

# Systemic Bioinformatics Computational Analysis of Hazard Ratio (HR) Level of RNA-Binding Proteins in Human Breast, Colon and Lung Cancer

Tala Bakheet<sup>1</sup>, Nada Al-Mutairi<sup>1</sup>, Mosaab Doubi<sup>1</sup>, Wijdan Al-Ahmadi<sup>1</sup>, Khaled Alhosaini<sup>2,3</sup>, Fahad Al-Zoghaibi<sup>1\*</sup>

<sup>1</sup>Department of Molecular BioMedicine, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; <sup>2</sup>Department of Pharmacology and Toxicology, King Saud University, Riyadh, Saudi Arabia; <sup>3</sup>Department of Molecular and Cell Biology, University of Leicester, Leicester, UK

## ABSTRACT

Breast, colon and lung carcinomas are classified as aggressive tumors that have poor Relapse-Free Survival (RFS) or Progression-Free survival (PF) and poor Hazard Ratios (HRs) despite of extensive therapy. Therefore, it is essential to identify a gene expression signature correlating with RFS/PF and HR status to predict the efficiency of treatment. RNA Binding Proteins (RBPs) play a critical role in RNA metabolic activities including RNA transcription, maturation and posttranslational regulation. However, their particular involvement in cancers is not yet understood. In this study, we used computational bioinformatics to classify the function and the correlation of RBPs among solid cancers. We aimed to identify the molecular biomarker that would help in disease prognosis prediction or improve therapeutic efficiency in treated patients. The intersection analysis summarized more than 1659 RBPs across three recently updated RNA databases. The bioinformatics analysis showed that 58 RBPs were common in breast, colon and lung cancers with HR values <1 and >1 and a significant Q-value<0.0001. RBP gene clusters were identified based on RFS/PF, HR, P-value and fold of induction. In order to define union RBPs, the common genes were subjected to hierarchical clustering and classified into two groups. Poor survival with high-risk HR genes included *CDKN2A*, *MEX3A*, *RPL39L* and *VARS* (Valine cytoplasmic-localized Aminoacyl-tRNA Synthetase) and poor survival with low-risk HR genes included *GSPT1*, *SNRPE*, *SSR1* and *TIA1*, *PPARGC1B*, *EIF4E3* and *SMAD9*. This study may highlight the significant contribution of the 11 RBP genes as prognostic predictors in breast, colon and lung cancer patient and their potential application in personalized therapy.

**Keywords:** Breast; Colon; Cancer; Lung; Monotherapy

**Abbreviations:** AREs: AU-Rich Elements; HRs: Poor Hazard Ratios; OGs: Oncogenes; PF: Progression-Free survival; RBPs: RNA-Binding Proteins; RFS: Relapse-Free Survival; TRAP: Translocon-Associated Protein complex; TS: Tumor Suppressor genes; VARS: Valine cytoplasmic-localized Aminoacyl-tRNA Synthetase

## INTRODUCTION

Cancer is one of the deadliest illnesses globally. According to a World Health Organization (WHO) report in 2020, 10 million new cancers are diagnosed globally each year and this number is expected to rise to 20 million in 17 years. The most common life-

threatening tumors are breast, colorectal, lung, prostate and stomach tumors and they do not respond well to treatment.

Breast, colon and lung carcinomas are classified as aggressive tumors that have poor Relapse-Free Survival (RFS) or poor Progression-Free survival (PF) and poor Hazard Ratios (HRs) in spite of extensive therapy. Recent report on cancer burden on member states of the European Union suggest that 4 million of

**Correspondence to:** Fahad Al-Zoghaibi, Department of Molecular BioMedicine, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; E-mail: zoghaibi@kfshrc.edu.sa

**Received:** 20-Mar-2024, Manuscript No. JTCO-24-30276; **Editor assigned:** 25-Mar-2024, PreQC No. JTCO-24-30276 (PQ); **Reviewed:** 08-Apr-2024, QC No. JTCO-24-30276; **Revised:** 13-May-2025, Manuscript No. JTCO-24-30276 (R); **Published:** 20-May-2025, DOI: 10.35248/2376-130X.25.11.242

**Citation:** Bakheet T, Al-Mutairi N, Doubi M, Al-Ahmadi W, Alhosaini K, Al-Zoghaibi F (2024) Systemic Bioinformatics Computational Analysis of Hazard Ratio (HR) Level of RNA-Binding Proteins in Human Breast, Colon and Lung Cancer. J Theor Comput Sci. 11:242.

**Copyright:** © 2024 Bakheet T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

new cancer cases (excluding non-melanoma skin cancer) and 1.9 million cancer-related casualties occur each year [1]. The most common cause of cancer deaths are lung (0.38 million), colorectal (0.25 million) and breast (0.14 million).

Consequently, experts have always argued that research, information and awareness are crucial in cancer prevention, early detection and strategic decisions on treatment options. Global gene analyses confirm the association between genes, diseases and drugs [2]. Therefore, it is essential to identify a gene expression signature that correlates with RFS or PF and HR status to predict treatment efficiency.

Precision medicine is used to assess the epigenetic regulation of disease at the molecular level in an individual patient and this helps researchers to tailor appropriate and optimal therapies that can be used in addition to current therapies or as a monotherapy based on each patient's unique omics features, maximizing drug efficacy and minimizing adverse drug reactions. However, the fragmentation and heterogeneity of the available data make it difficult to obtain first-hand information.

Regulation of gene expression is mostly performed by RNA-Binding Proteins (RBPs), which bind to unique RNA binding sites and alter the fate or function of the bound RNAs. Over the years, several hundred RBPs have been identified and studied for their critical roles in regulating transcriptional and posttranscriptional gene expression and their unique involvement in cellular processes. RBPs contribute to RNA processing in major human diseases including neurodegenerative diseases, cancer and muscular atrophies [3-7]. However, their particular involvement in cancers is not yet understood.

In this study, we used computational bioinformatics to classify the correlation between the expression level, survival and the HR risk factors of RBPs across solid cancers. We aimed to identify the molecular biomarker that would help in disease prognosis prediction or improve therapeutic efficiency in patients.

A total of 1659 RBP gene summaries were obtained from three different RNA databases: RBPome (1344 genes), census (1542 genes) and RBPDB (416 genes) [8,9]. A total of 58 common RBP gene signatures were collected across breast, colon and lung cancers. Union RBP gene signatures of 11 genes were defined by exposing the common (58) genes to hierarchical clustering with RFS, PF, HR, P-value and fold of induction. Based on the results of the clustering, four clusters were identified. In these clusters, RBPs were classified as poor survival with high-risk HR genes (*CDKN2A*, *MEX3A*, *RPL39L* and *VARS*) and poor survival with low-risk HR genes (*GSPT1*, *SNRPE*, *SSR1* and *TIA1*, *PPARGC1B*, *EIF4E3* and *SMAD9*).

