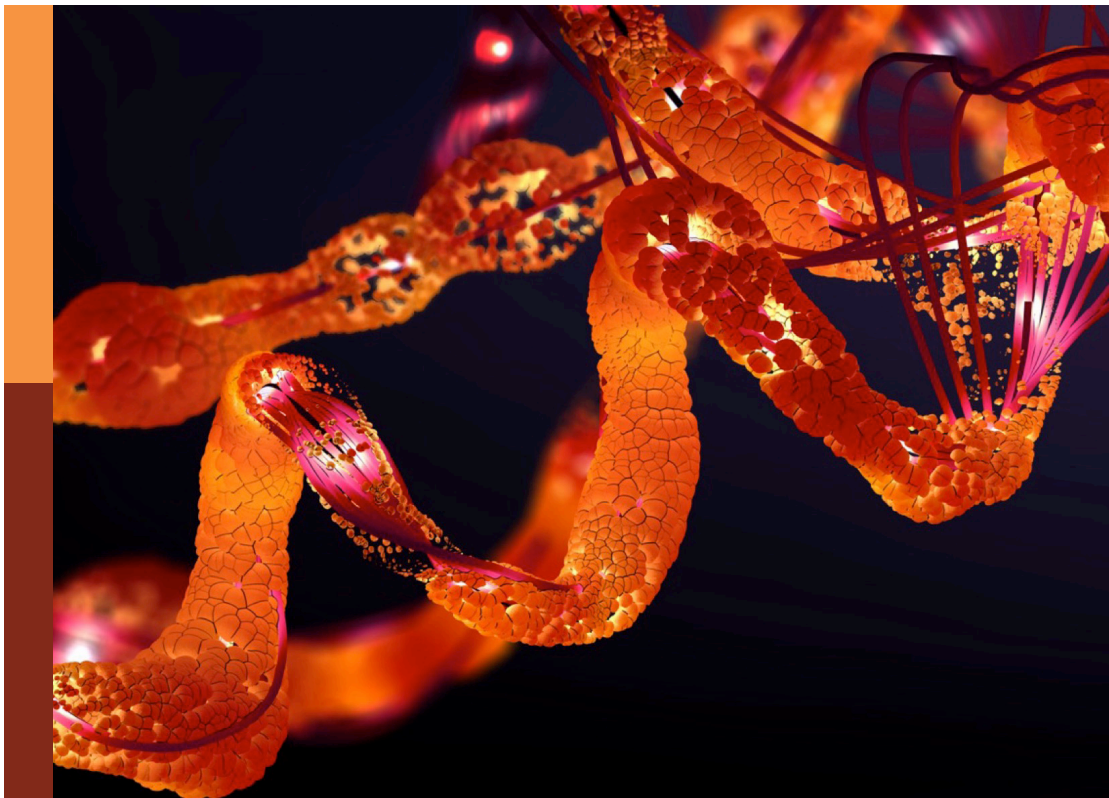
frontiers | Research Topics

Insights in RNA networks and biology

Edited by
André P. Gerber and Gian Gaetano Tartaglia

Published in
Frontiers in Molecular Biosciences



ISSN 1664-8714
ISBN 978-2-83251-823-6
DOI 10.3389/978-2-83251-823-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Insights in RNA networks and biology

Topic editors

André P. Gerber — University of Surrey, United Kingdom

Gian Gaetano Tartaglia — Italian Institute of Technology (IIT), Italy

Citation

Gerber, A. P., Tartaglia, G. G., eds. (2023). *Insights in RNA networks and biology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-823-6

Table of contents

frontiers | Frontiers in Molecular Biosciences

TYPE Editorial
PUBLISHED 17 February 2023
DOI 10.3389/fmolb.2023.1156841

OPEN ACCESS

EDITED AND REVIEWED BY
Graça Soveral,
University of Lisbon, Portugal

*CORRESPONDENCE
André P. Gerber,
a.gerber@surrey.ac.uk
Gian Gaetano Tartaglia,
gian.tartaglia@iit.it

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 01 February 2023
ACCEPTED 10 February 2023
PUBLISHED 17 February 2023

CITATION
Gerber AP and Tartaglia GG (2023), Editorial: Insights into RNA networks and biology.
Front. Mol. Biosci. 10:1156841.
doi: 10.3389/fmolb.2023.1156841

Editorial: Insights into RNA networks and biology

André P. Gerber[1]* and Gian Gaetano Tartaglia[2]*

[1]Department Microbial Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom, [2]Center for Human Technologies (CHT), Istituto Italiano di Tecnologia (IIT), Genova, Italy

KEYWORDS

RNA, modelling, biochemistry, high-throughput sequencing, metabolism, translation

Editorial on the Research Topic
Insights into RNA networks and biology

In our Research Topic "Insights into RNA Networks and Biology" we gathered a number of articles that relate to the newest developments in the RNA field. Starting from computational works (Pepe et al. and Vandelli et al.) and passing through network analysis (Yang et al. and Zhang et al.), we present approaches for biochemical characterization of protein-RNA interactions (Giambruno et al. and Kieffer et al.), implication of long non-coding RNAs (lncRNAs) in disease (Fan et al. and Li et al.) and investigation of the role of RNA molecules and translation in physiology (Buchanan et al. and Zhang et al.).

Several theoretical methods for studying physical interactions between RNA and proteins are currently being developed. The review by Pepe et al. summarizes the most recent advances in the algorithms, especially to predict circular RNA (circRNA) and long non-coding RNA (lncRNA) interactions with RNA-binding proteins (RBPs). The authors emphasize the need for developing larger interaction datasets to increase performance and expand on approaches amenable to machine learning. In their Perspective, Vandelli et al. provide an example on how the integration of RNA-protein interaction data with computational analysis tools can lead to new hypotheses. The authors used data from four experimental and one computational study of interactors of SARS-CoV-2 genomic RNA. Although hundreds of proteins have been identified across the different studies, only twenty-one appear in all the experiments to interact with SARS-CoV-2 RNA. The identified proteins refer to known or predicted stress granule forming proteins, which leads to the intriguing hypothesis that a mechanism of action for SARS-CoV 2 could be to target proteins that attract other partners through phase separation.

Two articles focus on transcriptome analysis for construction of RNA-RNA networks. Yang et al. performed a systematic analysis of mRNA, microRNA (miRNA) and circRNA expression in type II alveolar epithelial cells (ATII) in Tibetan pigs. These pigs live at high altitude and developed an efficient metabolism for living at lower oxygen levels than "standard" landrace pigs. The authors simulated a hypoxic environment for cultured ATII cells of tibetan pigs and landrace pigs grown under normoxic and hypoxic conditions. The differential expression of circRNA, miRNAs and mRNAs were combined to construct competing endogenous RNA (ceRNA) networks for these cells, thereby identifying critical regulatory modes for metabolic control. A similar approach was used by Zhang et al. that explored the transcriptomic network of Rotator cuff tears (RCT), the most common cause of shoulder dysfunction. Specifically, the authors performed RNA-seq analysis of five patient samples with RTC in supraspinatus muscles and matched

unharmed subscapularis muscles from the same individual to dissect the dysregulated transcriptome including mRNAs, miRNAs, lncRNAs, and circRNAs. Based on network construction considering the ceRNA hypothesis, they suggested dysregulation of the ceRNA network in RTC through several ncRNAs. Hence, the authors propose that these ncRNAs could play a role in the development of RCT, possibly adding a new angle for therapeutic approaches.

Cross-linking is often used to preserve functional protein-RNA interactions present in the cell. The review by Giambruno et al. describes recently developed complementary proximity labeling techniques to map RNA-protein interactions in mammalian cells. This technology permits the identification of relatively transient interactions as well as poorly expressed molecules by relying on the inducible activity of enzymes (biotin ligases or peroxidases) that are expressed in living cells and biotinylate amino acids and nucleic acids of binding factors in proximity–within 20 nM range. Besides the interaction with their target RNAs, RBPs often interact with other proteins which further influence the fate of ribonucleoprotein (RNP) complexes. Kieffer et al. combined affinity purification and mass spectrometry (MS) to define the network of nuclear proteins interacting with the N-terminal region of FMRP. FMRP is encoded by the *FMR1* gene (Fragile X messenger ribonucleoprotein 1) whose absence of expression results in Fragile X-syndrome (FXS), the most common inherited form of intellectual disability. While it was well known that FMRP regulates translation of subsets of mRNA in the cytoplasm, the protein shuttles between the nucleus and cytoplasm but little was known about its nuclear functions. Hence, this study adds a base to further mechanistic studies on its nuclear functions in neuronal physiology.

Two review articles highlight the specific and broader implications of lncRNAs in disease. Fan et al. discuss the role of lncRNAs and aerobic glycolysis in tumorigenesis and tumor progression and explore the interaction networks to provide insights into therapeutic targets for treatment. The authors suggest that lncRNAs could regulate key enzymes related to glycolysis to promote aerobic glycolysis (examples include *PKM2, LDH, HK2* and *GLUT*). An important highlight of the work is that lncRNAs can act as sponges for miRNAs to regulate expression of enzymes or modulate signaling pathways. The association of lncRNAs with enzymes could also promote formation of protein complexes (examples include lncRNA *HULC* binding to LDHA, PKM2) thus conferring additional functions. The article by Li et al. elaborates on the relationship between lncRNAs and diseases. In the case of tumor development, the authors report that a lncRNA can act either as proto-oncogene or tumor suppressor. As for neurodegenerative and psychiatric diseases, examples are provided for *DISC1* (schizophrenia); *PINK1* (endocrine diseases), *MALAT-1* and *NEAT1/2* and others. One important case is *BACE1* antisense (*BACE1AS*) that forms complex with *BACE1* messenger RNA to increase the stability of the latter thus promoting aggregation of β-amyloid protein in Alzheimer's disease.

Besides the many ncRNA and viral RNAs concerned in this Research Topic, two articles focus on mRNAs for cell fate determination and in translation. In their perspective, Buchanan et al. discuss the potential role of ribosome heterogeneity in brain plasticity. After briefly reviewing the importance of translation regulation in neuroplasticity, the authors highlight recent results showing that complex brain functions such as sleep and learning have a significant impact on ribosomal protein (RP) expression. As sleep is critical in promoting brain plasticity, the authors then propose that changes in RP stoichiometry during sleep could establish a mechanism for regulation of translation of subsets of mRNAs that code for proteins mediating brain plasticity. Zhang et al. investigate the regulation of Hsp70 chaperones *Ophraella communa*. In concert with many cofactors, Hsp mediates essential activities such as protein folding and assembly. Normally Hsp70 is located in the cytoplasm, however, when cells are stimulated by heat stress, the protein is transferred to the nucleus. In the case of *O. communa*, an important biological control agent of the ragweed *Ambrosia artemisiifolia* worldwide, Hsp70 was found to be highly expressed in the female ovaries and male testes and induced by mating. The regulatory mechanisms of Hsp70 and the investigation of its RNA stability and translation should be further investigated.

Author contributions

AG and GT wrote the piece together.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Molecular Biosciences

PERSPECTIVE
published: 11 May 2022
doi: 10.3389/fmolb.2022.893067

Phase Separation Drives SARS-CoV-2 Replication: A Hypothesis

Andrea Vandelli[1,2†], *Giovanni Vocino*[3†] *and Gian Gaetano Tartaglia*[4,5,6*]

[1]Department of Biochemistry and Molecular Biology, Universitat Autònoma de Barcelona, Barcelona, Spain, [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain, [3]Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy, [4]Center for Human Technologies, Istituto Italiano di Tecnologia, Genova, Italy, [5]Department of Biology 'Charles Darwin', Sapienza University of Rome, Rome, Italy, [6]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Identifying human proteins that interact with SARS-CoV-2 genome is important to understand its replication and to identify therapeutic strategies. Recent studies have unveiled protein interactions of SARS-COV-2 in different cell lines and through a number of high-throughput approaches. Here, we carried out a comparative analysis of four experimental and one computational studies to characterize the interactions of SARS-CoV-2 genomic RNA. Although hundreds of interactors have been identified, only twenty-one appear in all the experiments and show a strong propensity to bind. This set of interactors includes stress granule forming proteins, pre-mRNA regulators and elements involved in the replication process. Our calculations indicate that DDX3X and several editases bind the 5′ end of SARS-CoV-2, a regulatory region previously reported to attract a large number of proteins. The small overlap among experimental datasets suggests that SARS-CoV-2 genome establishes stable interactions only with few interactors, while many proteins bind less tightly. In analogy to what has been previously reported for *Xist* non-coding RNA, we propose a mechanism of phase separation through which SARS-CoV-2 progressively sequesters human proteins hijacking the host immune response.

Keywords: viral RNA, phase separation, stress granules, protein-RNA interactions, RNA-binding proteins

OPEN ACCESS

Edited by:
Barbara Bardoni, UMR7275 Institut de Pharmacologie Moléculaire et Cellulaire (IPMC), France

Reviewed by:
Venu Raman, Johns Hopkins Medicine, United States
Lorena Zubovic, University of Trento, Italy

***Correspondence:**
Gian Gaetano Tartaglia
gian.tartaglia@iit.it

†*These authors share first authorship*

Specialty section:
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *09 March 2022*
Accepted: *25 April 2022*
Published: *11 May 2022*

Citation:
Vandelli A, Vocino G and Tartaglia GG (2022) Phase Separation Drives SARS-CoV-2 Replication: A Hypothesis. Front. Mol. Biosci. 9:893067. doi: 10.3389/fmolb.2022.893067

INTRODUCTION

Identification of viral interactions within the host cell can lead to the design of novel strategies against infection. Recently, different high-throughput strategies have been implemented to characterize host interactions with SARS-CoV-2 proteins and genomic RNA.

Non-structural proteins of SARS-CoV-2 have been used for affinity purification to retrieve host binding partners using mass spectrometry in HEK-293T/17 cells (Gordon et al., 2020). A total of 332 interactions between human and SARS-CoV-2 proteins have been identified. Around 40% of SARS-CoV-2 interacting proteins are associated with vesicle trafficking pathways and endomembrane compartments.

Here, we focus on four experimental studies aiming to characterize interactions with SARS-CoV-2 genomic RNA.

In one experiment, a multi-omic approach was employed to identify which viral and human RNA-binding proteins (RBPs) are involved in SARS-CoV-2 infection (Kamel et al., 2021). The "comparative RNA interactome capture" (cRIC) method was developed to find in which way the RNA-bound proteome responds to the infection. The results show that SARS-CoV-2 genome is the epicenter of critical interactions with host proteins: many cellular RBP networks are remodeled upon SARS-CoV-2 infection and around 300 proteins are affected, mostly related to RNA metabolic

processes and antiviral defenses. A second approach called "viral RNA interactome capture" (vRIC) was employed to identify cellular and viral proteins interacting with SARS-CoV-2 genomic RNA (Kamel et al., 2021). Inhibition of specific proteins interacting with viral RNA was shown to impair SARS-CoV-2 infection.

In another study (Lee et al., 2021), the repertoire of host proteins associated with SARS-CoV-2 and HCoV-OC43 genomes was identified. The work relies on a robust nucleoprotein (RNP) capture protocol. More than 100 host factors directly binding to SARS-CoV-2 RNA were detected. By applying RNP capture on HCoV-OC43, evolutionary conserved interactions between the viral RNAs and the host proteins could be identified. Upon knockdown experiments and transcriptome analysis, Lee *et al.* identified 17 antiviral and 8 pro-viral RBPs that have a role in several steps of the mRNA life cycle. The authors identified La-related protein 1 (LARP1), a downstream target of the mTOR signaling pathway, as an important antiviral host factor that interacts with SARS-CoV-2 RNA.

Another group exploited an approach in which a comprehensive identification of RBPs followed by mass spectrometry (ChIRP-MS) led to the identification of host proteins that bind SARS-CoV-2 genomic RNA during active infection (Flynn et al., 2021). The results were corroborated with analyses from three RNA viruses and contributed to characterize the specificity of virus-host interactions. Flynn *et al.* also carried out a series of targeted CRISPR screens that highlighted the fact that a big portion of functional RNA-binding proteins act as host's protectors from virus-induced cell death. Comparative CRISPR screens, performed across seven RNA viruses, reveal both shared and SARS-specific antiviral factors. By combining the RNA-centric approach and the functional CRISPR screens, the authors found a functional connection between SARS-CoV-2 and mitochondria, showing that this organelle is a platform for antiviral activity.

A slightly different experiment led to the identification of more than 100 human proteins that directly and specifically bind to SARS-CoV-2 RNAs in infected cells, performing RNA antisense purification and mass spectrometry. Schmidt et al. linked SARS-CoV-2 interactome with changes in proteome abundance induced by viral infection, identifying cellular pathways relevant to SARS-CoV-2 infections. The authors demonstrated by genetic perturbation that both Cellular Nucleic-acid Binding Protein (CNBP) and LARP1, which are two of the most enriched viral RNA binders, have the ability to restrict SARS-CoV-2 replication in infected cells and provide a general map of their direct RNA contact sites. The authors demonstrated a reduced viral replication rate in two human cell lines after a pharmacological inhibition of three other binding partners (PPIA, ATP1A1, ARP2/3 complex).

As experimental studies require time and resources and are affected by intrinsic limitations (for instance mass-spec cannot identify every protein with the same efficiency), computational methods can be exploited to prioritize candidate targets. We previously used the CROSS method (Delli Ponti et al., 2017) to predict secondary structure content of and the *cat*RAPID approach (Bellucci et al., 2011; Agostini et al., 2013b; Cirillo et al., 2017) to compute >100000 human protein interactions with SARS-CoV-2 genomic RNA (Vandelli et al., 2020). The 5′ and 3′ end of SARS-CoV-2 were found to be highly structured, in agreement with subsequent experimental reports (Manfredonia et al., 2020) and show strong propensity to interact with human proteins. Among the identified interactors we identified there are several RNA editases and ATP-dependent RNA helicases that play a role in viral RNA processing and have a high propensity to participate in large macromolecular complexes. A number of proteins are predicted to be sequestered by SARS-CoV-2 genome and their recruitment contributes is thought to modify both the transcriptional and post-transcriptional regulations of host cells.

Here, we analyzed four experimental and one computational studies on human RBPs interactions with SARS-CoV-2 genomic RNA. We exploited the *cat*RAPID algorithm to estimate the ability of proteins to bind SARS-CoV-2 and identified a tight correlation between the number of experiments in which a specific protein is detected experimentally and its predicted binding affinity. Finally, we propose a model in which SARS-CoV-2 RNA promotes the formation of a phase-separated assembly by sequestering specific human proteins.

RESULTS

Interactomes Comparison

To retrieve interactions relevant for SARS-CoV-2 infection, we analysed four protein-RNA interactome experiments (**Supplementary Material S1**).

