



Article

# Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments

Fanchi Meng <sup>1</sup>, Insung Na <sup>2</sup>, Lukasz Kurgan <sup>1,3,\*</sup> and Vladimir N. Uversky <sup>2,4,5,6,7,\*</sup>

Received: 11 October 2015; Accepted: 18 December 2015; Published: 25 December 2015

Academic Editors: Christo Z. Christov and Mark L. Richter

<sup>1</sup> Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada; fanchi@ualberta.ca

<sup>2</sup> Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA; ina2@health.usf.edu

<sup>3</sup> Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23219, USA

<sup>4</sup> University of South Florida Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

<sup>5</sup> Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region 142292, Russian

<sup>6</sup> Biology Department, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia

<sup>7</sup> Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, Saint Petersburg 194064, Russian

\* Correspondence: lkurgan@ualberta.ca (L.K.); vuversky@health.usf.edu (V.N.U.); Tel.: +1-804-827-3986 (L.K.); +1-813-974-5815 (V.N.U.); Fax: +1-804-828-2771 (L.K.); +1-813-974-7357 (V.N.U.)

**Abstract:** The cell nucleus contains a number of membrane-less organelles or intra-nuclear compartments. These compartments are dynamic structures representing liquid-droplet phases which are only slightly denser than the bulk intra-nuclear fluid. They possess different functions, have diverse morphologies, and are typically composed of RNA (or, in some cases, DNA) and proteins. We analyzed 3005 mouse proteins localized in specific intra-nuclear organelles, such as nucleolus, chromatin, Cajal bodies, nuclear speckles, promyelocytic leukemia (PML) nuclear bodies, nuclear lamina, nuclear pores, and perinuclear compartment and compared them with ~29,863 non-nuclear proteins from mouse proteome. Our analysis revealed that intrinsic disorder is enriched in the majority of intra-nuclear compartments, except for the nuclear pore and lamina. These compartments are depleted in proteins that lack disordered domains and enriched in proteins that have multiple disordered domains. Moonlighting proteins found in multiple intra-nuclear compartments are more likely to have multiple disordered domains. Protein-protein interaction networks in the intra-nuclear compartments are denser and include more hubs compared to the non-nuclear proteins. Hubs in the intra-nuclear compartments (except for the nuclear pore) are enriched in disorder compared with non-nuclear hubs and non-nuclear proteins. Therefore, our work provides support to the idea of the functional importance of intrinsic disorder in the cell nucleus and shows that many proteins associated with sub-nuclear organelles in nuclei of mouse cells are enriched in disorder. This high level of disorder in the mouse nuclear proteins defines their ability to serve as very promiscuous binders, possessing both large quantities of potential disorder-based interaction sites and the ability of a single such site to be involved in a large number of interactions.

**Keywords:** intrinsically disordered protein; cell nucleus; membrane-less organelles; intra-nuclear compartments; DNA-binding protein; RNA-binding protein

## 1. Introduction

The nucleus is a double membrane-bound organelle, which is the most characteristic and important component of any eukaryotic cell, serving as the “container” of genetic information that is responsible for the storage of hereditary material (in chromosomes and genes), storage of proteins and RNA (specifically in the nucleolus), exchange of hereditary molecules (DNA and RNA), production of ribosomes, and coordination of the gene expression [1–4]. Although the nucleus typically takes ~6% of the total cellular volume, in some cell types, nucleus occupies 30%–50% of the cellular volume. The nucleus is a highly dynamic organelle, and its morphology, size, and shape are tightly regulated and are noticeably changed during the cell cycle. Development of several diseases, such as cancer, is correlated with the altered nuclear morphology [4].

It is recognized now that the cell nucleus is a non-homogeneous entity filled with a transparent, semi-solid, granular substance known as nucleoplasm, nuclear sap or karyolymph, that has a complex architecture and contains several sub-nuclear structures alias nuclear domains, or sub-nuclear organelles, or nuclear compartments, or sub-nuclear bodies [5]. These sub-nuclear compartments are temporary and are classified as highly dynamic, membrane-less organelles, which do not exist all the time, appear only during certain stages of the cell cycle, and are dynamically formed via recruitment of proteins, RNA and DNA. At the moment, little information is available on the highly dynamic nuclear environment, on the roles of various nuclear organelles, domains and compartments, and on the existence and functional roles of interactions between these organelles and compartments. Furthermore, the molecular mechanisms defining the ability of these nuclear domains to maintain their specific structures in the absence of membranes also remain mostly unknown [6]. Since they lack membranes, these organelles possess a highly dynamic nature, and their components are involved in direct contact with the surrounding nucleoplasm or cytoplasm [7,8]. These membrane-less compartments with their high level of internal dynamics are marginally denser than the bulk intra-nuclear/intra-cellular fluid, and are, therefore, considered as liquid-droplet phases of the nucleoplasm/cytoplasm [9–14].

Among these sub-nuclear organelles are Cajal bodies, chromatin, cleavage bodies, nuclear gems (Gemini or coiled bodies), nuclear pores, nuclear speckles, nucleolus, Oct1/PTF/transcription (OPT) domains, PcG bodies (sub-nuclear organelles containing polycomb group proteins), perinuclear compartment, promyelocytic leukemia (PML) nuclear bodies, SAM68 nuclear bodies, and a few others [15]. Several of these sub-nuclear structures (such as nucleolus, chromatin, Cajal bodies, nuclear speckles, PML bodies, nuclear lamina, nuclear pores, and perinuclear compartment) are subjects of the analysis conducted in our study, and therefore they are briefly introduced below.

Nucleolus is one of the most important sub-nuclear organelle which is clearly visible in the nucleoplasm as a large, dense, spherical structure with pronounced acidophilic character. The main function of the nucleolus is the biogenesis of ribosomal subunits, which, being synthesized, are taken to the cytoplasm for the translation of RNA. Therefore, the size of nucleolus is related to the ribosomal demands of the cell, where cells producing large quantities of proteins (and, thereby, requiring more ribosomes) have large nucleolus, and cells with no synthetic activities (such as spermatozooids or muscle cells) have smaller nucleolus or do not have nucleolus at all [16]. The typical nucleolus is composed of three regions: a fibrillar center containing DNA that is not actively transcribed, a dense fibrillar component containing RNA molecules which are actively transcribed, and a granular component containing maturing ribosomal precursor particles [17]. The nucleolar proteins represent the largest subset of nuclear proteins analyzed in this study.

Chromatin or chromatin fibers are thread-like, coiled, elongated structures made of DNA and proteins and found in the nucleoplasm. The primary functions ascribed to chromatin are: (i) DNA packaging into a smaller volume to fit in the cell; (ii) DNA reinforcement to allow mitosis; (iii) prevention of the DNA damage; and (iv) providing means to control DNA replication and gene expression [18]. Therefore, chromatin is a way of dynamic storage of long DNA molecules. In the eukaryotic nucleus, histone proteins participate in the hierarchical packaging of genomic DNA. Here, the (H3-H4)<sub>2</sub> tetramer capped with two dimers of H2A-H2B constitute a core histone octamer, around

which the 146 base pairs of DNA are wrapped in 1.7 super-helical turns to form the nucleosome, chromatin's fundamental repeating unit [19–21]. The resulting “beads-on-a-string” fiber that has a diameter of 11 nm serves as the first level of organization for chromatin [20]. Meanwhile, the subsequent binding of the linker histone (H1 or H5) forms a more condensed 30-nm fiber, which generates the second structural level of DNA organization [21,22]. Multiple mechanisms ranging from histone modification to chromatin remodeling, to histone eviction, and to histone variant incorporation are involved in tight regulation of the chromatin structure dynamics [18]. Chromatin is unevenly distributed within the nucleus, being organized into the chromosome territories [15,23,24]. Nuclear speckles, which are also known as interchromatin granule clusters, are the storage sites for the pre-mRNA splicing factors that can be later recruited by RNA polymerase II transcription sites in the nucleoplasm. The cell nucleus contains 25–50 such speckles which are diffusely distributed throughout the nucleoplasm and that include a set of pre-mRNA splicing factors [25].

Nucleus typically has one to five Cajal bodies (CBs), which are roughly spherical 0.1–2.0  $\mu\text{M}$  sub-nuclear structures. Although CBs are not easily detected in all cell types, they have very prominent appearance in the highly active cells such as cancer cells or neurons [26,27]. CBs are known as nucleolar accessory bodies or coiled bodies because of the presence of the coiled threads of the specific protein, coilin [26,27]. Since CBs are able to concentrate some of the components related to the several nuclear processes (e.g., they are involved in the assembly of U small nuclear ribonucleoproteins (snRNPs) needed for the formation of the spliceosome), these sub-organelles are assumed to increase efficiency of the related processes [27].

The average mammalian nucleus contains between 10 and 30 of promyelocytic leukemia (PML) bodies. These organelles with the diameter ranging from 0.3 to 1.0  $\mu\text{m}$  are also known as ND10, Kr bodies, and PODs (PML oncogenic domains) [28]. PML bodies can be found in proximity of other nuclear organelles (e.g., Cajal bodies and nuclear gems). However, functional significance of this juxtaposition is unknown. Although the central core of the PML bodies are devoid of chromatin and RNA, the newly synthesized RNAs are found at their periphery [29]. On late-replicated satellite DNA, these sub-nuclear organelles are believed to re-establish the condensed heterochromatic state [30].

The nuclear envelope separates the content of the nucleus, which is the double membrane-enclosed cellular organelle, from the cytoplasm, and which is known to disappear during cell division. The nuclear envelope has two layers, an outer membrane and an inner membrane (also known as nuclear lamina). Each of these lipoproteinaceous layers is about 75 to 90  $\text{\AA}$  thick, with the space between them being known as the perinuclear space [31]. Both these layers enable the exchange of ions between the nucleus and cytoplasm, and in some cell types (such as salivary glands) these two structures work as a barrier for the diffusion of substances and some ions, such as  $\text{K}^+$ ,  $\text{Na}^+$ , and chloride  $\text{Cl}^-$  [32–34].

Nuclear pores are abundantly found in the nuclear envelope. The nuclear pore complex is a large, multichain proteinaceous machine (124 MDa in mammals), which is arranged hexagonally along the membrane, crosses the nuclear envelope and contains approximately 30 different protein components (nucleoporins), each in multiple copies [35]. On average, these pores cover 10% of the surface of the nucleus in mammalian cells. In some cases, the nuclear pore complex is additionally covered by a thin membrane which is potentially involved in selective permeability [35–37].

An irregularly shaped nuclear body known as the perinucleolar compartment (PNC) is found at the periphery of the nucleolus [38–40]. The length of PNC ranges from 0.25 to 4.0  $\mu\text{m}$ , and this organelle contains several short RNAs and the heterogeneous nuclear ribonucleoprotein I, which is a polypyrimidine tract-binding protein [38].

It is of interest to point out that the membrane-less compartments are not unique for the nucleus, and several of the non-membrane bound organelles are found in the cytoplasm of eukaryotic cells. Often, these organelles (such as centrosomes [41], spindles [42], and stress granules [14], to name a few) are comprised of proteins and RNA [42].

Although these sub-nuclear organelles are rather different morphologically and have very different functions, they are typically composed of RNA (or, in some cases, DNA) and proteins. Recent studies revealed that in contrast to the classical protein structure paradigm stating that a protein must have a defined 3D-structure in order to be functional, many biologically active proteins with a wide spectrum of crucial functions are characterized by the lack of such specific tertiary structures [43–49]. Studies of different proteomes suggested that both intrinsically disordered proteins (IDPs) and hybrid proteins containing intrinsically disordered protein regions (IDPRs) as well as ordered domains are very prevalent in nature and that proteins from eukaryotes have more intrinsic disorder than proteins of bacteria or archaea [46,50–56]. Disorder-based functions are typically related to control pathways, signaling, and regulation [57–61], being therefore complementary to the catalytic and transport functions of ordered proteins [62–65]. Since among the disorder-specific functions are DNA- and RNA-interactions, and since almost all sub-nuclear organelles contain RNA- and DNA-binding proteins, our goal was to study the nuclear proteins of the mouse cell, to empirically evaluate their level of disorder, to investigate whether there is a connection between intrinsic disorder and the functional roles of these proteins in the cell nucleus, and, thereby, to better understand the role of the intrinsic disorder in the context of compartmentalization of the nucleus. We also empirically evaluated certain characteristics of protein-protein interaction (PPI) network, such as density and presence of highly connected protein hubs, in the considered intra-nuclear compartments and analyzed the relation between these characteristics and the intrinsic disorder.

The current work extends our earlier study on the analysis of disorder status of 185 human nuclear proteins [66]. Here, we use a much larger dataset (3005 *vs.* 185 nuclear proteins), apply multiple disorder prediction tools *vs.* a single tool to strengthen confidence in the estimation of the abundance of disorder, conduct analysis of PPIs and hubs, perform analysis of the proteins that are co-localized in multiple compartments, and provide richer set of measures to quantify abundance of disorder (proteins with multiple disordered domains and normalized content of disordered domains).

## 2. Results and Discussion

### 2.1. Intrinsic Disorder in the Annotated and Predicted Intra-Nuclear Localizations Is Equivalent

Table 1 lists datasets used in our study. First, we compared the disorder content (Figure S1) and other characteristics of the intrinsic disorder (Tables S1 and S2) across the experimentally annotated intra-nuclear localizations ( $\text{NUCLEAR}_a$ ), annotated and predicted intra-nuclear localizations ( $\text{NUCLEAR}_{ap}$ ), and the annotated and predicted intra-nuclear localizations that were mapped into the PPI network ( $\text{PPI}_{\text{NUCLEAR}_{ap}}$ ). Figure S1 reveals that distributions of the disorder content are very similar between these three sets of annotated proteins.

