J Human Gen Genom. 2025 May;9(1): A-10-354-3 doi: 10.6118/jhgg.9.1.131
Published online 2026 Jul 3 Research Article

Integrated Transcriptomic and Network Analysis Identifies MILIP as a Central lncRNA Hub in the RNA-Binding Protein Regulatory Landscape of Colorectal Cancer

Ava Asadi[1], Mehrdad Hashemi[2,3*], Maryam Peymani[4], Najmeh Ranji[1], and Maliheh Entezari[3,5]

[1.] Department of Biology, Ra.C., Islamic Azad University, Rasht, Iran.

[2]Department of Genetics, TeMS.C., Islamic Azad University, Tehran, Iran.

[3]Farhikhtegan Medical Convergence Sciences Research Center, Farhikhtegan Hospital, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran.

[4]Department of Biology, Shk.C., Islamic Azad University, Shahrekord, Iran.

[5]Gastroenterology and Liver Disease Research Center, Research Institute for Gastroenterology and Liver Disease, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

***Corresponding Author:** Mehrdad Hashemi, **Email:** mhashemi@iau.ac.ir

Received: 2026/03/16 Accepted: 2026/07/2

Abstarct

Colorectal cancer (CRC) is one of the leading causes of cancer-related mortality worldwide, characterized by complex molecular heterogeneity and dysregulated gene expression programs. Increasing evidence suggests that post-transcriptional regulation mediated by RNA-binding proteins (RBPs) and long non-coding RNAs (lncRNAs) plays a crucial role in tumor initiation and progression. However, the global regulatory architecture of RBP–lncRNA interactions in CRC remains incompletely understood.

In this study, we integrated gene expression profiles of RBPs and lncRNAs to construct a comprehensive co-expression network based on Pearson correlation analysis. Significant RBP–lncRNA interactions were identified using thresholds of $|r| > 0.5$ and $p < 0.05$. The resulting regulatory network was visualized using a force-directed layout, and topological analysis was performed to identify hub genes based on node degree.

Network analysis revealed a highly structured interaction landscape with distinct hub RBPs and lncRNAs, suggesting the presence of coordinated post-transcriptional regulatory modules in CRC. To validate network-derived findings, differential expression analysis was performed on five selected lncRNAs from TCGA datasets (EMSLR, MILIP, RIPK2-DT, RNASEH1-AS1, and . Among these, MILIP was identified as the only overlapping gene between differentially expressed lncRNAs and network hub nodes.

Further analysis of a MILIP-centered subnetwork demonstrated strong and multiple interactions between MILIP and RBPs, indicating its central position within the regulatory network. These findings suggest that MILIP may function as a key regulatory lncRNA in CRC, integrating both transcriptional dysregulation and network-level centrality.

Overall, this integrative analysis provides new insights into the RBP–lncRNA regulatory landscape in CRC and highlights MILIP as a potential candidate for further functional investigation and biomarker development. These findings are hypothesis-generating and require experimental validation in future studies.

Keywords: Colorectal cancer; RNA-binding proteins; long non-coding RNAs; gene co-expression network; hub genes

Copyright© 2022, Journal of Human Genetics and Genomics. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/) which permits copy and redistribute the material just in noncommercial usages, provided the original work is properly cited.

Introduction

Colorectal cancer (CRC) is one of the most frequently diagnosed malignancies and remains a leading cause of cancer-related mortality worldwide [1]. Despite significant advances in screening, surgical techniques, and systemic therapies, a substantial proportion of patients are still diagnosed at advanced stages, where therapeutic options are limited, and prognosis remains poor [2]. This underscores the urgent need to better understand the molecular mechanisms underlying CRC development and progression.

In recent years, cancer research has increasingly shifted toward unraveling the complexity of post-transcriptional gene regulation. Among the key regulators of this layer of gene expression are RNA-binding proteins (RBPs), which play essential roles in RNA splicing, transport, stability, localization, and translation [3]. RBPs are now recognized as critical modulators of cellular homeostasis, and their dysregulation has been widely implicated in tumor initiation, progression, and therapeutic resistance across multiple cancer types, including CRC [4].