Twenty-one proteins were found in common to the four datasets (Flynn et al., 2021; Kamel et al., 2021; Lee et al., 2021; Schmidt et al., 2021) (**Figure 1A**). The list includes PABPC1 (Polyadenylate-binding protein 1), SND1 (Staphylococcal nuclease domain-containing protein 1), PPIA (Peptidyl-prolyl cis-trans isomerase A), DDX3X (ATP-dependent RNA helicase DDX3X), HNRNPA2B1 (Heterogeneous nuclear ribonucleoproteins A2/B1), HNRNPA0 (Heterogeneous nuclear ribonucleoprotein A), G3BP1 (Ras GTPase-activating protein-binding protein 1), G3BP2 (Ras GTPase-activating protein-binding protein 2), EIF4B (Eukaryotic translation initiation factor 4B), RPS2 (40S ribosomal protein S2), RPS3 (40S ribosomal protein S3), EIF3G (Eukaryotic translation initiation factor 3 subunit G) and YBX1 (Y-box-binding protein 1), **Supplementary Tables S1, S2**).

These proteins form a dense protein-protein network (**Figure 1B**) containing several stress granule components (G3BP1, G3BP2, EIF4B, DDX3X, YBX1, PABPC1), ribosomal units (RPS2 and RPS3) and pre-mRNA processing units (HNRNPA1/B2, HNRNPA0, YBX1) (Warde-Farley et al., 2010). The biological relevance of these interactions is confirmed by the fact that SARS-CoV-2 N protein impairs stress granule by sequestering G3BP1 (Lu et al., 2021; Zheng et al., 2021). RPS2 and RPS3 are important because the NSP1 protein of SARS-CoV-2 is responsible for the impairment of mRNA translation by blocking the entry access to the ribosome. The docking within the ribosomal entry channel occurs through binding with RPS2 and RPS3 together with 18S RNA (Mendez et al., 2021).

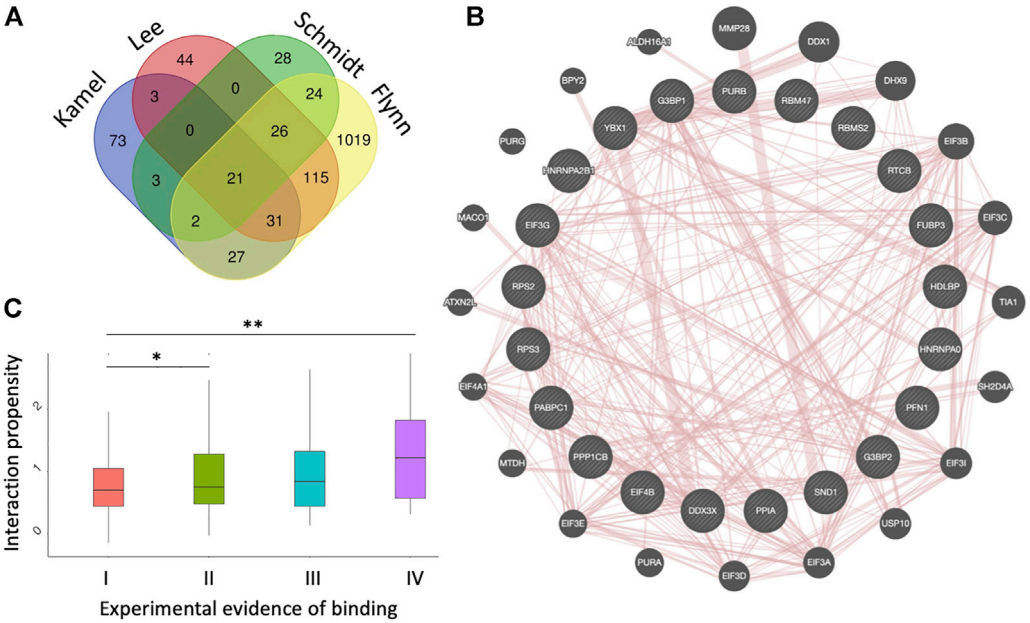


FIGURE 1 | Datasets of protein interactions with SARS-CoV-2 genome. **(A)** experimental datasets (Flynn et al., 2021; Kamel et al., 2021; Lee et al., 2021; Schmidt et al., 2021). The name of each dataset is shown above the diagrams. **(B)** diagram showing the protein-protein interactions among the 21 proteins identified in the four experiments, as annotated by GeneMANIA (Warde-Farley et al., 2010). **(C)** *cat*RAPID interaction scores (Agostini et al., 2013b; Armaos et al., 2021) correlate with the number of experiments reporting a protein to interact with SARS-CoV-2, indicating that strong binding proteins are more likely to be identified; **p*-value < 0.05; ***p*-value < 0.01 (Wilcoxon rank sum test); I II, II, III and IV indicate proteins detected in 1,2,3 or 4 experiments, respectively.

Some of these proteins have been shown to be also relevant for other viruses' infection. SND1 is involved in Epstein-Barr infection (Tong et al., 1995); PABPC1 positively regulates Dengue virus infection (Suzuki et al., 2016); PPIA acts as a mediator for SARS-CoV nucleoprotein during the cell invasion process and stimulates RNA-binding ability of HCV NS5A (Chen et al., 2005; Foster et al., 2011); EIF3G is involved in FCV infection process (Pöyry et al., 2007) and DDX3X has been shown to facilitate the viral replication of other several viruses, such as HIV-1, Dengue, Zyka, Venezuelan equine encephalitis and hepatitis C virus (Yedavalli et al., 2004; Amaya et al., 2016; Doñate-Macián et al., 2018). DDX3X has been identified as a suitable target to fight against SARS-CoV-2 infection by Ciccosanti et al. (2021). More precisely, DDX3X has the capability of unfolding viral RNA secondary structures (Kukhanova et al., 2020) as reported for HIV-1 (Brai et al., 2020) in which it enhances both translation and nucleus-to-cytoplasm transport (Stunnenberg et al., 2018), and West Nile (Brai et al., 2019). DDX3X belongs to the DEAD-box family of ATP-dependent RNA helicases and assumes a crucial role in an important variety of processes concerning RNA metabolism, including transcription, splicing, and initial phase of translation (Ariumi, 2014). Importantly, DDX3X interacts with the N protein of SARS-CoV-2 and is required to infect both Vero E6 and Calu-3 cells (Ciccosanti et al., 2021). Additionally, SARS-CoV-2 protein N interacts with DDX3X to inhibit its activity in the antiviral response (Winnard et al., 2021). For these reasons, treating cells with DDX3X inhibitors represents a promising approach to block SARS-CoV-2 replication and viral production (Maga et al., 2011; Brai et al., 2020).

Relationship Between Experimental Interactomes and Computational Predictions

We used the *cat*RAPID method to understand the relationship between experimental evidence of binding and predicted interaction propensity that estimates interaction affinity (Agostini et al., 2013a; Cid-Samper et al., 2018). For this analysis we followed a procedure previously introduced to study the interactome of the long non-coding RNA *Xist* (Cirillo et al., 2017). We computed all SARS-CoV-2 interactions with proteins reported in the four experimental datasets and counted how many times they were identified (**Supplementary Material S1**). We observed a distinct correlation between occurrence and strength of interactions, indicating that high-affinity interactions are more likely to be detected (**Figure 1C**). We note that in the case of *Xist*, strong interaction proteins were predicted to initiate the formation of a phase-separated assembly (Cerase et al., 2019, 2022), as recently confirmed experimentally (Markaki et al., 2021; Jachowicz et al., 2022).

Evaluation of the Predictions of SARS-CoV-2 Protein Interactions

The vRIC dataset by Kamel *et al.* contains both enriched and depleted interactions (Kamel et al., 2021) and thus can be used to assess the ability of *cat*RAPID to distinguish between binding and non-binding proteins. To analyze the vRIC interactome, we

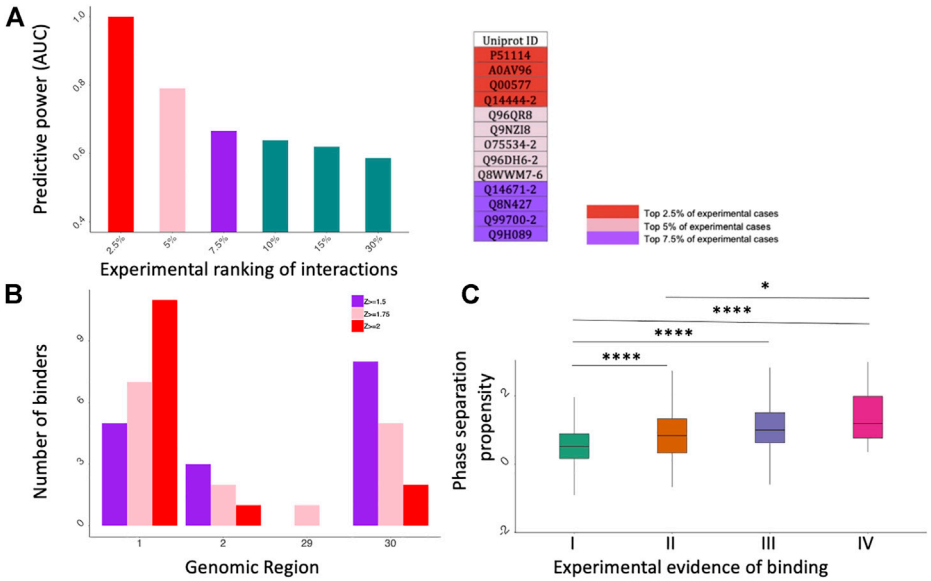


FIGURE 2 | catRAPID and catGRANULE predictions of protein interactions. **(A)** catRAPID performance evaluation. On the *X* axis we report different portions of the experimental dataset ranked by fold change and on the *Y* axis there is the corresponding predictive power (Area Under the ROC Curve, AUC). On the right, we report a summary table showing the Uniprot IDs of top 2.5%, 5% and 7.5% experimental cases. **(B)** Distribution of specific binders for Kamel et al. dataset (Kamel et al., 2021). The most contacted SARS-CoV-2 genomic regions correspond to the 5′ (first fragment) e 3′ (30th fragment). **(C)** catGRANULE phase separation propensity scores correlate with the number of experiments reporting a protein to interact with SARS-CoV-2 (Bolognesi et al., 2016; Cid-Samper et al., 2018); **p*-value < 0.05; *****p*-value < 0.0001 (Wilcoxon rank sum test); I II, II, III and IV indicate proteins detected in 1,2,3 or 4 experiments, respectively.

computed *cat*RAPID predictions of interactions with an experimental FDR <0.10 for SARS-CoV-2 RNA following a procedure detailed in a previous work (Vandelli et al., 2020) (**Supplementary Material S1**).

As shown in **Figure 2A**, *cat*RAPID performs extremely well when the proteins are ranked according to their experimental scores (fold change; **Supplementary Table S3**): the predictive power is proportional to the significance of protein interactions: the Area Under the ROC Curve (AUC) increases from 0.60 to 0.99 while the experimental scores move from 30% (i.e., the 30% strongest positives vs. the 30% strongest negatives) to 2.5% (i.e., the 2.5% strongest positives vs. the 2.5% strongest negatives). Thus, in agreement with the results presented in **Figure 1C**, computational approaches such as *cat*RAPID can be exploited to address the problem of which proteins bind more tightly to SARS-CoV-2 genome.

Specific Binders to SARS-CoV-2 Genomic Fragments

*cat*RAPID was employed for the localization of protein binding sites on SARS-CoV-2 genomic RNA. To identify which regions of SARS-CoV-2 bind to specific proteins, we computed interactions for the four experimental protein datasets (30 fragments; **Supplementary Material S4**), a procedure already proven to be efficient in a previous work (Vandelli et al., 2020).

For each dataset the proteins bound to one fragment at a certain interaction threshold were retained as interactors. We applied three Z-score thresholds ($Z \geq 1.5$, $Z \geq 1.75$ and $Z \geq 2$) in order to evaluate the binding at the different levels of stringency. Higher Z-scores correspond to higher interaction strength (**Supplementary Material S5**).

Regions encompassing nucleotides 1–1000, 1001–2000, 22001–23000, 26001–27000, 28001–29000, 29001–29903 (Fragments 1, 2, 23, 27, 29 and 30 respectively) are the most contacted SARS-CoV-2 regions, with a high number of interactors in fragments 1, 2 and 30. (**Figure 2B**; **Supplementary Figures S1–S3**). In particular, fragment 1, corresponding to the 5′ end of SARS-CoV-2 genome, is the region showing the highest number of specific interactors in all four datasets, as previously discovered (Vandelli et al., 2020). DDX3X is the only common interactor reported in the experimental and computational studies. At a $Z \geq 1.75$ we DDX3X is found to bind specifically to fragment 1 of SARS-CoV-2.

Experimental Interactors Have a High Propensity to Phase-Separate

Stress granules facilitate the establishment of an antiviral state by limiting viral protein accumulation and regulating signaling cascades that affect replication (McCormick and Khaperskyy, 2017). The sequestration of G3BP1, G3BP2, EIF4B, DDX3X, YBX1, PABPC1, among other proteins, is part of a mechanism through which SARS-CoV-2 eludes the host immune response by weakening the formation of stress granules (Lu et al., 2021; Zheng et al., 2021). Biochemically, stress granule proteins form labile protein-protein and protein-RNA interactions (Balcerak et al., 2019; Vandelli et al., 2022), which induces the condensation in liquid-liquid phase

separated assemblies (Gotor et al., 2020). We reasoned that the relatively small overlap among experimental datasets (**Figure 1A**) could be caused by the establishment of weak molecular interactions with SARS-CoV-2 RNA. In agreement with this observation, previous studies have suggested that phase separation could be a mechanism through which SARS-CoV-2 attracts host proteins (Iserman et al., 2020; Vandelli et al., 2020).

Using the *cat*GRANULE algorithm to predict phase separation propensities (Bolognesi et al., 2016; Cid-Samper et al., 2018) we analyzed the interactomes of the four experimental datasets. We discovered that the phase separation propensity correlates with how many times proteins are identified experimentally (**Figure 2C**). Considering that strong binding propensities are associated with proteins reported in the four experiments (**Figure 1C**) and the reliability of our approach (**Figure 2A**), we speculate that a possible mechanism of action for SARS-CoV-2 is to target proteins that attract other partners through phase separation.

DISCUSSION

This work is a comparative analysis on protein-RNA interactomes reported in experimental and computational studies. We found several proteins shared by the four experiments, including PABPC1, SND1, PPIA, EIF3G and DDX3X, which previous studies have shown to regulate replication of viruses.

DDX3X is found in all the experimental studies and it has been proven fundamental in SARS-CoV-2 biological processes and in the replication process of other viruses (Maga et al., 2011; Ariumi, 2014; Stunnenberg et al., 2018; Brai et al., 2019, 2020; Kukhanova et al., 2020; Ciccosanti et al., 2021; Winnard et al., 2021). *cat*RAPID predictions of human protein interactions with SARS-CoV-2 showed a prevalence of specific binders to the 5′ end of the virus, with DDX3X being one of them. Since *cat*RAPID reproduces experimental data to a remarkable extent, as assessed by directly comparing performances at different cut-offs, we believe that this information on the localization of protein interactions is to be taken into account for future analyses.

Predictive studies always have a margin of error, so further work will be necessary for a complete understanding of the specific binding sites and the role(s) of proteins in the context of infection.

In a recent study (Cirillo et al., 2017), we reported that the long non-coding RNA *Xist* physically interacts with few specific proteins that attract several other proteins (Cerase et al., 2019) forming a phase-separated assembly that silences the X chromosome (Cerase et al., 2022; Jachowicz et al., 2022). The relatively poor overlap of interactors among SARS-CoV-2 studies (only 21 proteins in common out of hundreds identified in total) suggests a mechanism similar to the one identified for *Xist*. The fact that SARS-CoV-2 binding proteins are either stress granules components or have high phase separation propensity supports our hypothesis. Indeed, phase separation is caused by weak protein-protein or protein-RNA interactions (Balcerak et al., 2019; Vandelli et al., 2022), which renders the identification of binding partners particularly difficult at the experimental level (Tartaglia, 2016; Cerase and Tartaglia, 2020) and could hamper their reproducibility. Moreover, the fact that proteins with the highest interaction and phase separation propensities were identified in all experimental studies suggests that they could act as the primary attractors to ignite the formation of an assembly that is capable of using host elements for replication. Further work is needed to study this fundamental aspect of SARS-CoV-2 biology and how it could be exploited to prevent viral infection. For example, molecular chaperones (Tartaglia et al., 2010; Alagar Boopathy et al., 2022) could be important players (Guihur et al., 2020) to be investigated in more detail.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AV and GT conceived the study. AV and GV carried out the analysis. AV, GV, and GT wrote the paper.

FUNDING

The research leading to these results has been supported by European Research Council (RIBOMYLOME_309545 and ASTRA_855923), the H2020 projects (IASIS_727658 and INFORE_825080) and the Spanish Ministry of Science and Innovation (RYC 2019-026752-I and PID 2020-117454RA-I00).

ACKNOWLEDGMENTS

The authors would like to thank Prof. Alberto Danielli and Prof. Marc Torrent Burgas for illuminating discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.893067/full#supplementary-material

Supplementary Table S1 | Twenty-one proteins shared by the four experimental studies. Uniprot IDs and gene names of the proteins are provided.

Supplementary Table S2 | Proteins found in each different experimental study.

Supplementary Table S3 | Dataset used for the evaluation of *cat*RAPID performances. For each protein, its Uniprot ID, experimental fold-change and normalized *cat*RAPID score are reported. The top experimental cases are highlighted in different colors according to the relative portion of the dataset.

Supplementary Table S4 | Sequence and genomic location of each of the 30 fragments of SARS-CoV-2.

Supplementary Table S5 | List of interactors for each experimental dataset, predicted to bind univocally each of the 30 SARS-CoV-2 fragments.

