Molecular BioSystems

PAPER

Cite this: *Mol. BioSyst.*, 2016, 12, 2798

# Autophagy-related intrinsically disordered proteins in intra-nuclear compartments†

Insung Na,‡[a] Fanchi Meng,‡[b] Lukasz Kurgan*[c] and Vladimir N. Uversky*[adef]

Received 27th January 2016,
Accepted 20th June 2016

DOI: 10.1039/c6mb00069j

www.rsc.org/molecularbiosystems

Recent analyses indicated that autophagy can be regulated *via* some nuclear transcriptional networks and many important players in the autophagy and other forms of programmed cell death are known to be intrinsically disordered. To this end, we analyzed similarities and differences in the intrinsic disorder distribution of nuclear and non-nuclear proteins related to autophagy. We also looked at the peculiarities of the distribution of the intrinsically disordered autophagy-related proteins in various intra-nuclear organelles, such as the nucleolus, chromatin, Cajal bodies, nuclear speckles, promyelocytic leukemia (PML) nuclear bodies, nuclear lamina, nuclear pores, and perinucleolar compartment. This analysis revealed that the autophagy-related proteins constitute about 2.5% of the non-nuclear proteins and 3.3% of the nuclear proteins, which corresponds to a substantial enrichment by about 32% in the nucleus. Curiously, although, in general, the autophagy-related proteins share similar characteristics of disorder with a generic set of all non-nuclear proteins, chromatin and nuclear speckles are enriched in the intrinsically disordered autophagy proteins (29 and 37% of these proteins are disordered, respectively) and have high disorder content at 0.24 and 0.27, respectively. Therefore, our data suggest that some of the nuclear disordered proteins may play important roles in autophagy.

## Introduction

The nucleus is a eukaryote-specific organelle that is used for the storage of hereditary material and for the coordination of various cellular activities, such as protein synthesis, cell division, cell growth, and metabolism.[1–4] Structurally and morphologically, the cell nucleus is a double membrane-bound, large, non-homogeneous organelle equipped with a set of highly dynamic, membrane-less sub-nuclear organelles or compartments, many of which appear only during certain stages of the cell cycle *via* dynamic recruitment of proteins, RNA and DNA.[5] Because of the lack of protective outer membranes, the content of these organelles is involved in direct contact with the surrounding nucleoplasm.[6,7] Similar to membrane-less organelles found in the cytoplasm, these sub-nuclear organelles are described as liquid-droplet phases of the nucleoplasm.[8–13] Currently, the exact functions of many of these sub-nuclear compartments, as well as biological roles of interactions between them, and the molecular mechanisms defining the ability of these organelles to maintain their specific structures in the absence of membranes remain mostly unknown.[14] Among these subnuclear organelles are Cajal bodies, chromatin, cleavage bodies, nuclear gems (Gemini of coiled bodies), nuclear pores, nuclear speckles, nucleolus, Oct1/PTF/ transcription (OPT) domains, PcG bodies (subnuclear organelles containing polycomb group proteins), perinucleolar compartment, promyelocytic leukemia (PML) nuclear bodies, SAM68 nuclear bodies, and a few others.[15] The introductory description of some of these organelles (nucleolus, chromatin, Cajal bodies, nuclear speckles, PML bodies, nuclear lamina, nuclear pores, and perinucleolar compartment) is provided in our recent studies.[16,17]

Proteins (and RNA or, in some cases, DNA) represent the common denominator of various sub-nuclear membrane-less organelles. Many of the proteins found there are intrinsically

[a] *Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA. E-mail: vuversky@health.usf.edu; Fax: +1-813-974-7357; Tel: +1-813-974-5815*
[b] *Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, T6G 2V4, Canada*
[c] *Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23219, USA. E-mail: lkurgan@vcu.edu; Fax: +1-804-828-2771; Tel: +1-804-827-3986*
[d] *USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA*
[e] *Biology Department, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia*
[f] *Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg 194064, Russia*

† Electronic supplementary information (ESI) available: For several illustrative autophagy-related IDPs found in various nuclear compartments, supplementary materials provide results of the evaluation of their per-residue intrinsic disorder propensities by members of the PONDR family, information on functional disorder in these proteins retrieved from the D2P2 database, and the results of the analysis of the interactivity of these autophagy-related nuclear proteins by STRING. See DOI: 10.1039/c6mb00069j

‡ These authors contributed equally to this study.

disordered;[16,17] *i.e.*, they have a wide spectrum of crucial functions, being characterized by the lack of specific tertiary structures.[18–24] Such intrinsically disordered proteins (IDPs) and hybrid proteins containing ordered domains and intrinsically disordered protein regions (IDPRs) are abundantly found in all proteomes,[25–31] where they complement the functionality of ordered proteins,[32–34] being engaged in various activities related to regulation, control, recognition, and signaling.[35–37]

Recently, it has been shown that proteins involved in the regulation and execution of programmed cell death (PCD) possess substantial amounts of intrinsic disorder.[38] PCD is a suicidal response of a cell to some environmental factors mediated by a set of specific intracellular programs. Depending on the type such a program, PCD, is realized in the form of apoptosis, autophagy, or necroptosis, with different PCD routes assumed to have very different biological roles. In fact, apoptosis (which is important for the development, immune regulation, and homeostasis of a multi-cellular organism) and necroptosis (that plays a role in the modulation of inflammatory response, serves as a backup mechanism to clear pathogens, and is involved in the immunologically silent maintenance of T cell homeostasis[39]) are typically considered as mechanisms contributing to the cell death. On the other hand, autophagy could have either pro-survival or pro-death functions,[40–42] since this program typically regulates the turnover of long-lived proteins, the disposal of the damaged organelles and misfolded proteins, and the turnover of the cellular building blocks following nutrient deprivation. Therefore, it controls starvation, cell differentiation, and cell survival,[43–45] and under some conditions it can initiate cell death.[40,43]

Autophagy is traditionally considered as a cytoplasmic process where the unnecessary or dysfunctional cellular cytoplasmic components are first isolated from the rest of the cell within a specialized double-membrane vesicle, autophagosome, which then fuses with the lysosome, where these unnecessary or dysfunctional cellular components are degraded and recycled.[46–50] Furthermore, recent studies clearly indicated that this important mechanism is controlled *via* a transcriptional network that includes tightly controlled transcription factors (such as TFEB and ZKSCAN3).[51] Furthermore, it has been emphasized that the autophagy in yeast and mammals is regulated by various transcriptional, post-transcriptional, and post-translational means.[52] Based on these and other observations it has been concluded that although the role of nuclear events in autophagy was underappreciated, the nucleus has to be considered as a major regulator of autophagy.[51,53] To fill this gap, the current work extends our earlier studies on the analysis of the disorder status of human[16] and mouse[17] nuclear proteins. Here, we analyzed the disorder status of mouse and yeast nuclear proteins that are involved in autophagy to shed more light on the prevalence and potential roles of autophagy-related intrinsically disordered nuclear proteins.

## Results and discussion

### Global analysis of the prevalence of intrinsic disorder in autophagy-related nuclear proteins

We investigated whether autophagy-related mouse proteins are preferentially localized in the nucleus. We do not consider the Cajal bodies and perinucleolar compartment when assessing the characteristics of these proteins in the individual intra-nuclear compartments since the corresponding protein counts are too low at 2 and 1, respectively. The other compartments include between 6 and 56 autophagy-related mouse proteins (Table 1).

We found that the autophagy-related proteins constitute about 2.5% of the non-nuclear mouse proteins and 3.3% of the nuclear mouse proteins, which corresponds to a substantial enrichment by about 32% in the nucleus. This raises an important question as to whether the enrichment in nuclear IDPs observed here is specific to the autophagy process or can be a more general phenomenon found under various stressed conditions. This question deserves a detailed analysis. However, based on the well-known involvement of IDPs in the regulation of various cellular processes,[35–37] one can argue that any conditions leading to the changes in cellular regulatory networks might cause an increase in the overall cellular or nuclear intrinsic disorder status.

Fig. 1A reveals that the number of autophagy-related proteins in the mouse nucleus varies widely between different intra-nuclear compartments, with particularly high fractions being found in the PML nuclear bodies, nuclear pores, and lamina. This enrichment of autophagy-related intrinsically disordered proteins in three nuclear compartments in the mouse cell can be related to the roles of these compartments in the regulation of autophagy. For example, besides playing important roles as a provider of the mechanical support to the nucleus, the nuclear lamina is known to be involved in most activities taking place in the nucleus.

Table 1 Characteristics of mouse autophagy-related proteins

| Intra-nuclear compartment | Number of proteins | Averaged length | Averaged PONDR® VSL2 score | Averaged PONDR® VLXT score | Averaged ΔCH | Averaged ΔCDF |
|---|---|---|---|---|---|---|
| Non-nuclear | 748 | 871 ± 800 | 0.49 ± 0.18 | 0.38 ± 0.15 | −0.088 ± 0.116 | −0.052 ± 0.246 |
| Nuclear | 123 | 928 ± 898 | 0.52 ± 0.18 | 0.41 ± 0.18 | −0.059 ± 0.111 | −0.089 ± 0.237 |
| Chromatin | 21 | 1591 ± 1316 | 0.56 ± 0.16 | 0.45 ± 0.14 | −0.042 ± 0.095 | −0.146 ± 0.223 |
| Cajal bodies | 2 | 643 ± 575 | 0.61 ± 0.34 | 0.53 ± 0.31 | −0.034 ± 0.211 | −0.245 ± 0.442 |
| Nucleolus | 56 | 641 ± 538 | 0.49 ± 0.15 | 0.37 ± 0.13 | −0.071 ± 0.109 | −0.047 ± 0.219 |
| Nuclear lamina | 9 | 1249 ± 1060 | 0.55 ± 0.20 | 0.39 ± 0.15 | −0.073 ± 0.081 | −0.135 ± 0.293 |
| Nuclear pore | 6 | 1154 ± 1252 | 0.38 ± 0.02 | 0.31 ± 0.01 | −0.135 ± 0.016 | 0.104 ± 0.021 |
| Nuclear speckle | 19 | 682 ± 464 | 0.58 ± 0.18 | 0.48 ± 0.17 | −0.005 ± 0.154 | −0.179 ± 0.253 |
| Perinuclear | 1 | 1274 | 0.52 | 0.47 | −0.062 | −0.067 |
| PML nuclear bodies | 9 | 1251 ± 1079 | 0.53 ± 0.18 | 0.39 ± 0.16 | −0.085 ± 0.081 | −0.101 ± 0.268 |

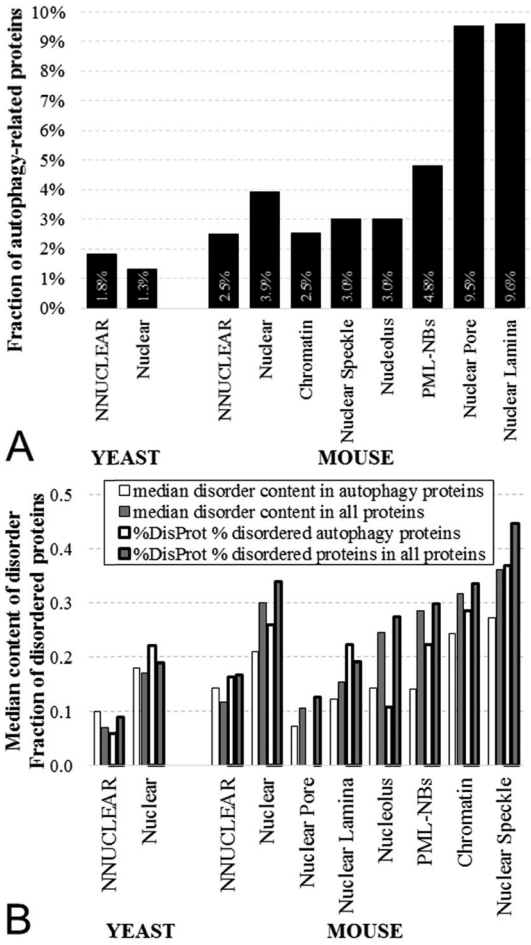


Fig. 1 Fraction of autophagy-related proteins (panel A) and median disorder content and fraction of disordered proteins (panel B) for the autophagy-related non-nuclear and nuclear yeast proteins (left sides of plots) and the autophagy-related proteins and the non-nuclear proteins from the NNUCLEAR dataset and from the considered intra-nuclear compartment from the $NUCLEAR_{ap}$ dataset (for mouse proteins) (right sides of plots). Disorder was annotated with the consensus of Espritz and IUPred.

The related processes include DNA replication, RNA transcription, and nuclear and chromatin organization, as well as several important cellular processes, such as apoptosis, cell cycle regulation, cell development and differentiation, and nuclear migration.[54] PML (promyelocytic leukemia) nuclear bodies belong to the nuclear matrix that plays a role in the regulation of many nuclear functions, such as DNA replication, transcription, and epigenetic silencing.[55] PMLs contribute to the formation of a favorable nuclear environment for the expression of specific genes, thereby playing a role in the regulation of gene expression.[56] They are regulated by various types of cellular stresses, ranging from viral infection to DNA-damage, transformation, and oxidative stress,[57] are commonly associated with viral infection and oncogenesis,[57,58] and are known to recruit a very large number ($\sim$100) of different partner proteins.[57] Finally, nuclear pores, which are one of the largest protein complexes in the cell, represent a specific aqueous channel providing access to the nucleus. It is filled with multiple copies of several evolutionarily conserved proteins known as nucleporins that form a complex network regulating the transport of proteins and RNA across the nuclear envelope.[59]

Next, we compared disorder content and fractions of disordered proteins between the autophagy-related proteins and all considered proteins both inside and outside the nucleus (Fig. 1B). The results for the non-nuclear proteins (NNUCLEAR set of bars in Fig. 1B) demonstrate that they are typically less disordered than nuclear proteins in general. Furthermore, Fig. 1B illustrates that among the non-nuclear mouse proteins the autophagy-related proteins share similar characteristics of disorder with a generic set of all non-nuclear proteins (the corresponding bars have comparable heights). On the other hand, the results for the intra-nuclear compartments, which are sorted by the increasing median disorder content over all nuclear proteins (grey bars), show that the autophagy-related proteins are characterized by relatively low levels of intrinsic disorder, except for the chromatin and nuclear speckles. The latter two compartments are enriched in the disordered autophagy proteins (29 and 37% of these proteins are disordered, respectively) and have high disorder content at 0.24 and 0.27, respectively. These high levels of disorder are similar to the overall disorder levels in proteins from these two intra-nuclear compartments. Fig. 1B also shows that in comparison with other proteins found in the corresponding nuclear compartments, autophagy-related mouse proteins (except for lamina) are not enriched in intrinsic disorder. To summarize this part, Fig. 1A illustrates how autophagy-related proteins are distributed inside and outside the nucleus. It clearly shows that in the mouse cell, autophagy-related proteins are more common among the nuclear proteins than among the cytoplasmic proteins. Fig. 1B shows that the nuclear proteins are, in general, more disordered than the non-nuclear proteins, and this observation agrees well with the results of previous studies where nuclear proteins were systematically reported to be characterized by high levels of intrinsic disorder.[16,17] Furthermore, based on the comparison of the levels of intrinsic disorder in autophagy-related and autophagy-unrelated proteins in various nuclear compartments of the mouse nucleus one can conclude (see Fig. 1B) that the autophagy-related proteins are characterized by relatively low levels of intrinsic disorder, except for the chromatin and nuclear speckles, whose corresponding bars are taller than those of the nuclear proteins in general. Finally, among the autophagy-related proteins found in various compartments of the mouse nucleus, only lamina is characterized by an apparent enrichment of disorder. The observations that IDPs are more commonly found in the autophagy-unrelated proteins of various nuclear compartments than in their autophagy-related proteins do not mean that IDPs are not important. In addition to the information on how many IDPs are associated with autophagy in various nuclear compartments, a very important consideration is what these autophagy-related IDPs are and what they do.

