



Open Access

INVITED ORIGINAL ARTICLE

Prostate Cancer

Unfoldomics of prostate cancer: on the abundance and roles of intrinsically disordered proteins in prostate cancer

Kevin S Landau¹, Insung Na¹, Ryan O Schenck¹, Vladimir N Uversky^{1,2,3}

Prostatic diseases such as prostate cancer and benign prostatic hyperplasia are highly prevalent among men. The number of studies focused on the abundance and roles of intrinsically disordered proteins in prostate cancer is rather limited. The goal of this study is to analyze the prevalence and degree of disorder in proteins that were previously associated with the prostate cancer pathogenesis and to compare these proteins to the entire human proteome. The analysis of these datasets provides means for drawing conclusions on the roles of disordered proteins in this common male disease. We also hope that the results of our analysis can potentially lead to future experimental studies of these proteins to find novel pathways associated with this disease.

Asian Journal of Andrology (2016) 18, 662–672; doi: 10.4103/1008-682X.184999; published online: 19 July 2016

Keywords: intrinsically disordered proteins; posttranslational modifications; prostate cancer; protein-protein interaction; proteomics; proteome; unfoldome

INTRODUCTION

Prostate cancer is the most prevalent form of cancer in males. It is estimated that in 2016, it will account for 180 890 (21%) of new cancer cases. It is also estimated that 26 120 (8%) of all cancer deaths which will occur in the United States male population in the year 2016 will be caused by prostate cancer.¹ Although there has been progress in the detection and treatment of prostate cancer, it is clear that more research is needed to make the diagnosis more reliable. The need for the constant improvement of the diagnostics tools is given, for example, by a recently established effect of obesity (measured in terms of the body mass index) on the predictive performance of the well-established and widely used biomarker, prostate-specific antigen (PSA), and PSA-related markers for prostate cancer.² These and similar facts alone should be sufficient to spark the scientific community's interest in studying more on possible causes of this very common male disease. Sometimes, a new approach needs to be taken to further the advancements in biomedical research.

One of such approaches is considering roles of intrinsically disordered proteins, or IDPs, which are proteins that lack a stable secondary and tertiary structure and have been shown to constitute a noticeable part of any proteome of interest.^{3–15} At the primary sequence level, IDPs are characterized by noticeable compositional biases, being noticeably depleted in order-promoting amino acids (Cys, Trp, Tyr, Phe, Ile, Leu, Val, and Asn) and being found to be enriched with disorder-promoting residues (Pro, Arg, Gly, Gln, Ser, Glu, Lys, and Ala).^{3,16–18} This peculiar physicochemical nature of these proteins, therefore, does not favor spontaneous folding into well-defined globular structures, and they, in turn, remain “floppy.”^{19,20} Among other

important features, this “floppiness” of IDPs provides them with an ability to be promiscuous binders and be involved in regulation and control of various signaling processes, being controlled themselves at multiple levels.^{4,5,21,22} Furthermore, intrinsically disordered protein regions (IDPRs) are often targeted for posttranslational modifications (PTMs)^{23–26} and are known to be involved in a myriad of biological processes.^{4,5,16,18,20,27–32} Proteins that are involved in multiple functions and processes are sometimes referred to as “moonlighting” proteins, and many of these moonlighting proteins are shown to be either completely disordered or possess long IDPRs.³³ Often, IDPs are specifically compartmentalized, e.g., IDPs and proteins with IDPRs in various nuclear membrane-less organelles.³⁴ Since cells in various cancers, not just of the prostate, grow and divide in an uncontrolled manner, it would be reasonable to assume that these moonlighting proteins with the multiple disordered regions have a significant role in many oncological processes. The validity of this hypothesis was demonstrated in earlier bioinformatics study where the majority of human cancer-associated proteins (HCAP) were shown to contain long IDPRs (i.e., regions possessing ≥ 30 consecutive disordered residues).³⁵ More generally, a D² (disorder in disorders) concept was introduced to emphasize that many proteins related to various human diseases such as cancer, neurodegeneration, diabetes, cardiovascular disease, and amyloidosis are intrinsically disordered.³⁶

This two-part study was dedicated to the analysis of the prevalence and functionality of disordered proteins in prostate cancer cells and to compare them to the abundance and functionality of IDPs in the human proteome as a whole. The present paper, which represents the first part

¹Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA; ²USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, Florida 33612, USA; ³Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia.

Correspondence: Dr. VN Uversky (vversky@health.usf.edu)

Received: 29 February 2016; Revised: 20 April 2016; Accepted: 21 June 2016

of this two-part study, reports the results of the global bioinformatics analysis of the disorderedness of proteins that were previously shown to be associated with the prostate cancer pathogenesis and compares these proteins with the entire human proteome. The second part continues this analysis, being focused on more detailed characterization of the prostate cancer-associated proteins from KEGG database.

MATERIALS AND METHODS

Acquisition of protein datasets

UniProt IDs of human proteins were obtained from the reviewed UniProt release 2015–09.³⁷ From those, reviewed IDs without uncommon amino acid codes (J, X, O, U, B, and Z) and proteins shorter than 30 amino acids were removed. The remaining proteins totaled to 20 120 and were considered as the human proteome set.

To obtain an experimentally validated dataset of proteins that were already shown to experimentally be involved in prostate cancer, we used results of a proteomic study that reported on the characterization of a total of 359 proteins, including 17 potential biomarkers of prostate cancer from prostatic cell lines, entered into the prostate cancer proteomics (PCP) database (<http://ef.inbi.ras.ru>).³⁸ This database is described as a multilevel informational database that was created using the results of a proteomic study of human prostate carcinoma and benign hyperplasia tissues, and of some human-cultured cell lines. Prostate cancer-related proteins entered to this database were first separated by 2D electrophoresis and subsequently identified by mass-spectrometry.³⁸ Proteins are extensively annotated using data from published articles and existing databases and contain direct Internet links to the information in the NCBI and UniProt databases.³⁸ Although this PCP website was originally designed in Russian, it can be toggled into English. The PCP database consists of 7 interrelated modules that contain proteomic data from different cell lines.³⁸ The seven modules that were used in our analysis are prostate proteins (hyperplasia, cancer), LNCAP (IPG-2DE), LNCAP (IEF-2DE), proteins of Rhabdomyosarcoma A-204, normal human myoblasts proteins, PC3, and BPH-1. The gene names were retrieved from the PCP database, and then were used as identifiers to find corresponding human proteins in UniProtKB. Some gene product names could not be mapped to UniProt IDs and were not included in our analysis. Once these UniProt IDs were found, obvious duplicates were removed, which decreased this dataset to 291 proteins. Since the PCP dataset had 7 modules that were interrelated, there were many repeated UniProt IDs including those corresponding to various isoforms. Therefore, a series of programs were written to obtain the unique representatives for subsequent analysis. After parsing the files, a dataset consisted of 251 proteins was assembled. This dataset included 196 reviewed UniProt proteins and 55 unreviewed TrEMBL proteins. The isoforms (designated with a dash and a number in their UniProt IDs) fell into the unreviewed category. Collectively, dataset containing 196 UniProt-reviewed proteins from the PCP database is referred to as the “Russian dataset” in our analysis and corresponding files.

An independent search of KEGG database resource (<http://www.genome.jp/kegg/>)³⁹ conducted in September 2015 provided us with a set of 48 proteins experimentally shown to be involved in prostate cancer. This dataset referred to as the “KEGG dataset” was used in a global analysis of abundance of intrinsic disorder in this set of prostate cancer-related proteins described in section of Results. A new search of KEGG database conducted on May 28, 2016, gave 40 additional proteins with experimentally validated connection to prostate cancer. The resulting set of 88 proteins constituted the “extended KEGG dataset”, and proteins from these datasets were subjected to a more

focused and detailed analysis of their predisposition for intrinsic disorder in the companion paper (**Supplementary Table 1**).

Proteins from Russian and KEGG datasets that were not found in the whole human proteome dataset were filtered out. For comparison purpose, a modified human proteome dataset was generated from which proteins included into Russian and KEGG datasets were excluded. Since transmembrane domains of membrane proteins are typically enriched in hydrophobic, order-promoting residues, Russian, KEGG, and human proteome datasets were further divided into two categories: membrane and nonmembrane. Membrane and nonmembrane proteins were classified using the QuickGO gene association file (gene_association.goa_human.gz, updated on September 14, 2015), and 41 Cellular Component Gene Ontology (GO) terms with ‘integral’ and ‘membrane’ on their names.^{40,41}

Functional annotation using gene ontology (GO) terms

The GO terms for the human proteome were downloaded on September 14, 2015, from QuickGO (<https://www.ebi.ac.uk/QuickGO/>). The resulting file shows all of the biological processes (P), molecular functions (F), and cellular component (C) for all of the human proteins in UniProt format. All of the GO terms that are specific for annotating the cellular component were downloaded from the Gene Ontology Consortium (<http://www.geneontology.org>).

Evaluation of intrinsic disorder propensity and disorder-based functionality

Intrinsic disorder propensities of human proteins were evaluated using PONDR-FIT, PONDR[®] VLXT, and PONDR[®] VSL2 algorithms.^{42–46} For each protein, after obtaining an average disorder score by each predictor, all three predictor-specific average scores were averaged again to generate a per-protein intrinsic disorder score. The 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.^{47–49} The dataset average disorder scores (AVG scores) were then calculated for human membrane and nonmembrane proteins based on the corresponding per-protein scores, and those AVG scores were used to prepare contingency tables.

The UniProt IDs in the set of human proteins were then used to match the disorder scores to the UniProt IDs of proteins in Russian and KEGG datasets to generate disorder scores for all of the proteins in our datasets. Since the human proteome disorder scores were only retrieved for the SwissProt reviewed human proteins, only the disorder propensities of the reviewed proteins in the Russian dataset were analyzed.

Contingency table constructions, statistical test, and common GO term search

Fisher’s exact test was used to find significance of the frequency (number of proteins) that were observed in the prostate cancer protein datasets (Russian and KEGG) to have an average disorder score calculated based on scores all three disorder predictors (PONDR-FIT, PONDR[®] VSL2, and PONDR[®] VLXT) greater than or equal to the AVG score of the entire human proteome. These data were used to construct the 2 × 2 contingency tables (**Supplementary Table 1** represents an example of such a contingency table), so the subsequent statistical analyses can be performed.⁵⁰ The null (H_0) and research hypotheses (H_1) shown below for both membrane and nonmembrane subsets of all datasets were tested using $\alpha = 0.01$ as the statistical significance level.

$$H_0: a/(a + b) \leq c/(c + d)$$

$$H_1: a/(a + b) > c/(c + d)$$

IBM SPSS 22 software (IBM Corporation, USA) was used to perform the Fisher's exact test whereas the R statistical package was used to create figures to present the data.

Similar approach was also used to find GO terms significantly associated with intrinsically disordered proteins in all datasets. Here, after categorization of the protein sets in contingency tables, Fisher's exact test was performed to find GO terms significantly associated with proteins possessing per-protein intrinsic disorder score greater than AVG score in Russian or KEGG datasets separately when those were compared with the whole human proteome. In these studies, membrane and nonmembrane proteins were treated separately. Fisher's exact test *P* values were retained through Python 3 programming, and its module, *scipy*.