Parallel to RBPs, long non-coding RNAs (lncRNAs) have emerged as fundamental regulators of gene expression, despite lacking protein-coding capacity. These transcripts are involved in a wide range of biological processes, including chromatin remodeling, transcriptional regulation, and post-transcriptional control [5]. Importantly, accumulating evidence suggests that lncRNAs can function either as oncogenes or tumor suppressors depending on cellular context, thereby contributing to cancer heterogeneity and progression [6].

Recent studies have highlighted that RBPs and lncRNAs do not function independently, but instead form intricate regulatory networks that collectively shape gene expression programs. RBPs can regulate lncRNA stability and processing, while lncRNAs can, in turn, modulate RBP activity or serve as molecular scaffolds, creating complex feedback and feedforward regulatory loops [7]. However, despite growing interest in these interactions, the global architecture and functional significance of RBP–lncRNA regulatory networks in CRC remain incompletely understood.

With the rapid expansion of high-throughput transcriptomic datasets, network-based approaches have become powerful tools for deciphering complex molecular interactions in cancer. In particular, correlation-based network modeling enables the identification of co-expression patterns that may reflect functional regulatory relationships. Integrating these approaches with large-scale cancer datasets, such as The Cancer Genome Atlas (TCGA), provides an opportunity to identify clinically relevant regulatory molecules and potential biomarkers.

In this study, we constructed a comprehensive RBP–lncRNA interaction network in CRC based on Pearson correlation analysis of expression profiles. We further performed topological analysis to identify hub genes and integrated these findings with TCGA-derived differential expression analysis. Our goal was to uncover key regulatory lncRNAs and RBPs that may play central roles in CRC progression, with a particular focus on identifying functionally significant overlapping candidates across network and expression-based analyses.

METHODS

Data acquisition and preprocessing

Gene expression data of CRC were obtained from The Cancer Genome Atlas (TCGA) COADREAD cohort (combined colon adenocarcinoma and rectal adenocarcinoma) and processed using R software (version 4.x) within the RStudio environment. The dataset comprised RNA-seq expression profiles (HTSeq-FPKM-normalized counts) of 594 tumor samples and 51 adjacent normal samples, including expression data for RBPs and lncRNAs. Expression matrices were imported into R from standard tabular formats (.txt and .csv files).

Data preprocessing included removal of missing values, standardization of gene identifiers, and log2 transformation where applicable. All datasets were checked for consistency prior to downstream analysis.

Differential expression analysis of RBPs

A total of 22 RBPs were selected for differential expression analysis based on their prior documented involvement in post-transcriptional regulation and cancer-relevant functions, as established in the published literature. These candidates were not derived from a genome-wide filtering step but were curated from a predefined panel of RBPs known to regulate RNA splicing, stability, translation, or modification in cancer contexts. Log2 fold change (logFC) values were calculated from tumor versus normal comparisons. RBPs were classified as upregulated (logFC > 0) or downregulated (logFC < 0). Differential expression patterns were visualized using heatmaps and bar plots to represent both global trends and gene-level variation

Visualization of gene expression

Gene expression patterns of RBPs were visualized using heatmap-based clustering approaches. Heatmaps were generated using the *pheatmap* package in R, with log2 fold change values used as the input matrix. A blue–white–red color gradient was applied to represent downregulated, neutral, and upregulated expression levels, respectively. Where applicable, hierarchical clustering was performed to explore expression pattern similarity among genes.

Bar plots were additionally generated using the *ggplot2* package to provide complementary visualization of RBP expression profiles.

Correlation analysis between lncRNAs and RBPs

To investigate potential regulatory relationships between RBPs and lncRNAs, Pearson correlation analysis was performed using gene expression using base R functions. Significant RBP–lncRNA interactions were selected based on predefined thresholds of $|r| > 0.5$, with p-value < 0.05 considered statistically significant. Only significant correlations were retained for downstream analyses.

Construction of RBP–lncRNA regulatory network

A regulatory interaction network was constructed based on significant RBP–lncRNA correlation pairs.