The point that mouse autophagy-related nuclear proteins contain substantial amounts of intrinsic disorder is further illustrated by Fig. 2A. Here, the per-protein propensities for disorder (computed by averaging the corresponding per-residue propensities) were evaluated by PONDR® VSL2[60] and PONDR®

 

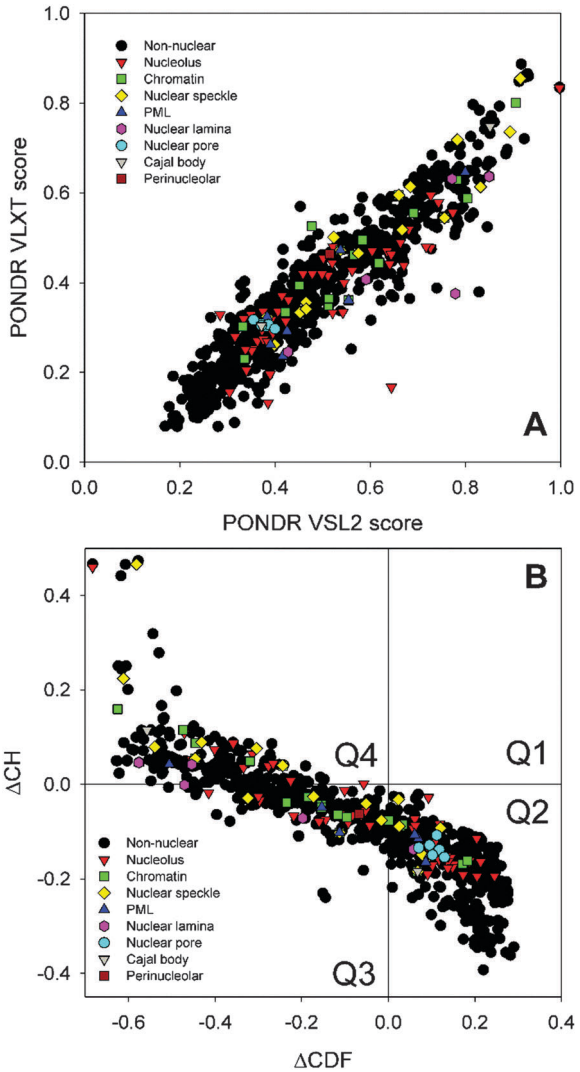


Fig. 2 Prevalence of intrinsic disorder in autophagy-related mouse proteins. (A) Per-protein propensities for disorder (average of the corresponding per-residue propensities) evaluated by PONDR® VSL2[60] (*x*-axis) and PONDR® VLXT[21,61] (*y*-axis) for the autophagy-related non-nuclear proteins and autophagy-related nuclear proteins in various intra-nuclear compartments listed in Table 1. (B) Evaluation of intrinsic disorder in autophagy-related mouse proteins by CH-CDF analysis.[20,28,62] Here, the coordinates of each point were calculated as a distance of the corresponding protein in the CH-plot from the boundary (*y*-axis) and an average distance of the respective CDF curve from the CDF boundary (*x*-axis). The four quadrants correspond to the following predictions: Q1, proteins predicted to be disordered by CH-plots, but ordered by CDFs; Q2, ordered proteins; Q3, proteins predicted to be disordered by CDFs, but compact by CH-plots (*i.e.*, putative molten globules or hybrid proteins); Q4, proteins predicted to be disordered by both methods. Proteins found in different sub-nuclear compartments are indicated by differently colored symbols.

VLXT.[21,61] Fig. 2A and Table 1 illustrate that mouse autophagy-related nuclear proteins are generally more disordered than the autophagy-related non-nuclear proteins. In fact, our analysis using PONDR® VSL2 (PONDR® VLXT) indicates that the autophagy-related non-nuclear proteins are characterized by the mean per-protein propensity for disorder of 0.49 ± 0.18 (0.38 ± 0.15), whereas for all nuclear autophagy-related proteins the corresponding values were 0.52 ± 0.18 (0.41 ± 0.18). Curiously, autophagy-related proteins found in different membrane-less nuclear organelles were disordered to a different extent, with the nuclear pore proteins (0.377 ± 0.015 (0.306 ± 0.007)) being noticeably more ordered than non-nuclear proteins and proteins in other nuclear compartments, and with autophagy-related proteins in nuclear speckles and chromatin being noticeably more disordered than proteins in other datasets (0.58 ± 0.18 (0.48 ± 0.17), and 0.56 ± 0.16 (0.45 ± 0.14), respectively, see Table 1).

Overall, according to the results of the analysis of the average per-protein intrinsic disorder propensities evaluated by PONDR® VSL2, the autophagy-related proteins in intra-nuclear compartments can be arranged in the following order: nuclear speckles > chromatin > nuclear lamina > PML nuclear bodies > nucleus = non-nuclear > nuclear pore. Similar but not identical order is generated based on the results of the PONDR® VLXT-based analysis: nuclear speckles > chromatin > nuclear lamina = PML nuclear bodies > non-nuclear > nucleus > nuclear pore (see Table 1). These results are also consistent with the analysis from Fig. 1B where we used a different set of methods to putatively annotate intrinsic disorder. Curiously, this order is noticeably different from that generated in our earlier analysis of the disorder status in all mouse nuclear proteins:[17] perinucleolar compartment > nuclear speckles > chromatin > nucleolus = PML nuclear bodies > Cajal bodies > nuclear lamina > non-nuclear proteins > nuclear pore. This observation indicates that autophagy-related IDPs are differently distributed within the membrane-less nuclear organelles, suggesting that these organelles might be differently involved in autophagy. In our view, this is an interesting hypothesis that requires further analysis in order to find supporting evidence.

Next, the distribution of autophagy-related nuclear proteins within the CH-CDF phase space was analyzed. Fig. 2B shows that these proteins occupy quadrants Q2 (that contains "true" ordered proteins; *i.e.*, proteins predicted to be ordered by both CH and CDF), Q3 (where native molten globular or mixed proteins containing comparable amounts of order and disorder are located; *i.e.*, proteins predicted to be disordered by CDF but compact by CH-plot), and Q4 (that includes native coils or native pre-molten globules; *i.e.*, proteins predicted to be disordered as a whole by both CH and CDF), whereas there are no proteins in quadrant Q1 (that contains proteins predicted to be extended by CH but ordered by CDF analysis). Therefore, in this analysis, quadrants Q1, Q3, and Q4 contain proteins predicted to be disordered as a whole by at least one of the methods, with disordered proteins located in quadrants Q3 and Q4 being further classified as proteins with compact/mixed or extended disorder, respectively.

Overall, of 871 autophagy-related proteins analyzed in this study, 467 (54%) are predicted to be ordered by both CH and CDF predictors. Fig. 2B also shows that the autophagy-related non-nuclear mouse proteins are distributed within the four quadrants of the CH-CDF plot as follows: 0% (Q1), 54.1% (Q2), 19.7% (Q3), and 26.2% (Q4). On the other hand, the situation is quite different for the autophagy-related nuclear mouse proteins, which are split between the four quadrants of the

CH-CDF plot in the 0.0%(Q1) : 50.4%(Q2) : 26.0%(Q3) : 23.6%(Q4) proportion. Therefore, this analysis indicates that in comparison with the autophagy-related non-nuclear mouse proteins, in a set of mouse nuclear proteins associated with autophagy, there is a noticeable decrease in the "truly" ordered proteins compensated by the noticeable increase in native molten globules, suggesting that nuclear IDPs with a molten-globule type of disorder might have some autophagy-related functions. Obviously, this interesting observation requires further analysis.

The content of mostly disordered autophagy-related proteins (*i.e.*, Q3 + Q4 proteins located within the quadrants 3 and 4 of the CH-CDF plot) in different sub-nuclear compartments ranges from 67% in chromatin to 0% in nuclear pore. By their content of mostly disordered autophagy-related proteins, the nuclear compartments can be arranged in the following order: chromatin (67%) > nuclear speckle (63%) > non-nuclear (46%) > nucleolus (45%) > PML nuclear bodies (44%) = nuclear lamina (44%) > nuclear pore (0%). Therefore, with the exception for proteins found in nuclear pores, a significant fraction of autophagy-related proteins is expected to be either extended IDPs (behave as native coils or native pre-molten globules) or the potential native molten globules/hybrid proteins.

Altogether, our results that employ multiple methodologies clearly and consistently demonstrate that in the mouse cells, the autophagy-related intra-nuclear compartments (with the exception of the nuclear pore) are enriched in IDPs, with particularly high prevalence of disorder in the nuclear speckle and chromatin.

To evaluate the evolutionary conservation of the observed phenomenon (where many nuclear autophagy-related proteins are enriched in IDPRs), we collected the autophagy related proteins in the proteome of *Saccharomyces cerevisiae*. First, we collected the complete and reviewed yeast proteome (6721 proteins) from UniProt. The same as for mouse, we extracted 112 autophagy related yeast proteins from the Autophagy Database which we mapped into the complete proteome. Next, we used the GO (gene ontology) annotations in the UniProt to annotate nuclear proteins in yeast (we could not use Nsort/DB since it is specific to mouse). We generated a list of GO cellular component terms that are associated with nucleus; they include the term nucleus and its child nodes in the ontology. Yeast proteins that include at least one of these terms are assumed to be localized in the nucleus. Overall, we found 2040 nuclear proteins in yeast, and 27 nuclear proteins among the 112 autophagy-related yeast proteins. Using the GO-based annotations we were able to associate only four autophagy related proteins with specific intra-nuclear localizations: two in chromatin, one in the nuclear pore and one in the nucleolus. Therefore, we decided not to split nuclear yeast proteins according to their subnuclear locations and analyzed them as two sets, autophagy-relate and unrelated yeast nuclear proteins (see Fig. 1). Fig. 1A shows that contrarily to mouse, yeast contains relatively less autophagy-related proteins in the nucleus than in the cytoplasm. Next, we compared disorder content and fractions of disordered proteins between the autophagy-related proteins and all considered yeast proteins. Fig. 1B shows that irrespectively of their relation to autophagy, nuclear proteins in yeast are expected to be significantly more disordered than the non-nuclear proteins. Furthermore, in the yeast nucleus, autophagy-related proteins were predicted to be noticeably more disordered than other nuclear proteins. Next, we compared various yeast datasets non-nuclear autophagy unrelated (set 1, 4681 proteins), non-nuclear autophagy related (set 2, 85 proteins), nuclear autophagy unrelated (set 3, 2040 proteins), and nuclear autophagy related (set 4, 27 proteins), looking at percentages of their proteins that have at least one, three, or five long IDPRs, and also calculated the number of long IDPRs per 1000 residues. According to this analysis, set 1 contains 27%, 6%, and 1% of proteins with at least one, three, or five long IDPRs, whereas in sets 2, 3, and 4, the corresponding percentages are: 49%, 7%, and 1%; 58%, 15%, and 4%; and 78%, 26% and 11%, respectively. In these four datasets, the average numbers of long IDPRs per 1000 residues are 1.23, 1.25, 2.05, and 1.92. Therefore, these data clearly show that the most disordered proteins in yeast are nuclear autophagy related proteins. Fig. 1B also shows that there is a proportional increase in the disorder content among nuclear and non-nuclear proteins related and unrelated to autophagy while moving from yeast to mouse. In fact, disorder content of nuclear proteins in either yeast or mouse is ~2-fold higher than the disorder level of their cytoplasmic proteins. This evolutionary conserved enrichment in disorder content suggests the functional importance of nuclear IDPs.

### Analysis of the autophagy-related transcription factors and nuclear IDPs associated with autophagy

Next, we analyzed a set of transcription factors involved in the regulation of autophagy.[51] The results of the analysis of the disorder status of these transcription factors are summarized in Table 2, which clearly shows that all these proteins are predicted to have high contents of intrinsic disorder and contain numerous functional IDPRs. These observations are in good agreement with the results of previous large-scale computational studies showing that intrinsic disorder is crucial for the function of transcription factors.[63,64] Furthermore, several proteins included in Table 2, such as p53,[38,65–67] p63,[66] p73,[66] CHOP,[68,69] NF-κB,[67] β-catenin,[70,71] and FOXO3,[72,73] represent well-characterized examples of functional IDPs.

By applying a set of computational tools we revealed that two important transcription factors, TFEB and ZKSCAN3, which are known to be located at the center of the transcriptional network that regulates autophagy,[51] are expected to be highly disordered, contain numerous disorder-based binding sites and various post-translational modifications, and to be involved in a multitude of interactions with numerous proteins (see Fig. S1 and S2, ESI†). Also, our analysis revealed that not all autophagy-related nuclear IDPs are transcription factors, and that such autophagy-related nuclear IDPs can be found in all nuclear compartments.

To illustrate this point, Table 3 represents some peculiarities of functional intrinsic disorder in illustrative representatives of the autophagy-related nuclear IDPs. Here, these proteins are grouped according to their association with various nuclear compartments. Table 3 shows that nuclear IDPs associated with

 

Table 2 Peculiarities of functional intrinsic disorder in mouse transcription factors involved in the regulation of autophagy

| Name (UniProt ID)[a,b] | Core autophagy genes regulated at the transcriptional level[a] | Effect on autophagy[a] | Length | pI | MobiDB consensus[d] | PONDR VSL2[e] | $N_{LIDR}$[f] (PLD[g]) | $N_{AIBS}$[h] |
|---|---|---|---|---|---|---|---|---|
| CHOP (P35639) | *ATG5* and *LC3* | Enhanced | 168 | 4.65 | 83.93 | 86.31 | 1 (61.90) | 5 |
| FOXO3 (Q9WVH4) | *ATG4, ATG12, BECN1, BNIP3, LC3, ULK1, ULK2, and VPS34* | Enhanced or suppressed | 672 | 4.92 | 72.62 | 85.86 | 3 (21.58) | 16 |
| FOXO1 (Q9R1E0) | *ATG5, ATG12, ATG14, BECN1, BNIP3, LC3, and VPS34* | Enhanced | 652 | 6.47 | 69.48 | 71.47 | 3 (24.39) | 20 |
| TFEB (Q9R210) | *ATG4, ATG9, BCL2, LC3, SQSTM1, UVRAG, and WIPI* | Enhanced | 475 | 5.86 | 64.63 | 87.79 | 3 (21.11) | 15 |
| ATF5 (O70191) | mTOR | Suppressed | 283 | 4.87 | 61.48 | 85.51 | 2 (33.92) | 6 |
| ATF4 (Q06507) | *ATG5*, *BH3*-only *LC3*, and *ULK1* | Enhanced | 349 | 4.71 | 58.74 | 91.69 | 1 (12.89) | 10 |
| SOX2 (P48432) | ATG10 | Enhanced | 319 | 9.81 | 51.41 | 100.00 | 1 (13.17) | 10 |
| JUN (P05627) | *BECN1 and LC3* | Enhanced | 334 | 8.88 | 45.21 | 77.84 | 1 (30.84) | 7 |
| p73 (Q9JJP2) | *ATG5, ATG7, and UVRAG* | Enhanced | 631 | 6.49 | 43.42 | 59.43 | 2 (8.08) | 11 |
| p63 (O88898) | *ATG3, ATG4, ATG5, ATG7, ATG9, ATG10, BECN1, LC3, and ULK1* | Enhanced | 680 | 6.19 | 40.00 | 59.12 | 3 (15.59) | 9 |
| HIF1 (Q61221) | *BNIP3* | Enhanced | 836 | 5.14 | 39.00 | 60.53 | 3 (12.56) | 15 |
| ZKSCAN3 (Q91VW9) | *LC3*,[c] *ULK1*,[c] and *WIPI* | Suppressed | 553 | 5.66 | 38.89 | 75.59 | 4 (58.95) | 10 |
| GATA1 (P17679) | *LC3* | Enhanced | 413 | 8.91 | 38.26 | 64.89 | 1 (5.08) | 6 |
| E2F1 (Q61501) | *ATG5, BNIP3, LC3, and ULK1* | Enhanced | 430 | 4.92 | 36.05 | 63.95 | 2 (16.28) | 9 |
| C/EBPβ (P28033) | *BNIP3, LC3, and ULK1* | Enhanced | 296 | 8.78 | 34.12 | 89.53 | 1 (8.11) | 7 |
| NF-κB (P25799, Q9WTK5) | *BCL2, BECN1, BNIP3*,[c] and *SQSTM1* | Enhanced or suppressed | 971 | 5.20 | 17.20 | 33.26 | 1 (2.68) | 6 |
| | | | 899 | 5.93 | 32.59 | 60.29 | 3 (13.01) | 16 |
| p53 (P02340) | *ATG2*, *ATG4*, *ATG7*, *ATG10*, *BCL2*,[c] *BH3*-only, *ULK1*, and *UVRAG* | Enhanced (nucleus) or suppressed (cytosole) | 387 | 6.83 | 30.23 | 64.60 | 1 (11.11) | 6 |
| SREBP2 (Q3U1N2) | *LC3*, *ATG4B*, and *ATG4D* | Enhanced | 1130 | 8.75 | 18.32 | 45.22 | 1 (3.98) | 6 |
| β-catenin (Q02248) | *SQSTM1*[c] | Suppressed | 781 | 5.53 | 17.31 | 29.96 | 2 (14.98) | 9 |
| STAT1 (P42225) | *ATG12*[c] and *BECN1*[c] | Suppressed | 749 | 5.42 | 11.48 | 37.12 | 4 (29.91)[i] | 4 |
| STAT3 (P42227) | *ATG3*, *BCL2*, and *BNIP3* | Suppressed | 770 | 5.94 | 9.74 | 32.34 | 2 (19.61)[i] | 3 |