REFERENCES

Agostini, F., Cirillo, D., Bolognesi, B., and Tartaglia, G. G. (2013a). X-inactivation: Quantitative Predictions of Protein Interactions in the Xist Network. *Nucleic Acids Res.* 41, e31. doi:10.1093/nar/gks968

Agostini, F., Zanzoni, A., Klus, P., Marchese, D., Cirillo, D., and Tartaglia, G. G. (2013b). catRAPID Omics: a Web Server for Large-Scale Prediction of Protein-RNA Interactions. *Bioinformatics* 29, 2928–2930. doi:10.1093/bioinformatics/btt495

Alagar Boopathy, L. R., Jacob-Tomas, S., Alecki, C., and Vera, M. (2022). Mechanisms Tailoring the Expression of Heat Shock Proteins to Proteostasis Challenges. *J. Biol. Chem.*, 101796. doi:10.1016/j.jbc.2022.101796

Amaya, M., Brooks-Faulconer, T., Lark, T., Keck, F., Bailey, C., Raman, V., et al. (2016). Venezuelan Equine Encephalitis Virus Non-structural Protein 3 (nsP3) Interacts with RNA Helicases DDX1 and DDX3 in Infected Cells. *Antivir. Res.* 131, 49–60. doi:10.1016/j.antiviral.2016.04.008

Ariumi, Y. (2014). Multiple Functions of DDX3 RNA Helicase in Gene Regulation, Tumorigenesis, and Viral Infection. *Front. Genet.* 5. 423. doi:10.3389/fgene.2014.00423

Armaos, A., Colantoni, A., Proietti, G., Rupert, J., and Tartaglia, G. G. (2021). catRAPID Omics v2.0: Going Deeper and Wider in the Prediction of Protein-RNA Interactions. *Nucleic Acids Res.* 49, W72–W79. gkab393. doi:10.1093/nar/gkab393

Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M., and Grzybowska, E. A. (2019). RNA-protein Interactions: Disorder, Moonlighting and Junk Contribute to Eukaryotic Complexity. *Open Biol.* 9, 190096. doi:10.1098/rsob.190096

Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G. G. (2011). Predicting Protein Associations with Long Noncoding RNAs. *Nat. Methods* 8, 444–445. doi:10.1038/nmeth.1611

Bolognesi, B., Lorenzo Gotor, N., Dhar, R., Cirillo, D., Baldrighi, M., Tartaglia, G. G., et al. (2016). A Concentration-dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Rep.* 16, 222–231. doi:10.1016/j.celrep.2016.05.076

Brai, A., Martelli, F., Riva, V., Garbelli, A., Fazi, R., Zamperini, C., et al. (2019). DDX3X Helicase Inhibitors as a New Strategy to Fight the West Nile Virus Infection. *J. Med. Chem.* 62, 2333–2347. doi:10.1021/acs.jmedchem.8b01403

Brai, A., Riva, V., Saladini, F., Zamperini, C., Trivisani, C. I., Garbelli, A., et al. (2020). DDX3X Inhibitors, an Effective Way to Overcome HIV-1 Resistance Targeting Host Proteins. *Eur. J. Med. Chem.* 200, 112319. doi:10.1016/j.ejmech.2020.112319

Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., and Tartaglia, G. G. (2019). Phase Separation Drives X-Chromosome Inactivation: a Hypothesis. *Nat. Struct. Mol. Biol.* 26, 331–334. doi:10.1038/s41594-019-0223-0

Cerase, A., Calabrese, J. M., and Tartaglia, G. G. (2022). Phase Separation Drives X-Chromosome Inactivation. *Nat. Struct. Mol. Biol.* 29, 183–185. doi:10.1038/s41594-021-00697-0

Cerase, A., and Tartaglia, G. G. (2020). Long Non-coding RNA-Polycomb Intimate Rendezvous. *Open Biol.* 10, 200126. doi:10.1098/rsob.200126

Chen, Z., Mi, L., Xu, J., Yu, J., Wang, X., Jiang, J., et al. (2005). Function of HAb18G/CD147 in Invasion of Host Cells by Severe Acute Respiratory Syndrome Coronavirus. *J. Infect. Dis.* 191, 755–760. doi:10.1086/427811

Ciccosanti, F., Di Rienzo, M., Romagnoli, A., Colavita, F., Refolo, G., Castilletti, C., et al. (2021). Proteomic Analysis Identifies the RNA Helicase DDX3X as a Host Target against SARS-CoV-2 Infection. *Antivir. Res.* 190, 105064. doi:10.1016/j.antiviral.2021.105064

Cid-Samper, F., Gelabert-Baldrich, M., Lang, B., Lorenzo-Gotor, N., Ponti, R. D., Severijnen, L. W. F. M., et al. (2018). An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Rep.* 25, 3422–e7. e7. doi:10.1016/j.celrep.2018.11.076

Cirillo, D., Blanco, M., Armaos, A., Buness, A., Avner, P., Guttman, M., et al. (2017). Quantitative Predictions of Protein Interactions with Long Noncoding RNAs. *Nat. Methods* 14, 5–6. doi:10.1038/nmeth.4100

Delli Ponti, R., Marti, S., Armaos, A., and Tartaglia, G. G. (2017). A High-Throughput Approach to Profile RNA Structure. *Nucleic Acids Res.* 45, e35. doi:10.1093/nar/gkw1094

Doñate-Macián, P., Jungfleisch, J., Pérez-Vilaró, G., Rubio-Moscardo, F., Perálvarez-Marín, A., Diez, J., et al. (2018). The TRPV4 Channel Links Calcium Influx to DDX3X Activity and Viral Infectivity. *Nat. Commun.* 9, 2307. doi:10.1038/s41467-018-04776-7

Flynn, R. A., Belk, J. A., Qi, Y., Yasumoto, Y., Wei, J., Alfajaro, M. M., et al. (2021). Discovery and Functional Interrogation of SARS-CoV-2 RNA-Host Protein Interactions. *Cell* 184, 2394–2411. doi:10.1016/j.cell.2021.03.012

Foster, T. L., Gallay, P., Stonehouse, N. J., and Harris, M. (2011). Cyclophilin A Interacts with Domain II of Hepatitis C Virus NS5A and Stimulates RNA Binding in an Isomerase-dependent Manner. *J. Virol.* 85, 7460–7464. doi:10.1128/JVI.00393-11

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* 583, 459–468. doi:10.1038/s41586-020-2286-9

Gotor, N. L., Armaos, A., Calloni, G., Torrent Burgas, M., Vabulas, R. M., De Groot, N. S., et al. (2020). RNA-binding and Prion Domains: the Yin and Yang of Phase Separation. *Nucleic Acids Res.* 48, 9491–9504. doi:10.1093/nar/gkaa681

Iserman, C., Roden, C. A., Boerneke, M. A., Sealfon, R. S. G., McLaughlin, G. A., Jungreis, I., et al. (2020). Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell* 80, 1078–1091. e6. doi:10.1016/j.molcel.2020.11.041

Jachowicz, J. W., Strehle, M., Banerjee, A. K., Blanco, M. R., Thai, J., and Guttman, M. (2022). Xist Spatially Amplifies SHARP/SPEN Recruitment to Balance Chromosome-wide Silencing and Specificity to the X Chromosome. *Nat. Struct. Mol. Biol.* 29, 239–249. doi:10.1038/s41594-022-00739-1

Kamel, W., Noerenberg, M., Cerikan, B., Chen, H., Järvelin, A. I., Kammoun, M., et al. (2021). Global Analysis of Protein-RNA Interactions in SARS-CoV-2-Infected Cells Reveals Key Regulators of Infection. *Mol. Cell* 81, 2851–2867. doi:10.1016/j.molcel.2021.05.023

Kukhanova, M. K., Karpenko, I. L., and Ivanov, A. V. (2020). DEAD-box RNA Helicase DDX3: Functional Properties and Development of DDX3 Inhibitors as Antiviral and Anticancer Drugs. *Molecules* 25, 1015. doi:10.3390/molecules25041015

Lee, S., Lee, Y.-s., Choi, Y., Son, A., Park, Y., Lee, K.-M., et al. (2021). Young-sukThe SARS-CoV-2 RNA Interactome. *Mol. Cell* 81, 2838–2850. doi:10.1016/j.molcel.2021.04.022

Lu, S., Ye, Q., Singh, D., Cao, Y., Diedrich, J. K., Yates, J. R., et al. (2021). The SARS-CoV-2 Nucleocapsid Phosphoprotein Forms Mutually Exclusive Condensates with RNA and the Membrane-Associated M Protein. *Nat. Commun.* 12, 502. doi:10.1038/s41467-020-20768-y

Maga, G., Falchi, F., Radi, M., Botta, L., Casaluce, G., Bernardini, M., et al. (2011). Toward the Discovery of Novel Anti-HIV Drugs. Second-Generation Inhibitors of the Cellular ATPase DDX3 with Improved Anti-HIV Activity: Synthesis, Structure-Activity Relationship Analysis, Cytotoxicity Studies, and Target Validation. *ChemMedChem* 6, 1371–1389. doi:10.1002/cmdc.201100166

Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T. K., Marinus, T., et al. (2020). Genome-wide Mapping of SARS-CoV-2 RNA Structures Identifies Therapeutically-Relevant Elements. *Nucleic Acids Res.* 48, 12436–12452. doi:10.1093/nar/gkaa1053

Markaki, Y., Gan Chong, J., Wang, Y., Jacobson, E. C., Luong, C., Tan, S. Y. X., et al. (2021). Xist Nucleates Local Protein Gradients to Propagate Silencing across the X Chromosome. *Cell* 184, 6174–6192. doi:10.1016/j.cell.2021.10.022

McCormick, C., and Khaperskyy, D. A. (2017). Translation Inhibition and Stress Granules in the Antiviral Immune Response. *Nat. Rev. Immunol.* 17, 647–660. doi:10.1038/nri.2017.63

Mendez, A. S., Ly, M., González-Sánchez, A. M., Hartenian, E., Ingolia, N. T., Cate, J. H., et al. (2021). The N-Terminal Domain of SARS-CoV-2 Nsp1 Plays Key Roles in Suppression of Cellular Gene Expression and Preservation of Viral Gene Expression. *Cell Rep.* 37, 109841. doi:10.1016/j.celrep.2021.109841

Pöyry, T. A. A., Kaminski, A., Connell, E. J., Fraser, C. S., and Jackson, R. J. (2007). The Mechanism of an Exceptional Case of Reinitiation after Translation of a Long ORF Reveals Why Such Events Do Not Generally Occur in Mammalian mRNA Translation. *Genes Dev.* 21, 3149–3162. doi:10.1101/gad.439507

Schmidt, N., Lareau, C. A., Keshishian, H., Ganskih, S., Schneider, C., Hennig, T., et al. (2021). The SARS-CoV-2 RNA-Protein Interactome in Infected Human Cells. *Nat. Microbiol.* 6, 339–353. doi:10.1038/s41564-020-00846-z

Stunnenberg, M., Geijtenbeek, T. B. H., and Gringhuis, S. I. (2018). DDX3 in HIV-1 Infection and Sensing: A Paradox. *Cytokine. Growth Factor Rev.* 40, 32–39. doi:10.1016/j.cytogfr.2018.03.001

Suzuki, Y., Chin, W.-X., Han, Q. E., Ichiyama, K., Lee, C. H., Eyo, Z. W., et al. (2016). Characterization of RyDEN (C19orf66) as an Interferon-Stimulated Cellular Inhibitor against Dengue Virus Replication. *PLOS Pathog.* 12, e1005357. doi:10.1371/journal.ppat.1005357

Tartaglia, G. G., Dobson, C. M., Hartl, F. U., and Vendruscolo, M. (2010). Physicochemical Determinants of Chaperone Requirements. *J. Mol. Biol.* 400, 579–588. doi:10.1016/j.jmb.2010.03.066

Tartaglia, G. G. (2016). The Grand Challenge of Characterizing Ribonucleoprotein Networks. *Front. Mol. Biosci.* 3, 24. doi:10.3389/fmolb.2016.00024

Tong, X., Drapkin, R., Yalamanchili, R., Mosialos, G., and Kieff, E. (1995). The Epstein-Barr Virus Nuclear Protein 2 Acidic Domain Forms a Complex with a Novel Cellular Coactivator that Can Interact with TFIIE. *Mol. Cell. Biol.* 15, 4735–4744. doi:10.1128/MCB.15.9.4735

Vandelli, A., Cid Samper, F., Torrent Burgas, M., Sanchez de Groot, N., and Tartaglia, G. G. (2022). The Interplay between Disordered Regions in RNAs and Proteins Modulates Interactions within Stress Granules and Processing Bodies. *J. Mol. Biol.* 434, 167159. doi:10.1016/j.jmb.2021.167159

Vandelli, A., Monti, M., Milanetti, E., Armaos, A., Rupert, J., Zacco, E., et al. (2020). Structural Analysis of SARS-CoV-2 Genome and Predictions of the Human Interactome. *Nucleic Acids Res.* 48, 11270–11283. doi:10.1093/nar/gkaa864

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucleic Acids Res.* 38, W214–W220. doi:10.1093/nar/gkq537

Winnard, P. T., Vesuna, F., and Raman, V. (2021). Targeting Host DEAD-Box RNA Helicase DDX3X for Treating Viral Infections. *Antivir. Res.* 185, 104994. doi:10.1016/j.antiviral.2020.104994

Yedavalli, V. S. R. K., Neuveut, C., Chi, Y.-h., Kleiman, L., and Jeang, K.-T. (2004). Requirement of DDX3 DEAD Box RNA Helicase for HIV-1 Rev-RRE Export Function. *Cell* 119, 381–392. doi:10.1016/j.cell.2004.09.029

Zheng, Z.-Q., Wang, S.-Y., Xu, Z.-S., Fu, Y.-Z., and Wang, Y.-Y. (2021). SARS-CoV-2 Nucleocapsid Protein Impairs Stress Granule Formation to Promote Viral Replication. *Cell Discov.* 7, 38. doi:10.1038/s41421-021-00275-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Molecular Biosciences

TYPE Review
PUBLISHED 22 August 2022
DOI 10.3389/fmolb.2022.941653

OPEN ACCESS

EDITED BY
Gian Gaetano Tartaglia,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Yiran Liang,
Qilu Hospital, Shandong University, China
Xueping Lei,
Guangzhou Medical University, China
Han Jin,
Science for Life Laboratory (SciLifeLab), Sweden

*CORRESPONDENCE
Yuhong Bian,
bianyuhong_2012@163.com
Yingpeng Li,
liyingpeng@tjutcm.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 11 May 2022
ACCEPTED 18 July 2022
PUBLISHED 22 August 2022

CITATION
Fan N, Fu H, Feng X, Chen Y, Wang J, Wu Y, Bian Y and Li Y (2022), Long non-coding RNAs play an important regulatory role in tumorigenesis and tumor progression through aerobic glycolysis.
Front. Mol. Biosci. 9:941653.
doi: 10.3389/fmolb.2022.941653

Long non-coding RNAs play an important regulatory role in tumorigenesis and tumor progression through aerobic glycolysis

Ni Fan[1†], Hui Fu[2†], Xuchen Feng[1], Yatong Chen[1], Jingyu Wang[1], Yuqi Wu[1], Yuhong Bian[2*] and Yingpeng Li[1,3*]

[1]College of Chinese Materia Medica, Tianjin University of Traditional Chinese Medicine, Tianjin, China, [2]College of Integrated Chinese and Western Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, China, [3]Engineering Research Center of Modern Chinese Medicine Discovery and Preparation Technique, Ministry of Education, Tianjin University of Traditional Chinese Medicine, Tianjin, China

Compared to normal cells, cancer cells generate ATP mainly through aerobic glycolysis, which promotes tumorigenesis and tumor progression. Long non-coding RNAs (LncRNAs) are a class of transcripts longer than 200 nucleotides with little or without evident protein-encoding function. LncRNAs are involved in the ten hallmarks of cancer, interestingly, they are also closely associated with aerobic glycolysis. However, the mechanism of this process is non-transparent to date. Demonstrating the mechanism of lncRNAs regulating tumorigenesis and tumor progression through aerobic glycolysis is particularly critical for cancer therapy, and may provide novel therapeutic targets or strategies in cancer treatment. In this review, we discuss the role of lncRNAs and aerobic glycolysis in tumorigenesis and tumor progression, and further explore their interaction, in hope to provide a novel therapeutic target for cancer treatment.

KEYWORDS

lncRNAs, aerobic glycolysis, tumorigenesis, tumor progression, enzyme

1 Introduction

In recent years, the incidences of tumors have been increasing dramatically, the morbidity rates are increasing in the younger population (Sung et al., 2021). Cancer has become one of the leading causes of death in the human population (Feng et al., 2019; Mattiuzzi and Lippi, 2020). However, the occurrence of cancer involves a complex biological mechanism (Ping et al., 2021), and one of the critical mechanisms of cancer is altered metabolism (Shuvalov et al., 2021), which is cancer metabolic reprogramming. It provides tumor cells with the energy and structural resources necessary for excessive cell proliferation and growth, and has been widely regulated by activation of oncogenes or loss of tumor suppressors (Liu et al., 2019). The uniquely reprogrammed metabolic phenotype exhibited by tumor cells known as the Warburg effect or aerobic glycolysis, which is

characterized by high rates of aerobic glycolysis, leads to the production of lactic acid and reduced mitochondrial oxidative phosphorylation even in the presence of oxygen (Warburg, 1956; Liberti and Locasale, 2016). Some signaling pathways play important roles in glucose metabolism, such as PI3K/Akt/mTOR pathway, JAK/STAT pathway, P53 pathway, and so on.

Long non-coding RNAs (LncRNAs) are a heterogeneous group of transcripts of more than 200 nucleotides in length. They regulate numerous cellular processes, primarily through physical interaction with other molecules. LncRNAs are also involved in the ten hallmarks of cancer, including enabling replicative immortality (Chu et al., 2017), sustaining proliferative signaling (Wang et al., 2018b), evading growth suppressors (Hu et al., 2019), resisting cell death (Zhang et al., 2020b), tissue invasion and metastasis (Hanniford et al., 2020), inducing angiogenesis (Wang et al., 2020b), genome instability and mutation (Elguindy and Mendell, 2021), tumor-promoting inflammation (Ahmad et al., 2021), deregulating cell energetics (Chen et al., 2019b), avoiding immune destruction (Jiang et al., 2017). Multiple studies have documented an aberrant lncRNA expression in various cancers where they act as oncogenes or tumor suppressors (Fu et al., 2019; Barth et al., 2020).

Some lncRNAs contributed to cancers are involved in metabolic alterations (Balihodzic et al., 2021). LncRNAs-derived metabolism reprogramming allows cancer cells to maintain deregulated proliferation and withstand hostile microenvironments such as energy stress (Liu et al., 2019), and the probable mechanism of this is that lncRNAs can upregulate metabolic enzymes, disturb metabolic signaling pathways, and modulate the expression of oncogenic or tumor-suppressive genes (Balihodzic et al., 2021). Additionally, lncRNAs are also increasingly being considered potential therapeutic targets (Chandra Gupta and Nandan Tripathi, 2017). Thus, the interaction between lncRNAs and metabolic reprogramming, especially aerobic glycolysis needs further studies. In this review, we mainly discuss the connection between lncRNAs, aerobic glycolysis, and cancer, primarily illustrating the mechanism of lncRNAs regulating tumorigenesis and tumor progression through aerobic glycolysis.

2 Long non-coding RNAs play an important regulatory role in tumorigenesis and tumor progression

About 75% of the human genome can be transcribed into RNA, less than 2% of this group encode proteins, and the vast majority of transcripts are non-coding RNAs (ncRNAs) (2012). The proportion of ncRNAs in the human genome is much lower than that in low-level organisms, which suggests the importance of non-coding RNAs in biological evolution (Birney et al., 2007; ENCODE Project Consortium, 2012).

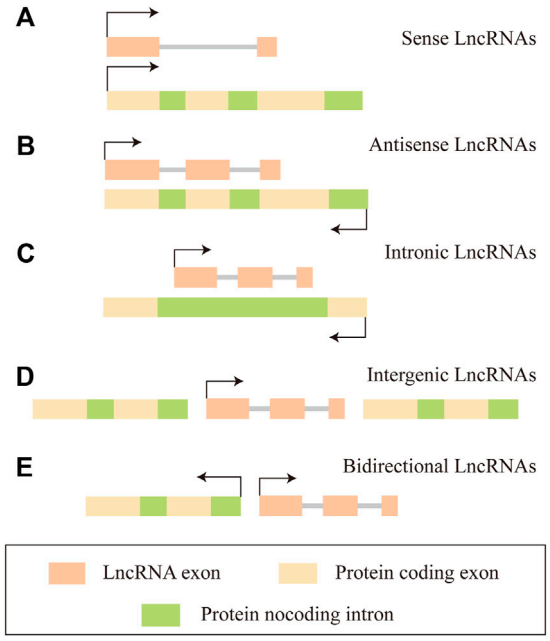


FIGURE 1
The classification of Long non-coding RNAs (LncRNAs). **(A)** Sense LncRNAs, it is transcribed by sense chain of protein coding gene, overlaps with at least one exon of protein coding gene on the same chain. **(B)** Antisense LncRNAs, it is transcribed by the complementary DNA strand of protein coding gene, overlaps with at least one exon of the positive gene. **(C)** Intronic LncRNAs, it is located in the intron region of a protein coding gene without overlapping with its exons. **(D)** Intergenic LncRNAs, it is located between two protein coding genes and can be transcribed independently. **(E)** Bidirectional LncRNAs, they share the same promoter with protein coding genes, but the transcription direction is opposite to protein coding genes.