In this network, nodes represent RBPs and lncRNAs, while edges represent significant co-expression relationships (both positive and negative correlations). Network construction and visualization were performed using the *igraph* package in R.

The Fruchterman–Reingold force-directed layout algorithm was applied to optimize node positioning for improved visualization of network topology. Edge weights were defined based on the absolute values of correlation coefficients to reflect interaction strength between nodes.

Identification of hub genes

Network topology was analyzed by calculating node degree, defined as the number of direct connections for each node. Nodes with higher degree values were considered hub genes, representing highly connected regulatory elements within the network. Hub lncRNAs and hub RBPs were analyzed separately to identify key regulatory components within each class of molecules.

Differential expression analysis of selected lncRNAs

To validate network-derived findings, differential expression analysis was performed on five selected lncRNAs identified from TCGA datasets: EMSLR, MILIP, RIPK2-DT, RNASEH1-AS1, and SNHG32. Expression levels were compared between tumor and normal samples to assess their potential involvement in cancer-associated regulatory processes.

Software and packages

All analyses were performed in R (version 4.x) using the following packages:

- *pheatmap* for heatmap visualization
- *igraph* for network construction and visualization
- *ggplot2* for data visualization (bar plots)
- *dplyr* for data manipulation
- *tidyr* for data reshaping
- base R functions for statistical analysis

Statistical analysis

Pearson correlation analysis was used to assess relationships between RBPs and lncRNAs. Statistical significance was defined as $p < 0.05$ combined with a correlation magnitude threshold of $|r| > 0.5$. Although the analysis involved multiple pairwise comparisons, formal multiple testing correction (e.g., Benjamini–Hochberg FDR adjustment) was not applied, as this study is primarily exploratory in nature and the applied dual-threshold approach (combining both correlation strength and nominal significance) was intended to limit false discovery while retaining biologically meaningful interactions. We acknowledge this as a limitation; application of FDR correction in future studies with larger datasets would further strengthen reproducibility. Only interactions meeting both thresholds were included in the final network.

Results

1. Seventeen RNA-binding proteins were upregulated in CRC

Differential expression analysis identified 22 RBPs exhibiting significant expression alterations between CRC and normal tissues. Among these, 17 RBPs were upregulated, whereas 5 RBPs were downregulated in tumor samples. Notably, EIF4A3, SRSF7, NOP56, NOP58, IGF2BP3, METTL3, ELAVL1, and YTHDF1 demonstrated increased expression levels, while RBM47, QKI, ELAVL3, AUH, and VIM were among the downregulated RBPs. The bar plot and heatmap analyses revealed distinct expression patterns across the selected RBPs, supporting their potential involvement in CRC-associated post-transcriptional regulatory mechanisms (Figure 1).

2. Correlation analysis revealed predominantly positive lncRNA–RBP associations

To investigate the regulatory associations between lncRNAs and RBPs, Pearson correlation analysis was performed. A total of 66 significant correlations were identified, comprising 53 positive and 13 negative associations. The predominance of positive correlations suggests coordinated expression patterns between lncRNAs and RBPs within CRC tissues. The correlation matrix further demonstrated heterogeneous interaction patterns, indicating that individual RBPs may participate in distinct lncRNA-associated regulatory programs (Figure 2).

3. The lncRNA–RBP interaction network exhibited a highly connected regulatory architecture

Based on significant correlation pairs, an lncRNA–RBP interaction network was constructed. The resulting network consisted of 41 nodes and 66 edges, reflecting extensive connectivity between RBPs and lncRNAs. Visualization of the network revealed several highly interconnected regions, suggesting the existence of regulatory modules potentially involved in RNA metabolism, post-transcriptional regulation, and CRC progression. Both positive and negative interactions contributed to the overall network architecture, highlighting the complexity of lncRNA-mediated regulation of RBP activity (Figure 3).

