

RESEARCH

Open Access

Dissecting the expression landscape of RNA-binding proteins in human cancers

Bobak Kechavarzi¹ and Sarath Chandra Janga^{1,2,3*}

Abstract

Background: RNA-binding proteins (RBPs) play important roles in cellular homeostasis by controlling gene expression at the post-transcriptional level.

Results: We explore the expression of more than 800 RBPs in sixteen healthy human tissues and their patterns of dysregulation in cancer genomes from The Cancer Genome Atlas project. We show that genes encoding RBPs are consistently and significantly highly expressed compared with other classes of genes, including those encoding regulatory components such as transcription factors, miRNAs and long non-coding RNAs. We also demonstrate that a set of RBPs, numbering approximately 30, are strongly upregulated (SUR) across at least two-thirds of the nine cancers profiled in this study. Analysis of the protein–protein interaction network properties for the SUR and non-SUR groups of RBPs suggests that path length distributions between SUR RBPs is significantly lower than those observed for non-SUR RBPs. We further find that the mean path lengths between SUR RBPs increases in proportion to their contribution to prognostic impact. We also note that RBPs exhibiting higher variability in the extent of dysregulation across breast cancer patients have a higher number of protein–protein interactions. We propose that fluctuating RBP levels might result in an increase in non-specific protein interactions, potentially leading to changes in the functional consequences of RBP binding. Finally, we show that the expression variation of a gene within a patient group is inversely correlated with prognostic impact.

Conclusions: Overall, our results provide a roadmap for understanding the impact of RBPs on cancer pathogenesis.

Background

RNA-binding proteins (RBPs) have been identified as key regulatory components interacting with the RNA within a cell. Their function is largely dependent on their expression and localization within a cell. They may be involved in processes ranging from alternative splicing to RNA degradation. Combining together, RBPs form dynamic ribonucleoprotein (RNP) complexes, often in a highly combinatorial fashion that can affect all aspects of the life of RNA [1-3]. Due to their central role in controlling gene expression at the post-transcriptional level, alterations in expression or mutations in either RBPs or their binding sites in target transcripts have

been reported to be the cause of several human diseases such as muscular atrophies, neurological disorders and cancer (reviewed in [4-7]). These studies suggest there is precise regulation of expression levels of RBPs in a cell. In fact, a recent system-wide study of the dynamic expression properties of yeast RBPs showed that RBPs with a high number of RNA targets are likely to be tightly regulated, since significant changes in their expression levels can bring about large-scale changes in the post-transcriptional regulatory networks controlled by them [8]. RBPs have also been shown to autoregulate their expression levels. Fluctuations in the expression of autoregulatory RBPs are significantly decreased [9]. These results show that a low degree of expression noise for RBPs is a characteristic feature of their normal state.

Cancer is a complex genetic disease and many of its regulatory factors have been identified as being irregularly expressed. In particular, changes in the normal expression of RBPs have been shown to alter their function leading to a cancer phenotype [10]. Enhanced eIF4E

* Correspondence: scjanga@iupui.edu

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

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

Full list of author information is available at the end of the article

and HuR expression levels have been implicated in initiating translation of mRNAs encoding mostly for pro-oncogenic proteins and other cancer-promoting processes. For instance, Sam68 regulates the alternative splicing of cancer-related mRNAs [10]. Yet another example is the cell-specific alternative splicing of FAS (Fas cell surface death receptor, a member of the TNF receptor superfamily) mRNA. This has been linked to cancer predisposition depending on whether the pro- or anti-apoptotic protein form is produced as a result of the interplay between various RBPs on the FAS transcript [11-14]. In some cases, disruption of the functionality of RBPs, although without directly acting on oncogenic genes, has been shown to affect alternative splicing regulation or the regulation of alternative cleavage mechanisms on transcripts, which can lead to the development of cancer [15,16].

In a recent study, Castello and co-workers [17] utilized cross-linking and immunoprecipitation (CLIP) and photoactivatable-ribonucleoside-enhanced CLIP (PAR-CLIP) to isolate and validate, via proteomics, a set of approximately 850 high-confidence RBPs in humans. These approaches can be used to catalogue and study RBPs and their post-transcriptional networks in healthy and diseased states. By knowing the low degree of expression variation that is tolerated by RBPs in a healthy state and identifying them in mammalian systems, we can begin to investigate their dysregulation profiles in various disease conditions.

In this study, we analyzed the expression patterns of RBPs in a set of 16 healthy human tissues and compared their fold change in expression levels in nine human cancers using the high-resolution expression profiles based on RNA sequencing (RNA-seq) available from the Human BodyMap (HBM) [18] and the Cancer Genome Atlas (TCGA) [19] (see Figure 1, which outlines the different steps, and Materials and methods). We also compared the network properties of a set of 31 RBPs, which were found to be strongly upregulated (SUR) for most of the cancers studied. The network properties may help to determine the cause of the altered expression for the RBPs. Finally, a subset of RBPs was identified based on their expression profiles and network metrics and their contribution to the survival of patients with breast cancer was investigated.

Results and discussion

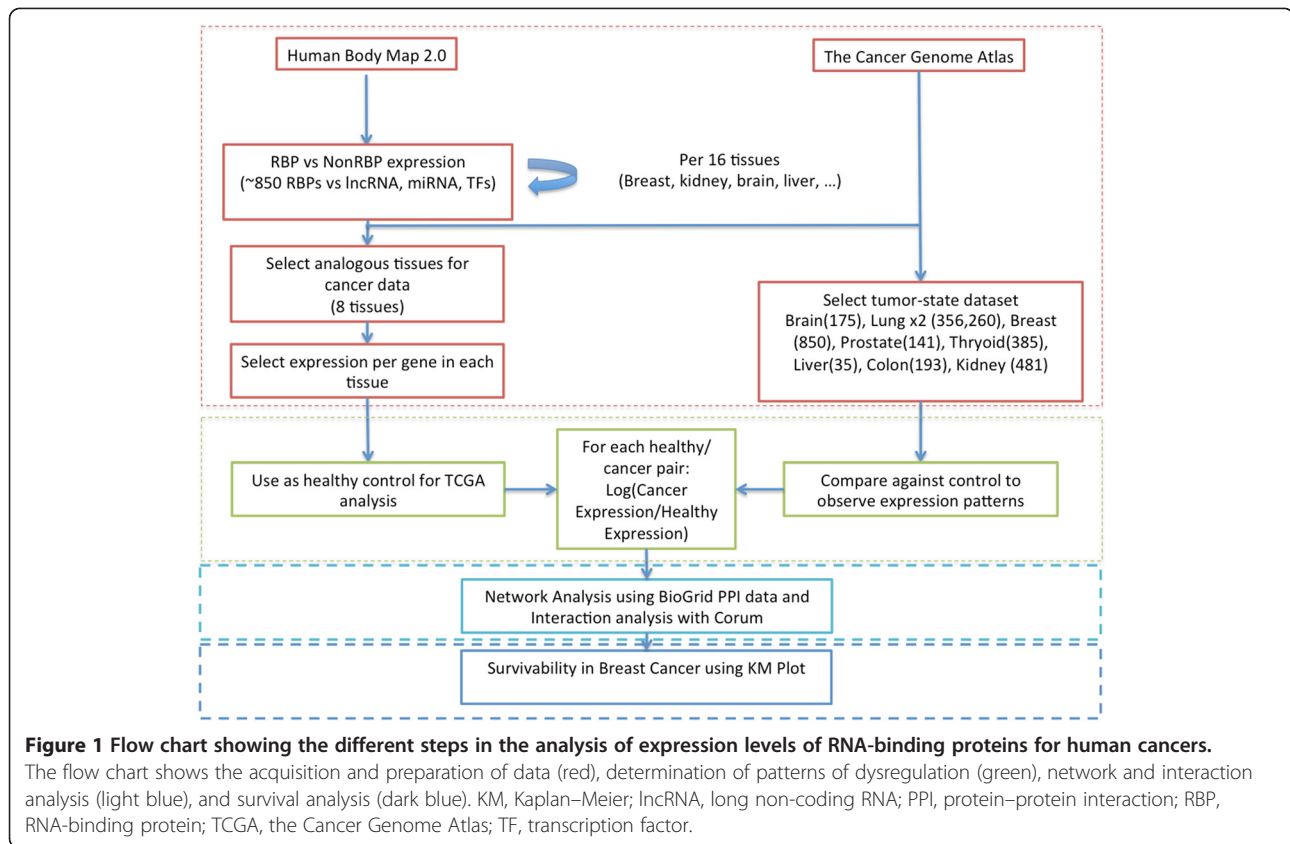
RNA-binding proteins show significantly higher expression than non-RNA-binding proteins and other regulatory factors for 16 human tissues

In eukaryotes, transcription and translation occur in different compartments. This gives a plethora of options for controlling RNA at the post-transcriptional level, including splicing, polyadenylation, transport, mRNA stability, localization and translational control [1,2]. Although some early studies revealed the involvement of RBPs in the transport of mRNA from the nucleus to the

translation site, increasing evidence now suggests that RBPs regulate almost all of these post-transcriptional steps [1-3,20]. RBPs have a central role in controlling gene expression at the post-transcriptional level. Alterations in expression and mutations in either RBPs or their RNA targets (the transcripts that physically associate with the RBP) have been reported to be the cause of several human diseases, such as muscular atrophies, neurological disorders and cancer [4-6,21].