NcRNAs can be divided into two main categories according to the molecular structure, linear RNA and circular RNA. The linear RNAs are further classified as short non-coding RNAs (ncRNAs) and long non-coding RNAs (lncRNAs). LncRNAs are composed of six types according to the genomic location: sense lncRNA, antisense lncRNA, intronic lncRNA, intergenic lncRNA, and bidirectional lncRNA (Zhang et al., 2020c) (Figure 1). In summary, lncRNAs regulating gene expressions at multiple levels, are defined as a group of transcripts longer than 200 nucleotides with little or without evident protein-encoding function (Batista and Chang, 2013). In the past decades, lncRNAs have gradually become one of the most popular molecules in biomedical research. Recent studies have shown that lncRNAs are associated with many important physiologic and pathologic processes, including differentiation, development, and disease. The molecular mechanism involved has been concluded as follows (Huang et al., 2020; Statello et al., 2021): 1) LncRNAs interact with specific protein to participate in global cellular processes by regulating protein activity or modulating protein-protein interactions as scaffolds to

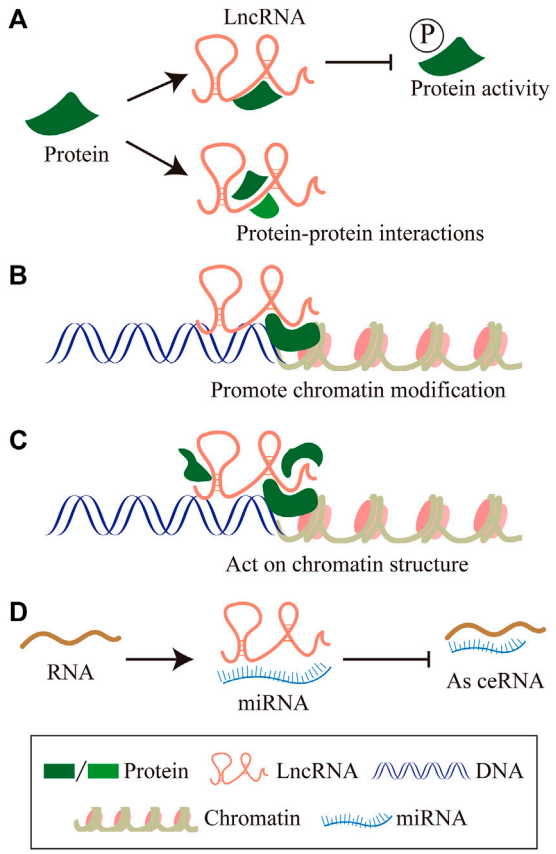


FIGURE 2
The molecular mechanism of LncRNAs. **(A)** LncRNAs interact with specific protein to participate in global cellular processes by regulating protein activity or modulating protein-protein interactions as scaffolds to facilitating the formation of protein complex. **(B)** LncRNAs recruit chromatin modification complexes to the promoter region of chromatin, and promote chromatin modification. **(C)** LncRNAs recruit the DNA of RNA binding proteins (RBPs), remodel chromatin structure, thus regulating the expression of target genes. **(D)** LncRNAs sequester miRNAs from target mRNAs as competitive endogenous RNA.

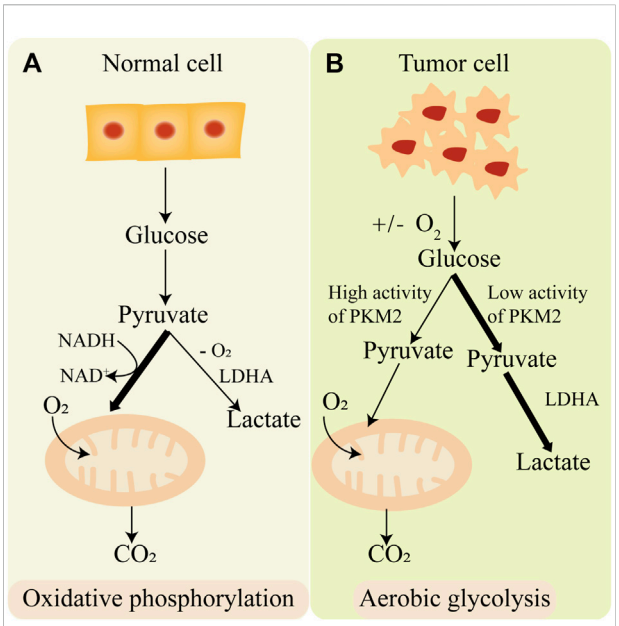


FIGURE 3
The difference of oxidative phosphorylation and aerobic glycolysis. **(A)** Normal cell produces energy mainly through oxidative phosphorylation. **(B)** Cancer cell produces energy mainly through aerobic glycolysis.

facilitating the formation of the protein complex. 2) LncRNAs recruit chromatin modification complexes to the promoter region of chromatin and promote chromatin modification. 3) LncRNAs recruit the DNA of RNA binding proteins (RBPs) and remodel chromatin structure, thus regulating the expression of target genes. 4) LncRNAs sequester miRNAs from target mRNAs as competitive endogenous RNA (Figure 2).

LncRNAs have been determined to be involved in regulating a variety of physiological processes during tumor development, including proliferation, apoptosis, metastasis, maintaining the stemness property of cancer stem cells (CSCs), tumor-related inflammation, etc. For example, LINC00926 inhibits breast cancer cell proliferation, invasion, and metastasis both *in vivo* and vitro (Chu et al., 2021). LncRNA HOTAIR functions as a ceRNA for tumor-suppressive miRNAs to induce the CSC phenotype in hepatocellular carcinoma (HCC) under hypoxia (Hu et al., 2020). LncRNA GNAS-AS1 redirected the polarization of macrophages in tumor microenvironment and promoted the migration and invasion of non-small cell lung cancer (Li et al., 2020c). LncRNA FENDRR potentiates tumorigenicity and cell growth in hepatocellular carcinoma mainly through suppressing Treg-mediated immune evasion of cancer cells by competitively bounding to miR-423-5p (Yu et al., 2019). LncRNA GAS5 could enhance the killing effect of NK cells on liver cancer by sponging miR-544 to target RUNX3, which augments NK cell cytotoxicity, IFN-γ secretion, and the proportion of $CD107a^+$ NK cells (Fang et al., 2019). Therefore, lncRNAs are promising targets in tumors because of their important roles in tumorigenesis and tumor progression.

Unfortunately, even though more than 40,000 lncRNAs have been found in human tissues, statistics from Human GENCODE suggest that the human genome contains more than 16,000 lncRNA genes (Fang et al., 2018; Uszczynska-Ratajczak et al., 2018), however, most of them have not been reported in the literature. Until now, the molecular mechanisms of lncRNAs in regulating biological processes are unknown to a large extent. Exploring the role of lncRNAs in eukaryotic cells, especially in cancer cells may reveal new rules and mechanisms for regulating physiological processes. It will annotate and clarify the structure

and function of the genome from a different perspective from protein-coding genes, and analyze the essence of life activity more comprehensively.

3 Aerobic glycolysis is closely associated with tumorigenesis and tumor progression

Glycolysis is an important process of cell glucose metabolism, during the process glucose degrades into acetone acid and synthesis ATP in cytoplasm (Gatenby and Gillies, 2004). Glycolysis is the initial enzymatic decomposition reaction of glucose in eukaryotic cells and bacteria. It has important physiological significance for cells in hypoxia and some special physiological and pathological conditions. In the 1920s, Warburg found that tumor cells employ aerobic glycolysis coupling with reduced mitochondrial oxidative phosphorylation for energy instead of oxidative phosphorylation, even in a sufficient oxygen state. This phenomenon is called "Warburg effect" (Warburg, 1956). On the contrary, normal cells produce ATP mainly through oxidative phosphorylation (Figure 3). One glucose molecule can synthesize 38 ATP molecules through oxidative phosphorylation, while synthesis of 2 ATP molecules through aerobic glycolysis (Feng et al., 2020). Although aerobic glycolysis synthesizes far less energy than oxidative phosphorylation, aerobic glycolysis synthesizes a large number of intermediates necessary for anabolism. Aerobic glycolysis has an important effect on tumor metabolic pathways, tumor microenvironment, and signaling pathways in tumors. Therefore, it is of great significance to explore the mechanisms of aerobic glycolysis in tumorigenesis and tumor progression, which may provide new therapeutic targets or strategies in tumor treatment.

3.1 Aerobic glycolysis laid the foundation of proliferation in tumors

Tumorigenesis is a relatively long-term and steady pathological process, phenotypically characterized by uncontrolled cell proliferation (Coleman and Parlo, 2021). Altered metabolisms, which are one of the critical hallmarks of cancer, provide tumor cells with the necessary energy and structural resources for rapid proliferation (Shuvalov et al., 2021). Besides, the large number of intermediates produced during aerobic glycolysis provides the necessary substrates for the rapid and sustained proliferation of tumor cells (Vaupel et al., 2019). Furthermore, compared with oxidative phosphorylation, aerobic glycolysis produces fewer reactive oxygen species to circumvent the deleterious effects of cancer cells. Numerous studies have documented that aerobic glycolysis is pivotal to support the rapid proliferation of malignant cells. The glycolytic enzyme Enolase 1 (ENO1) promotes serine/threonine kinase (AKT) activation to exert its metabolic effects, and the AKT/mTOR signaling pathway is essential for glucose metabolism. It is quite a coincidence that circRPN2 could bind to ENO1 and accelerates its degradation to promote glycolytic reprogramming through the AKT/mTOR pathway, thereby inhibiting hepatocellular carcinoma (HCC) proliferation and metastasis (Li et al., 2022). This indicates that aerobic glycolysis is closely related to tumor proliferation. Besides, pyruvate dehydrogenase kinase 1 (PDK1) was demonstrated to be overexpressed in non-small cell lung cancer tissues and promote cell proliferation and migration by modulating aerobic glycolysis (Liu and Yin, 2017). Similarly, the expression of pyruvate kinase M2 (PKM2) and glucose transporter1 (GLUT1), two important enzymes in aerobic glycolysis, was upregulated by Lin28A/SNHG14/IRF6 axis, thereby reprogramed glucose metabolism and stimulate cell proliferation in glioma cells (Lu et al., 2020). Hexokinase3 (HK3) is a key gene in aerobic glycolysis and participates in the first step of aerobic glycolysis (Kudryavtseva et al., 2016), it could promote the rapid proliferation, invasion, and metastasis of clear cell renal cancer cells by suppressing apoptosis and enhancing epithelial-mesenchymal transition (EMT) (Zhang et al., 2021).

In summary, aerobic glycolysis plays a necessary role in tumor proliferation. In contrast, inhibition of aerobic glycolysis maybe a promising method to suppress tumor growth and proliferation. For example, miR-30a-5p inhibits breast cancer cell proliferation, invasion, and metastasis both *in vitro* and *in vivo* by dampening aerobic glycolysis (Li et al., 2017).

3.2 Aerobic glycolysis laid the foundation of invasion and metastasis in tumors

Invasion and metastasis are the leading causes of cancer death (Romano et al., 2014). The ratio between glycolytic and oxidative ATP flux rate is potently associated with cancer invasion and metastasis behavior (Yizhak et al., 2014). Aerobic glycolysis provides nicotinamide adenine dinucleotide phosphate (NADPH) and ATP, as well as the carbon skeletons and intermediates obtained from its high rate of glucose fermentation, as components of biomolecular synthesis that are used by cancer cells to support their rapid growth and metastasis, invasiveness, and chemoresistance (Mullarky and Cantley, 2015; Abdel-Wahab et al., 2019). Unsurprisingly, a recent study showed that glycolysis is the major source of ATP production in endothelial cells and that the silencing of the glycolytic regulator 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3) impairs the cell metastasis capacity and interferes with vessel sprouting (De Bock et al., 2013). In addition, the expression changes of rate-limiting enzymes in aerobic glycolytic can greatly affect cancer metastasis. For example, the glycolytic enzyme ENO2 promotes the growth and invasion of clear cell renal cancer cells through aerobic

glycolysis (Zhang et al., 2021). Hexokinase (HK) is the first rate-limiting enzyme in aerobic glycolytic. It has been reported that the increased expression of HK3 is related to EMT in colorectal cancer, which is involved in the rapid growth and metastasis of colorectal cancer (Pudova et al., 2018). Moreover, the abnormally high expression of fructose-bisphosphatase 1 (FBP1), a downstream glycolysis enzyme and tumor suppressor which was recognized as a glycolysis inhibitor inhibits the invasion and metastasis of breast cancer (Shi et al., 2017). In contrast, the low expression of FBP1 promoted hepatocellular carcinoma cells metastasis through aerobic glycolysis (Yang et al., 2017). These enzymes accelerate the rate of glycolysis, producing faster and more ATP.

Interestingly, the generation of ATP and intermediates also promotes the generation of toxic levels of lactate (Doherty et al., 2014), which aids cancer metastasis and invasion mainly by causing an acidic environment. Many studies have found that a higher lactate level accompanied by enhanced aerobic glycolytic is significantly correlated with recurrence and high metastatic potential of tumors resulting in poor patient outcomes (Walenta et al., 2000). Another study reveals that lactate increases the uptake of folate and glucose, and further increases breast cancer cell T47D metastasis capacity (Guedes et al., 2016).

Consequently, the demand for developing drugs that could attenuate metastasis of malignancies *via* aerobic glycolysis is highly appreciated. Yi et al. reported in 2019 that betulinic acid, a pentacyclic triterpene widely found in birch bark extracts, could restrain breast cancer invasion and metastasis by inhibiting aerobic glycolysis through GRP78/PERK/β-catenin/c-Myc signaling pathway (Zheng et al., 2019b). Collectively, these findings provide a further prospect that the study of aerobic glycolysis inhibition is a promising new strategy for anti-invasion and anti-metastasis in malignancies.

3.3 Aerobic glycolysis laid the foundation of epithelial-mesenchymal transition in tumors

The initiation of metastasis is closely associated with EMT (Ribatti et al., 2020). EMT is a process in which epithelial cells lose their intercellular adherence and cellular polarity and acquire the mesenchymal phenotype. EMT is a crucial process in embryogenesis, organ fibrosis, and cancer metastasis (Greenburg and Hay, 1982). Recently, it has become apparent that EMT is tightly associated with aerobic glycolysis. Induction of EMT is associated with heightened rates of glycolysis and lactate production (Liu et al., 2016). Several studies have demonstrated that the master-regulators of EMT transcriptional factors such as zinc finger E-box binding homeobox 1 (Zeb1), snail zinc finger protein (Snail), and twist protein (Twist) are also able to regulate metabolic modulations (Georgakopoulos-Soares et al., 2020). Zeb1 was shown to induce aerobic glycolysis in pancreatic cancer cell models by repressing mitochondrial-localized tumor suppressor sirtuin 3 (SIRT3) (Krebs et al., 2017). Twist was shown to enhance glucose consumption and lactate production. Yang et al. (2015) found that Twist augments PKM2, Lactate dehydrogenase A (LDHA), and glucose-6-phosphate dehydrogenase (G6PD) by activating β1-integrin/FAK/PI3K/AKT/mTOR axis in MCF10A mammary epithelial cells. Similarly, Snail also promotes the metastatic spread by affecting glucose metabolism. Kim and his colleagues showed that Snail reprograms glucose metabolism by repressing phosphofructokinase (PFKP) which switches the glucose flux towards the pentose phosphate pathway in breast cancer (Jo et al., 2020). Besides, Snail-mediated increase in glucose uptake and lactate production was also demonstrated in gastric cancer. Snail-FBP1 signaling axis serves as an effective therapeutic target for primary tumor EMT and glucose metabolism reprogramming (Yu et al., 2017). Consequently, inhibiting aerobic glycolysis maybe a new avenue for suppressing EMT in cancer.

3.4 Aerobic glycolysis laid the foundation of apoptosis in tumors

The orderly and delicate regulation of apoptosis under physiological and pathological conditions is an autonomous clearance mechanism adopted by cells to maintain their homeostasis (Galluzzi et al., 2018). However, insufficient apoptosis induces neoplastic diseases, such as cancer. Increasing evidence indicates the close link between aerobic glycolysis and apoptosis resistance in tumor progression as well as poor patient outcomes (Gu et al., 2017). Metabolism can directly or indirectly regulate the apoptotic machinery, cancer cells regulate aerobic glycolysis to escape apoptosis (Matsuura et al., 2016). Lactate acid, the end product of aerobic glycolysis, induces the expression of glycolytic enzymes PFK1 in tumor cells, enhance the supply of ATP and resist cellular apoptosis. PFK15, a glycolytic inhibitor, could rapidly reduce the glucose uptake and induce apoptosis of lung carcinomas cancer cells both *in vitro* and vivo (Clem et al., 2013). Similarly, S100A10 activated mTOR pathway by interacting with annexin A2 to accelerate tumor aerobic glycolysis, promoted malignant proliferation, and suppressed cell apoptosis in gastric cancer (Li et al., 2020b).

In summary, aerobic glycolysis could affect cancer cell apoptosis, and aerobic glycolysis inhibition maybe a novel way to promote cancer cell apoptosis. Such as, several HK inhibitors, including the catalytic inhibitors 3-Bromopyruvate (3-BrPyr), Lonidamine, and the glucose-analogue, competitive inhibitor 2-Deoxyglucose (2-DG) both target HK2 in many tumor models, detach it from mitochondria and elicit tumor cell death (Garcia et al., 2019).

3.5 Aerobic glycolysis maintains the stemness property of cancer stem cells

Cancer stem cells are a rare subpopulation of cells that exhibit self-renewal properties and higher tumorigenicity compared with normal tumor cells (Magee et al., 2012). They are recognized as driving forces behind tumor growth (Mamouni et al., 2021). Emerging evidence indicates that metabolic reprogramming, especially the shift of glucose metabolism from mitochondrial oxidative phosphorylation to aerobic glycolysis which is known as Warburg effect, is a prerequisite step for the generation of the cancer stem cells, modulates the phenotype of cancer stem cells, and reshapes the tumor microenvironment (Chen et al., 2019a; Thakur and Chen, 2019). Recently, some studies revealed that aerobic glycolysis is tightly associated with the stemness property of cancer stem cells. And the product of aerobic glycolysis lactate further enhanced the stemness properties of cancer stem cells. Malignancies are recognized as a kind of devastating disease characterized by persistent hypoxia. Hypoxia-inducible factor-2α (HIF-2α), which is critical for tumor cells to adapt to the hypoxic microenvironment, enhances tumor stemness by elevating the expression level of stemness-associated transcriptional factors Nanog and Oct4 through classic Wnt/β-catenin signaling pathway (Zhang et al., 2017). β-arrestin1 (ARRB1) regulated the metabolic preference of bladder cancer stem cells and functioned as a molecular switch which promoted reprogramming towards glycolysis by negatively regulating mitochondrial pyruvate carrier 1 (MPC1) and positively regulating GLUT1 along with glucose uptake (Mamouni et al., 2021). The glycolytic enzyme ENO1 can improve the stemness of gastric cancer stem cells by prominently enhancing the cell's aerobic glycolysis (Yang et al., 2020). Besides, accumulating research have shown that chemoresistance of cancer stem cell results from dysregulation of glucose metabolism. Tao et al. demonstrated the chemoresistance of gemcitabine (GEM) in pancreatic cancer treatment is due to its metabolic reprogramming and cancer cell stemness enhancement. GEM treatment induced a metabolic shift from mitochondrial oxidation to aerobic glycolysis, which further promoted cancer cell stemness through KRAS/AMPK signaling (Zhao et al., 2019).