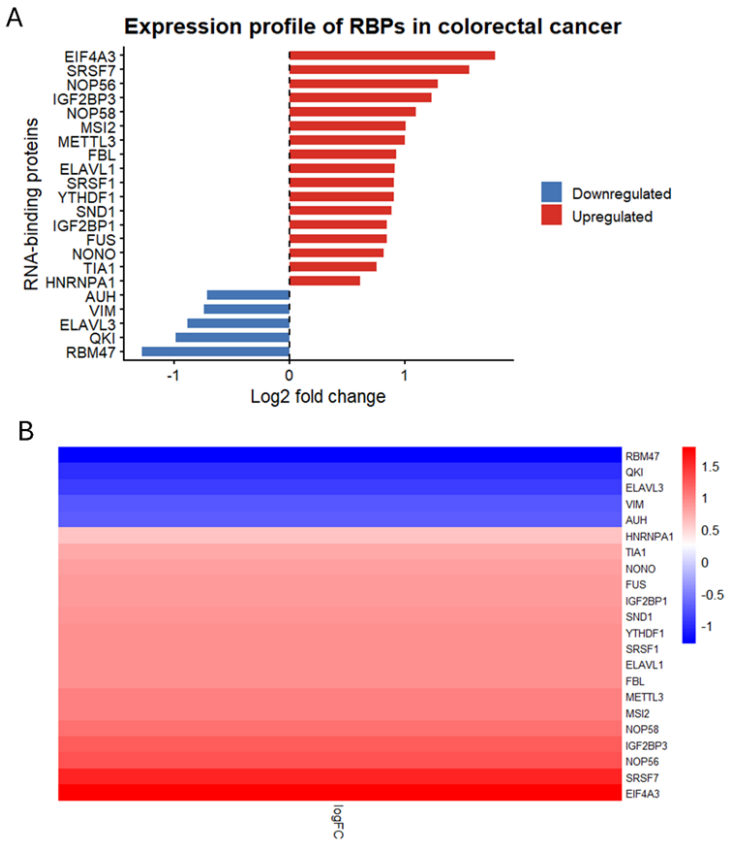

Figure 1. Differential expression profile of selected RNA-binding proteins in colorectal cancer.
(A) Bar plot showing log2 fold-change values of 22 significantly dysregulated RBPs between tumor and normal tissues. (B) Heatmap illustrating relative expression patterns of the same RBPs based on logFC values.

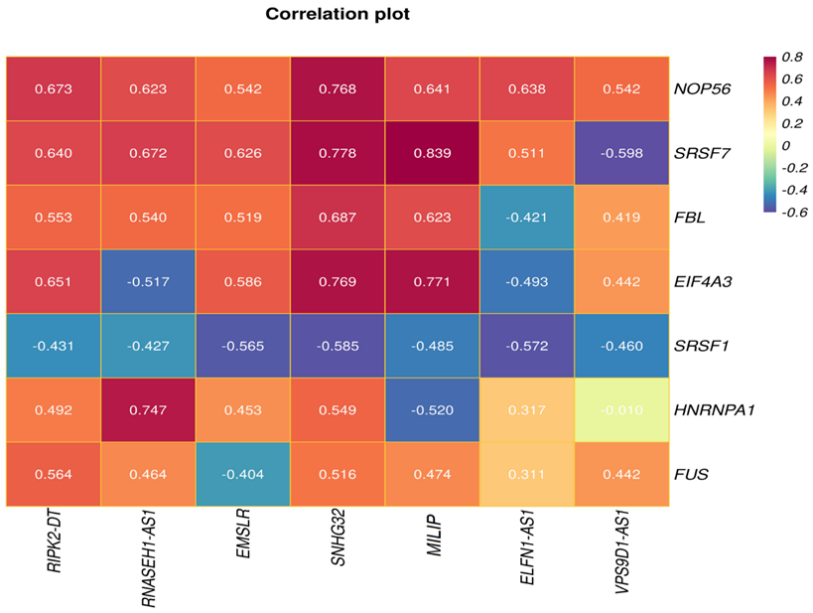

Figure 2. Correlation matrix of candidate lncRNAs and RNA-binding proteins. Heatmap showing Pearson correlation coefficients between selected lncRNAs and RBPs. Positive and negative correlations are represented by red and blue colors, respectively.