Therefore, we first chose to study the mRNA expression levels of a repertoire of approximately 850 experimentally determined RBPs for all 16 human tissues for which expression data are available from the Human BodyMap 2.0 Project [18,22] (see Materials and methods). This analysis clearly showed that RBPs are significantly more highly expressed ($P < 2 \times 10^{-16}$, Wilcoxon test) than non-RBPs in all of the tissues (Figure 2). Closer inspection of the trends also revealed that some tissues, such as those from the testes, lymph and ovary, had particularly high RBP expression compared to non-RBPs. To determine the regulatory effect of RBPs at the post-transcriptional level compared to other regulatory factors, such as transcription factors (TFs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), their expression levels were compared for different human tissues (see Additional file 1: Figure S1, Additional file 2: Table S1 and Materials and methods). This analysis further revealed that the expression levels of RBPs are significantly different for these 16 tissues compared to these families of regulatory factors ($P < 2 \times 10^{-16}$, Kruskal-Wallis test). Further analysis to compare the expression levels of RBPs and TFs across tissues revealed that except for the heart, kidney, ovary and testis, RBPs are significantly more highly expressed than TFs ($P < 0.05$, Wilcoxon test) (Additional file 2: Table S1). These observations suggest that in most tissues, the magnitude of expression of RBPs is more prominent than even TFs, possibly indicating their central role in controlling gene expression than previously anticipated. Our observation that RBPs are not significantly more highly expressed than TFs in heart, kidney and gonadal tissues like the testis and ovary suggests that both transcriptional and post-transcriptional regulators are equally important in terms of their expression levels in these tissues. In contrast, tissues like the liver ($P < 3.57 \times 10^{-11}$, Wilcoxon test) and white blood cells ($P < 3.85 \times 10^{-5}$, Wilcoxon test) were found to have significantly higher expression for RBPs compared to TFs, possibly indicating the importance of post-transcriptional regulation in the regenerative capabilities of a tissue or in monitoring inflammation and immune response.

The fact that RBPs exhibit a particularly high level of expression in some tissues suggests a need for extensive post-transcriptional control of gene expression in them. For example, the coordinated and cyclic processes of



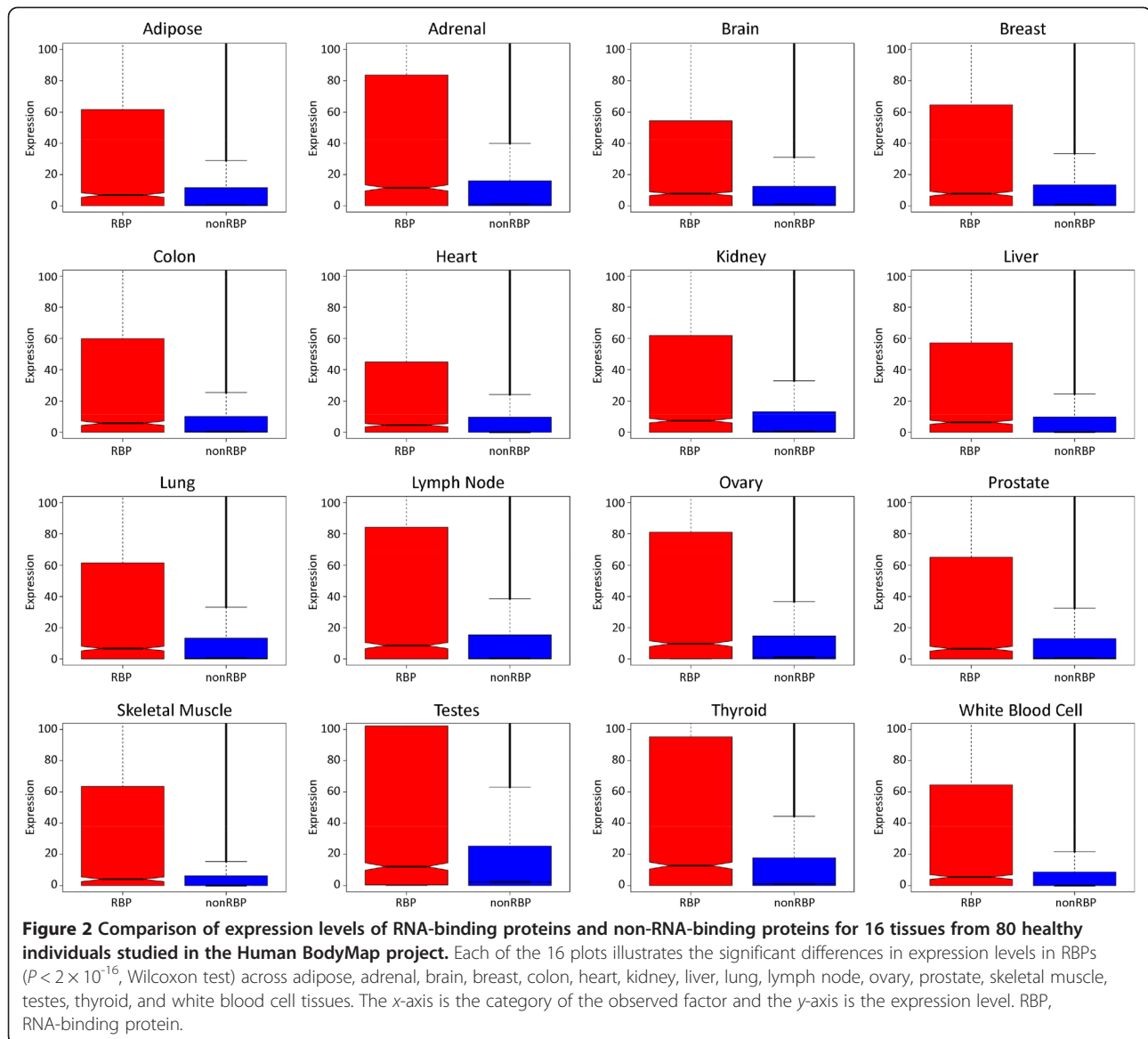
spermatogenesis in testes necessitate the essential temporal and spatial expression of pertinent genes [23]. In the human prostate, slight alterations to the androgen receptor functionality [24] or transcription factors [25] have been shown to lead to a cancerous state. These trends suggest that a significant fraction of the RBPome might play an important regulatory role in diverse human tissues, although in some gonadal and developed tissues, RBPs and TFs had similar levels of expression. Our results show that the high expression of RBPs is especially important in developmentally important tissues suggesting that any patterns of dysregulation could strongly effect these tissues [8].

RNA-binding proteins are dysregulated across cancers and a subset are strongly upregulated across a majority of cancers

Based on our understanding of the expression landscape of RBPs in healthy human tissues, we next asked whether RBPs are dysregulated across cancers (see Materials and methods). Since expression data for healthy tissue was available for eight tissues from the Human BodyMap project corresponding to a set of nine different cancers profiled in the Cancer Genome Atlas (TCGA), we calculated the log-ratio of expression levels of RBPs in the healthy to cancerous states in each of the nine cancers

(Materials and methods). Positive values represent a shift towards upregulation, or, more generally, increased transcript abundance. Negative log-ratios represent a trend of downregulation or decreased abundance. The log-ratio expression profile matrix for the nine cancers was hierarchically clustered to show patterns of similar dysregulation (Additional file 3: Figure S2 and Additional file 2: Table S1 includes log-ratio expression of RBPs). We observed that cancers in similar tissues (lung adenocarcinoma and lung squamous carcinoma) are clustered together suggesting a similar degree of dysregulation of the RBP repertoire. Our analysis also revealed that similar cancers, such as adenocarcinomas were clustered together. These trends indicate that expression ratios are reliable for profiling cancers with unique morphologies in various body locations.

An analysis of the log-ratios representing the fold changes in expression of RBPs between healthy and cancerous states for nine different cancers allowed us to define a criterion for classifying RBPs as strongly upregulated (SUR) or not (non-SUR) (Figure 3, Materials and methods). If an RBP, across six of the nine cancers, was found to have a log-ratio for expression level change of at least nine, it was classified as highly dysregulated, otherwise it was not considered to be a significantly dysregulated RBP. This also corresponded to the RBPs that belonged to the upper



quartile of the fold changes in expression across cancers. According to this criterion, all the RBPs that had at least a ninefold change in expression were found to be only upregulated and hence this group was termed SUR RBPs (Figure 3). Table 1 lists these 31 SUR RBPs (Additional file 4: Table S2 provides detailed information).

We then asked whether tumor-matched normal expression data for TCGA samples can further support the set of SUR RBPs identified here. Although 'normal' site tissue samples from TCGA cannot provide an adequate control, since these samples are collected from a cancerous tissue and it is entirely feasible that the expression levels would still be in a state of dysregulation at the neighboring sites, this analysis can still provide an additional level of support for SUR RBPs. Additionally it

is not possible to control for morphological types of tumors, which, depending on their type, can affect more than just the site of the tumor growth. Nevertheless, we profiled the tumor-matched normal expression levels that are available for eight of the nine cancer types with varying number of samples for breast (106 patients), colon (20 patients), kidney (69 patients), liver (49 patients), two types of lung cancers (57 and 50 patients), prostate (45 patients) and thyroid (58 patients). As suspected, we found the fold changes in expression for all the genes across eight cancers to be minimal (median [IQR] 0.055 [-0.28-0.39]), suggesting that tumor-matched normal expression data may not reflect a true healthy control. However, when we compared the fold changes in expression levels for RBPs and non-RBPs in

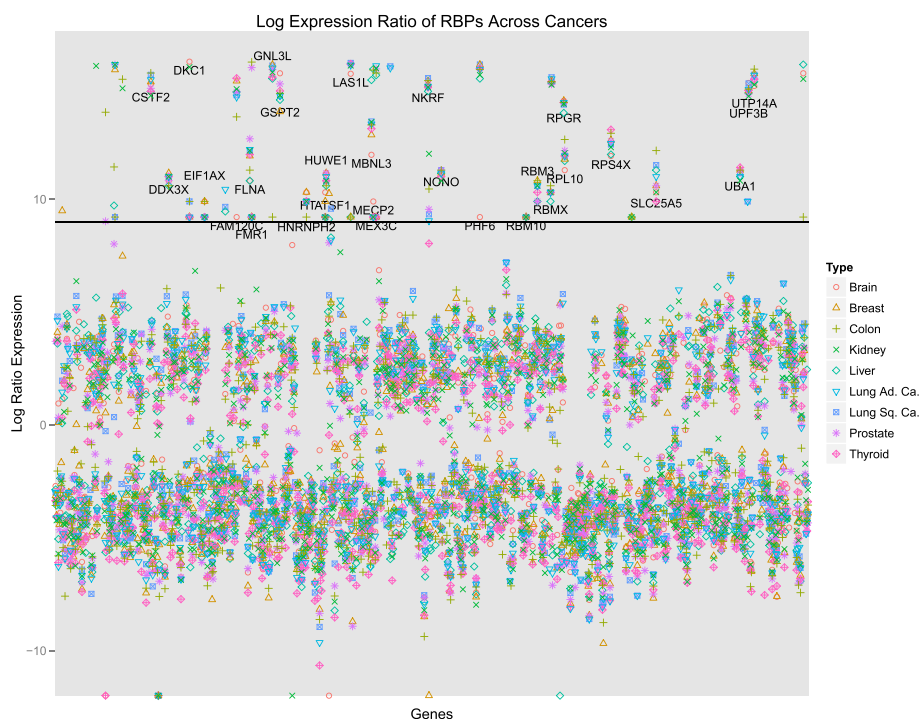


Figure 3 Log-ratio of expression for cancer to healthy expression for RNA-binding proteins in nine human cancers. The x-axis is an index of all the RNA-binding proteins that could be extracted from the expression data in the Cancer Genome Atlas. The y-axis is the ratio of the median expression level for each gene across patients versus the observed expression in the Human BodyMap data. Marked are the 31 strongly upregulated RBPs that have an expression ratio over nine across more than half of the studied cancers. Lung Ad. Ca., lung adenocarcinoma; Lung Sq. Ca., lung squamous carcinoma; RBP, RNA-binding protein.

the tumor-matched samples across cancers, we found that RBPs exhibited significantly higher fold changes compared to non-RBPs (median [IQR] 0.104 [-0.07:0.29] for RBPs versus median [IQR] -0.034 [-0.39:0.25] for non-RBPs, $P < 2.2 \times 10^{-16}$, Wilcoxon test) clearly indicating that RBPs are still significantly upregulated in tumors.

