *2*, e100. [[CrossRef](#)] [[PubMed](#)]
77. Patil, A.; Kinoshita, K.; Nakamura, H. Hub promiscuity in protein-protein interaction networks. *Int. J. Mol. Sci.* **2010**, *11*, 1930–1943. [[CrossRef](#)] [[PubMed](#)]
78. Collins, M.O.; Yu, L.; Campuzano, I.; Grant, S.G.; Choudhary, J.S. Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. *Mol. Cell. Proteomics* **2008**, *7*, 1331–1348. [[CrossRef](#)] [[PubMed](#)]
79. Brangwynne, C.P.; Eckmann, C.R.; Courson, D.S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Julicher, F.; Hyman, A.A. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **2009**, *324*, 1729–1732. [[CrossRef](#)] [[PubMed](#)]
80. Brangwynne, C.P.; Mitchison, T.J.; Hyman, A.A. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4334–4339. [[CrossRef](#)] [[PubMed](#)]
81. Li, P.; Banjade, S.; Cheng, H.C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J.V.; King, D.S.; Banani, S.F.; et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **2012**, *483*, 336–340. [[CrossRef](#)] [[PubMed](#)]
82. Kato, M.; Han, T.W.; Xie, S.; Shi, K.; Du, X.; Wu, L.C.; Mirzaei, H.; Goldsmith, E.J.; Longgood, J.; Pei, J.; et al. Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **2012**, *149*, 753–767. [[CrossRef](#)] [[PubMed](#)]
83. Banjade, S.; Wu, Q.; Mittal, A.; Peeples, W.B.; Pappu, R.V.; Rosen, M.K. Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6426–E6435. [[CrossRef](#)] [[PubMed](#)]
84. Nott, T.J.; Petsalaki, E.; Farber, P.; Jarvis, D.; Fussner, E.; Plochowitz, A.; Craggs, T.D.; Bazett-Jones, D.P.; Pawson, T.; Forman-Kay, J.D.; et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **2015**, *57*, 936–947. [[CrossRef](#)] [[PubMed](#)]
85. Pak, C.W.; Kosno, M.; Holehouse, A.S.; Padrick, S.B.; Mittal, A.; Ali, R.; Yunus, A.A.; Liu, D.R.; Pappu, R.V.; Rosen, M.K. Sequence Determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **2016**, *63*, 72–85. [[CrossRef](#)] [[PubMed](#)]