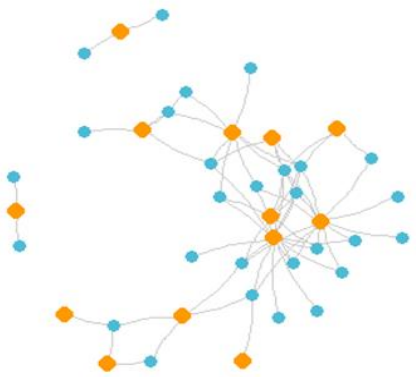

Figure 3. lncRNA-RBP interaction network in colorectal cancer. Network constructed from significant Pearson correlations between lncRNAs and RBPs. Node size is proportional to degree centrality. Blue nodes represent RBPs and orange nodes represent lncRNAs.

4. Highly connected hub RBPs and lncRNAs formed the core of the regulatory network

Topological analysis based on node degree identified several highly connected hub molecules within the lncRNA-RBP interaction network. As shown in Figure 4A, hub RBPs formed a densely interconnected regulatory core, with FBL, EIF4A3, QKI, and SND1 exhibiting the highest connectivity scores. These RBPs occupied central positions in the network and interacted with multiple lncRNAs, suggesting broad regulatory influence on post-transcriptional processes in CRC.

In parallel, Figure 4B highlights the most highly connected hub lncRNAs, which interacted with multiple RBPs and may function as key non-coding regulatory elements within the network. The coexistence of hub RBPs and hub lncRNAs indicates that CRC is characterized by a coordinated RNA-centered regulatory architecture rather than isolated molecular alterations.

Collectively, these findings suggest that the identified hub nodes may play pivotal roles in orchestrating post-transcriptional regulatory mechanisms and could represent promising candidates for future functional and therapeutic investigations in CRC.

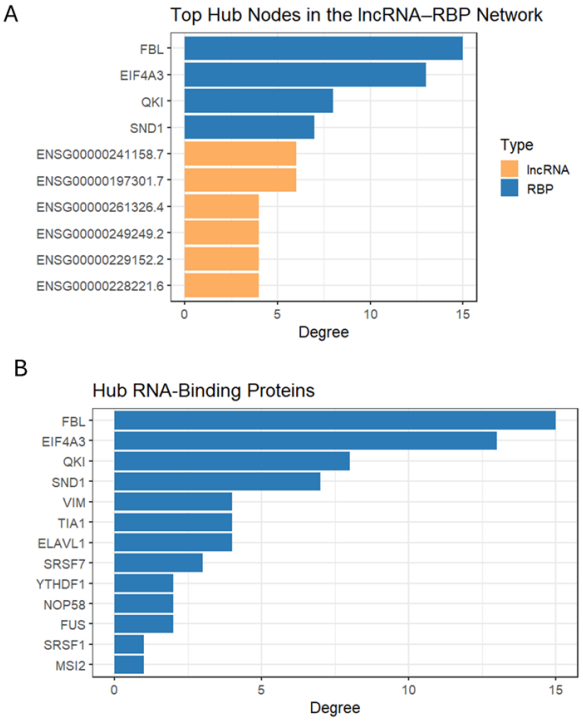

Figure 4. Identification of hub molecules within the lncRNA-RBP network.Bar plot showing the degree centrality of the most highly connected nodes. FBL, EIF4A3, QKI, and SND1 were identified as major RBP hubs.

5. Differential expression of selected lncRNAs in the TCGA dataset

To validate the biological relevance of network-derived candidates, expression levels of five lncRNAs identified from TCGA datasets (EMSLR, MILIP, RIPK2-DT, RNASEH1-AS1, and SNHG32) were compared between colorectal tumor and normal tissues (Figure 5). All five candidates showed distinct expression patterns, supporting their potential involvement in CRC-associated regulatory pathways.