Further analysis to test for the enrichment of RBPs in the top quartile of upregulated genes across cancers revealed that RBPs are strongly over-represented in this list ($P = 1.62 \times 10^{-93}$, hypergeometric test). We also found that all the SUR RBPs are significantly dysregulated ($P < 0.001$, *t*-test comparing tumor and matched normal samples) in at least four of the eight cancers profiled (Additional file 2: Table S1). When we raised the stringency to identify an RBP to be dysregulated in at least six or more cancer types, we still found 24 of the original 31 SUR RBPs to be detected at $P < 0.001$. Very few SUR RBPs from the cancer types Kidney renal cell carcinoma (KIRC) and Liver Hepatocellular Carcinoma (LIHC) were found to be significantly altered in the tumor-matched analysis. While most of the SUR RBPs were found to be upregulated in the tumor-matched analysis, we also found cases of downregulation (Additional file 2: Table S1). Nevertheless, SUR RBPs as a group were also found to be strongly over-represented in the top

quartile of the upregulated set in the tumor-matched analysis ($P = 2.16 \times 10^{-8}$, hypergeometric test), further supporting the notion that SUR RBPs identified using an external healthy control across a broad range of cancers are a confident set of dysregulated RBPs.

Non-RBP log-ratios showing the expression changes were also calculated using the external healthy data to determine if the proportion of strongly upregulated genes (SURs) in RBPs is significantly enriched. We found that the proportions were significantly different ($P < 0.05$, hypergeometric test) with RBPs having a higher proportion of SURs than non-RBPs. Several of these SUR RBPs were annotated to function in important biological processes, such as regulation of gene expression, transcriptional regulation and transport of biomolecules, although very few studies have explored their role in the context of post-transcriptional control, suggesting that their functional roles are far more diverse than previously understood and appreciated.

Of these RBPs classified as SUR RBPs, we note several that have already been implicated in complex genetic disorders and cancer or in cellular regulation and proliferation (Additional file 4: Table S2). Identified RBPs, such as NONO, are involved in RNA biogenesis and DNA double-strand break repair, and have been found

Table 1 Strongly upregulated RNA-binding proteins identified from nine cancers in humans and their cancer relevant references

Associated gene name	Description	References
CCDC124	Coiled-coil domain containing 124	
CSTF2	Cleavage stimulation factor, 3' pre-RNA, subunit 2, 64 kDa	[26]
DDX3X	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked	[27-29]
DKC1	Dyskeratosis congenita 1, dyskerin	[30-32]
EIF1AX	Eukaryotic translation initiation factor 1A, X-linked	
FAM120C	Family with sequence similarity 120C	
FLNA	Filamin A, alpha	[33-36]
FMR1	Fragile X mental retardation 1	
GNL3L	Guanine nucleotide binding protein-like 3 (nucleolar)-like	[37,38]
GSPT2	G1 to S phase transition 2	
HNRNPH2	Heterogeneous nuclear ribonucleoprotein H2 (H')	
HTATSF1	HIV-1 Tat specific factor 1	
HUWE1	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	[39]
LAS1L	LAS1-like (<i>Saccharomyces cerevisiae</i>)	
MBNL3	Muscleblind-like splicing regulator 3	
MECP2	Methyl CpG binding protein 2 (Rett syndrome)	
MEX3C	Mex-3 homolog C (<i>Caenorhabditis elegans</i>)	
NKRF	NFKB repressing factor	[40]
NONO	Non-POU domain containing, octamer-binding	[41,42]
PHF6	PHD finger protein 6	[43-45]
RBM10	RNA-binding motif protein 10	
RBM3	RNA-binding motif (RNP1, RRM) protein 3	[46-49]
RBMX	RNA-binding motif protein, X-linked	[50]
RBMX2	RNA-binding motif protein, X-linked 2	
RPGR	Retinitis pigmentosa GTPase regulator	
RPL10	Ribosomal protein L10	
RPS4X	Ribosomal protein S4, X-linked	
SLC25A5	Solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 5	
UBA1	Ubiquitin-like modifier activating enzyme 1	[51,52]
UPF3B	UPF3 regulator of nonsense transcripts homolog B (yeast)	
UTP14A	UTP14, U3 small nucleolar ribonucleoprotein, homolog A (yeast)	

to be regulated by other factors, when dysregulated potentially promote carcinogenesis [41]. DDX3X, a member of the DEAD box RNA helicase family, has been shown to affect Wnt pathways, which leads to the developments of cancers [27]. DDX3X has also been demonstrated to promote growth and neoplastic transformation of breast epithelial cells [28]. Another SUR RBP, LAS1L was identified to interact with PELP1, which is implicated in pancreatic cancers [53]. HUWE1 is a member of the HECT family of E3 ubiquitin ligases, which has been identified as being overexpressed in breast, lung and colorectal cancers [54]. Indeed, increasing evidence now points to the role of novel ubiquitin-protein ligases in binding to RNA [55,56]. For instance, ubiquitin-like fold has been recently shown to be independently enriched in novel unconventional RBPs identified in the yeast genome [57]. The RNA-binding protein RBM3 is associated with cisplatin sensitivity, the probability of a patient becoming resistant to cisplatin treatment and a positive prognosis in epithelial ovarian cancer [46]. RBM3 has seldom been found expressed in normal tissues, but it is more expressed in common cancers, particularly for the nuclear expression of Estrogen-Receptor (ER) positive tumors. These findings suggest the possible utility of the gene as a positive prognostic marker [47,48].

PHF6 encodes a plant homeodomain (PHD) factor containing four nuclear localization signals and two imperfect PHD zinc-finger domains and it has been proposed that it has a role in controlling gene expression [58]. Inactivating mutations in PHF6 cause Börjeson-Forssman-Lehmann syndrome, a relatively uncommon type of X-linked familial syndromic mental retardation [58-60]. Recent studies show that mutations of this gene are implicated in the development of T-cell acute lymphoblastic leukemia and mutations have been detected in other forms of leukemia as well, suggesting a strong role in tumorigenesis [43,61]. For other nucleolar proteins such as dyskerin (DKC1), which is responsible for the biogenesis of ribonucleoproteins and telomerase stability, the loss or gain of functions is associated with tumorigenesis [30-32]. Filamin A (FLNA) is an actin-binding protein, which interacts with a number of proteins including signaling molecules and membrane receptors, and its expression has been correlated with metastases in prostate and lung cancers [33,34]. A recent study demonstrated the role of FLNA as a nucleolar protein that associates with the RNA polymerase I (Pol I) transcription machinery to suppress rRNA gene transcription [62]. Although further confirmation of how the global RNA-binding role of unconventional RBPs, like the E3 ubiquitin ligase HUWE1, contribute to cancer is needed, increasing evidence suggests that several enzymes and kinases bind to RNAs to control numerous

cellular processes [57,63]. Recent genome-wide screens for novel RBPs further support these observations, suggesting that unconventional RBPs are enriched for enzymatic functions [57,64]. Functional enrichment analysis of SUR RBPs using the DAVID functional annotation system [65] revealed that RNA splicing, nucleotide binding and ribosome biogenesis were the common biological processes associated with these proteins, with a significant fraction of them associated with nucleolus and nuclear lumen cellular components (Additional file 4: Table S2).

Our observations combined with the existing corpus of literature in support of the roles for several of these SUR RBPs in cancerous states, suggest that their dysregulation could be the cause or result of the cancer phenotypes, especially given that even slight alterations in the expression levels of RBPs can bring about large-scale changes in the RBP–RNA interaction networks that they control [8]. It is important to note that although some of these SUR genes shown in Table 1 have been described in relation to cancer, there is little evidence in support of their contribution to either being RBPs or their post-transcriptional network as a contributing factor for the cancer phenotype. Our results in this study implicate them as a strongly upregulated set of RBPs across multiple cancers. Our analysis also corroborates that these significantly dysregulated RBPs are not an artifact of aberrations in calculations, or due to variability in patient expression data mainly because: (1) most of our patient sample sets are at least of the order of 100 for the cancers studied and (2) fold changes in expression levels between healthy and cancerous states for each patient were used to calculate the median fold change in expression of an RBP to account for extreme outliers. Our results also emphasize that these high expression levels may be indicative of a major dysfunction of these RBPs in addition to dysregulation. For example, the mutated form of PHF6, which is implicated in various forms of leukemia, has higher expression. Alternatively, the change in expression may be a result of an upstream alteration in the regulatory mechanisms, for example NONO; another example is that NKRF expression is regulated by miR-301a [40]. The high expression of some of these RBPs may be the result of their normal physiological levels being too low compared to a cancer context, as is the case for the proposed positive prognostic marker, RBM3. So a natural question to ask is whether RBPs have some prognostic impact for cancer, starting from the trends that have been observed in this expression analysis.