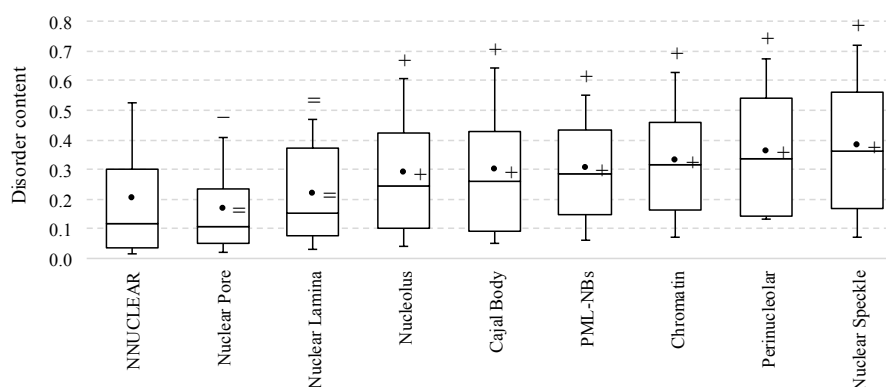
**Table 1.** Datasets used in this study.

Dataset	Description	Number of Proteins
NNUCLEAR	Complete mouse proteome excluding the nuclear proteins.	29,863
$\text{NUCLEAR}_a$	Nuclear mouse proteome with annotated intra-nuclear compartments.	1285
$\text{NUCLEAR}_{ap}$	Nuclear mouse proteome with annotated and predicted intra-nuclear compartments.	3005
PPI	Mouse proteins included in the protein-protein interaction (PPI) network.	8206
$\text{PPI}_{\text{NNUCLEAR}}$	Complete mouse proteome without nuclear proteins included in the PPI network (intersection of NNUCLEAR and PPI).	5687
$\text{PPI}_{\text{NUCLEAR}_{ap}}$	Nuclear mouse proteome with annotated and predicted intra-nuclear compartments included in the PPI network (intersection of $\text{NUCLEAR}_{ap}$ and PPI).	2519

The two supplementary Tables also show that the fraction of proteins with disordered domains and completely disordered proteins are similar. Therefore, in the subsequent analyses, we report our results on the larger combined set of annotated and predicted localization; the corresponding results on the smaller set of annotated localizations are given in the Supplement. Furthermore, this similarity also justifies our use of the datasets mapped into the PPI network to analyze the peculiarities of the protein-protein interactions in relation to the intrinsic disorder for the intra-nuclear compartments.

## 2.2. Intrinsic Disorder in Intra-Nuclear Compartments

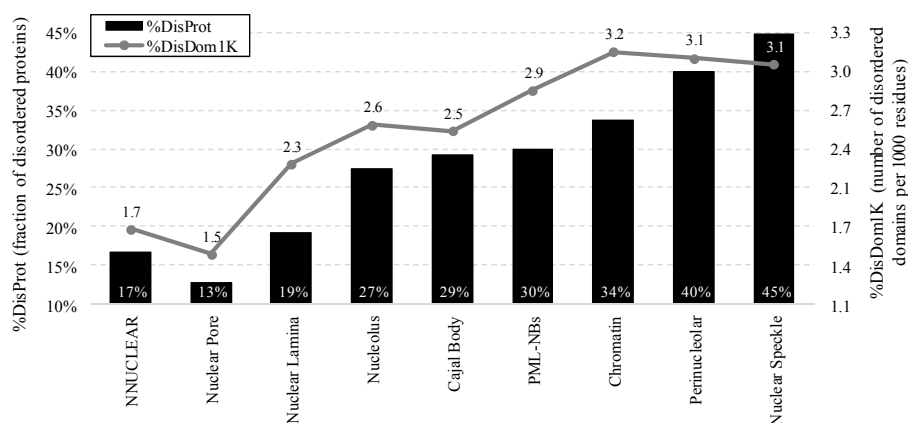
Figure 1 compares the disorder content between the non-nuclear proteins and the nuclear proteins localized in intra-nuclear compartments (Figure S2 shows the corresponding results based solely on the experimental annotations of localization). Nuclear proteins in most of the compartments, except for nuclear pore and lamina, are significantly ( $p$ -value  $< 0.01$ ) enriched in the disorder compared to the non-nuclear proteins. The median disorder content is higher by between 108% in the nucleolus (0.25 *vs.* 0.12 for the non-nuclear proteins) and 200% in the nuclear speckle (0.36 *vs.* 0.12). The lower disorder content in the nuclear pore could be explained by the highly structured composition of the nuclear pore complex [67]. Nuclear lamina has median content of 0.15 which is slightly higher ( $p$ -value = 0.07) than the overall content of 0.12 for the non-nuclear proteins.



**Figure 1.** Disorder content of the non-nuclear proteins from the NNUCLEAR dataset and proteins in the considered intra-nuclear compartment from the NUCLEAR<sub>ap</sub> dataset. The box plots include the first quartile, median and third quartile while whiskers correspond to the 10th and 90th centiles of the disorder content of proteins in a given set. The black circle marker is the mean value of the disorder content. The significance of the differences in the median (mean) disorder content between proteins in a given compartment and non-nuclear proteins is annotated above the whiskers (right of the marker); + and – mean that the content of the nuclear proteins is significantly higher and lower ( $p$ -value  $< 0.01$ ), respectively; “=” means that the difference is not significant ( $p$ -value  $\geq 0.01$ ). Disorder was annotated with the consensus of Espritz and IUPred.

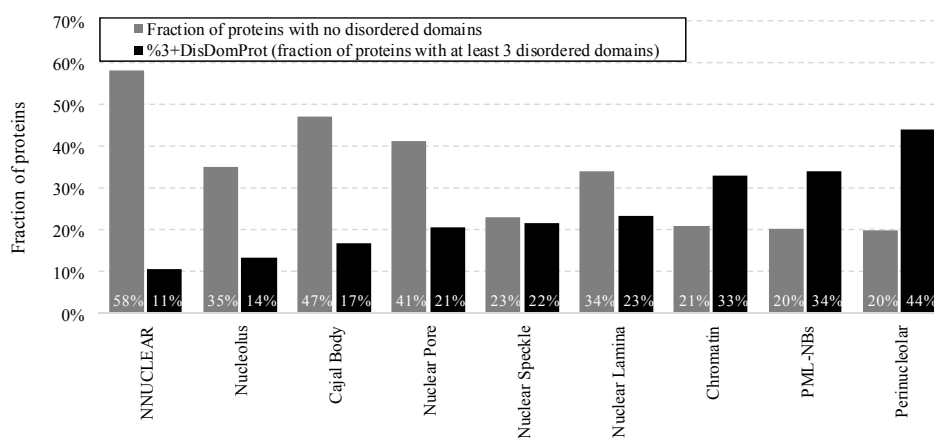
Figure 2 compares other characteristics of the intrinsic disorder including fraction of disordered proteins and normalized (by size of protein chains) fraction of disordered domains (corresponding Figure S3 is based only on the experimental annotations of localization).

We observe that most of the intra-nuclear compartments are enriched in disordered proteins; 27% of proteins in the nucleolus are disordered, around 30% in Cajal body, PML nuclear bodies and chromatin, and 40% or more in the perinucleolar compartment and nuclear speckle, compared to 17% in the non-nuclear proteins. Similarly, between 2.3 and 3.2 disordered domains per unit of 1000 residues can be found in the proteins from across all compartments, except for the nuclear pore (Figure 2). This is a large increase over the 1.7 domain per 1000 residues for the non-nuclear proteins.



**Figure 2.** Fraction of disordered proteins (black bars) and normalized number of disordered domains (gray line) for the non-nuclear proteins from the NNUCLEAR dataset and proteins in the considered intra-nuclear compartment from the NUCLEAR<sub>ap</sub> dataset. Disorder was annotated with the consensus of Espritz and IUPred.

Furthermore, Table S1 (see %DisDomProt values) reveals that between 59% (nuclear pore) and 80% (PML nuclear bodies and perinucleolar compartment) of proteins in the intra-nuclear compartments include disordered domains. Figure 3 compares the proportions of proteins that have no disordered domains and proteins that have at least three disordered domains (%3+DisDomProt) in each compartment and for the non-nuclear proteins (corresponding Figure S4 shows results on the NUCLEAR<sub>a</sub> dataset).

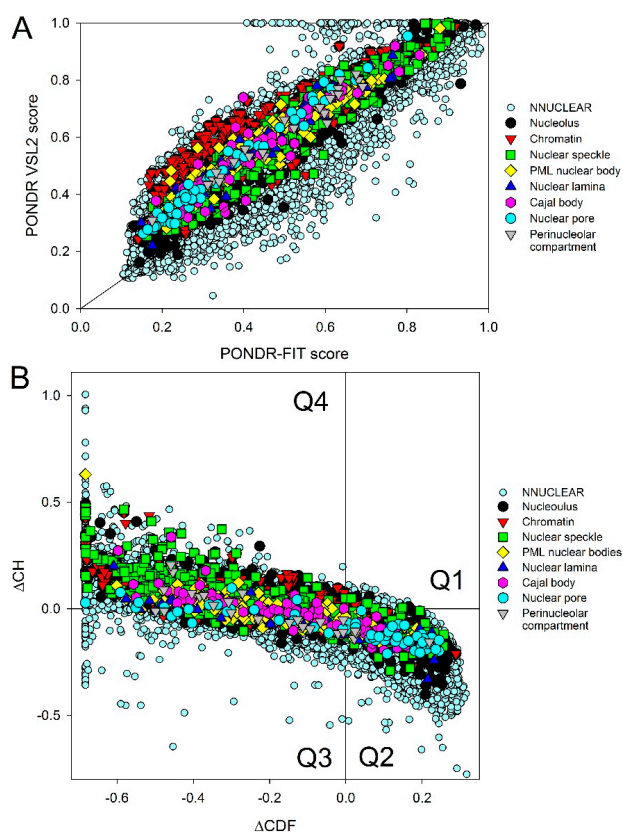


**Figure 3.** Fraction of proteins with no disordered domains (gray bars) and with at least 3 disordered domains (%3+DisDomProt) (black bars) for the non-nuclear proteins from the NNUCLEAR dataset and proteins in the considered intra-nuclear compartment from the NUCLEAR<sub>ap</sub> dataset. Disorder was annotated with the consensus of Espritz and IUPred.

The majority of the non-nuclear proteins are structured and have no disordered domains and only 11% of these proteins have at least three such domains. In contrast, only between 20% (PML nuclear bodies and perinucleolar compartment) and 47% (Cajal body) of proteins localized in the intra-nuclear compartments have no disordered domains. Moreover, in several compartments (nuclear pore, speckle, and lamina) over 20% of proteins have three or more disordered domains, while in chromatin, PML nuclear bodies and perinucleolar compartment, these fractions reach well over 30%.

The point that mouse nuclear proteins analyzed in this study are mostly disordered is further illustrated by Figure 4A that compares the per-protein propensities for disorder (computed as average of the corresponding per-residue propensities) evaluated for the considered nuclear proteins by

two additional predictors: PONDR-FIT [68] and by PONDR<sup>®</sup> VSL2 [69]. Figure 4A indicates that the disorder prediction results obtained by both computational tools mostly agree with each other. Figure 4A and Table 2 show that mouse proteins in various nuclear compartments (except for the nuclear pore) are noticeably more disordered than the non-nuclear proteins. In fact, our analysis with PONDR-FIT (PONDR<sup>®</sup> VSL2) indicates that the non-nuclear proteins are characterized by the mean per-protein propensity for disorder of  $0.36 \pm 0.17$  ( $0.46 \pm 0.18$ ), whereas the corresponding values for the nuclear proteins range between  $0.39 \pm 0.17$  ( $0.52 \pm 0.17$ ) for the nuclear lamina and  $0.49 \pm 0.15$  ( $0.61 \pm 0.15$ ) for the perinucleolar compartment, excluding nuclear pore where they equal  $0.34 \pm 0.16$  ( $0.47 \pm 0.18$ ) (Table 2).



**Figure 4.** Abundance of intrinsic disorder in mouse proteins. (A) Per-proteins propensities for disorder (average of the corresponding per-residue propensities) evaluated by PONDR-FIT ( $x$ -axis) and by PONDR<sup>®</sup> VSL2 (85) ( $y$ -axis) for the non-nuclear proteins from the NNUCLEAR dataset and proteins in the considered intra-nuclear compartment from the NUCLEAR<sub>ap</sub> dataset; (B) Evaluation of intrinsic disorder in mouse proteins (NNUCLEAR and NUCLEAR<sub>ap</sub> datasets) by CH-CDF analysis. In this figure, each point's coordinates were computed as follows: the  $y$ -axis represents the distance of the protein in the CH-plot from the boundary, while the  $x$ -axis shows the average distance between the CDF boundary and the respective CDF curve. The figure is split into four quadrants/sections, with each quadrant representing a different family of predictions. Q1 shows proteins predicted to be ordered by CDFs but disordered by CH-plots. Q2 displays proteins that are completely ordered. Q3 shows proteins that were predicted as compact by CH-plots but disordered by CDFs (*i.e.*, hybrid proteins or putative molten globules). Finally, Q4 portrays proteins which both methods predicted to be disordered. Proteins found in different sub-nuclear compartments are indicated by differently colored symbols. In both plots, mouse non-nuclear proteins are shown by small light blue circles. Proteins in different nuclear compartments are indicated as follows: nucleolus—black circles; chromatin—red inverted triangles; nuclear speckles—green squares; PML nuclear bodies—yellow diamonds; nuclear lamina—blue triangles; Cajal bodies—pink hexagons; nuclear pore—cyan circles; and perinucleolar compartment—gray inverted triangles.