Interestingly, lncRNAs can regulate the stemness of cancer stem cells by regulating aerobic glycolysis. LncRNAs described in EMT regulation or metabolic reprogramming are freshly discovered to contribute to CSC creation and stemness maintenance by interacting with self-renewal transcriptional factors such as Nanog homeobox (NANOG), organic cation/carnitine transporter4 (OCT4) and SRY-box transcription factor 2 (SOX2). For example, lncRNA NEAT1 promoted glycolysis by regulating miRNAs (Tan et al., 2020). In addition, NEAT1 regulated CSC properties including self-renew and chemo-resistance in triple-negative breast cancer and lung cancer stem cells by upregulating the expression of pluripotency regulators SOX2, CD44 molecule (CD44), and aldehyde dehydrogenase (ALDH) in a Wnt signaling dependent manner (Shin et al., 2019). Therefore, aerobic glycolysis is closely related to the stemness property of cancer stem cells, and lncRNA is emerging as a pivotal regulator of aerobic glycolysis in CSC, but its overall role in CSC reprogramming needs to be further explored in future studies.

3.6 Aerobic glycolysis is closely associated with tumor immune microenvironment

Tumor immune microenvironment (TIME) refers to distinct populations of innate and adaptive immune cells, accompanied by different degrees and types of immune cell infiltration (Togo et al., 2020), including mast cells, macrophages, neutrophils, myeloid-derived suppressor cells (MDSCs), B cells, $CD4^+$ T helper (Th) cells, regulatory T cells (Tregs), $CD8^+$ cytotoxic T lymphocytes (CTLs), natural killer (NK) cells, dendritic cells (DCs) and some other innate immune cells. TIME influences the immune escape of cancer, and the response and outcome of immunotherapy (Zhang et al., 2020c). The immune response is related to the dramatic modifications in tissue metabolism, including the depletion of nutrients, the increase of oxygen consumption, and the production of reactive nitrogen and oxygen intermediates (Terry et al., 2020). Cancer cells suppress anti-tumor immune response by competing for nutrients or depleting nutrients or reducing the metabolic fitness of tumor-infiltrating immune cells in TIME (Guerra et al., 2020). Moreover, metabolites in TIME also influence immune cells differentiation and effector function. In recent years, the intimate relationship between multifaceted alterations in tumor metabolism and their subsequent influence on immune regulation has become increasingly recognized as an important factor contributing to tumorigenesis and tumor progression (Kesarwani et al., 2017). As we all know, tumor cells generate ATP mainly through aerobic glycolysis, subsequently promoting their proliferation (Altman et al., 2016). However, a new study described that enhanced aerobic glycolysis not only supports the proliferation of cancer cells but also supports the bioenergetic and biosynthetic needs of immune cells (Pearce and Pearce, 2013). Moreover, several clinical studies indicated that aerobic glycolytic activity in human tumors is negatively associated with host antitumor immune responses and the prognosis of anticancer immunotherapy (Jia et al., 2020). Furthermore, some rate-limiting enzymes in the glycolytic pathway act as hallmarks of malignancies, such as HK2 and PKM2, which are responsible for the regulation of immune evasion (Feng et al., 2020).

To date, many studies revealed some immune cells in TIME, such as T cells, and TAMs, which could affect tumorigenesis and tumor progression linked to aerobic glycolysis. T cells play an important role in anti-tumor defense because they exert a powerful immunogens-specific response against cancer cells

(Guerra et al., 2020). The emerging evidence indicated that the metabolic status of T cells is crucial for their anti-tumor functions through aerobic glycolysis, tumor cells release lactate into the TIME, which interferes with T cell survival and activation (Brand et al., 2016), reversely, TIME has a significant impact on T cell metabolism, differentiation, and function. A study revealed that excessive glucose uptake by tumor cells restricts the anti-cancer activity of tumor-infiltrating T cells, which leads to increased glycolytic capacity, dampened the mechanistic target of rapamycin kinase (mTOR) activity, and allows tumor progression (Chang et al., 2015). LDHA could catalyze pyruvate to lactate in tumor cells through aerobic glycolysis, knockdown LDHA in tumor cells neutralized TIME acidity, and promoted tumor infiltration by $CD8^+$ T cells and NK cells while decreasing the number of immunosuppressive immune cells (Zhang et al., 2019). TAMs are the major immunosuppressive components in TIME, accounting for a large proportion of the tumor mass, in addition, they are highly glycolytic and produce large amounts of lactate. A study demonstrated that mTOR signaling is closely connected with the polarization of TAMs from anti-tumoral M1Φ to pro-tumoral M2Φ, anti-tumoral M1Φ regulates aerobic glycolysis in cancer cells leading to reduced proliferation and decreased production of lactate, and lactate was the potent immunosuppressor and angiogenesis stimulator in the tumor microenvironment (Chen et al., 2020a). Furthermore, genetic deletion of LDHA, 2-DG administration, or mTORC1 inhibition has been proposed as therapeutic avenues designed to decrease glycolytic metabolism in cancer cells, reduce lactate in the TIME, and repolarize TAMs to a pro-inflammatory state (Vitale et al., 2019). Thus, affecting aerobic glycolysis might not only improve immune cell responses against highly immunogenetic cancers but also increase the immunogenicity of cancer cells, and targeting aerobic glycolysis might be a novel insight to regulating tumor immune microenvironment, finally inhibit the immune escape of cancer, tumorigenesis, and tumor progression.

4 Long non-coding RNAs regulate tumorigenesis and tumor progression through glycolytic enzymes and glucose transporter

The "Warburg effect" has been discovered nearly a hundred years ago, however, there are still many mysteries about the molecular mechanisms of aerobic glycolysis in tumors, especially the roles of lncRNAs in the regulation of aerobic glycolysis. Recent studies have implicated the roles of lncRNAs in the regulation of glycolysis. Some lncRNAs are reported to modulate the expression levels of glycolytic enzymes (Fan et al., 2017). Some lncRNAs are known to regulate glycolysis-related transcription factors or signaling pathways. For example, lincRNA-p21 (Yang et al., 2014), linc-AC020978 (Hua et al., 2020), and LINK-A (Lin et al., 2016), have been reported to promote glycolysis under hypoxic conditions through HIF1α. The lncRNA-MEG3 can depress aerobic glycolysis *via* p53 and functions as a tumor suppressor in liver cancer cells (Zhang et al., 2003; Zhou et al., 2007). These studies indicated that lncRNAs may play an important role in glycolysis, but the biological functions and underlying mechanisms should be evaluated in-depth. Particularly, there is no evidence confirming that lncRNAs regulate the balance between aerobic glycolysis and oxidative phosphorylation by directly regulating related signaling pathways. Here, we reviewed that lncRNAs could regulate tumorigenesis and tumor progression through glycolytic enzymes and glucose transporter.

4.1 Long non-coding RNAs regulate tumor metabolism and tumorigenesis, and tumor progression through the enzymes of aerobic glycolysis

4.1.1 Long non-coding RNAs regulate tumor metabolism and tumorigenesis, and tumor progression through pyruvate kinase M2

Pyruvate Kinase (PK) is a class of known glycolytic enzymes involved in the last step of glycolysis by converting phosphoenolpyruvate (PEP) to pyruvate. Four isoforms of pyruvate kinase have been identified: PKM1, PKM2, PKR, and PKL (Chen et al., 2020b). PKL and PKR, which are encoded by the PKLR gene, are expressed in some special tissues, such as liver, kidney, and erythrocytes (Massari et al., 2016), respectively, whereas PKM1 and PKM2, are encoded by the same PKM gene *via* alternative RNA splicing (Taniguchi et al., 2020). Compared to the other three subtypes, PKM2 is an embryonic isoform that is highly expressed in embryos with lower enzymatic activity than PKM1 (Christofk et al., 2008a). During embryogenesis, PKM2 is progressively replaced by the other three isoenzymes. Conversely, during tumorigenesis, PKM2 gradually replace the other three subtypes and showed a tendency of high expression in tumor tissues (Mazurek, 2011). For example, there is a switch from the PKM1 isoform to the PKM2 isoform in various cancers like glioblastoma and hepatocellular carcinoma, which enhances the level of aerobic glycolysis in tumors and promotes tumor formation and proliferation (Christofk et al., 2008b). Therefore, suppressing PKM2 is significantly important for inhibiting carcinogenesis and the development of carcinomas. PKM2pS37 is the best-studied form of PKM2 phosphorylation to date and triple-negative breast cancer is the most aggressive breast cancer subtype. A study demonstrated that PKM2pS37, as the prognostic indicator of triple negative breast cancer outcomes, has the potential to impact triple-negative breast cancer patients directly with CDK inhibitors and pyruvate kinase activators alone or in combination (Apostolidi et al., 2021). Consistent with these studies, we defined that the expression level of PKM2 is significantly different in four subtypes of breast cancer. As shown in Figure 4, PKM2 is relatively highly expressed in

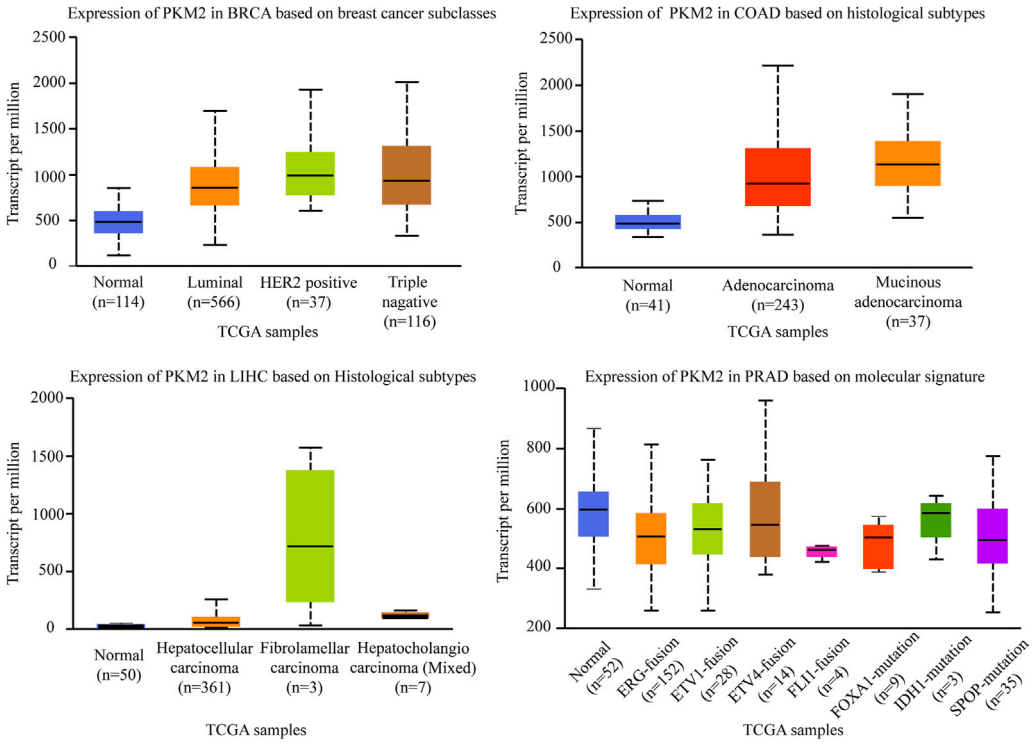


FIGURE 4
Association of PKM2 gene expression with histological subtypes and molecular signature in different cancers. BRCA, breast invasive carcinoma. COAD, colon adenocarcinoma. LIHC, liver hepatocellular carcinoma. PRAD, prostate adenocarcinoma.

luminal, HER2$^+$, and triple-negative breast cancers in the different subclasses of breast cancers. It is not the only case, in the different histological subtypes of colon adenocarcinomas, PKM2 is also relatively highly expressed in adenocarcinoma and mucinous adenocarcinoma. In the different histological subtypes of liver hepatocellular carcinomas, PKM2 especially showed the highest expression level in fibrolamellar carcinoma. However, although PKM2 showed relatively low expression levels in the ERG fusion, FOXA1 mutation, and SPOP mutation molecular signature of prostate adenocarcinomas, there were no significant differences in several other molecular signatures of prostate adenocarcinoma (The data comes from UALCAN).

According to recent studies, lncRNAs are involved in tumor metabolism and proliferation by regulating the enzymatic or transcriptional activity of PKM2 and related signaling pathways. LncRNA-FEZF1-AS1 could enhance aerobic glycolysis, promote proliferation and metastasis of colorectal cancer, and increase the stability and expression level of PKM2 in cytoplasmic and nucleus. And the increasing expression of PKM2 in cytoplasm promoted lactate production in colorectal cancer cells (Bian et al., 2018). LncRNA-WFDC21P markedly inhibited hepatocellular carcinoma cell's proliferation and metastasis *via* modulating the process of glycolysis by binding to PKM2 and suppressed its transcriptional activity, which inhibited the activity and nuclear translocation of HIF1α (Guan et al., 2020). LncRNA-HOXB-AS3 peptide suppressed colon cancer (CRC) growth through blocking hnRNPA1-mediated PKM splices, thereby inhibiting the formation of PKM2 and suppressing the reprogramming of the glucose metabolism (Huang et al., 2017). Several lncRNAs have been found to investigate their roles in liver cancer development and progression. Linc01554 downregulation empowers HCC cancer cells to acquire high aerobic glycolysis and sustain cells' growth advantages. Zheng YL et. demonstrated that Linc01554 plays a tumor suppressor role in regulating ubiquitin-mediated degradation of PKM2 and inhibiting Akt/mTOR signaling pathway (Zheng et al., 2019a). LncRNA-SOX2OT promotes HCC metastasis by upregulating PKM2, which increases the glycolytic pathway in HCC cells and thereby enhances EMT (Liang et al., 2020). Thus, PKM2 may be closely related to lncRNAs-mediated tumor pathogenesis and development. However, further studies are needed to explore the underlying molecular mechanism.

4.1.2 Long non-coding RNAs regulate tumor metabolism and tumorigenesis, tumor progression through lactate dehydrogenase A

Lactate dehydrogenase (LDH) comes from a family of NAD$^+$-dependent enzymes, which act as a tetrameric enzyme with distinct catalytic activity. The LDH family comprises two major subunits: A and/or B, resulting in five major isozymes: A4,

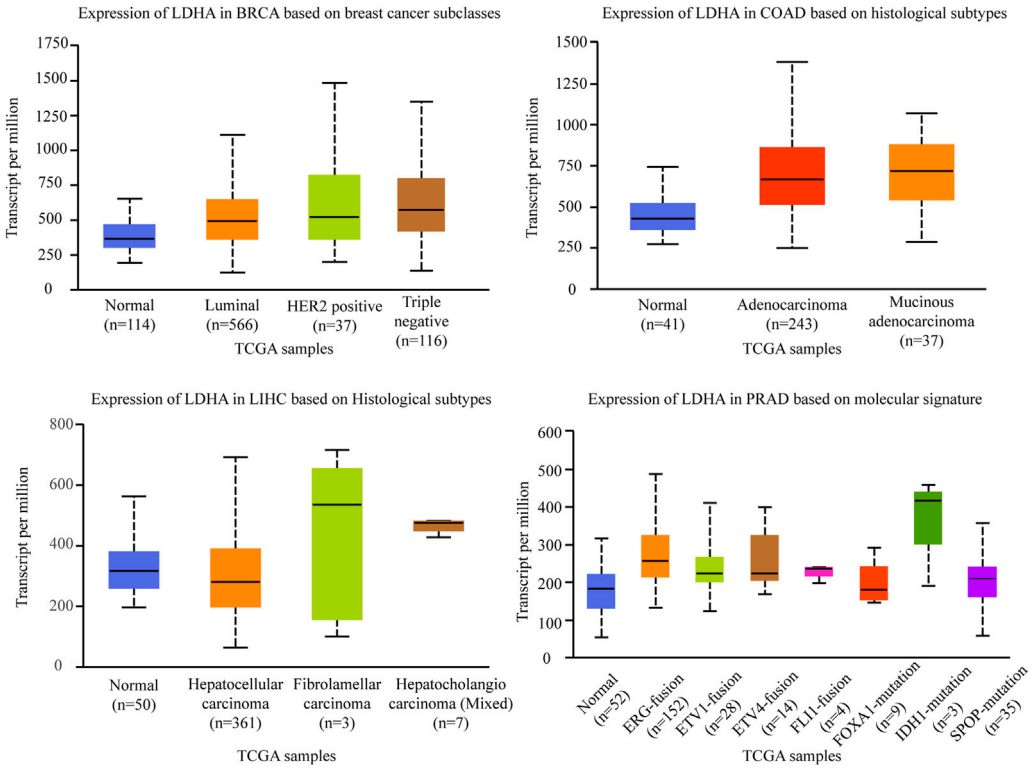


FIGURE 5
Association of LDHA gene expression with histological subtypes and molecular signature in different cancers. BRCA, breast invasive carcinoma. COAD, colon adenocarcinoma. LIHC, liver hepatocellular carcinoma. PRAD, prostate adenocarcinoma.