[a] Data are taken from ref. 51. [b] Abbreviated names (in order of appearance): ATF, activating transcription factor; *BECN1*, beclin 1; *BH3*-only, BCL-2 homology 3-only; C/EBPβ, CCAAT/enhancer-binding protein-β; CHOP, C/EBP-homologous protein; FOXO, forkhead box O; HIF1, hypoxia-inducible factor 1; *LC3*, light chain 3; NF-κB, nuclear factor-κB; RB1CC1, RB1-inducible coiled-coil 1; SOX2, SRY box-containing factor 2; *SQSTM1*, sequestosome 1; SREBP2, sterol regulatory element binding protein 2; STAT, signal transducer and activator of transcription; TFEB, transcription factor EB; *ULK*, UNC-51-like kinase; *UVRAG*, ultraviolet irradiation resistance-associated gene; *VPS34*, vacuolar protein sorting 34; *WIPI*, WD repeat domain, phosphoinositide interacting; ZKSCAN3, zinc-finger protein with KRAB and SCAN domains 3. [c] Genes are upregulated by the specified transcription factor unless marked, which indicates that the transcription factor downregulates them. [d] Content of predicted disordered residues in a given protein based on the MobiDB consensus score.[74,75] [e] Content of predicted disordered residues in a given protein based on the PONDR® VSL2 analysis. [f] $N_{LIDR}$, number of long disordered regions in a protein based MobiDB consensus analysis.[74,75] [g] PLD, percent of long disorder calculated as content of predicted disordered residues in long disordered regions in a given protein based on the MobiDB consensus analysis.[74,75] [h] $N_{AIBS}$, number of ANCHOR identified binding sites.[76,77] [i] Values calculated based on the results of the PONDR® VSL2 analysis.

autophagy have numerous functionally important IDPRs that are involved in protein–protein interactions. Curiously, even within the intra-nuclear compartments containing the least disordered autophagy-related proteins (*e.g.*, nuclear pore and perinucleolar compartment) these proteins still have numerous AIBSs, emphasizing their binding promiscuity. Also, two highly disordered autophagy-related nuclear proteins (Akap8 and RING1, see Table 3) were found to be associated with two different nuclear compartments (chromatin and nuclear speckles).

### Focused look at the illustrative examples of nuclear IDPs associated with autophagy

For all proteins listed in Table 3, we evaluated their per-residue intrinsic disorder propensities using PONDR® VLXT,[21] PONDR® VSL2,[60] PONDR® VL3,[78] and PONDR® FIT,[79] retrieved information on functional disorder in these proteins from the $D^2P^2$ database (http://d2p2.pro/),[30] and finally looked at the interactivity of these autophagy-related nuclear proteins using the STRING database.[80] The results of these analyses for eight autophagy-related nuclear proteins (one illustrative example for each membrane-less nuclear organelle considered in this study) are discussed below, whereas analogous data for the remaining proteins from Table 3 are shown in the ESI† (Fig. S3–S14).

### Cajal body: ataxin-2-like protein A2LP

Fig. 3 represents the results of computational analysis of the prevalence and functionality of intrinsic disorder in ataxin-2-like protein (A2LP, UniProt ID: Q7TQH0) found in Cajal body. According to this analysis, mouse A2LP is predicted to be a highly disordered protein (see Fig. 3A and B) that is heavily decorated with various post-translational modifications (PTMs) and has numerous disorder-based interaction sites (see Fig. 3B) and is shown to be involved in multiple protein–protein interactions (see Fig. 3C). The presence of multiple PTM sites[81,82] and binding promiscuity[21,24,36,37,65,83] are well-known functional features characteristic of IDPs.

A2LP is a paralogue of ataxin-2, which is a disease-causing protein involved in the pathogenesis of spinocerebellar ataxia type 2. Mouse protein has a proline-rich N-terminal region (residues 4–61). It has been established that human A2LP is able to interact with ataxin-2 itself and with some of the known binding partners of ataxin-2, such as the poly(A)-binding protein and the RNA helicase DDX6.[84] It is also known that mouse

Table 3 Peculiarities of functional intrinsic disorder in illustrative representatives of mouse autophagy-related nuclear IDPs

| Name (UniProt ID)[a] | Length | pI | MobiDB consensus[b] | PONDR VSL2[c] | $N_{LIDR}$[d] (PLD[e]) | $N_{AIBS}$[f] |
|---|---|---|---|---|---|---|
| Cajal body | | | | | | |
| A2LP (Q7TQH0) | 1049 | 8.94 | 77.22 | 91.13 | 6 (53.00) | 26 |
| Chromatin | | | | | | |
| α-NAC (P70670) | 2187 | 9.39 | 85.32 | 97.12 | 8 (55.05) | 65 |
| Champ1 (Q8K327) | 802 | 8.08 | 71.07 | 81.67 | 5 (56.61) | 11 |
| Piccolo (Q9QYX7) | 5068 | 6.10 | 63.24 | 84.65 | 17 (38.32) | 96 |
| Akap8 (Q9DBR0) | 687 | 5.03 | N.D. | 91.99 | 5 (93.67) | 15 |
| RING1 (O35730) | 406 | 5.54 | 51.23 | 65.02 | 2 (30.05) | 5 |
| Nuclear lamina | | | | | | |
| Cux1 (P53564) | 1515 | 6.02 | 62.05 | 84.09 | 6 (36.57) | 30 |
| Nuclear pore | | | | | | |
| SRP1-β (Q60960) | 538 | 4.93 | 14.68 | 31.41 | 1 (11.90) | 3 |
| Qip2 (O35344) | 521 | 4.80 | 14.97 | 26.30 | 1 (5.57) | 3 |
| Pendulin (P52293) | 529 | 5.49 | 14.37 | 27.98 | 2 (9.83) | 5 |
| Nuclear speckle | | | | | | |
| Bud13 (Q8R149) | 637 | 9.94 | 92.31 | 94.98 | 4 (73.94) | 14 |
| Srrm1 (Q52KI8) | 946 | 11.87 | 87.53 | 90.70 | 1 (82.45) | 17 |
| Pabp2 (Q8CCS6) | 302 | 5.13 | 59.93 | 76.82 | 1 (36.75) | 5 |
| Jip1 (Q9WVI9) | 707 | 4.84 | 54.15 | 68.74 | 3 (47.38) | 13 |
| Akap8 (Q9DBR0) | 687 | 5.03 | N.D. | 91.99 | 5 (93.67) | 15 |
| RING1 (O35730) | 406 | 5.54 | 51.23 | 65.02 | 2 (30.05) | 5 |
| Nucleolus | | | | | | |
| Tp2 (P11378) | 117 | 11.80 | 56.15 | 68.74 | 3 (47.38) | 3 |
| Nkx-3.1 (P97436) | 237 | 8.98 | 56.12 | 80.59 | 1 (40.51) | 5 |
| Eps15R (Q60902) | 907 | 4.86 | 57.11 | 71.44 | 2 (29.88) | 13 |
| Perinuclear compartment | | | | | | |
| Sin3a (Q60520) | 1274 | 6.82 | 30.14 | 48.90 | 5 (15.23) | 20 |
| PML | | | | | | |
| Grip1 (Q61026) | 1462 | 6.22 | 67.37 | 87.14 | 4 (15.39) | 28 |
| Numa1 (Q80Y35) | 2094 | 5.68 | 61.03 | 93.27 | 7 (23.40) | 44 |

[a] Abbreviated names (in order of appearance): A2LP, ataxin-2-like protein; α-NAC, nascent polypeptide-associated complex subunit alpha, muscle-specific form; Champ1, chromosome alignment-maintaining phosphoprotein 1; Piccolo, protein piccolo; Akap8, a-kinase anchor protein 8; RING1, E3 ubiquitin-protein ligase RING1; Cux1, homeobox protein cut-like 1; SRP1-β, importin subunit α-5; Qip2, importin subunit α-4; pendulin, importin subunit α-1; Srrm1, serine/arginine repetitive matrix protein 1; Pabp2, polyadenylate-binding protein 2; Jip1, C-Jun-amino-terminal kinase-interacting protein 1; Tp2, nuclear transition protein 2; Nkx-3.1, homeobox protein Nkx-3.1; Eps15R, epidermal growth factor receptor substrate 15-like 1; Sin3a, paired amphipathic helix protein Sin3a; Grip1, nuclear receptor coactivator 2; Numa1, nuclear mitotic apparatus protein 1. [b] Content of predicted disordered residues in a given protein based on the MobiDB consensus score.[74,75] [c] Content of predicted disordered residues in a given protein based on the PONDR® VSL2 analysis. [d] $N_{LIDR}$, number of long disordered regions in a protein based on the MobiDB consensus analysis.[74,75] [e] PLD, percent of long disorder calculated as content of predicted disordered residues in long disordered regions in a given protein based on the MobiDB consensus analysis.[74,75] [f] $N_{AIBS}$, number of ANCHOR identified binding sites.[76,77]

protein is phosphorylated at Ser27, Thr45, Ser109, Ser236, Ser407, Ser496, Ser499, Ser560, Ser597, and Ser637,[85] and is acetylated at the residue Lys205.[86] Human A2LP is involved in the regulation of stress granules and processing bodies.[84] Furthermore, in addition to the canonical form comprising 1049 residues, two more forms are produced by alternative splicing (AS). In one of these AS isoforms, the 439–444 region is missing, whereas in the second AS form, in addition to the removal of this region, the C-terminal 16 residues are changed to the GEQPGQAPGFPGGADDRIREFS LAGGIWHGRAEGLQVGQDA RVLGGD sequence. Also, residues 98–121 of human protein (which correspond to residues 96–119 in mouse A2LP) are involved in interaction with a cytokine receptor Mpl, thereby activating multiple downstream signal transduction pathways.[87]

Although currently available structural and functional information about A2LP is rather limited, some of the results of bioinformatics analysis are supported by experimental evidence. In fact, the indicated Mpl-binding region of mouse A2PL coincides with one of its 25 predicted disorder-based binding sites, whereas all the aforementioned PTM sites are located within the disordered regions of this protein. Also, earlier studies revealed that alternative splicing occurs mostly in regions of RNA that code for the disordered protein regions.[88,89]

### Chromatin: nascent polypeptide-associated complex subunit α

Fig. 4 depicts the results of the evaluation of the muscle-specific form of α-NAC protein (skNAC, UniProt ID: P70670) for intrinsic disorder and shows that this protein is expected to be highly disordered, have several isoforms produced by AS, and possess numerous PTM sites and disorder-based binding sites defining its ability to interact with various partners. Due to the length of this and several other proteins discussed below and due to the

 

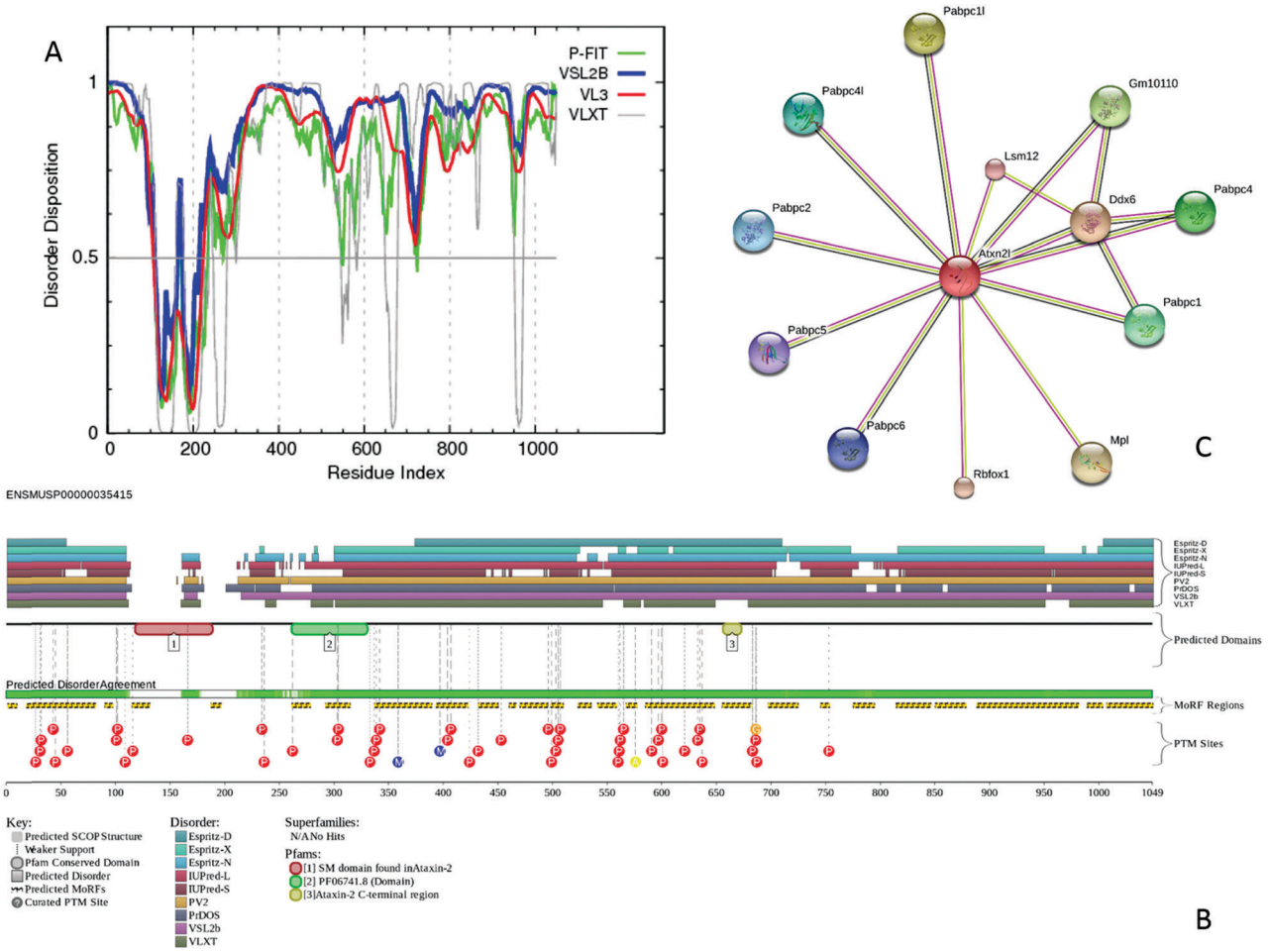


Fig. 3 Prevalence and functionality of intrinsic disorder in A2LP protein (UniProt ID: Q7TQH0) from Cajal body. (A) Evaluation of the per-residue disorder propensity by the members of the PONDR family of disorder predictors. A disorder threshold is indicated as a thin line (at score of 0.5) to show a boundary between disorder (>0.5) and order (<0.5). (B) Evaluation of the functional intrinsic disorder propensity by the $D^2P^2$ database (http://d2p2.pro/).[30] In the corresponding plot, top nine colored bars represent location of disordered regions predicted by different disorder predictors (Espritz-D, Espritz-N, Espritz-X, IUPred-L, IUPred-S, PV2, PrDOS, PONDR® VSL2b, and PONDR® VLXT, see keys for the corresponding color codes). Green-and-white bar in the middle of the plot shows the predicted disorder agreement between these nine predictors, with green parts corresponding to disordered regions by consensus. Yellow bar shows the location of the predicted disorder-based binding site (MoRF region), whereas colored circles at the bottom of the plot show locations of various PTM sites. (C) Analysis of the A2LP interactivity by STRING computational platform.[80] STRING produces the network of predicted associations for a particular protein and its interactome. The network nodes are proteins, whereas the edges represent the predicted or known functional associations. There are seven types of evidence used in predicting the associations which are indicated in the resulting network by the differently colored lines, where a red line indicates the presence of fusion evidence; a green line – neighborhood evidence; a blue line – co-occurrence evidence; a purple line – experimental evidence; a yellow line – text mining evidence; a light blue line – database evidence; a black line – co-expression evidence.[80]

high informational density of the corresponding STRING and $D^2P^2$ plots, it is almost impossible to extract exact information from these graphs. One should keep in mind though that the goal of the corresponding figures is to show how common is disorder in these proteins and what it can be used for. We expect that the interested readers will be able to obtain required information from the interactive plots available for these proteins at the STRING http://string-db.org/ and $D^2P^2$ (http://d2p2.pro/) websites.

