The human RBPome: from genes and proteins to human disease

Yaseswini Neelamraju[1], Seyedsasan Hashemikhabir[1], Sarath Chandra Janga[1, 2, 3,*]

[1]Department of Biohealth Informatics, School of Informatics and Computing, Indiana University Purdue University, 719 Indiana Ave Ste 319, Walker Plaza Building, Indianapolis, Indiana 46202

[2]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 5021 Health Information and Translational Sciences (HITS), 410 West 10th Street, Indianapolis, Indiana, 46202

[3]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Medical Research and Library Building, 975 West Walnut Street, Indianapolis, Indiana, 46202

[*]*Correspondence should be addressed to :*

Sarath Chandra Janga (scjanga@iupui.edu)

Tel: +1-317-278-4147, Fax: +1-317-278-9201

Manuscript details:

Abstract:230

Figures: 5

Supplementary Material: 5 Tables and 3 Figures

Running title: Characteristics of an atlas of mRNA-binding proteins in the human genome

Significance

RNA Binding proteins (RBPs) play a central role in mediating post transcriptional regulation of genes. However less is understood about them and their regulatory mechanisms. In the present study, we present an analysis of 1344 human RBPs identified from recent experimental studies. We analyse their domain architecture, intrinsic disorder state, evolutionary conservation, protein expression across tissues and disease associations. This study should form a foundation for elucidation and discovery of the functions of RBPs and the cellular regulatory networks they control.

Abstract

RNA Binding Proteins (RBPs) play a central role in mediating post transcriptional regulation of genes. However less is understood about them and their regulatory mechanisms. In this study, we construct a catalogue of 1344 experimentally confirmed RBPs. The domain architecture of RBPs enabled us to classify them into three groups - Classical (29%), Non-classical (19%) and Unclassified (52%). A higher percentage of proteins with unclassified domains reveals the presence of various uncharacterised motifs that can potentially bind RNA. RBPs were found to be highly disordered compared to non-RBPs ($p<2.2e-16$, Fisher's exact test), suggestive of a dynamic regulatory role of RBPs in cellular signalling and homeostasis. Evolutionary analysis in 62 different species showed that RBPs are highly conserved compared to non-RBPs ($p<2.2e-16$, Wilcox-test), reflecting the conservation of various biological processes like mRNA splicing and ribosome biogenesis. The expression patterns of RBPs from human proteome map revealed that ~40% of them are ubiquitously expressed and ~60% are tissue-specific. RBPs were also seen to be highly associated with several neurological disorders, cancer and inflammatory diseases. Anatomical contexts like B cells, T-cells, fetal liver and fetal brain were found to be strongly enriched for RBPs, implying a prominent role of RBPs in immune responses and different developmental stages. The catalogue and meta-analysis presented here should form a foundation for furthering our understanding of RBPs and the cellular networks they control, in years to come.

Introduction

Cellular processes are controlled by several genes; expression of which is regulated at different levels by various cellular entities. In eukaryotes, gene regulation is a complex multilevel process comprising of -- transcriptional, post-transcriptional and post-translational control. Although the regulation at transcriptional and post-translational levels is increasingly being understood, protein machinery and the mechanisms underlying the post-transcriptional regulation still remain to be elucidated. One of the pivotal players that are responsible for post-transcriptional regulatory control in eukaryotic organisms are the RNA binding proteins (RBPs). RBPs bind single or double stranded RNA and determine their fate from synthesis to decay [1-5]. They possess one or more domains that can recognize RNA in a sequence specific manner and hence conferring different binding affinities and specificities. In addition to these domains that can directly bind RNA, RBPs also contain auxiliary domains that mediate interactions with other proteins [6]. When bound to an mRNA, these proteins control all the major steps of an mRNA's life, including splicing, export, localization, translation and degradation [1, 3, 4]. Due to their multi-functionality, RBPs become the most prominent of the post transcriptional machinery and any alteration in their function can cause deleterious effects which could lead to numerous complex disorders [2, 7, 8]. Hence, it becomes important to understand the structural and functional characteristics of RBPs in humans. Increasing interest in RBPs has led to the development of various experimental protocols like SELEX (systematic evolution of ligands by exponential enrichment) [9, 10], CLIP [11], PAR-CLIP [12, 13], iCLIP [14] and RNA compete [15], to identify the binding specificities of RBPs; thus adding context and dynamics to the regulation of gene expression at post-transcriptional level.

In this study, we construct a catalogue of 1344 genes encoding for RBPs in the human genome (Supplementary Table 1): identified from recent high-throughput screens including the mRNA interactome of proliferating HeLa cells through interactome capture [1], mRNA –bound proteome in the human embryonic kidney cells identified using the photoreactive nucleotide-enhanced UV crosslinking and oligo(dT) purification approach , [16] , proteins with the ability to bind RNA from the RNA compete experiments [17], human orthologs of RBPs identified in the mouse embryonic stem cells through interactome capture [18] and RBPs with known binding specificities manually curated and reported in RBPDB [19] to perform a systematic survey of their domain composition, structural disorder, expression across 23 tissues, evolutionary conservation across 62 species and associated diseases by integrating diverse datasets in the public domain. This allowed us to not only uncover the domain architecture, expression and evolutionary dynamics of RNA-binding proteins but also provide novel insights into their roles in diverse human tissues and disease phenotypes.

Materials and Methods

Dataset of RNA-binding proteins

We construct a catalogue of 1344 genes encoding for RBPs, identified from high-throughput screens by Castello et. al [1] , Baltz et. al [16], Ray et. al [17], human orthologs of RBPs identified in mouse embryonic stem cells by Kwon et. al[18] and RBPs reported in RBPDB[19] (Supplementary Table 1) for analysis in this study. Proteins annotated in ENSEMBL's human genome build which were not identified as RBPs were considered as Non-RBPs.

Annotation of domains for human proteins

We used the ENSEMBL v73 biomart (http://www.ensembl.org) to annotate proteins with their corresponding Pfam domains [20]. The superfamily annotations were obtained from the Superfamily database [21]. Domains in RBPs were categorized as classical and non-classical based on the definitions proposed by Castello et. al [1]. A domain occurring in an RBP which could not be classified as either a classical or non-classical was defined as "Unclassified". The list of domains and superfamilies annotated for each RBP is shown in Supplementary Table 1. Further, the enrichment of a given superfamily in RBPs was calculated using Fishers exact test taking Non-RBPs as a control set ("*" above a bar plot in Figure 2C indicates a significant p value (<0.05)).

Intrinsic disorder of proteins

Intrinsic structural disorder of proteins was predicted using IUPRED, which predicts disorder on a per-residue basis [22, 23]. The disorder score predicted by IUPRED was normalized by protein length to account for variations in different protein lengths when comparing predictions for various protein sets. A protein with a normalized score higher than 0.5 was considered to be disordered; this resulted in 30% of RBPs being highly disordered (Supplementary Table 2). To test if RBPs are highly disordered than the rest of the proteome, we calculated an enrichment for disorder in RBPs taking Non-RBPs as a control.

Dataset of orthologs and evolutionary conservation

We identified one-to-one, one-to-many and many-to many orthologs of RBPs across 62 different species from ENSEMBL (v73) (http://www.ensembl.org) (Supplementary Table 3). A hierarchical complete linkage clustering of the data was performed to identify groups of species and genes exhibiting common patterns of evolutionary conservation. Furthermore, the genes were binned into three classes based on their conservation level - high, medium and low. Genes with orthologs in more than 80% of the species were considered to be highly conserved. The next level of conservation is medium which comprises of genes with orthologs in more than 50% but less than 80% of the species. The low conservation class is comprised of genes having orthologs in less than 50% of the species.

Tissue-wide expression profiles

We examined the expression patterns of the RBP catalogue across 17 adult tissues and 6 hematopoietic cells from the recently published human proteome catalogue [24]. Further, tissue specificity of the expression pattern was estimated using an index that varies between zero (for housekeeping genes) and one (observed for tissue specific genes)[25, 26]. We then calculated the tissue specificity scores for 17 well known housekeeping genes [27] and these resulted in an average tissue specificity score of 0.63. Based on this data, we termed genes with a tissue specificity score of at least 0.7 as tissue specific and those below 0.7 were termed ubiquitous. The difference between the expression levels of classical and non-classical RBPs was estimated using the Wilcox test.

