Scrutinizing functional interaction networks from RNA-binding proteins to their targets in cancer

Sajal Kumar*
Department of Computer Science
New Mexico State University
Las Cruces, USA
sajal49@nmsu.edu

Hua Zhong*
Department of Computer Science
New Mexico State University
Las Cruces, USA
huazhong@nmsu.edu

Ruby Sharma
Department of Computer Science
New Mexico State University
Las Cruces, USA
ruby49@nmsu.edu

Yiyi Li
Department of Computer Science
New Mexico State University
Las Cruces, USA
gtarex@nmsu.edu

Mingzhou Song
Department of Computer Science
Molecular Biology Graduate Program
New Mexico State University, Las Cruces, USA
joemsong@nmsu.edu

***Abstract*—RNA-binding proteins (RBPs) participate in all stages of RNA life cycle from transcription, splicing, to translation. Under the ENCODE project, a large number of RBPs were knocked down in human cancer cell lines, offering an excellent opportunity to infer targets of RBPs. Taking both RBP binding sites and RNA-seq profiles of RBP knockdown samples as input, we present a pipeline to identify causal RBP→RNA interactions. The pipeline employs a recent functional chi-square test (FunChisq) that deciphers directional association, and utilizes a novel functional index that measures the effect size of functional dependency. We examined ~45 million RBP→RNA pairs in leukemia (K562) and liver cancer (HepG2) cell lines for functional patterns as causal interaction candidates. Here, we report a total of 936,707 RBP→RNA pairs in the two cell lines that show statistically significant linear or nonlinear functional patterns. About 31% of these pairs have supportive biological evidence from other sources, suggesting the effectiveness of the pipeline. The interactions constitute RBP specific regulatory networks that may potentially represent core mechanisms in the two cancers. The pipeline is implemented through an R interface with precomputed results and data libraries for users to query specific networks and visualize RBP→RNA interactions. Such networks serve as a useful resource for studying RNA dysregulation in cancer.**

***Index Terms*—functional dependency, causal interactions, RNA-binding protein, RNA regulatory network, liver cancer, leukemia.**

I. Introduction

A transcribed pre-mRNA is subject to further processing mediated by RNA-binding proteins (RBPs) [1]. These processes include splicing, editing, and polyadenylation, in which an RBP plays important regulatory roles. Anomaly in RBP functioning within human cells can lead to disorders such as neurodegeneration, autoimmune diseases, and cancer [2].

By obtaining physical evidence of direct RBP-RNA binding via the cross-linking immunoprecipitation (CLIP) technology, previous studies inferred protein-RNA interactions [3]–[6]. However, binding activity does not necessarily lead to changed gene expression.

Transcriptome profiling in response to perturbed RBPs can offer necessary dynamics to identify direct/indirect causal RBP→RNA relationships. The ENCODE consortium [7] facilitates such dynamic RNA-seq profiles in liver cancer and leukemia cell lines by shRNA induced silencing of ~250 RBPs. They also mapped genome-wide binding sites of the RPBs with a high accuracy [2] using enhanced crosslinking immunoprecipitation (eCLIP).

To process the perturbed RBP datasets from ENCODE, we introduce a network inference pipeline to isolate potential causal RBP→RNA interactions. In the pipeline, we chose the functional chi-square (FunChisq) test [8], [9] for the task of functional dependency inference. FunChisq is a model-free method without an assumption on the parametric form of a function. Using FunChisq for causal inference is based on the *causality-by-functionality principle*, which Nobel laureate Herbert A. Simon and philosopher Nicholas Rescher stated as that a causal relation between variables is "a function of one variable (cause) on to another (the effect)" [10]. A related exact functional test [11] has been established to test directional association. It uses the same statistic but computes an exact p-value instead of estimating it from an asymptotic null distribution. We also utilize a novel functional index derived from FunChisq to measure the effect size—the strength of functional dependency. Using both the effect size and the significance level of functional association caters better functional patterns that are strong and unlikely arise by chance.

We examined a total of ~45 million RBP→RNA pairs and report top 936,707 significant functional patterns, out of which 259,979 RBP→RNA are direct interactions with eCLIP binding evidence, 25,019 are consistent with Pathway Commons interaction database and an additional 651,709 patterns are testable hypotheses of indirect/direct effect for future studies. We provide directed RNA regulatory networks of a few top-

*: Authors of equal contribution.

978-1-5386-5488-0/18/$31.00 ©2018 IEEE

tier interactions in leukemia and liver cancer cell lines. The entire analysis has been implemented into a pipeline and wrapped into an interactive R program that generates networks and interaction scatter plots for RBPs of interest using pre-processed data libraries. Our work thus provides a resource to study RNA regulatory networks centered around RBPs in cancer.