α-NAC is the α-subunit of the nascent polypeptide-associated complex (NAC). In yeast, together with the Hsp40/70-based chaperone system RAC-Ssb, NAC is involved in assisting cotranslational folding of nascent polypeptides at the ribosome.[90] In archaea, NAC is a homodimeric complex formed by two α-NAC subunits, whereas in eukaryotes, a stable αβ-NAC heterodimer is found. NAC serves as a dynamic component of the ribosomal exit tunnel that ensures appropriate nascent protein targeting by protecting the emerging polypeptides from interacting with other cytoplasmic proteins.[90] In mice, α-NAC serves as a cardiac- and muscle-specific transcription factor.[91]

It was shown that a muscle-specific, proline-rich isoform of α-NAC termed as skNAC is generated by an alternative splicing-in of a 6.0 kb-exon.[91] skNAC serves as a tissue-specific DNA-binding activator and participates in normal myogenic differentiation as well as in the regulation of myoblast fusion.[91] Also, this muscle-specific transcription factor skNAC interacts with muscle-restricted histone methyltransferase Smyd1 and is involved in cardiac development and skeletal muscle growth and regeneration,[92]

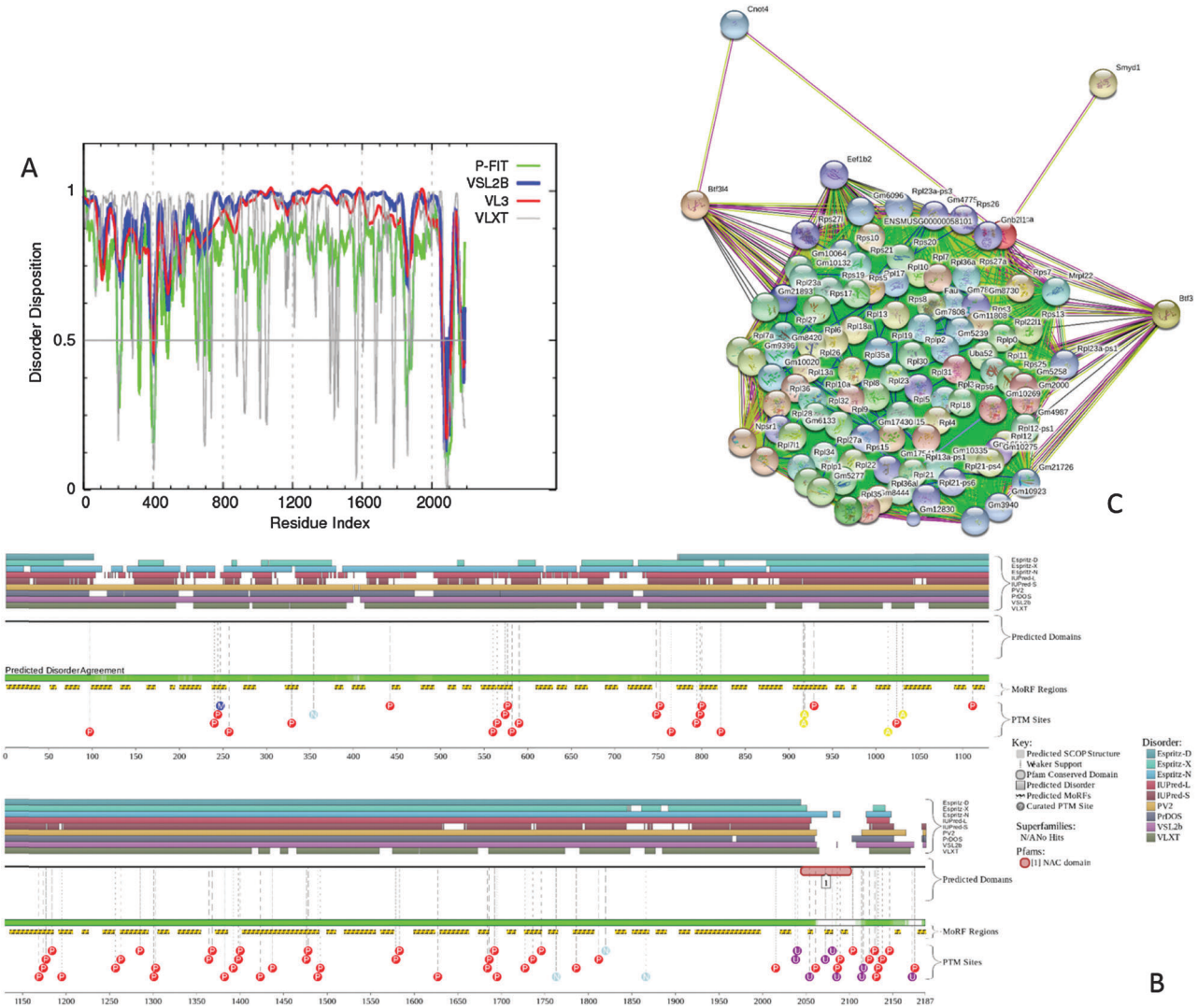


Fig. 4 Prevalence and functionality of intrinsic disorder in α-NAC protein (UniProt ID: P70670) from chromatin. (A) Evaluation of the per-residue disorder propensity by the members of the PONDR family of disorder predictors. (B) Evaluation of the functional intrinsic disorder propensity by the $D^2P^2$ database (http://d2p2.pro/). (C) Analysis of the α-NAC interactivity by STRING computational platform.

whereas skNAC depletion stimulates myoblast migration and perturbs sarcomerogenesis *via* enhancement of the activity of calpains 1 and 3.[93] skNAC can be found both in the cytoplasm and in the nucleus.[94] Also, *via* its PXLXP motif skNAC is able to interact with the evolutionarily conserved MYND domain of the m-Bop repressor protein, which is essential for cardiogenesis,[94] as well as with other MYND-containing transcriptional regulators linked to development, chromatin stability, and cancer.[94] High predisposition of skNAC for intrinsic disorder is not surprising since many transcription factors are known to be highly disordered.[63]

### Nuclear lamina: homeobox protein cut-like 1 (Cux1)

The results of the multiparametric disorder analysis of the homeobox protein cut-like 1 (Cux1, UniProt ID: P53564) from nuclear lamina are shown in Fig. 5, which clearly illustrate that Cux1 is predicted to be highly disordered (Fig. 5A) and can be involved in a multitude of interactions with various binding partners (Fig. 5B).

Because Cux1 is a transcription factor (see below), it was expected that it will be highly disordered. Although the results of $D^2P^2$-based analysis are not available for this protein as of yet, Table 3 shows that it is predicted to have 30 disorder-based binding sites. Furthermore, there are four DNA binding domains in mouse Cux1, residues 540–627, 929–1016, 1112–1199, and 1239–1298. Also, it is expected to be phosphorylated at Ser427, Ser761, Ser901, Ser1064, Ser1332, and Ser1506. There are six isoforms of Cux1 produced by alternative splicing: a canonical isoform (1515 residues); isoform 1 (678 residues) is known as the cytohesin-associated scaffolding protein (CASP) and has first 420 residues similar to those of isoform 3. In isoform 2 (1413 residues), residues 406–507 are missing.

In isoform 3 (1504 residues), residues 630–651 are missing and the N-terminal 10 residues are changed to the

 

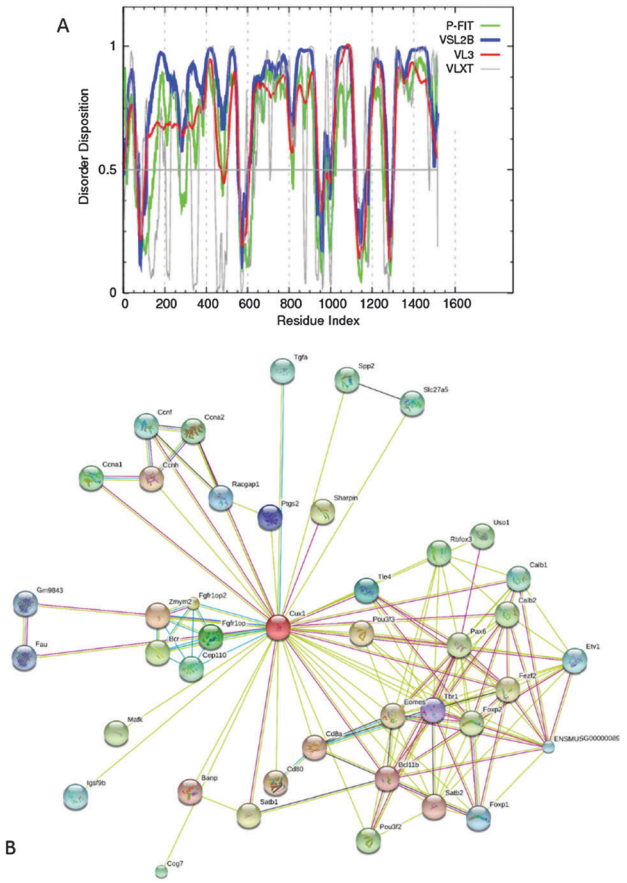


Fig. 5 Prevalence and functionality of intrinsic disorder in Cux1 protein (UniProt ID: P53564) from nuclear lamina. (A) Evaluation of the per-residue disorder propensity by the members of the PONDR family of disorder predictors. (B) Analysis of the Cux1 interactivity by STRING computational platform.

MAANVGSMFQYWKRFDLQQLQ sequence; isoforms 4 and 5 (1332 and 460 residues, respectively) are characterized by missing regions 1–183 and 1–1055, respectively. The canonical form of the mouse protein has a coiled–coil domain (residues 56–361) and the alanine-rich domain (residues 1358–1463).

On the functional site, Cux1 serves as a repressor of developmentally regulated gene expression.[95,96] Cux1, also known as the CCAAT displacement protein (CDP), was shown to act as a transcriptional repressor in various processes, ranging from terminal differentiation, to cell cycle progression, to DNA replication, to the spindle assembly checkpoint that controls progression through mitosis, and to the control of nuclear matrix attachment regions,[95,97] likely due to the ability of this transcription factor to regulate a large number of genes and microRNAs involved in multiple cellular processes.[97] CUX1 is found within a multi-protein complex known as the histone nuclear factor D (HiNF-D).[98] In humans, the CASP isoform of Cux1 acts as an important scaffold protein needed for the assembly and recruitment of various protein complexes associated with intracellular trafficking and signaling and plays a role in immunity.[99] Changes in the expression levels of various Cux1 AS isoforms were observed in many cancers, where this protein plays a role in the stimulation of cell migration and invasion, as well as tumor initiation and progression.[97] Besides carcinogenesis, Cux1 is involved in the etiology of polycystic kidney diseases, cognitive disorders, neurodegeneration, and obesity.[97]

### Nuclear pores: importin subunit α5 (SRP1-β)

Fig. 6 summarizes the results of computational analysis of the α5 subunit of importin (UniProt ID: Q60960), also known as importin α-S1, karyopherin subunit α-1 (Kpnα1), nucleoprotein interactor 1 (NPI-1), RAG cohort protein 2 (RCH2), or SRP1-β. It is seen that although SRP1-β is one of the most ordered proteins considered in our study, it has a long, highly disordered N-terminal tail (residues 1–120, Fig. 6A), which contains three disorder-based binding sites (Fig. 6B). SRP1-β is involved in a wide network of protein–protein interactions (Fig. 6C), and binding promiscuity helps SRP1-β to serve as an adapter protein for the nuclear receptor karyopherin subunit β-1 (Kpnβ1) and to bind specifically and directly to a multitude of protein-substrates containing nuclear localization signal (NLS) motifs.[100]

Being an active substrate recruiter of the karyopherin-nuclear pore complex (NPC) machinery, SRP1-β is needed for targeting specific cargos for nucleo-cytoplasmic transport, and for the relocation of essential targets into and out of the nucleus.[102] Since both NLS-binding sites (major, residues 149–241, and minor, residues 318–406) are located within the mostly ordered part of SRP1-β, it is likely that these localities serve as docking sites for the intrinsically disordered NLS motifs.

### Nuclear speckles: Bud13 homolog protein

Fig. 7 shows the PONDR-based disorder profile, $D^2P^2$ plot, and STRING-based interactome of the nuclear speckle-located Bud13 homolog protein (UniProt ID: Q8R149), which, in yeast, is a member of the precursor mRNA (pre-mRNA) retention and splicing (RES) complex.[101] This RES complex is composed of at least three proteins, snRNP-associated protein 17 (Snu17, also known as Ist3), pre-mRNA–leakage protein 1 (Pml1), and bud site–selection protein 13 (Bud13).[103] It has been shown that the entire Bud13 and the N-terminal third of Pml1 are intrinsically disordered in isolation.[104] Fig. 7A and B show that in agreement with these experimental data for the yeast Bud13, the mouse protein is predicted to be mostly disordered.

Therefore, if these extensions and insertions would be excluded, the actual sequence identity for these two proteins would be >60%. Curiously, the C-terminally located UHM-ligand motif (ULM) of yeast Bud13 is highly similar to the C-terminal region of mouse protein that contains two disorder-based binding sites (see Fig. 7B). Altogether, mouse Bud13 is predicted to have 14 such disorder-based interaction sites and therefore is expected to have a multitude of binding partners. In agreement with this hypothesis, Fig. 7D shows that this protein is indeed a promiscuous binder. Furthermore, according to Fig. 7B, functionality of Bud13 is regulated by intensive phosphorylation and acetylation. Importantly, all sites of posttranslational modifications are located within the intrinsically disordered regions. This is in agreement with the well-known fact that phosphorylation[81] and many other enzymatically catalyzed PTMs are preferentially located within the IDPRs.[82]