Disease Associations

Disease annotations for RBPs were obtained from Malacards [28]. Based on these annotations, diseases enriched for RBPs were identified through hypergeometric probability considering Non-RBPs as a background. Upon filtering at $p < 0.05$ (Corrected p-value by Benjamini-Hochberg (BH) method), we identified 165 diseases to be significantly enriched for their associations with RBPs. Further, the anatomical contexts annotated for these 165 diseases were obtained from Malacards. The enrichment calculated as the odds ratio and corresponding p-value for each anatomical context was computed using the Fisher's exact test.

Results and Discussion

Majority of the experimentally confirmed RBPs have uncharacterized RNA recognition domains

RNA binding proteins associate with nascent RNA to aid in processing, export, transport and localization. The heterogeneity in the functions of these proteins is due to the presence of different RNA binding domains that recognize RNA. RBPs are built with multiple copies of a unique domain or a mosaic of different domains that confer specificity and affinity [29]. Hence, identifying the presence of various domains can provide clues to novel functions of RBPs. We annotated the RBP catalogue with existing Pfam [20] and Superfamily [21] definitions (Materials and Methods, Supplementary Table 1). Depending on the type of domain associated with an RBP and definitions provided by Castello et al [1], we categorized these proteins into “classical” and “non-classical”. Proteins which could not be categorised into either of the classes were termed ‘Unclassified’ in the present work. Using this classification, 29% constituted classical proteins, 19% non-classical and the remaining 52% formed the unclassified group (Figure 1A). The domain distribution of this catalogue (Figure 1B) illustrates the presence of various known and well-studied RNA binding domains like RRM (RNA recognition motif) [30], K-homology (KH) [31, 32], DEAD/DEAH box [33], dsrm [34], WD40 [35] in addition to unclassified domains like the Pkinase domain of EIF2AK2 that auto phosphorylates upon binding of RNA to the dsRBD domain of EIF2AK2 [36, 37]. Also notable is the Calponin homology (CH) domain found in both cytoskeletal and signalling proteins [38]. Further, superfamilies like the RNA binding domain, KH domain type-1, CCCH zinc finger were found to be enriched in RBPs when compared to the Non-RBPs ($p<2.2e-16$, Fisher’s exact test) (Figure 1C). Interestingly, superfamilies like the P-loop containing nucleoside triphosphate hydrolase (P-loop NTH) (p-value<0.05, Fisher’s exact test), S-adenosyl,L-methionine dependent methyltransferases (p-value=9.52e-11, Fisher’s exact test), Spectrin repeat (p-value=3.92e-15, Fisher’s exact test) were found to be highly enriched in our RBP catalogue suggesting novel functions. For example, SRP54, a signal recognition peptide of the P-loop NTH superfamily was shown to play an important role in the splicing of tau gene [39]. Another example from the Spectrin repeat superfamily is RRBP1, a membrane protein of the Endoplasmic reticulum(ER) that plays a role in the ER proliferation, mediating ER-microtubule interactions and enhancing the association of certain mRNA to ER [40]. Additionally, superfamilies RBD, KH-domain type1 and CCCH zinc finger were found to be highly populated with classical proteins whereas WD40 repeat like and beta-beta-alpha zinc finger (BBA-Zinc finger), ubiquitin-like superfamilies were enriched with non-classical proteins (Figure 1D). These analyses together suggests that in addition to binding to the RNA molecules, RBPs could also function as an integral part in maintaining cellular integrity and architecture.

RBPs exhibit significant intrinsic disorder and are enriched among the hubs in protein interaction networks

Intrinsically disordered proteins or natively disordered proteins lack a stable secondary and/or tertiary structure either completely or in part. They are often observed to be playing a major role in signalling, control and regulation, where interaction with more than one protein becomes necessary. Intrinsically disordered proteins are characterised by a structural feature called “Intrinsic disorder” that enables them to participate in varied cellular functions [41, 42]. RBPs being diverse structurally and functionally, are known to be highly disordered [43, 44]. Intrinsic structural disorder of the RBPs was predicted using IUPRED, which predicts disorder on a per-residue basis [22, 23] (Materials and Methods). The disorder score predicted by IUPRED was normalized by protein length to account for variations in different protein lengths when comparing predictions for entire dataset. A protein with a normalized score higher than 0.5 was considered to be disordered; this resulted in 30% of RBPs being highly disordered (Supplementary Table 2). To test if RBPs are highly disordered than the rest of the proteome, we calculated an enrichment of disordered proteins in RBPs taking Non-RBPs as a control. This suggested that RBPs are highly unstructured when compared to the Non-RBPs ($p<2.2e-16$; Fisher’s exact test) (Figure 2A). Although RBPs were found to be significantly enriched for disorder compared to the non-RBPs, we also observed that the proportion of RBPs which are ordered is significantly higher than the disordered RBPs. Additionally, classical proteins were observed to be significantly disordered compared to the non-classical proteins (p-value < 0.001, Fishers exact test) (Figure 2B). Proteins, irrespective of whether ordered or disordered do not function in isolation; instead, they interact with other proteins or cofactors to perform a biological function. Thus, resulting in protein-protein interaction networks, the study of which was enabled by the advent of high-throughput technologies [45]. Advances in the field of structural biology of proteins and protein-protein interactions enabled researchers to put these networks in the context of protein 3D structure. Studies show that the network properties of disordered proteins is different from those which are ordered. Intrinsically disordered proteins are known to be highly interacting due to their unstable 3D structure and hence form hubs in their protein-protein interaction networks [44, 46, 47]. This property of disordered proteins being hubs in their protein-protein interaction network has been analysed in our current study. The protein-protein interaction network for proteins encoded in the human genome was constructed using the annotations available in BIOGRID (3.2.106) [48]. This resulted in a network of 14,897 proteins yielding 1,27,586 interactions with 1095 hubs. Hubs were defined as nodes which interact with at least 50 proteins. On comparison, it was observed that ~30% of the hubs constitute RBPs, of which ~25% are disordered. Also, the disordered hub proteins were seen to be enriched for RBPs when compared to the Non-RBPs (p = 0.0003, Fisher's exact test). The PPI network revealed that the hubs in the network are enriched for RBPs (p<2.2e-16, Wilcox test). These observations imply a highly connected and dynamic role of RBPs via the formation of RNP complexes in the regulation of cellular events.

RBPs are highly conserved across species

Genes and their functionality can vary across species. In particular, regulatory processes are subject to change during the course of evolution and could be a major basis of phenotypic diversity and evolutionary adaptation [49]. Therefore, we aimed to study the conservation of genes coding for RBPs across different species to gain insight into their regulatory functions. For all the RBPs, we identified one-to-one, one-many and many-to many orthologs across 62 different species from ENSEMBL(v73) (http://www.ensembl.org) (Materials and Methods, Supplementary Table 3). A hierarchical complete linkage clustering of the data (Figure 3A) revealed that 95% of the RBPs have orthologs in at least 50% of the species reflecting extensive conservation of various post-transcriptional processes like RNA splicing and ribosome biogenesis. This was especially evident because RBPs were significantly highly conserved compared to the rest of the genome (Median conservation 55 vs 0 species, p<2.2E-16, Wilcox test). We further binned RBPs into three classes based on their conservation level - high, medium and low. RBPs with orthologs in more than 80% of the species have high conservation level (Figure 3B shows selected set from this class). These include members of the PUM family (PUM1 and PUM2), which are a highly conserved family of eukaryotic RBPs [50] . Other examples of highly conserved RBPs include ELAVL2 – an RBP which has an important and evolutionary conserved role in embryogenesis [51], ADARs (ADAR1, ADAR2) - family of RNA editing enzymes, RBM19 [52] - a nucleolar protein that regulates ribosome biogenesis. The next level of conservation is medium (Figure 3C) which comprises of RBPs with orthologs in more than 50% but less than 80% of the species. Genes in this class include MECP2 - a protein important in the function of nerve cells, which is found to be conserved in Mammals, Primates and Vertebrates and ZFP36 - an important player in inflammatory responses, conserved in most species except aves (Chicken, Flycatcher, Turkey, Zebra finch, Duck) [53]. The low conservation class (Figure 3D) comprised of genes having orthologs in less than 50% of the species. Genes in this group included DCD- an antimicrobial peptide coding gene, member of the APOBEC family which are specific to primates and mammals [54]. Furthermore, nLRP11 – a member of the Nod-like receptor protein family was observed to be present only in mammals [55]. Analysis of the variations in the extent of conservation between classical and non-classical RBPs across species showed that there is no significant difference (p=0.1, Fishers exact test)(Figure 3E). These observations highlight that RBPs are evolutionarily highly conserved, with classical and non-classical RBPs exhibiting no significant difference in their evolutionary trajectories.