Next, significant GO terms were analyzed using AmiGo2.⁵¹ First, GO terms with *P* values smaller than 0.05 or 0.1 were obtained for KEGG or Russian dataset versus whole human proteome, separately. Those GO terms were uploaded to visualization server, and the resulting GO term maps were downloaded in the SVG format. These maps included significant GO terms specific for the protein in KEGG or Russian dataset and were used to find the intersection or common GO terms present in both datasets. Those intersection GO terms were analyzed using REVIGO.⁵²

Evaluation of intrinsic disorder propensity and disorder-based functionality

For prostate cancer-related proteins from the KEGG dataset, we analyzed the per-residue disorder propensities by PONDR-FIT, PONDR[®] VLXT, PONDR[®] VSL2 algorithms,^{42–46} and by the PONDR[®] VL3 predictor that possesses high accuracy in finding long IDPRs.⁵³ We also used a consensus approach MobiDB,^{54,55} applied a binary disorder classifier charge-hydrophathy plot (CH-plot) that evaluates the predisposition of a given protein to be ordered or disordered as a whole,^{18,56} predicted potential disorder-based binding sites using the ANCHOR algorithm,^{57,58} looked at the functional disorder using D²P² database,⁵⁹ and analyzed interactivity of these proteins by STRING.⁶⁰

The MobiDB database (<http://mobidb.bio.unipd.it/>),^{54,55} that generates consensus disorder scores by aggregating the output from ten predictors, such as two versions of IUPred (IUPred-long and IUPred-short),⁶¹ three versions of ESpritz (ESpritz-DisProt, ESpritz-NMR, and ESpritz-XRay),⁶² two versions of DisEMBL (DisEMBL-465 and DisEMBL-HL),⁶³ JRONN,⁶⁴ PONDR[®] VSL2B,^{46,65} and GlobPlot.⁶⁶

A CH-plot represents an input protein as a point within the 2D graph where the mean Kate-Doolittle hydrophobicity and the mean absolute net charge are used as the X- and Y-coordinates, respectively. In the corresponding CH-plot, fully structured proteins and fully disordered proteins can be separated by a boundary line. All proteins located above this boundary line are highly likely to be extended whereas proteins located below this line are likely to be compact.^{18,56}

In addition to CH-plot, another binary disorder predictor, cumulative distribution function (CDF) analysis was used.⁵⁶ It summarizes the per-residue disorder predictions by plotting PONDR scores against their cumulative frequency, which allows ordered and disordered proteins to be distinguished on the basis of the distribution of prediction scores.⁵⁶ At any given point on the CDF curve, the ordinate gives the proportion of residues with a PONDR score less than or equal to the abscissa. The optimal boundary that provided the most accurate order-disorder classification was shown to represent seven points located in the 12th through 18th bins.⁵⁶ Thus, in the CDF analysis, order-disorder classification is based on whether a CDF curve of a

given protein is above (ordered) or below (disordered) a majority of boundary points.⁵⁶

Disorder evaluations together with important disorder-related functional information were retrieved from the D²P² database (<http://d2p2.pro/>),⁵⁹ which is a database of predicted disorder for a large library of proteins from completely sequenced genomes.⁵⁹ D²P² database uses outputs of IUPred,⁶¹ PONDR[®] VLXT,⁴² PrDOS,⁶⁷ PONDR[®] VSL2B,^{46,65} PV2,⁵⁹ and ESpritz.⁶² 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 Search Tool for the Retrieval of Interacting Genes; STRING, <http://string-db.org/>, which based on predicted and experimentally validated information on the interaction partners of a protein of interest generates a network of predicted associations.⁶⁰ In the corresponding network, the nodes correspond to proteins whereas the edges show predicted or known functional associations. Seven types of evidence are used to build the corresponding network where they are indicated by the differently colored lines: 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; and a black line - co-expression evidence.⁶⁰ In our analysis, the most stringent criteria were used for selection of interacting proteins by choosing the highest cutoff of 0.9 as the minimal required confidence level.

Finally, potential disorder-based protein-binding sites of prostate cancer-related proteins from the KEGG dataset were identified by the ANCHOR algorithm.^{57,58} This algorithm utilizes the pair-wise energy estimation approach originally used by IUPred.^{61,68} 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.^{57,58}

Evaluation of interactability of prostate cancer-related proteins in the KEGG dataset

Interactability of human proteins related to prostate cancer from the KEGG dataset was further evaluated by the APID (Agile Protein Interactomes DataServer) web server (<http://apid.dep.usal.es>).⁶⁹ APID has information on 90 379 distinct proteins from more than 400 organisms (including *Homo sapiens*) and on the 678 441 singular protein-protein interactions. For each protein-protein interaction (PPI), the server provides currently reported information about its experimental validation. For each protein, APID unifies PPIs found in five major primary databases of molecular interactions, such as BioGRID,⁷⁰ Database of Interacting Proteins (DIP),⁷¹ Human Protein Reference Database (HPRD),⁷² IntAct,⁷³ and the Molecular Interaction (MINT) database,⁷⁴ as well as from the BioPlex (biophysical interactions of ORFeome-based complexes)⁷⁵ and from the protein databank (PDB) entries of protein complexes.⁷⁶ This server provides a simple way to evaluate the interactability of individual proteins in a given dataset and also allows researchers to create a specific protein-protein interaction network in which proteins from the query dataset are engaged. Of 88 prostate cancer-related proteins in the KEGG dataset, APID was able to find protein-protein-related information on 87 proteins. No such information was available for the 3-oxo-5- α -steroid 4-dehydrogenase 2 (SRD5A2, UniProt ID: P31213).

RESULTS

Abundance of intrinsic disorder in Russian and KEGG datasets of prostate cancer-related proteins

There were 20 120 proteins in the whole human proteome dataset used. This dataset was separated into membrane proteins and nonmembrane proteins:

- Membrane proteins: 5193
- Nonmembrane proteins: 14 927.

The Russian dataset included 196 unique proteins that were mined from the PCP database. This dataset was then separated into membrane proteins and nonmembrane proteins:

- Membrane proteins: 6
- Nonmembrane proteins: 190.

The KEGG dataset contained 48 proteins, 4 of which were classified as membrane and remaining 44 as nonmembrane proteins.

Peculiarities of distribution of disorder within the members of these six datasets are shown in **Figure 1** in a form of DS_{FIT} vs DS_{VSL2} plots where DS_{FIT} and DS_{VSL2} correspond to the mean disorder scores calculated for query proteins using PONDR-FIT and PONDR® VSL2 algorithms, respectively.

To avoid redundancy in the subsequent statistical analysis of GO terms, the membrane and nonmembrane subsets of the whole human proteome datasets were adjusted to exclude prostate cancer-associated proteins. **Supplementary Tables 2–5** show the resulting 2×2 contingency tables and the corresponding results are further shown in **Figure 2** where they are depicted as a bar-graph.

We recognize that the values of some cells in the 2×2 contingency tables (**Supplementary Materials**) are rather low, indicating that it is too risky to draw any solid statistical conclusions when there are not enough samples, especially when comparison is done between very small and very large samples. Furthermore, disorder propensities of proteins analyzed in this study were evaluated using a set of standard disorder predictors that are characterized by the accuracy of 80%–85%. This indicates that the confidence of conclusions outlined below (especially for the cases having very limited samples) is further influenced by the limited accuracy of predictors. Therefore, data presented below should be taken only as indication of potential tendencies in the disorder predisposition and not as the final statistically significant conclusions.

The average disorder scores between all three predictors (PONDR® VSL2, PONDR® VLXT, and PONDR-FIT) were calculated for nonadjusted subsets of membrane and nonmembrane proteins in whole human proteome. These average disorder scores of 0.301 and 0.437 for the membrane and nonmembrane dataset, respectively, were then used to perform the statistical analysis of the proportion of the disordered proteins in the KEGG and Russian datasets compared to the proportion of the disordered proteins in the entire human proteome.

The one-sided Fisher's exact test statistic value for the membrane proteins in Russian dataset is 0.363, and therefore, H_0 cannot be rejected at the significance level of $\alpha = 0.01$. This means that in the membrane subset of Russian dataset, the proportion of proteins that have per-protein disorder scores \geq to the AVG score is not significantly different from human membrane proteins that have per-protein disorder scores \geq to the AVG score.

On the other hand, the one-sided Fisher's exact test statistic value for the nonmembrane proteins from the Russian dataset is <0.001 . Therefore, we reject H_0 at the significance level of $\alpha = 0.01$. From these data, we conclude that the fraction of the nonmembrane proteins that have per-protein disorder scores exceeding the AVG score is significantly higher in the Russian dataset of prostate cancer-associated proteins than

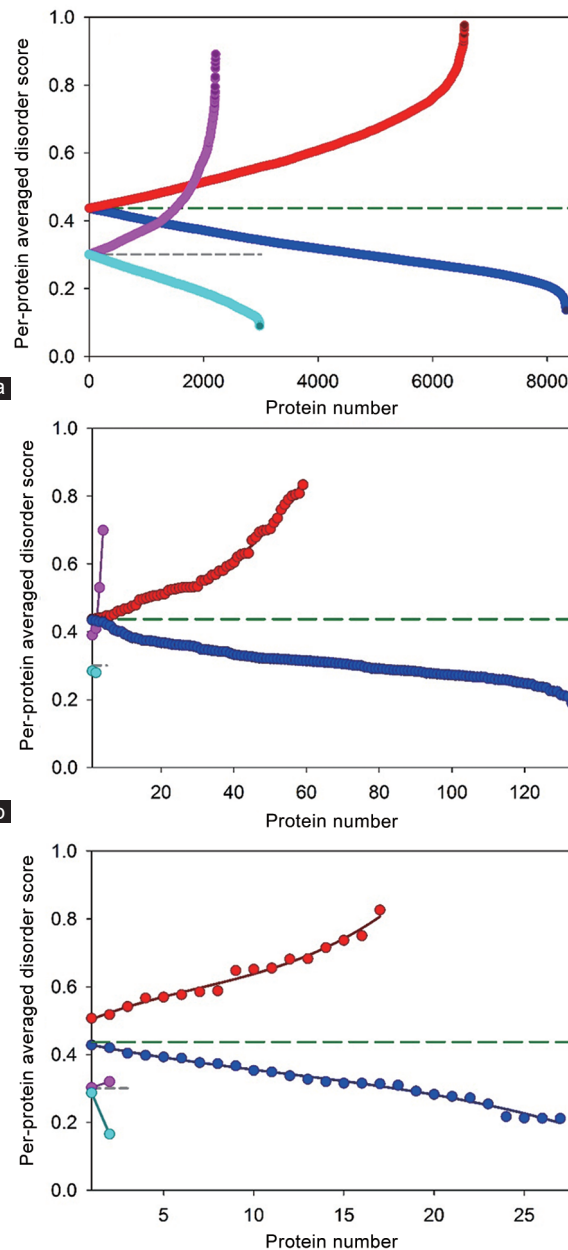


Figure 1: Abundance of intrinsic disorder in six datasets analyzed in this study - (a) human membrane (red and blue circles) and nonmembrane proteins (pink and cyan circles); (b) human membrane (red and blue circles) and nonmembrane prostate cancer-related proteins (pink and cyan circles) from Russian dataset; (c) human membrane (red and blue circles) and nonmembrane prostate cancer-related proteins (pink and cyan circles) from KEGG-dataset. Data for all these sets are shown in a form of proteins with per-protein disorder score above (red and pink circles) or below AVG (blue and cyan circles). AVGs for nonmembrane and membrane proteins are shown as medium dashed dark green and short dashed dark gray lines.

in the human proteome. This indicates that proteins with high intrinsic disorder levels are found significantly more often in the nonmembrane Russian dataset than in the whole human proteome, suggesting that higher abundance of IDPs/IDPRs can be related to cancer development. This observation is in agreement with the results of earlier studies on the high abundance of intrinsic disorder in cancer-related proteins.³⁵

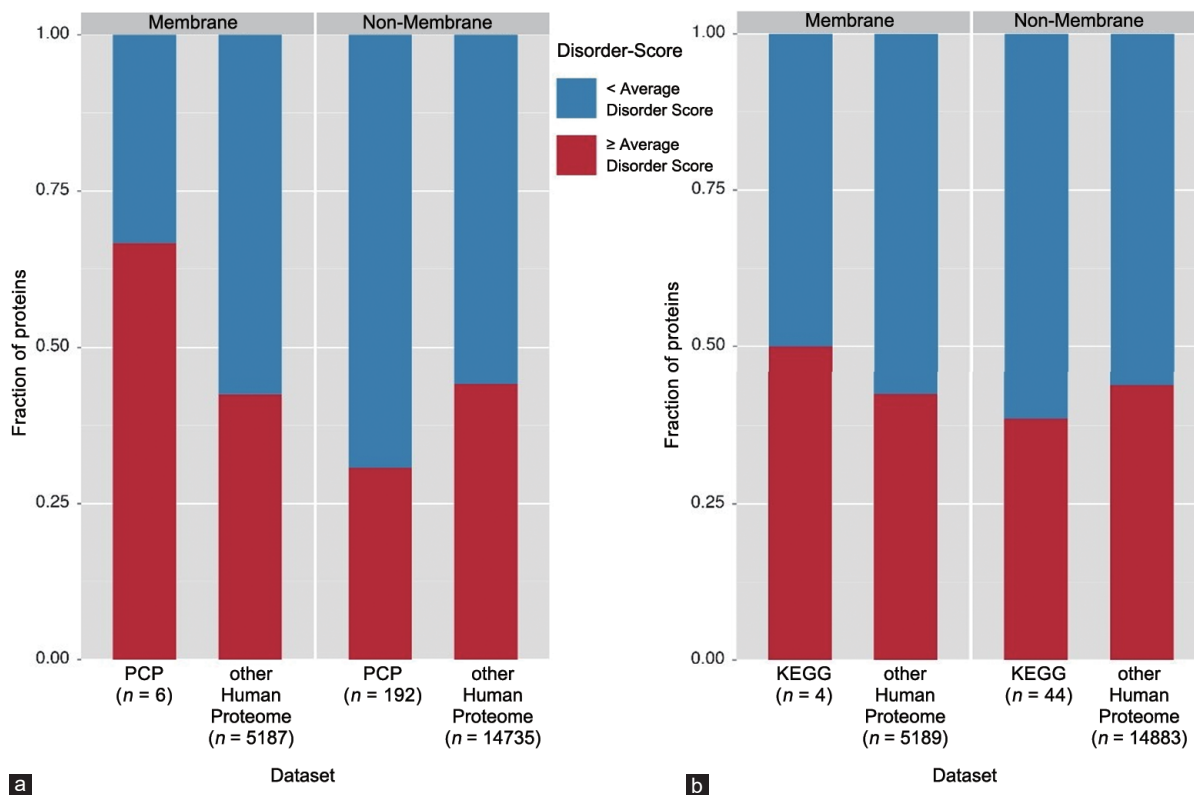


Figure 2: Bar graphs presenting data from **Supplementary Tables 2–5**. These graphs represent analyzed datasets that were split up into the membrane and nonmembrane groups, in which each bar represents either data for the human proteome, or the Russian (PCP) dataset (a), or KEGG dataset (b) within their respective group. The red section of each bar represents the fraction of proteins that have per-protein average disorder score \geq AVG and the blue sections represent the fraction of proteins that have per-protein average disorder score $<$ AVG.