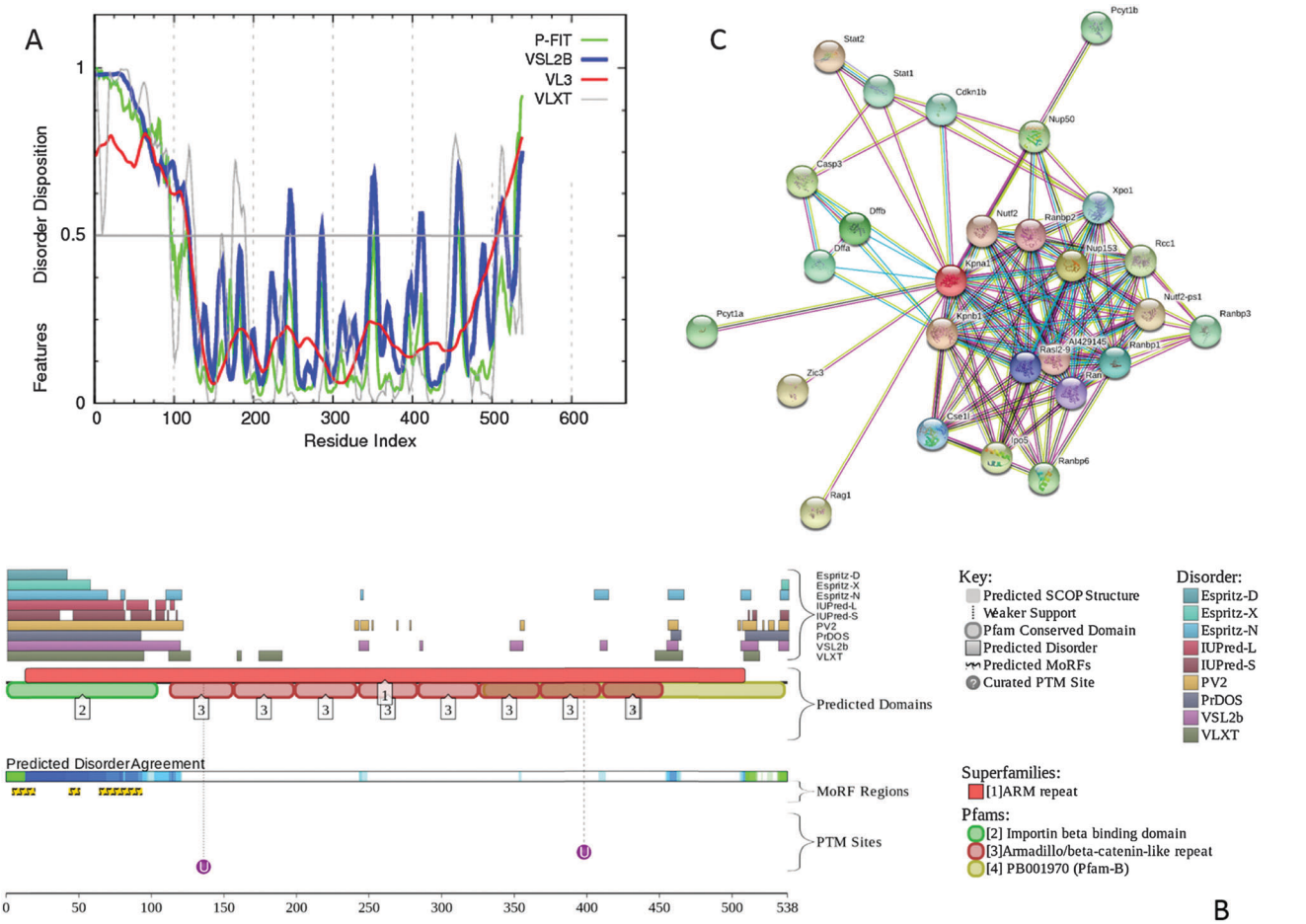


Fig. 6 Prevalence and functionality of intrinsic disorder in SRP1-β protein (UniProt ID: Q60960) from nuclear pore. (A) Evaluation of the per-residue disorder propensity by the members of the PONDR family of disorder predictors. (B) Evaluation of the functional intrinsic disorder propensity by the $D^2P^2$ database (http://d2p2.pro/). (C) Analysis of the SRP1-β interactivity by STRING computational platform.

Bud13 binds to Snu17 with nanomolar affinity.[104] The formation of this heterodimeric complex involves an interaction between the U2AF-homology motif (UHM) located within the RNA-recognition motif (RRM) domain of Snu17 and the C-terminal UHM-ligand motif (ULM) of Bud13.[104] The 50 residue-long IDPR located at the N-terminus of Pml1 is responsible for the engagement of this protein in the RES complex.[105] Therefore, the RES core complex in yeast involves the C-terminal domain (residues 200–266) of Bud13 and residues 1–60 of Pml1 bound to the RRM domain (residues 25–138) of Snu17.[101] Fig. 7C represents the solution structure of this RES core complex and shows that even in its bound form, the C-terminal domain of Bud13 preserves significant flexibility. Curiously, in addition to the aforementioned engagement in the RES complex formation, yeast Snu17 and Bud13 are known to be directly involved in mRNA splicing,[106,107] whereas in *Caenorhabditis elegans*, Bud13 also plays a role in the regulation of embryogenesis.[108] Although Bud13 in yeast is noticeably shorter than its orthologue in mouse (266 *vs.* 637 residue), these two proteins share 25.1% identical residues. The actual similarity is even higher, since in comparison with Bud13 from yeast, mouse protein has a 299 residue-long N-terminal extension and multiple long insertions that sum up to 92 residues.

### Nucleolus: nuclear transition protein 2 (Tp2)

Fig. 8 shows that the nuclear transition protein 2 (Tp2, UniProt ID: P11378), being shortest among the proteins considered here (it has just 117 residues), is predicted to be highly disordered (Fig. 8A) and is involved in a very dense network of protein–protein interactions (Fig. 8B). Table 3 also illustrates that Tp2 has three disorder-based potential binding sites. This high binding promiscuity and high disorderedness level are directly related to the functionality of this important protein found in the nuclei of male germ cells. In fact, together with two other testis-specific transition proteins, Tp1 and Tp4, Tp2 plays a crucial role in the global chromatin remodeling process that takes place during mammalian spermiogenesis, and where haploid round spermatids are transformed into a toroidal nucleoprotamine fiber in the mature spermatozoa.[109–115] In this two-step process, 90% of histones are first replaced by the moderately basic transition proteins Tp1 (55 residues, pI 12.09, net charge +19), Tp2 (117 residues, pI 11.80, net charge +25),

 

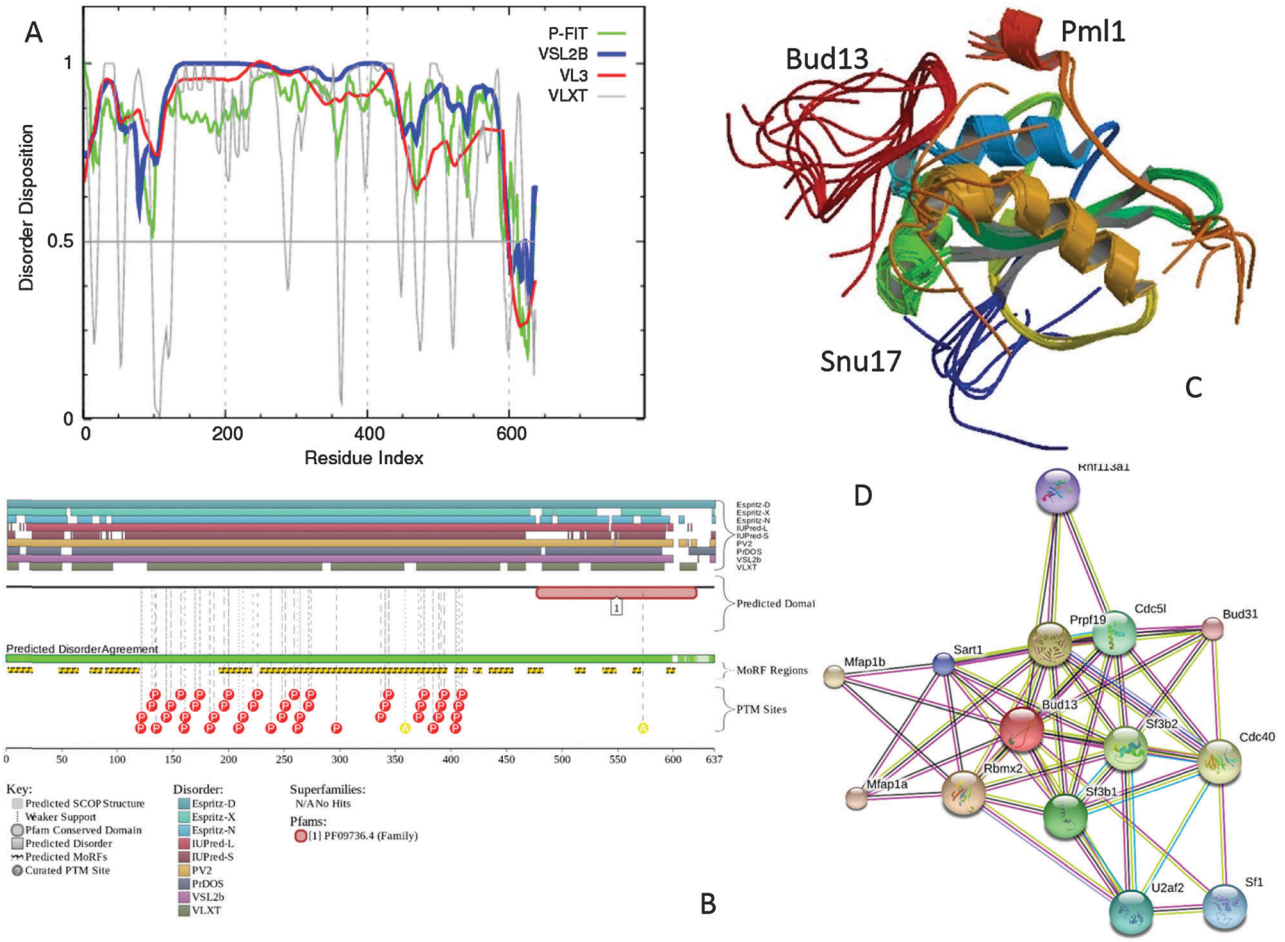


Fig. 7 Prevalence and functionality of intrinsic disorder in Bud13 (UniProt ID: Q8R149). (A) Evaluation of the per-residue disorder propensity by the members of the PONDR family of disorder predictors. (B) Evaluation of the functional intrinsic disorder propensity by the $D^2P^2$ database (http://d2p2.pro/). (C) Solution NMR structure of the RES core protein from yeast that involves C-terminal domain (residues 215–245) of Bud13 and residues 22–42 of Pml1 bound to the RRM domain (residues 25–138) of Snu17 (PDB ID: 2MKC).[101] (D) Analysis of the Bud13 interactivity by STRING computational platform.

and Tp4 (in *Sus scrofa*, 138 residues, pI 12.10, net charge +43), which at the later stage are replaced by highly basic proteins, sperm-specific protamines P1 (51 residues, pI 12.07, net charge +31) and P2 (107 residues, pI 12.05, net charge +33) to form a highly condensed sperm chromatin.[109–116] Recently it has been established that Tp1 and Tp2 from rat have 16 and 19 PTMs, respectively, and that Tp2 is methylated at Arg71, Arg75, and Arg92 residues by the arginine methyltransferase PRMT4 (CARM1), and is methylated at Lys88 and Lys91 residues by the lysine methyltransferase KMT7 (Set9).[116] In support of this Tp2 from rat was shown to be a zinc-metalloprotein, whose intrinsically disordered structure in the unbound form was moderately perturbed by interaction with zinc ions.[117]

### Perinucleolar compartment: paired amphipathic helix protein Sin3a

The results of the functional disorder evaluation in the paired amphipathic helix protein Sin3a (UniProt ID: Q60520) are shown in Fig. 9. It is seen that similar to the aforementioned karyopherin subunit SRP1-β from nuclear pore, Sin3a belongs to the category of hybrid proteins[118] that possess both functional IDPRs and ordered domains. Fig. 9A and B show that both N- and C-terminal domains of this protein are substantially disordered (residues 1–450 and 1050–1150, respectively). There are also several shorter disordered regions in the middle part of Sin3a. According to Fig. 9B, disordered regions of Sin3a contain numerous PTM sites. Fig. 9C shows that Sin3A is involved in numerous protein–protein interactions, whereas Fig. 9B illustrates that several potential binding sites are located within the terminally and centrally located IDPRs. Curiously, many of these predicted binding sites overlap or coincide with known binding regions of Sin3a. For example, residues 119–196 are involved in interaction with HCFC1, residues 205–479 bind to REST, whereas residues 688–830 are responsible for interaction with HDAC1 and ARID4B.

Sin3A serves as a co-repressor required for the regulation of functions of the Mad proteins (which are basic region-helix-loop-helix-leucine zipper (bHLHZip) proteins that act as transcriptional repressors and antagonize the transcriptional and transforming activity of the Myc proto-oncogenes[119–121]), where the Mad-Max transcriptional repression is mediated by the formation of the ternary complex with Sin3A,[122,123] and where

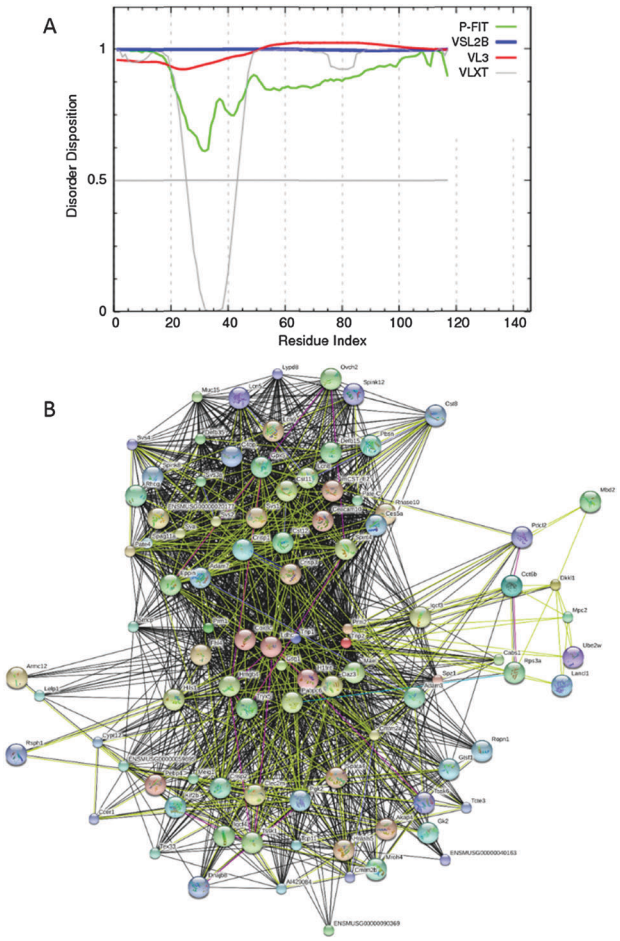


Fig. 8 Prevalence and functionality of intrinsic disorder in Tp2 protein (UniProt ID: P11378) from nucleolus. (A) Evaluation of the per-residue disorder propensity by the members of the PONDR family of disorder predictors. (B) Analysis of the Tp2 interactivity by STRING computational platform.

interaction with Sin3A is critical for the function of the Mad proteins as transcriptional repressors.[124] Sin3A also is involved in the formation of large multi-protein complexes that contain the histone deacetylases HDAC11 and HDAC2, and these histone deacetylases are needed for the mediation of the Mad-based transcriptional repression.[125,126] Additionally, these complexes contain SAP18 and SAP30 proteins, with both SAP18 and SAP30 being involved in direct interaction with Sin3A.[127] In addition to modulating the biological and transcriptional activities of Mad, the complex between Sin3A and HDAC was shown to serve as a co-repressor used by various transcriptional repressors, such as the estrogen receptor, MeCP2, Pit1, RPX, and RXR.[128–132]

Important functional units of the members of the Sin3 family are paired amphipathic α-helix (PAH) domains that contain two amphipathic α-helices separated by a flexible linker and that serve as specific protein–protein interaction domains.[122,133] In fact, PAH1 of Sin3A interacts with the repression domain of the nuclear hormone corepressor N-CoR,[128,134] PAH2 binds to Mad proteins,[122–124] and PAH3 is responsible for interaction with SAP30.[132]

### PML nuclear bodies: glucocorticoid receptor-interacting protein-1 Grip1

Glucocorticoid receptor-interacting protein-1 (Grip1, UniProt ID: Q61026) is also known as nuclear receptor coactivator-2 (NCoA-2), steroid receptor coactivator-2 (SRC2), and transcriptional intermediary factor-2 (TIF2). Fig. 10A and B shows that this transcriptional coactivator for steroid receptors and nuclear receptors is predicted to be highly disordered, and that disorder is unevenly distributed within the protein sequence, with C-terminal 1000 residues representing a highly disordered domain and with the N-terminal 400-residue-long domain containing a mixture of disordered and ordered regions. Although there are 28 potential disorder-based binding sites spread throughout the protein sequence, these sites are very dense within the disordered domain, which can be considered as a long binding platform (see Fig. 10B), and which can explain the high binding promiscuity of Grip1 and its involvement in the well-developed and highly dense interaction network (see Fig. 10C). Finally, Fig. 10C shows that functionality of Grip1 is regulated by numerous PTMs. In agreement with the prediction that many regions of Grip1 may undergo binding-induced folding, the 686–697 region folds into short α-helix upon binding to the human estrogen receptor α ligand-binding domain[135,136] or to the peroxisome proliferator-activated receptor γ (PPARγ),[137] and the 741–753 region also adopts an α-helical conformation as a result of binding to the nuclear receptor FXR,[138] or the androgen receptor,[139] or to the glucocorticoid receptor.[140]

High binding promiscuity and a wide spectrum of binding partners define the multifarious role of Grip1 in various biological processes ranging from orchestration of metabolism[141] to regulation of energy balance between white and brown adipose tissues and related coordination of energy homeostasis,[142] to control of circadian clock,[141] to regulation of fasting hepatic glucose release needed for basal brain function and survival when dietary glucose is unavailable,[143] to control of the thymic-specific retinoic acid-related orphan receptor γ (RORγ)t-regulated thymocyte survival,[144] and to regulation of chromatin structure remodeling, recruitment of RNA polymerase, and transcription activation *via* interaction with various hormone-activated nuclear receptors (NR).[145]

## Experimental

### Datasets

We collected a comprehensive set of autophagy-related proteins in mouse from the Autophagy Database.[146] We picked mouse since it is characterized by a good coverage of the autophagy-related proteins and annotations of intra-nuclear compartments.