II. Methods

Our overarching approach is based on a recent statistical method, FunChisq, which makes no assumption on the underlying functional form in inferring functional relationships. It identifies promising interaction patterns that are directional and maximize functional dependency. The statistical significance is based on the p-value of the FunChisq test and the effect size is measured by the functional index.

A. Functional chi-square test

Given n samples of discrete random variables X and Y of r and s levels respectively, we test whether Y is a non-constant function of X, i.e., $Y = f(X)$ and $Y \neq c$ ($c \in \mathbb{R}$). If Y is a function of X, we call X parent variable and Y child variable.

The test operates on contingency tables with observed counts of variable X and Y. In a $r \times s$ contingency table, let n_{ij} be the observed count of $X = i$ and $Y = j$, $i = 1, \ldots, r$ and $j = 1, \ldots, s$. Let $n_{i\cdot}$ be the sum of row i and $n_{\cdot j}$ be the sum of column j.

Definition 1. *Given n samples of variables X and Y, the FunChisq statistic is defined by [8], [11]*

$$\chi^2_f(X \to Y) = \left[\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(n_{ij} - n_{i\cdot}/s)^2}{n_{i\cdot}/s}\right] - \left[\sum_{j=1}^{s} \frac{(n_{\cdot j} - n/s)^2}{n/s}\right] \quad (1)$$

The theoretical foundation of FunChisq is described in [8], [11]. We summarize the main results here. $\chi^2_f(X \to Y)$ asymptotically follows a chi-square distribution with $(r-1) \times (s-1)$ degrees of freedom under the null hypothesis that Y is independent of X and the assumption that Y is uniformly distributed. The statistic is minimized to zero when X and Y are empirically independent.

The FunChisq statistic has asymmetric functional optimality—given the column (Y) sums, it is maximized only if Y is a non-constant function of X (Theorem 8 [8]). This property is not claimed by any other known causal inference method to the best of our knowledge. FunChisq has recently outperformed state-of-the-art techniques in a causal network inference challenge [9]. The functional chi-square test represents a paradigm shift from Pearson's chi-square test [12] by differentiating the direction of association.

B. Functional index

In addition to using the p-value for statistical significance, some measure of effect size is important in hypothesis testing [13]. In the FunChisq test, the effect size needs to reflect the extent of Y being a function of X. We introduce the functional index to measure the strength of functional dependency of Y on X.

Definition 2. *We define the functional index ξ_f by the ratio of FunChisq statistic $\chi^2_f(X \to Y)$ to the maximum attainable FunChisq subtracting the column chi-square:*

$$\xi_f = \sqrt{\frac{\chi^2_f(X \to Y)}{n(s-1) - \chi^2(Y)}} \quad (2)$$

where $\chi^2(Y)$ is the column marginal chi-square of Y defined within the second pair of square brackets in Eq. (1).

When a contingency table is multiplied by a linear factor, the FunChisq statistic in the numerator, the maximum attainable FunChisq, and the marginal chi-square of Y in the denominator all scale linearly with the factor. Thus $\xi_f \in [0, 1]$ is independent of linear scaling of a pattern, making it suitable to capture the effect size of a function pattern. A larger ξ_f indicates stronger functional dependency.

C. A pipeline to scrutinize RBP→RNA interactions

We integrate the two statistics into a pipeline for delineating RNA regulatory networks. In the context of RBP→RNA interactions, the pipeline promotes patterns where the transcript abundance of an RNA is similar to a non-constant function of the transcript abundance of a RBP. This is distinctively different from associative statistical methods such as Pearson's chi-square, correlation, or mutual information. Figure 1 shows the pipeline for scrutinizing RBP→RNA interactions in a given cell line. The input to the pipeline includes:

i Transcript quantification files obtained from the RNA-seq data of each RBP knockdown sample.
ii RBP binding sites from eCLIP data.

Six main steps of the pipeline are summarized below:

1) Form a matrix of transcript expression (in TPM) for all RNA transcripts.
2) Log-transform the data iteratively until a followup Shapiro test [14] reports a statistic greater than or equal to 0.9 which indicates that the log-transformed gene expression is normally distributed. If log-transform decreases or makes no change to the Shapiro test statistic of gene, no log-transform is applied.
3) Discretize the log-transformed data for each transcript using the R package '*Ckmeans.1d.dp*' [15], [16]. The number of discrete levels is computed by the R package '*mclust*' [17] based on Bayesian information criterion.
4) Pick a representative transcript for each parent RBP by choosing the major transcribed isoform with a maximum median expression. The representative transcript helps in integrating the transcript expression profile to the gene binding site profile offered by eCLIP.
5) Apply the *FunChisq* test on all possible pairs of RBP and RNA transcripts. Those pairs with p-value ≤ 0.05 and ξ_f in the top 1% of all interactions ($\xi_f \geq 0.48$) are considered potential causal relationships.