For the KEGG membrane proteins versus dataset of human membrane proteins, we cannot reject the H_0 hypothesis at $\alpha = 0.05$. Therefore, there is no statistically significant difference between the proportions of KEGG membrane proteins with the per-protein disorder scores \geq AVG and human membrane proteins with the per-protein disorder scores \geq AVG. Similarly, for the KEGG nonmembrane proteins versus human nonmembrane proteins, we also cannot reject the H_0 hypothesis at $\alpha = 0.05$. Although this analysis revealed that the intrinsic disorder propensities of the prostate cancer-related proteins in the KEGG dataset are not significantly different from the disorder predispositions of proteins in human proteome, we conducted more detailed disorder-oriented analysis of proteins from the extended KEGG dataset to illustrate peculiarities of disorder distribution in these proteins and to see how IDPRs can be related to function and pathology (see Supplementary Materials).

Of the 17 potential biomarkers of prostate cancer outlined in Shishkin *et al.* paper,³⁸ only 12 were included in the reviewed protein dataset analyzed in our study. All of these 12 potential biomarkers are in the nonmembrane subset members of which are characterized by the per-protein disorder scores below the AVG disorder score. Furthermore, the well-known prostate cancer biomarker PSA was also present in our set of nonmembrane prostate cancer-related proteins characterized by the disorder scores below the AVG disorder score of human nonmembrane proteins. These observations suggest that potential biomarkers of prostate cancer are characterized by lower disorder levels than an average human protein.

Finding GO terms significantly associated with intrinsically disordered prostate cancer-related proteins in the KEGG and Russian datasets

GO terms are based on three structured ontologies that are designed for consistent functional descriptions of proteins in a species-independent manner. These terms show the relations of the query proteins to the biological processes they are involved in, the molecular functions they conduct, and the cellular components where they can be found at. Obviously, because GO terms are specifically designed as general terms of functional classification of proteins, these terms are applicable to any annotated protein (not only to proteins found in various pathologies but also to the normally functioning proteins). However, GO terms can also be used to find a correlation between protein intrinsic disorder and functionality. In fact, it has been shown in several previous studies that some GO terms are preferentially found to be associated with IDPRs whereas other GO terms would be more suitable for characterization of ordered proteins. Our analysis was conducted to show that many of the GO terms ascribed to the IDPRs associated with prostate cancer describe disorder-related functions. This observation reemphasizes the importance of intrinsic disorder for these proteins.

GO term analysis using AmiGo2⁵¹ did not produce noticeable number of disorder-associated GO terms when the P value threshold of less than 0.05 was used. Since we wanted to see which functions can be assigned to the intrinsically disordered proteins related to prostate cancer, we decided to use a looser P value threshold of less than 0.1 to find disorder associated GO terms (**Supplementary Table 6**). One should keep in mind that although the use of the looser P value ($P < 0.1$)

does not provide statistically significant data, the corresponding analysis generates a notable statistical trend that can be used to look for weakly significant correlations between intrinsic disorder and function. Among those GO terms with tentatively significant correlation to intrinsic disorder, four process-oriented GO terms – GO: 0006915 (apoptotic process), GO: 0008219 (cell death), GO: 0012501 (programmed cell death), and GO: 0016265 (death) – were related with programmed cell death through REVIGO⁵² (**Supplementary Figure 1a**). Moreover, those four chosen GO terms showed a simple map through AmiGo2 visualize tool (**Supplementary Figure 1b**).

Bioinformatics analysis of the prostate cancer-related proteins in the KEGG dataset

To gain information on the disorder status of prostate cancer-related proteins in the KEGG dataset and on the potential functional roles of their predicted IDPRs, we looked at them using a set of disorder predictors of PONDR family (PONDR-FIT, PONDR[®] VSL2, PONDR[®] VL3, and PONDR[®] VLXT), a binary disorder predictor CH-plot, a consensus disorder evaluating internet tools MobiDB and D²P², a platform STRING for finding potential interaction partners of a protein of interest, and a tool for predicting the disorder-based binding sites (ANCHOR). Some of these results are summarized in **Supplementary Table 7** and **Figure 3**.

Figure 3 illustrates that disorder predictions generated for these KEGG proteins by PONDR-FIT, PONDR[®] VSL2, and MobiDB in a form of per-protein disorder scores (DSs) or the percentage of disordered residues (per-protein content of disordered residues, CDRs, for PONDR-FIT and MobiDB) are generally agreed. Furthermore, from data shown in **Supplementary Table 7**, it is clearly seen that many proteins in this dataset are predicted to be moderately or highly disordered using the classification of proteins as highly ordered, moderately disordered, or highly disordered, if their CDR <10%, 10% ≤ CDR <30%, and CDR ≥30%, respectively.⁷⁷ In fact, according to these criteria, the

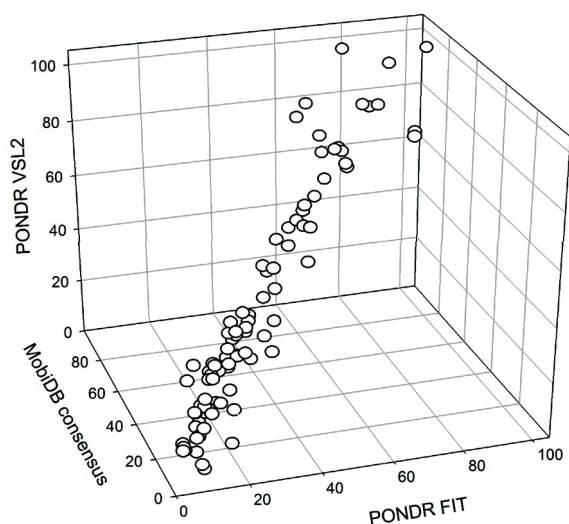


Figure 3: 3D representation of the results of evaluation of disorder levels in human prostate cancer-related proteins from the extended KEGG dataset. Here, the percentages of residues in these proteins predicted to be disordered by PONDR-FIT and PONDR[®] VSL2 are compared with the predicted percentages of disordered residues predicted by the MobiDB platform that aggregates the output from ten disorder predictors. The overall goal of this plot is to show the overall agreement between the outputs of different disorder predictors used in this study.

KEGG dataset includes 34 (38.6%), 38 (43.2%), and 16 (18.2%) highly disordered, moderately disordered, and highly ordered proteins, respectively. This suggests that almost 82% of prostate cancer-related proteins in this dataset are very noticeably disordered.

Supplementary Table 7 shows that although the disorder propensities of these proteins are spread over a wide range (e.g., by PONDR-FIT, from 3% of disordered residues in PI3K- α to 100% of such residues in BAD), the vast majority of them possess sizable IDPRs containing at least 10 consecutive residues predicted to be disordered, with many of these proteins having several such regions. In fact, only eleven proteins from the KEGG dataset were shown not to have such regions, and the remaining 77 proteins possessed 293 IDPRs (i.e., on average, each of these 77 proteins is expected to contain 3.81 such regions).

Often, IDPRs contain local regions with a strong tendency to become ordered at interaction with specific binding partners. Therefore, these regions might undergo coupled folding and binding, as shown for many of them by the NMR studies.^{78–83} Furthermore, such local short segments of order located within long disordered regions were shown to often coincide with the potential-binding sites.⁸⁴ Therefore, a number of computational tools for finding such molecular recognition features (MoRFs) were developed (e.g., a tool for predicting short binding regions with high α -helix-forming propensity, α -MoRFs^{85,86} or the more general ANCHOR algorithm for finding potential disorder-based binding sites,^{57,58} which are termed below AIBS for ANCHOR-identified binding sites). Curiously, in earlier studies, a systematic application of such computational tools indicated that α -MoRFs are likely to play important roles in protein-protein interactions involved in signaling events.⁸⁵ Our analysis revealed that the majority of proteins in the KEGG dataset (71 of 88) are predicted to have at least one AIBS, and that many of these proteins are expected to have multiple such binding regions each (**Supplementary Table 7**). Furthermore, AIBSs are found in almost each KEGG proteins that have at least one disordered region, and many of these proteins are shown to contain multiple disorder-based binding sites, with 34 and 30 AIBSs being found in CREB-binding protein (Q92793) and histone acetyltransferase p300 (Q09472), respectively. In fact, we found 472 AIBSs in 71 human proteins associated with prostate cancer, suggesting that on average, each of these protein contains >6.6 disorder-based binding sites. **Supplementary Table 7** also shows that the length of AIBSs is ranging from 6 to 150 residues, and the overall content of residues involved in the disorder-based interactions ranges from 0% to 70.8%.

The presence of more than one AIBS in a protein suggests that many prostate cancer-related proteins in the KEGG dataset commonly utilize disorder for their interactions with binding partners, and that these proteins are involved either in the polyvalent interactions using multiple binding sites to interact with one binding partner or in scaffolding-like interactions using multiple binding sites to interact with multiple binding partners. The wide spread of lengths of identified AIBSs also suggests the presence of multiple disorder-based binding mechanisms (ranging from local folding-on-binding of short regions to wrapping around binding mode to global binding-induced folding of large regions).

We established that many of the proteins in the KEGG dataset are predicted to have noticeable amounts of intrinsic disorder. Furthermore, this analysis revealed that 14 of these proteins (P05019, P14625, P36402, P46527, Q00987, Q92934, Q99801, Q9UJU2, P38936, P07900, Q92569, P08238, Q9Y6K9, and Q02930) are expected to

be disordered as a whole according to the CH-plot, i.e., are located above the boundary line separating compact and extended disordered proteins, and 16 more of these proteins, being located below this boundary, are found in its proximity (P04085, P04637, P25963, P60484, Q09472, Q92793, Q9NQB0, P62993, Q9Y243, P01127, Q9GZP0, P27986, O00459, Q9HCS4, Q96BA8, and P18848). Analogous analysis using another binary predictor, cumulative distribution function (CDF) plot (where proteins expected to be disordered or ordered as a whole are found based on the position of their CDF curve relative to the boundary separating mostly ordered and disordered proteins), revealed that 39 are expected to be mostly disordered (O00716, P04085, P04637, P05019, P06400, P10275, P10415, P16220, P25963, P36402, P46527, P49841, Q00987, Q01094, Q07889, Q09472, Q12778, Q14209, Q92793, Q92934, Q99801, Q9NQB0, Q8WYR1, P38936, O43889, Q9UJU2, P01127, O00459, P15056, Q07890, Q04206, Q9Y6K9, Q9HCS4, Q02930, Q8TEY5, Q68CJ9, Q70SY1, Q96BA8, and P18848) since the majority of their CDF curves are located below the boundary, whereas the CDF curves of 11 additional proteins (O14920, P01308, P10398, P14625, P24864, P36507, Q02750, P07900, O15530, P27986, and Q92569) follows boundary almost exactly, suggesting that these proteins are definitely not ordered as whole, as their overall disorder status is “undecided”. It was pointed out that combined analysis of protein disorder status using CH-plot and CDF analysis simultaneously can provide additional important information on the classification of protein disorder. This combined approach is known as the CH-CDF analysis,^{87–89} and it is based on the presence of a principle difference between the sensitivity of the CH-plot and the CDF analysis to different types of disorder. Here, the CH-plot can discriminate proteins with substantial amount of extended disorder (random coils and pre-molten globules) from proteins with compact conformations (molten globule-like and rigid well-structured proteins) whereas the CDF analysis may discriminate all disordered conformations, including molten globules and mixed proteins containing both disordered and ordered regions, from rigid well-folded proteins. Therefore, the CH-CDF analysis represents a computational tool to discriminate proteins with extended disorder from potential molten globules and mixed proteins containing comparable amounts of ordered and disordered regions.^{87–89} Therefore, based on the combination of outputs of their CH-plot and CDF analyses, proteins can be classified as follows: proteins predicted to be disordered by CH-plots, but ordered by CDF; ordered proteins (i.e., proteins predicted as ordered by both tools); putative molten globules or mixed proteins (i.e., proteins predicted to be disordered by CDF, but compact by CH-plot); and proteins with extended disorder (i.e., proteins predicted to be disordered by both methods).^{87–89} Based on this classification, there are 13 prostate cancer-associated proteins with extended disorder (P05019, P14625, P36402, P46527, Q00987, Q92934, Q99801, Q9UJU2, P38936, P07900, Q92569, Q9Y6K9, and Q02930) and there are at least 29 putative native molten globules or mixed proteins containing comparable amounts of ordered and disordered regions (O00716, P04085, P04637, P06400, P10275, P10415, P16220, P25963, P49841, Q01094, Q07889, Q09472, Q12778, Q14209, Q92793, Q9NQB0, Q8WYR1, O43889, P01127, O00459, P15056, Q07890, Q04206, Q9HCS4, Q8TEY5, Q68CJ9, Q70SY1, Q96BA8, and P18848).