The autophagy-related proteins include experimentally annotated proteins, orthologs of the autophagy-related proteins and proteins that share high sequence similarity to these 'reviewed' and 'orthologous' proteins (based on PSI-BLAST with 3 iterations and *E*-value $\leq 1e^{-100}$).

We collected nuclear proteins and their annotations of the intra-nuclear compartments from the Nsort/DB database[147]

 

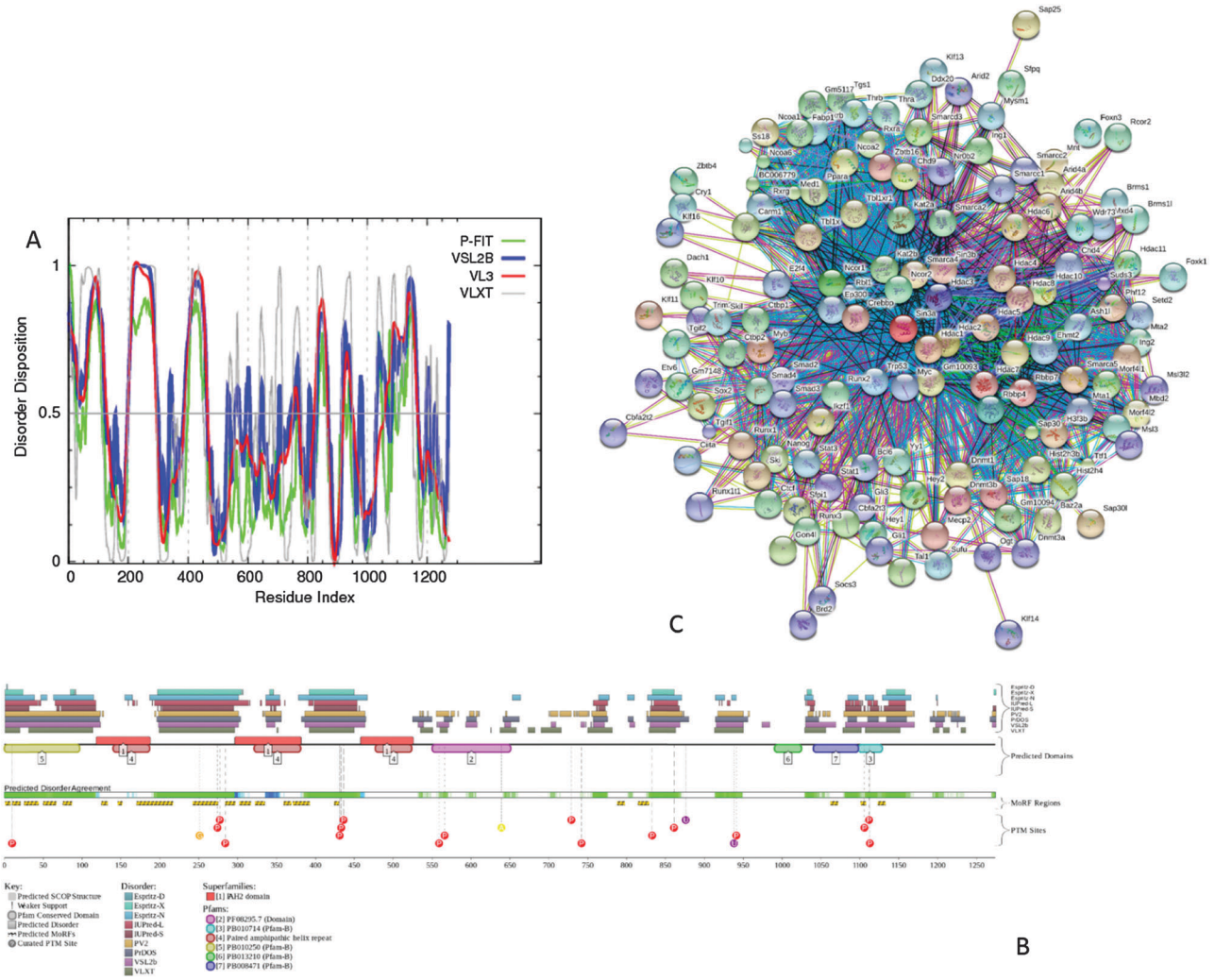


Fig. 9 Prevalence and functionality of intrinsic disorder in Sin3a protein (UniProt ID: Q60520) from Perinuclear compartment. (A) Evaluation of the per-residue disorder propensity by the members of the PONDR family of disorder predictors. (B) Evaluation of the functional intrinsic disorder propensity by the $D^2P^2$ database (http://d2p2.pro/). (C) Analysis of the Sin3a interactivity by STRING computational platform.

that combines data from a comprehensive set of sources: NPD,[148] NOPdb,[149] NMPdb,[150] UniProt,[151] and HPRD.[152] We mapped the autophagy-related and nuclear proteins into a complete mouse proteome collected from UniProt,[151] using the same procedure as in ref. 17. Briefly, mapping primarily relied on the comparison of accession numbers and in the case where these numbers were not available we matched proteins using sequence identity. This way the autophagy-related proteins are assigned to specific intra-nuclear localizations or to the set of non-nuclear proteins. We include 748 and 123 non-nuclear and nuclear autophagy-related proteins, respectively (see Table 1).

### Annotation and characterization of intrinsic disorder

We collected putative annotations of intrinsic disorder based on a majority vote consensus of two high-throughput predictors of intrinsically disordered residues. We utilized the consensus since this was demonstrated to increase predictive performance as compared to using a single predictor.[153–156] Our consensus considers a diverse set of methods that use complementary definitions of disorder and predict different types of disordered regions. We used three versions of the Espritz method[157] that predict disorder annotated based on crystal structures, NMR-generated structures, and using annotations from the Disprot database.[158] We also used two versions of the IUPred method[159] that were optimized to predict long and short disordered regions. A recent comparative assessment found that these methods are characterized by high predictive performance, with AUC values of about 0.77.[153] The same consensus was applied in a few related studies.[17,31,160–162] We obtained the putative disorder annotations for each residue and computed the following two measures by combining these predictions per protein or a group of proteins:

– Disorder content: fraction of disordered residues in a given protein;

– %DisProt (fraction of disordered proteins): fraction of proteins with disorder content $\geq 0.4$ in a given protein set.

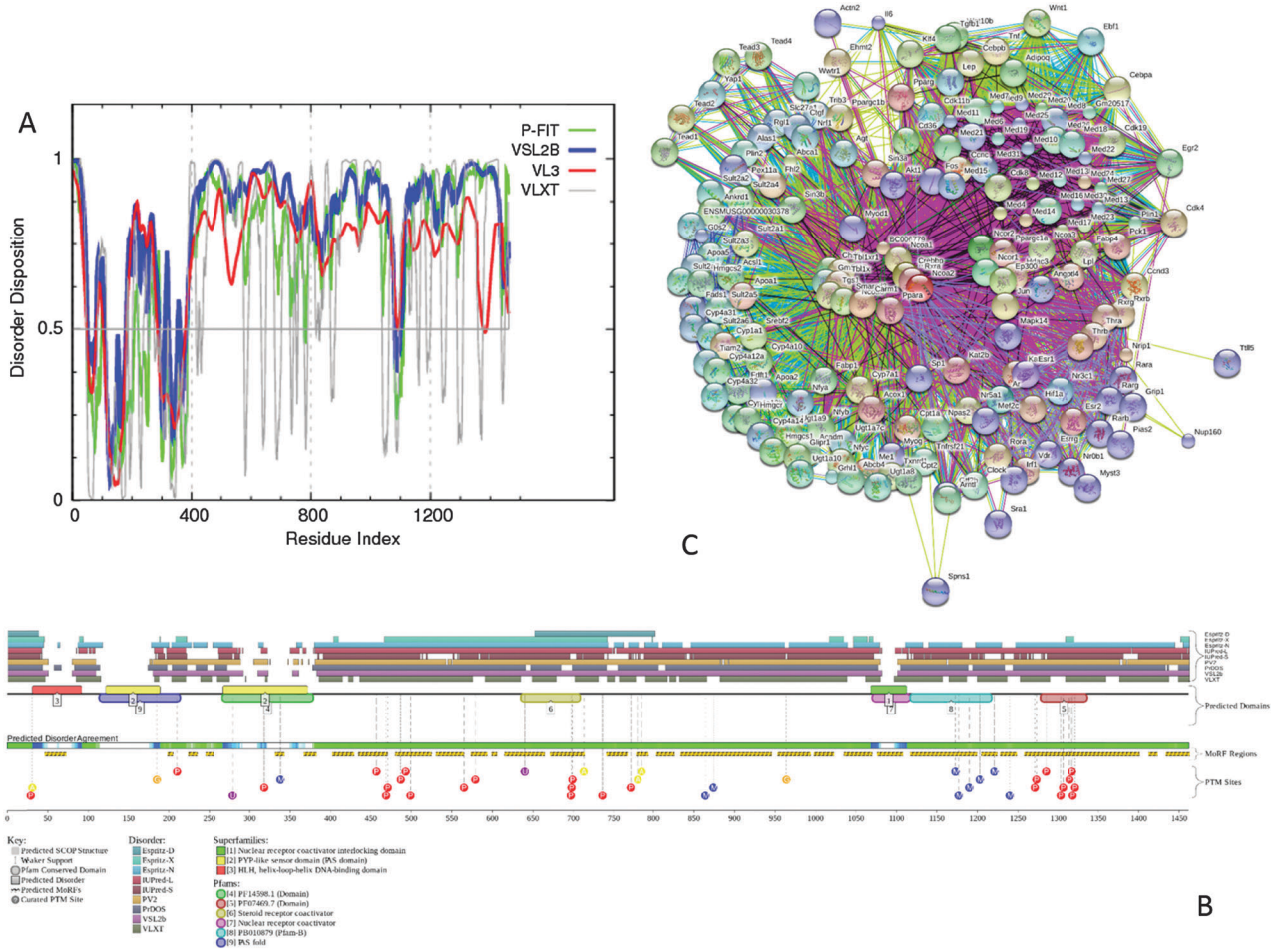


Fig. 10 Prevalence and functionality of intrinsic disorder in Grip1 protein (UniProt ID: Q61026) from PML nuclear bodies. (A) Evaluation of the per-residue disorder propensity by the members of the PONDR family of disorder predictors. (B) Evaluation of the functional intrinsic disorder propensity by the $D^2P^2$ database (http://d2p2.pro/). (C) Analysis of the Grip1 interactivity by STRING computational platform.

We calculated these measures for the autophagy-related non-nuclear proteins, and the autophagy-related nuclear proteins in each intra-nuclear compartment.

### Disorder sub-classification of autophagy-related proteins based on charge/hydropathy-cumulative distribution function analysis

One of the visually attractive ways to classify intrinsic disorder in whole proteins is given by the charge-hydropathy/cumulative distribution function (CH-CDF) analysis, which utilizes the ability of the CH-plot to separate proteins with extended disorder from compact proteins[20,28] as opposed to the CDF plot that discriminates all types of IDPs, including molten globules, from the ordered proteins.[28] Here, the disorderedness of a protein is characterized by its position within the two-dimensional plot, where a distance of this protein in the CH-plot (charge-hydropathy plot)[20,28] from the boundary separating compact (molten globule-like and well-structured globular proteins) and intrinsically disordered proteins with extended disorder (random coils and pre-molten globules) is used as the *Y*-coordinate, whereas an average distance of the protein's CDF curve[28] from the CDF boundary separating ordered and disordered proteins serves as its *X*-coordinate.[62,163,164] The resulting CH-CDF plot contains four quadrants and provides useful means for the visual sub-classification of proteins into ordered proteins, IDPs with extended disorder, compact molten globular IDPs or hybrid proteins containing ordered domains and IDPRs. Here, quadrant Q1 contains D/O proteins predicted to be disordered by CH-plots, but ordered by CDFs; quadrant Q2 correspond to the O/O proteins predicted to be ordered by both binary classifiers; quadrant Q3 includes O/D proteins predicted to be compact by CH-plots (*i.e.*, putative molten globules or hybrid proteins with ordered domains and IDPRs) and disordered by CDF analysis; whereas quadrant Q4 contains D/D proteins predicted to be disordered using both methods.[62,165]

### Disorder and disorder-based functional analysis of selected autophagy-related nuclear proteins using consensus disorder predictors, PONDR predictors, $D^2P^2$, STRING, and ANCHOR

To strengthen confidence of our analysis, disorder evaluations for several selected proteins (which are mouse autophagy-related nuclear proteins found in specific sub-nuclear compartments) are further analyzed by two consensus-based computational tools

 

for evaluation of intrinsic disorder, PONDR-FIT[79] and MobiDB.[74,75] The MobiDB database (http://mobidb.bio.unipd.it/)[74,75] uses consensus-derived disorder scores based on ten disorder predictors: ESpritz in its two flavors,[157] IUPred in its two versions,[159] DisEMBL in two of its flavors,[166] GlobPlot,[167] PONDR® VSL2B,[78,168] and JRONN.[169]

Furthermore, disorder was also evaluated using the methods from the PONDR family: PONDR® VLXT,[21] PONDR® VSL2,[60] PONDR® VL3,[78] and PONDR® FIT.[79] PONDR® VSL2 was shown to be among the accurate disorder predictors.[60,154,156] PONDR® VLXT has high sensitivity to local sequence peculiarities associated with disorder-based interaction sites.[21] PONDR® VL3 was assessed to provide accurate predictions of long disordered regions,[78] whereas PONDR-FIT is one of the recent consensus predictors. Again, we use multiple disorder predictors here to increase confidence of our results.

For selected mouse autophagy-related nuclear proteins additional information on the prevalence of disorder together with important disorder-related functional information was collected from the $D^2P^2$ database (http://d2p2.pro/).[30] This database utilizes the data pre-generated by PONDR® VLXT,[61] IUPred,[159] PONDR® VSL2B,[78,168] PrDOS,[170] ESpritz,[157] and PV2.[30] It also provides access to relevant functional information, such as location of various curated posttranslational modifications and predicted disorder-based protein binding sites.[30]

Additional functional information was obtained with the STRING (Search Tool for the Retrieval of Interacting Genes; http://string-db.org/) platform. STRINGS includes experimental and predicted interactions for proteins[80] We collected these interactions based on stringent cut-off of 0.9 as the minimal required level of confidence.

Putative protein binding sites that are localized in disordered regions of selected mouse nuclear proteins and which are co-localized in at least three sub-nuclear compartments and have at least five disordered domains (defined as regions of at least 30 consecutive disordered residues[171,172]) were identified using the ANCHOR algorithm.[76,77] ANCHOR relies on the estimation of pair-wise energy and hypothesis that long disordered regions contain localized potential binding sites that cannot form favorable intra-chain interactions to fold on their own but which are likely to fold by interacting with a globular protein partner.[76,77] Regions of a protein suggested by the ANCHOR algorithm to have significant potential to be binding sites are the ANCHOR-indicated binding sites (AIBS).

### Statistical analysis

We computed and compared the median disorder content and the %DisProt values computed for the autophagy-related nuclear proteins in the intra-nuclear compartments with the corresponding values for the non-nuclear autophagy-related proteins. We evaluated the statistical significance of the differences between these values following ref. 17, 31 and 161. We assessed the significance between two groups of values of a given measure estimated over ten subsets of randomly chosen sets of half of proteins in the corresponding two protein sets. We chose half of proteins in a given intra-nuclear compartment and the same number of non-nuclear proteins that have a similar size (±10% tolerance) compared to the selected nuclear proteins; we repeated that ten times. The protein size should be matched since prior research shows that the amount of disorder is dependent on the length of eukaryotic proteins.[160] We used the *t*-test for normal measurements (normality was assessed by the Anderson-Darling test[173] at the 0.05 significance) and otherwise we used the Wilcoxon rank sum test.[174] A given difference was deemed significant if the corresponding $p$-value $< 0.01$.