**Table 2.** Evaluation of the disorder status for the non-nuclear proteins from the NNUCLEAR dataset and proteins in the considered intra-nuclear compartment from the NUCLEAR<sub>ap</sub> dataset.

Intra-Nuclear Compartments	Disorder Characteristics by the Per-Residue Predictors			Distribution within the CH-CDF Plot			
	PONDR-FIT <sup>a</sup>	PONDR VSL2 <sup>a</sup>	Two HT Consensus <sup>b</sup>	Q1 (%)	Q2 (%)	Q3 (%)	Q4 (%)
Perinucleolar	0.49 ± 0.15	0.61 ± 0.15	36 ± 21	0.0	16.0	48.0	36.0
Nuclear speckle	0.48 ± 0.17	0.60 ± 0.17	38 ± 25	0.8	25.3	28.5	45.4
Chromatin	0.45 ± 0.15	0.61 ± 0.14	33 ± 21	0.7	15.7	33.9	49.7
PML	0.43 ± 0.14	0.57 ± 0.13	31 ± 19	0.0	22.4	48.7	28.9
Nucleolus	0.43 ± 0.16	0.55 ± 0.16	29 ± 22	1.0	34.6	29.4	35.0
Cajal body	0.42 ± 0.14	0.54 ± 0.13	30 ± 23	0.0	37.5	26.4	36.1
Nuclear lamina	0.39 ± 0.17	0.52 ± 0.17	22 ± 17	0.0	44.7	35.1	20.2
Nuclear pore	0.34 ± 0.16	0.47 ± 0.18	17 ± 16	0.0	61.9	23.8	14.3
Non-nuclear	0.36 ± 0.17	0.46 ± 0.18	20 ± 22	0.3	58.0	23.5	19.2
Total	0.37 ± 0.17	0.47 ± 0.18	21 ± 22	0.4	55.4	24.1	20.1

<sup>a</sup> For PONDR-FIT and PONDR<sup>®</sup> VSL2, Table 2 reports the average of the per-protein propensities for disorder and the corresponding standard deviations calculated for each dataset; <sup>b</sup> Content of predicted disordered residues in a given set of proteins based on a majority vote consensus of two high-throughput predictors, Espritz [70] and IUPred [71].

Therefore, according to these average per-protein propensities evaluated by PONDR-FIT, the intra-nuclear compartments can be arranged in the following order: perinucleolar compartment > nuclear speckles > chromatin > nucleolus = PML nuclear bodies > Cajal bodies > nuclear lamina > non-nuclear proteins > nuclear pore. These results are analogous to the results generated with the consensus of Espritz and IUPred (Table 2 and Figure 1).

The point that mouse nuclear proteins analyzed in this study are mostly disordered is further illustrated by Figure 4A that compares the per-protein propensities for disorder (computed as average of the corresponding per-residue propensities) evaluated for the considered nuclear proteins by two additional predictors: PONDR-FIT [68] and by PONDR<sup>®</sup> VSL2 [69]. Figure 4A shows that the results of the disorder predictions obtained by these two computational tools mostly agree with each other. Figure 4A and Table 2 show that mouse proteins in various nuclear compartments (except for the nuclear pore) are noticeably more disordered than the non-nuclear proteins. In fact, our analysis with PONDR-FIT (PONDR<sup>®</sup> VSL2) indicates that the non-nuclear proteins are characterized by the mean per-protein propensity for disorder of  $0.36 \pm 0.17$  ( $0.46 \pm 0.18$ ), whereas the corresponding values for the nuclear proteins range between  $0.39 \pm 0.17$  ( $0.52 \pm 0.17$ ) for the nuclear lamina and  $0.49 \pm 0.15$  ( $0.61 \pm 0.15$ ) for the perinucleolar compartment, excluding nuclear pore where they equal  $0.34 \pm 0.16$  ( $0.47 \pm 0.18$ ) (Table 2). Therefore, according to these average per-protein propensities evaluated by PONDR-FIT, the intra-nuclear compartments can be arranged in the following order: perinucleolar compartment > nuclear speckles > chromatin > nucleolus = PML nuclear bodies > Cajal bodies > nuclear lamina > non-nuclear proteins > nuclear pore. These results are analogous to the results generated with the consensus of Espritz and IUPred (Table 2 and Figure 1).

To understand how mouse nuclear proteins can be grouped into different disorder categories, we examine the details behind their distribution within the CH-CDF phase space. Figure 4B shows that, based on their location in this space, mouse nuclear proteins can be subdivided in three major groups, with the vast majority of proteins in almost all nuclear suborganelles (except for nuclear pore proteins) are predicted as disordered by both CDF and CH, or predicted as compact by CH but disordered by CDF (see also Table 2). In fact, these mostly disordered proteins are located within the quadrants 3 and 4 of the CH-CDF plot, respectively, and the overall content of such Q3 + Q4 proteins in different sub-nuclear compartments ranges from 84% in the perinucleolar compartment to 83.6% in chromatin, to 77.6% in PML nuclear bodies, to 73.9% in nuclear speckles, to 64.4% in nucleolus, to 62.5% in Cajal bodies, to 55.3% in nuclear lamina, and to 38.1% in nuclear pore. Therefore, the vast majority of nuclear proteins are expected to be either extended IDPs (behave as native pre-molten globules or native coils), or the potential native molten globules/hybrid proteins. Furthermore, statistical analysis (Pearson's Chi-squared test with the Yates' continuity correction) of these data revealed that the portion of Q4



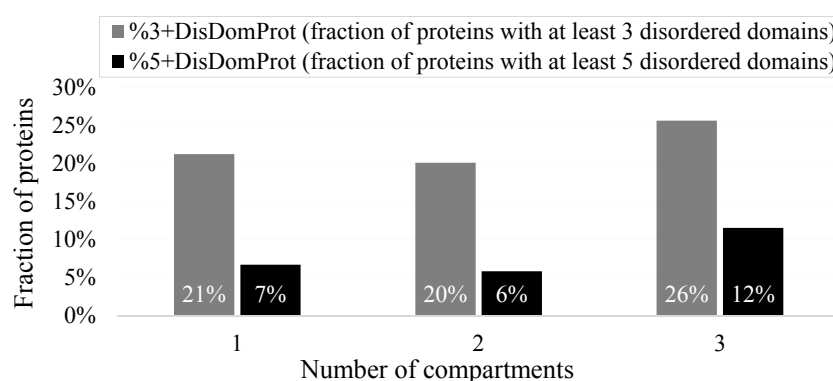
proteins (*i.e.*, proteins that are expected to behave as native pre-molten globules or native coils) in Cajal body, chromatin, nucleolus, PML nuclear bodies, nuclear speckles, and perinucleolar compartment is significantly larger than the portions of proteins in other quadrants. Table 2 also shows that on average only ~29% of mouse proteins in various sub-nuclear domains are predicted as ordered by both CH and CDF and therefore are expected to be mostly ordered. There are only a very few proteins in this set which are predicted to be ordered by CDF and disordered by CH-plot analysis (typically, less than 1%). Figure 4B and Table 2 also show that the situation is very different for the mouse non-nuclear proteins, which are split between the four quadrants of the CH-CDF plot in the 0.3% (Q1):58% (Q2):23.5% (Q3):19.2% (Q4) proportion.

Altogether, our results that employ multiple methodologies clearly demonstrate that the intra-nuclear compartments (with the exception of the nuclear pore) are highly enriched in disordered residues, disordered domains, and disordered proteins.

### 2.3. Intrinsic Disorder in Proteins Co-Localized in Multiple Intra-Nuclear Compartments

Table S3 shows that while the majority of the nuclear proteins are assigned to a single intra-nuclear compartment, a large number is co-localized in multiple compartments. Approximately 19% of the nuclear proteins are assigned to two compartments and another 2.5% (78 proteins) to three compartments; only 10 proteins are associated with over three compartments and we have not considered them in our analysis due to the low count. The disorder content and fraction of disordered proteins is similar between the proteins assigned to one and to multiple compartments (Table S3). However, the fraction of proteins that have several disordered domains that are co-localized in three compartments is much higher compared to the proteins localized in one and two compartments (Figure 5). More specifically, about 26% proteins that are co-localized in three compartments have at least three disordered domains compared to 20% for proteins assigned to fewer compartments, which corresponds to the enrichment by 30%. The difference is even larger when considering proteins that have at least five disordered domains where the enrichment reaches 100%, from 6% to 12%. This result suggests the disordered domains play a role in the nuclear protein moonlighting.

Table S4 provides more detailed characterization of several mouse nuclear proteins that are co-localized in at least three sub-nuclear compartments and have at least five disordered domains. For this sample set, we predicted potential disorder-based binding sites using the ANCHOR algorithm [72,73] and also analyzed disorder content by three consensus approaches, MobiDB [74,75], PONDR-FIT [68], and the majority vote consensus of two high-throughput predictors, Espritz [70] and IUPred [71].



**Figure 5.** Fraction of proteins with 3 or more (%3+DisDomProt) and 5 or more (%5+DisDomProt) disordered domains for the nuclear proteins from the NUCLEAR<sub>ap</sub> dataset localized in 1, 2, and 3 intra-nuclear compartments. Disorder was annotated with the consensus of Espritz and IUPred.

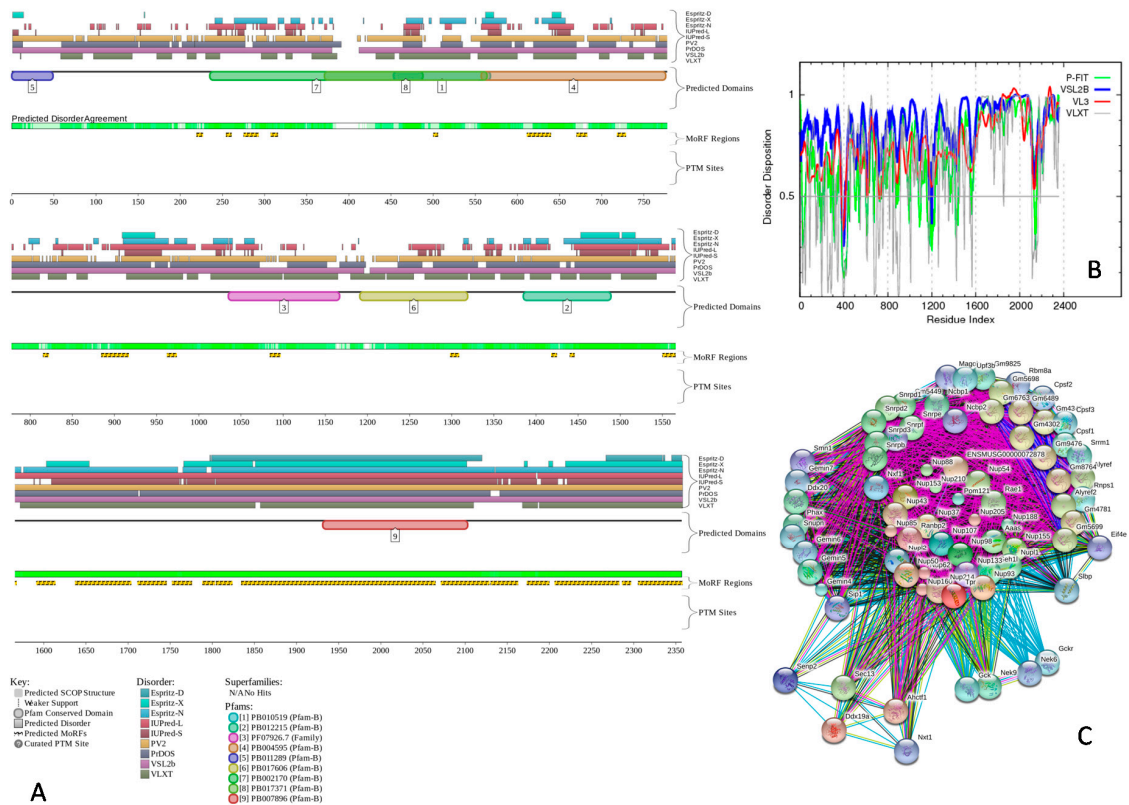
Table S4 shows a strong consensus between these three consensus-based approaches, which all show that proteins in the sample set are highly disordered and contain numerous potential disorder-based protein binding sites, AIBSs (ANCHOR-identified binding sites). In fact, the number of AIBSs ranges from 11 in transcription factor Sp7 (UniProt ID: Q8VI67) to 77 in Msx2-interacting protein (UniProt ID: Q62504) and the percentage of residues predicted to be involved in disorder-based protein interactions ranges from 12.7% in E3 SUMO-protein ligase RanBP2 (UniProt ID: Q9ERU9) to 51.1% in Msx2-interacting protein (UniProt ID: Q62504). On average, these selected nuclear multifunctional proteins are predicted to contain 16.4 AIBSs per 1000 residues, with 29.9% of residues in these proteins being predicted to be interaction-prone. The presence of so many AIBSs in all proteins analyzed here suggests that these mouse nuclear proteins that are co-localized in at least three sub-nuclear compartments and have at least five disordered domains are promiscuous binders that can be involved in various interactions with multiple unrelated partners and/or participate in the multivalent binding by wrapping around partners.