Strongly upregulated and non-strongly upregulated RNA-binding proteins exhibit significantly different within-group path lengths and variability in expression is related to the number of interactions

To identify further characteristics that differentiate SUR RBPs in cancer, we calculated the network properties of

all the RBPs using a network constructed from the experimentally reported set of protein–protein interactions in the human genome obtained from the BioGRID database [66] (see Materials and methods). In particular, we computed the shortest paths between pairs of proteins within SUR and non-SUR RBP groups (that is, distances from SUR RBPs to SUR RBPs and distances from non-SUR RBPs to non-SUR RBPs) (Figure 4A). SUR RBPs were found to have significantly shorter path lengths to each other when compared to non-SUR RBP path lengths ($P < 2 \times 10^{-16}$, Wilcoxon test). Other network metrics such as normalized degree distribution, normalized closeness, normalized betweenness and mean path lengths for RBPs in each group were also calculated (see Materials and methods). However, we found no significant difference between SUR and non-SUR RBPs for these properties (Additional file 5: Figure S3). This suggests that the interaction properties of an individual RBP (whether it is a hub and so on) do not relate to its dysregulation but rather the set of SUR RBPs are closely intertwined in the physical interaction network compared to the non-SUR RBPs. Although our observations on dysregulation are at the RNA level, it is possible to speculate, from the shorter path lengths observed, that the interaction network and crosstalk between SUR RBPs could also be perturbed in cancer genomes, with one or more of the SUR RBPs predominantly contributing to this perturbation.

Since our analysis of the shortest path lengths between RBPs from SUR and non-SUR groups suggested that the particular protein interaction partners of RBPs might play an important role in mediating or cascading the effect of dysregulation, we rationalized that the protein complex size and a RBP's occurrence frequency in protein complexes would be related to their sensitivity to dysregulation. RBPs long have been known to form protein complexes, and if a key component within a complex is dysregulated or malformed, it would affect its overall functionality. If a SUR RBP was very prolific we would expect that many patterns of dysregulation would occur downstream as a result of the formation of a faulty complex. Furthermore, if these SUR RBPs participate in smaller complexes, it may be that their dysfunction will not be regulated or counteracted by other members within the complex. From the CORUM data [67] (see Materials and methods), five SUR RBPs were identified and 172 non-SUR RBPs were identified. We found that for the two classifications of RBPs (SUR vs non-SUR), there were no significant differences in distributions for either complex size or complex frequency nor was there any correlation with expression levels (Additional file 6: Figure S4 and Additional file 7: Figure S5). While the current coverage of the experimentally characterized human protein complexes is very limited, these results

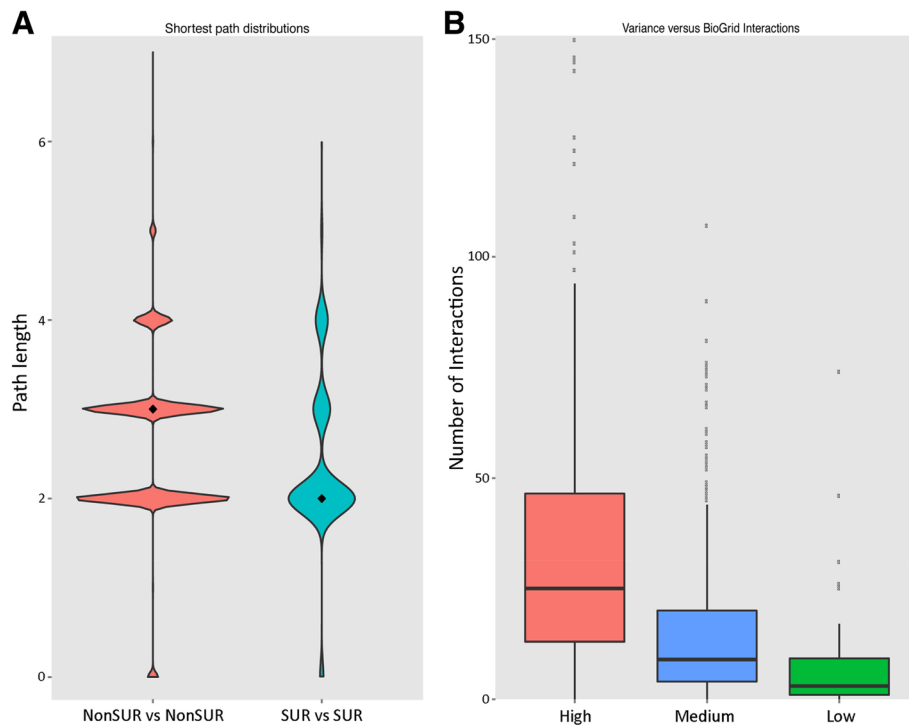


Figure 4 Interaction profiles of RBPs. (A) Distribution of shortest path lengths between every pair of RBPs belonging to SUR and non-SUR RBP groups using the protein–protein interactions documented in the BioGRID database [66], shown as violin plots. The width of each plot is the frequency distribution and the diamond is the median value for the category. SUR RBPs were found to have significantly shorter path lengths amongst themselves in comparison to non-SUR RBPs ($P < 2 \times 10^{-16}$, Wilcoxon test). (B) Box plot showing the number of interactions identified in BioGRID data for RBPs classified by variability levels defined by observed percentiles. The higher the variability for a RBP, the higher the observed number of protein interactions ($P = 9.247 \times 10^{-16}$, low vs medium; $P < 2.226 \times 10^{-16}$, low vs high; $P = 6.6556 \times 10^{-16}$, medium vs high, KS test). RBP, RNA-binding protein; SUR, strongly upregulated; Kolmogorov–Smirnov test (KS test).

indicate that SUR and non-SUR RBPs do not have significant differences in terms of their protein complex membership.

We next asked whether the variability in expression levels of an RBP across cancer patients is different between SUR and non-SUR RBPs. To address this question, we choose breast cancer as our disease model due to the fact that it is the cancer with the most patient samples in TCGA and would naturally be the most robust dataset for identifying variation in the fold changes in expression levels of a RBP. We found that SUR and non-SUR RBPs did not exhibit significantly different expression variation ($P = 0.1212$, KS test), which was measured as the median absolute deviation (MAD) in the expression fold changes between healthy and cancerous tissue across all the patients (see Materials and methods). However, an analysis to test the relation between expression variation and the number of protein interactions of an RBP revealed that the higher the expression variation, the higher the number of protein interaction partners of the RBP (Figure 4B). Indeed, we noticed a significant difference in the number of interactions in the classified levels of variability for RBPs

($P = 9.247 \times 10^{-16}$, low vs medium; $P < 2.226 \times 10^{-16}$, low vs high; $P = 6.6556 \times 10^{-16}$, medium vs high, KS test). In contrast, TFs did not exhibit such significant differences in the number of interactions with the classified levels of variability ($P = 0.8931$, low vs medium; $P = 0.0014$, low vs high; $P = 0.01$, medium vs high, KS test). However, for non-RBPs a significant difference was found between medium and high as well as between high and low levels of variability ($P = 0.7519$, low vs medium; $P < 2.2 \times 10^{-16}$, low vs high; $P < 2.2 \times 10^{-16}$, medium vs high, KS test). The observation that the higher the variability in expression of a RBP the more interactions it has, suggests that fluctuating RBPs whose expression is not tightly controlled might have more promiscuous (non-specific) protein interactions (and protein complexes) thereby leading to RNA off-targets at post-transcriptional level. Our results also suggest that such dysregulation may be suppressed or is minimal due to the lower number of interactions for RBPs with less variability in expression. Our analysis here has focused on the RNA expression levels of RBPs though it is likely that there will be influences from diverse post-transcriptional regulatory phenomena like alternative splicing, translation control and

post-translational modifications, which will affect the ultimate protein levels. Our observations do provide evidence that RBPs with high variability in expression have a higher number of protein interactions.

Survival contributions of RNA-binding proteins in breast cancer is related to network proximity to strongly upregulated RBPs and variability in expression across patients

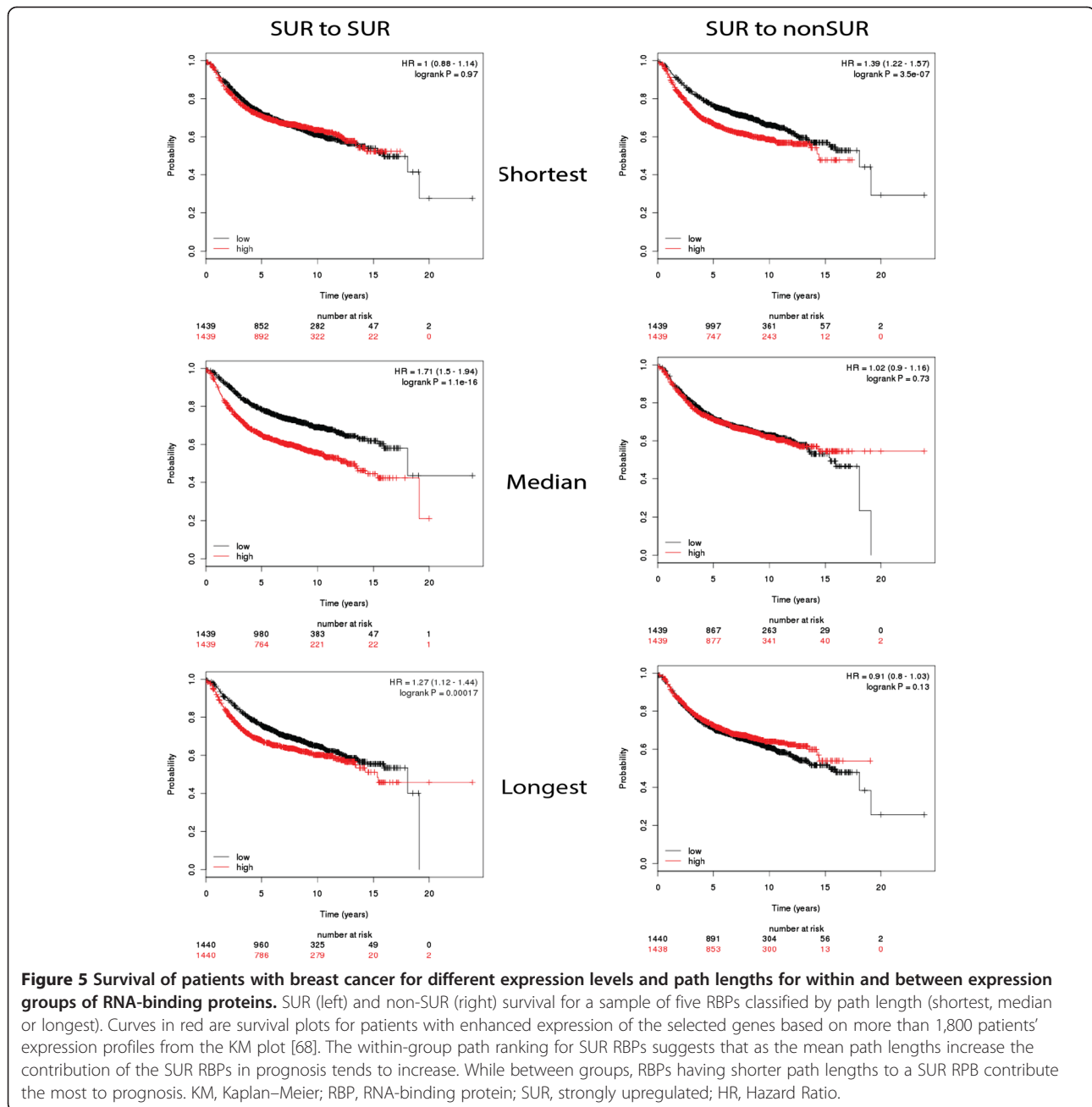
Based on our observation that SUR and non-SUR RBPs significantly differ in their within-group shortest path lengths, we questioned whether the path length of an RBP within the protein–protein interaction network might contribute to its prognostic impact for a cancer. We ranked each RBP in each classification based on the mean path lengths to all connected nodes in the BioGRID protein interaction network and also computed the mean shortest paths to other nodes belonging to SUR RBPs and non-SUR RBPs. This allowed the construction of profiles for overall mean path lengths, lengths within-group for members of the SUR and non-SUR groups, and between the groups. The top five genes with the shortest and longest mean path lengths, and a randomly selected set of genes with intermediate mean path lengths, were selected for the survival analyses (Figure 5) (see Materials and methods). We found that as the mean path lengths between SUR RBPs increased, their contribution to prognostic impact increased. This suggests that SUR RBPs with longer path lengths, that is, those with higher network distances with respect to other SUR RBPs, are more likely to contribute independently to survival as they might influence a larger fraction of the dysregulated network of SUR RBPs. On the other hand, when non-SUR RBPs were sorted by rank based on their mean path lengths with respect to SUR RBPs, we found the opposite trend. This suggests that non-SUR RBPs with shorter distances to SUR RBPs contribute to the perturbation of an important section of the RBP protein interaction network. In particular, if a non-SUR RBP has a shorter path length, it has a good prognostic impact on survival for patients with breast cancer due to its lower expression. SUR RBPs are potentially in a malfunctioning state, and the closer a RBP is to them, the more the prognostic impact influenced by the SUR RBP interactions.