This study may highlight the significant contribution of the 11 RBP genes as prognostic predictors in breast, colon and lung

cancer patient and their potential application in personalized therapy. Here, we present the correlations of the up and down-regulation of RBPs in cancer development [10].

## MATERIALS AND METHODS

### Availability of data

The datasets generated and/or analyzed in this study were obtained from three different RNA databases: RBPome (a catalogue of 1344 experimentally confirmed RBP genes) (Supplementary Table 1), census (manually curated RBPs of 1542 genes) (Supplementary Table 2) and the RBPDB database (experimental RBPs with known RNA-binding domains of 416 genes that were manually curated from the literature) (Supplementary Table 3) [11].

### Union RBP intersection master list and gene clustering

A total of 3336 genes were obtained and compiled from three gene databases including RBPome, census and the RBPDB database. The intersection of these genes comprised the RBP master list of 1659 genes.

Gene expression microarray datasets of the RBP master list were downloaded from Oncomine ([www.oncomine.com](http://www.oncomine.com)). The analyzed datasets were from breast, colon and lung cancers. Up-regulated and down-regulated gene expressions of each dataset were downloaded with matching criteria of P-value<0.05 and Q-value<0.001 (Table 1).

Common RBP signatures across breast, colon and lung cancers were identified and the data of the union RBP master list including the gene expression levels, RFS, PF and P-values for each cancer type were compiled from the work of Balazs Gyorffy (Supplementary Table 4) then exposed to further filtrations. Gene lists were classified based on the value of fold induction: >1.5 and <-1.5. This was followed by filtration on HR values greater than 1 and less than 1 along with significant P-values<0.001. A total of 58 genes were identified across breast, colon and lung cancers.

Supervised hierarchical clustering visualization was performed in JMP® (Version 12, SAS Institute Inc., Cary, NC, 1989-2019) on the common RBP genes. The input to the model was up-/down-regulated RNA data for each cancer. Specifically, the fold change of tumor to normal tissues (T/N), HR and the P-value for each RBP served as the input to the clustering.

**Table 1:** The details list of gene expression microarray datasets downloaded from of Oncomine.

Cancer type	Dataset	No. of samples	No of array genes	Portal	Platform
Breast					

Invasive ductal breast carcinoma	TCGA breast	593	20423	Oncomine	Not defined
Colon					
Colon adenocarcinoma	TCGA colorectal	237	20423	Oncomine	Not defined
Lung					
Lung adenocarcinoma	Okayama	246	19574	Oncomine	Human genome U133 Plus 2.0 Array

## Functional and pathway enrichment analysis of union RBPs

In order to comprehensively analyze the biological functions of union RBPs, we used the Protein Analysis Through Evolutionary Relationships (PANTHER) Gene Ontology (GO) software website to visualize the biological and molecular process and integration of the genes. A pie graph was then constructed using Graph Pad Prism 6.

## Protein-protein interaction network construction and interrelation analysis between pathways

STRING version 11 (<https://string-db.org>) was used to evaluate the current interaction networks and experiments on protein-protein interactions. Then, the interaction networks of these proteins were visualized by executing the list of 11 protein identifiers in the multiple searches and selecting homo sapiens as the organism. The protein network was analyzed to show the interactions at the protein level.

## RBP cross-correlation

Gene expressions for the 11 genes were downloaded from Oncomine, TCGA breast, TCGA colon and Yokohama lung. The diseased gene expressions of 389, 102 and 226 for breast colon and lung cancer, respectively, were collected and multivariate correlation was calculated in SAS Institute Inc. 2013 (SAS® Enterprise Guide™ 6.1) across multiple genes. The correlation values and P-values were input to Graph Pad Prism version 6.05 to construct a volcano plot [12].

## RBP Co-expression with oncogene/tumor suppressor genes

A co-expression analysis was performed in oncomine on the genes in the curtis breast database, as it is a comprehensive database that includes information on many patients. A list of correlated genes was downloaded for each gene of interest in each cancer. This list was searched for matching Oncogenes (OGs) or Tumor Suppressor (TS) genes. The OG database (378 genes) was downloaded from <http://ongene.bioinforminzhao.org/> and the TS database (540 genes) from <https://bioinfo.uth.edu/TSGene/>. In order to get statistically significant correlations, the list of matched OGs and TS genes was filtered

out to extract only the genes with correlation coefficients,  $R \geq 0.20$ .

## Kaplan-Meier plot

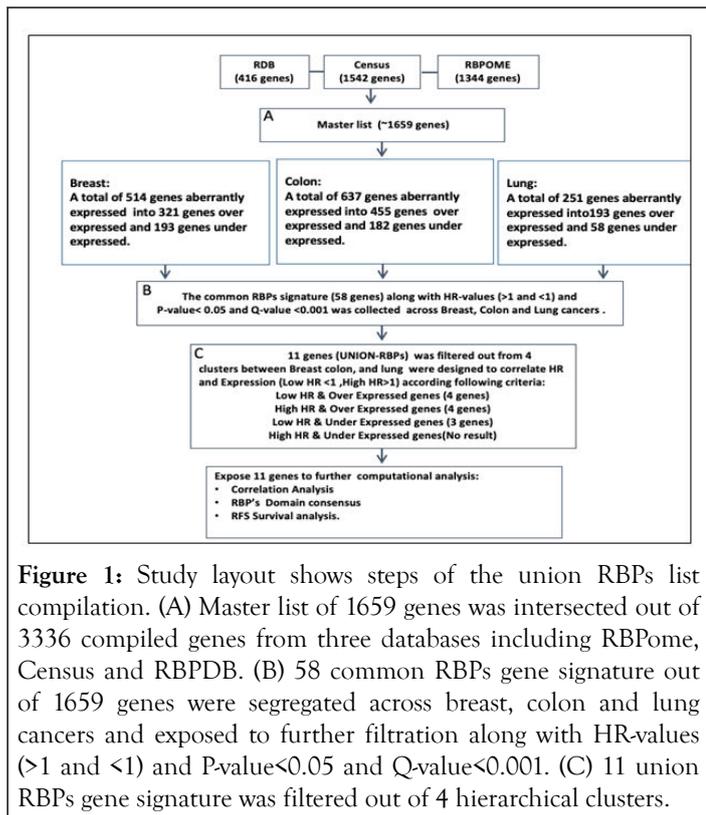
The Kaplan-Meier survival analyses were performed using the Kaplan-Meier Plotter (<http://kmplot.com/analysis/>), a comprehensive dataset for survival analysis that includes the cross-normalized expression data of 54,675 genes in 4,142 breast cancer patients. The database was built from the gene expression and survival data from the European Genome-Phenome Archive (EGA) and the Gene Expression Omnibus (GEO) repositories. Recurrence-free survival was determined using gene cluster stratification. Associations between gene expression and patient survival were assessed using Kaplan-Meier method (log-rank test, GraphPad 6.0) assessed associations between gene expression and patient survival. The percentile threshold algorithm (25) was used to determine the optimal cutoff of the members of the RBP cluster. The jetset best probe set was selected in case multiple probe sets measured the same gene to ensure the optimal probe set for each gene. HRs and P-values were determined using Cox proportional hazards [13].