Additional analysis to study the relation between conservation levels of RBPs and their intrinsic disorder indicated that while majority (71%) of the highly conserved RBPs are highly ordered we found that 29% of the highly conserved RBPs (See Supplementary Figure 1) were found to be significantly disordered (versus 14% for non-RBPs, Odds Ratio = 2.6, p< 2.2e-16, Fishers exact test) suggesting that even the highly conserved RBPs are significantly over-represented for structural disorder.

Majority of the RBPs exhibit tissue-specific protein expression levels

We examined the expression patterns of RBPs across 17 adult tissues and 6 hematopoietic cells from recently published human proteome catalogue [24]. A complete linkage hierarchical clustering of the expression levels (Figure 4A) groups tissues exhibiting similar levels of expression. Further, the tissue specificity of the expression pattern was estimated using an index [25, 26] that varies between zero (for housekeeping genes) and one (observed for tissue specific genes). Genes that resulted in an index of 0.7 or higher were termed tissue-specific and the rest of them were classified as ubiquitous (Materials and Methods, Supplementary Table 4). The ubiquitous category (Figure 4B shows selected set) comprised 40% of the RBP catalogue and included several well-known proteins such as the polypyimidine tract binding protein PTBP1 [56], polyadenylate binding nuclear protein PABPN1 [57], member of the 14-3-3 family YWHAE [58] and Decorin(DCN), a proteoglycan important in collagen fibrillogenesis [59]. In addition to these proteins, the ubiquitous category is highly enriched with the components of the spliceosome (SRSF9, SRSF2, U2AF1, HNRNPL) and proteasome (UBC, UBE2I, UBE2D3). Majority of the RBPs (60%) constituted tissue-specific category (Figure 4C) and included heat response protein 12, HRSP12, known to be expressed in kidney and liver [60], DAZL - a germ cell specific RBP [61], member of the CELF family CELF3 that is highly expressed in brain [62], PUF60-poly-U binding splicing factor 60kDa, that was recently shown to be required for the splicing of a subset of tissue-specific splicing events which when deregulated in the absence of PUF60 affect the development of organs such as brain, heart, kidney and eye [63]. Analysis of the comparison of expression patterns of classical and non-classical RBPs in each of the 16 tissues revealed that in ~50% of the tissues, the expression of non-classical RBPs was significantly higher than the classical RBPs ($p<0.05$, Wilcoxon test) (Figure 4D). While this relative higher expression of non-classical RBPs could be largely attributed to the presence of ribosomal proteins in the non-classical group, these observations suggest the prominent role likely played by both groups of proteins in diverse tissues in post-transcriptional regulatory control (See Supplementary Table 4).

RBPs are significantly associated with inflammatory diseases and immune responses

Aberrant expression of RBPs is associated with several disorders including cancer and neurodegenerative diseases [5, 64]. So in order to better understand the disease associations of RBPs, we obtained diseases annotations for RBPs from Malacards database [28] (Materials and Methods). We evaluated the enrichment of RBPs in various disorders to identify disorders that are highly associated with RBPs. Of all the disorders annotated for human genes in malacards, 165 were found to be highly enriched for RBPs ($p<1e-05$, FDR<1% and number of annotated RBPs >10) which included all major types of cancers -breast, lung, prostate and liver as well as neurodegenerative diseases- Parkinson's disease and down syndrome (Figure 5A). For example, genes like ELAV1, which regulate mRNA stability are known to contribute to breast cancer [65], RBM5, a tumor suppressor gene is known to control cell growth in lung cancer [66], UPF1, subunit of the post splicing multi protein complex is shown to be dysregulated in prostate cancer [67]. RBPs that are known to be dysregulated in neurodegenerative diseases include members of the NOVA family [68], QKI, a candidate gene for schizophrenia [69] and ELAVL4, an important player in parkinson's disease [70]. In addition to these disorders, RBPs were enriched in various inflammatory diseases such as neuronitis, prostatitis, esophagitis suggesting an important role for RBPs in mediating inflammatory responses. For example, a cold inducible RNA binding protein (CIRBP) that triggers inflammatory responses in hemarragic shock and sepsis was observed to be associated with endothelitis and hypoxia [71]. These results suggest that RBPs are not only implicated in cancers and neurodegenerative diseases but also play an important role in mediating various immunological responses. The anatomical contexts associated with the 165 diseases were studied to get an insight into the cells/tissues that are majorly affected by an abnormal expression of RBPs. Anatomical contexts like B cells, T cells , monocytes, fetal liver ,fetal brain were observed to be enriched($p<0.001$, Hypergeometric test) (Figure 5B). For example, ELAVL1 is known to regulate various gene expression programs during the embryonic development in mouse [72]. Recently, the phenomenon of intron retention was observed at a high level in T cells with increased expression of hnRNPL, whereas the introns were efficiently spliced out in cells that had normal/less expressed hnRNPL [73]. Additionally, another independent study identified a RNA Binding domain in the thyroid receptor and these receptors are single stranded RNA Binding Proteins [74]. All these observations further emphasize the under-appreciated role of RBPs in mediating inflammatory and immune responses. Further, we have also analysed the proportion of each class of RBPs in each of the enriched diseases and identified that diseases like Pancreatitis and Endotheliitis have higher proportion of non-classical RBPs (See Supplementary Table 5 and Supplementary Figure 2). This observation suggests an increased prevalence of non-classical RBPs in some inflammatory diseases. In addition, we also calculated the proportion of RBPs that are ubiquitously expressed or tissue specific in their expression patterns, for each enriched disease. This analysis revealed 55 enriched diseases - one-third of the total diseases significantly enriched for RBPs, to have a significant proportion of ubiquitously expressed RBPs ($p < 0.01$, Fishers exact test, Supplementary Figure 3). We also found that only 16 diseases exhibited a prevalence for tissue-specific RBPs compared to ubiquitous ones. Since many of the 165 diseases enriched for RBPs are complex multigenic disorders, these results support the notion that majority of these diseases are enriched with RBPs which are expressed in multiple tissues and hence their observed phenotypic contributions might extend beyond the tissue of origin.

Conclusions

RNA Binding proteins are major players mediating the post-transcriptional regulation and dearth in data has limited our understanding on their regulatory mechanisms and interactions for several years. Recent advances in various experimental methods has led to an expansion of the RBPome and thus unravelling the RNA binding ability of several proteins. In this study, we present a meta-analysis of the RBP catalogue to study their domain architecture, protein structural disorder, tissue wide expression, evolutionary conservation and their role in the disease context. We identified several RBPs to have domains that are not known for their role in facilitating the function of an RBP (termed as “Unclassified” in the present study). This analysis would broaden the scope for researchers in the field of proteomics to explore the unknown functions of these domains in mediating post-transcriptional regulatory control. For instance, several genes currently annotated as bonafide kinases can also function as RNA-binding proteins, so it remains unclear how their interplay between post-translational and post-transcriptional regulation would drive specific cellular signalling events. It is also unclear whether such signalling proteins would also form RNP complexes as do the canonical RNA-binding proteins. Our results also show that RBPs are highly unstructured when compared to the non-RBPs and are over-represented to be occurring as the hubs of the protein-protein interaction networks, suggesting a likely occurrence of their dynamic RNP complexes in the cell. Analysis of the extent of conservation of RBPs clearly revealed that RBPs are preserved across majority of the species studied here suggestive of a wider conservation of post-transcriptional processes. Our data also supports that while majority of the highly conserved RBPs are highly ordered, we found that even the highly conserved RBPs are significantly over-represented for structural disorder compared to non-RBPs indicating the importance of their disorder in the formation and maintenance of RNPs. Since RBPs predominantly function by forming RNP complexes it would be interesting to dissect their interplay with other protein partners in various cell types such as immune cells, where currently there is limited understanding of either their RNP complexes or post-transcriptional regulatory networks they govern. Our analysis of expression data revealed that RBPs are largely tissue specific in their protein levels. Disease association analysis of RBPs showed a clear enrichment for their association with several inflammatory diseases and immune responses among other complex disease phenotypes such as cancer and neurological disorders. RBPs associated with these complex disorders were also found to be over-represented for ubiquitously expressed RBPs. Together, this analyses uncovers several characteristics of ~1300 RBPs which would help researchers in the field of proteomics to develop strategies that can target specific RBPs of interest to gain further insights into their function in specific tissues and disease states.