A3B1, A2B2, A1B3, and B4. LDH catalyzes the conversion of pyruvate to lactate and NADPH to Nicotinamide adenine dinucleotide (NADH) and produces energy in the form of ATP, which is the end product of glycolysis. Lactate dehydrogenase A (LDHA), also known as LDH-5, M-LDH, or A4, which is the predominant form in skeletal muscle, kinetically prefers the converting pyruvate to lactate in aerobic glycolysis in tumors. The overexpression of LDHA has been established in various solid cancers, including renal, pancreatic, non-small cell lung cancer, colorectal cancer, breast cancer, and other gynecologic cancers (Goldman et al., 1964). Molecular mechanism studies indicated that LDHA plays critical roles in tumor maintenance and aggravation, including promoting cancer cell proliferation, epithelial to mesenchymal transition (Jiang et al., 2016), angiogenesis (Giatromanolaki et al., 2006), cytoskeletal remodeling (Arseneault et al., 2013), cell motility, invasion and metastasis (Liu et al., 2015). Correspondingly, LDHA inhibition is shown to impair tumorigenesis and tumor growth (Le et al., 2010; Baig et al., 2019). Besides, it has been observed in several studies that inhibition of LDHA causes no significant toxic effect on normal tissue, which makes LDHA a promising therapeutic target in cancer (Tachtsidis et al., 2016). Such as, inhibiting LDHA with FX11 (LDHA inhibitor) suppressed pyruvate to lactate conversion, and caused reductions in ATP levels and substantial oxidative stress in cancer cells (Le et al., 2010). We evaluated the gene expression level of LDHA within different molecular or histological subtypes in four cancer types in TCGA publications. As shown in Figure 5, the gene expression analyzed are: 1) LDHA is relatively highly expressed in luminal, HER2^{+}, and triple-negative breast cancers in the different subclasses of breast cancers. 2) In different histological subtypes of colon adenocarcinomas, LDHA is also relatively highly expressed in adenocarcinoma and mucinous adenocarcinoma. 3) Similarly, in the different histological subtypes of liver hepatocellular carcinomas, LDHA is relatively highly expressed in hepatocholangio carcinoma (Mixed). 4) Similar result was also found in prostate adenocarcinoma, LDHA is especially highly expressed in ERG fusion, ETV1 fusion, ETV4 fusion, IDH1 mutation, and SPOP mutation subtype, there were no significant differences in several other molecular signatures of prostate adenocarcinomas such as FLI1 fusion and FOXA1 mutation (The data comes from UALCAN).

Recent studies have focused on the regulatory function of lncRNAs in cancer cell glucose metabolism through LDHA. LncRNA-IGFBP4-1 has been demonstrated to increase ATP production in lung cancer cells by upregulating metabolism

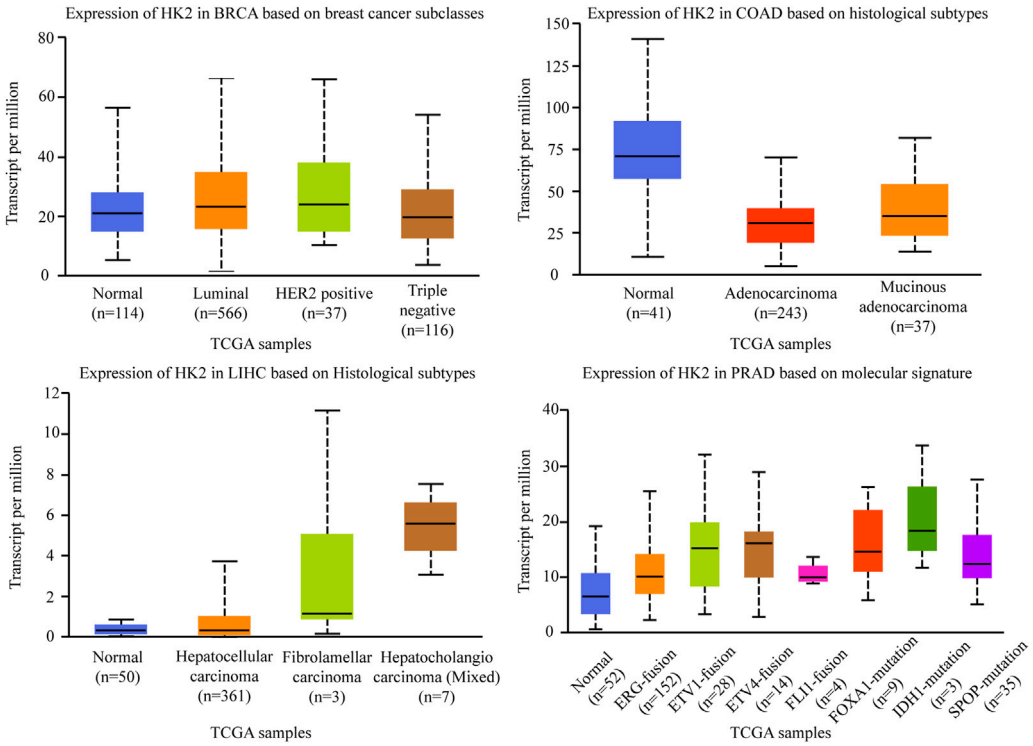


FIGURE 6
Association of HK2 gene expression with histological subtypes and molecular signature in different cancers. BRCA, breast invasive carcinoma. COAD, colon adenocarcinoma. LIHC, liver hepatocellular carcinoma. PRAD, prostate adenocarcinoma.

enzymes expression, including HK2, LDHA, and PDK1 (Yang et al., 2017a). Wang et al. (2020a) revealed that lncRNA HULC orchestrates enzymatic activities of two glycolytic enzymes, LDHA and PKM2, by directly adapting them to binding to fibroblast growth factor receptor type 1 (FGFR1), and eventually promoting Warburg effect in hepatocellular carcinoma cells. Over-activated metabolic signaling pathways were found to be involved in lncRNAs related to rapid ATP production and proliferation in cancer cells. For example, lncRNA ANRIL up-regulated the expression of Glut1 and LDHA by promoting the phosphorylation of Akt to activate the mTOR signal pathway (Zou et al., 2016). LncRNA RAET1K upregulated the expression of LDHA, decreased the expression of miR-100-5p, then upregulated the ICMT-Rac1 signaling pathway, thus promoting HCC metastasis (Zhou et al., 2020). Thus, LDHA may be another target that is related to lncRNAs and mediated tumor pathogenesis and development, but the underlying mechanism needs to be urgently investigated.

4.1.3 Long non-coding RNAs regulate tumor metabolism and tumorigenesis, tumor progression through Hexokinase2

Hexokinase (HK) is a group of rate-limiting enzymes in the glycolysis pathway, which using ATP as a phosphate donor, catalyzes the phosphorylation of glucose to produce glucose 6-phosphate (Lu and Hunter, 2018). There are four isoforms of HK: HK-I, HK-II, HK-III, and HK-IV. HK-I, and HK-III are ubiquitously expressed, while the expression of HK-IV, glucokinase, is restricted primarily to the liver and pancreas. Different from these three isoforms, HK-II (HK2) is the predominant isoform in insulin-sensitive tissues such as adipose, skeletal, and cardiac muscle. HK2 catalyzes the first committed step of glucose metabolism and initiates the major pathways of glucose utilization, it is confirmed to be a tumor promoter and plays an important regulating role in glucose metabolism in multiple malignancies, including breast cancer, lung cancer, and liver cancers (Kawai et al., 2005; Zhang et al., 2020a). Besides, HK2 has been recognized to regulate the malignant phenotype of cancer cells, such as cellular apoptosis and migratory capabilities (Lis et al., 2016; Chen J. et al., 2019c; Feng et al., 2020). Among the different subtypes of cancers, the expression of HK2 is diverse. As shown in Figure 6, there was no significant correlation between HK2 expression and molecular or histological subtypes in breast cancers. In the different molecular of colon adenocarcinomas, HK2 is relatively lowly expressed in adenocarcinoma and mucinous adenocarcinoma. While in the different histological subtypes of liver hepatocellular carcinomas, HK2 is relatively highly expressed in hepatocellular carcinoma. A

similar result was also found in prostate adenocarcinoma, HK2 is especially highly expressed in ERG fusion, ETV1 fusion, ETV4 fusion, FOXA1 mutation, IDH1 mutation, and SPOP mutation subtype, there was no significant difference in FLI1 fusion (The data comes from UALCAN).

Recently, there are several studies revealed that lncRNAs can regulate tumor metabolism and proliferation through HK2. Some studies indicated that lncRNA could act as a sponge of miRNA, enhance the expression of HK2, and promote cancer cell proliferation, metastasis, and invasion. For example, lncRNA PVT1 could act as a molecular sponge of miR-497, promote the expression of HK2, enhance the uptake of glucose and production of lactate, and promote osteosarcoma cell proliferation (Song et al., 2017). LncRNA C1QTNF1 could sponge miR-484 and consequently increase HK2 expression, promoting colorectal cancer cell proliferation, migration, and invasion (Jin et al., 2020a). LncRNA DUXAP8 could directly sponge miR-409-3p to regulate HK2 expression, and promote non-small-cell lung cancer cell growth, and metastasis (Yin et al., 2020). LncRNA HOTTIP promoted glycolysis under hypoxia by directly binding to miR-615-3p, acting as a molecular sponge, thus regulating the protein expression of hexokinase 2 (HK2) and high mobility group box 3 (HMGB3) (Shi et al., 2019). Interestingly, HOTTIP has also been demonstrated to stimulate CSC properties by mediating the activation of stemness and self-renewal transcriptional factors NANOG, OCT4, and SOX2 through Wnt/β-catenin signaling pathway (Han et al., 2020). Some studies indicated that lncRNA could regulate signaling pathways, and affect aerobic glycolysis and tumorigenesis. For example, lncRNA UCA1 could activate the mTOR pathway, mediate the regulation of urothelial cancer associated 1 (UCA1) to HK2 through activation of signal transducer and activator of transcription 3 (STAT3), then promote glycolysis, exert promotion effects on tumorigenesis in bladder cancer (Li et al., 2014). LncRNA BCAR4 coordinated the Hedgehog signaling pathway to enhance the transcription of glycolysis activators HK2, facilitating tumorigenesis in breast cancer (Zheng et al., 2017). Accordingly, HK2 may be related to lncRNA and co-regulated tumor pathogenesis and development, but the underlying mechanism needed to be further studied.

4.2 Long non-coding RNAs regulate tumor metabolism and tumorigenesis, tumor progression through glucose transporters1

Tumor cells have an increased dependence on extracellular glucose, thus glucose transporters (GLUTs) constitute also an anticancer target (Barbosa and Martel, 2020). The GLUT family is facultative transporter that play a vital role in glucose transport across the plasma membrane, which is the initial step of glycolysis. The family of transporters is composed of 14 members: GLUT1-GLUT12, GLUT14, and the H^+/myo-inositol transporter. Each of the GLUT transport proteins possesses different affinities for glucose. Among these, GLUT1, GLUT3, and GLUT4 have a high affinity for glucose, allowing the transport of glucose at a higher rate under normal physiological conditions (Mueckler and Thorens, 2013). GLUT2 found predominantly in liver, intestine, kidney, and pancreatic β-cells is a low-affinity glucose transport protein that is part of the glucose sensor in pancreatic β-cells and facilitates either glucose uptake or efflux from the liver depending on the nutritional state. GLUT3 is the glucose transporter responsible for maintaining an adequate that is responsible for insulin-regulated glucose disposal. Distinguished from several other glucose transporters, GLUT1 is a facilitative glucose transporter that belongs to the solute-linked carrier gene family SLC2 and is overexpressed ubiquitously in human cancer cells (Ganapathy et al., 2009), including breast, lung, renal, colorectal, and pancreatic cancers (Szablewski, 2013), which has potential effects on glycolysis process in cancer (Shang et al., 2020). Consistent with GLUT1's overexpression, GLUT1 is crucial for the uptake of glucose by breast cancer cells and is also the main glucose transporter in breast cancer cell lines (Wuest et al., 2018). Besides, the deregulation of GLUT1 is involved in the biological processes of tumor cells, including survival, growth, and metastasis (Goos et al., 2016).

Several studies have shown that lncRNAs can regulate tumor metabolism and proliferation through GLUT1. Some studies indicated that LncRNA could regulate signaling pathways to affect the expression of GLUT1, tumorigenesis, and tumor progression. For example, LncRNA HOTAIR induced GLUT1 expression *via* activating the mTOR pathway, promoting cell proliferation in hepatocellular carcinoma cells and tissues (Wei et al., 2017). LncRNA NBR2 promoted protein kinase AMP-activated catalytic subunit alpha 1 (AMPK) pathway to down-regulate GLUT1 expression and the EMT process, suppressed tumor progression in osteosarcoma cells (Liu and Gan, 2016). LncRNA SLC2A1-AS1 negatively regulated GLUT1 expression then inhibited STAT3 signaling pathway, and markedly decreased the proliferation and metastasis of hepatocellular carcinoma cells (Shang et al., 2020). Some studies indicated that lncRNAs associated with ceRNAs, based on the LncRNA-miRNA-mRNA network to exert their biological functions on the tumorigenesis and tumor progression. For example, Inc-p23154 could inhibit miR-378a-3p transcription, thereby enhancing GLUT1 expression and promoting oral squamous cell carcinoma metastasis (Wang et al., 2018a). LncRNA RAD51-AS1 acted as a sponge of miR-29b/c-3p, which in turn inhibited the expression of GLUT1, ultimately inhibiting proliferation, invasion, and glycolytic metabolism of colorectal cancer cells (Li et al., 2020a). LncRNA XIST functioned as a ceRNA to regulate the IRS1/PI3K/Akt pathway by sponging miR-126,

TABLE 1 Overview of LncRNAs related to the enzymes of aerobic glycolysis in cancer.

Enzyme	LncRNAs	Target	Function	Cancer type
Pyruvate kinase M2 (PKM2)	LncRNA-FEZF1-AS1	PKM2/STAT3	Increase the stability and expression level of PKM2 in cytoplasmic and nucleus, promote colorectal cancer cell proliferation and metastasis	Colorectal cancer
	LncRNA-WFDC21P	PKM2/HIF1α	Bind to PKM2 and suppress its transcriptional activity, inhibit hepatocellular carcinoma cell's proliferation and metastasis	Hepatocellular carcinoma
	LncRNA-HOXB-AS3	hnRNP A1/PKM2	Ensure the formation of lower PKM2, suppress colon cancer growth	Colon cancer
	Linc01554	PI3K/Akt/mTOR	Regulate ubiquitin-mediated degradation of PKM2, suppress hepatocellular carcinoma cancer cell growth	Hepatocellular carcinoma
	LncRNA-SOX2OT	miR-122-5p	Upregulate PKM2, promote hepatocellular carcinoma cancer cell metastasis	Hepatocellular carcinoma
Lactate dehydrogenase A (LDHA)	LncRNA-IGFBP4-1	IGFBP4	Upregulate LDHA expression, promote lung cancer progression	Lung cancer
	LncRNA HULC	FGFR1	Elevate phosphorylation and activity of LDHA, enhance hepatocellular carcinoma cancer cell proliferation	Hepatocellular carcinoma
	LncRNA ANRIL	PI3K/Akt/mTOR	Upregulate the expression of LDHA, promote nasopharyngeal carcinoma progression	Nasopharyngeal carcinoma
	LncRNA RAET1K	miR-100-5p	Upregulate the expression of LDHA, promote hepatocellular carcinoma cell metastasis	Hepatocellular carcinoma
Hexokinase 2 (HK2)	LncRNA PVT1	miR-497	Enhance the expression of HK2, promote osteosarcoma cell proliferation	Osteosarcoma
	LncRNA C1QTNF1	miR-484	Increase HK2 expression, promote colorectal cancer cell proliferation, migration, and invasion	Colorectal cancer
	LncRNA DUXAP8	miR-409-3p	Regulate HK2 expression, promote non-small-cell lung cancer cell growth, metastasis	Non-small-cell lung cancer
	LncRNA UCA1	mTOR-STAT3/miR-143	Mediate the regulation of UCA1 to HK2, exert promotion effects on tumorigenesis in bladder cancer	Bladder cancer
	LncRNA BCAR4	Hedgehog	Enhance the transcription of HK2, facilitate tumorigenesis in breast cancer	Breast cancer
Glucose transporter1 (GLUT1)	LncRNA HOTAIR	PI3K/Akt/mTOR	induce GLUT1 expression, promote hepatocellular carcinoma cell proliferation	Hepatocellular carcinoma
	LncRNA NBR2	AMPK	Down-regulate GLUT-1 expression and the EMT process, suppress tumor progression in osteosarcoma cell	Osteosarcoma
	LncRNA SLC2A1-AS1	STAT3	Negatively regulate GLUT1 expression, decrease the proliferation and metastasis of hepatocellular carcinoma cells	Hepatocellular carcinoma
	Lnc-P23154	miR-378a-3p	Enhance GLUT1 expression and promote oral squamous cell carcinoma metastasis	Oral squamous cell carcinoma
	LncRNA RAD51-AS1	miR-29b/c-3p	Inhibit the expression of GLUT1, inhibit proliferation, invasion and glycolytic metabolism of colorectal cancer cell	Colorectal cancer
	LncRNA XIST	miR-126/IRS1/PI3K/Akt	Elevate the expression of GLUT1, promote glioblastoma cell viability, migration, invasion, and resistance to apoptosis	Glioblastoma

elevated the expression of GLUT1, and promoted glioblastoma cell viability, migration, invasion, and resistance to apoptosis (Cheng et al., 2020) (Table 1; Figure 7). Thus, GLUT1 is also closely related to lncRNA and affects tumor metabolism and proliferation, but the underlying mechanism needed to be further studied.

5 Long non-coding RNAs serve as a promising target in the treatment of malignancies

To date, the biomarkers are commonly used in early screening for all kinds of cancers (Yuan et al., 2020), and lncRNAs have attracted increasing attention as cancer biomarkers for early screening, diagnosis, prognosis, and responses to drug treatment (Zhang et al., 2018; Necula et al., 2019; Zhuo et al., 2019). LncRNAs constitute an ever-growing category of functional RNA species known to impinge on all hallmarks of cancer (Napoli and Flores, 2020). Many studies revealed that lncRNAs could regulate many important pathological processes in cancer, such as tumorigenesis, tumor progression, proliferation, metastasis, and drug resistance (Fang and Fullwood, 2016; Schmitt and Chang, 2016), suggesting an enormous potential for further development of lncRNA biomarkers in specific cancer histologic analysis. The followings are some evidence that lncRNA can be used as a biomarker and is strongly associated with poor prognosis in

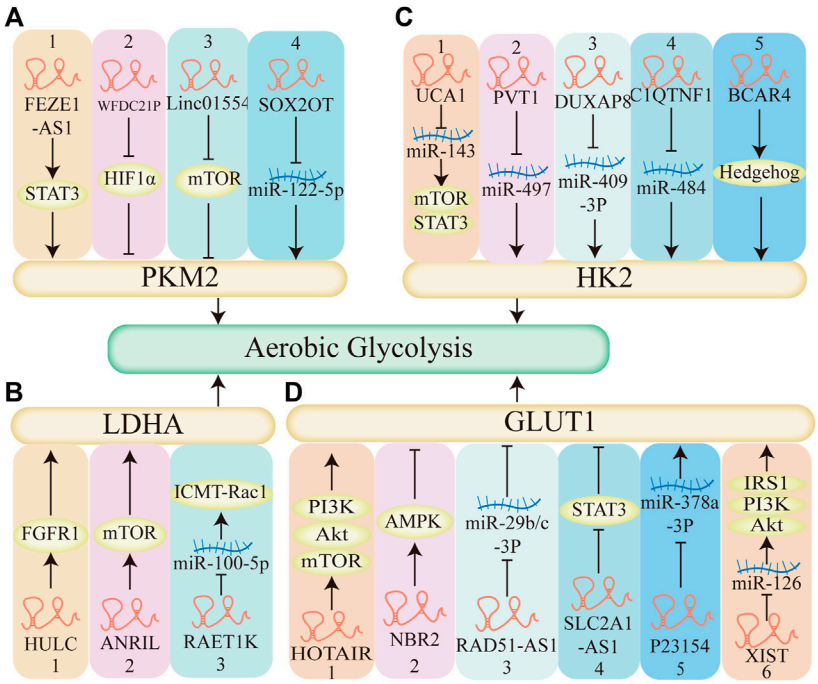


FIGURE 7
The mechanism of LncRNAs regulate aerobic glycolysis. **(A)** LncRNAs regulate aerobic glycolysis through pyruvate kinase M2 (PKM2). 1–3. LncRNA-FEZF1-AS1, LncRNA-WFDC21P, Linc01554 regulates STAT3, HIF1α, mTOR pathway respectively, affects the expression of PKM2, regulates aerobic glycolysis. 4. LncRNA-SOX2OT decreases the expression of miR-122-5p, upregulates PKM2, promotes aerobic glycolysis. **(B)** LncRNAs regulate aerobic glycolysis through Lactate dehydrogenase A (LDHA). 1–2. LncRNA HULC, LncRNA ANRIL elevates FGFR1 and mTOR, respectively, regulates LDHA, enhances aerobic glycolysis. 3.LncRNA RAET1K decreases the expression of miR-100-5P, upregulates the expression of LDHA, promotes aerobic glycolysis. **(C)** LncRNAs regulate aerobic glycolysis through hexokinase 2 (HK2). 1. LncRNA UCA1 acts as a molecular sponge of miR-143, activates the mTOR pathway, promotes aerobic glycolysis. 5. LncRNA BCAR4 coordinates the Hedgehog signaling pathway, enhances the HK2 and aerobic glycolysis. 2–4. LncRNA PVT1, LncRNA DUXAP8, LncRNA C1QTNF1 acts as a molecular sponge of miR-497, miR-409-3P, miR-484 respectively, promotes the expression of HK2, enhances aerobic glycolysis. **(D)** LncRNAs regulate aerobic glycolysis through glucose transporter1 (GLUT1). 1,2,4: LncRNA HOTAIR, LncRNA NBR2, LncRNA SLC2A1-AS1 regulates mTOR, AMPK, STAT3 pathway respectively and affects the expression of GLUT1, affects aerobic glycolysis. 3,5: LncRNA RAD51-AS1, LncRNA P23154, acts as a molecular sponge of miR-29b/c-3P, miR-378a-3P respectively, regulates the expression of GLUT1, affects aerobic glycolysis. 6. LncRNA XIAT acts as a molecular sponge of miR-126, activates the PI3K pathway and the expression of GLUT1, promotes aerobic glycolysis.

breast cancer, colorectal cancer, medulloblastoma, and renal cell carcinoma.