Furthermore, our analysis not only indicated the disorder status of these 88 proteins but also showed that disorder is crucial for functionality of many of the prostate cancer-related proteins. This conclusion follows from the analysis of the data generated for each protein in the KEGG dataset by the D^2P^2 that, in the visually attractive

form, provides an access to the precomputed disorder predictions⁵⁹ generated by PONDR[®] VLXT,⁴² IUPred,⁶¹ PONDR[®] VSL2B,^{46,65} PrDOS,⁶⁷ ESpritz,⁶² and PV2⁵⁹ and also provides information on the curated cites of various posttranslational modifications and on the location of predicted potential disorder-based binding sites. Our analysis clearly shows that many of human prostate cancer-related proteins from the KEGG dataset are predicted to have disordered regions of various lengths, often possess numerous potential disorder-based binding motifs (**Supplementary Table 7**) and contain multiple sites of various posttranslational modifications (PTMs). The finding that the IDPRs of these proteins have a multitude of PTMs is in agreement with the well-known fact that phosphorylation²³ and many other enzymatically catalyzed PTMs are preferentially located within the IDPRs.^{23–26}

The interactivity of prostate cancer-associated proteins from the KEGG dataset was further evaluated by the STRING computational platform that provides information on both experimentally validated and predicted interactions of query proteins.⁶⁰ The corresponding STRING-generated protein-protein interaction (PPI) networks of these proteins (data not shown) indicate that all KEGG proteins are predicted to have very well-developed interactomes. This observation suggests that these proteins can serve as hubs in their functional PPI networks. These findings are in accord with earlier observations that intrinsic disorder plays an important role in functionality of hubs, defining their ability to be promiscuous binders engaged in interactions with a multitude of often unrelated partners.^{90–96} Furthermore, previous studies showed that many hubs are intrinsically disordered or contain functional IDPRs, and that the partners of ordered hubs are preferentially intrinsically disordered.^{90–96} It is likely that this binding promiscuity of the prostate cancer-related proteins can be at least partially attributed to the fact that many of them have numerous AIBSs. Also, this astonishing capability of mostly ordered, hybrid, and mostly disordered proteins from the KEGG dataset to be heavily connected hubs represents a major hurdle for the development of drugs targeting these proteins.

DISCUSSION

Interaction network of prostate cancer-related proteins

Results of the application of the APID server for evaluation of the interactivity of the 87 prostate cancer-related proteins from the KEGG dataset (as it was mentioned, no protein-protein interaction-related information was available for the 3-oxo-5- α -steroid 4-dehydrogenase 2 [SRD5A2, UniProt ID: P31213]) clearly showed that each of these proteins is known to be engaged in multiple protein-protein interactions (PPIs) (**Supplementary Table 7**). In fact, the number of PPIs ranges from 4 (TCF7L1 and INSRR) to >1,000 (e.g., p53, a 393-residue-long protein for which the content of disordered residues predicted PONDR FIT (CDR_{FIT}) is 54.96%, and EGFR with the length of 1021 residues and CDR_{FIT} of 9.1%, interact with 1072 and 1031 partners, respectively). In fact, 81 proteins in this dataset are able to interact with more than 10 partners each (**Supplementary Table 7**). This observation suggests that the vast majority of the prostate cancer-related proteins from the KEGG dataset can be considered as hub proteins. **Figure 4** shows that the interactivity of these proteins is not correlated with their length (**Figure 4a**) or their intrinsic disorder content (**Figure 4b**), suggesting that these two features do not directly determine the ability of a prostate cancer-related protein to be involved in many PPIs and to serve as a hub. The aforementioned lack of a correlation between the interactability of a given protein and its disorder content seems to be in contradiction with the claim that the ability of a protein to be engaged in many PPIs relies on intrinsic disorder. However, it is

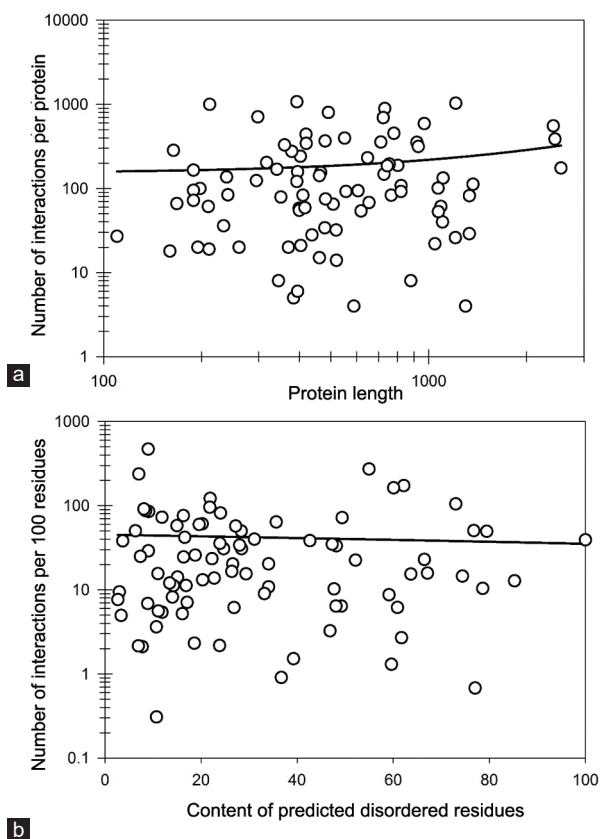


Figure 4: Characterization of the interactability of prostate cancer-related proteins from the extended KEGG dataset based on the results of their analysis by APID server. (a) Correlation between the number of PPIs found by APID for individual proteins and protein length. Note logarithmic scale of this plot. (b) Correlation between interactability (evaluated as the number of PPIs per 100 residues of a given protein) and intrinsic disorder (measured as POND FIT-based content of predicted disordered residues in a query protein). Note semi-logarithmic scale of this plot.

known that some hub proteins can be entirely disordered, other hubs may contain both ordered and disordered regions, and still other hubs can be highly structured.⁹⁷ However, the binding regions of the partner proteins of ordered hubs were found to be intrinsically disordered.^{98,99} These observations suggested two primary mechanisms by which disorder is utilized in protein-protein interaction networks, namely, one disordered region binding to many partners and many disordered region binding to one partner.^{91–94,97,100,101}

As it follows from brief description of several highly disordered proteins related to prostate cancer, it is not uncommon to find them to be involved in interaction with each other. To understand how common this phenomenon is, we used the ability of the APID web server (<http://apid.dep.usal.es>) to build a specific PPI network between proteins included in a query list.⁶⁹ Figure 5 represents the results of application of this tool to prostate cancer-related proteins from the KEGG dataset and shows that all proteins with known interactions are involved in the formation of a common interactive cluster, where each prostate cancer-related protein interacts with at least one other prostate cancer-related protein. Figure 5 shows results of this analysis in a form of a grid, where each node corresponds to a protein from the KEGG dataset, and where PPIs are shown as corresponding edges, thickness of which reflects the reliability of a

given interaction. The resulting prostate cancer-related interactome clearly shows that almost all proteins currently known to be related to the pathogenesis of this disease are talking to each other. Therefore, both internal (interactions with other prostate cancer-related proteins) and external connectivities (interaction with other proteins) are high for many prostate cancer-related proteins.

Functionality of intrinsic disorder in a major player, androgen receptor

Androgen receptor (AR, CDR_{FIT} is 42.7%) is one of the steroid hormone receptors. AR is a 920 residue-long ligand-activated transcription factor controlling expression of various eukaryotic genes and affecting proliferation and differentiation of cells in target tissues. AR plays a key role in the development and progression of prostate cancer and contributes to the development of resistance to androgen deprivation leading to the formation of the castration-resistant form of prostate cancer.¹⁰² There are several functional regions/domains in the AR. For example, the most AR transcriptional activity is due to the N-terminal domain (NTD) transcriptional activation domain whereas interaction with DNA is attributed to the central DNA-binding domain that contains 2 zinc-finger motifs. Nuclear localization upon activation is driven by a short, flexible, hinge located after the DNA-binding domain. Finally, interaction of AR with ligands is conducted through its C-terminal ligand-binding domain (LBD).

Structural information is available for the LBD (residues 659–920) and the N-terminal peptide, AR_{20–30}.¹⁰³ Analysis of the solution structure of AR NTD (residues 1–537) by circular dichroism revealed that this domain has a relatively limited amount of stable secondary structure.¹⁰⁴ The lack of stable structure in NTD is further supported by the results of evaluation of disorder predisposition in this protein as shown in **Supplementary Table 7**. As typical for IDPs and hybrid proteins containing ordered domains and IDPRs, AR has several regions with compositional biases, such as Gln-rich region (residues 58–120), poly-Gln regions (residues 58–80, 86–91, and 195–199), a poly-Pro region (residues 374–383), and poly-Ala and poly-Gly regions (residues 398–404 and 451–473, respectively). There are four alternatively spliced isoforms of this protein, where in comparison with the canonical form, isoform 2 misses residues 1–532 and has a GPYGDMR → MILWLHS substitution at position 533–539; isoform 3 has a substitution ARKLLKLGNLKQEEG → EKFRVGNCKHLKMTRP (residues 629–644) and misses region 645–920; finally, in the isoform 4, residues 649–920 are missing and region 630–648 has a RKLKLGNLKQEEGEASS → AVVVSERILRVFGVSEWLP substitution. As follows from our analysis, the majority of regions affected by alternative splicing is predicted to be disordered.

CONCLUSIONS

This study shows that intrinsically disordered proteins and proteins with long disordered regions are commonly found in prostate cancer. Many of these proteins are predicted to be moderately or highly disordered using either the per-protein disorder score (DS) for classification of proteins as highly ordered ($DS < 0.1$), moderately disordered ($0.1 \leq DS < 0.3$), and highly disordered ($DS \geq 0.3$) or looking at the per-protein content of disordered residues (CDRs), with proteins being classified as highly ordered, moderately disordered, or highly disordered, if their $CDR < 10\%$, $10\% \leq CDR < 30\%$, and $CDR \geq 30\%$, respectively. Functions of these proteins are regulated by various posttranslational modifications. Furthermore, many of these proteins are promiscuous binders and contain numerous disorder-based binding sites. We also show that irrespectively of their

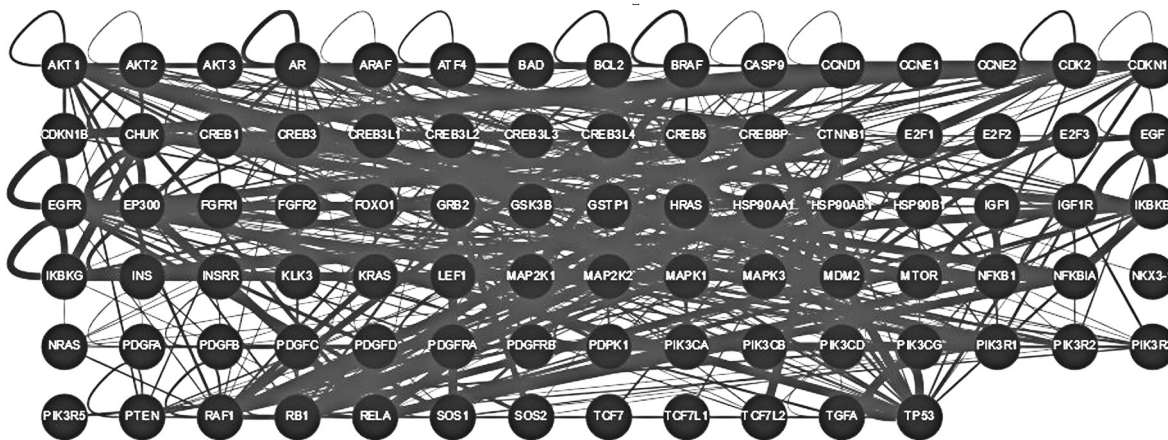


Figure 5: Evaluation of the inter-set interactivity of prostate cancer-related proteins from the KEGG dataset using the APID web server (<http://apid.dep.usal.es>). This tool builds a PPI network between proteins included in a query list. Results are shown in a form of a grid, where each node corresponds to a protein from the KEGG dataset, and where PPIs are shown as corresponding edges, thickness of which reflects the reliability of a given interaction.

disorder status (i.e., irrespectively of being mostly ordered, mixed, or mostly disordered), these proteins are characterized by an astonishing capability to be heavily connected hubs involved in a broad range of interactions. This binding promiscuity might represent a major hurdle for the development of drugs targeting these proteins.