We then compared the overall significance of the Kaplan–Meier P values ($-\log[P]$) for groups of RBPs classified by their level of dysregulation (SUR versus non-SUR) and their levels of variability in expression across patients (high, medium and low variability determined by quartiles, see Materials and methods) in breast cancer (Figure 6). We observed that for both RBPs and non-RBPs, there was no significant difference between SUR and non-SUR genes in terms of prognosis

for survival ($P = 0.12$ and $P = 0.06$, KS test) (Figure 6A,B). However, when we compared the significance of the P values for survival between SURs from RBP and non-RBP groups we found them to be significantly different ($P = 0.05$, KS test). We noted that in the comparison between variability levels of genes in RBPs, there was no significant difference between the Kaplan–Meier (KM) analysis significance levels ($P = 0.945$, low vs medium; $P = 0.3566$, low vs high; $P = 0.1478$, medium vs high, KS test) (Figure 6C). For non-RBPs, we found that the levels of variability did have a very significant difference in the significance of KM-plotter survival P values ($P < 2.226 \times 10^{-16}$, low vs medium; $P < 2.226 \times 10^{-16}$, low vs high; $P = 6.6556 \times 10^{-16}$, medium vs high, KS test) suggesting that, in general, the higher the expression variation of a group of genes, the smaller is their contribution to prognosis for survival (Figure 6D). While there was no significant difference in RBPs we did observe a similar weak trend where the lower the variance in expression across patients, the greater the KM-plotter significance. A highly variable RBP has less effect on survival because it could potentially be regulated by a number of other factors and could be the result of an indirect effect, whereas low variability RBPs have a less but more direct effect on the prognosis for an individual and hence could be the actual drivers. This also corroborates our notion after observing variability versus the number of protein interactions (Figure 4B). More generally, our results suggest that while we observe a larger proportion of SUR RBPs, their elevated expression alone does not necessarily mean they have a direct effect on positive or negative prognoses.

Conclusions

In this study, we investigated the gene expression profiles of RBPs in healthy humans for 16 tissues and found that RBPs are consistently and significantly highly expressed compared to other classes of genes (non-RBPs) as well as in comparison to well-documented groups of regulatory factors like transcription factors, miRNAs and lncRNAs. This, in concordance with previous research, emphasizes their importance in post-transcriptional regulatory control across all the tissues. To understand the expression profile changes in a disease state for hundreds of RBPs in the human genome, we obtained analogous RNA-sequencing-based expression data for a total of 2,876 patient samples spanning nine cancers from TCGA and calculated a log-ratio for expression between cancer and healthy states. We showed that there is a unique signature of approximately 30 RBPs that had significantly increased expression levels across six out of nine (two-thirds) cancers profiled. These could be clearly labeled as a set of SUR RBPs delineating them from the rest of the RBPs based on the change in expression levels. This proportion of SUR RBPs in



the RBP population is greater than the proportion of SUR non-RBPs suggesting for the first time that the expression levels of a significant fraction of the RBPs are affected in cancerous states. Analysis of the protein–protein interaction network properties for SUR and non-SUR group of RBPs, suggested that the shortest path length distributions between SUR RBPs is significantly lower than that observed for non-SUR RBPs. This observation together with survival analysis based on path lengths suggests that not all the SUR RBPs might be directly implicated in cancer but rather that a cause-and-effect relation might hold between some of

the SUR RBPs. This observation was further supported by the fact that the higher the expression variation of a RBP in breast cancer patients, the higher the number of protein–protein interactions. This indicates that fluctuating RBPs whose expression is not tightly controlled (with differing fold changes in expression levels across patients) might be involved in more promiscuous (non-specific) protein interactions thereby leading to variable RNA off-targets at the post-transcriptional level.

To further determine the prognostic impact in breast cancer patients we ranked the SUR and non-SUR RBPs based on path length. The two RBP groups had different

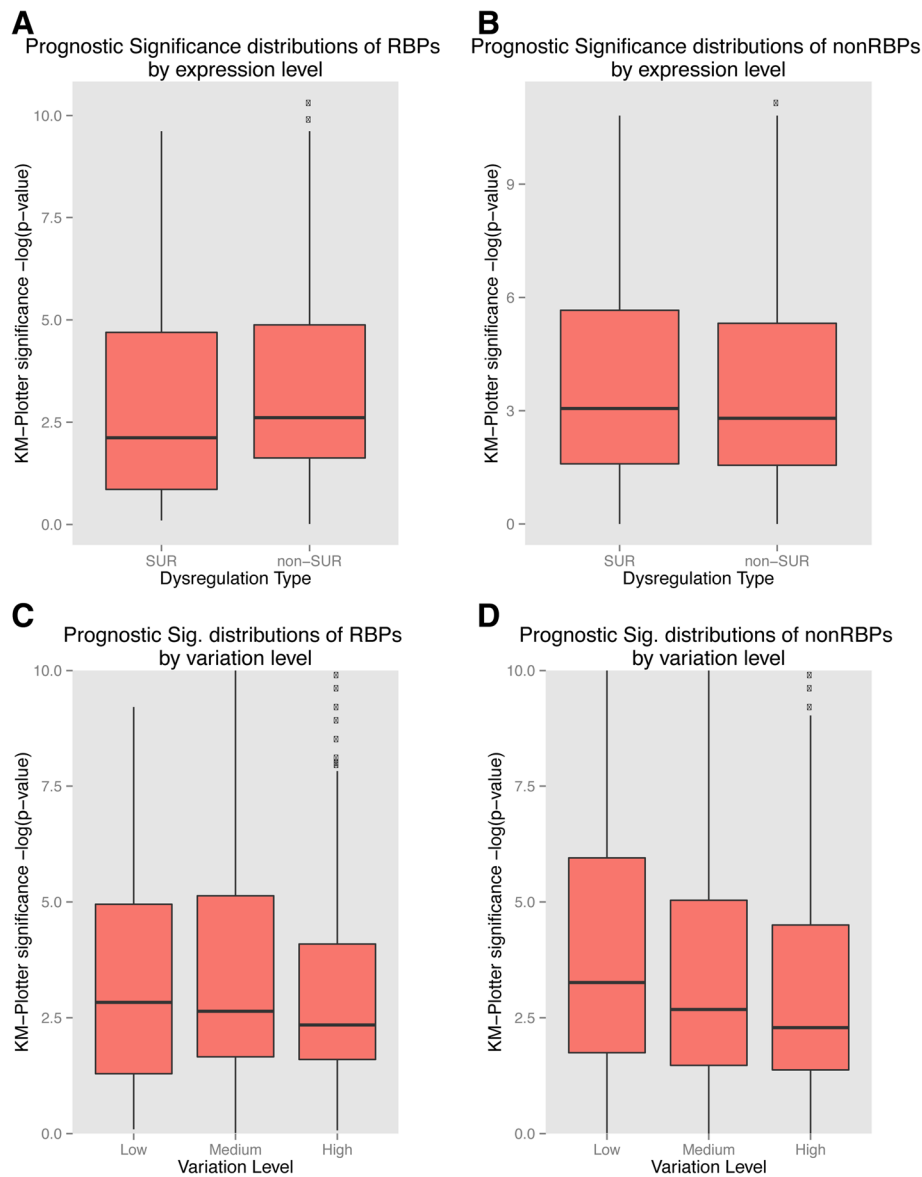


Figure 6 Comparison and distribution of prognostic impact based on expression dysregulation and expression variability in breast tissue. RNA-binding proteins (**A, C**) and non-RNA-binding proteins (**B, D**) were categorized based on their level of dysregulation as healthy or cancer expression (SUR or non-SUR) and the variability of expression levels (high, medium or low) in patients with breast cancer. The statistical significances for the differences in the distributions of prognostic impact are discussed in the main text. KM, Kaplan–Meier; RBP, RNA-binding protein; Sig., significance; SUR, strongly upregulated.