To give a further illustration of the abundance and functionality of intrinsic disorder in nuclear proteins, Figures 6 and 7 represent the results of the analysis of two of the proteins found in multiple sub-nuclear compartments, nuclear pore complex-associated intra-nuclear coiled-coil protein TPR (UniProt ID: Q7M739), and bloom syndrome protein homolog (UniProt ID: O88700) by D<sup>2</sup>P<sup>2</sup>, multiple per-residue predictors of the PONDR family, and STRING. These two proteins were selected from the subset of nuclear proteins shown in Table 3 as one of the most disordered and one of the least disordered proteins, respectively. The results of the analogous analyses for the remaining proteins from Table S4 are presented in Figures S5–S11.

First, the data corresponding to these two proteins were retrieved from the D<sup>2</sup>P<sup>2</sup> database [55] that contains information on predicted disorder and selected disorder-related functions. Besides providing information on the distribution of disordered regions predicted by nine tools in a query protein, the D<sup>2</sup>P<sup>2</sup> visual console is further enhanced by providing information on the location of predicted disorder-based protein binding sites and on the curated sites of various posttranslational modifications (see Materials and Methods). This analysis confirmed that the selected nuclear proteins are predicted to possess long disordered regions, contain numerous binding sites and multiple sites of various posttranslational modifications (see Figures 6A and 7A, and Figures S5–S11). The fact that disordered domains/regions of these nuclear proteins have several posttranslational modification sites is in agreement with the well-known fact that phosphorylation [76] and many other enzymatically catalyzed posttranslational modifications are preferentially located within the IDPRs [77].

**Table 3.** Number of proteins and hubs in the intra-nuclear compartments.

Intra-Nuclear Compartment	Number of Proteins in NUCLEAR <sub>a</sub>	Number of Proteins in NUCLEAR <sub>ap</sub>	PPI <sub>NUCLEARap</sub>		
			Number of Proteins	Number of Hubs	Number of Intra-Compartment Hubs
Cajal Body	49	72	50	15	11
Chromatin	323	828	500	192	145
Nuclear Lamina	77	94	70	19	14
Nuclear Pore	51	63	49	15	11
Nuclear Speckle	403	632	459	138	104
Nucleolus	598	1860	1227	392	335
Perinucleolar	24	25	24	14	9
PML nuclear bodies	91	187	140	71	49

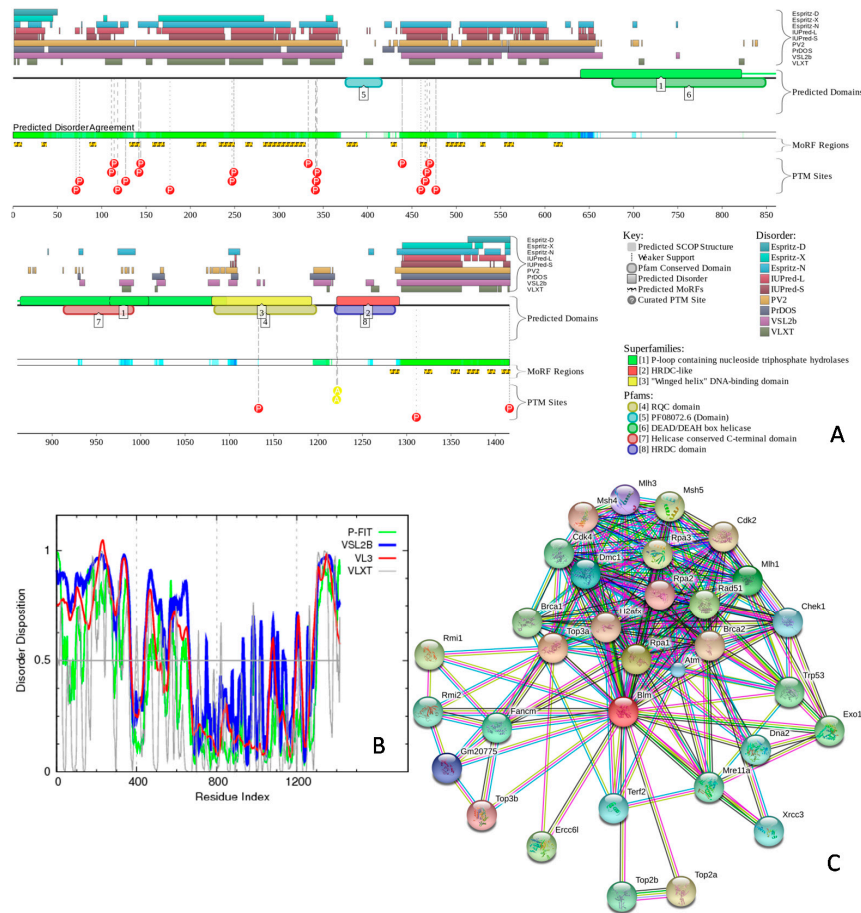


A

B

C

**Figure 6.** Abundance and functionality of intrinsic disorder in the nuclear pore complex-associated intra-nuclear coiled-coil protein TPR (UniProt ID: Q7M739). (A) Evaluation of the functional intrinsic disorder propensity by the D<sup>2</sup>P<sup>2</sup> database (<http://d2p2.pro/>) [55]. In the corresponding plot, the colored bars at the top are representative of the location of disordered regions obtained from different disorder predictors (including IUPred-L, IUPred-S, Espritz-D, Espritz-N, Espritz-X, PONDR<sup>®</sup> VSL2b, PONDR<sup>®</sup> VLXT, PV2, and PrDOS). Location of predicted and known domains are shown by numbered colored bars. The white-and-green bars display the agreement in the predicted disorder among the aforementioned predictors, with green parts indicating regions that are disordered by consensus. The yellow bar indicates the location of the predicted disorder-based binding site (MoRF region), whereas red and yellow circles at the bottom of the plots show locations of phosphorylation and acetylation sites, respectively. Vertical dashed lines show actual positions of the phosphorylation sites; (B) Evaluation of the per-residue disorder propensity based on predictors from the PONDR family. A thin line at a score of 0.5 represents the disorder threshold between disorder (>0.5) and order (<0.5); (C) Analysis of the interactivity of the nuclear pore complex-associated intra-nuclear coiled-coil protein TPR (UniProt ID: Q7M739) by STRING computational platform [78]. STRING produces the network of predicted associations for a particular protein and its interactome. The nodes in this network are proteins, while the predicted or known functional associations are the edges. When predicting the associations, seven types of evidence are used, which are indicated in the resulting network by the differently colored lines. Here, a green line represents neighborhood evidence; a red line—the presence of fusion evidence; a purple line—experimental evidence; a blue line—co-occurrence evidence; a light blue line—database evidence; a yellow line—text mining evidence; a black line—co-expression evidence [78].



**Figure 7.** Functionality and prevalence of intrinsic disorder in the bloom syndrome protein homolog (UniProt ID: O88700). (A) Evaluation of the functional intrinsic disorder propensity by the D<sup>2</sup>P<sup>2</sup> database (<http://d2p2.pro/>) [55]. In the corresponding plot, the colored bars at the top are representative of the location of disordered regions obtained from different disorder predictors (including IUPred-L, IUPred-S, Espritz-D, Espritz-N, Espritz-X, PONDR<sup>®</sup> VSL2b, PONDR<sup>®</sup> VLXT, PV2, and PrDOS). Location of predicted and known domains are shown by numbered colored bars. The white-and-green portion displays the agreement in the predicted disorder among the aforementioned predictors, with green parts indicating regions that are disordered by consensus. The yellow bar indicates the location of the predicted disorder-based binding site (MoRF region), whereas red circles with “P” inside and yellow circles with “A” inside at the bottom of the plots show locations of phosphorylation and acetylation sites, respectively. Vertical dashed lines show actual positions of these phosphorylation and acetylation sites; (B) Evaluation of the per-residue disorder propensity based on predictors from the PONDR family. A thin line at a score of 0.5 represents the disorder threshold between disorder (>0.5) and order (<0.5); (C) Analysis of the interactivity of the nuclear pore complex-associated intra-nuclear coiled-coil protein TPR (UniProt ID: Q7M739) by STRING computational platform [78]. STRING produces the network of predicted associations for a particular protein and its interactome. The nodes in this network are proteins, while the predicted or known functional associations are the edges. When predicting the associations, seven types of evidence are used, which are indicated in the resulting network by the differently colored lines. Here a green line represents neighborhood evidence; a red line—the presence of fusion evidence; a purple line—experimental evidence; a blue line—co-occurrence evidence; a light blue line—database evidence; a yellow line—text mining evidence; a black line—co-expression evidence [78].

Next, the per-residue intrinsic disorder propensities of the selected two proteins (see Figures 6B and 7B) and remaining proteins from Table S4 (see Figures S5–S11) were analyzed by various predictors

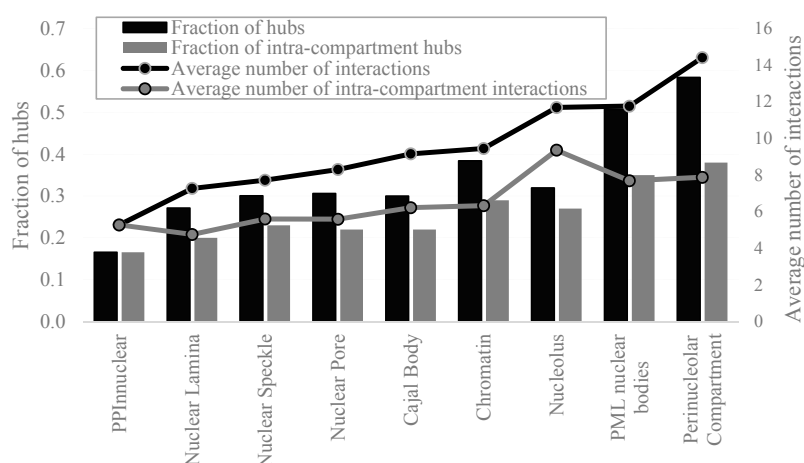
from the PONDR family, such as PONDR<sup>®</sup> FIT [68], PONDR<sup>®</sup> VSL2 [69], PONDR<sup>®</sup> VLXT [79], and PONDR<sup>®</sup> VL3 [80]. Here, a disorder score above 0.5 corresponds to disordered residues/regions. This analysis provides additional support to the idea that the selected mouse nuclear proteins co-localized in several nuclear compartments are predicted to contain numerous IDPRs.

Finally, the interactivity of these selected nuclear proteins was further evaluated by STRING (Search Tool for the Retrieval of Interacting Genes) database [78]. This database produces the network of predicted associations for a particular protein and its interactome (see Materials and Methods). Figures 6C and 7C represent the results of STRING analysis of nuclear pore complex-associated intra-nuclear coiled-coil protein TPR and bloom syndrome protein homolog, whereas Figures S5–S11 for the remaining proteins from Table S4. This analysis shows that all proteins are expected to have rather well developed interactomes, likely serving as hub proteins in their functional protein-protein interaction networks (PPI). It was pointed out that intrinsic disorder represents an important feature defining the ability of a hub protein to be a promiscuous binder [43,59,81–85]. In fact, intrinsic disorder is always associated with the ability to serve as a hub, and many hubs are intrinsically disordered or contain functional IDPRs, whereas partners of ordered hubs are known to be abundantly disordered [43,59,81–85].

#### 2.4. Protein-Protein Interactions and Intrinsic Disorder in Intra-Nuclear Compartments

Motivated by the abundance of protein-binding regions and the large sizes of interactomes of proteins that are co-localized in multiple intra-nuclear compartments (Figures 6 and 7 and Figures S5–S11), we analyze PPIs in the considered compartments.

Table 3 shows number of proteins and hubs in the intra-nuclear compartments, whereas Table S2 summarizes results for the nuclear and non-nuclear proteins that are included in the PPI network (PPI<sub>NUCLEARap</sub> and PPI<sub>NNUCLEAR</sub> datasets). We also count intra-compartment hubs, *i.e.*, proteins that are hubs when excluding their interactions with proteins located in other intra-nuclear compartments (interactions across compartments). The nuclear proteins are characterized by a substantial enrichment in hubs and average (per protein) number of PPIs. While the non-nuclear proteins include about 17% of hubs, proteins localized in the intra-nuclear compartments have between 27% (nuclear lamina) and 58% of hubs (perinucleolar compartment) (Figure 8). The average number of PPIs for the non-nuclear proteins is at 5.3 and it goes up to between 7.3 and 14.4 for the proteins localized in the intra-nuclear compartments (Figure 8).

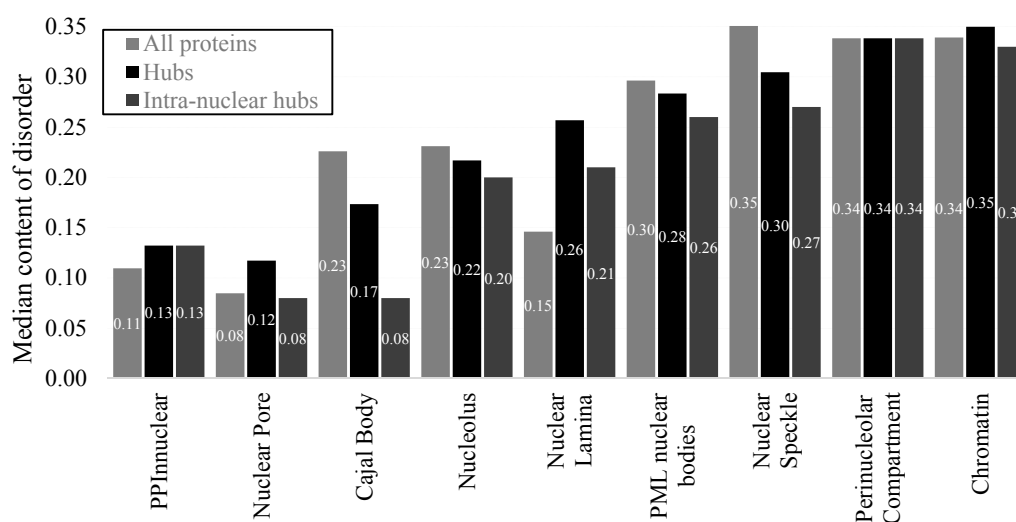


**Figure 8.** Fraction of hubs (black bars), intra-compartment hubs (gray bars), average number of interactions (black line), and average number of intra-compartment interactions (gray line) for the non-nuclear proteins from the PPI<sub>NNUCLEAR</sub> dataset and proteins in the considered intra-nuclear compartment from the PPI<sub>NUCLEARap</sub> dataset. Disorder was annotated with the consensus of Espritz and IUPred.