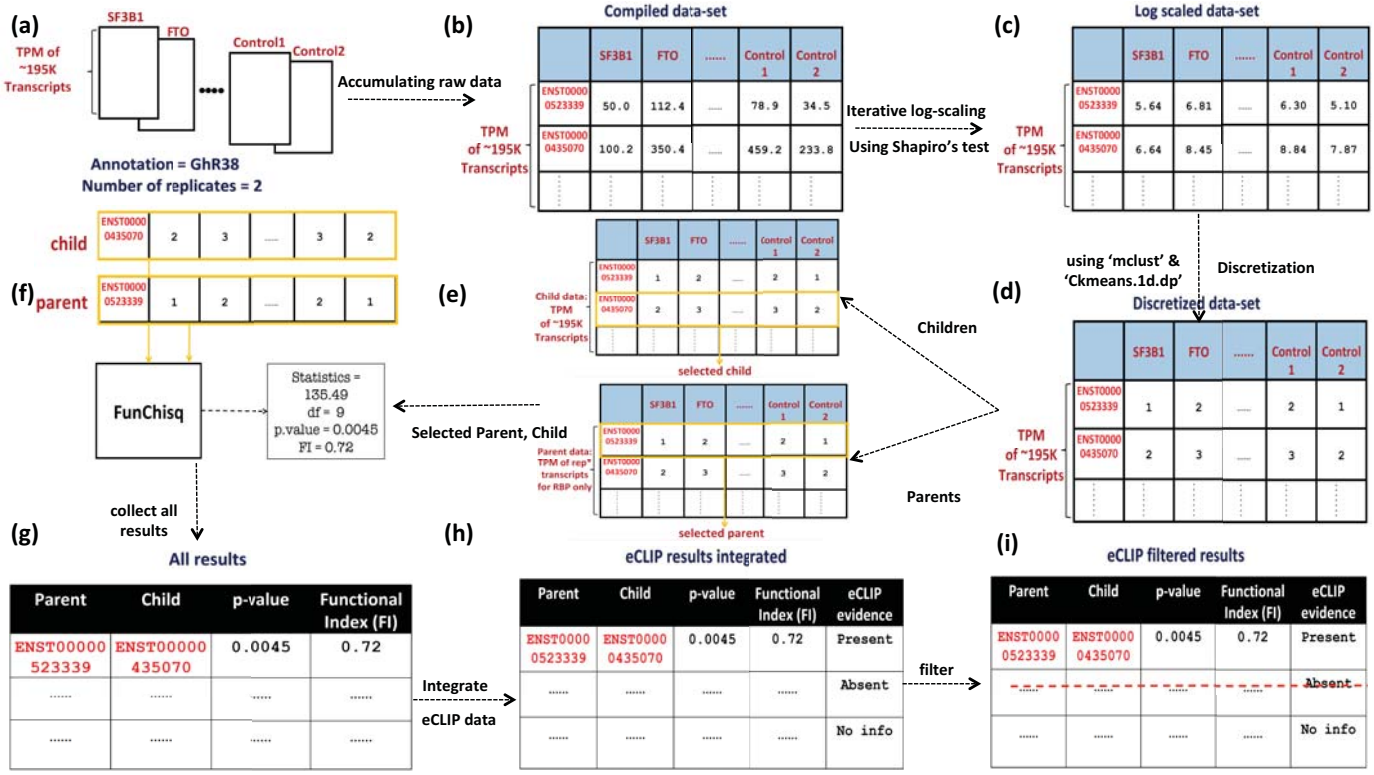

Fig. 1. **Overview of major steps taken to scrutinize RBP→RNA interactions**. The star * representative transcript for each RBP was picked from the log transformed data by selecting the transcript with the maximum median expression.

6) The filtered RBP-RNA pairs were placed into one of the three categories: 1) absolute direct interactions, if the eCLIP data for the RBP shows binding evidence to the RNA; 2) hypothetical indirect interaction, if the eCLIP data for the RBP shows no binding to the RNA (these were removed from consideration) and 3) hypothetical direct/indirect interactions, if there is no eCLIP available for the RBP.

The output of the pipeline is a list of statistically significant RBP → RNA interactions.