## RESULTS

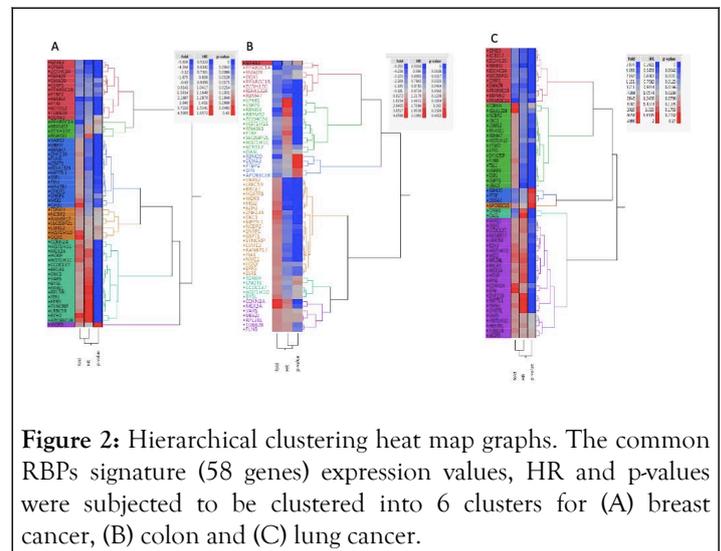
### Union RBP intersection master list and gene clustering

A total of 3336 genes from three RBP databases were compiled and intersected to form the master list, which comprised 1659 genes identified across breast, colon and lung cancers. The layout of the analysis is illustrated in Figure 1.

To segregate the list of common RBP genes, the intersection master list of RBPs was further filtered using 50%-fold induction or reduction to define the list of up and down-regulated genes of each cancer. About 63% of the 514 breast cancer genes were classified as up-regulated genes and 37% were down-regulated. For colon cancer, 71% of the 637 genes were up-regulated and 29% were down-regulated. Of the 251 lung cancer genes, 77% were up-regulated and 23% were down-regulated.



genes were identified for breast, colon and lung cancers. For each cancer, a set of 58 gene expression values, HR and P-values were subjected to hierarchical clustering into six clusters (Figure 2 A-C). The six clusters were identified for breast, colon and lung cancers (Table 2). After clustering each cancer, genes were in clusters of either good/poor survival or up/down-regulated genes [14].



In the microarray gene expression data of the common RBP gene list, representative HR values and P-values were aligned (Supplemental Table 1). An HR value>1 was considered high risk and a HR value<1, was low risk. A total of 58 common

**Table 2:** Union RBP genes clustering: Tables representing the six clusters including the number of genes, gene expression fold change, HR and p-values of each cluster across breast, colon and lung cancer.

Cancer	Cluster	Count	Fold	HR	P-value
Breast	1.	13	-3.0899	0.7487	0.0053
	2.	4	-3.0786	1.1749	0.0531
	3.	15	1.8787	0.7468	0.0018
	4.	7	2.42	1.0572	0.0715
	5.	18	2.2835	1.4626	0.000123
	6.	1	1.8811	1.06	0.43
Colon	1	8	-3.402	0.675	0.0347
	2	12	-3.022	1.6226	0.009
	3	5	-2.8213	0.9812	0.3408
	4	21	1.9641	0.632	0.0042
	5	5	1.7852	0.8115	0.113
	6	7	2.4175	1.328	0.0573
Lung	1	11	-2.2714	0.5686	0.000788

2	17	1.7615	0.6425	0.0046
3	3	-2.0165	0.8606	0.2137
4	1	1.5968	1.12	0.27
5	2	-2.0638	1.7699	0.00018
6	24	1.9161	1.538	0.0086

Since the study aimed to identify key prognostic predictor genes that are associated in terms of survival conditions, HR status and gene expression, the first filtration was based on HR values followed by gene expressions. Four classes were designed to categorize these genes based on their HR values and gene expression (Table 3). Class 1 (highest HR values and up-regulated genes (HR>1 and FI>1.5)), class 2 (highest HR values and down-regulated genes (HR>1 and FI<1.5)), class 3 (lowest HR values and up-regulated genes (HR<1 and FI>1.5)) and class 4 (lowest HR values and down-regulated genes (HR<1 and FI<1.5)).

For example, class 1 in the breast cancer cluster represented in cluster 5 contains 18 genes with average FI and HR values of 2.2835 and 1.4626, respectively and a P-value<0.001. Similarly, in colon cancer, cluster 6 contains 7 genes with average FI and HR values of 22.4175 and 1.3280, respectively and a P-value ~ 0.05 and lung cancer cluster 6 contains 24 genes with average FI and HR values of 1.9161 and 1.5380, respectively and a P-value<0.01 [15].

**Table 3:** The union RBP genes class contributions: Class 1 shows genes with criteria of HR-value>1 and FI>1.5. Class 2 shows genes with criteria of HR-value<1 and FI>1.5. Class 3 shows genes with criteria of HR-value>1 and FI<1.5. Class 4 shows genes with criteria of HR-value<1 and FI<1.5.

	Upregulated Breast/Colon/Lung (FI>1.5)	Downregulated Breast/Colon/Lung (FI<1.5)
High risk (HR>1)	Class 1	Class 3
	MEX3A	NA
	CDKN2A	
	RPL39L	
	VARS	
Low risk (HR<1)	Class 2	Class 4
	GSPT1	PPARGC1B
	SNRPE	EIF4E3
	SSR1	SMAD9
	TIA1	

In class 1, CDKN2A, MEX3A, RPL39L and VARS had HRs>1 and FIs>1.5. Class 2 also had common genes (GSPT1, SNRPE, SSR1 and TIA1) with HRs<1 and FIs>1.5. Class 3 had no common genes with HRs>1 and high FIs<1.5. Class 4 had common genes (PPARGC1B, EIF4E3 and SMAD9) with HRs<1 and FIs<1.5. Therefore, the following 11 genes were compiled from four classes: MEX3A, CDKN2A, RPL39L, VARS, GSPT1, SNRPE, SSR1, TIA1, PPARGC1B, EIF4E3 and SMAD9. The

mean expression levels of each gene across the three types of cancers and their HR values and consensus targets are shown in Table 4 [16].