Even when excluding the interactions across compartments, the fraction of intra-compartment hubs is substantially higher in all nuclear compartments, between 20% and 38%, as compared to the fraction of hubs among the non-nuclear proteins (Figure 8). Similarly, the average number of PPIs for all compartments except for the nuclear lamina is higher even when excluding the inter-compartment interactions. The corresponding highest enrichment in the PPIs is for the nucleolus, PML nuclear bodies, and perinucleolar compartment (Figure 8).

Table S5 provides details of the abundance of the disorder in the hub proteins across all considered intra-nuclear compartments and the non-nuclear proteins. Interestingly, the non-nuclear hubs are represented by heightened amounts of disorder when viewed against the overall population of non-nuclear proteins (median disorder content of 0.13 *vs.* 0.11, Figure 9), which agrees with previously reported data [83]. We observe that the proteins localized in the intra-nuclear compartments have already high amounts of disordered and similarly high disorder content is characteristic for the corresponding hubs (Figure 9). Hubs in all compartments, except for the nuclear pore, are substantially enriched in the disorder compared to both non-nuclear hubs and all non-nuclear proteins. Their median disorder content varies between 0.17 (Cajal body) and 0.35 (chromatin). Over 75% of hubs have disordered domains and over 30% of hubs are disordered in several intra-nuclear compartments including nuclear speckle, PML nuclear bodies, perinucleolar compartment and chromatin (Table S5). Figure 9 also shows that the intra-nuclear hubs (when we exclude the inter-nuclear compartment interactions) are also enriched in disorder for the majority of the compartments, except for the nuclear pore and cajal body. The intra-nuclear hubs in these two compartments are depleted in disorder as compared to all hubs, which means that disorder is likely involved in the inter-compartment interactions. The disorder content in the other six inter-nuclear compartments is comparable between all hubs and the intra-nuclear hubs.

We conclude that the density of the PPIs for the nuclear proteins is unusually high. The intra-nuclear compartments are rich in hub proteins and these hubs are substantially enriched in disorder and include a large fraction of disordered proteins when compared to the non-nuclear hubs. When excluding interactions across nuclear compartments, we still observe enrichment in the abundance of hub proteins and their disorder content for the majority of the compartments.



**Figure 9.** Median disorder content for all proteins (light gray bars), hubs (black bars), and intra-nuclear hubs (dark gray bars) from a given protein set including the non-nuclear proteins in the PPI<sub>NNUCLEAR</sub> dataset and proteins in the considered intra-nuclear compartment from the PPI<sub>NUCLEAR<sub>app</sub></sub> dataset. Disorder was annotated with the consensus of Espritz and IUPred.

### 3. Materials and Methods

#### 3.1. Datasets

Our analysis compares nuclear proteins and subgroups of nuclear proteins localized in specific intra-nuclear compartments with other proteins in the same proteome. Thus, we collected a complete set of nuclear and non-nuclear proteins. We use mouse proteome due to the availability of well-annotated intra-nuclear compartments for its nuclear proteins and a relatively comprehensive PPI network. We collected nuclear proteins from the intra-nuclear compartment protein database Nsort/DB [86]. This comprehensive database aggregates data from multiple sources including Nuclear Protein Database (NPD) [87], Nuclear Proteome Database (NOPdb) [88], Nuclear Matrix Protein Database (NMPdb) [89], UniProt [90] and Human Protein Reference Database (HPRD) [91]. The complete nuclear proteome contains 3566 proteins where 1285 proteins are assigned into at least one intra-nuclear compartment based on experimental evidence [86] (they compose  $NUCLEAR_n$  dataset, see Table 1) and another 1720 proteins for which the assignment was predicted [92]. The  $NUCLEAR_{ap}$  dataset combines proteins with annotated and predicted intra-nuclear localization. We mapped the complete set of nuclear proteins into the complete mouse proteome to create the set of non-nuclear proteins. The mouse proteome was collected from UniProt (release from August 2013) and we excluded protein fragments. The mapping was done primarily based on the accession numbers; in 233 cases where accession numbers did not match we used sequence identity. The resulting NNUCLEAR dataset includes 29,863 mouse proteins that exclude the nuclear proteins.

Furthermore, we analyzed differences in PPIs between nuclear proteins, including subsets of nuclear proteins localized in specific intra-nuclear compartments, and the non-nuclear proteins. We collected the mouse PPI network from the Mentha database (release from October 2013) [93] which includes 8206 proteins that have at least one interaction. We mapped these proteins into the NNUCLEAR and  $NUCLEAR_{ap}$  datasets and collected a set of 5687 non-nuclear proteins ( $PPI_{NNUCLEAR}$  dataset) and 2519 nuclear proteins ( $PPI_{NUCLEAR_{ap}}$  dataset) that are included in the PPI network, respectively (Table 1). We annotate hubs [83], highly connected proteins in the PPI networks, in the  $PPI_{NNUCLEAR}$  and  $PPI_{NUCLEAR_{ap}}$  datasets, including the intra-nuclear compartments in the latter dataset. Based on previously reported criteria [94–96], hubs are defined as the top 20% most connected proteins in the PPI network based on the PPI dataset. The number of hubs in the  $PPI_{NNUCLEAR}$  dataset is 942. The number of proteins and hubs in individual intra-nuclear compartments is summarized in Table 2; note that we cannot annotate hubs in the  $NUCLEAR_n$  and  $NUCLEAR_{ap}$  datasets.

#### 3.2. Annotation and Characterization of Intrinsic Disorder

We generated annotations of intrinsic disorder using a majority vote consensus of two high-throughput predictors. Use of the consensus was shown to lead to an increase in the predictive performance compared to the use of a single predictor [97–99]. We used three versions of Espritz [70], which were designed to predict disorder in proteins annotated based on X-ray crystal structures, NMR-derived structures, and the Disprot database [100]; and two versions of IUPred [71], for prediction of long and short disordered regions. A recent large-scale assessment shows that these predictors offer good predictive performance, with AUC values around 0.77 [97]. We also note that the same consensus was recently utilized in a few related works [56,101,102]. We computed the putative annotations of the disorder for each residue and aggregated them for a given protein or a group of proteins using the following measures:

- 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
- %DisDomProt (fraction of proteins with disordered domains): fraction of proteins that have at least one long disorder region ( $\geq 30$  residues long) in a given protein group. Such long regions are

recognized as functional protein domains [52,103,104], which means that proteins with the long disorder regions are likely to carry functions through disorder

- %3+DisDomProt (fraction of proteins with at least three disordered domains)
- %5+DisDomProt (fraction of proteins with at least five disordered domains)
- %DisDom1 K (number of disordered domains per 1000 residues). This measure normalizes the count of long disordered regions to a size of a given protein or a set of proteins to accommodate for a potential bias that longer proteins may have more disordered domains because of their length.

We calculated these measures for the non-nuclear proteins in NNUCLEAR and PPI<sub>NNUCLEAR</sub>, and the nuclear proteins in each intra-nuclear compartment in NUCLEAR<sub>a</sub>, NUCLEAR<sub>ap</sub>, and PPI<sub>NUCLEARap</sub>. We also did the same calculation for hubs from PPI<sub>NNUCLEAR</sub> and each compartment in PPI<sub>NUCLEARap</sub>.

To further boost confidence of the putative annotations of disorder, we also calculated disorder content for the mouse non-nuclear proteins in the NNUCLEAR dataset and the nuclear proteins in each intra-nuclear compartment in NUCLEAR<sub>ap</sub> dataset using a different metapredictor PONDR-FIT [68]. This method was shown to be moderately more accurate than each of its six component predictors, PONDR<sup>®</sup> VSL2 [69], PONDR<sup>®</sup> VLXT [79], FoldIndex [105], PONDR<sup>®</sup> VL3 [80], TopIDP [106], and IUPred [71].

### 3.3. Disorder Sub-Classification of Mouse Proteins Based on the Charge/Hydrophathy-Cumulative Distribution Function Analysis

The charge/hydrophathy-cumulative distribution function (CH-CDF) analysis represents a simple graphic way of intrinsic disorder classification in whole proteins. In the resulting CD-CDF plot, each point corresponds to a protein with coordinates computed in the following manner: the Y coordinate represents the distance between the boundary and the corresponding protein in the CH-plot (charge-hydrophathy plot) [45,53], while the X coordinate represents the average distance between the CDF boundary and the corresponding cumulative distribution function (CDF) curve [53,107–109]. The use of the CH-CDF plot utilizes the principle difference in the logistics of how the two component binary predictors evaluate the predisposition of a given protein to be disordered or ordered or overall. Here, the CH-plot serves as a linear classifier utilizing the charge and hydrophathy of a particular protein sequence [45,53]. On the other hand, analysis of the CDF relies on the output of a PONDR<sup>®</sup> predictor, which is a nonlinear disorder classifier trained to distinguish order and disorder based on a significantly larger sequence feature space [53]. According to these methodological differences, the CH-plot can discriminate proteins with compact conformations (well-structured globular and molten globule-like proteins) from proteins with substantial amount of extended disorder (pre-molten globules and random coils), whereas CDF analysis may discriminate all types of IDPs, including molten globules, from the ordered proteins. The aforementioned methodological differences constitute a means for the simple visual sub-classification of proteins into ordered proteins, IDPs with extended disorder, molten globular compact IDPs, and hybrid proteins containing ordered domains and IDPRs. In the resulting CH-CDF plot, positive Y values correspond to proteins predicted to be extended IDPs by the CH-plot analysis, while negative Y values indicate proteins predicted to be compact. Meanwhile, positive X values indicate proteins predicted to be ordered by CDF analysis, and negative X values represent predicted IDPs. The resultant quadrants of CDF-CH plots can therefore be summarized as follows: Q1 contains proteins predicted to be ordered by CDFs but disordered by CH-plots. Q2 displays proteins predicted to be completely ordered by both methods. Q3 shows proteins that were predicted as compact by CH-plots but disordered by CDFs (*i.e.*, hybrid proteins or putative molten globules). Finally, Q4 contains proteins which both methods predicted to be disordered [109,110].



### 3.4. Disorder and Disorder—Based Functional Analysis of Selected Mouse Nuclear Proteins with Consensus Disorder Predictors, PONDR Predictors, D<sup>2</sup>P<sup>2</sup>, STRING, and ANCHOR

Disorder evaluations for several selected proteins (which are mouse nuclear proteins that are co-localized in at least three sub-nuclear compartments and have at least five disordered domains) are further analyzed by three consensus-based computational tools for evaluation of intrinsic disorder, PONDR-FIT [68] and the majority vote consensus approach that were described above and MobiDB [74,75]. The MobiDB database (<http://mobidb.bio.unipd.it/>) [74,75], that generates consensus disorder scores by aggregating the output from ten predictors, such as two versions of IUPred [71], two versions of ESpritz [70], two versions of DisEMBL [111], JRONN [112], PONDR<sup>®</sup> VSL2B [80,113], and GlobPlot [114]. Use of consensus for evaluation of intrinsic disorder is motivated by empirical observations that this leads to an increase in the predictive performance compared to the use of a single predictor [97–99].

Per-residue disorder distribution was evaluated by PONDR predictors, including PONDR<sup>®</sup> FIT [68], PONDR<sup>®</sup> VSL2 [69], PONDR<sup>®</sup> VLXT [79], and PONDR<sup>®</sup> VL3 [80]. Here, a disorder score above 0.5 indicated disordered residues/regions. These four predictors were chosen for several reasons. PONDR<sup>®</sup> VSL2 was chosen due to being one of the best standalone predictors [69,98,115]. PONDR<sup>®</sup> VLXT is useful for identifying disorder-based interaction sites since it is sensitive to minute aberrations in local sequences [79]. PONDR<sup>®</sup> VL3 was used since it is very accurate when evaluating long disordered regions [80], whereas PONDR-FIT is already described above. Again, we use multiple disorder predictors here to increase confidence of our results.

For the same selected mouse nuclear proteins co-localized in several sub-nuclear compartments, disorder evaluations together with important disorder-related functional information were retrieved from the D<sup>2</sup>P<sup>2</sup> database (<http://d2p2.pro/>) [55]. D<sup>2</sup>P<sup>2</sup> is a database of predicted disorder that serves as a collaborative resource for pre-computed disorder predictions. Its contents include results for a large library of proteins from completely sequenced genomes [55]. D<sup>2</sup>P<sup>2</sup> database uses outputs of IUPred [71], PONDR<sup>®</sup> VLXT [79], PrDOS [116], PONDR<sup>®</sup> VSL2B [80,113], PV2 [55], and ESpritz [70]. The database is further supplemented by data concerning location of various curated posttranslational modifications and predicted disorder-based protein binding sites.