D. Implementation

Both the FunChisq test and functional index are available in the R package '*FunChisq*' [18], freely downloadable from Comprehensive R Archive Network (CRAN). The pipeline was implemented in the R programming language [19]. It offers an easy to use interface with pre-computed results and data libraries that can generate and visualize networks for RBPs and RNAs of interest. The source code and auxiliary data are provided in **Additional file 2**.

III. Results

A. Cancer transcriptome in response to RNA-binding protein knockdown

The shRNA knockdown RNA-seq dataset from ENCODE [7] has a total of 221 unique RBPs knocked down in myelogenous leukemia (K562) with 22 wild type samples (control) and 225 unique RBPs knocked down in hepatocellular carcinoma (HepG2) with 24 wild type samples. Each sample in both cell lines had 2 or 4 replicates with RNA abundance in the unit of TPM measured for about 195K RNA transcripts. We only considered samples that were annotated with the latest GRCh38 human reference genome.

We integrated eCLIP data from ENCODE to filter RBP→RNA interactions that also have evidence of physical binding. The eCLIP data include 174 samples collected from 87 knockdown RBPs in K562 cell line and 140 samples for 70 RBPs in HepG2 cell line. We removed those RBP→RNA interactions that had insignificant binding or no binding according to the available eCLIP data. The interactions that had no eCLIP data available were considered predictions. Finally we report a list of all interactions with functional index ξ_f in the top 1% of all interactions ($\xi_f \geq 0.48$) and adjusted p-value ≤ 0.05 in **Additional file 1**.

B. Summary of statistically significant RBP→RNA interactions in cancer

Out of about 45 million total RBP-RNA pairs, approximately 13 million in the HepG2 cell line were found to have a p-value less than 0.05 after Benjamini-Hochberg correction [20]. Taking the top 1% based on the effect size, functional index ($\xi_f \geq 0.48$), gave 390,166 pairs, out of which 83,490 had eCLIP binding evidence and the rest 306,676 with no binding information available are hypothesized to be indirect/direct interactions.

In the K562 cell line, another approximately 13 million RBP-RNA pairs were found to have an adjusted p-value of less than 0.05. The top 1% ($\xi_f \geq 0.48$) of 546,541 pairs contained 176,489 pairs with eCLIP binding evidence for direct interactions and 370,052 without binding information hypothesized to be direct/indirect interactions.

Out of the 676,728 interaction patterns with no binding information, 25,019 patterns were found consistent with interaction patterns in Pathway Commons interaction database [21]. To report the most promising potential interaction patterns we used the ξ_f scores of top 2162 known interactions (top 5% of 25,019 interactions from Pathway Commons) to further threshold ξ_f at 0.74 which resulted in a list of 16,100 putative interactions (**Additional file 1**).

C. An RNA regulatory network inferred for hepatocellular carcinoma

Figure 2a illustrates a subset of an RNA regulatory network, by choosing top 50 RBP→RNA interactions sorted in decreasing order of functional index (ξ_f), in HepG2 cell line. The network contains five RBPs and 44 target RNAs. DDX6 has strong directional interactions to 28 RNA targets, followed by HNRNPA1 with 18 RNA targets, DDX3X with three RNA targets, and a pair of RBPs, SRSF1 and HNRNPM, that are direct targets of each other. Increasing the number of top interactions picked for the network would subsequently add more RBP→RNA edges to figure 2a. DEAD-box helicase 6 (DDX6) is a known oncogene to B-cell non-Hodgkin lymphoma in COSMIC Cancer Gene Census version 85, but its role in liver cancer is not clear except that it promotes hepatitis C virus replication [22]. Mediated by EGFR, heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1) up-regulates the IR-A:IR-B ratio which signals the proliferative effect in human hepatocellular carcinoma [23]. DEAD-box helicase 3 X-linked (DDX3X) is a known tumor suppressor gene to chronic lymphocytic leukemia and medulloblastoma. Proteomics analysis suggests that DDX3X is up-regulated in liver diseases including liver cancer [24].

Figure 2b–e shows four RBP→RNA interaction patterns that exhibit functional dependency. They are reinforced by eCLIP physical binding and have literature support from previous studies. Reportedly, HNRNPA1 and SRSF11 (figure 2b) are both splicing repressors [25] and can inhibit SMN2 exon 7 inclusion [26]. In Figure 2c, when HNRNPK is knocked down, the child RPL10 has no consequential expression which indicates causal association. The overall pattern suggests that HNRNPK may function as a switch. Indeed, RPL10 mutation carries 62 significantly regulated candidates, HNRNPK being one of them [27]. In Figure 2d, both HNRNPA1 and NCBP2 belong to the human spliceosome pathway (hsa03040) [28]. In Figure 2e, over-expression of HNRNPM promotes exon 7 inclusion of both SMN2 and SMN1 pre-mRNA. It contacts an enhancer on exon 7, which was previously shown to provide a binding site for TRA2B. Evidently HNRNPM and TRA2B contact an overlapped sequence on exon 7 [29].