Majority of the RNA binding proteins in this study are unclassified and uncharacterised for their protein domains which bind to RNA or mediate RNA binding. So future computational and experimental efforts should be able to uncover the specific protein domains which can bind and recognize the RNA species in the cell. Our results also suggest that RBPs are highly unstructured when compared to the rest of the proteome implying an active role of RBPs in regulating cellular events. This study also expands the existing knowledge on tissue expression patterns of RBPs, their evolutionary conservation and disease associations. Altogether, providing a snapshot of the characteristics of RNA Binding proteins and will be a valuable resource to unravel many unconventional, novel functions and dynamics of RNP complexes in the context of post-transcriptional regulatory networks.

Acknowledgements

The authors would also like to thank members of the Janga Lab for critically reading a previous version of the manuscript and providing helpful comments. The authors report no financial or other conflict of interest relevant to the subject of this article.

Conflict of Interest

The authors have no competing interests to declare.

Funding

SCJ acknowledges support from the School of Informatics and Computing at IUPUI in the form of start-up funds.

References

[1] Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell. 2012;149:1393-406.
[2] Kechavarzi B, Janga SC. Dissecting the expression landscape of RNA-binding proteins in human cancers. Genome biology. 2014;15:R14.
[3] Mittal N, Roy N, Babu MM, Janga SC. Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. Proceedings of the National Academy of Sciences of the United States of America. 2009;106:20300-5.
[4] Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. FEBS letters. 2008;582:1977-86.
[5] Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. Trends in genetics : TIG. 2008;24:416-25.
[6] Mitchell SF, Parker R. Principles and properties of eukaryotic mRNPs. Molecular cell. 2014;54:547-58.
[7] Castello A, Fischer B, Hentze MW, Preiss T. RNA-binding proteins in Mendelian disease. Trends in genetics : TIG. 2013;29:318-27.
[8] Lukong KE, Fatimy RE. Implications of RNA-binding Proteins for Human Diseases. eLS: John Wiley & Sons, Ltd; 2001.
[9] Riordan DP, Herschlag D, Brown PO. Identification of RNA recognition elements in the Saccharomyces cerevisiae transcriptome. Nucleic acids research. 2011;39:1501-9.
[10] Galarneau A, Richard S. Target RNA motif and target mRNAs of the Quaking STAR protein. Nature structural & molecular biology. 2005;12:691-8.

[11] Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. Science. 2003;302:1212-5.

[12] Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010;141:129-41.

[13] Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T. Identification of RNA-protein interaction networks using PAR-CLIP. Wiley interdisciplinary reviews RNA. 2012;3:159-77.

[14] Huppertz I, Attig J, D'Ambrogio A, Easton LE, Sibley CR, Sugimoto Y, et al. iCLIP: protein-RNA interactions at nucleotide resolution. Methods. 2014;65:274-87.

[15] Ray D, Kazan H, Chan ET, Pena Castillo L, Chaudhry S, Talukder S, et al. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nature biotechnology. 2009;27:667-70.

[16] Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Molecular cell. 2012;46:674-90.

[17] Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A compendium of RNA-binding motifs for decoding gene regulation. Nature. 2013;499:172-7.

[18] Kwon SC, Yi H, Eichelbaum K, Fohr S, Fischer B, You KT, et al. The RNA-binding protein repertoire of embryonic stem cells. Nature structural & molecular biology. 2013;20:1122-30.

[19] Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. RBPDB: a database of RNA-binding specificities. Nucleic acids research. 2011;39:D301-8.

[20] Sonnhammer EL, Eddy SR, Durbin R. Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins. 1997;28:405-20.

[21] Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Journal of molecular biology. 2001;313:903-19.

[22] Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. Journal of molecular biology. 2005;347:827-39.

[23] Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 2005;21:3433-4.

[24] Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. Nature. 2014;509:575-81.

[25] Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 2014;505:635-40.

[26] Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005;21:650-9.

[27] Lee PD, Sladek R, Greenwood CM, Hudson TJ. Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. Genome research. 2002;12:292-7.

[28] Rappaport N, Nativ N, Stelzer G, Twik M, Guan-Golan Y, Stein TI, et al. MalaCards: an integrated compendium for diseases and their annotation. Database : the journal of biological databases and curation. 2013;2013:bat018.

[29] Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. Nature reviews Molecular cell biology. 2007;8:479-90.

[30] Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. The FEBS journal. 2005;272:2118-31.

[31] Grishin NV. KH domain: one motif, two folds. Nucleic acids research. 2001;29:638-43.

[32] Garcia-Mayoral MF, Hollingworth D, Masino L, Diaz-Moreno I, Kelly G, Gherzi R, et al. The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. Structure. 2007;15:485-98.

[33] Rocak S, Linder P. DEAD-box proteins: the driving forces behind RNA metabolism. Nature reviews Molecular cell biology. 2004;5:232-41.

[34] Stefl R, Skrisovska L, Allain FH. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. EMBO reports. 2005;6:33-8.

[35] Lau CK, Bachorik JL, Dreyfuss G. Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. Nature structural & molecular biology. 2009;16:486-91.

[36] Masliah G, Barraud P, Allain FH. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. Cellular and molecular life sciences : CMLS. 2013;70:1875-95.

[37] Jammi NV, Beal PA. Phosphorylation of the RNA-dependent protein kinase regulates its RNA-binding activity. Nucleic acids research. 2001;29:3020-9.

[38] Castresana J, Saraste M. Does Vav bind to F-actin through a CH domain? FEBS letters. 1995;374:149-51.

[39] Wu JY, Kar A, Kuo D, Yu B, Havlioglu N. SRp54 (SFRS11), a regulator for tau exon 10 alternative splicing identified by an expression cloning strategy. Molecular and cellular biology. 2006;26:6739-47.

[40] Cui XA, Zhang H, Palazzo AF. p180 promotes the ribosome-independent localization of a subset of mRNA to the endoplasmic reticulum. PLoS biology. 2012;10:e1001336.

[41] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annual review of biophysics. 2008;37:215-46.

[42] Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. Current opinion in structural biology. 2011;21:432-40.

[43] Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry. 2002;41:6573-82.

[44] Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, et al. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS computational biology. 2006;2:e100.

[45] Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature reviews Genetics. 2004;5:101-13.

[46] Hsu WL, Oldfield C, Meng J, Huang F, Xue B, Uversky VN, et al. Intrinsic protein disorder and protein-protein interactions. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing. 2012:116-27.

[47] Kim PM, Sboner A, Xia Y, Gerstein M. The role of disorder in interaction networks: a structural analysis. Molecular systems biology. 2008;4:179.

[48] Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic acids research. 2006;34:D535-9.

[49] Kirschner M, Gerhart J. Evolvability. Proceedings of the National Academy of Sciences of the United States of America. 1998;95:8420-7.

[50] Spassov DS, Jurecic R. The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? IUBMB life. 2003;55:359-66.

[51] Wiszniak SE, Dredge BK, Jensen KB. HuB (elavl2) mRNA is restricted to the germ cells by post-transcriptional mechanisms including stabilisation of the message by DAZL. PloS one. 2011;6:e20773.

[52] Kallberg Y, Segerstolpe A, Lackmann F, Persson B, Wieslander L. Evolutionary conservation of the ribosomal biogenesis factor Rbm19/Mrd1: implications for function. PloS one. 2012;7:e43786.