Additional functional information for these proteins was retrieved using the STRING platform (Search Tool for the Retrieval of Interacting Genes; <http://string-db.org/>) which offers both predicted and information on the interactions of a protein of interest [78]. For a query protein, STRING produces the network of predicted associations with a particular group of proteins. The nodes in this network are proteins, while the predicted or known functional associations are the edges. When predicting the associations, seven types of evidence are used, which are indicated in the resulting network by the differently colored lines. Here a green line represents neighborhood evidence; a red line—the presence of fusion evidence; a purple line—experimental evidence; a blue line—co-occurrence evidence; a light blue line—database evidence; a yellow line—text mining evidence; a black line—co-expression evidence [78]. In our analysis, the most stringent criteria were used for selection of interacting proteins by choosing the highest cut-off of 0.9 as the minimal required confidence level.

Potential protein binding sites in disordered regions of selected mouse nuclear proteins that are co-localized in at least three sub-nuclear compartments and have at least five disordered domains were identified by the ANCHOR algorithm [72,73]. This algorithm utilizes the pair-wise energy estimation approach originally used by IUPred [71,117]. This approach acts on the hypothesis that long regions of disorder include localized potential binding sites which are not capable of folding on their own due to not being able to form enough favorable intrachain interactions, but can obtain the energy to stabilize via interaction with a globular protein partner [72,73]. The ANCHOR algorithm generates regions with a high likelihood to be binding sites, which are referred to as ANCHOR-indicated binding site (AIBS).

### 3.5. Statistical Analysis

We compared the median and mean disorder content and the values of %DisProt, %DisDomProt, %3+DisDomProt, %5+DisDomProt, and %DisDom1K computed for the nuclear proteins in the intra-nuclear compartments from NUCLEAR<sub>a</sub> or NUCLEAR<sub>ap</sub> (PPI<sub>NUCLEARap</sub>) with the corresponding values for the non-nuclear proteins from the NNUCLEAR (PPI<sub>NNUCLEAR</sub>) dataset. We assessed statistical significance of the differences in these values between proteins in a given intra-nuclear compartment and the corresponding non-nuclear proteins following the procedure from Refs [56,118]. We determined the significance between two sets of values of a given measurement computed over ten subsets of randomly chosen half of proteins in the corresponding two protein sets. More specifically, we selected half of the proteins in a given intra-nuclear compartment and the same number of non-nuclear proteins with similar size (with  $\pm 10\%$  tolerance) compared to the selected nuclear proteins; we repeated that ten times. We matched the size since the amount of disorder was shown to be dependent on the protein chain length in eukaryotes [101]. We used *t*-test for measurements that follows normal distribution as evaluated with the Anderson-Darling test [119] at the 0.05 significance; otherwise, we used the Wilcoxon rank sum test [120]. We assumed that a given difference is significant if the *p*-value < 0.01.

## 4. Conclusions

This work provides support to the idea of the functional importance of IDPs in the cell nucleus and shows that many sub-nuclear organelles in nuclei of mouse cells are enriched in IDPs and hybrid proteins containing disordered and ordered domains. We also show that the mouse nuclear proteins are very promiscuous binders possessing both large quantities of potential disorder-based interaction sites and the ability of a single such site to be involved in a large number of interactions. In fact, we reveal that intrinsic disorder (overall content, number of disordered proteins, and number of disordered domains) is enriched in the majority of the mouse intra-nuclear compartments, except for the nuclear pore and lamina. Intra-nuclear compartments are depleted in proteins that lack disordered domains and enriched in proteins that have multiple disordered domains. Proteins that are co-localized in multiple intra-nuclear compartments are more likely to have multiple disordered domains, which could be related to their moonlighting functions. Protein-protein interaction networks in the mouse intra-nuclear compartments are denser and include more hub proteins compared to the non-nuclear proteins. Hubs in the intra-nuclear compartments (except for the nuclear pore) are enriched in disorder and include more disordered proteins compared with non-nuclear hubs and non-nuclear proteins. These results were obtained based on analysis with several different tools that provided unanimous support for the above observations.

Since the dataset analyzed here contains more than 3000 entries, it is physically impossible to provide even a brief description of all these mouse nuclear proteins. However, our observations are in agreement with known experimental and computational data for several nuclear proteins. For example, the exceptional functional importance of intrinsic disorder has been shown for histones [20,121–138], nucleoporins [139–142], as well as for ribosomal [102] and spliceosomal proteins [143,144]. These proteins are found in chromatin, nuclear pore, nucleolus, and Cajal bodies, respectively. Intrinsic disorder is also known to play an important role in function of transcription factors [145,146].

Lastly, we comment on the potential role of intrinsic disorder in assembly and disassembly of the sub-nuclear membrane-less organelles. Recently, it has been proposed that IDPs may play an important role in driving the intracellular liquid-liquid phase separations generating various nucleoplasmic and cytoplasmic membrane-less organelles [147]. Although these organelles are found in different cellular and nuclear locations, and although they are composed of rather different proteins and nucleic acids, it is believed that such membrane-less organelles are formed via a common mechanism related to the intracellular phase transitions [42]. Such transitions occur due to the changes to the structure and solvent properties of water that are driven by macromolecules, and can usually be found in high concentrations of macromolecular solutes. This happens because the solution exists

as a single phase at low macromolecule concentrations but experiences phase separation at higher concentrations [148]. An earlier study revealed that several of the proteins that play a role in the formation of the nucleoplasmic or cytoplasmic membrane-less organelles are in fact intrinsically disordered [66,147]. It was also hypothesized that because the IDPs are known to be engaged in various weak interactions of a different physico-chemical nature and because these proteins are commonly seen in different cytoplasmic and nuclear membrane-less organelles, IDPs might serve as perfect regulators and controllers of the formation of these organelles via the aforementioned phase separation [147]. Data reported in our work provide indirect support to this hypothesis by showing that many membrane-less sub-nuclear organelles are enriched in IDPs and by emphasizing that these disordered nuclear proteins are highly promiscuous binders.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/17/1/24/s1>.

**Acknowledgments:** We are thankful to Alexey Uversky for careful reading and editing this manuscript. This work was supported in part by the Discovery grant (298328) from the Natural Sciences and Engineering Research Council (NSERC) of Canada to Lukasz Kurgan and by the China Scholarship Council scholarship to Fanchi Meng.

**Author Contributions:** Fanchi Meng and Insung Na—collected datasets, performed computational analysis of sequences, conducted statistical analysis of the data, and helped with the manuscript preparation. Lukasz Kurgan—analyzed data and wrote the review. Vladimir N. Uversky—conceived the idea, collected and analyzed literature data and wrote the review.

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

## References

1. Goodsell, D.S. Miniseries: Illustrating the machinery of life: Eukaryotic cell panorama. *Biochem. Mol. Biol. Educ.* **2011**, *39*, 91–101. [[CrossRef](#)] [[PubMed](#)]
2. Tripathi, V.; Prasanth, K.V. Cell nucleus. In *eLS*; John Wiley & Sons Ltd.: Chichester, UK, 2011.
3. Hancock, R.; Jeon, K. *New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals*; Academic Press: Amsterdam, The Netherlands; Boston, MA, USA; Heidelberg, Germany; London, UK; New York, NY, USA; Oxford, UK; Paris, France; San Diego, CA, USA; San Francisco, CA, USA; Singapore; Sydney, Australia; Tokyo, Japan, 2014; p. 512.
4. Jevtic, P.; Edens, L.J.; Vukovic, L.D.; Levy, D.L. Sizing and shaping the nucleus: Mechanisms and significance. *Curr. Opin. Cell Biol.* **2014**, *28*, 16–27. [[CrossRef](#)] [[PubMed](#)]
5. Fakan, S. The functional architecture of the nucleus as analysed by ultrastructural cytochemistry. *Histochem. Cell Biol.* **2004**, *122*, 83–93. [[CrossRef](#)] [[PubMed](#)]
6. Handwerker, K.E.; Gall, J.G. Sub-nuclear organelles: New insights into form and function. *Trends Cell Biol.* **2006**, *16*, 19–26. [[CrossRef](#)] [[PubMed](#)]
7. Phair, R.D.; Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **2000**, *404*, 604–609. [[PubMed](#)]
8. Pederson, T. Protein mobility within the nucleus—What are the right moves? *Cell* **2001**, *104*, 635–638. [[CrossRef](#)]
9. Brangwynne, C.P.; Eckmann, C.R.; Courson, D.S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Julicher, F.; Hyman, A.A. Germline p granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **2009**, *324*, 1729–1732. [[CrossRef](#)] [[PubMed](#)]
10. Brangwynne, C.P.; Mitchison, T.J.; Hyman, A.A. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4334–4339. [[CrossRef](#)] [[PubMed](#)]
11. Li, P.; Banjade, S.; Cheng, H.C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J.V.; King, D.S.; Banani, S.F.; *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **2012**, *483*, 336–340. [[CrossRef](#)] [[PubMed](#)]
12. Aggarwal, S.; Snaidero, N.; Pahler, G.; Frey, S.; Sanchez, P.; Zweckstetter, M.; Janshoff, A.; Schneider, A.; Weil, M.T.; Schaap, I.A.; *et al.* Myelin membrane assembly is driven by a phase transition of myelin basic proteins into a cohesive protein meshwork. *PLoS Biol.* **2013**, *11*, e1001577. [[CrossRef](#)] [[PubMed](#)]
13. Feric, M.; Brangwynne, C.P. A nuclear F-actin scaffold stabilizes ribonucleoprotein droplets against gravity in large cells. *Nat. Cell Biol.* **2013**, *15*, 1253–1259. [[CrossRef](#)] [[PubMed](#)]

14. Wippich, F.; Bodenmiller, B.; Trajkovska, M.G.; Wanka, S.; Aebersold, R.; Pelkmans, L. Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. *Cell* **2013**, *152*, 791–805. [[CrossRef](#)] [[PubMed](#)]
15. Spector, D.L. Nuclear domains. *J. Cell Sci.* **2001**, *114*, 2891–2893. [[PubMed](#)]
16. Hernandez-Verdun, D.; Roussel, P.; Thiry, M.; Sirri, V.; Lafontaine, D.L. The nucleolus: Structure/function relationship in RNA metabolism. *Wiley Interdiscip. Rev. RNA* **2010**, *1*, 415–431. [[CrossRef](#)] [[PubMed](#)]
17. Boisvert, F.M.; van Koningsbruggen, S.; Navascues, J.; Lamond, A.I. The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 574–585. [[CrossRef](#)] [[PubMed](#)]
18. Li, B.; Carey, M.; Workman, J.L. The role of chromatin during transcription. *Cell* **2007**, *128*, 707–719. [[CrossRef](#)] [[PubMed](#)]
19. Kornberg, R.D. Chromatin structure: A repeating unit of histones and DNA. *Science* **1974**, *184*, 868–871. [[CrossRef](#)] [[PubMed](#)]
20. Luger, K.; Mader, A.W.; Richmond, R.K.; Sargent, D.F.; Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, *389*, 251–260. [[PubMed](#)]
21. Li, G.; Reinberg, D. Chromatin higher-order structures and gene regulation. *Curr. Opin. Genet. Dev.* **2011**, *21*, 175–186. [[CrossRef](#)] [[PubMed](#)]
22. Robinson, P.J.; Fairall, L.; Huynh, V.A.; Rhodes, D. EM measurements define the dimensions of the “30-nm” chromatin fiber: Evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6506–6511. [[CrossRef](#)] [[PubMed](#)]
23. Cremer, T.; Cremer, M.; Dietzel, S.; Muller, S.; Solovei, I.; Fakan, S. Chromosome territories—A functional nuclear landscape. *Curr. Opin. Cell Biol.* **2006**, *18*, 307–316. [[CrossRef](#)] [[PubMed](#)]
24. Cremer, T.; Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a003889. [[CrossRef](#)] [[PubMed](#)]
25. Spector, D.L. Macromolecular domains within the cell nucleus. *Annu. Rev. Cell Biol.* **1993**, *9*, 265–315. [[CrossRef](#)] [[PubMed](#)]
26. Ogg, S.C.; Lamond, A.I. Cajal bodies and coilin—Moving towards function. *J. Cell Biol.* **2002**, *159*, 17–21. [[CrossRef](#)] [[PubMed](#)]
27. Morris, G.E. The cajal body. *Biochim. Biophys. Acta* **2008**, *1783*, 2108–2115. [[CrossRef](#)] [[PubMed](#)]
28. Maul, G.G.; Negorev, D.; Bell, P.; Ishov, A.M. Review: Properties and assembly mechanisms of ND10, PML bodies, or PODs. *J. Struct. Biol.* **2000**, *129*, 278–287. [[CrossRef](#)] [[PubMed](#)]
29. Boisvert, F.M.; Hendzel, M.J.; Bazett-Jones, D.P. Promyelocytic leukemia (PML) nuclear bodies are protein structures that do not accumulate RNA. *J. Cell Biol.* **2000**, *148*, 283–292. [[CrossRef](#)] [[PubMed](#)]
30. Luciani, J.J.; Depetris, D.; Usson, Y.; Metzler-Guillemain, C.; Mignon-Ravix, C.; Mitchell, M.J.; Megarbane, A.; Sarda, P.; Sirma, H.; Moncla, A.; *et al.* PML nuclear bodies are highly organised DNA-protein structures with a function in heterochromatin remodelling at the G2 phase. *J. Cell Sci.* **2006**, *119*, 2518–2531. [[CrossRef](#)] [[PubMed](#)]
31. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Robers, K.; Walter, P. Intracellular compartments and protein sorting. In *Molecular Biology of the Cell*; Garland Science: New York, NY, USA, 2008; pp. 695–724.
32. Mor, A.; White, M.A.; Fontoura, B.M. Nuclear trafficking in health and disease. *Curr. Opin. Cell Biol.* **2014**, *28*, 28–35. [[CrossRef](#)] [[PubMed](#)]
33. Laba, J.K.; Steen, A.; Veenhoff, L.M. Traffic to the inner membrane of the nuclear envelope. *Curr. Opin. Cell Biol.* **2014**, *28*, 36–45. [[CrossRef](#)] [[PubMed](#)]
34. Amendola, M.; van Steensel, B. Mechanisms and dynamics of nuclear lamina-genome interactions. *Curr. Opin. Cell Biol.* **2014**, *28*, 61–68. [[CrossRef](#)] [[PubMed](#)]
35. Grossman, E.; Medalia, O.; Zwerger, M. Functional architecture of the nuclear pore complex. *Annu. Rev. Biophys.* **2012**, *41*, 557–584. [[CrossRef](#)] [[PubMed](#)]
36. Strambio-De-Castillia, C.; Niepel, M.; Rout, M.P. The nuclear pore complex: Bridging nuclear transport and gene regulation. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 490–501. [[CrossRef](#)] [[PubMed](#)]
37. Hoelz, A.; Debler, E.W.; Blobel, G. The structure of the nuclear pore complex. *Annu. Rev. Biochem.* **2011**, *80*, 613–643. [[CrossRef](#)] [[PubMed](#)]
38. Matera, A.G.; Frey, M.R.; Margelot, K.; Wolin, S.L. A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. *J. Cell Biol.* **1995**, *129*, 1181–1193. [[CrossRef](#)] [[PubMed](#)]