## Conclusions

Overall, our study revealed that autophagy-related nuclear proteins analyzed here are expected to possess multiple long disordered regions, contain numerous disorder-based binding sites and multiple sites of various posttranslational modifications and have multiple alternatively spliced isoforms. They are also expected to be promiscuous interactors, typically possessing rather well-developed interactomes.

Since autophagy-related IDPs were found in different membraneless nuclear organelles, our data are consistent with the conclusion that various nuclear compartments can be related to the autophagy regulation. This conclusion is in accord with the earlier observations that a nuclear pore protein, nucleoporin Tpr, is involved in the control of autophagy,[175] that a multifunctional and highly-conserved nuclear protein HMGB1 (high mobility group box 1), which typically serves as an architectural chromatin-binding protein, is associated with cell autophagy,[176] that the histone deacetylases sirtuins Sirt1 and Sirt2 have multifaceted roles in the regulation of autophagy,[177] as well as with the presence of numerous autophagy-related histone modifications,[53] and with known involvement of various transcription factors in controlling autophagy *via* transcriptional events.[51]

In conclusion, this work provides support to the idea of the functional importance of intrinsic disorder for the functionality of autophagy-related proteins. This conclusion follows from the fact that the autophagy-related nuclear IDPs are not only transcription factors but proteins that can be found in all nuclear compartments. These findings clearly indicate the presence of strong involvement of the cell nucleus in the control of autophagy and suggest that various nuclear events can be tightly connected to autophagy.

## Acknowledgements

This work was supported in part by the Qimonda Endowed Chair to L. K., the China Scholarship Council [scholarship to F. M.], and the Russian Science Foundation [14-24-00131 to V. N. U.].

## Notes and references

1 D. S. Goodsell, *Biochem. Mol. Biol. Educ.*, 2011, **39**, 91–101.
2 V. Tripathi and K. V. Prasanth, *eLS*, John Wiley & Sons Ltd, Chichester, Editon edn, 2011.
3 R. Hancock and K. Jeon, in *New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals*, ed.R. Hancock and K. Jeon, Academic Press, Amsterdam,

Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapoe, Sydney, Tokyo, 2014.
4 P. Jevtic, L. J. Edens, L. D. Vukovic and D. L. Levy, *Curr. Opin. Cell Biol.*, 2014, **28**, 16–27.
5 S. Fakan, *Histochem. Cell Biol.*, 2004, **122**, 83–93.
6 R. D. Phair and T. Misteli, *Nature*, 2000, **404**, 604–609.
7 T. Pederson, *Cell*, 2001, **104**, 635–638.
8 C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher and A. A. Hyman, *Science*, 2009, **324**, 1729–1732.
9 C. P. Brangwynne, T. J. Mitchison and A. A. Hyman, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 4334–4339.
10 P. Li, S. Banjade, H. C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q. X. Jiang, B. T. Nixon and M. K. Rosen, *Nature*, 2012, **483**, 336–340.
11 S. Aggarwal, N. Snaidero, G. Pahler, S. Frey, P. Sanchez, M. Zweckstetter, A. Janshoff, A. Schneider, M. T. Weil, I. A. Schaap, D. Gorlich and M. Simons, *PLoS Biol.*, 2013, **11**, e1001577.
12 M. Feric and C. P. Brangwynne, *Nat. Cell Biol.*, 2013, **15**, 1253–1259.
13 F. Wippich, B. Bodenmiller, M. G. Trajkovska, S. Wanka, R. Aebersold and L. Pelkmans, *Cell*, 2013, **152**, 791–805.
14 K. E. Handwerger and J. G. Gall, *Trends Cell Biol.*, 2006, **16**, 19–26.
15 D. L. Spector, *J. Cell Sci.*, 2001, **114**, 2891–2893.
16 T. Frege and V. N. Uversky, *Biochem. Biophys. Rep.*, 2015, **1**, 33–51.
17 F. Meng, I. Na, L. Kurgan and V. N. Uversky, *Int. J. Mol. Sci.*, 2016, **17**, 24.
18 A. K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger and J. E. Villafranca, *Pac. Symp. Biocomput.*, 1998, 473–484.
19 P. E. Wright and H. J. Dyson, *J. Mol. Biol.*, 1999, **293**, 321–331.
20 V. N. Uversky, J. R. Gillespie and A. L. Fink, *Proteins*, 2000, **41**, 415–427.
21 A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner and Z. Obradovic, *J. Mol. Graphics Modell.*, 2001, **19**, 26–59.
22 P. Tompa, *Trends Biochem. Sci.*, 2002, **27**, 527–533.
23 G. W. Daughdrill, G. J. Pielak, V. N. Uversky, M. S. Cortese and A. K. Dunker, in *Handbook of Protein Folding*, ed. J. Buchner and T. Kiefhaber, Wiley-VCH, Verlag GmbH & Co. KGaA, Weinheim, Germany, Editon edn, 2005, pp. 271–353.
24 V. N. Uversky and A. K. Dunker, *Biochim. Biophys. Acta*, 2010, **1804**, 1231–1264.
25 A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner and C. J. Brown, *Genome Inform.*, 2000, **11**, 161–171.
26 P. Romero, Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guilliot and A. K. Dunker, *Pac. Symp. Biocomput.*, 1998, 437–448.
27 J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton and D. T. Jones, *J. Mol. Biol.*, 2004, **337**, 635–645.
28 C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2005, **44**, 1989–2000.
29 B. Xue, A. K. Dunker and V. N. Uversky, *J. Biomol. Struct. Dyn.*, 2012, **30**, 137–149.
30 M. E. Oates, P. Romero, T. Ishida, M. Ghalwash, M. J. Mizianty, B. Xue, Z. Dosztanyi, V. N. Uversky, Z. Obradovic, L. Kurgan, A. K. Dunker and J. Gough, *Nucleic Acids Res.*, 2013, **41**, D508–D516.
31 Z. Peng, J. Yan, X. Fan, M. J. Mizianty, B. Xue, K. Wang, G. Hu, V. N. Uversky and L. Kurgan, *Cell. Mol. Life Sci.*, 2015, **72**, 137–151.
32 P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky and A. K. Dunker, *Biophys. J.*, 2007, **92**, 1439–1456.
33 H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, V. N. Uversky and Z. Obradovic, *J. Proteome Res.*, 2007, **6**, 1882–1898.
34 H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic and V. N. Uversky, *J. Proteome Res.*, 2007, **6**, 1917–1932.
35 L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradovic and A. K. Dunker, *J. Mol. Biol.*, 2002, **323**, 573–584.
36 A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva and V. N. Uversky, *FEBS J.*, 2005, **272**, 5129–5148.
37 V. N. Uversky, C. J. Oldfield and A. K. Dunker, *J. Mol. Recognit.*, 2005, **18**, 343–384.
38 Z. Peng, B. Xue, L. Kurgan and V. N. Uversky, *Cell Death Differ.*, 2013, **20**, 1257–1267.
39 A. Kaczmarek, P. Vandenabeele and D. V. Krysko, *Immunity*, 2013, **38**, 209–223.
40 S. Bialik, E. Zalckvar, Y. Ber, A. D. Rubinstein and A. Kimchi, *Trends Biochem. Sci.*, 2010, **35**, 556–564.
41 L. Galluzzi, T. Vanden Berghe, N. Vanlangenakker, S. Buettner, T. Eisenberg, P. Vandenabeele, F. Madeo and G. Kroemer, *Int. Rev. Cell Mol. Biol.*, 2011, **289**, 1–35.
42 L. Ouyang, Z. Shi, S. Zhao, F. T. Wang, T. T. Zhou, B. Liu and J. K. Bao, *Cell Proliferation*, 2012, **45**, 487–498.
43 A. Eisenberg-Lerner, S. Bialik, H. U. Simon and A. Kimchi, *Cell Death Differ.*, 2009, **16**, 966–975.
44 C. He and D. J. Klionsky, *Annu. Rev. Genet.*, 2009, **43**, 67–93.
45 B. Liu, Y. Cheng, Q. Liu, J. K. Bao and J. M. Yang, *Acta Pharmacol. Sin.*, 2010, **31**, 1154–1164.
46 N. Y. Lin, C. Beyer, A. Giessl, T. Kireva, C. Scholtysek, S. Uderhardt, L. E. Munoz, C. Dees, A. Distler, S. Wirtz, G. Kronke, B. Spencer, O. Distler, G. Schett and J. H. Distler, *Ann. Rheum. Dis.*, 2012, **72**, 761–768.
47 A. S. Patel, L. Lin, A. Geyer, J. A. Haspel, C. H. An, J. Cao, I. O. Rosas and D. Morse, *PLoS One*, 2012, **7**, e41394.
48 K. R. Parzych and D. J. Klionsky, *Antioxid. Redox Signaling*, 2014, **20**, 460–473.
49 D. J. Klionsky and S. D. Emr, *Science*, 2000, **290**, 1717–1721.
50 N. Mizushima, *Genes Dev.*, 2007, **21**, 2861–2873.
51 J. Fullgrabe, D. J. Klionsky and B. Joseph, *Nat. Rev. Mol. Cell Biol.*, 2014, **15**, 65–74.

 

52 Y. Feng, Z. Yao and D. J. Klionsky, *Trends Cell Biol.*, 2015, **25**, 354–363.
53 J. Fullgrabe, N. Heldring, O. Hermanson and B. Joseph, *Autophagy*, 2014, **10**, 556–561.
54 Y. Gruenbaum, R. D. Goldman, R. Meyuhas, E. Mills, A. Margalit, A. Fridkin, Y. Dayani, M. Prokocimer and A. Enosh, *Int. Rev. Cytol.*, 2003, **226**, 1–62.
55 N. Stuurman, A. M. Meijne, A. J. van der Pol, L. de Jong, R. van Driel and J. van Renswoude, *J. Biol. Chem.*, 1990, **265**, 5460–5465.
56 F. M. Boisvert, M. J. Hendzel and D. P. Bazett-Jones, *J. Cell Biol.*, 2000, **148**, 283–292.
57 V. Lallemand-Breitenbach and H. de The, *Cold Spring Harbor Perspect. Biol.*, 2010, **2**, a000661.
58 T. Sternsdorf, T. Grotzinger, K. Jensen and H. Will, *Immunobiology*, 1997, **198**, 307–331.
59 G. Kabachinski and T. U. Schwartz, *J. Cell Sci.*, 2015, **128**, 423–429.
60 K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker and Z. Obradovic, *J. Bioinf. Comput. Biol.*, 2005, **3**, 35–60.
61 P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown and A. K. Dunker, *Proteins*, 2001, **42**, 38–48.
62 F. Huang, C. J. Oldfield, B. Xue, W. L. Hsu, J. Meng, X. Liu, L. Shen, P. Romero, V. N. Uversky and A. Dunker, *BMC Bioinf.*, 2014, **15**(suppl 17), S4.
63 J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2006, **45**, 6873–6888.
64 Y. Minezaki, K. Homma, A. R. Kinjo and K. Nishikawa, *J. Mol. Biol.*, 2006, **359**, 1137–1149.
65 C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky and A. K. Dunker, *BMC Genomics*, 2008, **9**(suppl 1), S1.
66 B. Xue, C. J. Brown, A. K. Dunker and V. N. Uversky, *Biochim. Biophys. Acta*, 2013, **1834**, 725–738.
67 A. V. Uversky, B. Xue, Z. Peng, L. Kurgan and V. N. Uversky, *F1000Research*, 2013, **2**, 190.
68 V. K. Singh, I. Pacheco, V. N. Uversky, S. P. Smith, R. J. MacLeod and Z. Jia, *J. Mol. Biol.*, 2008, **380**, 313–326.
69 V. K. Singh, M. N. Rahman, K. Munro, V. N. Uversky, S. P. Smith and Z. Jia, *PLoS One*, 2012, **7**, e34680.
70 B. Xue, A. K. Dunker and V. N. Uversky, *J. Biomol. Struct. Dyn.*, 2012, **29**, 843–861.
71 B. Xue, P. R. Romero, M. Noutsou, M. M. Maurice, S. G. Rudiger, A. M. William, Jr., M. J. Mizianty, L. Kurgan, V. N. Uversky and A. K. Dunker, *FEBS Lett.*, 2013, **587**, 1587–1591.
72 F. Wang, C. B. Marshall, K. Yamamoto, G. Y. Li, M. J. Plevin, H. You, T. W. Mak and M. Ikura, *J. Mol. Biol.*, 2008, **384**, 590–603.
73 F. Wang, C. B. Marshall, G. Y. Li, K. Yamamoto, T. W. Mak and M. Ikura, *ACS Chem. Biol.*, 2009, **4**, 1017–1027.
74 T. Di Domenico, I. Walsh, A. J. Martin and S. C. Tosatto, *Bioinformatics*, 2012, **28**, 2080–2081.
75 E. Potenza, T. D. Domenico, I. Walsh and S. C. Tosatto, *Nucleic Acids Res.*, 2015, **43**, D315–D320.
76 B. Meszaros, I. Simon and Z. Dosztanyi, *PLoS Comput. Biol.*, 2009, **5**, e1000376.
77 Z. Dosztanyi, B. Meszaros and I. Simon, *Bioinformatics*, 2009, **25**, 2745–2746.
78 K. Peng, P. Radivojac, S. Vucetic, A. K. Dunker and Z. Obradovic, *BMC Bioinf.*, 2006, **7**, 208.
79 B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker and V. N. Uversky, *Biochim. Biophys. Acta*, 2010, **1804**, 996–1010.
80 D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen and C. von Mering, *Nucleic Acids Res.*, 2011, **39**, D561–D568.
81 L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic and A. K. Dunker, *Nucleic Acids Res.*, 2004, **32**, 1037–1049.
82 V. Pejaver, W. L. Hsu, F. Xin, A. K. Dunker, V. N. Uversky and P. Radivojac, *Protein Sci.*, 2014, **23**, 1077–1093.
83 V. N. Uversky, *Curr. Pharm. Des.*, 2013, **19**, 4191–4213.
84 C. Kaehler, J. Isensee, U. Nonhoff, M. Terrey, T. Hucho, H. Lehrach and S. Krobitsch, *PLoS One*, 2012, **7**, e50134.
85 E. L. Huttlin, M. P. Jedrychowski, J. E. Elias, T. Goswami, R. Rad, S. A. Beausoleil, J. Villen, W. Haas, M. E. Sowa and S. P. Gygi, *Cell*, 2010, **143**, 1174–1189.
86 J. Park, Y. Chen, D. X. Tishkoff, C. Peng, M. Tan, L. Dai, Z. Xie, Y. Zhang, B. M. Zwaans, M. E. Skinner, D. B. Lombard and Y. Zhao, *Mol. Cell*, 2013, **50**, 919–930.
87 C. Meunier, D. Bordereaux, F. Porteu, S. Gisselbrecht, S. Chretien and G. Courtois, *J. Biol. Chem.*, 2002, **277**, 9139–9147.
88 P. R. Romero, S. Zaidi, Y. Y. Fang, V. N. Uversky, P. Radivojac, C. J. Oldfield, M. S. Cortese, M. Sickmeier, T. LeGall, Z. Obradovic and A. K. Dunker, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 8390–8395.
89 E. Kovacs, P. Tompa, K. Liliom and L. Kalmar, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 5429–5434.
90 A. K. Ott, L. Locher, M. Koch and E. Deuerling, *PLoS One*, 2015, **10**, e0143457.
91 W. V. Yotov and R. St-Arnaud, *Genes Dev.*, 1996, **10**, 1763–1772.
92 C. Y. Park, S. A. Pierce, M. von Drehle, K. N. Ivey, J. A. Morgan, H. M. Blau and D. Srivastava, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 20750–20755.
93 J. Berkholz, A. Zakrzewicz and B. Munz, *Biochem. J.*, 2013, **453**, 303–310.
94 R. J. Sims, 3rd, E. K. Weihe, L. Zhu, S. O'Malley, J. V. Harriss and P. D. Gottlieb, *J. Biol. Chem.*, 2002, **277**, 26524–26529.
95 T. Ellis, L. Gambardella, M. Horcher, S. Tschanz, J. Capol, P. Bertram, W. Jochum, Y. Barrandon and M. Busslinger, *Genes Dev.*, 2001, **15**, 2307–2319.
96 Z. Wang, A. Goldstein, R. T. Zong, D. Lin, E. J. Neufeld, R. H. Scheuermann and P. W. Tucker, *Mol. Cell. Biol.*, 1999, **19**, 284–295.
97 L. Hulea and A. Nepveu, *Gene*, 2012, **497**, 18–26.
98 L. Sansregret and A. Nepveu, *Gene*, 2008, **412**, 84–94.
99 A. J. MacNeil and B. Pohajdak, *Immunol. Cell Biol.*, 2009, **87**, 72–80.
100 K. E. Park, H. D. Inerowicz, X. Wang, Y. Li, S. Koser and R. A. Cabot, *PLoS One*, 2012, **7**, e38990.