**Table 4:** The mean expression levels of union RBP genes across the three type of cancers along their HR values and their consensus targets.

	Average fold induction breast, colon and lung	Average HR-value breast, colon and lung	Consensus target	Function
<i>CDKN2A</i>	2.967172	1.36098	unknown	TS that encoding p14 and p16 which are involved in different cellular processes
<i>MEX3A</i>	3.228625	1.500737	mRNA	Putative RBP involve in polarity and stress that contributes with cellular homeostasis and carcinogenesis
<i>RPL39L</i>	1.889479	1.49	ribosome	Ribosomal protein paralogs that are involved in gene translations
<i>VAR5</i>	1.594438	1.318373	tRNA	Charging and catalyzing the bond between tRNA and designated amino acid
<i>GSPT1</i>	1.717739	0.613333	mRNA	Termination of protein translation
<i>SNRPE</i>	1.809188	0.686667	snRNA	Cellular spliceosome complex that involve in mRNA maturation process
<i>SSR1</i>	1.674405	0.683333	unknown	Proteins-specific transportation across ER membrane
<i>TIA1</i>	1.636625	0.66	mRNA	Consider as a TS that involved in controlling the translation and co-localization of target genes in SGs
<i>PPARGC1B</i>	-2.50662	0.6805	unknown	unknown
<i>EIF4E3</i>	-2.99672	0.605903	mRNA	Play essential roles in initiation the protein translation that are involved in mRNA metabolism
<i>SMAD9</i>	-2.42747	0.663987	putative miRNA	Belong to receptor SMAD protein complex, binds to DNA in process of suppressing of target gene transcription

### The functional analysis of the Union RBP consensus targets and the suggested protein-protein network interactions

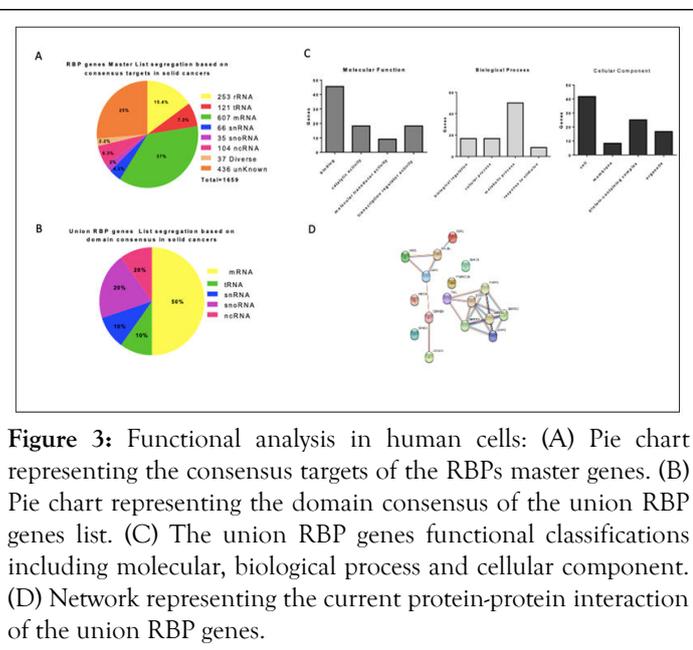
A pie chart was used to visualize the consensus targets of the master RBPs and the gene signatures of the union RBPs in human cells (Figure 3A and B). Interestingly, there were increases in the mRNA, tRNA, snRNA and ncRNA consensus targets in the union RBPs compared to the master RBP list. However, there was a reduction in rRNA percentages and

unknown consensus targets in the union RBPs compared to the master RBP list. This may explain the diversity of consensus domains; therefore, there are no specific consensus targets for the RBPs that are involved in cancer development or therapy rejection.

To determine the molecular and biological function and mechanisms of the union RBP gene signatures, we used the PANTHER GO unifying biology analysis and STRING11 software.

Three functional classifications were defined in panther: Molecular, biological and cellular processes. Binding indirect targets, catalytic activity and transduction regulator activity enhanced the molecular function of the union RBPs. In the biological process, 50% of the RBPs were involved in metabolic processes, and 41% were found in the cytoplasmic compartment (Figure 3C). Union genes were input to STRING11 to analyze possible interactions between the network and experimental protein-protein interactions in order to better visualize and understand the functions of the union RBPs. Molecular functions have shown that they can bind to several RNAs, heterocyclic compounds, organic, cyclic compounds and translation factors. In the biological process, there was mainly an improvement in the regulation of translation, the biosynthesis of cellular nitrogen compounds and macromolecules, gene expression and peptide metabolism. Cellular component cytoplasmic and ribonucleic protein and protein-containing complex [17].

Moreover, the union RBPs obtained 11 PPI nodes, 5 edges, and a PPI enrichment P-value <0.019. In addition, different tandem affinity purification assays, co-immunoprecipitation assays, yeast two-hybrid assays and affinity chromatography assays demonstrated the affinity protein-protein interactions in two main clusters. The first cluster consisted of five genes: *CDKN2A*, *GSPT1*, *SSR1*, *RPL39L* and *VARS*. The second cluster demonstrated protein interactions between *TIA1*, *SNRPE* and others. *MEX3A*, *SMAD9*, *PPARGC1B* and *EIF4E3* were not involved in the protein network interactions (Figure 3D).



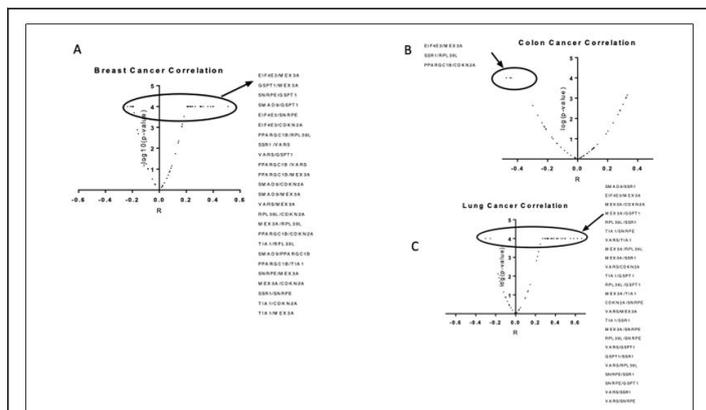
**Figure 3:** Functional analysis in human cells: (A) Pie chart representing the consensus targets of the RBPs master genes. (B) Pie chart representing the domain consensus of the union RBP genes list. (C) The union RBP genes functional classifications including molecular, biological process and cellular component. (D) Network representing the current protein-protein interaction of the union RBP genes.