Although our analysis revealed that potential biomarkers of prostate cancer are characterized by lower disorder levels than an average human protein, in our view, there is no contradiction between these observations and our discussion of the fact that IDPs and proteins with long disordered regions are commonly found in prostate cancer, where they might play a number of important roles. In fact, what we are showing is that intrinsically disordered regions are commonly present among prostate cancer-related proteins. Furthermore, the fact that disorder might be more commonly found in the whole human proteome than in a set of proteins related to prostate cancer does not mean that IDPs are not important for progression of this disease. In addition to the information on how many IDPs are associated with prostate cancer, a very important consideration is who these cancer-related IDPs are and what they do. For example, AR, PTEN, and p53 are known as major players in prostate cancer development (and, as a matter of fact, PTEN and p53 are involved in the development of many other pathological conditions). Loss of function of these three proteins (all of which have long disordered regions) contributes greatly to disease progression. Another example is given by NKX3.1, which is also a very disordered protein, whose loss is associated with the prostate cancer development. Therefore, we think that our data make an important contribution to the field by bringing attention to IDPs potentially related to prostate cancer.

AUTHOR CONTRIBUTIONS

KSL collected datasets, conducted computational analysis, participated in data analysis, and drafted the manuscript. IN participated in collecting datasets, conducted computational analysis, participated in data analysis, conducted statistical analysis, and drafted the manuscript. ROS participated in data analysis and contributed to drafting the manuscript. VNU conceived the idea of the study, designed and coordinated experiments, analyzed data, conducted computational analysis, and drafted the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declared lack of any competing financial interest.

Supplementary information is linked to the online version of the paper on the *Asian Journal of Andrology* website.

REFERENCES

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016; 66: 7–30.
- Zhu Y, Han CT, Zhang GM, Liu F, Ding Q, *et al*. Effect of body mass index on the performance characteristics of PSA-related markers to detect prostate cancer. *Sci Rep* 2016; 6: 19034.
- Uversky VN. Dancing protein clouds: the strange biology and chaotic physics of intrinsically disordered proteins. *J Biol Chem* 2016; 29: 6681–8.
- van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, *et al*. Classification of intrinsically disordered regions and proteins. *Chem Rev* 2014; 114: 6589–631.
- Habchi J, Tompa P, Longhi S, Uversky VN. Introducing protein intrinsic disorder. *Chem Rev* 2014; 114: 6561–88.
- Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 2000; 11: 161–71.
- Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, *et al*. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* 2015; 72: 137–51.
- Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 2004; 337: 635–45.
- Xue B, Dunker AK, Uversky VN. 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: 137–49.
- Tompa P, Dosztanyi Z, Simon I. Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. *J Proteome Res* 2006; 5: 1996–2000.
- Tompa P, Davey NE, Gibson TJ, Babu MM. A million peptide motifs for the molecular biologist. *Mol Cell* 2014; 55: 161–9.
- Vincent M, Whidden M, Schnell S. Quantitative proteome-based guidelines for intrinsic disorder characterization. *Biophys Chem* 2016; 213: 6–16.
- Charon J, Theil S, Nicaise V, Michon T. Protein intrinsic disorder within the *Potyvirus* genus: from proteome-wide analysis to functional annotation. *Mol Biosyst* 2016; 12: 634–52.
- Skupien-Rabian B, Jankowska U, Swiderska B, Lukaszewicz S, Ryszawy D, *et al*. Proteomic and bioinformatic analysis of a nuclear intrinsically disordered proteome. *J Proteomics* 2016; 130: 76–84.
- Panca R, Tompa P. Structural disorder in eukaryotes. *PLoS One* 2012; 7: e34687.
- Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, *et al*. Intrinsically disordered protein. *J Mol Graph Model* 2001; 19: 26–59.
- Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, *et al*. Intrinsic disorder and functional proteomics. *Biophys J* 2007; 92: 1439–56.
- Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 2000; 41: 415–27.

- 19 Uversky VN. Unusual biophysics of intrinsically disordered proteins. *Biochim Biophys Acta* 2013; 1834: 932–51.
- 20 Uversky VN. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci* 2013; 22: 693–724.
- 21 Uversky VN. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem Soc Rev* 2011; 40: 1623–34.
- 22 Uversky VN. Intrinsic disorder-based protein interactions and their modulators. *Curr Pharm Des* 2013; 19: 4191–213.
- 23 Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, *et al*. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 2004; 32: 1037–49.
- 24 Pejaver V, Hsu WL, Xin F, Dunker AK, Uversky VN, *et al*. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci* 2014; 23: 1077–93.
- 25 Zerze GH, Mittal J. Effect of O-linked glycosylation on the equilibrium structural ensemble of intrinsically disordered polypeptides. *J Phys Chem B* 2015; 119: 15583–92.
- 26 Kurotani A, Sakurai T. In silico analysis of correlations between protein disorder and post-translational modifications in algae. *Int J Mol Sci* 2015; 16: 19812–35.
- 27 Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. *Biochemistry* 2002; 41: 6573–82.
- 28 Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. *Adv Protein Chem* 2002; 62: 25–49.
- 29 Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 2002; 11: 739–56.
- 30 Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem* 2002; 269: 2–12.
- 31 Uversky VN. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? *Cell Mol Life Sci* 2003; 60: 1852–71.
- 32 Tompa P, Schad E, Tantos A, Kalmar L. Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struct Biol* 2015; 35: 49–59.
- 33 Tompa P, Szasz C, Buday L. Structural disorder throws new light on moonlighting. *Trends Biochem Sci* 2005; 30: 484–9.
- 34 Meng F, Na I, Kurgan L, Uversky VN. Compartmentalization and functionality of nuclear disorder: intrinsic disorder and protein-protein interactions in intra-nuclear compartments. *Int J Mol Sci* 2015; 17: 24.
- 35 Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 2002; 323: 573–84.
- 36 Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 2008; 37: 215–46.
- 37 Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, *et al*. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2004; 32: D115–9.
- 38 Shishkin SS, Kovalyov LI, Kovalyova MA, Lisitskaya KV, Eremina LS, *et al*. "Prostate cancer proteomics" database. *Acta Naturae* 2010; 2: 95–104.
- 39 Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010; 38: D355–60.
- 40 Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, *et al*. QuickGO: a web-based tool for gene ontology searching. *Bioinformatics* 2009; 25: 3045–6.
- 41 Huntley RP, Sawford T, Mutowo-Muullenet P, Shypitsyna A, Bonilla C, *et al*. The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res* 2015; 43: D1057–63.
- 42 Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, *et al*. Sequence complexity of disordered protein. *Proteins* 2001; 42: 38–48.
- 43 Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and internal regions. *Genome Inform Ser Workshop Genome Inform* 1999; 10: 30–40.
- 44 Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 2010; 1804: 996–1010.
- 45 Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, *et al*. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol* 2005; 3: 35–60.
- 46 Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* 2005; 61 Suppl 7: 176–82.
- 47 Walsh I, Giollo M, Di Domenico T, Ferrari C, Zimmermann O, *et al*. Comprehensive large-scale assessment of intrinsic protein disorder. *Bioinformatics* 2015; 31: 201–8.
- 48 Fan X, Kurgan L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *J Biomol Struct Dyn* 2014; 32: 448–64.
- 49 Peng Z, Kurgan L. On the complementarity of the consensus-based disorder prediction. *Pac Symp Biocomput* 2012; 176–87.
- 50 Pietrosemoli N, Garcia-Martin JA, Solano R, Pazos F. Genome-wide analysis of protein disorder in *Arabidopsis thaliana*: implications for plant environmental adaptation. *PLoS one* 2013; 8: e55524.
- 51 Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, *et al*. AmiGO: online access to ontology and annotation data. *Bioinformatics* 2009; 25: 288–9.
- 52 Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS one* 2011; 6: e21800.
- 53 Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 2006; 7: 208.
- 54 Di Domenico T, Walsh I, Martin AJ, Tosatto SC. MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics* 2012; 28: 2080–1.
- 55 Potenza E, Domenico TD, Walsh I, Tosatto SC. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res* 2015; 43 (Database issue): D315–20.
- 56 Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, *et al*. Comparing and combining predictors of mostly disordered proteins. *Biochemistry* 2005; 44: 1989–2000.
- 57 Meszaros B, Simon I, Dosztanyi Z. Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* 2009; 5: e1000376.
- 58 Dosztanyi Z, Meszaros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 2009; 25: 2745–6.
- 59 Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, *et al*. D (2) P (2): database of disordered protein predictions. *Nucleic Acids Res* 2013; 41: D508–16.
- 60 Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, *et al*. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011; 39: D561–8.
- 61 Dosztanyi 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: 3433–4.
- 62 Walsh I, Martin AJ, Di Domenico T, Tosatto SC. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* 2012; 28: 503–9.
- 63 Lindner R, Jensen LJ, Diella F, Bork P, Gibson TJ, *et al*. Protein disorder prediction: implications for structural proteomics. *Structure* 2003; 11: 1453–9.
- 64 Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 2005; 21: 3369–76.
- 65 Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 2006; 7: 208.
- 66 Lindner R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 2003; 31: 3701–8.
- 67 Ishida T, Kinoshita K. PRDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* 2007; 35: W460–4.
- 68 Dosztanyi Z, Csizmek 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: 827–39.
- 69 Alonso-Lopez D, Gutierrez MA, Lopes KP, Prieto C, Santamaria R, *et al*. APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. *Nucleic Acids Res* 2016. [pii: gkw363]. [Epub ahead of print].
- 70 Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, *et al*. The BioGRID interaction database: 2015 update. *Nucleic Acids Res* 2015; 43: D470–8.
- 71 Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, *et al*. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 2004; 32: D449–51.
- 72 Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, *et al*. Human protein reference database-2009 update. *Nucleic Acids Res* 2009; 37: D767–72.
- 73 Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, *et al*. The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 2012; 40: D841–6.
- 74 Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, *et al*. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 2012; 40: D857–61.
- 75 Huttlin EL, Ting L, Bruckner RJ, Gebreb F, Gygi MP, *et al*. The BioPlex network: a systematic exploration of the human interactome. *Cell* 2015; 162: 425–40.
- 76 Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, *et al*. The RCSB protein data bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res* 2015; 43: D345–56.
- 77 Rajagopalan K, Mooney SM, Parekh N, Getzenberg RH, Kulkarni P. A majority of the cancer/testis antigens are intrinsically disordered proteins. *J Cell Biochem* 2011; 112: 3256–67.
- 78 Bienkiewicz EA, Adkins JN, Lumb KJ. Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). *Biochemistry* 2002; 41: 752–9.
- 79 Chi SW, Kim DH, Lee SH, Chang I, Han KH. Pre-structured motifs in the natively unstructured preS1 surface antigen of hepatitis B virus. *Protein Sci* 2007; 16: 2108–17.
- 80 Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, *et al*. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J Biol Chem* 2000; 275: 29426–32.
- 81 Ramelot TA, Gentile LN, Nicholson LK. Transient structure of the amyloid precursor protein cytoplasmic tail indicates preordering of structure for binding to cytosolic factors. *Biochemistry* 2000; 39: 2714–25.
- 82 Sayers EW, Gerstner RB, Draper DE, Torchia DA. Structural preordering in the

- N-terminal region of ribosomal protein S4 revealed by heteronuclear NMR spectroscopy. *Biochemistry* 2000; 39: 13602–13.
- 83 Zitzewitz JA, Ibarra-Molero B, Fishel DR, Terry KL, Matthews CR. Prefolded secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J Mol Biol* 2000; 296: 1105–16.
- 84 Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting Binding regions within disordered proteins. *Genome Inform Ser Workshop Genome Inform* 1999; 10: 41–50.
- 85 Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, *et al*. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 2005; 44: 12454–70.
- 86 Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, *et al*. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* 2007; 46: 13468–77.
- 87 Mohan A, Sullivan WJ Jr., Radivojac P, Dunker AK, Uversky VN. Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. *Mol Biosyst* 2008; 4: 328–40.
- 88 Xue B, Oldfield CJ, Dunker AK, Uversky VN. CDF it all: consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. *FEBS Lett* 2009; 583: 1469–74.
- 89 Huang F, Oldfield C, Meng J, Hsu WL, Xue B, *et al*. Subclassifying disordered proteins by the CH-CDF plot method. *Pac Symp Biocomput* 2012: 128–39.
- 90 Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 2005; 18: 343–84.
- 91 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: 2041–5.
- 92 Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, *et al*. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* 2006; 2: e100.
- 93 Ekman D, Light S, Bjorklund AK, Elofsson A. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol* 2006; 7: R45.
- 94 Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* 2006; 5: 2985–95.
- 95 Singh GP, Ganapathi M, Sandhu KS, Dash D. Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins* 2006; 62: 309–15.
- 96 Dunker AK, Garner E, Guillot S, Romero P, Albrecht K, *et al*. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput* 1998: 473–84.
- 97 Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets: the roles of intrinsic disorder in protein interaction networks. *FEBS J* 2005; 272: 5129–48.
- 98 Bustos DM, Iglesias AA. Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins* 2006; 63: 35–42.
- 99 Radivojac P, Vucetic S, O'Connor TR, Uversky VN, Obradovic Z, *et al*. Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. *Proteins* 2006; 63: 398–410.
- 100 Singh GP, Dash D. Intrinsic disorder in yeast transcriptional regulatory network. *Proteins* 2007; 68: 602–5.
- 101 Singh GP, Ganapathi M, Dash D. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins* 2007; 66: 761–5.
- 102 Anantharaman A, Friedlander TW. Targeting the androgen receptor in metastatic castrate-resistant prostate cancer: a review. *Urol Oncol* 2015; Doi:10.1016/j.urolonc.2015.11.003. [Epub ahead of print].
- 103 He B, Gampe RT Jr., Kole AJ, Hnat AT, Stanley TB, *et al*. Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell* 2004; 16: 425–38.
- 104 Davies P, Watt K, Kelly SM, Clark C, Price NC, *et al*. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *J Mol Endocrinol* 2008; 41: 301–14.