Importantly, MILIP was the only lncRNA that overlapped between TCGA-derived differentially expressed lncRNAs and hub nodes identified in the global lncRNA–RBP interaction network. This dual evidence—transcriptional dysregulation together with network centrality—strongly suggests that MILIP may represent a key regulatory lncRNA in CRC.

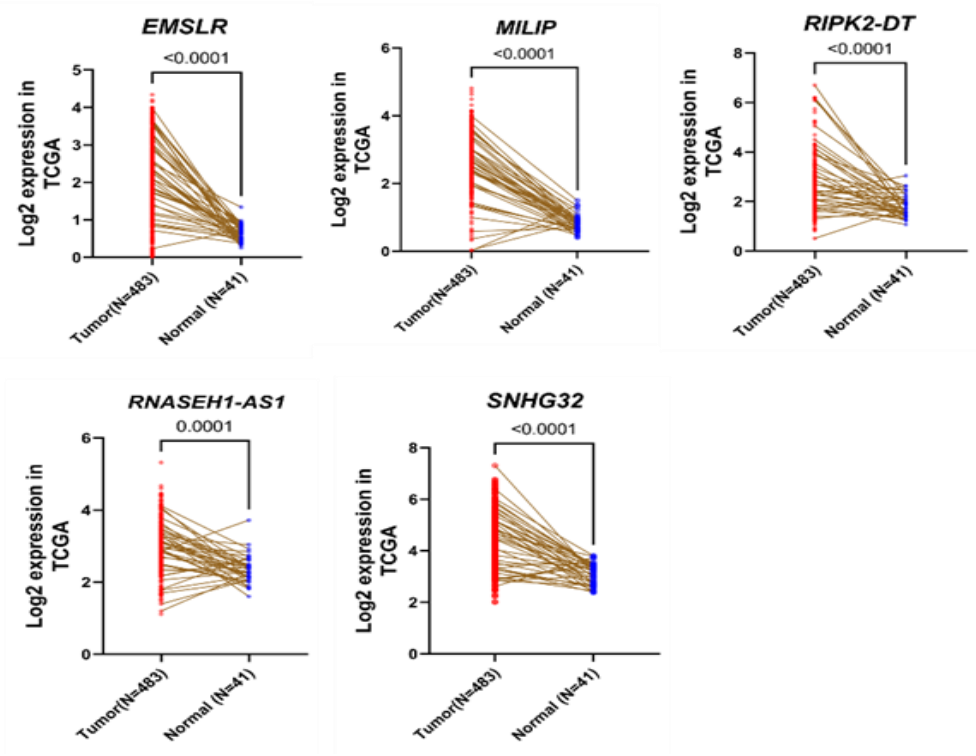

Figure 5. Differential expression of selected lncRNAs in tumor and normal tissues

Box/violin plot showing the expression differences of selected lncRNAs between tumor and normal colorectal tissues. Statistical significance was assessed using the Wilcoxon rank-sum test; asterisks indicate significance thresholds (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). Significant dysregulation of multiple lncRNAs was observed in tumor samples, suggesting their potential involvement in colorectal cancer pathogenesis

6. A MILIP-centered subnetwork highlighted extensive interactions with RNA-binding proteins

To further explore the regulatory context of MILIP, a MILIP-centered subnetwork was extracted from the global lncRNA–RBP interaction network (Figure 6). This focused analysis revealed multiple significant correlations between MILIP and several RNA-binding proteins, including FBL, EIF4A3, SND1, and QKI, indicating that MILIP is embedded within a densely connected post-transcriptional regulatory module.

Within this local network, MILIP occupied a central topological position characterized by high connectivity and strong correlation strengths with multiple RBPs. These observations support the hypothesis that MILIP may function as a regulatory scaffold or coordinator of RBP-mediated RNA regulatory processes in CRC. The convergence of differential expression, hub-node status, and subnetwork centrality collectively highlights MILIP as a promising candidate for further mechanistic and functional studies in CRC.