D. An RNA regulatory network inferred for leukemia

Figure 3a illustrates a subset of an RNA regulatory network, by choosing top 50 RBP→RNA interactions sorted in the decreasing order of ξ_f, in K562 cell line. The network highlights NONO and NPM1 as the two RBPs having the strongest functional influence on 32 and 18 RNAs, respectively. Their targets share 13 common RNAs suggesting some synergy between the two RBPs. Increasing the number of top interactions picked for the network would subsequently add more RBP→RNA edges to figure 3a. Non-POU domain containing octamer binding (NONO) gene is recently shown to promote migration and invasion of THP-1 cell line [30] derived from an acute monocytic leukemia patient. Nucleophosmin 1 (NPM1) is a clinically tested molecular genetic marker for acute myeloid leukemia [31]. It is categorized as oncogene and fusion gene in COSMIC Cancer Gene Census version 85.

Figure 3b–e shows visualizes four RBP→RNA interaction patterns that exhibit functional dependency, are reinforced by eCLIP evidence and have literature support. In Figure 3b, both NONO and EIF3L are involved in the functional network of molecular transportation of human high choroid plexus epithelium expression gene sub-dataset [32]. In Figure 3c, EIF4G2 and SRRM1 are both included in the human RNA-transport pathway (hsa030313) [33]. Related to the NPM1→RPL31 and NPM1→RPL10A interactions in Figure 3d and 3e, three RNAs coding for ribosomal protein, RPL31, RPL10A and RPL36A were found to be exclusively present in chronic myelogenous leukemia-specific NPM1 co-expressed gene pairs and were absent in normal-specific co-expressed gene pairs. This is interesting as NPM1 protein is a well-recognized key player in ribosome biogenesis and transport [34].

It is interesting that the top 50 interactions centers around only two RBPs in RBP→RNA networks of both liver cancer (Figure 2a) and leukemia cell line (Figure 3a). The two networks apparently do not share the most significant RBP→RNA interactions.

IV. Conclusions

In summary, we provide a pipeline that exploits the dynamics in ENCODE shRNA knockdown data sets to identify causal relationships. The pipeline uses FunChisq and a novel functional index that reveal functional RBP→RNA interaction patterns in HepG2 and K562 cell lines. With adjusted p-value ≤ 0.05, we chose top 1% of all interactions ($\xi_f \geq 0.48$) to recapture 259,979 direct RBP→RNA interactions with eCLIP binding evidence, 25,019 known interactions and 651,709 predicted putative interactions out of which 16,100 are most promising. The networks provide a map of how RNAs are regulated by RBPs in cancer from the ENCODE data. The most active interactions differ greatly between the two different cancer types we studied. We revealed potentially new RBPs involved in cancer and the mechanism by which they may influence other genes. Lastly, the R interface provided is a useful visualization tool to study the behavior of RBP(s) and RNA(s) of interest as captured in the ENCODE data sets. Both the computational tools and the RNA regulatory networks