### Cross-correlation of RBPs

The expression of 11 genes was obtained from the oncomine database to identify the potential cross-correlation between the RBP gene signatures. The multivariate similarity was calculated in SAS Institute Inc. 2013 SAS® Enterprise Guide TM 6.1 across multiple genes. A volcano plot was obtained by plotting the R-values of the common 58 up-regulated and down-regulated

RBP genes on the x-axis and the base 10 logarithm of their corresponding P-values on the y-axis. P-values<0.00001 were considered and reported as base 10 logarithms (P-value of 0.00001=4).

In colon cancer and most probably because of the lack of data, the cross-correlation identified only three correlating groups (*EIF4E3/MEX3A*, *SSR1/RPL39L* and *PPARGC1B/CDKN2A*). The cross-correlation of *EIF4E3/MEX3A* was defined in breast and lung cancer. The cross-correlation of *SSR1/RPL39L* was found only in colon and lung cancer and the cross-correlation of *PPARGC1B/CDKN2A* was found only in colon and breast cancer (Figure 4 A-C) [18].



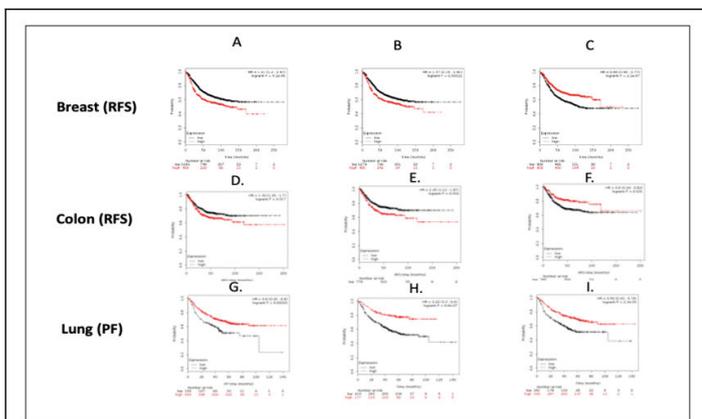
**Figure 4:** Cross correlations between the union RBP genes signature across multiple genes were determined for; (A) Breast, (B) Colon and (D) Lung cancer. Volcano plot is shown by plotting the R-values on the x-axis and their significance P-values as corresponding  $-\log_{10}(p\text{-values})$  on the y-axis.

### Survival analysis

The main aim of this study was to identify the survival-associated factors of cancer treatment plans with the gene expression of union RBPs. Therefore, we extracted the union RBPs that matched the representative treated patient data as explained above from the list of common RBPs in order to assess the prognostic value of the gene signatures of the 11 RBPs. The gene expression of the union RBPs of each cancer was divided into three subgroups (up-/down-regulated genes, up-regulated genes and down-regulated genes). The RFS or PF of each subgroup for each cancer type was then examined using the Kaplan-Meier estimation method and the log-rank test to assess the significant differences of the two-group survival curves. In all RBP gene signature subgroups, patients with up-regulated gene expression had significantly lower survival rates than patients with down-regulated gene expression. This indicated poor survival rates in patients who were treated for high-risk breast cancer (HR=1.4, CI=1.2-1.67, P=.2e-05), colon cancer (HR=1.34, CI=1.05-1.7, P=0.017) and lung cancer (HR=0.6, CI=0.45-0.8, P=0.00045) (Figure 5 A, D and G). In order to interpret the correlation between the gene expression levels of the union RBPs and the survival prognosis in breast, colon and lung cancers, the up-regulated and down-regulated gene clusters were used [19].

In the up-regulated RBP subgroup, patients with up-regulated gene expression had significantly lower survival rates than patients with down-regulated gene expression. This indicated poor survival rates in patients who were treated for high-risk breast cancer (HR=1.37, CI=1.16–1.61, P=0.00022), colon cancer (HR=1.45, CI=1.12–1.87, P=0.004) and lung cancer (HR=0.42, CI=0.3–0.6, P=6.6e-07) (Figure 5 B, E and H). However, in the down-regulated RBP subgroup, patients with down-regulated gene expressions had significantly lower survival rates than patients with up-regulated gene expressions. This indicated poor survival rates in patients who were treated for low-risk breast cancer (HR=0.56, CI=0.56–0.77, P=2.2e-07), colon cancer (HR=0.6, CI=0.44–0.82, P=0.001) and lung cancer (HR=0.56, CI=0.43–0.74, P=2.3e-05) (Figure 5 C, F and I).

Hence, the up-regulation of *CDKN2A*, *MEX3A*, *RPL39L* and *VARS* in breast, colon and lung cancer patients indicated poor survival with high risks and inefficient treatment plans. However, the up-regulation of *GSPT1*, *SNRPE*, *SSR1* and *TIA1* and the down-regulation of *PPARGC1B*, *EIF4E3* and *SMAD9* in breast, colon and lung cancer patients indicated poor survival with low risks and efficient treatment plans.



**Figure 5:** Kaplan-Meier plot representing Relapse-Free Survival (RFS) of breast and colon and Progression-Free survival (PF) of Lung patient across the union RBP genes signature subgroups: (A, D and G) Representing all up and down-regulated genes, (B, E and H) Representing the up-regulated genes and (C, F and I) Representing down-regulated genes across breast, colon and lung respectively.

## DISCUSSION

With the introduction of genomic profiling data and selective molecular-targeted approaches to identify effective therapeutic alternatives, biomarkers have become increasingly important targets in cancer patients clinical diagnosis and treatment. Single gene/protein or multi-gene signature-based assays have been developed to test particular molecular pathway deregulations that direct therapeutic decision-making as predictive biomarkers. For example, the six-gene signature for survival prediction in patients with glioblastoma can be used in personalized therapy and promoting drug efficiency. Gene expression and computational analysis can be used in adjuvant therapy and gene profiling of non-small cell lung cancer patients at high risk of relapse. In addition, the effect of CpG-methylation on gene

expression is a functional and effective prognostic tool for clear cell Renal Cell Carcinoma (ccRCC) that can add prognostic value to the staging system.