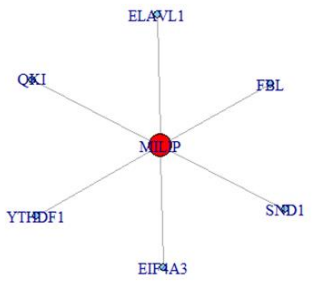

Figure 6. MILIP-centered RBP–lncRNA interaction network.

A MILIP-centered RBP–lncRNA subnetwork was extracted from the global correlation network. Nodes represent RBPs and lncRNAs, and edges indicate significant co-expression ($|r| > 0.5$, $p < 0.05$). MILIP is centrally located with strong connections to multiple RBPs, including FBL, EIF4A3, SND1, and QKI, suggesting a potential key regulatory role in colorectal cancer. Node size reflects degree centrality, and edge width represents correlation strength.

Discussion

In this study, we constructed an integrated RBP–lncRNA regulatory network in CRC using correlation-based analysis and combined it with TCGA-derived expression profiling to identify key regulatory molecules. Our findings highlight widespread dysregulation of RBPs and their coordinated interactions with lncRNAs, suggesting a complex post-transcriptional regulatory landscape in CRC.

Our differential expression analysis revealed that several RBPs, including EIF4A3, IGF2BP1/3, METTL3, SRSF7, and ELAVL1, were significantly upregulated, whereas QKI, RBM47, and VIM were downregulated. These observations are consistent with previous studies reporting that RBPs play dual roles in cancer biology, acting either as oncogenic drivers or tumor suppressors depending on context [3,4]. For example, IGF2BP proteins have been shown to stabilize oncogenic transcripts and promote tumor progression in multiple cancers, including CRC [11]. Similarly, METTL3-mediated RNA methylation has been widely associated with enhanced tumor growth and metastasis [12].

Correlation analysis between RBPs and lncRNAs revealed both positive and negative associations, indicating a complex regulatory landscape. Such bidirectional relationships suggest that RBPs may regulate lncRNA stability and processing, while lncRNAs may reciprocally modulate RBP function. This aligns with previous evidence showing that RBPs and lncRNAs frequently participate in feedback regulatory loops that control gene expression programs in cancer [7,13].

Network-based analysis further demonstrated a non-random interaction structure, where highly connected nodes occupied central positions within the RBP–lncRNA network. The identification of hub genes supports the concept that biological networks are scale-free, with a small number of highly connected nodes playing central regulatory roles [9]. In cancer systems biology, such hub genes are often functionally essential and may represent potential therapeutic targets.

Importantly, integration of TCGA expression data with network analysis identified MILIP as the only lncRNA overlapping between differential expression and hub node classification. This dual-level evidence suggests that MILIP may play a central regulatory role in CRC. Although MILIP has been less extensively studied compared to other oncogenic lncRNAs, emerging evidence suggests its involvement in cancer-related pathways and interaction with RNA regulatory machinery [14]. Its identification as a hub node further strengthens its potential biological relevance.

The MILIP-centered subnetwork analysis demonstrated strong connectivity between MILIP and multiple RBPs, including FBL, EIF4A3, SND1, and QKI. FBL is a nucleolar methyltransferase involved in ribosome biogenesis and has been linked to oncogenic transformation through dysregulation of ribosomal RNA processing [15]. EIF4A3, a member of the eIF4A helicase family, plays a key role in RNA splicing as a core component of the exon junction complex and has been reported to promote tumor progression in CRC [16]. SND1 (staphylococcal nuclease domain-containing protein 1) functions in RNA stability and miRNA biogenesis and has oncogenic roles in multiple cancers [17]. QKI, in contrast, is a well-characterized tumor-suppressive RBP that regulates alternative splicing and mRNA stability, with frequent downregulation observed in CRC and other malignancies [18]. The co-occurrence of MILIP with both oncogenic (FBL, EIF4A3, SND1) and tumor-suppressive (QKI) RBPs within the same subnetwork suggests that MILIP may be embedded in a complex regulatory module with opposing functional roles, acting as a potential coordinator of post-transcriptional gene expression dynamics. Such scaffold-like lncRNAs have been shown to coordinate protein–RNA interactions and regulate gene expression in cancer [5,7]. This supports the hypothesis that MILIP may contribute to CRC progression through modulation of RBP-mediated RNA regulatory processes. Several limitations of the current study should be acknowledged. First, no survival analysis or correlation with clinicopathological features (such as tumor stage, grade, or metastasis status) was performed for MILIP expression. Future studies should integrate clinical outcome data to assess whether MILIP expression is prognostically relevant in CRC. Second, while our findings are based on robust publicly available transcriptomic data, this study is entirely computational and hypothesis-generating in nature.