Song E et al. reported that the high expression level of HIFAL, a HIF-1α anti-sense lncRNA, is associated with aggressive breast cancer phenotype and poor prognosis. Mechanistically, HIFAL overexpression promotes tumor progression by forming a positive feed-forward loop with HIF-1α to enhance HIF-1α-mediated transactivation and glycolysis (Zheng et al., 2021). LncRNA TROJAN is highly expressed in Estrogen receptor-positive (ER^+) breast cancer tissues and promotes cancer proliferation. Thus, lncRNA TROJAN may serve as a potential therapeutic target for ER^+ breast cancer (Jin et al., 2020b).

LncRNA RAMS11 directly affects colorectal cancer biology, including promoting an aggressive phenotype and correlating with treatment response and resistance, indicating the potential value of lncRNA RAMS11 as a biomarker and therapeutic target for colorectal cancer (Silva-Fisher et al., 2020). LncRNA SNHG11 has been reported as a potential biomarker for the early detection of colon cancer and a new therapeutic target for this disease (Xu et al., 2020).

Lnc-HLX-2-7 is highly upregulated in Group 3 Medulloblastoma (MB) cell lines, promoted cell proliferation and 3D colony formation, inhibited cell apoptosis, indicating that Lnc-HLX-2-7 is oncogenic in MB and represents a promising novel molecular marker and a potential therapeutic target in Group 3 MBs (Katsushima et al., 2021).

LncRNA TRAF31P2-AS1 functions as a tumor suppressor in NONO-TFE3 translocation renal cell carcinoma progression and may serve as a novel target for NONO-TFE3 translocation renal cell carcinoma therapy (Yang et al., 2021).

Therefore, targeting lncRNAs may be a promising strategy to enhance chemosensitivity and improve the efficacy of chemotherapy (Arun et al., 2018). However, there is not enough clinical evidence indicating that lncRNAs could act as biomarkers for the treatment of cancer, more in-depth studies are required to accelerate the clinical applications of lncRNAs.

6 Conclusion and prospects

It is now clear that aerobic glycolysis is involved in a range of events important in cancer, including initiation, progression, metastasis, drug resistance, immune evasion, and the dynamic changes in immune microenvironment. As glucose metabolic disarrangement provides substrates for the biosynthesis of biomolecules essential for the rapid development of tumor. And the production of lactate also leads to a lower environmental pH and benefits tumor cell metastasis, invasion, and drug resistance. Moreover, there is ample evidence that high lactate levels have immune-modulatory properties. LncRNAs participate in each of these events by transcriptional, post-transcriptional, and epigenetic gene-regulatory mechanisms.

As master regulators of gene expression, the mis-regulation of lncRNAs expression has been demonstrated to be the driver of tumorigenesis and development associated with metabolic disarrangement. Most commonly, lncRNAs sponge to corresponding miRNA and mRNA that target critical metabolic enzymes, such as PKM2, LDAH, HK2, and glucose transporter GLUT1, to modulate the expression of numerous oncogenes and tumor-suppression genes. Furthermore, lncRNAs also play critical roles in generating an immune-permissive microenvironment by modulating aerobic glycolysis. The reversible transition of EMT and mesenchymal-epithelial transition (MET) is a key event in tumor progression, metastasis, and invasiveness into normal tissues. And the processes of EMT and MET are also highly regulated by lncRNAs. The discovery of lncRNAs in the regulation of "stemness" further broadens its opportunities for the treatment of cancer. Therefore, continued study of lncRNAs in preclinical research will yield new insights into RNA-based therapeutics in cancer.

Author contributions

Writing—original draft preparation, NF; writing—review and editing, HF; editing the final version of the manuscript, NF, HF, XF, YC, JW, YW, YB, and YL. Financial support for the project, HF and YL. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by National Natural Science Foundation of China, grant numbers 82074030 and 82104568.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abdel-Wahab, A. F., Mahmoud, W., and Al-Harizy, R. M. (2019). Targeting glucose metabolism to suppress cancer progression: Prospective of anti-glycolytic cancer therapy. *Pharmacol. Res.* 150, 104511. doi:10.1016/j.phrs.2019.104511

Ahmad, S., Abbas, M., Ullah, M. F., Aziz, M. H., Beylerli, O., Alam, M. A., et al. (2021). Long non-coding RNAs regulated NF-κB signaling in cancer metastasis: Micromanaging by not so small non-coding RNAs. *Semin. Cancer Biol.* S1044-579X (21), 00210–00218. doi:10.1016/j.semcancer.2021.07.015

Altman, B., Stine, Z., and Dang, C. (2016). From krebs to clinic: Glutamine metabolism to cancer therapy. *Nat. Rev. Cancer* 16 (11), 749. doi:10.1038/nrc.2016.114

Apostolidi, M., Vathiotis, I. A., Muthusamy, V., Gaule, P., Gassaway, B. M., Rimm, D. L., et al. (2021). Targeting pyruvate kinase M2 phosphorylation reverses aggressive cancer phenotypes. *Cancer Res.* 81 (16), 4346–4359. doi:10.1158/0008-5472.CAN-20-4190

Arseneault, R., Chien, A., Newington, J., Rappon, T., Harris, R., and Cumming, R. (2013). Attenuation of LDHA expression in cancer cells leads to redox-dependent alterations in cytoskeletal structure and cell migration. *Cancer Lett.* 338 (2), 255–266. doi:10.1016/j.canlet.2013.03.034

Arun, G., Diermeier, S., and Spector, D. (2018). Therapeutic targeting of long non-coding RNAs in cancer. *Trends Mol. Med.* 24 (3), 257–277. doi:10.1016/j.molmed.2018.01.001

Baig, M., Adil, M., Khan, R., Dhadi, S., Ahmad, K., Rabbani, G., et al. (2019). Enzyme targeting strategies for prevention and treatment of cancer: Implications for cancer therapy. *Semin. Cancer Biol.* 56, 1–11. doi:10.1016/j.semcancer.2017.12.003

Balihodzic, A., Barth, D., Prinz, F., and Pichler, M. (2021). Involvement of long non-coding RNAs in glucose metabolism in cancer. *Cancers* 13 (5), 977. doi:10.3390/cancers13050977

Barbosa, A., and Martel, F. (2020). Targeting glucose transporters for breast cancer therapy: The effect of natural and synthetic compounds. *Cancers* 12 (1), E154. doi:10.3390/cancers12010154

Barth, D., Juracek, J., Slaby, O., Pichler, M., and Calin, G. (2020). lncRNA and mechanisms of drug resistance in cancers of the genitourinary system. *Cancers* 12 (8), E2148. doi:10.3390/cancers12082148

Batista, P., and Chang, H. (2013). Long noncoding RNAs: Cellular address codes in development and disease. *Cell* 152 (6), 1298–1307. doi:10.1016/j.cell.2013.02.012

Bian, Z., Zhang, J., Li, M., Feng, Y., Wang, X., Zhang, J., et al. (2018). LncRNA-FEZF1-AS1 promotes tumor proliferation and metastasis in colorectal cancer by regulating PKM2 signaling. *Clin. Cancer Res.* 24 (19), 4808–4819. doi:10.1158/1078-0432.Ccr-17-2967

Birney, E., Stamatoyannopoulos, J., Dutta, A., Guigó, R., Gingeras, T., Margulies, E., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447 (7146), 799–816. doi:10.1038/nature05874

Brand, A., Singer, K., Koehl, G., Kolitzus, M., Schoenhammer, G., Thiel, A., et al. (2016). LDHA-associated lactic acid production blunts tumor immunosurveillance by T and NK cells. *Cell Metab.* 24 (5), 657–671. doi:10.1016/j.cmet.2016.08.011

Chandra Gupta, S., and Nandan Tripathi, Y. (2017). Potential of long non-coding RNAs in cancer patients: From biomarkers to therapeutic targets. *Int. J. Cancer* 140 (9), 1955–1967. doi:10.1002/ijc.30546

Chang, C., Qiu, J., O'Sullivan, D., Buck, M., Noguchi, T., Curtis, J., et al. (2015). Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell* 162 (6), 1229–1241. doi:10.1016/j.cell.2015.08.016

Chen, B., Gao, A., Tu, B., Wang, Y., Yu, X., Wang, Y., et al. (2020a). Metabolic modulation *via* mTOR pathway and anti-angiogenesis remodels tumor microenvironment using PD-L1-targeting codelivery. *Biomaterials* 255, 120187. doi:10.1016/j.biomaterials.2020.120187

Chen, C., Bai, L., Cao, F., Wang, S., He, H., Song, M., et al. (2019a). Targeting LIN28B reprograms tumor glucose metabolism and acidic microenvironment to suppress cancer stemness and metastasis. *Oncogene* 38 (23), 4527–4539. doi:10.1038/s41388-019-0735-4

Chen, F., Chen, J., Yang, L., Liu, J., Zhang, X., Zhang, Y., et al. (2019b). Extracellular vesicle-packaged HIF-1α-stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells. *Nat. Cell Biol.* 21 (4), 498–510. doi:10.1038/s41556-019-0299-0

Chen, J., Yu, Y., Li, H., Hu, Q., Chen, X., He, Y., et al. (2019c). Long non-coding RNA PVT1 promotes tumor progression by regulating the miR-143/HK2 axis in gallbladder cancer. *Mol. Cancer* 18 (1), 33. doi:10.1186/s12943-019-0947-9

Chen, X., Chen, S., and Yu, D. (2020b). Protein kinase function of pyruvate kinase M2 and cancer. *Cancer Cell Int.* 20 (1), 523. doi:10.1186/s12935-020-01612-1

Cheng, Z., Luo, C., and Guo, Z. (2020). LncRNA-XIST/microRNA-126 sponge mediates cell proliferation and glucose metabolism through the IRS1/PI3K/Akt pathway in glioma. *J. Cell. Biochem.* 121 (3), 2170–2183. doi:10.1002/jcb.29440

Christofk, H. R., Vander Heiden, M. G., Harris, M. H., Ramanathan, A., Gerszten, R. E., Wei, R., et al. (2008b). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* 452 (7184), 230–233. doi:10.1038/nature06734

Christofk, H., Vander Heiden, M., Harris, M., Ramanathan, A., Gerszten, R., Wei, R., et al. (2008a). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* 452 (7184), 230–233. doi:10.1038/nature06734

Chu, H.-P., Cifuentes-Rojas, C., Kesner, B., Aeby, E., Lee, H.-G., Wei, C., et al. (2017). TERRA RNA antagonizes ATRX and protects telomeres. *Cell* 170 (1), 86–101. doi:10.1016/j.cell.2017.06.017

Chu, Z., Huo, N., Zhu, X., Liu, H., Cong, R., Ma, L., et al. (2021). FOXO3A-induced LINC00926 suppresses breast tumor growth and metastasis through inhibition of PGK1-mediated Warburg effect. *Mol. Ther.* 29 (9), 2737–2753. doi:10.1016/j.ymthe.2021.04.036

Clem, B., O'Neal, J., Tapolsky, G., Clem, A., Imbert-Fernandez, Y., Kerr, D., et al. (2013). Targeting 6-phosphofructo-2-kinase (PFKFB3) as a therapeutic strategy against cancer. *Mol. Cancer Ther.* 12 (8), 1461–1470. doi:10.1158/1535-7163.Mct-13-0097

Coleman, P., and Parlo, R. (2021). Warburg's ghost-cancer's self-sustaining phenotype: The aberrant carbon flux in cholesterol-enriched tumor mitochondria *via* deregulated cholesterogenesis. *Front. Cell Dev. Biol.* 9, 626316. doi:10.3389/fcell.2021.626316

De Bock, K., Georgiadou, M., Schoors, S., Kuchnio, A., Wong, B., Cantelmo, A., et al. (2013). Role of PFKFB3-driven glycolysis in vessel sprouting. *Cell* 154 (3), 651–663. doi:10.1016/j.cell.2013.06.037

Doherty, J. R., Yang, C., Scott, K. E. N., Cameron, M. D., Fallahi, M., Li, W., et al. (2014). Blocking lactate export by inhibiting the Myc target MCT1 Disables glycolysis and glutathione synthesis. *Cancer Res.* 74 (3), 908–920. doi:10.1158/0008-5472.CAN-13-2034

Elguindy, M. M., and Mendell, J. T. (2021). NORAD-induced Pumilio phase separation is required for genome stability. *Nature* 595 (7866), 303–308. doi:10.1038/s41586-021-03633-w

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489 (7414), 57–74. doi:10.1038/nature11247

Fan, C., Tang, Y., Wang, J., Xiong, F., Guo, C., Wang, Y., et al. (2017). Role of long non-coding RNAs in glucose metabolism in cancer. *Mol. Cancer* 16 (1), 130. doi:10.1186/s12943-017-0699-3

Fang, P., Xiang, L., Chen, W., Li, S., Huang, S., Li, J., et al. (2019). LncRNA GAS5 enhanced the killing effect of NK cell on liver cancer through regulating miR-544/RUNX3. *Innate Immun.* 25 (2), 99–109. doi:10.1177/1753425919827632

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., et al. (2018). NONCODEV5: A comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* 46 (D1), D308–D314. doi:10.1093/nar/gkx1107

Fang, Y., and Fullwood, M. (2016). Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genomics Proteomics Bioinforma.* 14 (1), 42–54. doi:10.1016/j.gpb.2015.09.006

Feng, J., Li, J., Wu, L., Yu, Q., Ji, J., Wu, J., et al. (2020). Emerging roles and the regulation of aerobic glycolysis in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* 39 (1), 126. doi:10.1186/s13046-020-01629-4

Feng, R., Zong, Y., Cao, S., and Xu, R. (2019). Current cancer situation in China: Good or bad news from the 2018 global cancer statistics? *Cancer Commun.* 39 (1), 22. doi:10.1186/s40880-019-0368-6

Fu, P., Zheng, X., Fan, X., and Lin, A. (2019). Role of cytoplasmic lncRNAs in regulating cancer signaling pathways. *J. Zhejiang Univ. Sci. B* 20 (1), 1–8. doi:10.1631/jzus.B1800254

Galluzzi, L., Vitale, I., Aaronson, S., Abrams, J., Adam, D., Agostinis, P., et al. (2018). Molecular mechanisms of cell death: Recommendations of the nomenclature committee on cell death 2018. *Cell Death Differ.* 25 (3), 486–541. doi:10.1038/s41418-017-0012-4

Ganapathy, V., Thangaraju, M., and Prasad, P. (2009). Nutrient transporters in cancer: Relevance to Warburg hypothesis and beyond. *Pharmacol. Ther.* 121 (1), 29–40. doi:10.1016/j.pharmthera.2008.09.005

Garcia, S. N., Guedes, R. C., and Marques, M. M. (2019). Unlocking the potential of HK2 in cancer metabolism and therapeutics. *Curr. Med. Chem.* 26 (41), 7285–7322. doi:10.2174/0929867326666181213092652

Gatenby, R., and Gillies, R. (2004). Why do cancers have high aerobic glycolysis? *Nat. Rev. Cancer* 4 (11), 891–899. doi:10.1038/nrc1478

Georgakopoulos-Soares, I., Chartoumpekis, D., Kyriazopoulou, V., and Zaravinos, A. (2020). EMT factors and metabolic pathways in cancer. *Front. Oncol.* 10, 499. doi:10.3389/fonc.2020.00499

Giatromanolaki, A., Sivridis, E., Gatter, K., Turley, H., Harris, A., Koukourakis, M., et al. (2006). Lactate dehydrogenase 5 (LDH-5) expression in endometrial cancer relates to the activated VEGF/VEGFR2(KDR) pathway and prognosis. *Gynecol. Oncol.* 103 (3), 912–918. doi:10.1016/j.ygyno.2006.05.043

Goldman, R., Kaplan, N., and Hall, T. (1964). Lactic dehydrogenase in human neoplastic tissues. *Cancer Res.* 24, 389–399.