Survival and HR analyses are clinical and biostatistical methods used to assess the efficiency of treatment in patient groups. Our study focused on identifying the survival-associated risk factors and the gene signatures of common RBPs across the most common cancers: Breast, colon and lung cancers. We developed a statistical bioinformatics analysis method based on gene expression, RFS, PF, HR and P-values in the data of treated patients [20].

The union RBP gene signatures were classified into two subgroups (overexpression and underexpression) and each subgroup was further divided into genes with poor survival and low-risk HRs and those with poor survival and high-risk HRs. Interestingly, most of the down-regulated genes including *PPARGC1B*, *EIF4E3* and *SMAD9* were classified as poor survival with low risk. However, the up-regulated genes were divided into two subgroups: The poor survival with low risk genes (*GSPT1*, *SNRPE*, *SSR1* and *TIA1*) and the poor survival with high risk genes (*CDKN2A*, *MEX3A*, *RPL39L* and *VARS*) (Table 4). Moreover, these genes play diverse roles in mRNA metabolism and the roles of most of these genes in cancer have not yet been identified.

Based on predicted and experimental STRING data, we identified two signaling pathway relation clusters and non-signaling pathway-related genes of the union RBP gene signatures. The first signaling pathway relation cluster of union RBPs comprised of *CDKN2A*, *GSPT1*, *SSR1*, *RPL39L* and *VARS*. These genes interacted and were collocated and overexpressed in cancer patients. Surprisingly, in this cluster, there were discrepancies between the cellular function, expression correlations, survival-associated RISK factors and HRs of genes. For instance, the overexpression of *CDKN2A*, *RPL39L* and *VARS* overexpression were associated with poor survival and high risk. However, *GSPT1* and *SSR1* were associated with poor survival and low risk. This may elucidate the competition relation in end gene function.

*CDKN2A* is a TS gene that encodes the p14 and p16 proteins and is involved in different cellular processes. It had high fold induction and was associated with poor survival and high risk HR in breast, colon and lung cancer dataset analyses. According to previous studies, the encoded protein level of *CDKN2A* is almost undetectable. However, it has the dual role of blocking tumor development and cell proliferation and in the oncogenic condition, its level increases and stimulates p53-dependent and/or independent pathways. According to STRING software analysis, *CDKN2A* binds directly to *GSPT1* (Figure 3D). Based on Curtis analysis, *CDKN2A* had an 11.4% correlation with OGs and only a 4.0% correlation with TS genes (Table 5). *CDKN2A* may perform its dual function by controlling the expression or suppression of OGs and TSs.

*RPL39L* is a ribosomal protein paralog that is abundantly expressed in cancer and embryonic stem cells. Recently, 2D gel and proteomics analyses have suggested that *RPL39L* and other RPs are involved in gene translation [14]. The hypo-methylation

of cancer-specific CpG Islands (CGIs) and RPL39L reactivation is important for the treatment and risk stratification of lung adenocarcinoma. Based on Curtis analysis, 0.26% correlation with OGs and 4.26% with TS genes, which points to its role in TS gene expression levels with a strong correlation coefficient, R (0.2–4.26) (Table 5).

VARS is one of the 37 Aminoacyl-Trna Synthetases (ARSs). VARS and ARSs mainly charge the tRNA and catalyze the bond between the tRNA and the designated amino acid. VARS mutations are associated with a loss of enzymatic activity and the development of a spectrum of global developmental delays, epileptic encephalopathy and primary or progressive microcephaly [15]. According to our analysis, VARS binds to the GST c-terminal of the target gene (Table 5). In Curtis, VARS had a 2.4% correlation with OGs and a 0.74% correlation with TS genes (Table 5). The expression level of VARS in cancer patients and the high correlation coefficient, R (0.2–0.48), with OGs in contrast to the correlation coefficient, R, of 0.2–0.3 with TS genes, may explain the critical role of this gene in catalyzing the bond between the tRNA and amino acid to translate recovery proteins in the treated cancer cells. This mechanism requires further investigation.

In eukaryotic cells, the stable G1 to S phase transition protein/eukaryotic Release Factor (eRF1) (GSPT1/eRF3a) complex is involved in translation termination. GSPT1 depletion causes cell cycle arrest at the G1 phase *via* inhibition of the mTOR pathway. There is a statistically significant relationship between rs4561483 risk genotype and increased GSPT1 expression in Testicular Germ Cell Tumors (TGCT). Nicotine and EGF induce genes, including GSPT1, to promote the proliferation, invasion and migration of non-small cell lung cancers, thus enhancing its tumorigenic activity and revealing the central role of the inhibitor of DNA binding/differentiation 1 (ID1) and its downstream targets in facilitating lung cancer progression [16]. In this study, the overexpression of GSPT1 in treated cancer patients was associated with poor survival and low-risk HRs (Table 4). Using Curtis, there was a 2.4% correlation with OGs and a 3.9% correlation with TSs (Table 5).

**Table 5:** Union RBP genes summary; binding site, AU-rich and correlation to oncogenes and tumor suppression genes.

Symbol	Descriptions	Motif	Binding site	STRING: PPI with elavl/TTP	% ONCO correlation	% TS correlation
CDKN2A	Cyclin dependent kinase inhibitor 2A	unknown	N/A	none	11.37	4.07%
MEX3A	RNA-binding protein MEX3A	AURICH motif	KHx2; Znf_CCCHx1	none	**	**
RPL39L	Ribosomal protein L39 like	unknown	N/A	none	0.26	4.26
VARS	Valine-tRNA ligase, Aminoacyl tRNA synthetases, class 1	unknown	GST c-terminal	none	2.38	0.74
GSPT1	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A	AURICH motif	N/A	none	2.38	3.89
SNRPE	Small nuclear ribonucleoprotein E	unknown	LSmx1	elavl	**	**
SSR1	Translocon associated protein subunit alpha	AURICH motif	N/A	none	6.61	7.04
TIA1	Nucleolysin isoform p40	TIA-1 AURICH motif	UUUUUGU/RRMX3	elavl	8.99	6.11
PPARGC1B	Peroxisome proliferator-activated receptor gamma coactivator 1-beta	AURICH motif	RRMx1	none	**	**

<i>EIF4E3</i>	Karyotic translation initiation factor 4E type 3	AURICH motif	N/A	none	7.14	8.15
<i>SMAD9</i>	Mothers against decapentaplegic homolog 9	AURICH motif	N/A	none	**	**

Signal-Sequence Receptor 1 (SSR1) is part of an SSR complex known as the Translocon-Associated Protein Complex (TRAP). SSR1 or the TRAP- $\alpha$  subunit is one of four TRAP subunits. The primary function of TRAP is protein-specific transportation across the Endoplasmic Reticulum (ER) membrane. The overexpression of SSR1 in treated cancers may lead to the release of the translated genes through the ER rather than to their final destinations, where they would usually play specific roles. TRAP had a 6.6% correlation with OGs and a 7.0% correlation with TSs, which could be the target genes for SSR1. SSR1 controls these TSs by accumulating them in the ER under certain pathological and/or physiological conditions (Table 5). The discrepancy between the genes cellular function, expression correlations, survival-associated factors and HRs necessitates further studies to understand the cross-functional correlation.