39. Huang, S.; Deerinck, T.J.; Ellisman, M.H.; Spector, D.L. The dynamic organization of the perinucleolar compartment in the cell nucleus. *J. Cell Biol.* **1997**, *137*, 965–974. [[CrossRef](#)] [[PubMed](#)]
40. Pollock, C.; Huang, S. The perinucleolar compartment. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000679. [[CrossRef](#)] [[PubMed](#)]
41. Decker, M.; Jaensch, S.; Pozniakovsky, A.; Zinke, A.; O'Connell, K.F.; Zachariae, W.; Myers, E.; Hyman, A.A. Limiting amounts of centrosome material set centrosome size in *C. elegans* embryos. *Curr. Biol.* **2011**, *21*, 1259–1267. [[CrossRef](#)] [[PubMed](#)]
42. Brangwynne, C.P. Phase transitions and size scaling of membrane-less organelles. *J. Cell Biol.* **2013**, *203*, 875–881. [[CrossRef](#)] [[PubMed](#)]
43. Dunker, A.K.; Garner, E.; Guillot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J.E. Protein disorder and the evolution of molecular recognition: Theory, predictions and observations. *Pac. Symp. Biocomput.* **1998**, 473–484.
44. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331. [[CrossRef](#)] [[PubMed](#)]
45. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415–427. [[CrossRef](#)]
46. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001**, *19*, 26–59. [[CrossRef](#)]
47. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533. [[CrossRef](#)]
48. Daughdrill, G.W.; Pielak, G.J.; Uversky, V.N.; Cortese, M.S.; Dunker, A.K. Natively disordered proteins. In *Handbook of Protein Folding*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH, Verlag GmbH & Co.: Weinheim, Germany, 2005; pp. 271–353.
49. Uversky, V.N.; Dunker, A.K. Understanding protein non-folding. *Biochim. Biophys. Acta* **2010**, *1804*, 1231–1264. [[CrossRef](#)] [[PubMed](#)]
50. Dunker, A.K.; Obradovic, Z.; Romero, P.; Garner, E.C.; Brown, C.J. Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* **2000**, *11*, 161–171. [[PubMed](#)]
51. Romero, P.; Obradovic, Z.; Kissinger, C.R.; Villafranca, J.E.; Garner, E.; Guillot, S.; Dunker, A.K. Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* **1998**, 437–448.
52. Ward, J.J.; Sodhi, J.S.; McGuffin, L.J.; Buxton, B.F.; Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004**, *337*, 635–645. [[CrossRef](#)] [[PubMed](#)]
53. Oldfield, C.J.; Cheng, Y.; Cortese, M.S.; Brown, C.J.; Uversky, V.N.; Dunker, A.K. Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **2005**, *44*, 1989–2000. [[CrossRef](#)] [[PubMed](#)]
54. Xue, B.; Dunker, A.K.; Uversky, V.N. Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **2012**, *30*, 137–149. [[CrossRef](#)] [[PubMed](#)]
55. Oates, M.E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M.J.; Xue, B.; Dosztanyi, Z.; Uversky, V.N.; Obradovic, Z.; Kurgan, L.; *et al.* D<sup>2</sup>P<sup>2</sup>: Database of disordered protein predictions. *Nucleic Acids Res.* **2013**, *41*, D508–D516. [[CrossRef](#)] [[PubMed](#)]
56. Peng, Z.; Yan, J.; Fan, X.; Mizianty, M.J.; Xue, B.; Wang, K.; Hu, G.; Uversky, V.N.; Kurgan, L. Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **2015**, *72*, 137–151. [[CrossRef](#)] [[PubMed](#)]
57. Iakoucheva, L.M.; Brown, C.J.; Lawson, J.D.; Obradovic, Z.; Dunker, A.K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573–584. [[CrossRef](#)]
58. Dunker, A.K.; Cortese, M.S.; Romero, P.; Iakoucheva, L.M.; Uversky, V.N. Flexible nets: The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **2005**, *272*, 5129–5148. [[CrossRef](#)] [[PubMed](#)]
59. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Showing your ID: Intrinsic disorder as an id for recognition, regulation and cell signaling. *J. Mol. Recognit.* **2005**, *18*, 343–384. [[CrossRef](#)] [[PubMed](#)]
60. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. [[CrossRef](#)] [[PubMed](#)]
61. Marasco, D.; Scognamiglio, P.L. Identification of inhibitors of biological interactions involving intrinsically disordered proteins. *Int. J. Mol. Sci.* **2015**, *16*, 7394–7412. [[CrossRef](#)] [[PubMed](#)]

62. Radivojac, P.; Iakoucheva, L.M.; Oldfield, C.J.; Obradovic, Z.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder and functional proteomics. *Biophys. J.* **2007**, *92*, 1439–1456. [[CrossRef](#)] [[PubMed](#)]
63. Vucetic, S.; Xie, H.; Iakoucheva, L.M.; Oldfield, C.J.; Dunker, A.K.; Obradovic, Z.; Uversky, V.N. Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *J. Proteome Res.* **2007**, *6*, 1899–1916. [[CrossRef](#)] [[PubMed](#)]
64. Xie, H.; Vucetic, S.; Iakoucheva, L.M.; Oldfield, C.J.; Dunker, A.K.; Obradovic, Z.; Uversky, V.N. Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J. Proteome Res.* **2007**, *6*, 1917–1932. [[CrossRef](#)] [[PubMed](#)]
65. Xie, H.; Vucetic, S.; Iakoucheva, L.M.; Oldfield, C.J.; Dunker, A.K.; Uversky, V.N.; Obradovic, Z. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* **2007**, *6*, 1882–1898. [[CrossRef](#)] [[PubMed](#)]
66. Frege, T.; Uversky, V.N. Intrinsically disordered proteins in the nucleus of human cells. *Biochem. Biophys. Rep.* **2015**, *1*, 33–51. [[CrossRef](#)]
67. Hetzer, M.W.; Walther, T.C.; Mattaj, I.W. Pushing the envelope: Structure, function, and dynamics of the nuclear periphery. *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 347–380. [[CrossRef](#)] [[PubMed](#)]
68. Xue, B.; Dunbrack, R.L.; Williams, R.W.; Dunker, A.K.; Uversky, V.N. Ponder-fit: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* **2010**, *1804*, 996–1010. [[CrossRef](#)] [[PubMed](#)]
69. Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C.J.; Dunker, A.K.; Obradovic, Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinform. Comput. Biol.* **2005**, *3*, 35–60. [[CrossRef](#)] [[PubMed](#)]
70. Walsh, I.; Martin, A.J.M.; di Domenico, T.; Tosatto, S.C.E. Espritz: Accurate and fast prediction of protein disorder. *Bioinformatics* **2012**, *28*, 503–509. [[CrossRef](#)] [[PubMed](#)]
71. Dosztányi, Z.; Csizmek, V.; Tompa, P.; Simon, I. Iupred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433–3434. [[CrossRef](#)] [[PubMed](#)]
72. Meszaros, B.; Simon, I.; Dosztanyi, Z. Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.* **2009**, *5*, e1000376. [[CrossRef](#)] [[PubMed](#)]
73. Dosztanyi, Z.; Meszaros, B.; Simon, I. Anchor: Web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **2009**, *25*, 2745–2746. [[CrossRef](#)] [[PubMed](#)]
74. Di Domenico, T.; Walsh, I.; Martin, A.J.; Tosatto, S.C. Mobidb: A comprehensive database of intrinsic protein disorder annotations. *Bioinformatics* **2012**, *28*, 2080–2081. [[CrossRef](#)] [[PubMed](#)]
75. Potenza, E.; Domenico, T.D.; Walsh, I.; Tosatto, S.C. Mobidb 2.0: An improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.* **2015**. [[CrossRef](#)] [[PubMed](#)]
76. Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049. [[CrossRef](#)] [[PubMed](#)]
77. Pejaver, V.; Hsu, W.L.; Xin, F.; Dunker, A.K.; Uversky, V.N.; Radivojac, P. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci.* **2014**, *23*, 1077–1093. [[CrossRef](#)] [[PubMed](#)]
78. Szklarczyk, D.; Franceschini, A.; Kuhn, M.; Simonovic, M.; Roth, A.; Minguetz, P.; Doerks, T.; Stark, M.; Muller, J.; Bork, P.; *et al.* The string database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **2011**, *39*, D561–D568. [[CrossRef](#)] [[PubMed](#)]
79. Romero, P.; Obradovic, Z.; Li, X.; Garner, E.C.; Brown, C.J.; Dunker, A.K. Sequence complexity of disordered protein. *Proteins* **2001**, *42*, 38–48. [[CrossRef](#)]
80. Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A.K.; Obradovic, Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinform.* **2006**, *7*, 208. [[CrossRef](#)] [[PubMed](#)]
81. Patil, A.; Nakamura, H. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett.* **2006**, *580*, 2041–2045. [[CrossRef](#)] [[PubMed](#)]
82. Haynes, C.; Oldfield, C.J.; Ji, F.; Klitgord, N.; Cusick, M.E.; Radivojac, P.; Uversky, V.N.; Vidal, M.; Iakoucheva, L.M. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput. Biol.* **2006**, *2*, e100. [[CrossRef](#)] [[PubMed](#)]