distributions. We found that as the mean path lengths between SUR RBPs increased their contribution to prognostic impact increased, suggesting that SUR RBPs with higher network distances with respect to other SUR RBPs, are more likely to contribute independently to survival as they might influence a larger fraction of the dysregulated network of SUR RBPs. In contrast, when a non-SUR RBP had a shorter path to a SUR RBP, there was a significant prognostic impact. This suggests that they are closer to the actual contributors of pathogenesis at the post-transcriptional level; however, the longer the

path lengths, the weaker the prognosis. To gain further insight into the contribution of these subsets of RBPs in the development of and survival with cancer, we compared the overall significance of the Kaplan–Meier P values ($-\log[P]$) for groups of RBPs classified by their level of dysregulation (SUR vs non-SUR). This analysis revealed no significant differences between groups of SUR and non-SUR RBPs in terms of their prognosis for survival. However, we found that, in general, the higher the expression variation across patients, the lower the prognostic impact of the protein. Our results suggest

that RBPs from our signature set with lower variation in expression levels across patients might be good starting points for studying the effect of RBPs in cancer pathogenesis since SUR RBPs with large expression fold changes might be downstream or there might be indirect effects (Additional file 8: Figure S6). Additionally, common factors that are dysfunctional along the shortest paths in the protein interaction networks of SUR RBPs could also provide clues for potential drug targets as they can act as regulators for rewiring the post-translational landscape of RBPs thereby affecting RNP complex formation. With increasing efforts to uncover the binding sites of RBPs in higher eukaryotes using a variety of high-throughput approaches [69,70], it should also become possible in the near future to study the differences in the target RNA pools between healthy and cancer genomes for several of these SUR RBPs. This would provide a global picture of the affected post-transcriptional regulatory networks. The global integration of networks governed by post-transcriptional players like miRNAs and RBPs together with signaling networks can provide a comprehensive picture of the cause of the dysregulation in these RBPs, which can be used to tease apart the contributions of local malfunctions and those due to an upstream or downstream effect in the cellular networks.

Materials and methods

Data for healthy expression of RNA-binding proteins in 16 human tissues

Our general workflow is illustrated in Figure 1. RNA-seq data for 16 different human tissues from ArrayExpress [71] (Accession no. E-MTAB-513), which is part of the Human BodyMap (HBM) 2.0 project [18,22], was obtained for expression profiling. This data represents the healthy RNA transcript levels of male and female individuals aged 19 to 86, for 16 tissues: adipose, adrenal, brain, breast, colon, heart, kidney, liver, lung, lymph node, ovary, prostate, skeletal muscle, testes, thyroid and white blood cells. Expression data from the HBM project was quantified per transcript using the current annotations of the human genome from the Ensembl. This is available as reads per kilobase per millions of reads (RPKM) for each sample and hence can be compared across and within tissues. Therefore, each of the 16 tissues has a single RPKM value for the expression level of each transcript. A total of 850 genes experimentally characterized as RBPs in the human genome were obtained from a previous publication [17] and 4,647 transcripts associated with these RBPs were identified within the HBM set. The remaining set of 102,462 transcripts were classified as non-RBPs in this study. To examine the other regulatory factors in humans we obtained a set of 9,440 long non-coding RNAs (lncRNAs) from a

Genecode study [18,72], 529 microRNAs (miRNAs) from miRBase [73] and 1,231 transcription factors (TFs) from the DBD database [74] (Additional file 2: Table S1). For each of the 16 tissues we compared the distribution of the RPKM values for transcripts associated with RBPs and non-RBPs, as well as the distribution of expression levels of transcripts associated with RBPs with other regulatory factors to study their relative effect on regulatory control at the tissue level.

Data for cancer expression of RNA-binding proteins for nine cancers in humans

The cancer expression data was downloaded from TCGA [19]. TCGA provides multi-level data (clinical, genome sequencing, microarray, RNA sequencing and so on) procured from a number of institutions, from a variety of patients, for over 25 cancers. In this study, we collected RNAseq V2.0 data for 2,876 patients spanning nine cancers analogous to eight of our tissues in the HBM dataset: breast (850 patients), brain (175 patients), colon (193 patients), kidney (481 patients), liver (35 patients), two for lung (356 and 260 patients), prostate (141 patients), and thyroid (385 patients). TCGA accession numbers for the patient samples used in this study are available in Additional file 9: Table S3. For each cancer we collected the expression levels for each gene for all patients and determined a median representative level and MAD. This defines the genes' RNA expression levels and variability in the relevant cancer state. Likewise, cancer expression and variation were determined for the group of non-RBP genes from HBM as a complementary group for later network, interaction, and expression analyses. Hierarchical clustering of RBP expression for these nine cancers was performed in R, to determine if similar cancers and tissues group together (Additional file 3: Figure S2). Clustering results verified that the collected and amalgamated data are an accurate representation of their anatomical origin, and can be utilized to draw further conclusions.

Profiling for dysregulation of RNA-binding proteins and identification of strongly upregulated RNA-binding proteins across human cancers

For each gene identified as an RBP, we calculated a median expression level of its transcript products in the HBM data when there were multiple protein coding transcripts. To determine the extent of dysregulation in RBPs across cancers, we calculated for each cancer the log-ratio of the median expression in the cancer state over its expression in the associated healthy state. This allowed us to determine for the nine cancers if a particular gene annotated as an RBP is upregulated, downregulated or does not change in expression level in cancer states. Based on this analysis, if an RBP has a log-ratio of

expression level greater than 9 across six or more of the studied cancers, we classified it as being SUR. Otherwise, it was categorized as non-SUR. We focused mainly on defining characteristics unique to these SUR RBPs that differentiate them from other RBPs and non-RBPs. SUR genes as defined here were also observed in non-RBPs and a hypergeometric test was performed to examine potential differences in the proportionality of SUR RBPs and non-SUR RBPs between the two functional classes. The genes associated with RBPs and non-RBPs were also classified by their level of expression variability in a cancer, measured as the MAD value of the fold change in expression for the profiled patients for the cancer. If a gene's variability within a cancer was above the 75th percentile, it was considered highly variable, below the 25th percentile it was considered least variable and the remainder were considered moderately variable.

Network and interaction properties of dysregulated RNA-binding proteins in human cancers

The most recent BioGRID [66] protein–protein interaction (PPI) information (version 3.2.97) was downloaded and used to construct an undirected network of interactions documented in humans. These interactions were used to determine if there were any differences in network properties between the two classifications of dysregulated RBPs, that is, SUR and non-SUR RBPs. This allowed the determination of the potential importance of the classifications for these RBPs. For example, if an SUR RBP forms a hub, it could cause patterns of dysregulation in other, associated interactors. We compared network centrality measures such as degree, closeness and betweenness as well as clustering coefficients and shortest paths between nodes, for different RBP classes utilizing the R package igraph [75]. For shortest paths, we calculated the mean shortest paths for a SUR RBP to other SUR RBPs and SUR RBPs to non-SUR RBPs. We also obtained the overall average path length between each RBP/non-RBP and SUR RBP/non-SUR RBP combination.

Manually curated experimentally characterized human protein complex data was obtained from CORUM [76], to determine the general promiscuity of RBPs in forming complexes. Then 5,217 protein complexes were mapped to the RBPs. We calculated for SUR RBPs and non-SUR RBPs the frequency of membership in CORUM complexes, as well as the mean complex size. This information together with the log-ratios of expression levels between healthy and cancer states in the tissues, allowed us to address whether SUR RBPs are enriched in protein complexes and/or occur in larger or smaller complexes. This analysis also allowed us to test the relation between the extent of an RBP's dysregulation in the context of its membership.

Determination of prognostic impact of RNA-binding proteins for breast cancer

A gene's prognostic impact is the gene's ability to impact positively or negatively patient survival. The prognostic impact for each gene was determined using data from the Kaplan–Meier (KM)-Plotter [68], which was determined from microarray experiments for over 20,000 genes for 1,800 breast cancer patients. For each gene in the RBP and non-RBP groups, we further categorized them as SUR or non-SUR and high or low variability in expression. We compared the significance [$-\log(\text{KM-plotted } P)$] of the prognostic impacts within and between these groups.

Based on the network analyses, the genes were ranked in descending order based on their mean path lengths to the classification of dysregulated genes (SUR vs non-SUR). Path length calculations were determined from a distance matrix generated by the network analysis. From the ranked list of genes we selected five genes with the shortest and longest mean path lengths, and took a random sample of five genes with intermediate mean path lengths. This provided information on the prognostic impact associated with increased gene expression.

Additional files

Additional file 1: Figure S1. Expression levels of RNA-binding proteins (RBPs), non-RBPs, lncRNAs, miRNAs and transcription factors (TFs) for 16 human tissues. Each of the 16 plots illustrates the significant differences in expression levels of RBPs ($P < 2 \times 10^{-16}$, Wilcoxon test) for adipose, adrenal, brain, breast, colon, heart, kidney, liver, lung, lymph node, ovary, prostate, skeletal muscle, testes, thyroid and white blood cell tissues, compared to the other regulatory factors. The x-axis is the category of the observed factor and the y-axis is the expression level.

Additional file 2: Table S1. Expression values for transcripts from HBM data for RBPs, lncRNAs, miRNAs and transcription factors. The statistical analysis compares the expression levels of RBPs and TFs for various tissues. Included are log-fold changes in expression levels for RBPs across nine cancers and P values from t -tests comparing tumor and tumor-matched normal cancer samples. Additionally the statistical significance for interaction and expression/variation classifications are shown.

Additional file 3: Figure S2. Correlation matrix of overall log-ratio expression of RBPs across nine cancers. The matrix shows the clustering of similar tissue sites and similar cancer types.

Additional file 4: Table S2. Strongly upregulated RNA-binding proteins (SUR RBPs), including functional description, currently annotated disease associations and additional database identifiers.

Additional file 5: Figure S3. Comparison of normalized network metrics (closeness, betweenness and degree) between strongly upregulated (SUR) and non-strongly upregulated (non-SUR) RNA-binding proteins. The median values for each property are the same and there are no significant differences ($P > 0.05$, Wilcoxon test).