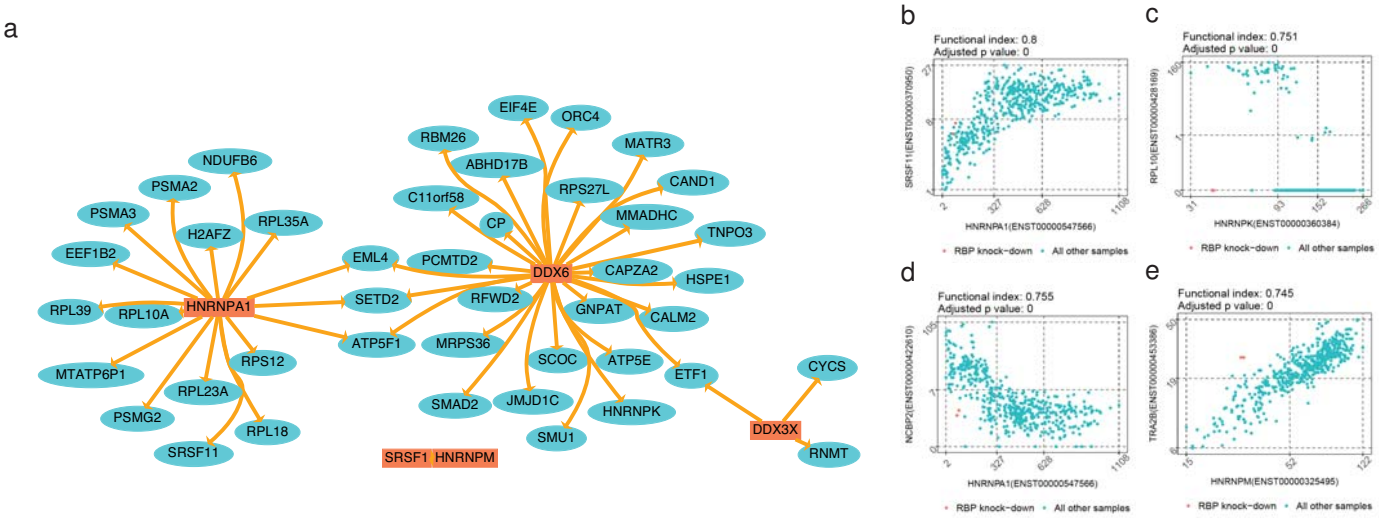

Fig. 2. **An RNA regulatory network and four significant functional RBP→mRNA interaction patterns detected in HepG2 liver cancer.** a) The directed network contains 50 direct interactions among five RBPs and 44 RNAs. The red box nodes represent RBPs. The light blue round nodes represent RNA transcripts. The orange solid arrows represent inferred direct RBP → RNA interactions with eCLIP binding evidence. b) HNRNPA1 promotes SRSF11. c) HNRNPK has a switch-like influence on RPL10. d) HNRNPA1 inhibits NCBP2. e) HNRNPM linearly promotes TRA2B.

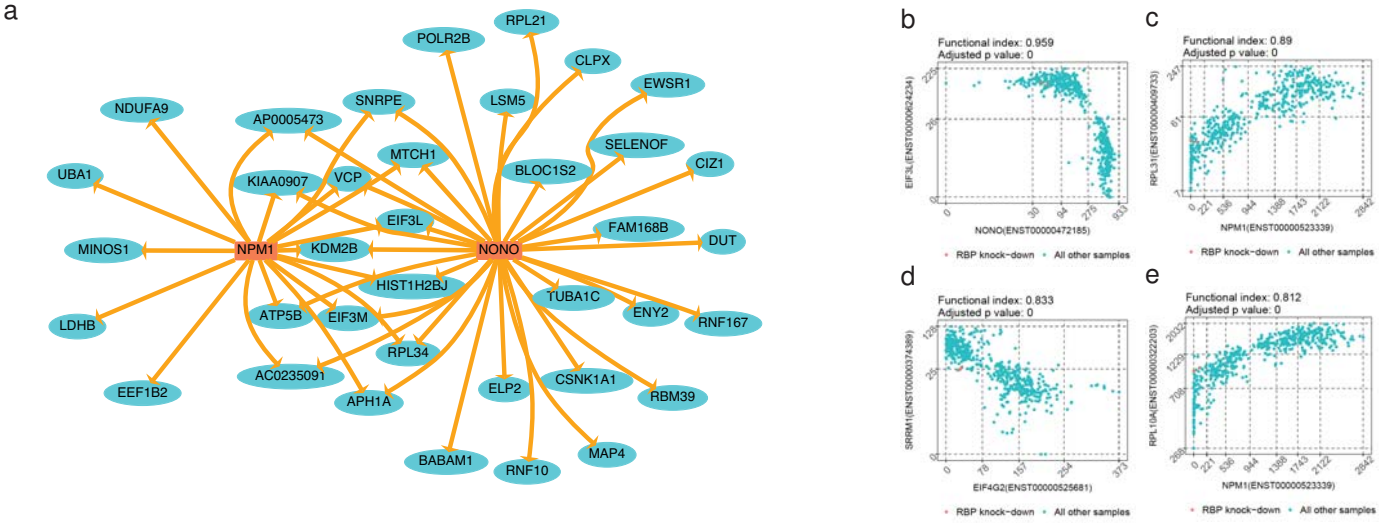

Fig. 3. **An RNA regulatory network and four significant RBP → mRNA functional interaction patterns detected in K562 leukemia cell line.** a) The directed network contains 50 direct interactions among two RBPs and 37 RNAs. The red box nodes represent RBPs. The light blue round nodes show RNA transcripts. The orange solid arrows represent inferred direct RBP → RNA interactions with eCLIP binding evidence. b) NONO inhibits EIF3L. c) EIF4G2 inhibits SRRM1. d) NPM1 promotes RPL31. e) NPM1 promotes RPL10A.

reported here serve as a resource to study gene regulation and cancer systems biology.

Acknowledgments

The reported work was supported by US National Science Foundation grant 1661331 and USDA grant 2016-51181-25408.