Supplementary materials

Unfoldomics of Prostate Cancer: On the Abundance and Roles of Intrinsically Disordered Proteins in Prostate Cancer

Kevin S. Landau,¹ Insung Na,¹ Ryan O. Schenck,¹ and Vladimir N. Uversky^{1,2,3,*}

¹Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA;

²USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, Florida 33612, USA;

³Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

***To whom correspondence should be addressed:** VNU, Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Blvd. MDC07, Tampa, Florida 33612, USA, Tel: 1-813-974-5816; E-mail: vuversky@health.usf.edu

Non-redundant list of human proteins related to prostate cancer from “Russian” dataset.

P32019, Q12907, P11021, P30533, Q9P2E9, P01033, P02787, P07288, Q16548, P07858, P09382, P07737, P04179, P04406, P68871, P15531, P22392, P02788, P06733, O75608, P13804, O75874, P60174, Q02252, P45880, P48735, Q03154, O43809, P14550, P21796, P15121, Q06830, P05164, P69905, P00558, P62937, P35228, P30084, P40939, P02768, P30041, P25787, P14174, Q99798, Q99714, P36952, P20618, Q9BUP3, P68371, P23284, P01009, P61457, P63261, P55084, P30044, Q9BPW8, P34897, P60709, P28838, P60981, P07954, P30101, P14618, P08758, Q9NR45, P04075, Q00796, P78371, O95994, O14933, P21291, P68104, P25398, Q13011, P49411, P49773, Q01469, P02766, P24752, P18669, P00367, P48047, P30042, P25705, P37802, P61088, P30040, Q16836, P00390, O75828, Q9BTZ2, P07355, O14818, P14854, O43678, Q9Y603, P50454, P00915, P04083, Q99497, P10809, P60903, P02794, Q15404, P23528, P61604, P02792, P30046, O60493, P30086, Q8N5N7, Q04760, Q00688, P07108, P17980, Q01995, P02675, P31949, P13693, P06703, P55072, P01857, P38117, P38646, Q9H4G4, P51570, P52597, P51911, P00441, Q7Z6V5, P54819, P62277, O00151, P31948, Q6DCA0, P62987, P31943, P01834, Q9UJ78, P02511, Q9NQ35, Q9HB71, P02647, P24844, O94776, P06702, Q9Y4L1, P20472, Q7LC44, P18206, Q9Y4X0, P04792, O60664, P22466, P11940, Q8N183, Q99878, Q32P51, P04908, P47813, P06899, Q96QV6, P02671, Q99879, Q16778, P62263, P62807, P68431, P17661, Q13526, P84243, O60841, P22626, P09651, P06748, P51991, P08670, Q9UII2, Q96AE4, Q96T37, Q9NX63, P02545, P09429, P06753, P07951, P09493, Q96T23, P67936, P23246, P63162, P38159, Q9H2D6, Q05682, P16949, P08123, Q96IZ7

Contingency tables

Supplementary Table 1. Illustrative representation of the 2×2 contingency table to find significantly associated GO terms

	\geq AVG (Membrane- / Non-membrane)	$<$ AVG (Membrane- / Non-membrane)
KEGG / Russian dataset	a	b
Whole human proteome	c	d

Supplementary Table 2. 2×2 contingency table of the membrane KEGG dataset and membrane proteins in whole human proteome (membrane AVG: 0.301)

Number of Proteins	\geq Membrane AVG	$<$ Membrane AVG
KEGG dataset	2	2
Whole human proteome	2210	2979

Supplementary Table 3. 2×2 contingency table of the non-membrane KEGG dataset and non-membrane proteins in whole human proteome (non-membrane AVG: 0.437).

Number of Proteins	\geq Non-Membrane AVG	$<$ Non-Membrane AVG
KEGG dataset	17	27
Whole human proteome	6555	8328

Supplementary Table 4. 2×2 contingency table of the membrane Russian dataset and membrane proteins in whole human proteome

Number of Proteins	\geq Membrane AVG	$<$ Membrane AVG
Russian dataset	4	2
Whole human proteome	2208	2981

Supplementary Table 5. 2×2 contingency table of the non-membrane Russian dataset and non-membrane proteins in whole human proteome

Number of Proteins	\geq Non-Membrane AVG	$<$ Non-Membrane AVG
Russian dataset	59	133
Whole human proteome	6513	8222

Supplementary Table 6. Prostate cancer-related proteins from the KEGG and Russian datasets with disorder-associated common Gene Ontology terms

p-value < 0.05			p-value < 0.1		
ID	Category	Name	ID	Category	Name
GO:0005575	C	cellular_component	GO:0003674	F	molecular_function
GO:0005623	C	cell	GO:0005575	C	cellular_component
GO:0006950	P	response to stress	GO:0005622	C	intracellular
GO:0008150	P	biological_process	GO:0005623	C	cell
GO:0044464	C	cell part	GO:0006915	P	apoptotic process
GO:0050896	P	response to stimulus	GO:0006950	P	response to stress
			GO:0008150	P	biological_process
			GO:0008219	P	cell death
			GO:0009987	P	cellular process
			GO:0012501	P	programmed cell death
			GO:0016043	P	cellular component organization
			GO:0016265	P	death
			GO:0031982	C	vesicle
			GO:0043226	C	organelle
			GO:0043227	C	membrane-bounded organelle
			GO:0043229	C	intracellular organelle
			GO:0043231	C	intracellular membrane-bounded organelle
			GO:0044422	C	organelle part
			GO:0044424	C	intracellular part
			GO:0044446	C	intracellular organelle part
			GO:0044464	C	cell part
			GO:0044699	P	single-organism process
			GO:0044763	P	single-organism cellular process
			GO:0050896	P	response to stimulus
			GO:0071840	P	cellular component organization or biogenesis

Supplementary Table 7. Characterization of human proteins involved in the prostate cancer pathway

Protein	UniProt ID	Protein length (N _{AIBS}) ^a	PONDR-FIT (%) ^b	MobiDB consensus (%) ^c	Location (length) of long disordered regions ^d	Location (length) of AIBSs ^e	N _{int} ^f
BAD, Bcl2-associated agonist of cell death	Q92934	168 (4/70.8)	100.00	84.54	1-105 (105) 122-147 (27) 158-168 (11)	1-53 (53) 57-80 (24) 100-129 (30) 146-157 (12)	66
CREB5; cyclic AMP-responsive element binding protein 5	Q02930	508 (7/67.9)	85.24	75.39	46-59 (14) 86-393 (308) 447-470 (24) 479-508 (31)	66-86 (21) 99-183 (85) 188-358 (171) 362-370 (9) 378-406 (29) 421-444 (24) 503-508 (6)	65
CREB1, cyclic AMP-responsive element-binding protein 1	P16220	341 (7/29.3)	79.47	40.47	1-32 (32) 40-50 (11) 102-132 (33) 138-171 (34) 271-285 (15)	32-44 (13) 89-104 (16) 128-145 (18) 166-191 (26) 265-270 (6) 307-314 (8) 329-341 (13)	169
FOXO1, Forkhead box protein O1	Q12778	655 (19/56.9)	78.63	72.82	1-69 (69) 74-101 (28) 105-160 (56) 199-210 (12) 229-336 (107) 385-450 (66) 463-488 (26) 498-569 (72) 644-655 (12)	1-32 (32) 54-82 (29) 88-118 (31) 160-172 (13) 182-196 (15) 216-226 (11) 258-280 (23) 289-297 (9) 306-314 (9) 323-365 (43) 371-388 (18) 301-409 (8) 447-469 (23) 483-517 (35) 528-545 (18) 550-565 (16) 570-592 (23) 605-612 (8)	68
TCF7L1, transcription factor 7 like 1	Q9HCS4	588 (16/54.5)	77.04	61.90	1-104 (104) 161-183 (23) 192-238 (47) 316-344 (29) 406-512 (107) 524-546 (21)	1-46 (46) 53-74 (22) 94-135 (42) 146-159 (14) 191-201 (11) 234-252 (19) 274-288 (15) 349-371 (23) 373-383 (11) 407-415 (9) 444-463 (20) 470-494 (25) 498-505 (8) 511-529 (19) 541-558 (18) 566-583 (18)	4
p27Kip1 (CDKN1B), cyclin-dependent kinase	P46527	198 (8/58.6)	76.77	90.40	1-37 (37) 50-67 (18)	1-8 (8) 27-37 (11)	100

inhibitor 1B					81-198 (118)	59-67 (9) 80-90 (11) 98-105 (8) 109-134 (26) 140-157 (18) 174-198 (35)	
LEF1, lymphoid enhancer-binding factor 1	Q9UJU2	399 (8/62.7)	74.4	66.92	1-104 (104) 136-148 (13) 161-201 (41) 240-301 (62) 369-399 (31)	1-84 (84) 99-121 (23) 125-131 (7) 142-165 (24) 193-247 (55) 258-269 (12) 276-283 (8) 300-336 (37)	58
IKBKG, inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma	Q9Y6K9	419 (6/15.8)	73.03	24.58	15-35 (21) 171-190 (20) 356-396 (41)	1-15 (15) 33-41 (9) 230-242 (13) 304-310 (7) 372-380 (9) 407-419 (13)	440
CREBBP, CREB-binding protein	Q92793	2,442 (34/53.4)	67.12	66.01	1-48 (48) 51-246 (196) 249-323 (75) 441-591 (151) 669-713 (45) 726-1091 (366) 1250-1276 (27) 1551-1619 (69) 1748-1769 (22) 1854-2066 (213) 2109-2442 (334)	1-7 (7) 17-102 (86) 109-140 (32) 145-164 (21) 171-221 (51) 232-247 (16) 254-283 (30) 295-314 (20) 353-365 (13) 479-500 (22) 515-578 (64) 599-613 (15) 628-635 (8) 649-662 (14) 673-683 (11) 707-768 (62) 776-879 (104) 881-898 (18) 902-1004 (103) 1017-1059 (43) 1084-1102 (19) 1121-1127 (7) 1537-1551 (15) 1616-1639 (24) 1844-1851 (8) 1857-1878 (22) 1911-1943 (33) 1952-2028 (77) 2031-2054 (23) 2060-2077 (18) 2084-2114 (31) 2132-2200 (69) 2217-2353 (137) 2361-2442 (82)	386
EP300, E1A binding protein p300, p300 HAT, histone acetyltransferase p300	Q09472	2,414 (30/52.9)	66.49	61.31	1-37 (37) 84-170 (87) 191-269 (79) 282-334 (53) 422-458 (37) 478-569 (02)	1-7 (7) 18-135 (118) 173-203 (31) 212-226 (15) 233-248 (16) 264-295 (32)	554