[53] Lai WS, Stumpo DJ, Kennington EA, Burkholder AB, Ward JM, Fargo DL, et al. Life without TTP: apparent absence of an important anti-inflammatory protein in birds. American journal of physiology Regulatory, integrative and comparative physiology. 2013;305:R689-700.

[54] Cullen BR. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. Journal of virology. 2006;80:1067-76.

[55] Tian X, Pascal G, Monget P. Evolution and functional divergence of NLRP genes in mammalian reproductive systems. BMC evolutionary biology. 2009;9:202.

[56] Sawicka K, Bushell M, Spriggs KA, Willis AE. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. Biochemical Society transactions. 2008;36:641-7.

[57] Apponi LH, Corbett AH, Pavlath GK. Control of mRNA stability contributes to low levels of nuclear poly(A) binding protein 1 (PABPN1) in skeletal muscle. Skeletal muscle. 2013;3:23.

[58] Toyo-oka K, Shionoya A, Gambello MJ, Cardoso C, Leventer R, Ward HL, et al. 14-3-3epsilon is important for neuronal migration by binding to NUDEL: a molecular explanation for Miller-Dieker syndrome. Nature genetics. 2003;34:274-85.

[59] Imai K, Hiramatsu A, Fukushima D, Pierschbacher MD, Okada Y. Degradation of decorin by matrix metalloproteinases: identification of the cleavage sites, kinetic analyses and transforming growth factor-beta1 release. The Biochemical journal. 1997;322 (Pt 3):809-14.

[60] Chong CL, Huang SF, Hu CP, Chen YL, Chou HY, Chau GY, et al. Decreased expression of UK114 is related to the differentiation status of human hepatocellular carcinoma. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008;17:535-42.

[61] Kim B, Cooke HJ, Rhee K. DAZL is essential for stress granule formation implicated in germ cell survival upon heat stress. Development. 2012;139:568-78.

[62] Ladd AN, Charlet N, Cooper TA. The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Molecular and cellular biology. 2001;21:1285-96.

[63] Dauber A, Golzio C, Guenot C, Jodelka FM, Kibaek M, Kjaergaard S, et al. SCRIB and PUF60 are primary drivers of the multisystemic phenotypes of the 8q24.3 copy-number variant. American journal of human genetics. 2013;93:798-811.

[64] Wurth L. Versatility of RNA-Binding Proteins in Cancer. Comparative and functional genomics. 2012;2012:178525.

[65] Upadhyay R, Sanduja S, Kaza V, Dixon DA. Genetic polymorphisms in RNA binding proteins contribute to breast cancer survival. International journal of cancer Journal international du cancer. 2013;132:E128-38.

[66] Shao C, Zhao L, Wang K, Xu W, Zhang J, Yang B. The tumor suppressor gene RBM5 inhibits lung adenocarcinoma cell growth and induces apoptosis. World journal of surgical oncology. 2012;10:160.

[67] Yang C, Strobel P, Marx A, Hofmann I. Plakophilin-associated RNA-binding proteins in prostate cancer and their implications in tumor progression and metastasis. Virchows Archiv : an international journal of pathology. 2013;463:379-90.

[68] Musunuru K. Cell-specific RNA-binding proteins in human disease. Trends in cardiovascular medicine. 2003;13:188-95.

[69] Radomska KJ, Halvardson J, Reinius B, Lindholm Carlstrom E, Emilsson L, Feuk L, et al. RNA-binding protein QKI regulates Glial fibrillary acidic protein expression in human astrocytes. Human molecular genetics. 2013;22:1373-82.

[70] Noureddine MA, Qin XJ, Oliveira SA, Skelly TJ, van der Walt J, Hauser MA, et al. Association between the neuron-specific RNA-binding protein ELAVL4 and Parkinson disease. Human genetics. 2005;117:27-33.

[71] Qiang X, Yang WL, Wu R, Zhou M, Jacob A, Dong W, et al. Cold-inducible RNA-binding protein (CIRP) triggers inflammatory responses in hemorrhagic shock and sepsis. Nature medicine. 2013;19:1489-95.

[72] Katsanou V, Milatos S, Yiakouvaki A, Sgantzis N, Kotsoni A, Alexiou M, et al. The RNA-binding protein Elavl1/HuR is essential for placental branching morphogenesis and embryonic development. Molecular and cellular biology. 2009;29:2762-76.

[73] Cho V, Mei Y, Sanny A, Chan S, Enders A, Bertram EM, et al. The RNA-binding protein hnRNPLL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA. Genome biology. 2014;15:R26.

[74] Xu B, Koenig RJ. An RNA-binding domain in the thyroid hormone receptor enhances transcriptional activation. The Journal of biological chemistry. 2004;279:33051-6.

Figure legends

Figure 1: Domain Architecture of RNA Binding Proteins| Pfam domains listed by Castello et al (2012) to define classical and non-classical RBPs were used to annotate the RBP catalogue. Those that could not be classified into either of the categories were termed 'Unclassified'. Figure A shows the overall domain distribution in RNA binding proteins. Figure B shows the distribution of individual pfam domains in RBPs. Figure C shows the distribution of various superfamilies' of RBPs (* against the Superfamily name indicates that the name has been abbreviated). Superfamilies' like P-loop nucleoside triphosphate hydrolases (P-loop NTH), beta-beta-alpha zinc finger (BBA-Zinc finger) were found to be enriched in RBPs when compared to the Non-RBPs (** indicates $p<0.05$, Fisher's exact test). Figure D shows the distribution of classical and non-classical RBPs in each superfamily. In all the figures 1B,1C,1D domains and superfamilies' associated with less than 1% of the RBPs are not shown.

Figure 2: Intrinsic Disorder of RNA Binding Proteins| The extent of disorder was predicted using IUPred. A protein was considered to be disordered if the normalized disorder score is greater than 0.5. Figure A compares the intrinsic disorder between RBPs and Non-RBPs suggesting RBPs to be more disordered than Non-RBPs ($p<2.2e-16$, Fisher's exact test). Figure B compares the intrinsic disorder between the classical and non-classical RBPs showing classical RBPs to be more disordered than the non-classical ($p<2.2e-16$, Fisher's exact test).

Figure 3: Evolutionary conservation of RBPs| A complete linkage hierarchical clustering of RBP orthologs in 62 different species illustrating high conservation levels (Figure A). Based on their conservation levels, RBPs were categorized into three levels - high (conservation in >=80% of the species), medium (conservation in >=50% and <=80% of the species), low (conservation in <=50% of the species). Figure B shows a subset of RBPs that are highly conserved. Figure C, a subset of RBPs which exhibited medium level of conservation. Figure D shows RBPs that are poorly conserved. Figure E compares the extent of conservation in classical and non-classical RBPs suggesting similar conservation patterns in the two groups ($p=0.1$, Fisher's exact test).

Figure 4: Tissue-wide expression patterns of RBPs| Expression of RBPs across 25 different tissues obtained from the human protein catalogue [24]. Complete linkage hierarchical clustering of the expression data is shown in Figure A. Based on the expression patterns, RBPs were classified as ubiquitous and tissue-specific [26]. Figure B and C show a subset of RBPs that are ubiquitous and tissue-specific respectively. Figure D compares the expression levels of classical and non-classical RBPs in each tissue (* indicates $p<0.05$, Wilcoxon test).

Figure 5: Diseases and anatomical contexts associated with RBPs| Diseases enriched for RBPs were obtained by using the annotations available from the malacards database [28] and filtered for at least 10 RBP associations, p-value < 1e-05 and FDR <1%. Among these, top 50 are displayed in the figure (Figure A) and the complete list is available as Supplementary Table 5. Figure 5B shows the top 20 anatomical contexts significantly associated with diseases enriched for RBPs ($p<0.001$, Fishers exact test). A complete list is available as Supplementary Table 5.