Additional file 6: Figure S4. CORUM complex membership and complex size distribution for strongly upregulated (SUR) and non-strongly upregulated (non-SUR) RNA-binding proteins. There were no significant differences between the two groups ($P > 0.05$, Wilcoxon test).

Additional file 7: Figure S5. CORUM complex membership and complex size distribution vs expression for strongly upregulated (SUR) and non-strongly upregulated (non-SUR) RNA-binding proteins.

No trends were observed when comparing the CORUM characteristics with expression.

Additional file 8: Figure S6. Heat map showing the variation in expression level measured as median absolute deviation (MAD) values for SUR RNA-binding proteins for nine types of cancer.

Additional file 9: Table S3. Accession numbers of TCGA patient samples for all nine cancers analyzed in this study.

Abbreviations

CLIP: cross-linking and immunoprecipitation; HBM: Human BodyMap; KM: Kaplan–Meier; Kolmogorov–Smirnov test: KS test; lncRNA: long non-coding RNA; MAD: median absolute deviation; miRNA: microRNA; PAR-CLIP: photoactivatable-ribonucleoside-enhanced CLIP; PHD: plant homeodomain; PPI: protein–protein interaction; RBP: RNA-binding protein; RNA-seq: RNA sequencing; RNP: ribonucleoprotein; RPKM: reads per kilobase per millions of reads; SUR: strongly upregulated; TCGA: the Cancer Genome Atlas; TF: transcription factor; TNF: tumor necrosis factor; KS test: Kolmogorov–Smirnov test.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

BK performed the experiments. SCJ conceived and designed the experiments. BK and SCJ analyzed the data, contributed reagents, materials and analysis tools and drafted the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

SCJ acknowledges support from the School of Informatics and Computing at Indiana University Purdue University Indianapolis (IUPUI) in the form of start-up funds. The authors would also like to thank members of the Janga Lab for providing helpful feedback in the course of this study and Sasan Hashemi for help with survival analysis using KM plotter.

Author details

¹Department of Biohealth Informatics, School of Informatics and Computing, Indiana University – Purdue University, 719 Indiana Ave Ste 319, Walker Plaza Building, Indianapolis, IN 46202, USA. ²Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 5021 Health Information and Translational Sciences (HITS), 410 West 10th Street, Indianapolis, IN 46202, USA. ³Department of Medical and Molecular Genetics, Indiana University School of Medicine, Medical Research and Library Building, 975 West Walnut Street, Indianapolis, IN 46202, USA.

Received: 1 August 2013 Accepted: 10 January 2014

Published: 10 January 2014

References

1. Glisovic T, Bachorik JL, Yong J, Dreyfuss G: RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* 2008, **582**:1977–1986.
2. Keene JD: RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* 2007, **8**:533–543.
3. Janga SC: From specific to global analysis of posttranscriptional regulation in eukaryotes: posttranscriptional regulatory networks. *Brief Funct Genomics* 2012, **11**:505–521.
4. Lukong KE, Chang KW, Khandjian EW, Richard S: RNA-binding proteins in human genetic disease. *Trends Genet* 2008, **24**:416–425.
5. Musunuru K: Cell-specific RNA-binding proteins in human disease. *Trends Cardiovasc Med* 2003, **13**:188–195.
6. Kim MY, Hur J, Jeong S: Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB Rep* 2009, **42**:125–130.
7. Castello A, Fischer B, Hentze MW, Preiss T: RNA-binding proteins in Mendelian disease. *Trends Genet* 2013, **29**:318–327.
8. Mittal N, Roy N, Babu MM, Janga SC: Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. *Proc Natl Acad Sci* 2009, **106**:20300–20305.
9. Mittal N, Scherrer T, Gerber AP, Janga SC: Interplay between posttranscriptional and posttranslational interactions of RNA-binding proteins. *J Mol Biol* 2011, **409**:466–479.
10. Wurth L: Versatility of RNA-binding proteins in cancer. *Comparative Functional Genomics* 2012, **2012**:11.
11. Lima L, Morais A, Lobo F, Calais-da-Silva FM, Calais-da-Silva FE, Medeiros R: Association between FAS polymorphism and prostate cancer development. *Prostate Cancer Prostatic Dis* 2008, **11**:94–98.
12. Izquierdo JM: Hu antigen R (HuR) functions as an alternative pre-mRNA splicing regulator of Fas apoptosis-promoting receptor on exon definition. *J Biol Chem* 2008, **283**:19077–19084.
13. Izquierdo JM: Cell-specific regulation of Fas exon 6 splicing mediated by Hu antigen R. *Biochem Biophys Res Commun* 2010, **402**:324–328.
14. Izquierdo JM, Majos N, Bonnal S, Martinez C, Castelo R, Guigo R, Bilbao D, Valcarcel J: Regulation of Fas alternative splicing by antagonistic effects of TI1-A and PTB on exon definition. *Mol Cell* 2005, **19**:475–484.
15. Lee M, Dworkin AM, Gildea D, Trivedi NS, Moorhead GB, Crawford NPS: RRP1B is a metastasis modifier that regulates the expression of alternative mRNA isoforms through interactions with SRSF1. *Oncogene* 2013, online.
16. Mayr C, Bartel DP: Widespread shortening of 3' UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 2009, **138**:673–684.
17. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann Benedikt M, Strein C, Davey Norman E, Humphreys David T, Preiss T, Steinmetz Lars M, Krijgsvelde J, Hentze Matthias W: Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 2012, **149**:1393–1406.
18. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R: The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012, **22**:1775–1789.
19. The Cancer Genome Atlas (TCGA). [<https://tcga-data.nci.nih.gov/tcga>]
20. Darnell RB: HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* 2010, **1**:266–286.
21. Cooper TA, Wan L, Dreyfuss G: RNA and disease. *Cell* 2009, **136**:777–793.
22. Human BodyMap (HBM). [<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513>]
23. Lui W-Y, Cheng CY: Transcriptional regulation of cell adhesion at the blood–testis barrier and spermatogenesis in the testis. *Adv Exp Med Biol* 2012, **763**:281.
24. Green SM, Mostaghel EA, Nelson PS: Androgen action and metabolism in prostate cancer. *Mol Cell Endocrinol* 2012, **360**:3–13.
25. Baena E, Shao Z, Linn DE, Glass K, Hamblen MJ, Fujiwara Y, Kim J, Nguyen M, Zhang X, Godinho FJ, Bronson RT, Mucci LA, Loda M, Yuan G-C, Orkin SH, Li Z: ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* 2013, **27**:683–686.
26. Aragaki M, Takahashi K, Akiyama H, Tsuchiya E, Kondo S, Nakamura Y, Daigo Y: Characterization of a cleavage stimulation factor, 3' pre-RNA, subunit 2, 64 kDa (CSTF2) as a therapeutic target for lung cancer. *Clin Cancer Res* 2011, **17**:5889–5900.
27. Cruciat C-M, Dolde C, de Groot REA, Ohkawara B, Reinhard C, Korswagen HC, Niehrs C: RNA helicase DDX3 is a regulatory subunit of casein kinase 1 in Wnt- β -catenin signaling. *Science* 2013, **339**:1436–1441.
28. Botlagunta M, Vesuna F, Mironchik Y, Raman A, Lisok A, Winnard P Jr, Mukadam S, Van Diest P, Chen JH, Farabaugh P, Patel AH, Raman V: Oncogenic role of DDX3 in breast cancer biogenesis. *Oncogene* 2008, **27**:3912–3922.
29. Wu D-W, Liu W-S, Wang J, Chen C-Y, Cheng Y-W, Lee H: Reduced p21WAF1/CIP1 via alteration of p53-DDX3 pathway is associated with poor relapse-free survival in early-stage human papillomavirus-associated lung cancer. *Clin Cancer Res* 2011, **17**:1895–1905.
30. Alawi F, Lin P: Dyskerin is required for tumor cell growth through mechanisms that are independent of its role in telomerase and only partially related to its function in precursor rRNA processing. *Mol Carcinog* 2011, **50**:334–345.
31. Katunaric M, Zamolo G: Modulating telomerase activity in tumor patients by targeting dyskerin binding site for hTR. *Med Hypotheses* 2012, **79**:319–320.
32. Liu B, Zhang J, Huang C, Liu H: Dyskerin overexpression in human hepatocellular carcinoma is associated with advanced clinical stage and poor patient prognosis. *PLoS One* 2012, **7**:e43147.
33. Bedolla RG, Wang Y, Asuncion A, Chamie K, Siddiqui S, Mudryj MM, Prihoda TJ, Siddiqui J, Chinnaiyan AM, Mehra R, de Vere White RW, Ghosh PM: Nuclear versus cytoplasmic localization of filamin A in prostate cancer:

- immunohistochemical correlation with metastases. *Clin Cancer Res* 2009, **15**:788–796.
34. Uramoto H, Akyürek LM, Hanagiri T: **A positive relationship between filamin and VEGF in patients with lung cancer.** *Anticancer Res* 2010, **30**:3939–3944.
35. Ai J, Huang H, Lv X, Tang Z, Chen M, Chen T, Duan W, Sun H, Li Q, Tan R, Liu Y, Duan J, Yang Y, Wei Y, Li Y, Zhou Q: **FLNA and PGK1 are two potential markers for progression in hepatocellular carcinoma.** *Cell Physiol Biochem* 2011, **27**:207–216.
36. Nallapalli R, Ibrahim M, Zhou A, Bandaru S, Sunkara SN, Redfors B, Pazooki D, Zhang Y, Boren J, Cao Y, Bergo M, Akyurek L: **Targeting filamin A reduces K-RAS-induced lung adenocarcinomas and endothelial response to tumor growth in mice.** *Mol Cancer* 2012, **11**:50.
37. Okamoto N, Yasukawa M, Nguyen C, Kasim V, Maida Y, Possemato R, Shibata T, Ligon KL, Fukami K, Hahn WC, Masutomi K: **Maintenance of tumor initiating cells of defined genetic composition by nucleostemin.** *Proc Natl Acad Sci* 2011, **108**:20388–20393.
38. Rao MRKS, Kumari G, Balasundaram D, Sankaranarayanan R, Mahalingam S: **A novel lysine-rich domain and GTP binding motifs regulate the nucleolar retention of human guanine nucleotide binding protein, GNL3L.** *J Mol Biol* 2006, **364**:637–654.
39. Kurokawa M, Kim J, Geradts J, Matsuura K, Liu L, Ran X, Xia W, Ribar TJ, Henaio R, Dewhirst MW, Kim W-J, Lucas JE, Wang S, Spector NL, Kornbluth S: **A network of substrates of the E3 ubiquitin ligases MDM2 and HUWE1 control apoptosis independently of p53.** *Sci Signal* 2013, **6**:ra32.
40. Lu Z, Li Y, Takwi A, Li B, Zhang J, Conklin DJ, Young KH, Martin R, Li Y: **miR-301a as an NF- κ B activator in pancreatic cancer cells.** *EMBO J* 2011, **30**:57–67.
41. Krietsch J, Caron M-C, Gagné J-P, Ethier C, Vignard J, Vincent M, Rouleau M, Hendzel MJ, Poirier GG, Masson J-Y: **PARP activation regulates the RNA-binding protein NONO in the DNA damage response to DNA double-strand breaks.** *Nucleic Acids Res* 2012, **40**:10287–10301.
42. Tsofack S, Garand C, Sereduk C, Chow D, Aziz M, Guay D, Yin H, Lebel M: **NONO and RALY proteins are required for YB-1 oxaliplatin induced resistance in colon adenocarcinoma cell lines.** *Mol Cancer* 2011, **10**:145.
43. Van Vlierberghe P, Palomero T, Khiabani H, Van der Meulen J, Castillo M, Van Roy N, De Moerloose B, Philippe J, Gonzalez-Garcia S, Toribio ML, Taghon T, Zuurbier L, Cauwelier B, Harrison CJ, Schwab C, Pisecker M, Strehl S, Langerak AW, Geetz J, Sonneveld E, Pieters R, Paietta E, Rowe JM, Wiernik PH, Benoit Y, Soulier J, Poppe B, Yao X, Cordon-Cardo C, Meijerink J: **PHF6 mutations in T-cell acute lymphoblastic leukemia.** *Nat Genet* 2010, **42**:338–342.
44. Yoo NJ, Kim YR, Lee SH: **Somatic mutation of PHF6 gene in T-cell acute lymphoblastic leukemia, acute myelogenous leukemia and hepatocellular carcinoma.** *Acta Oncol* 2011, **51**:107–111.
45. Wang J, Leung JW-c, Gong Z, Feng L, Shi X, Chen J: **PHF6 regulates cell cycle progression by suppressing ribosomal RNA synthesis.** *J Biol Chem* 2013, **288**:3174–3183.
46. Ehlen A, Brennan D, Nodin B, O'Connor D, Eberhard J, Alvarado-Kristensson M, Jeffrey I, Manjer J, Brandstedt J, Uhlen M, Ponten F, Jirstrom K: **Expression of the RNA-binding protein RBM3 is associated with a favourable prognosis and cisplatin sensitivity in epithelial ovarian cancer.** *J Transl Med* 2010, **8**:78.
47. Jogi A, Brennan DJ, Ryden L, Magnusson K, Ferno M, Stal O, Borgquist S, Uhlen M, Landberg G, Pahlman S, Ponten F, Jirstrom K: **Nuclear expression of the RNA-binding protein RBM3 is associated with an improved clinical outcome in breast cancer.** *Mod Pathol* 2009, **22**:1564–1574.
48. Jonsson L, Bergman J, Nodin B, Manjer J, Ponten F, Uhlen M, Jirstrom K: **Low RBM3 protein expression correlates with tumour progression and poor prognosis in malignant melanoma: an analysis of 215 cases from the Malmo Diet and Cancer Study.** *J Transl Med* 2011, **9**:114.
49. Zeng Y, Wodzinski D, Gao D, Shiraishi T, Terada N, Li Y, Griend DJV, Luo J, Kong C, Getzenberg RH, Kulkarni P: **Stress-response protein RBM3 attenuates the stem-like properties of prostate cancer cells by interfering with CD44 variant splicing.** *Cancer Res* 2013, online.
50. Adamson B, Smogorzewska A, Sigoillot FD, King RW, Elledge SJ: **A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response.** *Nat Cell Biol* 2012, **14**:318–328.
51. Moudry P, Lukas C, Macurek L, Hanzlikova H, Hodny Z, Lukas J, Bartek J: **Ubiquitin-activating enzyme UBA1 is required for cellular response to DNA damage.** *Cell Cycle* 2012, **11**:1573–1582.
52. Xu GW, Ali M, Wood TE, Wong D, Maclean N, Wang X, Gronda M, Skrtic M, Li X, Hurren R, Mao X, Venkatesan M, Zavareh RB, Ketela T, Reed JC, Rose D, Moffat J, Batey RA, Dhe-Paganon S, Schimmer AD: **The ubiquitin-activating enzyme E1 as a therapeutic target for the treatment of leukemia and multiple myeloma.** *Blood* 2010, **115**:2251–2259.
53. Kashiwaya K, Nakagawa H, Hosokawa M, Mochizuki Y, Ueda K, Piao L, Chung S, Hamamoto R, Eguchi H, Ohigashi H, Ishikawa O, Janke C, Shinomura Y, Nakamura Y: **Involvement of the tubulin tyrosine ligase-like family member 4 polyglutamylase in PELP1 polyglutamylation and chromatin remodeling in pancreatic cancer cells.** *Cancer Res* 2010, **70**:4024–4033.
54. Bernassola F, Karin M, Ciechanover A, Melino G: **The HECT Family of E3 ubiquitin ligases: multiple players in cancer development.** *Cancer Cell* 2008, **14**:10–21.
55. Kreft SG, Nassal M: **hRUL138, a novel human RNA-binding RING-H2 ubiquitin-protein ligase.** *J Cell Sci* 2003, **116**:605–616.
56. Cano F, Miranda-Saavedra D, Lehner PJ: **RNA-binding E3 ubiquitin ligases: novel players in nucleic acid regulation.** *Biochem Soc Trans* 2010, **38**:1621–1626.
57. Scherrer T, Mittal N, Janga SC, Gerber AP: **A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes.** *PLoS One* 2010, **5**:e15499.
58. Lower KM, Turner G, Kerr BA, Mathews KD, Shaw MA, Gedeon AK, Schelley S, Hoyme HE, White SM, Delatycki MB, Lampe AK, Clayton-Smith J, Stewart H, van Ravenswaay CM, de Vries BB, Cox B, Grompe M, Ross S, Thomas P, Mulley JC, Geetz J: **Mutations in PHF6 are associated with Borjeson-Forsman-Lehmann syndrome.** *Nat Genet* 2002, **32**:661–665.
59. Borjeson M, Forsman H, Lehmann O: **An X-linked, recessively inherited syndrome characterized by grave mental deficiency, epilepsy, and endocrine disorder.** *Acta Med Scand* 1962, **171**:13–21.
60. Turner G, Lower KM, White SM, Delatycki M, Lampe AK, Wright M, Smith JC, Kerr B, Schelley S, Hoyme HE, De Vries BB, Kleefstra T, Grompe M, Cox B, Geetz J, Partington M: **The clinical picture of the Borjeson-Forsman-Lehmann syndrome in males and heterozygous females with PHF6 mutations.** *Clin Genet* 2004, **65**:226–232.
61. Yoo NJ, Kim YR, Lee SH: **Somatic mutation of PHF6 gene in T-cell acute lymphoblastic leukemia, acute myelogenous leukemia and hepatocellular carcinoma.** *Acta Oncol* 2012, **51**:107–111.
62. Deng W, Lopez-Camacho C, Tang JY, Mendoza-Villanueva D, Maya-Mendoza A, Jackson DA, Shore P: **Cytoskeletal protein filamin A is a nucleolar protein that suppresses ribosomal RNA gene transcription.** *Proc Natl Acad Sci USA* 2012, **109**:1524–1529.
63. Ciesla J: **Metabolic enzymes that bind RNA: yet another level of cellular regulatory network?** *Acta Biochim Pol* 2006, **53**:11–32.
64. Tsvetanova NG, Klass DM, Salzman J, Brown PO: **Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*.** *PLoS One* 2010, **5**:e12671.
65. da Huang W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**:44–57.
66. Chatr-aryamontri A, Breitkreutz B-J, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, Reguly T, Breitkreutz A, Sellam A, Chen D, Chang C, Rust J, Livstone M, Oughtred R, Dolinski K, Tyers M: **The BioGRID interaction database: 2013 update.** *Nucleic Acids Res* 2013, **41**:D816–D823.
67. Ruepp A, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Strassky M, Waegele B, Schmidt T, Doudieu ON, Stümpflen V, Mewes HW: **CORUM: the comprehensive resource of mammalian protein complexes.** *Nucleic Acids Res* 2008, **36**:D646–D650.
68. Györfy B, Lanczky A, Eklund A, Denkert C, Budczies J, Li Q, Szallasi Z: **An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients.** *Breast Cancer Res Treat* 2010, **123**:725–731.
69. Konig J, Zarnack K, Luscombe NM, Ule J: **Protein–RNA interactions: new genomic technologies and perspectives.** *Nat Rev Genet* 2011, **13**:77–83.
70. Ray D, Kazan H, Chan ET, Pena Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR: **Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins.** *Nat Biotechnol* 2009, **27**:667–670.
71. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, Kurbatova N, Malone J, Mani R, Mupo A, Pedro Pereira R, Pilicheva E, Rung J, Sharma A, Tang YA, Ternent T, Tikhonov A, Welter D, Williams E, Brazma A, Parkinson H, Sarkans U: **ArrayExpress update – trends in database growth and links to data analysis tools.** *Nucleic Acids Res* 2013, **41**:D987–D990.

72. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M: **GENCODE: the reference human genome annotation for the ENCODE Project.** *Genome Res* 2012, **22**:1760–1774.
73. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic Acids Res* 2011, **39**:D152–D157.
74. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA: **DBD–taxonomically broad transcription factor predictions: new content and functionality.** *Nucleic Acids Res* 2008, **36**:D88–D92.
75. Csardi G, Nepusz T: **The igraph software package for complex network research.** *Inter Journal* 2006, *Complex Systems*:1695.
76. Ruepp A, Waegle B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW: **CORUM: the comprehensive resource of mammalian protein complexes–2009.** *Nucleic Acids Res* 2010, **38**:D497–D501.

doi:10.1186/gb-2014-15-1-r14

Cite this article as: Kechavarzi and Janga: Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biology* 2014 **15**:R14.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