Online Supplementary Files

1) **Additional_file_1_interactions-signif.tsv (79MB)** — All significant interactions between RBPs and RNAs, with adjusted p-value ≤ 0.05 and conditional function index $\xi_f \geq 0.48$. The file can be downloaded at
https://www.cs.nmsu.edu/~joemsong/BIBM2018/Additional_file_1_interactions-signif.tsv
2) **Additional_file_2_ENCODE-RBP-to-mRNA_V1.zip (333MB)** — This file contains an R interface with pre-compiled data libraries and functions that can be used to generate scatterplots and network of RBPs and RNAs specified by user input. The file can be downloaded at
https://www.cs.nmsu.edu/~joemsong/BIBM2018/Additional_file_2_ENCODE-RBP-to-mRNA_V1.zip

References

[1] T. Glisovic, J. L. Bachorik, J. Yong, and G. Dreyfuss, "Rna-binding proteins and post-transcriptional gene regulation," *FEBS Letters*, vol. 582, no. 14, pp. 1977–1986, 2008.

[2] E. L. Van Nostrand, G. A. Pratt, A. A. Shishkin, C. Gelboin-Burkhart, M. Y. Fang, B. Sundararaman, S. M. Blue, T. B. Nguyen, C. Surka, K. Elkins *et al.*, "Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)." *Nature Methods*, vol. 13, no. 6, pp. 508–514, 2016.

[3] J.-H. Li, S. Liu, H. Zhou, L.-H. Qu, and J.-H. Yang, "starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-seq data," *Nucleic Acids Research*, vol. 42, no. D1, pp. D92–D97, 2013.

[4] J. König, K. Zarnack, N. M. Luscombe, and J. Ule, "Protein–RNA interactions: new genomic technologies and perspectives," *Nature Reviews Genetics*, vol. 13, no. 2, p. 77, 2012.

[5] R. B. Darnell, "HITS-CLIP: panoramic views of protein–RNA regulation in living cells," *Wiley Interdisciplinary Reviews: RNA*, vol. 1, no. 2, pp. 266–286, 2010.

[6] M. Ascano, M. Hafner, P. Cekan, S. Gerstberger, and T. Tuschl, "Identification of RNA–protein interaction networks using PAR-CLIP," *Wiley Interdisciplinary Reviews: RNA*, vol. 3, no. 2, pp. 159–177, 2012.

[7] E. P. Consortium *et al.*, "An integrated encyclopedia of dna elements in the human genome," *Nature*, vol. 489, no. 7414, pp. 57–74, 2012.

[8] Y. Zhang and M. Song, "Deciphering interactions in causal networks without parametric assumptions," *arXiv Molecular Networks*, 2013. [Online]. Available: http://arxiv.org/abs/1311.2707

[9] S. M. Hill, L. M. Heiser, T. Cokelaer, M. Unger, N. K. Nesser, D. E. Carlin, Y. Zhang, A. Sokolov, E. O. Paull, C. K. Wong, K. Graim, A. Bivol, H. Wang, F. Zhu, B. Afsari, L. V. Danilova, A. V. Favorov, W. S. Lee, D. Taylor, C. W. Hu, B. L. Long, D. P. Noren, A. J. Bisberg, The HPN-DREAM Consortium, G. B. Mills, J. W. Gray, M. Kellen, T. Norman, S. Friend, A. A. Qutub, E. J. Fertig, Y. Guan, M. Song, J. M. Stuart, P. T. Spellman, H. Koeppl, G. Stolovitzky, J. Saez-Rodriguez, and S. Mukherjee, "Inferring causal molecular networks: empirical assessment through a community-based effort." *Nat Methods*, vol. 13, no. 4, pp. 310–318, Apr 2016.

[10] H. A. Simon and N. Rescher, "Cause and counterfactual," *Philosophy of Science*, vol. 33, no. 4, pp. 323–340, 1966.

[11] H. Zhong and M. Song, "A fast exact functional test for directional association and cancer biology applications," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, (in press) 10.1109/TCBB.2018.2809743.

[12] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[13] G. M. Sullivan and R. Feinn, "Using effect size—or why the p value is not enough," *Journal of graduate medical education*, vol. 4, no. 3, pp. 279–282, 2012.

[14] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[15] H. Wang and M. Song, "Ckmeans.1d.dp: Optimal k-means clustering in one dimension by dynamic programming," *The R Journal*, vol. 3, no. 2, pp. 29–33, 2011.

[16] J. Song and H. Wang, *Ckmeans.1d.dp: Optimal and Fast Univariate Clustering*, 2018, R package version 4.2.2. https://cran.r-project.org/package=Ckmeans.1d.dp.