Supplementary Tables

Supplementary Table 1: Detailed catalogue of RBPs and their domain architecture

Supplementary Table 2: Intrinsic Disorder of RNA Binding Proteins

Supplementary Table 3: Ortholog Information

Supplementary Table 4: Expression levels of RBPs across tissues

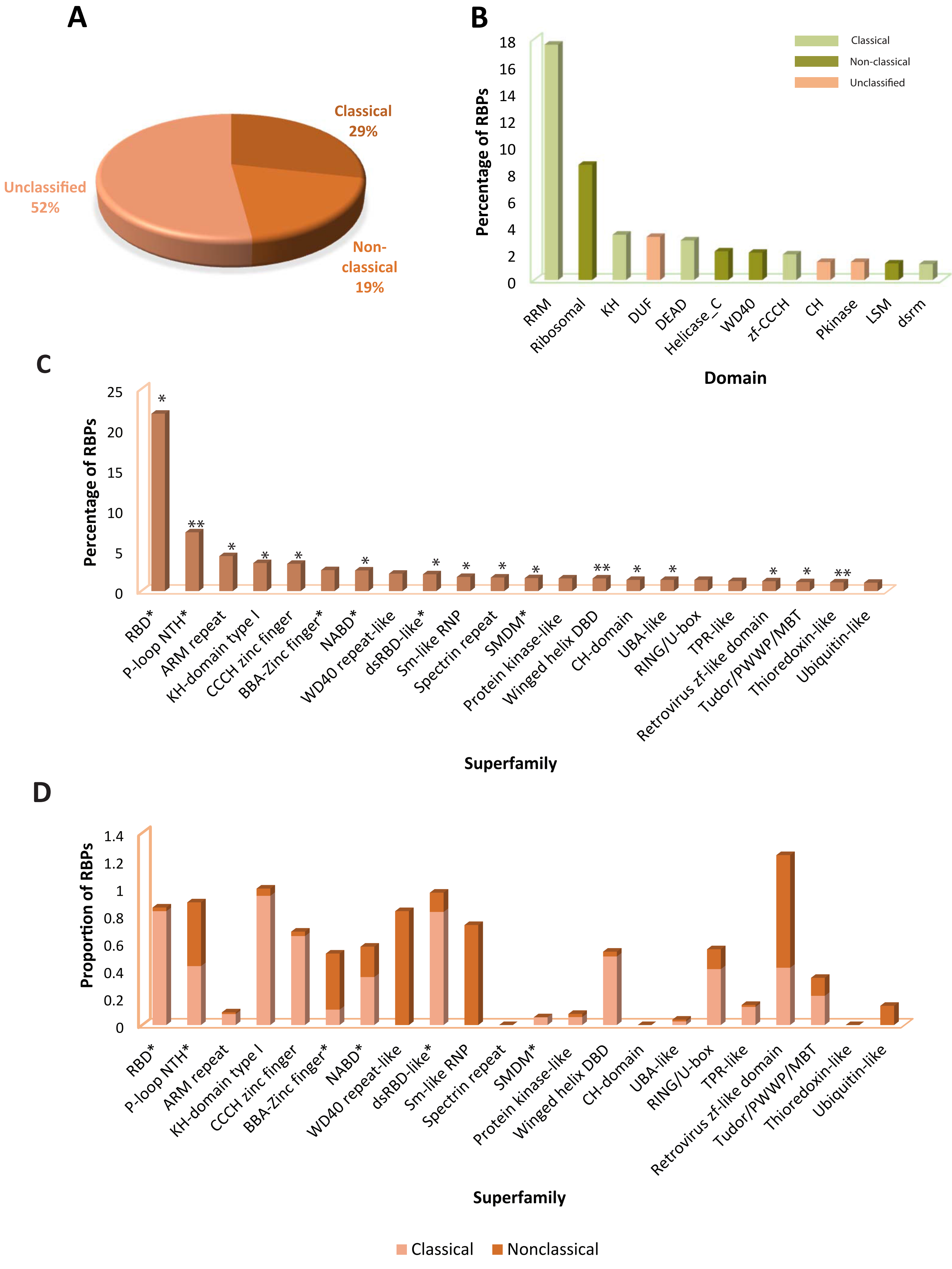
Supplementary Table 5: Disease associations

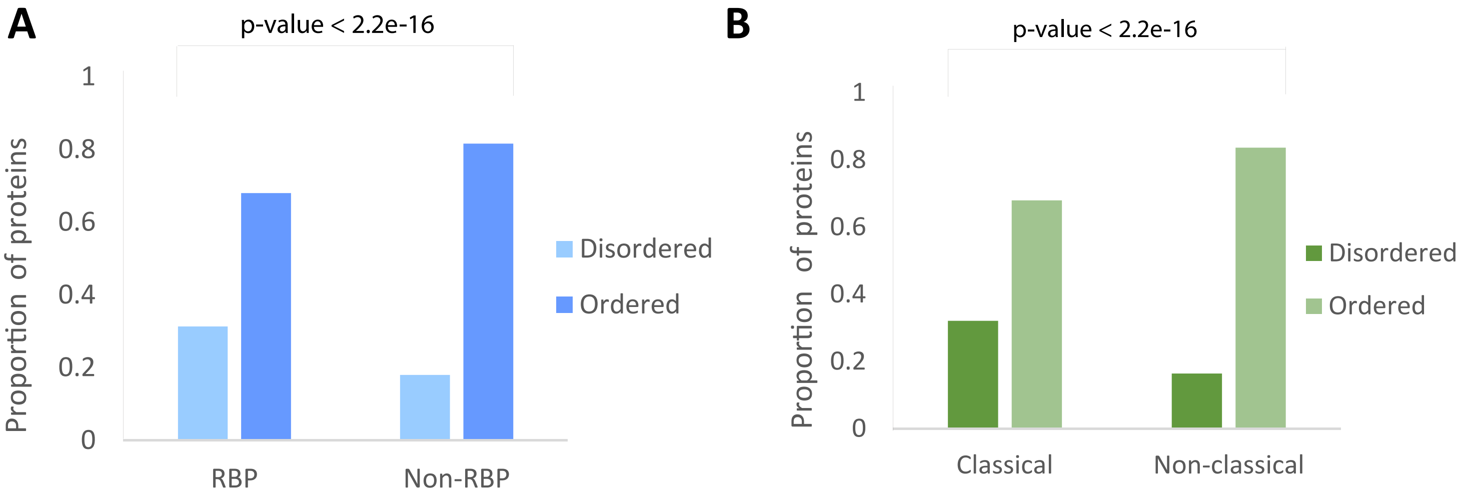
Supplementary Figures

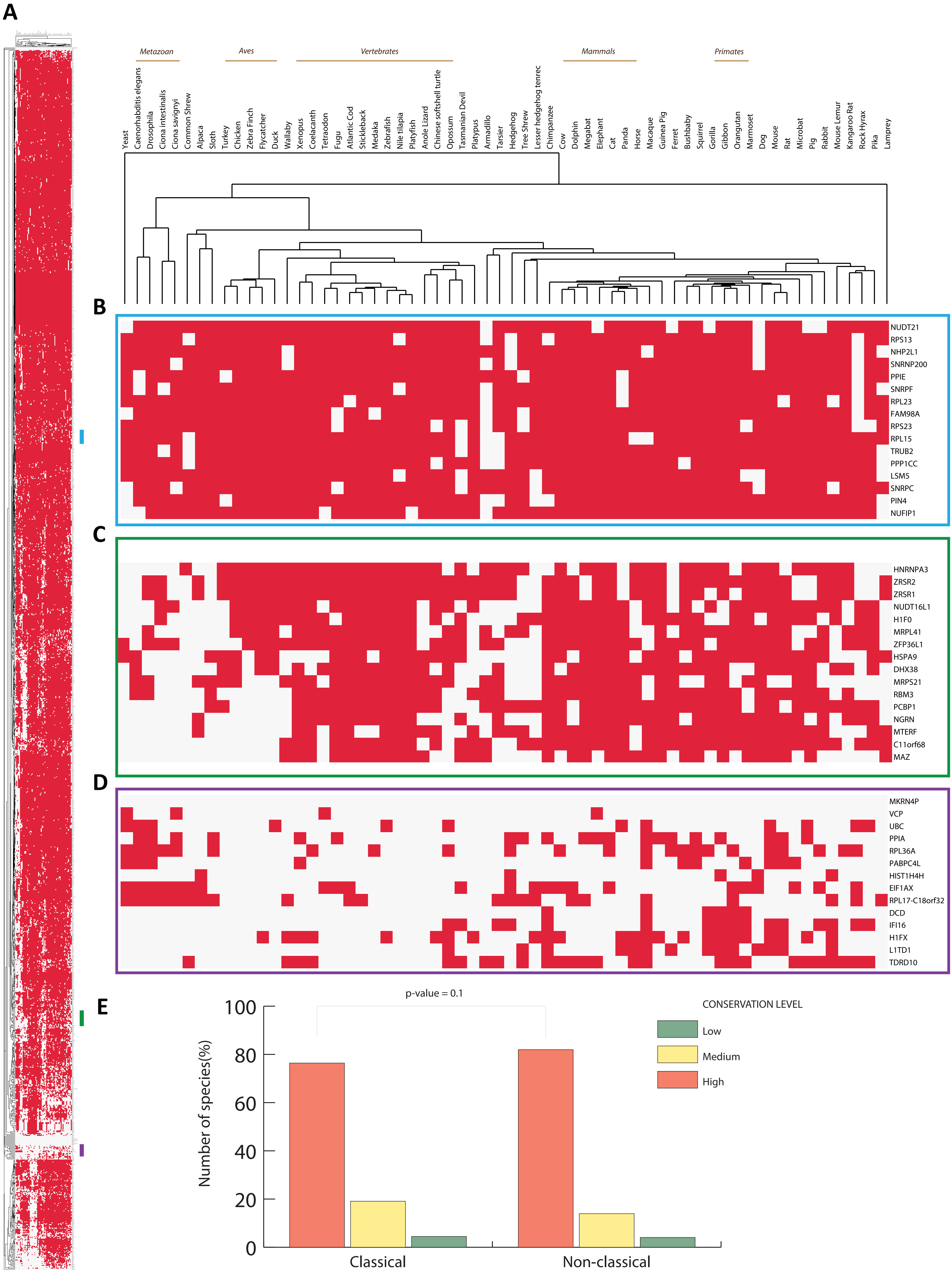
Figure 1: Scatter plot showing the conservation levels vs the disorder score for all the RBPs.