The second signaling pathway relation cluster of the union RBPs was composed of T-cell Intracellular Antigen 1 (TIA1) and Small Nuclear Ribonuclear Protein Polypeptide E (SNRPE). Both were associated with poor survival and low-risk HRs. The network represented the binding of TIA1 with the SNRP family, showing the critical function of TIA1 in the complex formation of the SNRP family. TIA1 is an RNA-binding protein that is considered to be TS and is involved in carcinogenesis. MiR-19a is involved in the destabilization of TIA1 mRNA by binding directly to the 3'UTR of TIA1 mRNA. It controls the translation of the target genes by binding and collocating those genes into the Stress Granules (SGs). DNA damage leads to the release of p53 from SGs due to the dissociation of TIA1. TIA1 mutation is implicated in the delay of SG disassembly and the accumulation of non-dynamic SGs and it is involved in neurodegenerative diseases, such as Amyotrophic Lateral Sclerosis (ALS). TIA1 is also directly involved in the tau oligomer-mediated pathway. TIA1 reduces the number and size of SGs, protecting against neurodegeneration and prolonging the survival of transgenic O301S tau mice and tau oligomer aggregation [18]. It is AU-Rich at the UTR site and has the potential to interact with Elavl1/HuR. It also recognizes the UUUUUGUI RRMX3 binding site motifs. It has a 9.0% correlation with OGs and a 6.11% correlation with TGs (Table 5).

SNRPE is part of a cellular spliceosome complex and plays a critical role in mRNA maturation. The down-regulation of SNRPE is implicated in the dramatic reduction in mTOR mRNA and protein levels and the induction of autophagy. It also plays a role in cell proliferation in prostatic cancers, which indicates its oncogenic effects. It directly regulates the Androgen Receptor (AR) and this explains its involvement in cellular proliferation [17]. SNRPE overexpression is also associated with

lung (adenocarcinoma) prognosis and pathogenesis. Its expression level and role in cell proliferation and invasion in treated cancer patients explain its association with poor survival and low-risk HR. This study found a potential protein-protein interaction between SNRPE and Elavl1/HuR was found, but this requires more experimental investigation (Table 5).

The non-signaling pathway-relation genes included *MEX3A*, *PPARGC1B*, *EIF4E3* and *SMAD9*. These genes were all associated with poor survival and low-risk HR except *MEX3A*, which was associated with poor survival and high-risk HR. *MEX3A* is a putative RBP that regulates CDX2 levels and plays a key role in intestinal differentiation, polarity and stemness contributing to cellular homeostasis and carcinogenesis [13]. *MEX3A* reverses the effects of chemotherapy and irradiation by regenerating the damaged crypts. This may explain why *MEX3A* levels are high in treated patients compared to untreated patients and why treated patients have low-risk HR when undergo this type of treatment. *MEX3A* contains an AU-rich motif at the UTR site, and it binds to KHx2, znf and CCChx1 binding sites of target genes.

The peroxisome Proliferator-activated receptor Gamma Co-Activator 1 Beta (PGC1B) is encoded by the *PPARGC1B* gene. AMP-activated kinase promotes aberrant PGC1B expression in human colon cancer cells. PGC1A and PGC1B methylation are early cancer risk biomarkers; it recognizes the RRMK1 binding site motif. Here, we showed a correlation between the downregulation of PGC1B and poor survival associated with a low risk of death and cancer relapse (Table 5). Further investigation is needed.

Eukaryotic translation Initiation Factor 4E type 3 (*EIF4E3*) belongs to the *EIF4E* protein family, which comprises *EIF4E 1*, *2* and *3*. They play essential roles in the initiation of protein translation that occurs in mRNA metabolism, proliferation, survival, invasion and metastases. In particular, *EIF4E3* binds to the positively charged m7G cap to compete with other factors and function as TS. The reduction of *EIF4E3* in high *EIF4E* cancers suggests that *EIF4E3* is a clinically relevant inhibitory mechanism lacking in some malignancies. In parallel survival analysis in breast cancer patients, the overexpression of certain genes, including *EIF4E3*, improved survival rates. In fact, the phosphorylation of *EIF4E1* has been implicated in the initiation of oncogenic mRNA translation. Enhanced *EIF4E3* expression competes with *EIF4E1* and suppresses *EIF4E1*-driven translation, which reveals a novel role of *EIF4E3* in translation initiation biology [19]. This study showed a 7.0% correlation with OGs and an 8.0% correlation with TSs. This correlation competition could explain the dual role of *EIF4E3* in

controlling the induction of genes under certain physiological and pathological conditions.

SMAD9 belongs to a family of nine proteins (SMAD 1-9), which is divided into three subgroups: receptor SMAD proteins (R-SMADs), which comprise of SMADs 1, 2, 3, 5 and 9 (8), cofactor SMAD proteins (Co-SMADs), comprising only of SMAD4 and inhibitor SMAD proteins (I-SMADs), which comprise of SMADs 6 and 7. SMADs are transcriptional regulators involved in intracellular TGF- $\beta$  family signaling. The major role of SMAD9 is to suppress target gene transcription by competing with SMAD1 to form a complex to bind to DNA [20]. We found that the downregulation of SMAD9 in treated cancer patients was associated with poor survival and a low risk of death.

*MEX3A*, *GSPT1*, *SSR1*, *TIA1*, *PGC1B*, *EIF4E3* and *SMAD9* contained AU-rich elements (AREs) at the 3'UTR mRNA site (Table 5). AREs are cis-acting instability and translation inhibition elements present in the 3'UTR and introns of most inducible genes. HuR/Elavl1 is associated with intronic AREs that are involved in posttranscriptional regulation in more than half of human genes. Abnormal ARE-mediated posttranscriptional control is associated with several abnormal cellular processes that underlie carcinogenesis. RBPs, TTP and HuR have antagonistic roles in mRNA regulation in metastatic breast cancer; TTP destabilizes mRNA and suppresses protein translation, while HuR is an mRNA-binding and translation-promoting component. Low TTP/HuR mRNA ratios are associated with poor survival in breast cancer patients and high levels of mitotic ARE-mRNA signatures, which highlighting the role of AREs and their binding proteins in cancer. Further investigation is required to understand better the relationship between the TTP/HuR ratio and the stability of *MEX3A*, *GSPT1*, *SSR1*, *TIA1*, *PGC1B*, *EIF4E3* and *SMAD9* expression levels.