					649-1054 (406) 1213-1233 (21) 1515-1582 (68) 1715-1739 (25) 1817-2044 (228) 2093-2414 (322)	314-323 (10) 336-349 (14) 456-479 (24) 504-554 (51) 578-592 (15) 608-613 (6) 627-642 (16) 654-668 (15) 687-834 (148) 844-883 (39) 890-988 (99) 994-1010 (17) 1018-1023 (7) 1501-1515 (15) 1579-1603 (25) 1822-1848 (27) 1881-1908 (28) 1914-2063 (150) 2111-2191 (81) 2194-2314 (121) 2324-2414 (91)	
NKX3-1, homeobox protein Nkx-3.1	Q99801	234 (3/50.4)	63.67	59.40	1-133 (133)	1-46 (46) 66-116 (51) 129-149 (21)	36
CDKN1A, cyclin-dependent kinase inhibitor 1	P38936	164 (5/36.0)	62.2	65.9	1-17 (17) 73-115 (43) 117-164 (48)	36-41 (6) 68-79 (12) 100-105 (6) 109-123 (15) 145-164 (20)	284
CREB3L2, cAMP responsive element binding protein 3-like 2	Q70SY1	520 (8/32.3)	61.73	44.04	61-79 (19) 120-139 (20) 149-167 (19) 195-265 (71) 303-327 (25) 433-452 (20)	30-58 (29) 76-88 (13) 109-122 (14) 138-153 (16) 163-202 (40) 217-232 (16) 239-267 (29) 273-283 (11)	14
CREB3L1, cAMP responsive element binding protein 3-like 1	Q96BA8	519 (9/32.4)	60.89	42.77	69-112 (44) 196-261 (66) 300-312 (13) 446-519 (74)	53-61 (9) 115-129 (15) 142-168 (27) 178-204 (27) 230-264 (35) 268-280 (13) 424-435 (12) 479-487 (9) 499-519 (21)	32
MDM2, E3 ubiquitin-protein ligase Mdm2	Q00987	491 (8/21.8)	60.08	57.43	1-21 (21) 112-193 (82) 209-244 (36) 251-295 (45) 331-365 (35) 370-428 (59)	100-108 (9) 189-208 (20) 235-261 (27) 273-284 (12) 296-306 (11) 322-329 (8) 359-367 (9) 399-409 (11)	801
TCF7, transcription factor 7	P36402	384 (7/61.7)	59.63	71.88	1-115 (115) 128-190 (63) 230-271 (42) 332-384 (53)	1-31 (31) 41-133 (93) 156-167 (12) 178-224 (47) 272-305 (34)	5

						328-336 (9) 374-384 (11)	
TCF7L2, Transcription factor 7-like 2	Q9NQB0	619 (13/46.2)	59.13	58.97	1-96 (96)	1-25 (25) 36-52 (17) 88-115 (28) 145-162 (18) 200-209 (10) 243-265 (23) 268-293 (26) 308-317 (10) 351-387 (37) 446-452 (7) 487-492 (6) 531-585 (55) 596-619 (24)	54
p53, cellular tumor antigen p53	P04637	393 (8/39.4)	54.96	50.89	1-23 (23) 33-97 (65) 277-328 (52) 346-393 (46)	11-57 (47) 106-115 (10) 132-141 (10) 232-239 (8) 251-258 (8) 265-277 (13) 322-355 (34) 363-387 (25)	1072
ATF4, activating transcription factor 4	P18848	351 (9/33.6)	52.14	57.55	66-77 (12) 131-149 (19) 168-194 (27) 207-315 (109)	94-102 (9) 115-127 (13) 154-169 (16) 177-187 (11) 196-209 (14) 219-230 (12) 246-257 (12) 267-284 (18) 326-338 (13)	79
RELA, v-rel avian reticuloendotheliosis viral oncogene homolog A	Q04206	551 (12/38.5)	49.36	37.93	39-64 (26) 75-92 (18) 273-348 (76) 376-398 (23) 414-437 (24) 505-530 (26)	1-11 (11) 31-41 (11) 62-76 (15) 98-103 (6) 110-118 (9) 285-290 (6) 305-317 (13) 350-380 (31) 398-414 (17) 433-483 (51) 492-504 (13) 523-551 (29)	397
Transcription factor E2F2	Q14209	437 (10/33.6)	49.20	43.48	1-26 (26) 56-66 (11) 109-130 (22) 196-210 (15) 301-389 (89)	1-9 (9) 37-47 (11) 75-82 (8) 135-145 (11) 228-235 (8) 248-256 (9) 285-307 (23) 319-331 (13) 335-349 (15) 364-403 (40)	28
Transcription factor E2F1	Q01094	437 (10/33.4)	48.08	38.90	37-129 (93) 299-356 (68)	16-39 (24) 57-71 (15) 87-105 (19) 132-143 (12) 245-251 (7)	28

						250-265 (6) 281-305 (25) 325-332 (8) 345-365 (21) 376-384 (9)	
Transcription factor E2F3	O00716	465 (6/28.8)	48.17	40.00	67-180 (114) 350-396 (47) 402-419 (18)	95-122 (28) 132-161 (30) 180-195 (16) 333-253 (21) 386-407 (22) 421-437 (17)	155
IGF1, insulin-like growth factor I	P05019	195 (2/13.8)	47.69	51.28	77-87 (11) 113-195 (83)	166-177 (12) 181-195 (15)	20
PDGFB, platelet-derived growth factor subunit B	P01127	241 (2/12.4)	47.30	33.20	50-65 (16) 67-76 (10) 188-241 (61)	208-216 (9) 221-241 (21)	84
CREB3L3, cAMP responsive element binding protein 3-like 3	Q68CJ9	461 (11/35.6)	46.85	41.87	51-120 (70) 124-139 (16) 251-268 (18) 369-414 (46) 442-461 (20)	20-36 (17) 60-93 (34) 121-127 (7) 140-157 (18) 178-186 (9) 198-204 (7) 356-364 (9) 367-376 (10) 399-412 (14) 416-445 (30) 453-461 (9)	15
AR, androgen receptor	P10275	919 (12/28.1)	42.66	35.80	1-18 (18) 31-167 (137) 183-227 (45) 293-312 (20) 368-388 (21) 415-433 (19) 445-479 (35) 635-659 (25)	1-59 (59) 94-112 (19) 118-150 (33) 155-194 (40) 236-252 (17) 264-273 (10) 342-351 (10) 356-368 (13) 391-412 (22) 429-447 (19) 476-481 (6) 495-504 (10)	353
CREB3L4, cAMP responsive element binding protein 3-like 4	Q8TEY5	395 (8/23.8)	39.24	35.95	43-59 (17) 81-109 (29) 221-241 (21) 318-345 (27) 351-395 (45)	1-13 (13) 20-31 (12) 65-81 (17) 110-121 (12) 135-143 (9) 180-188 (9) 368-377 (10) 384-395 (12)	6
PIK3R5, phosphoinositide 3-kinase regulatory subunit 5	Q8WYR1	880 (7/13.6)	36.7	29.2	312-349 (38) 351-370 (20) 386-418 (33) 437-515 (79) 563-603 (41) 783-793 (11) 838-858 (21)	339-364 (26) 371-387 (17) 420-435 (16) 451-463 (13) 485-495 (11) 516-536 (21) 547-562 (16)	8
NFKBIA, NF-kappa-B inhibitor alpha	P25963	317 (3/9.8)	35.65	26.81	1-42 (42) 52-67 (16) 280-298 (19)	1-6 (6) 76-64 (9) 302-317 (16)	203
PIK3R2, phosphoinositide-3-	O00459	728 (12/)	34.07	27.65	81-120 (40) 140-153 (14)	29-43 (15) 70-84 (15)	148

kinase regulatory subunit 2					189-213 (25) 251-285 (35) 289-324 (36) 544-560 (17) 714-728 (15)	109-118 (10) 122-136 (15) 169-183 (15) 218-231 (14) 241-256 (16) 277-298 (22) 312-320 (9) 325-335 (11) 395-400 (6) 574-581 (8)	
BRAF, B-Raf proto-oncogene, serine/threonine kinase	P15056	766 (8/17.1)	34.07	31.07	1-40 (41) 110-132 (23) 141-156 (16) 305-455 (151)	16-21 (6) 46-67 (22) 78-86 (9) 130-135 (6) 326-332 (7) 339-347 (9) 356-419 (64) 467-474 (8)	83
PDGF-1 (PDGFA), platelet-derived growth factor subunit A	P04085	211 (0/0)	33.18	28.44	61-82 (22) 181-211 (31)	N.P. ^g	19
CREB3, cyclic AMP-responsive element-binding protein 3	O43889	395 (2/4.6)	31.1	12.9	120-131 (12) 182-196 (15)	11-17 (7) 105-115 (11)	158
ARAF, Proto-oncogene A-Raf, serine/threonine-protein kinase A-Raf	P10398	606 (2/11.4)	29.37	29.04	1-17 (17) 156-224 (69) 231-294 (64) 579-595 (17)	193-247 (55) 290-303 (14)	94
PIK3R1, phosphoinositide-3-kinase regulatory subunit 1	P27986	712 (8/8.3)	28.37	12.57	78-112 (35) 301-322 (22) 608-631 (24)	7-12 (6) 26-43 (18) 69-79 (11) 112-119 (8) 131-136 (6) 331-336 (6) 580-585 (6) 657-662 (6)	356
PIK3R3, phosphoinositide-3-kinase regulatory subunit 3	Q92569	461 (3/5.4)	28.42	6.29	39-53 (15)	23-31 (9) 218-226 (9) 257-263 (7)	142
RB1, retinoblastoma-associated protein	P06400	928 (7/10.6)	28.02	25.32	1-48 (48) 246-261 (16) 348-367 (20) 577-590 (14) 609-635 (27) 775-786 (12) 816-827 (12) 859-928 (70)	7-20 (14) 41-49 (9) 57-64 (8) 830-840 (11) 844-860 (17) 872-881 (10) 891-919 (29)	315
BCL2, B-cell CLL/lymphoma 2, apoptosis regulator Bcl-2	P10415	239 (2/10.9)	27.20	25.52	39-87 (49)	13-32 (20) 46-51 (6)	137
SOS1, Son of sevenless homolog 1	Q07889	1,333 (8/16.7)	26.86	26.48	656-665 (10) 744-757 (14) 1019-1120 (102) 1125-1333 (209)	1005-1014 (10) 1027-1037 (11) 1058-1070 (13) 1094-1146 (53) 1166-1181 (16) 1188-1205 (18) 1218-1254 (37)	82

						1269-1333 (65)	
CCNE1, G1/S-specific cyclin-E1	P24864	410 (5/11.2)	26.59	22.93	1-45 (46) 78-87 (10) 377-410 (34)	21-26 (6) 31-42 (12) 64-74 (11) 370-379 (10) 389-395 (7)	83
PDPK1, 3-phosphoinositide-dependent protein kinase 1	O15530	556 (4/11.3)	26.4	18.4	1-10 (10) 24-82 (59) 389-412 (24)	1-29 (29) 48-53 (6) 82-98 (17) 376-386 (11)	92
MAP2K1, dual specificity mitogen-activated protein kinase kinase 1	Q02750	393 (4/7.9)	24.7	17.6	1-28 (28) 281-303 (23) 383-393 (11)	8-16 (9) 30-36 (7) 272-279 (8) 310-316 (7)	121
GSK3B, GSK3 β , glycogen synthase kinase-3 beta	P49841	420 (7/15.0)	24.05	22.38	1-54 (54) 382-420 (39)	9-17 (9) 21-26 (6) 316-321 (6) 333-341 (8) 358-366 (9) 371-388 (17) 411-417 (7)	343
RAF1, Raf-1 proto-oncogene, serine/threonine kinase	P04049	648 (3/6.3)	23.92	26.54	39-55 (17) 218-335 (118) 636-648 (13)	262-289 (28) 291-296 (6) 340-346 (7)	231
SOS2, SOS Ras/Rho guanine nucleotide exchange factor 2	Q07890	1,332 (10/13.7)	23.87	23.12	742-757 (16) 1017-1064 (48) 1075-1100 (26) 1140-1332 (193)	460-465 (6) 1003-1012 (10) 1030-1035 (6) 1063-1074 (12) 1098-1141 (44) 1155-1168 (14) 1188-1200 (13) 1223-1255 (33) 1270-1285 (16) 1302-1329 (28)	29
MAP2K2, dual specificity mitogen-activated protein kinase kinase 2	P36507	400 (5/14.0)	22.75	18.50	1-10 (10) 13-34 (22) 282-311 (30)	1-16 (16) 251-250 (10) 315-325 (11) 342-351 (10) 379-387 (9)	55
Enpl (HSP90B1), endoplasmin	P14625	803 (4/10.2)	22.29	17.81	30-50 (21) 286-326 (41) 598-608 (11) 747-803 (57)	331-338 (8) 646-651 (6) 711-753 (43) 779-803 (24)	188
HSP90AA1, heat shock protein HSP 90-alpha	P07900	732 (3/4.9)	21.90	23.2	1-18 (18) 166-181 (16) 223-288 (66) 547-565 (19) 698-732 (35)	208-223 (16) 594-602 (9) 688-698 (11)	892
HSP90AB1, heat shock protein 90kDa alpha family class B member 1	P08238	724 (3/4.8)	21.82	20.99	218-278 (61) 239-255 (17) 693-724 (32)	205-218 (14) 585-594 (10) 681-691 (11)	694
FGFR1, fibroblast growth factor receptor 1	P11362	822 (8/9.6)	20.32	18.98	23-33 (11) 80-89 (10) 105-160 (56) 777-822 (46)	93-106 (14) 110-122 (13) 150-155 (6) 165-177 (13) 208-216 (9) 754-760 (7) 766-773 (8)	108