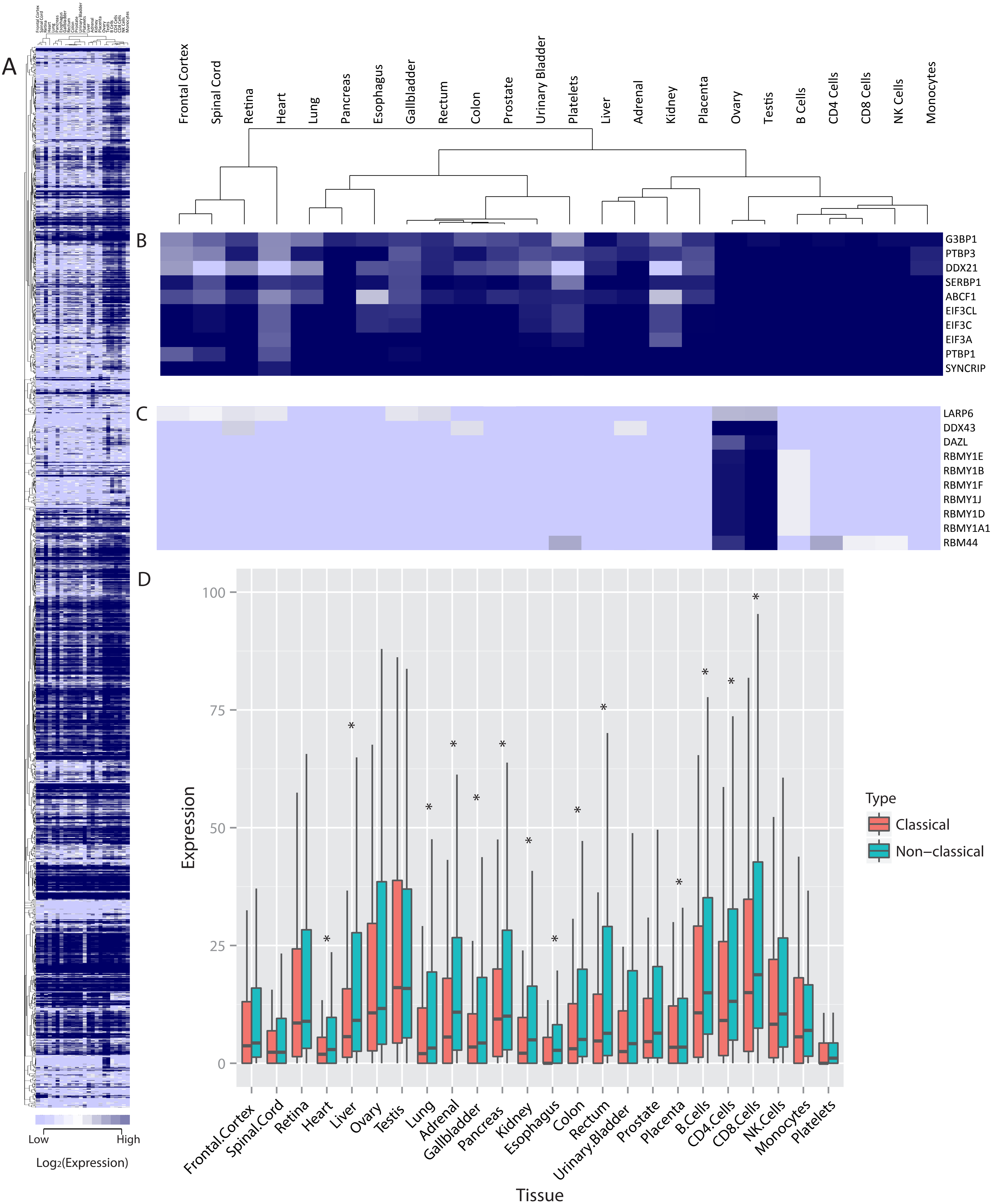
Figure 2: Top 50 enriched diseases with the proportion of Classical, Non-classical and Unclassified RBPs.

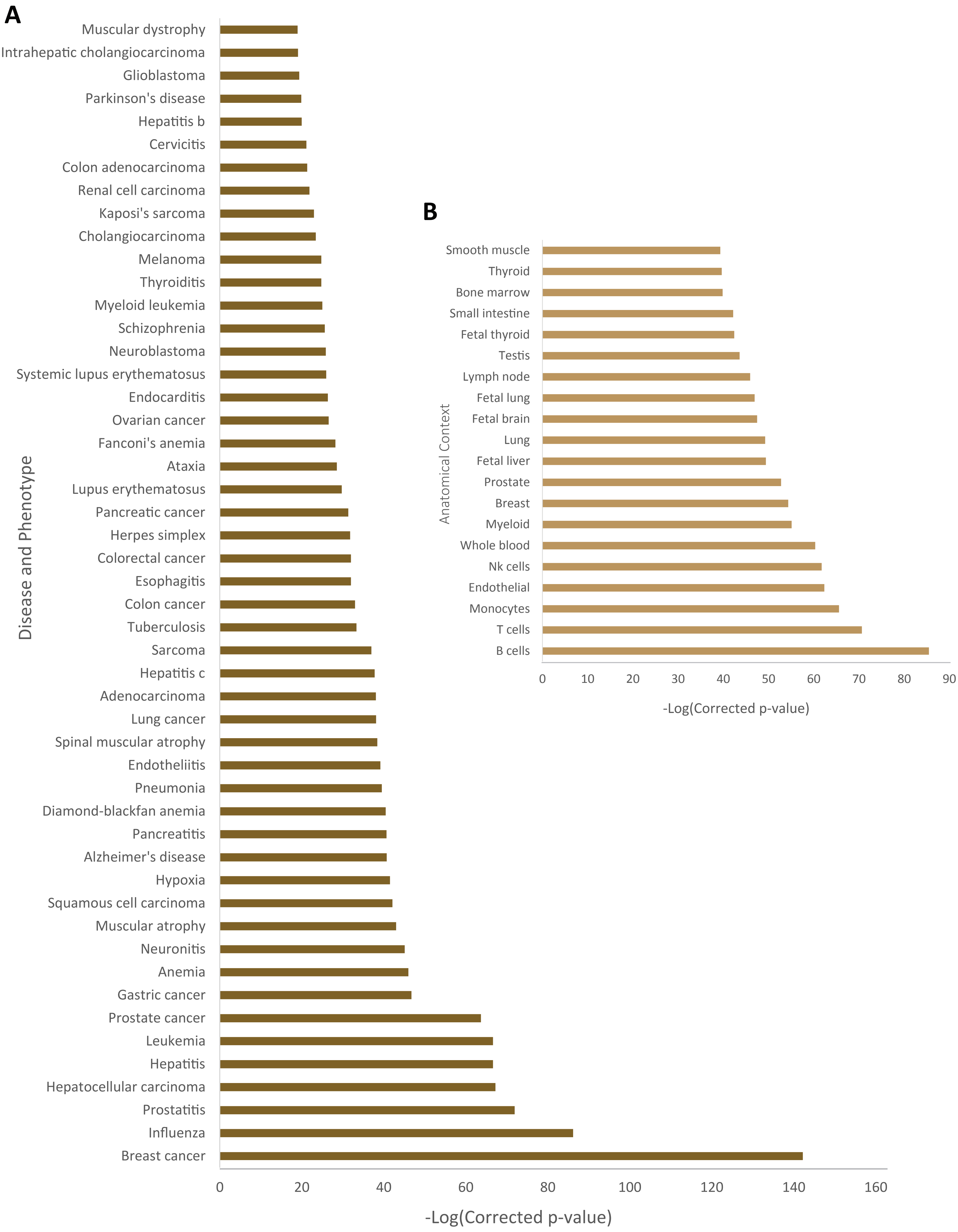
Figure 3: Diseases with significant over-representation of ubiquitous RBPs compared to tissue specific ones ($p < 0.01$, Fishers exact test).