All conclusions regarding the regulatory roles of MILIP and the identified RBPs require experimental validation through functional studies such as knockdown/overexpression assays, RNA immunoprecipitation, or patient-derived tissue analyses. Overall, our integrative approach highlights the importance of combining network-based modeling with differential expression analysis to uncover key regulatory molecules in cancer. The identified RBP–lncRNA network provides new insights into post-transcriptional regulation in CRC and suggests that MILIP may serve as a promising candidate for further functional validation and potential biomarker development.

Acknowledgements

The authors acknowledge Mr. Ali Valipour Motlagh for his contribution to RNA-seq and microarray analysis. We also thank our coworkers for their helpful discussions in this study.

Author contributions statement

A. A., and M. P., were responsible for designing and conceptualizing the study and methodology. Data mining, formal analysis, and investigation were performed by A. A. Supervision, validation, and visualization were carried out by M. H., and M. P. and interpretation of the obtained information was done by M. P. M. P. wrote the manuscript, while M. E. and N. R. reviewed, edited, and approved it. All authors read and approved the final version of the manuscript.

Funding

No funding is required.

Ethical approval

The presented research involving human tissue samples was approved by the Biomedical Ethics Committee of the Islamic Azad University of Rasht with the Ethics Code IR.IAU.RASHT.REC.1402.022. Written informed consent was obtained from all participants prior to sample collection. All methods were performed in accordance with the relevant guidelines and regulations.

Consent to publish

Not applicable.

Competing Interests

Data Availability

All gene expression data used in this study are publicly available from The Cancer Genome Atlas (TCGA) database. The R scripts used for network construction, differential expression analysis, and visualization are available from the corresponding author upon reasonable request.

The authors declare that they have no competing interests.

Informed consent

Informed consent was obtained from all participants and/or their legal guardians prior to their inclusion in the study.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide. CA Cancer J Clin. 2021;71(3):209–249.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):xx–xx.
3. Pereira B, Billaud M, Almeida R. RNA-binding proteins in cancer: old players and new actors. Nat Rev Cancer. 2017;17(12):757–771.
4. Gebauer F, Hentze MW. Molecular mechanisms of RNA-binding proteins in cancer. Nat Rev Cancer. 2021;21(9):631–646.
5. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22(2):96–118.
6. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. Cancer Cell. 2016;29(4):452–463.
7. Dassi E. Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins. Front Mol Biosci. 2017 ;4:67.
8. He S, Valkov E, Cheloufi S, Murn J. The nexus between RNA-binding proteins and their effectors. Nat Rev Genet. 2023;24(5):276-294.
9. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2021;12(1):56–68.
10. Peng WX, Koirala P, Mo YY. LncRNA-mediated regulation of cell signaling in cancer. Oncogene. 2017 Oct 12;36(41):5661-5667.
11. Huang H, Weng H, Sun W, Qin X, Shi H, Wu H, et al. Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. Nat Cell Biol. 2018;20(3):285–295.
12. Lin S, Choe J, Du P, Triboulet R, Gregory RI. The m6A methyltransferase METTL3 promotes translation in cancer cells. Mol Cell. 2016;62(3):335–345.
13. Yao RW, Wang Y, Chen LL. Cellular functions of long noncoding RNAs. Nat Cell Biol. 2019;21(5):542–551.
14. Feng YC, Liu XY, Teng L, et al. c-Myc inactivation of p53 through the pan-cancer lncRNA MILIP drives cancer pathogenesis. Nat Commun. 2020 Oct 5;11(1):4980.