83. Ekman, D.; Light, S.; Bjorklund, A.K.; Elofsson, A. What properties characterize the hub proteins of the protein-protein interaction network of *saccharomyces cerevisiae*? *Genome Biol.* **2006**, *7*, R45. [[CrossRef](#)] [[PubMed](#)]
84. Dosztanyi, Z.; Chen, J.; Dunker, A.K.; Simon, I.; Tompa, P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.* **2006**, *5*, 2985–2995. [[CrossRef](#)] [[PubMed](#)]
85. Singh, G.P.; Ganapathi, M.; Sandhu, K.S.; Dash, D. Intrinsic unstructuredness and abundance of pest motifs in eukaryotic proteomes. *Proteins* **2006**, *62*, 309–315. [[CrossRef](#)] [[PubMed](#)]
86. Willadsen, K.; Mohamad, N.; Bodén, M. NSort/DB: An intra-nuclear compartment protein database. *Genom. Proteom. Bioinform.* **2012**, *10*, 226–229. [[CrossRef](#)] [[PubMed](#)]
87. Dellaire, G.; Farrall, R.; Bickmore, W.A. The nuclear protein database (NPD): Sub-nuclear localisation and functional annotation of the nuclear proteome. *Nucleic Acids Res.* **2003**, *31*, 328–330. [[CrossRef](#)] [[PubMed](#)]
88. Leung, A.K.L.; Trinkle-Mulcahy, L.; Lam, Y.W.; Andersen, J.S.; Mann, M.; Lamond, A.I. NOPdb: Nucleolar proteome database. *Nucleic Acids Res.* **2006**, *34*, D218–D220. [[CrossRef](#)] [[PubMed](#)]
89. Mika, S.; Rost, B. Nmpdb: Database of nuclear matrix proteins. *Nucleic Acids Res.* **2005**, *33*, D160–D163. [[CrossRef](#)] [[PubMed](#)]
90. Consortium, T.U. Activities at the universal protein resource (UniProt). *Nucleic Acids Res.* **2014**, *42*, D191–D198. [[CrossRef](#)] [[PubMed](#)]
91. Keshava-Prasad, T.S.; Goel, R.; Kandasamy, K.; Keerthikumar, S.; Kumar, S.; Mathivanan, S.; Telikicherla, D.; Raju, R.; Shafreen, B.; Venugopal, A.; *et al.* Human protein reference database—2009 update. *Nucleic Acids Res.* **2009**, *37*, D767–D772. [[CrossRef](#)] [[PubMed](#)]
92. Bauer, D.C.; Willadsen, K.; Buske, F.A.; Lê Cao, K.-A.; Bailey, T.L.; Dellaire, G.; Bodén, M. Sorting the nuclear proteome. *Bioinformatics* **2011**, *27*, i7–i14. [[CrossRef](#)] [[PubMed](#)]
93. Calderone, A.; Castagnoli, L.; Cesareni, G. Mentha: A resource for browsing integrated protein-interaction networks. *Nat. Methods* **2013**, *10*, 690–691. [[CrossRef](#)] [[PubMed](#)]
94. Han, J.D.; Bertin, N.; Hao, T.; Goldberg, D.S.; Berriz, G.F.; Zhang, L.V.; Dupuy, D.; Walhout, A.J.; Cusick, M.E.; Roth, F.P.; *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **2004**, *430*, 88–93. [[CrossRef](#)] [[PubMed](#)]
95. Bertin, N.; Simonis, N.; Dupuy, D.; Cusick, M.E.; Han, J.D.; Fraser, H.B.; Roth, F.P.; Vidal, M. Confirmation of organized modularity in the yeast interactome. *PLoS Biol.* **2007**, *5*, e153. [[CrossRef](#)] [[PubMed](#)]
96. Yu, H.; Kim, P.M.; Sprecher, E.; Trifonov, V.; Gerstein, M. The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* **2007**, *3*, e59. [[CrossRef](#)] [[PubMed](#)]
97. Walsh, I.; Giollo, M.; di Domenico, T.; Ferrari, C.; Zimmermann, O.; Tosatto, S.C. Comprehensive large-scale assessment of intrinsic protein disorder. *Bioinformatics* **2015**, *31*, 201–208. [[CrossRef](#)] [[PubMed](#)]
98. Fan, X.; Kurgan, L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *J. Biomol. Struct. Dyn.* **2014**, *32*, 448–464. [[CrossRef](#)] [[PubMed](#)]
99. Peng, Z.; Kurgan, L. On the complementarity of the consensus-based disorder prediction. *Pac. Symp. Biocomput.* **2012**, 176–187.
100. Sickmeier, M.; Hamilton, J.A.; LeGall, T.; Vacic, V.; Cortese, M.S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V.N.; *et al.* Disprot: The database of disordered proteins. *Nucleic Acids Res.* **2007**, *35*, D786–D793. [[CrossRef](#)] [[PubMed](#)]
101. Howell, M.; Green, R.; Killeen, A.; Wedderburn, L.; Picascio, V.; Rabionet, A.; Peng, Z.L.; Larina, M.; Xue, B.; Kurgan, L.; *et al.* Not that rigid midglets and not so flexible giants: On the abundance and roles of intrinsic disorder in short and long proteins. *J. Biol. Syst.* **2012**, *20*, 471–511. [[CrossRef](#)]
102. Peng, Z.; Oldfield, C.J.; Xue, B.; Mizianty, M.J.; Dunker, A.K.; Kurgan, L.; Uversky, V.N. A creature with a hundred waggly tails: Intrinsically disordered proteins in the ribosome. *Cell. Mol. Life Sci.* **2014**, *71*, 1477–1504. [[CrossRef](#)] [[PubMed](#)]
103. Pentony, M.M.; Jones, D.T. Modularity of intrinsic disorder in the human proteome. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 212–221. [[CrossRef](#)] [[PubMed](#)]
104. Tompa, P.; Fuxreiter, M.; Oldfield, C.J.; Simon, I.; Dunker, A.K.; Uversky, V.N. Close encounters of the third kind: Disordered domains and the interactions of proteins. *BioEssays* **2009**, *31*, 328–335. [[CrossRef](#)] [[PubMed](#)]

105. Prilusky, J.; Felder, C.E.; Zeev-Ben-Mordehai, T.; Rydberg, E.H.; Man, O.; Beckmann, J.S.; Silman, I.; Sussman, J.L. Foldindex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **2005**, *21*, 3435–3438. [[CrossRef](#)] [[PubMed](#)]
106. Campen, A.; Williams, R.M.; Brown, C.J.; Meng, J.; Uversky, V.N.; Dunker, A.K. Top-idp-scale: A new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* **2008**, *15*, 956–963. [[CrossRef](#)] [[PubMed](#)]
107. Mohan, A.; Sullivan, W.J., Jr.; Radivojac, P.; Dunker, A.K.; Uversky, V.N. Intrinsic disorder in pathogenic and non-pathogenic microbes: Discovering and analyzing the unfoldomes of early-branching eukaryotes. *Mol. Biosyst.* **2008**, *4*, 328–340. [[CrossRef](#)] [[PubMed](#)]
108. Xue, B.; Oldfield, C.J.; Dunker, A.K.; Uversky, V.N. CDF it all: Consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. *FEBS Lett.* **2009**, *583*, 1469–1474. [[CrossRef](#)] [[PubMed](#)]
109. Huang, F.; Oldfield, C.J.; Xue, B.; Hsu, W.L.; Meng, J.; Liu, X.; Shen, L.; Romero, P.; Uversky, V.N.; Dunker, A. Improving protein order-disorder classification using charge-hydrophathy plots. *BMC Bioinform.* **2014**, *15*, S4. [[CrossRef](#)] [[PubMed](#)]
110. Huang, F.; Oldfield, C.; Meng, J.; Hsu, W.L.; Xue, B.; Uversky, V.N.; Romero, P.; Dunker, A.K. Subclassifying disordered proteins by the CH-CDF plot method. *Pac. Symp. Biocomput.* **2012**, 128–139.
111. Linding, R.; Jensen, L.J.; Diella, F.; Bork, P.; Gibson, T.J.; Russell, R.B. Protein disorder prediction: Implications for structural proteomics. *Structure* **2003**, *11*, 1453–1459. [[CrossRef](#)] [[PubMed](#)]
112. Yang, Z.R.; Thomson, R.; McNeil, P.; Esnouf, R.M. Ronn: The bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **2005**, *21*, 3369–3376. [[CrossRef](#)] [[PubMed](#)]
113. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Dunker, A.K. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* **2005**, *61*, 176–182. [[CrossRef](#)] [[PubMed](#)]
114. Linding, R.; Russell, R.B.; Neduva, V.; Gibson, T.J. Globplot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **2003**, *31*, 3701–3708. [[CrossRef](#)] [[PubMed](#)]
115. Peng, Z.L.; Kurgan, L. Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr. Protein Pept. Sci.* **2012**, *13*, 6–18. [[CrossRef](#)] [[PubMed](#)]
116. Ishida, T.; Kinoshita, K. Prdos: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **2007**, *35*, W460–W464. [[CrossRef](#)] [[PubMed](#)]
117. Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **2005**, *347*, 827–839. [[CrossRef](#)] [[PubMed](#)]
118. Peng, Z.; Sakai, Y.; Kurgan, L.; Sokolowski, B.; Uversky, V. Intrinsic disorder in the bk channel and its interactome. *PLoS ONE* **2014**, *9*, e94331. [[CrossRef](#)] [[PubMed](#)]
119. Stephens, M.A. Edf statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* **1974**, *69*, 730–737. [[CrossRef](#)]
120. Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1945**, *1*, 80–83. [[CrossRef](#)]
121. Boublik, M.; Bradbury, E.M.; Crane-Robinson, C. An investigation of the conformational changes in histones F1 and F2A1 by proton magnetic resonance spectroscopy. *Eur. J. Biochem.* **1970**, *14*, 486–497. [[CrossRef](#)] [[PubMed](#)]
122. Boublik, M.; Bradbury, E.M.; Crane-Robinson, C.; Johns, E.W. An investigation of the conformational changes of histone F2B by high resolution nuclear magnetic resonance. *Eur. J. Biochem.* **1970**, *17*, 151–159. [[CrossRef](#)] [[PubMed](#)]
123. D’Anna, J.A., Jr.; Isenberg, I. A complex of histones IIB2 and IV. *Biochemistry* **1973**, *12*, 1035–1043. [[CrossRef](#)] [[PubMed](#)]
124. D’Anna, J.A., Jr.; Isenberg, I. Conformational changes of histone LAK (F2A2). *Biochemistry* **1974**, *13*, 2093–2098. [[CrossRef](#)] [[PubMed](#)]
125. D’Anna, J.A., Jr.; Isenberg, I. Conformational changes of histone ARE(F3, III). *Biochemistry* **1974**, *13*, 4987–4992. [[CrossRef](#)] [[PubMed](#)]



126. Hansen, J.C. Conformational dynamics of the chromatin fiber in solution: Determinants, mechanisms, and functions. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 361–392. [[CrossRef](#)] [[PubMed](#)]
127. Hansen, J.C.; Lu, X.; Ross, E.D.; Woody, R.W. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.* **2006**, *281*, 1853–1856. [[CrossRef](#)] [[PubMed](#)]
128. Isenberg, I. Histones. *Annu. Rev. Biochem.* **1979**, *48*, 159–191. [[CrossRef](#)] [[PubMed](#)]
129. Kato, H.; Gruschus, J.; Ghirlando, R.; Tjandra, N.; Bai, Y. Characterization of the N-terminal tail domain of histone H3 in condensed nucleosome arrays by hydrogen exchange and NMR. *J. Am. Chem. Soc.* **2009**, *131*, 15104–15105. [[CrossRef](#)] [[PubMed](#)]
130. Li, H.J.; Isenberg, I.; Johnson, W.C., Jr. Absorption and circular dichroism studies on nucleohistone IV. *Biochemistry* **1971**, *10*, 2587–2593. [[PubMed](#)]
131. Li, H.J.; Wickett, R.; Craig, A.M.; Isenberg, I. Conformational changes in histone IV. *Biopolymers* **1972**, *11*, 375–397. [[CrossRef](#)] [[PubMed](#)]
132. Luger, K.; Richmond, T.J. The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* **1998**, *8*, 140–146. [[CrossRef](#)]
133. Munishkina, L.A.; Fink, A.L.; Uversky, V.N. Conformational prerequisites for formation of amyloid fibrils from histones. *J. Mol. Biol.* **2004**, *342*, 1305–1324. [[CrossRef](#)] [[PubMed](#)]
134. Peng, Z.; Mizianty, M.J.; Xue, B.; Kurgan, L.; Uversky, V.N. More than just tails: Intrinsic disorder in histone proteins. *Mol. Biosyst.* **2012**, *8*, 1886–1901. [[CrossRef](#)] [[PubMed](#)]
135. Wickett, R.R.; Li, H.J.; Isenberg, I. Salt effects on histone iv conformation. *Biochemistry* **1972**, *11*, 2952–2957. [[CrossRef](#)] [[PubMed](#)]
136. Zheng, C.; Hayes, J.J. Structures and interactions of the core histone tail domains. *Biopolymers* **2003**, *68*, 539–546. [[CrossRef](#)] [[PubMed](#)]
137. Placek, B.J.; Gloss, L.M. The N-terminal tails of the H2A-H2B histones affect dimer structure and stability. *Biochemistry* **2002**, *41*, 14960–14968. [[CrossRef](#)] [[PubMed](#)]
138. Potoyan, D.A.; Papoian, G.A. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.* **2011**, *133*, 7405–7415. [[CrossRef](#)] [[PubMed](#)]
139. Denning, D.P.; Patel, S.S.; Uversky, V.; Fink, A.L.; Rexach, M. Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2450–2455. [[CrossRef](#)] [[PubMed](#)]
140. Denning, D.P.; Uversky, V.; Patel, S.S.; Fink, A.L.; Rexach, M. The *Saccharomyces cerevisiae* nucleoporin NUP2P is a natively unfolded protein. *J. Biol. Chem.* **2002**, *277*, 33447–33455. [[CrossRef](#)] [[PubMed](#)]
141. Fuxreiter, M.; Toth-Petroczy, A.; Kraut, D.A.; Matouschek, A.T.; Lim, R.Y.; Xue, B.; Kurgan, L.; Uversky, V.N. Disordered proteinaceous machines. *Chem. Rev.* **2014**, *114*, 6806–6843. [[CrossRef](#)] [[PubMed](#)]
142. Yamada, J.; Phillips, J.L.; Patel, S.; Goldfien, G.; Calestagne-Morelli, A.; Huang, H.; Reza, R.; Acheson, J.; Krishnan, V.V.; Newsam, S.; *et al.* A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteom.* **2010**, *9*, 2205–2224. [[CrossRef](#)] [[PubMed](#)]
143. Korneta, I.; Bujnicki, J.M. Intrinsic disorder in the human spliceosomal proteome. *PLoS Comput. Biol.* **2012**, *8*, e1002641. [[CrossRef](#)] [[PubMed](#)]
144. Coelho-Ribeiro-Mde, L.; Espinosa, J.; Islam, S.; Martinez, O.; Thanki, J.J.; Mazariegos, S.; Nguyen, T.; Larina, M.; Xue, B.; Uversky, V.N. Malleable ribonucleoprotein machine: Protein intrinsic disorder in the *Saccharomyces cerevisiae* spliceosome. *PeerJ* **2013**, *1*, e2. [[CrossRef](#)] [[PubMed](#)]
145. Liu, J.; Perumal, N.B.; Oldfield, C.J.; Su, E.W.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder in transcription factors. *Biochemistry* **2006**, *45*, 6873–6888. [[CrossRef](#)] [[PubMed](#)]
146. Minezaki, Y.; Homma, K.; Kinjo, A.R.; Nishikawa, K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* **2006**, *359*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]

147. Uversky, V.N.; Kuznetsova, I.M.; Turoverov, K.K.; Zaslavsky, B. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS Lett.* **2015**, *589*, 15–22. [[CrossRef](#)] [[PubMed](#)]
148. Keating, C.D. Aqueous phase separation as a possible route to compartmentalization of biological molecules. *Acc. Chem. Res.* **2012**, *45*, 2114–2124. [[CrossRef](#)] [[PubMed](#)]



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).