[17] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.

[18] Y. Zhang, H. Zhong, R. Sharma, S. Kumar, and J. Song, *FunChisq: Chi-Square and Exact Tests for Model-Free Functional Dependency*, 2018, R package version 2.4.5-2. https://CRAN.R-project.org/package=FunChisq.

[19] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[20] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300, 1995.

[21] E. G. Cerami, B. E. Gross, E. Demir, I. Rodchenkov, Ö. Babur, N. Anwar, N. Schultz, G. D. Bader, and C. Sander, "Pathway commons, a web resource for biological pathway data," *Nucleic acids research*, vol. 39, no. suppl 1, pp. D685–D690, 2011.

[22] R. K. Jangra, M. Yi, and S. M. Lemon, "DDX6 (Rck/p54) is required for efficient hepatitis C virus replication but not for internal ribosome entry site-directed translation," *J Virol*, vol. 84, no. 13, pp. 6810–6824, 2010.

[23] H. Chettouh, L. Fartoux, L. Aoudjehane, D. Wendum, A. Claperon, Y. Chretien, C. Rey, O. Scatton, O. Soubrane, F. Conti, F. Praz, C. Housset, O. Rosmorduc, and C. Desbois-Mouthon, "Mitogenic insulin receptor-A is overexpressed in human hepatocellular carcinoma due to EGFR-mediated dysregulation of RNA splicing factors," *Cancer Res*, vol. 73, no. 13, pp. 3974–3986, Jul 2013.

[24] P. C. Schroder, J. Fernandez-Irigoyen, E. Bigaud, A. Serna, R. Renandez-Alcoceba, S. C. Lu, J. M. Mato, J. Prieto, and F. J. Corrales, "Proteomic analysis of human hepatoma cells expressing methionine adenosyltransferase i/III: Characterization of DDX3X as a target of S-adenosylmethionine." *J Proteomics*, vol. 75, no. 10, pp. 2855–2868, 2012.

[25] D. Dominguez and C. B. Burge, "Interactome analysis brings splicing into focus," *Genome Biology*, vol. 16, no. 1, p. 135, 2015.

[26] C. D. Wee, M. A. Havens, F. M. Jodelka, and M. L. Hastings, "Targeting sr proteins improves smn expression in spinal muscular atrophy cells," *PLoS One*, vol. 9, no. 12, p. e115205, 2014.

[27] A. G. Chiocchetti, D. Haslinger, M. Boesch, T. Karl, S. Wiemann, C. M. Freitag, F. Poustka, B. Scheibe, J. W. Bauer, H. Hintner *et al.*, "Protein signatures of oxidative stress response in a patient specific cell line model for autism," *Molecular Autism*, vol. 5, no. 1, p. 10, 2014.

[28] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "Kegg: New perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Research*, vol. 45, no. D1, pp. D353–D361, 2017.

[29] S. Cho, H. Moon, T. J. Loh, H. K. Oh, S. Cho, H. E. Choy, W. K. Song, J.-S. Chun, X. Zheng, and H. Shen, "hnrnp m facilitates exon 7 inclusion of smn2 pre-mrna in spinal muscular atrophy by targeting an enhancer on exon 7," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1839, no. 4, pp. 306–315, 2014.

[30] X. Zhang, C. Wu, W. Xiong, C. Chen, R. Li, and G. Zhou, "Knockdown of p54nrb inhibits migration, invasion and TNF-alpha release of human acute monocytic leukemia THP1 cells," *Oncol Rep*, vol. 35, no. 6, pp. 3742–3748, Jun 2016.

[31] S. Yohe, "Molecular genetic markers in acute myeloid leukemia," *J Clin Med*, vol. 4, no. 3, pp. 460–478, Mar 2015.

[32] S. F. Janssen, S. J. van der Spek, B. Jacoline, A. H. Essing, T. G. Gorgels, P. J. van der Spek, N. M. Jansonius, and A. A. Bergen, "Gene expression and functional annotation of the human and mouse choroid plexus epithelium," *PLoS One*, vol. 8, no. 12, p. e83345, 2013.

[33] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "Kegg as a reference resource for gene and protein annotation," *Nucleic Acids Research*, vol. 44, no. D1, pp. D457–D462, 2016.

[34] L. W. Chan, X. Lin, G. Yung, T. Lui, Y. M. Chiu, F. Wang, N. B. Tsui, W. C. Cho, S. Yip, P. M. Siu *et al.*, "Novel structural co-expression analysis linking the NPM1-associated ribosomal biogenesis network to chronic myelogenous leukemia," *Scientific Reports*, vol. 5, 2015.