101 P. Wysoczanski, C. Schneider, S. Xiang, F. Munari, S. Trowitzsch, M. C. Wahl, R. Luhrmann, S. Becker and M. Zweckstetter, *Nat. Struct. Mol. Biol.*, 2014, **21**, 911–918.

102 B. P. Mihalas, P. S. Western, K. L. Loveland, E. A. McLaughlin and J. E. Holt, *Reproduction*, 2015, **150**, 485–496.

103 A. Dziembowski, A. P. Ventura, B. Rutz, F. Caspary, C. Faux, F. Halgand, O. Laprevote and B. Seraphin, *EMBO J.*, 2004, **23**, 4847–4856.

104 S. Trowitzsch, G. Weber, R. Luhrmann and M. C. Wahl, *J. Biol. Chem.*, 2008, **283**, 32317–32327.

105 M. A. Brooks, A. Dziembowski, S. Quevillon-Cheruel, V. Henriot, C. Faux, H. van Tilbeurgh and B. Seraphin, *Nucleic Acids Res.*, 2009, **37**, 129–143.

106 A. Gottschalk, C. Bartels, G. Neubauer, R. Luhrmann and P. Fabrizio, *Mol. Cell. Biol.*, 2001, **21**, 3037–3046.

107 F. W. Scherrer, Jr. and M. Spingola, *RNA*, 2006, **12**, 1361–1372.

108 M. Jiang, J. Ryu, M. Kiraly, K. Duke, V. Reinke and S. K. Kim, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 218–223.

109 P. Sassone-Corsi, *Science*, 2002, **296**, 2176–2178.

110 M. L. Meistrich, B. Mohapatra, C. R. Shirley and M. Zhao, *Chromosoma*, 2003, **111**, 483–488.

111 M. M. Pradeepa and M. R. Rao, *Soc. Reprod. Fertil. Suppl.*, 2007, **63**, 1–10.

112 S. S. Hammoud, D. A. Nix, H. Zhang, J. Purwar, D. T. Carrell and B. R. Cairns, *Nature*, 2009, **460**, 473–478.

113 U. Brykczynska, M. Hisano, S. Erkek, L. Ramos, E. J. Oakeley, T. C. Roloff, C. Beisel, D. Schubeler, M. B. Stadler and A. H. Peters, *Nat. Struct. Mol. Biol.*, 2010, **17**, 679–687.

114 J. Gaucher, N. Reynoird, E. Montellier, F. Boussouar, S. Rousseaux and S. Khochbin, *FEBS J.*, 2010, **277**, 599–604.

115 C. Rathke, W. M. Baarends, S. Awe and R. Renkawitz-Pohl, *Biochim. Biophys. Acta*, 2014, **1839**, 155–168.

116 N. Gupta, M. P. Madapura, U. A. Bhat and M. R. Rao, *J. Biol. Chem.*, 2015, **290**, 12101–12122.

117 T. K. Kundu and M. R. Rao, *FEBS Lett.*, 1994, **351**, 6–10.

118 A. K. Dunker, M. Babu, E. Barbar, M. Blackledge, S. E. Bondos, Z. Dosztányi, H. J. Dyson, J. Forman-Kay, M. Fuxreiter, J. Gsponer, K.-H. Han, D. T. Jones, S. Longhi, S. J. Metallo, K. Nishikawa, R. Nussinov, Z. Obradovic, R. Pappu, B. Rost, P. Selenko, V. Subramaniam, J. L. Sussman, P. Tompa and V. N. Uversky, *Instrum. Anal. Intrinsically Disord. Proteins*, 2013, **1**, e24157.

119 D. E. Ayer, L. Kretzner and R. N. Eisenman, *Cell*, 1993, **72**, 211–222.

120 P. J. Hurlin, C. Queva, P. J. Koskinen, E. Steingrimsson, D. E. Ayer, N. G. Copeland, N. A. Jenkins and R. N. Eisenman, *EMBO J.*, 1995, **14**, 5646–5659.

121 P. J. Koskinen, D. E. Ayer and R. N. Eisenman, *Cell Growth Differ.*, 1995, **6**, 623–629.

122 D. E. Ayer, Q. A. Lawrence and R. N. Eisenman, *Cell*, 1995, **80**, 767–776.

123 N. Schreiber-Agus, L. Chin, K. Chen, R. Torres, G. Rao, P. Guida, A. I. Skoultchi and R. A. DePinho, *Cell*, 1995, **80**, 777–786.

124 A. L. Eilers, A. N. Billin, J. Liu and D. E. Ayer, *J. Biol. Chem.*, 1999, **274**, 32750–32756.

125 C. A. Hassig, T. C. Fleischer, A. N. Billin, S. L. Schreiber and D. E. Ayer, *Cell*, 1997, **89**, 341–347.

126 C. D. Laherty, W. M. Yang, J. M. Sun, J. R. Davie, E. Seto and R. N. Eisenman, *Cell*, 1997, **89**, 349–356.

127 Y. Zhang, R. Iratni, H. Erdjument-Bromage, P. Tempst and D. Reinberg, *Cell*, 1997, **89**, 357–364.

128 T. Heinzel, R. M. Lavinsky, T. M. Mullen, M. Soderstrom, C. D. Laherty, J. Torchia, W. M. Yang, G. Brard, S. D. Ngo, J. R. Davie, E. Seto, R. N. Eisenman, D. W. Rose, C. K. Glass and M. G. Rosenfeld, *Nature*, 1997, **387**, 43–48.

129 L. Nagy, H. Y. Kao, D. Chakravarti, R. J. Lin, C. A. Hassig, D. E. Ayer, S. L. Schreiber and R. M. Evans, *Cell*, 1997, **89**, 373–380.

130 P. L. Jones, G. J. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis and A. P. Wolffe, *Nat. Genet.*, 1998, **19**, 187–191.

131 X. Nan, H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman and A. Bird, *Nature*, 1998, **393**, 386–389.

132 C. D. Laherty, A. N. Billin, R. M. Lavinsky, G. S. Yochum, A. C. Bush, J. M. Sun, T. M. Mullen, J. R. Davie, D. W. Rose, C. K. Glass, M. G. Rosenfeld, D. E. Ayer and R. N. Eisenman, *Mol. Cell*, 1998, **2**, 33–42.

133 H. Wang, I. Clark, P. R. Nicholson, I. Herskowitz and D. J. Stillman, *Mol. Cell. Biol.*, 1990, **10**, 5927–5936.

134 L. Alland, R. Muhle, H. Hou, Jr., J. Potes, L. Chin, N. Schreiber-Agus and R. A. DePinho, *Nature*, 1997, **387**, 49–55.

135 A. K. Shiau, D. Barstad, J. T. Radek, M. J. Meyers, K. W. Nettles, B. S. Katzenellenbogen, J. A. Katzenellenbogen, D. A. Agard and G. L. Greene, *Nat. Struct. Biol.*, 2002, **9**, 359–364.

136 H. B. Zhou, K. W. Nettles, J. B. Bruning, Y. Kim, A. Joachimiak, S. Sharma, K. E. Carlson, F. Stossi, B. S. Katzenellenbogen, G. L. Greene and J. A. Katzenellenbogen, *Chem. Biol.*, 2007, **14**, 659–669.

137 T. Ostberg, S. Svensson, G. Selen, J. Uppenberg, M. Thor, M. Sundbom, M. Sydow-Backman, A. L. Gustavsson and L. Jendeberg, *J. Biol. Chem.*, 2004, **279**, 41124–41130.

138 L. Z. Mi, S. Devarakonda, J. M. Harp, Q. Han, R. Pellicciari, T. M. Willson, S. Khorasanizadeh and F. Rastinejad, *Mol. Cell*, 2003, **11**, 1093–1100.

139 E. Estebanez-Perpina, L. A. Arnold, P. Nguyen, E. D. Rodrigues, E. Mar, R. Bateman, P. Pallai, K. M. Shokat, J. D. Baxter, R. K. Guy, P. Webb and R. J. Fletterick, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16074–16079.

140 T. Seitz, R. Thoma, G. A. Schoch, M. Stihle, J. Benz, B. D'Arcy, A. Wiget, A. Ruf, M. Hennig and R. Sterner, *J. Mol. Biol.*, 2010, **403**, 562–577.

141 E. Stashi, R. B. Lanz, J. Mao, G. Michailidis, B. Zhu, N. M. Kettner, N. Putluri, E. L. Reineke, L. C. Reineke, S. Dasgupta, A. Dean, C. R. Stevenson, N. Sivasubramanian, A. Sreekumar, F. Demayo, B. York, L. Fu and B. W. O'Malley, *Cell Rep.*, 2014, **6**, 633–645.

142 F. Picard, M. Gehin, J. Annicotte, S. Rocchi, M. F. Champy, B. W. O'Malley, P. Chambon and J. Auwerx, *Cell*, 2002, **111**, 931–941.

143 A. R. Chopra, J. F. Louet, P. Saha, J. An, F. Demayo, J. Xu, B. York, S. Karpen, M. Finegold, D. Moore, L. Chan,


 

C. B. Newgard and B. W. O'Malley, *Science*, 2008, **322**, 1395–1399.
144 H. Xie, M. S. Sadim and Z. Sun, *J. Immunol.*, 2005, **175**, 3800–3809.
145 Y. H. Lee, S. S. Koh, X. Zhang, X. Cheng and M. R. Stallcup, *Mol. Cell. Biol.*, 2002, **22**, 3621–3632.
146 K. Homma, K. Suzuki and H. Sugawara, *Nucleic Acids Res.*, 2011, **39**, D986–D990.
147 K. Willadsen, N. Mohamad and M. Boden, *Genomics, Proteomics Bioinf.*, 2012, **10**, 226–229.
148 G. Dellaire, R. Farrall and W. A. Bickmore, *Nucleic Acids Res.*, 2003, **31**, 328–330.
149 Y. Ahmad, F. M. Boisvert, P. Gregor, A. Cobley and A. I. Lamond, *Nucleic Acids Res.*, 2009, **37**, D181–D184.
150 S. Mika and B. Rost, *Nucleic Acids Res.*, 2005, **33**, D160–D163.
151 R. Apweiler, A. Bateman, M. J. Martin, C. O'Donovan, M. Magrane, Y. Alam-Faruque, E. Alpi, R. Antunes, J. Arganiska, E. B. Casanova, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, W. M. Chan, G. Chavali, E. Cibrian-Uhalte, A. Da Silva, M. De Giorgi, F. Fazzini, P. Gane, L. G. Castro, P. Garmiri, E. Hatton-Ellis, R. Hieta, R. Huntley, D. Legge, W. D. Liu, J. Luo, A. MacDougall, P. Mutowo, A. Nightingale, S. Orchard, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Y. Qi, S. Rosanoff, T. Sawford, A. Shypitsyna, E. Turner, V. Volynkin, T. Wardell, X. Watkins, H. Zellner, M. Corbett, M. Donnelly, P. Van Rensburg, M. Goujon, H. McWilliam, R. Lopez, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, P. A. Binz, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. de Castro, L. Cerutti, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, J. James, F. Jungo, G. Keller, V. Lara, P. Lemercier, J. Lew, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A. L. Veuthey, C. H. Wu, C. N. Arighi, L. Arminski, C. M. Chen, Y. X. Chen, J. S. Garavelli, H. Z. Huang, K. Laiho, P. McGarvey, D. A. Natale, B. E. Suzek, C. R. Vinayaka, Q. H. Wang, Y. Q. Wang, L. S. Yeh, M. S. Yerramalla, J. Zhang and U. Consortium, *Nucleic Acids Res.*, 2014, **42**, D191–D198.
152 T. S. K. Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. H. Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, *Nucleic Acids Res.*, 2009, **37**, D767–D772.
153 I. Walsh, M. Giollo, T. Di Domenico, C. Ferrari, O. Zimmermann and S. C. Tosatto, *Bioinformatics*, 2015, **31**, 201–208.
154 X. Fan and L. Kurgan, *J. Biomol. Struct. Dyn.*, 2014, **32**, 448–464.
155 Z. Peng and L. Kurgan, *Pac. Symp. Biocomput.*, 2012, 176–187.
156 Z. L. Peng and L. Kurgan, *Curr. Protein Pept. Sci.*, 2012, **13**, 6–18.
157 I. Walsh, A. J. Martin, T. Di Domenico and S. C. Tosatto, *Bioinformatics*, 2012, **28**, 503–509.
158 M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic and A. K. Dunker, *Nucleic Acids Res.*, 2007, **35**, D786–D793.
159 Z. Dosztanyi, V. Csizmok, P. Tompa and I. Simon, *Bioinformatics*, 2005, **21**, 3433–3434.
160 M. Howell, R. Green, A. Killeen, L. Wedderburn, V. Picascio, A. Rabionet, Z. L. Peng, M. Larina, B. Xue, L. Kurgan and V. N. Uversky, *J. Biol. Syst.*, 2012, **20**, 471–511.
161 Z. Peng, C. J. Oldfield, B. Xue, M. J. Mizianty, A. K. Dunker, L. Kurgan and V. N. Uversky, *Cell. Mol. Life Sci.*, 2014, **71**, 1477–1504.
162 G. Hu, Z. Wu, K. Wang, V. N. Uversky and L. Kurgan, *Curr. Drug Targets*, 2016, **17**, 1198–1205.
163 A. Mohan, W. J. Sullivan, Jr., P. Radivojac, A. K. Dunker and V. N. Uversky, *Mol. BioSyst.*, 2008, **4**, 328–340.
164 B. Xue, C. J. Oldfield, A. K. Dunker and V. N. Uversky, *FEBS Lett.*, 2009, **583**, 1469–1474.
165 F. Huang, C. Oldfield, J. Meng, W. L. Hsu, B. Xue, V. N. Uversky, P. Romero and A. K. Dunker, *Pac. Symp. Biocomput.*, 2012, 128–139.
166 R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson and R. B. Russell, *Structure*, 2003, **11**, 1453–1459.
167 R. Linding, R. B. Russell, V. Neduva and T. J. Gibson, *Nucleic Acids Res.*, 2003, **31**, 3701–3708.
168 Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac and A. K. Dunker, *Proteins*, 2005, **61**(Suppl 7), 176–182.
169 Z. R. Yang, R. Thomson, P. McNeil and R. M. Esnouf, *Bioinformatics*, 2005, **21**, 3369–3376.
170 T. Ishida and K. Kinoshita, *Nucleic Acids Res.*, 2007, **35**, W460–464.
171 M. M. Pentony and D. T. Jones, *Proteins*, 2010, **78**, 212–221.
172 P. Tompa, M. Fuxreiter, C. J. Oldfield, I. Simon, A. K. Dunker and V. N. Uversky, *BioEssays*, 2009, **31**, 328–335.
173 M. A. Stephens, *J. Am. Stat. Assoc.*, 1974, **69**, 730–737.
174 F. Wilcoxon, *Biom. Bull.*, 1945, 80–83.
175 T. Funasaka, E. Tsuka and R. W. Wong, *Sci. Rep.*, 2012, **2**, 878.
176 X. Sun and D. Tang, *Autophagy*, 2014, **10**, 1873–1876.
177 F. Ng and B. L. Tang, *J. Cell. Physiol.*, 2013, **228**, 2262–2270.