We used a computational bioinformatics analysis tool on patient data to identify the union RBP gene signatures and their association with disease relapse and treatment failure in breast, colon and lung cancer. These RBPs may be involved in the occurrence, development, invasion, metastasis and drug resistance of cancer. The antiangiogenic treatments could have negative or positive results depending on ethnicity, gender and epigenetic diversity. For instance, Sunitinib, an antiangiogenic drug, has been implicated in stimulating the VEGFC-dependent pathway, activation of HuR and the inactivation of TTP in promoting lymphangiogenesis in ccRCC treated patients, leading to disease relapse, treatment failure and death.

The survival analysis suggested that *CDKN2A*, *MEX3A*, *RPL39L*, *VARS*, *GSPT1*, *SNRPE*, *SSR1*, *TIA1*, *PPARGC1B*, *EIF4E3* and *SMAD9* might have prognostic values for treated cancer patients [17]. The upregulation of *CDKN2A*, *MEX3A*, *RPL39L* and *VARS* in treated patients indicated poor survival, a high risk of death and inefficient treatment plans. The up-regulation of *GTP1*, *SNRPE*, *SSR1* and *TIA1* and/or down-regulation of *PPARGC1B*, *EIF4E3* and *SMAD9* indicated poor survival, a low risk of death and efficient treatment plans. However, the role of these genes in cancer development and drug resistance is still unclear. Further *in vitro* and *in vivo* studies are needed to verify

their functions and contribution to cancer development and drug resistance and the influence of ethnicity, gender and epigenetic diversity on drug efficiency.

## CONCLUSION

In conclusion, the results illustrate that our extended subtyping framework, by combining subtyping and subtype-specific biomarkers, may lead to improved patient prognostication, may form a strong basis for future studies and could potentially be applied as a personalized diagnostic test panel for routine laboratory tests.

## ACKNOWLEDGEMENT

The support of the research and innovation administration at King Faisal Specialist Hospital and Research Centre is highly appreciated. We grateful to the generosity of Dr. Balazs Gyorffy in sharing the RBPs data in breast, colon and lung cancers patient. We also appreciate all the online dataset providers and software who help us to conduct our study.

## ETHICAL APPROVAL

Not applicable.

## COMPETING INTERESTS

The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this manuscript.

## AVAILABILITY OF DATA AND MATERIALS

The Authors confirm that the data used in supporting this study are either cited or acknowledging within the article and/or its supplementary materials.

## REFERENCES

1. Dyba T, Randi G, Bray F, Martos C, Giusti F, Nicholson N, et al. The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. *Eur J Cancer*. 2021;157:308-347.
2. Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, et al. Extraction and analysis of signatures from the gene expression omnibus by the crowd. *Nat Commun*. 2016;7:12846.
3. Carioli G, Malvezzi M, Rodriguez T, Bertuccio P, Negri E, La Vecchia C. Trends and predictions to 2020 in breast cancer mortality in Europe. *Breast*. 2017;36:89-95.
4. Gerstberger S, Hafner M, Ascano M, Tuschl T. Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Adv Exp Med Biol*. 2014;825:1-55.
5. Kim MY, Hur J, Jeong S. Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB Rep*. 2009;42:125-130.
6. Lukong KE, Chang K, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. *Trends Genet*. 2008;24:416-425.

7. Musunuru K. Cell-specific RNA-binding proteins in human disease. *Trends Cardiovasc Med.* 2003;13:188–195.
8. Neelamraju Y, Hashemikhabir S, Janga SC. The human RBPome: From genes and proteins to human disease. *J Proteomics.* 2015;127:61–70.
9. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet.* 2014;15:829–845.
10. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012;487:330–337.
11. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* 2012;72:100–111.
12. Xu J, Liu H, Yang Y, Wang X, Liu P, Li Y, et al. Genome-wide profiling of cervical RNA-binding proteins identifies human papillomavirus regulation of RNASEH2A expression by viral E7 and E2F1. *mBio.* 2019;10:e02687-e02718.
13. Pereira B, Sousa S, Barros R, Carreto L, Oliveira P, Oliveira C, et al. CDX2 regulation by the RNA-binding protein MEX3A: impact on intestinal differentiation and stemness. *Nucleic Acids Res.* 2013;41:3986–3999.
14. Wong QW-L, Li J, Ng SR, Lim SG, Yang H, Vardy LA. RPL39L is an example of a recently evolved ribosomal protein paralog that shows highly specific tissue expression patterns and is upregulated in ESCs and HCC tumors. *RNA Biol.* 2014;11:33–41.
15. Friedman J, Smith DE, Issa MY, Stanley V, Wang R, Mendes MI, et al. Biallelic mutations in valyl-tRNA synthetase gene VARS are associated with a progressive neurodevelopmental epileptic encephalopathy. *Nat Commun.* 2019;10:707.
16. Nair S, Bora-Singhal N, Perumal D, Chellappan S. Nicotine-mediated invasion and migration of non-small cell lung carcinoma cells by modulating STMN3 and GSPT1 genes in an ID1-dependent manner. *Mol Cancer.* 2014;13:173.
17. Anchi T, Tamura K, Furihata M, Satake H, Sakoda H, Kawada C, et al. SNRPE is involved in cell proliferation and progression of high-grade prostate cancer through the regulation of androgen receptor expression. *Oncol Lett.* 2012;3:264–268.
18. Apicco DJ, Ash PEA, Maziuk B, LeBlang C, Medalla M, Al Abdullatif A, et al. Reducing the RNA binding protein TIA1 protects against tau-mediated neurodegeneration in vivo. *Nat Neurosci.* 2018;21:72–80.
19. Landon AL, Muniandy PA, Shetty AC, Lehrmann E, Volpon L, Houg S, et al. MNKs act as a regulatory switch for eIF4E1 and eIF4E3 driven mRNA translation in DLBCL. *Nat Commun.* 2014;5:5413.
20. Tsukamoto S, Mizuta T, Fujimoto M, Ohte S, Osawa K, Miyamoto A, et al. Smad9 is a new type of transcriptional regulator in bone morphogenetic protein signaling. *Sci Rep.* 2014;4:7596.