A
Classical 29%
Non-classical 19%
Unclassified 52%
B
Percentage of RBPs
18
16
14
12
10
8
6
4
2
0
Classical
Non-classical
Unclassified
RRM
Ribosomal
KH
DUF
DEAD
Helicase_C
WD40
zf-CCCH
CH
Pkinase
LSM
dsrm
Domain
C
Percentage of RBPs
25
20
15
10
5
0
*
**
*
*
*
*
*
*
*
*
**
*
*
*
*
**
RBD*
P-loop NTH*
ARM repeat
KH-domain type I
CCCH zinc finger
BBA-Zinc finger*
NABD*
WD40 repeat-like
dsRBD-like*
Sm-like RNP
Spectrin repeat
SMDM*
Protein kinase-like
Winged helix DBD
CH-domain
UBA-like
RING/U-box
TPR-like
Retrovirus zf-like domain
Tudor/PWWP/MBT
Thioredoxin-like
Ubiquitin-like
Superfamily
D
Proportion of RBPs
1.4
1.2
1
0.8
0.6
0.4
0.2
0
RBD*
P-loop NTH*
ARM repeat
KH-domain type I
CCCH zinc finger
BBA-Zinc finger*
NABD*
WD40 repeat-like
dsRBD-like*
Sm-like RNP
Spectrin repeat
SMDM*
Protein kinase-like
Winged helix DBD
CH-domain
UBA-like
RING/U-box
TPR-like
Retrovirus zf-like domain
Tudor/PWWP/MBT
Thioredoxin-like
Ubiquitin-like
Superfamily
Classical
Nonclassical


A
p-value < 2.2e-16
Proportion of proteins
1
0.8
0.6
0.4
0.2
0
RBP
Non-RBP
Disordered
Ordered
B
p-value < 2.2e-16
Proportion of proteins
1
0.8
0.6
0.4
0.2
0
Classical
Non-classical
Disordered
Ordered


A
Metazoan
Aves
Vertebrates
Mammals
Primates
Yeast
Caenorhabditis elegans
Drosophila
Ciona intestinalis
Ciona savignyi
Common Shrew
Alpaca
Sloth
Turkey
Chicken
Zebra Finch
Flycatcher
Duck
Wallaby
Xenopus
Coelacanth
Tetraodon
Fugu
Atlantic Cod
Stickleback
Medaka
Zebrafish
Nile tilapia
Platyfish
Anole Lizard
Chinese softshell turtle
Opossum
Tasmanian Devil
Platypus
Armadillo
Tarsier
Hedgehog
Tree Shrew
Lesser hedgehog tenrec
Chimpanzee
Cow
Dolphin
Megabat
Elephant
Cat
Panda
Horse
Macaque
Guinea Pig
Ferret
Bushbaby
Squirrel
Gorilla
Gibbon
Orangutan
Marmoset
Dog
Mouse
Rat
Microbat
Pig
Rabbit
Mouse Lemur
Kangaroo Rat
Rock Hyrax
Pika
Lamprey
B
NUDT21
RPS13
NHP2L1
SNRNP200
PPIE
SNRPF
RPL23
FAM98A
RPS23
RPL15
TRUB2
PPP1CC
LSM5
SNRPC
PIN4
NUFIP1
C
HNRNPA3
ZRSR2
ZRSR1
NUDT16L1
H1F0
MRPL41
ZFP36L1
HSPA9
DHX38
MRPS21
RBM3
PCBP1
NGRN
MTERF
C11orf68
MAZ
D
MKRN4P
VCP
UBC
PPIA
RPL36A
PABPC4L
HIST1H4H
EIF1AX
RPL17-C18orf32
DCD
IFI16
H1FX
L1TD1
TDRD10
E
p-value = 0.1
CONSERVATION LEVEL
Low
Medium
High
Number of species(%)
100
80
60
40
20
0
Classical
Non-classical


A
Frontal Cortex
Spinal Cord
Retina
Heart
Lung
Pancreas
Esophagus
Gallbladder
Rectum
Colon
Prostate
Urinary Bladder
Platelets
Liver
Adrenal
Kidney
Placenta
Ovary
Testis
B Cells
CD4 Cells
CD8 Cells
NK Cells
Monocytes
Low
High
Log2(Expression)
B
G3BP1
PTBP3
DDX21
SERBP1
ABCF1
EIF3CL
EIF3C
EIF3A
PTBP1
SYNCRIP
C
LARP6
DDX43
DAZL
RBMY1E
RBMY1B
RBMY1F
RBMY1J
RBMY1D
RBMY1A1
RBM44
D
100
75
50
25
0
Expression
*
Frontal.Cortex
Spinal.Cord
Retina
Heart
Liver
Ovary
Testis
Lung
Adrenal
Gallbladder
Pancreas
Kidney
Esophagus
Colon
Rectum
Urinary.Bladder
Prostate
Placenta
B.Cells
CD4.Cells
CD8.Cells
NK.Cells
Monocytes
Platelets
Tissue
Type
Classical
Non−classical


A
Muscular dystrophy
Intrahepatic cholangiocarcinoma
Glioblastoma
Parkinson's disease
Hepatitis b
Cervicitis
Colon adenocarcinoma
Renal cell carcinoma
Kaposi's sarcoma
Cholangiocarcinoma
Melanoma
Thyroiditis
Myeloid leukemia
Schizophrenia
Neuroblastoma
Systemic lupus erythematosus
Endocarditis
Ovarian cancer
Fanconi's anemia
Ataxia
Lupus erythematosus
Pancreatic cancer
Herpes simplex
Colorectal cancer
Esophagitis
Colon cancer
Tuberculosis
Sarcoma
Hepatitis c
Adenocarcinoma
Lung cancer
Spinal muscular atrophy
Endotheliitis
Pneumonia
Diamond-blackfan anemia
Pancreatitis
Alzheimer's disease
Hypoxia
Squamous cell carcinoma
Muscular atrophy
Neuronitis
Anemia
Gastric cancer
Prostate cancer
Leukemia
Hepatitis
Hepatocellular carcinoma
Prostatitis
Influenza
Breast cancer
Disease and Phenotype
0
20
40
60
80
100
120
140
160
-Log(Corrected p-value)
B
Smooth muscle
Thyroid
Bone marrow
Small intestine
Fetal thyroid
Testis
Lymph node
Fetal lung
Fetal brain
Lung
Fetal liver
Prostate
Breast
Myeloid
Whole blood
Nk cells
Endothelial
Monocytes
T cells
B cells
Anatomical Context
0
10
20
30
40
50
60
70
80
90
-Log(Corrected p-value)