						812-820 (9)	
NFKB1, nuclear factor NF-kappa-B p105 subunit	P19838	968 (5/6.7)	20.25	20.14	1-10 (11) 66-78 (13) 359-405 (47) 420-453 (34) 463-480 (18) 892-924 (33) 946-968 (23)	369-374 (6) 403-421 (19) 454-463 (10) 781-787 (7) 931-953 (23)	590
PTEN, phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	P60484	403 (2/5.5)	19.60	16.63	351-403 (53)	336-350 (15) 397-403 (7)	241
IKKB, inhibitor of nuclear factor kappa-B kinase subunit beta	O14920	756 (4/5.7)	18.78	13.10	1-9 (10) 400-409 (10) 549-562 (14) 572-586 (15) 666-706 (41)	384-392 (9) 655-664 (10) 709-723 (15) 734-742 (9)	195
PDGFC, platelet-derived growth factor subunit C	Q9NRA1	345 (0/0)	18.60	11.88	20-31 (12) 33-46 (14)	N.P.	8
AKT3, RAC-gamma serine/threonine-protein kinase	Q9Y243	479 (2/4.2)	17.1	16.7	109-140 (32) 445-479 (35)	72-82 (11) 469-477 (9)	34
TGFA, transforming growth factor alpha	P01135	160 (0/0)	16.88	5.00	N.P.	N.P.	18
BCL-1 (CCND1), G1/S-specific cyclin-D1	P24385	295 (0/0)	16.61	13.90	258-295 (38)	N.P.	124
Insulin	P01308	110 (0/0)	16.36	0.91	N.P.	N.P.	27
AKT1, RAC-alpha serine/threonine-protein kinase	P31749	480 (3/6.0)	16.33	19.38	111-140 (30) 424-433 (10) 449-480 (32)	73-84 (12) 441-447 (7) 471-480 (10)	366
CCNE2, G1/S-specific cyclin-E2	O96020	404 (2/3.7)	16.10	16.58	1-49 (49) 387-404 (18)	1-9 (9) 26-31 (6)	21
CASP9, Caspase-9	P55211	416 (2/3.1)	15.11	13.46	117-133 (17) 294-321 (28)	71-77 (7) 281-286 (6)	59
CTNB1, Catenin beta-1	P35222	781 (8/14.7)	14.98	18.18	17-27 (11) 30-59 (30) 547-557 (11) 709-724 (16) 727-773 (47)	1-14 (14) 22-38 (17) 60-75 (16) 96-101 (6) 133-140 (8) 695-702 (8) 719-739 (12) 748-781 (34)	452
FGFR2, fibroblast growth factor receptor 2	P21802	821 (4/4.3)	14.25	14.01	31-45 (15) 142-156 (15) 405-414 (10) 429-450 (22) 776-821 (46)	117-126 (10) 171-178 (8) 488-498 (11) 756-761 (6)	92
IGF1R, insulin-like growth factor 1 receptor	P08069	1,367 (8/6.6)	14.05	11.85	1-10 (10) 466-483 (18) 737-768 (32) 1284-1367 (84)	521-527 (7) 720-732 (13) 793-803 (11) 815-823 (9) 862-867 (6) 1308-1316 (9)	112

						1330-1352 (23) 1356-1367 (12)	
PDGFRB; platelet derived growth factor receptor beta	P09619	1,106 (5/8.5)	13.47	13.56	66-78 (13) 422-438 (15) 463-477 (15) 701-712 (12) 1019-1106 (88)	447-452 (6) 1003-1013 (11) 1017-1027 (11) 1035-1066 (32) 1073-1106 (34)	133
PDGFD, platelet-derived growth factor subunit D	Q9GZP0	370 (0/0)	11.89	5.95	190-200 (11)	N.P.	20
MAPK3, Mitogen-activated protein kinase 3	P27361	379 (0/0)	11.87	9.23	1-27 (27)	N.P.	276
PDGFRA, PDGFR- α (CD140 α), platelet-derived growth factor receptor alpha	P16234	1,089 (4/5.0)	11.11	9.27	61-71 (11) 1017-1089 (73)	133-139 (7) 1013-1025 (13) 1052-1061 (10) 1066-1089 (24)	61
AKT2, RAC-beta serine/threonine-protein kinase	P31751	481 (1/2.5)	11.02	10.81	41-50 (10) 113-137 (25)	73-84 (12)	75
INSRR, insulin receptor-related receptor	P14616	1,297 (3/3.3)	10.72	12.10	462-477 (16) 664-696 (33) 732-762 (31) 1074-1087 (14) 1265-1297 (33)	707-730 (24) 799-805 (7) 1253-1264 (12)	4
PIK3CG, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma	P48736	1,102 (1/1.2)	10.71	8.44	21-31 (11) 440-455 (16) 490-510 (21) 528-547 (20)	516-528 (13)	40
EGFR, epidermal growth factor receptor	P00533	1,210 (6/9.7)	9.1	12.4	1024-1038 (15) 1070-1082 (13) 1094-1139 (46) 1157-1181 (25) 1201-1210 (10)	1016-1023 (8) 1048-1069 (22) 1080-1095 (16) 1105-1112 (8) 1132-1180 (49) 1197-1210 (14)	1031
GSTP1, glutathione S-transferase pi 1	P09211	210 (0/0)	9.05	4.29	N.P.	N.P.	61
GRB2, growth factor receptor-bound protein 2	P62993	212 (0/0)	9.0	3.7	N.P.	N.P.	998
MTOR, mechanistic target of rapamycin, serine/ threonine-protein kinase mTOR	P42345	2,549 (4/1.3)	8.9	8.0	1-18 (18) 910-930 (21) 1815-1866 (52) 2437-2491 (55)	19-27 (9) 263-272 (10) 298-304 (7) 1748-1754 (7)	175
HRAS (p21ras), Harvey rat sarcoma viral oncogene homolog	P01112	189 (1/3.2)	8.47	9.52	170-181 (12)	78-83 (6)	165
SRD5A2, 3-oxo-5-alpha-steroid 4-dehydrogenase 2	P31213	254 (0/0)	8.27	3.54	N.P.	N.P.	N.P.
MAPK1, mitogen-activated protein kinase 1	P28482	360 (0/0)	8.1	5.3	N.P.	N.P.	329
PK3CD,	O00329	1,044	7.76	6.99	285-313 (29)	N.P.	22

phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta		(0/0)			404-414 (11) 448-461 (14)		
IKKA (CHUK), inhibitor of nuclear factor kappa-B kinase subunit alpha	O15111	745 (2/2.6)	7.38	6.31	1-15 (16)	713-723 (11) 738-745 (8)	186
CDK2, cyclin-dependent kinase 2	P24941	298 (0/0)	7.05	4.36	N.P.	N.P.	710
EGF, pro-epidermal growth factor	P01133	1,207 (4/5.4)	6.96	11.18	1062-1094 (33) 1108-1136 (29) 1168-1207 (40)	1093-1104 (12) 1135-1155 (20) 1157-1172 (16) 1181-1197 (17)	26
KRAS, Kirsten rat sarcoma viral oncogene homolog	P01116	189 (1/3.7)	6.34	1.06	N.P.	78-84 (7)	95
NRAS, neuroblastoma RAS viral (v-ras) oncogene homolog	P01111	189 (0/0)	3.70	5.29	172-178 (7)	N.P.	72
PIK3CB, PI3K- β , phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform	P42338	1,070 (0/0)	3.36	2.06	N.P.	N.P.	53
PIK3CA, PI3K- α , phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform	P42336	1,068 (0/0)	3.00	2.06	N.P.	N.P.	101
PSA, prostate-specific antigen or KLK3, kallikrein related peptidase 3	P07288	261 (0/0)	2.68	1.53	N.P.	N.P.	20

^a N_{AIBS} (A/B) represents the number of potential disorder-based binding sites identified by the ANCHOR algorithm (A) and the percentage of residues involved in disorder-based interactions (B).

^b Content of disordered residues (i.e., residues with the disorder propensity ≥ 0.5) in a protein based on the PONDR-FIT disorder prediction.

^c Content of predicted disordered residues in a protein based on the MobiDB consensus score.

^d Information on long disordered regions (i.e., disordered regions of at least 10 residues) was obtained based on the MobiDB consensus profile.

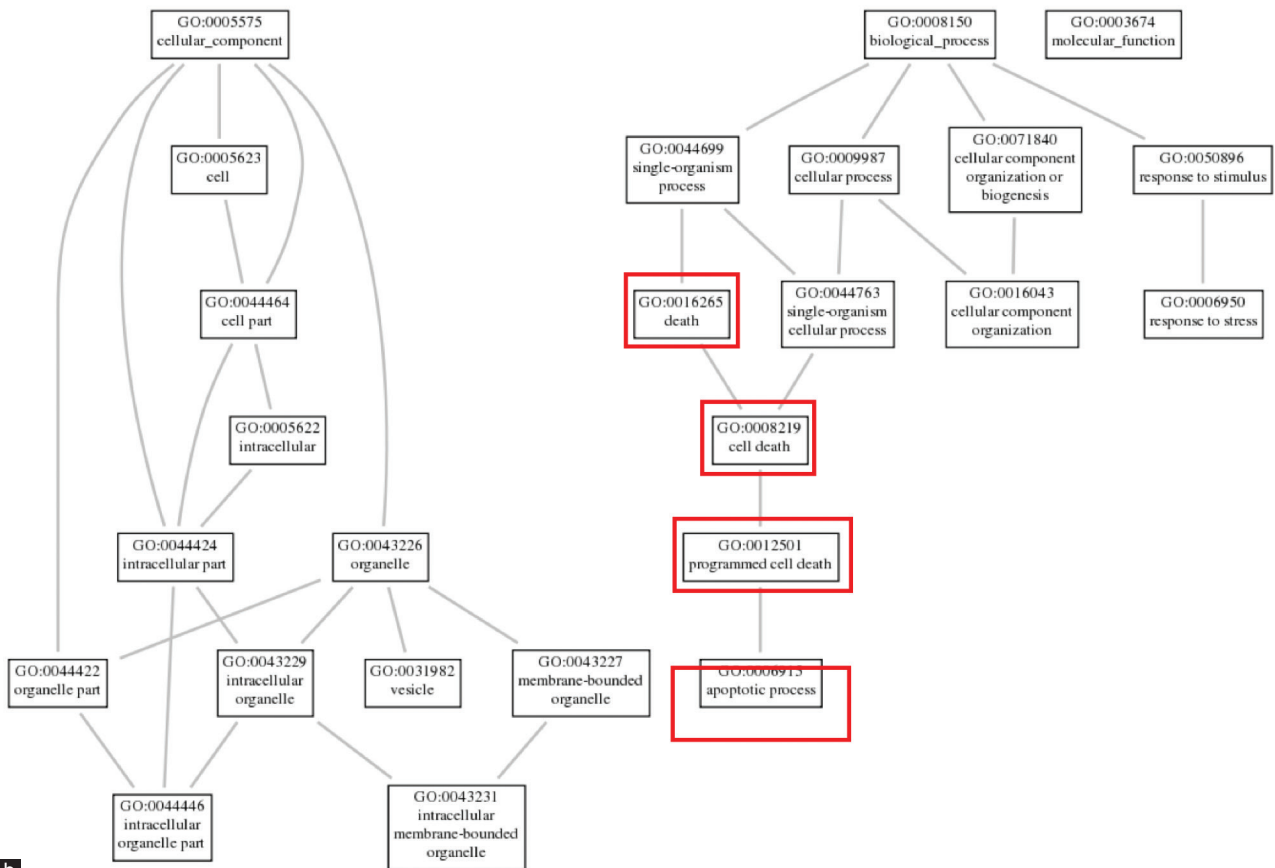
^e AIBSs are potential disorder-based binding sites identified by the ANCHOR algorithm.

^f N_{int}, number of interactions as found using the Agile Protein Interactomes DataServer (<http://cicblade.dep.usal.es:8080/APID/init.action>).

^g N.P., not present. This annotation is used for proteins that do not have predicted AIBSs or long disordered regions.



a



b

Supplementary Figure 1: Functions of intrinsically disordered proteins related to prostate cancer. **(a)** REVIGO analysis of the cellular process GO terms with $P < 0.1$. **(b)** AmiGo2 result for all GO terms with P value less than 0.1. Highlighted are the programmed cell death-related cellular process GO terms.