Goos, J., de Cuba, E., Coupé, V., Diosdado, B., Delis-Van Diemen, P., Karga, C., et al. (2016). Glucose transporter 1 (SLC2A1) and vascular endothelial growth factor A (VEGFA) predict survival after resection of colorectal cancer liver metastasis. *Ann. Surg.* 263 (1), 138–145. doi:10.1097/sla.0000000000001109

Greenburg, G., and Hay, E. (1982). Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. *J. Cell Biol.* 95 (1), 333–339. doi:10.1083/jcb.95.1.333

Gu, J., Li, Y., Zeng, J., Wang, B., Ji, K., Tang, Y., et al. (2017). Knockdown of HIF-1α by siRNA-expressing plasmid delivered by attenuated Salmonella enhances the antitumor effects of cisplatin on prostate cancer. *Sci. Rep.* 7 (1), 7546. doi:10.1038/s41598-017-07973-4

Guan, Y., Huang, Q., Ai, Y., Chen, Q., Zhao, W., Wang, X., et al. (2020). Nur77-activated lncRNA WFDC21P attenuates hepatocarcinogenesis *via* modulating glycolysis. *Oncogene* 39 (11), 2408–2423. doi:10.1038/s41388-020-1158-y

Guedes, M., Araújo, J., Correia-Branco, A., Gregório, I., Martel, F., and Keating, E. (2016). Modulation of the uptake of critical nutrients by breast cancer cells by lactate: Impact on cell survival, proliferation and migration. *Exp. Cell Res.* 341 (2), 111–122. doi:10.1016/j.yexcr.2016.01.008

Guerra, L., Bonetti, L., and Brenner, D. (2020). Metabolic modulation of immunity: A new concept in cancer immunotherapy. *Cell Rep.* 32 (1), 107848. doi:10.1016/j.celrep.2020.107848

Han, L., Yan, Y., Zhao, L., Liu, Y., Lv, X., Zhang, L., et al. (2020). LncRNA HOTTIP facilitates the stemness of breast cancer *via* regulation of miR-148a-3p/WNT1 pathway. *J. Cell. Mol. Med.* 24 (11), 6242–6252. doi:10.1111/jcmm.15261

Hanniford, D., Ulloa-Morales, A., Karz, A., Berzoti-Coelho, M. G., Moubarak, R. S., Sánchez-Sendra, B., et al. (2020). Epigenetic silencing of CDR1as drives IGF2BP3-mediated melanoma invasion and metastasis. *Cancer Cell* 37 (1), 55–70. doi:10.1016/j.ccell.2019.12.007

Hu, M., Fu, Q., Jing, C., Zhang, X., Qin, T., and Pan, Y. (2020). LncRNA HOTAIR knockdown inhibits glycolysis by regulating miR-130a-3p/HIF1A in hepatocellular carcinoma under hypoxia. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 125, 109703. doi:10.1016/j.biopha.2019.109703

Hu, Q., Ye, Y., Chan, L.-C., Li, Y., Liang, K., Lin, A., et al. (2019). Oncogenic lncRNA downregulates cancer cell antigen presentation and intrinsic tumor suppression. *Nat. Immunol.* 20 (7), 835–851. doi:10.1038/s41590-019-0400-7

Hua, Q., Mi, B., Xu, F., Wen, J., Zhao, L., Liu, J., et al. (2020). Hypoxia-induced lncRNA-AC020978 promotes proliferation and glycolytic metabolism of non-small cell lung cancer by regulating PKM2/HIF-1α axis. *Theranostics* 10 (11), 4762–4778. doi:10.7150/thno.43839

Huang, J.-Z., Chen, M., Chen, D., Gao, X.-C., Zhu, S., Huang, H., et al. (2017). A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol. Cell* 68 (1), 171–184. e176. doi:10.1016/j.molcel.2017.09.015

Huang, Z., Zhou, J.-K., Peng, Y., He, W., and Huang, C. (2020). The role of long noncoding RNAs in hepatocellular carcinoma. *Mol. Cancer* 19 (1), 77. doi:10.1186/s12943-020-01188-4

Jia, Y., Yang, Q., Wang, Y., Li, W., Chen, X., Xu, T., et al. (2020). Hyperactive PI3Kδ predisposes naive T cells to activation *via* aerobic glycolysis programs. *Cell. Mol. Immunol.* 18, 1783–1797. doi:10.1038/s41423-020-0379-x

Jiang, F., Ma, S., Xue, Y., Hou, J., and Zhang, Y. (2016). LDH-A promotes malignant progression *via* activation of epithelial-to-mesenchymal transition and conferring stemness in muscle-invasive bladder cancer. *Biochem. Biophys. Res. Commun.* 469 (4), 985–992. doi:10.1016/j.bbrc.2015.12.078

Jiang, R., Tang, J., Chen, Y., Deng, L., Ji, J., Xie, Y., et al. (2017). The long noncoding RNA lnc-EGFR stimulates T-regulatory cells differentiation thus promoting hepatocellular carcinoma immune evasion. *Nat. Commun.* 8, 15129. doi:10.1038/ncomms15129

Jin, S., Liu, Y., Wang, W., and Li, Z. (2020a). Long non-coding RNA C1QTNF1 antisense RNA 1 upregulates hexokinase 2 by sponging microRNA-484 to promote the malignancy of colorectal cancer. *Cancer Manag. Res.* 12, 12053–12066. doi:10.2147/cmar.S262096

Jin, X., Ge, L., Li, D., Shao, Z., Di, G., Xu, X., et al. (2020b). LncRNA TROJAN promotes proliferation and resistance to CDK4/6 inhibitor *via* CDK2 transcriptional activation in ER+ breast cancer. *Mol. Cancer* 19 (1), 87. doi:10.1186/s12943-020-01210-9

Jo, H., Lee, J., Jeon, J., Kim, S., Chung, J., Ko, H., et al. (2020). The critical role of glucose deprivation in epithelial-mesenchymal transition in hepatocellular carcinoma under hypoxia. *Sci. Rep.* 10 (1), 1538. doi:10.1038/s41598-020-58124-1

Katsushima, K., Lee, B., Kunhiraman, H., Zhong, C., Murad, R., Yin, J., et al. (2021). The long noncoding RNA lnc-HLX-2-7 is oncogenic in Group 3 medulloblastomas. *Neuro. Oncol.* 23 (4), 572–585. doi:10.1093/neuonc/noaa235

Kawai, S., Mukai, T., Mori, S., Mikami, B., and Murata, K. (2005). Hypothesis: Structures, evolution, and ancestor of glucose kinases in the hexokinase family. *J. Biosci. Bioeng.* 99 (4), 320–330. doi:10.1263/jbb.99.320

Kesarwani, P., Kant, S., Prabhu, A., and Chinnaiyan, P. (2017). The interplay between metabolic remodeling and immune regulation in glioblastoma. *Neuro. Oncol.* 19 (10), 1308–1315. doi:10.1093/neuonc/nox079

Krebs, A., Mitschke, J., Lasierra Losada, M., Schmalhofer, O., Boerries, M., Busch, H., et al. (2017). The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat. Cell Biol.* 19 (5), 518–529. doi:10.1038/ncb3513

Kudryavtseva, A., Fedorova, M., Zhavoronkov, A., Moskalev, A., Zasedatelev, A., Dmitriev, A., et al. (2016). Effect of lentivirus-mediated shRNA inactivation of HK1, HK2, and HK3 genes in colorectal cancer and melanoma cells. *BMC Genet.* 17, 156. doi:10.1186/s12863-016-0459-1

Le, A., Cooper, C., Gouw, A., Dinavahi, R., Maitra, A., Deck, L., et al. (2010). Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression. *Proc. Natl. Acad. Sci. U. S. A.* 107 (5), 2037–2042. doi:10.1073/pnas.0914433107

Li, C., Wang, P., Du, J., Chen, J., Liu, W., and Ye, K. (2020a). LncRNA RAD51-AS1/miR-29b/c-3p/NDRG2 crosstalk repressed proliferation, invasion and glycolysis of colorectal cancer. *IUBMB life* 73, 286–298. doi:10.1002/iub.2427

Li, J., Hu, Z.-Q., Yu, S.-Y., Mao, L., Zhou, Z.-J., Wang, P.-C., et al. (2022). CircRPN2 inhibits aerobic glycolysis and metastasis in hepatocellular carcinoma. *Cancer Res.* 82 (6), 1055–1069. doi:10.1158/0008-5472.CAN-21-1259

Li, L., Kang, L., Zhao, W., Feng, Y., Liu, W., Wang, T., et al. (2017). miR-30a-5p suppresses breast tumor growth and metastasis through inhibition of LDHA-mediated Warburg effect. *Cancer Lett.* 400, 89–98. doi:10.1016/j.canlet.2017.04.034

Li, Y., Li, X., Li, L., Zhou, R., Sikong, Y., Gu, X., et al. (2020b). S100A10 accelerates aerobic glycolysis and malignant growth by activating mTOR-signaling pathway in gastric cancer. *Front. Cell Dev. Biol.* 8, 559486. doi:10.3389/fcell.2020.559486

Li, Z., Feng, C., Guo, J., Hu, X., and Xie, D. (2020c). GNAS-AS1/miR-4319/NECAB3 axis promotes migration and invasion of non-small cell lung cancer cells by altering macrophage polarization. *Funct. Integr. Genomics* 20 (1), 17–28. doi:10.1007/s10142-019-00696-x

Li, Z., Li, X., Wu, S., Xue, M., and Chen, W. (2014). Long non-coding RNA UCA1 promotes glycolysis by upregulating hexokinase 2 through the mTOR-STAT3/microRNA143 pathway. *Cancer Sci.* 105 (8), 951–955. doi:10.1111/cas.12461

Liang, Y., Zhang, D., Zheng, T., Yang, G., Wang, J., Meng, F., et al. (2020). lncRNA-SOX2OT promotes hepatocellular carcinoma invasion and metastasis through miR-122-5p-mediated activation of PKM2. *Oncogenesis* 9 (5), 54. doi:10.1038/s41389-020-0242-z

Liberti, M. V., and Locasale, J. W. (2016). The Warburg effect: How does it benefit cancer cells? *Trends biochem. Sci.* 41 (3), 211–218. doi:10.1016/j.tibs.2015.12.001

Lin, A., Li, C., Xing, Z., Hu, Q., Liang, K., Han, L., et al. (2016). The LINK-A lncRNA activates normoxic HIF1α signalling in triple-negative breast cancer. *Nat. Cell Biol.* 18 (2), 213–224. doi:10.1038/ncb3295

Lis, P., Dyląg, M., Niedźwiecka, K., Ko, Y., Pedersen, P., Goffeau, A., et al. (2016). The HK2 dependent "Warburg effect" and mitochondrial oxidative phosphorylation in cancer: Targets for effective therapy with 3-bromopyruvate. *Mol. (Basel, Switz.* 21 (12), E1730. doi:10.3390/molecules21121730

Liu, H., Luo, J., Luan, S., He, C., and Li, Z. (2019). Long non-coding RNAs involved in cancer metabolic reprogramming. *Cell. Mol. Life Sci.* 76 (3), 495–504. doi:10.1007/s00018-018-2946-1

Liu, M., Quek, L., Sultani, G., and Turner, N. (2016). Epithelial-mesenchymal transition induction is associated with augmented glucose uptake and lactate production in pancreatic ductal adenocarcinoma. *Cancer Metab.* 4, 19. doi:10.1186/s40170-016-0160-x

Liu, T., and Yin, H. (2017). PDK1 promotes tumor cell proliferation and migration by enhancing the Warburg effect in non-small cell lung cancer. *Oncol. Rep.* 37 (1), 193–200. doi:10.3892/or.2016.5253

Liu, X., and Gan, B. (2016). lncRNA NBR2 modulates cancer cell sensitivity to phenformin through GLUT1. *Cell cycleGeorget. Tex.)* 15 (24), 3471–3481. doi:10.1080/15384101.2016.1249545

Liu, X., Yang, Z., Chen, Z., Chen, R., Zhao, D., Zhou, Y., et al. (2015). Effects of the suppression of lactate dehydrogenase A on the growth and invasion of human gastric cancer cells. *Oncol. Rep.* 33 (1), 157–162. doi:10.3892/or.2014.3600

Lu, J., Liu, X., Zheng, J., Song, J., Liu, Y., Ruan, X., et al. (2020). Lin28A promotes IRF6-regulated aerobic glycolysis in glioma cells by stabilizing SNHG14. *Cell Death Dis.* 11 (6), 447. doi:10.1038/s41419-020-2650-6

Lu, Z., and Hunter, T. (2018). Metabolic kinases moonlighting as protein kinases. *Trends biochem. Sci.* 43 (4), 301–310. doi:10.1016/j.tibs.2018.01.006

Magee, J., Piskounova, E., and Morrison, S. (2012). Cancer stem cells: Impact, heterogeneity, and uncertainty. *Cancer cell* 21 (3), 283–296. doi:10.1016/j.ccr.2012.03.003

Mamouni, K., Kim, J., Lokeshwar, B., and Kallifatidis, G. (2021). ARRB1 regulates metabolic reprogramming to promote glycolysis in stem cell-like bladder cancer cells. *Cancers* 13 (8), 1809. doi:10.3390/cancers13081809

Massari, F., Ciccarese, C., Santoni, M., Iacovelli, R., Mazzucchelli, R., Piva, F., et al. (2016). Metabolic phenotype of bladder cancer. *Cancer Treat. Rev.* 45, 46–57. doi:10.1016/j.ctrv.2016.03.005

Matsuura, K., Canfield, K., Feng, W., and Kurokawa, M. (2016). Metabolic regulation of apoptosis in cancer. *Int. Rev. Cell Mol. Biol.* 327, 43–87. doi:10.1016/bs.ircmb.2016.06.006

Mattiuzzi, C., and Lippi, G. (2020). Cancer statistics: A comparison between world health organization (WHO) and global burden of disease (GBD). *Eur. J. Public Health* 30 (5), 1026–1027. doi:10.1093/eurpub/ckz216

Mazurek, S. (2011). Pyruvate kinase type M2: A key regulator of the metabolic budget system in tumor cells. *Int. J. Biochem. Cell Biol.* 43 (7), 969–980. doi:10.1016/j.biocel.2010.02.005

Mueckler, M., and Thorens, B. (2013). The SLC2 (GLUT) family of membrane transporters. *Mol. Asp. Med.* 34, 121–138. doi:10.1016/j.mam.2012.07.001

Mullarky, E., and Cantley, L. C. (2015). "Diverting glycolysis to combat oxidative stress," in *Innovative medicine: Basic research and development*. Editors K. Nakao, N. Minato, and S. Uemoto (Tokyo: Springer), 3–23. 2015Copyright , The Author(s).

Napoli, M., and Flores, E. (2020). The p53 family reaches the final frontier: The variegated regulation of the dark matter of the genome by the p53 family in cancer. *RNA Biol.* 17 (11), 1636–1647. doi:10.1080/15476286.2019.1710054

Necula, L., Matei, L., Dragu, D., Neagu, A., Mambet, C., Nedeianu, S., et al. (2019). Recent advances in gastric cancer early diagnosis. *World J. Gastroenterol.* 25 (17), 2029–2044. doi:10.3748/wjg.v25.i17.2029

Pearce, E., and Pearce, E. (2013). Metabolic pathways in immune cell activation and quiescence. *Immunity* 38 (4), 633–643. doi:10.1016/j.immuni.2013.04.005

Ping, Q., Yan, R., Cheng, X., Wang, W., Zhong, Y., Hou, Z., et al. (2021). Cancer-associated fibroblasts: Overview, progress, challenges, and directions. *Cancer Gene Ther.* 28, 984–999. doi:10.1038/s41417-021-00318-4

Pudova, E., Kudryavtseva, A., Fedorova, M., Zaretsky, A., Shcherbo, D., Lukyanova, E., et al. (2018). HK3 overexpression associated with epithelial-mesenchymal transition in colorectal cancer. *BMC genomics* 19, 113. doi:10.1186/s12864-018-4477-4

Ribatti, D., Tamma, R., and Annese, T. (2020). Epithelial-mesenchymal transition in cancer: A historical overview. *Transl. Oncol.* 13 (6), 100773. doi:10.1016/j.tranon.2020.100773

Romano, S., D'Angelillo, A., Romano, A., Nappo, G., and Romano, M. F. (2014). Cellular and molecular background underlying the diversity in therapeutic responses between primary tumours and metastases. *Curr. Med. Chem.* 21 (14), 1631–1638. doi:10.2174/0929867313209990225

Schmitt, A., and Chang, H. (2016). Long noncoding RNAs in cancer pathways. *Cancer cell* 29 (4), 452–463. doi:10.1016/j.ccell.2016.03.010

Shang, R., Wang, M., Dai, B., Du, J., Wang, J., Liu, Z., et al. (2020). Long noncoding RNA SLC2A1-AS1 regulates aerobic glycolysis and progression in hepatocellular carcinoma *via* inhibiting the STAT3/FOXM1/GLUT1 pathway. *Mol. Oncol.* 14 (6), 1381–1396. doi:10.1002/1878-0261.12666

Shi, J., Wang, H., Feng, W., Huang, S., An, J., Qiu, Y., et al. (2019). Long non-coding RNA HOTTIP promotes hypoxia-induced glycolysis through targeting miR-615-3p/HMGB3 axis in non-small cell lung cancer cells. *Eur. J. Pharmacol.* 862, 172615. doi:10.1016/j.ejphar.2019.172615

Shi, L., He, C., Li, Z., Wang, Z., and Zhang, Q. (2017). FBP1 modulates cell metabolism of breast cancer cells by inhibiting the expression of HIF-1α. *Neoplasma* 64 (4), 535–542. doi:10.4149/neo_2017_407

Shin, V. Y., Chen, J., Cheuk, I. W.-Y., Siu, M.-T., Ho, C.-W., Wang, X., et al. (2019). Long non-coding RNA NEAT1 confers oncogenic role in triple-negative breast cancer through modulating chemoresistance and cancer stemness. *Cell Death Dis.* 10 (4), 270. doi:10.1038/s41419-019-1513-5

Shuvalov, O., Daks, A., Fedorova, O., Petukhov, A., and Barlev, N. (2021). Linking metabolic reprogramming, plasticity and tumor progression. *Cancers* 13 (4), 762. doi:10.3390/cancers13040762

Silva-Fisher, J., Dang, H., White, N., Strand, M., Krasnick, B., Rozycki, E., et al. (2020). Long non-coding RNA RAMS11 promotes metastatic colorectal cancer progression. *Nat. Commun.* 11 (1), 2156. doi:10.1038/s41467-020-15547-8

Song, J., Wu, X., Liu, F., Li, M., Sun, Y., Wang, Y., et al. (2017). Long non-coding RNA PVT1 promotes glycolysis and tumor progression by regulating miR-497/HK2 axis in osteosarcoma. *Biochem. Biophys. Res. Commun.* 490 (2), 217–224. doi:10.1016/j.bbrc.2017.06.024

Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22 (2), 96–118. doi:10.1038/s41580-020-00315-9

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Szablewski, L. (2013). Expression of glucose transporters in cancers. *Biochim. Biophys. Acta* 1835 (2), 164–169. doi:10.1016/j.bbcan.2012.12.004

Tachtsidis, A., McInnes, L., Jacobsen, N., Thompson, E., and Saunders, C. (2016). Minimal residual disease in breast cancer: An overview of circulating and disseminated tumour cells. *Clin. Exp. Metastasis* 33 (6), 521–550. doi:10.1007/s10585-016-9796-8

Tan, X., Wang, P., Lou, J., and Zhao, J. (2020). Knockdown of lncRNA NEAT1 suppresses hypoxia-induced migration, invasion and glycolysis in anaplastic thyroid carcinoma cells through regulation of miR-206 and miR-599. *Cancer Cell Int.* 20, 132. doi:10.1186/s12935-020-01222-x

Taniguchi, K., Uchiyama, K., and Akao, Y. (2020). PTBP1-targeting microRNAs regulate cancer-specific energy metabolism through the modulation of PKM1/M2 splicing. *Cancer Sci.* 112, 41–50. doi:10.1111/cas.14694

Terry, S., Engelsen, A., Buart, S., Elsayed, W., Venkatesh, G., and Chouaib, S. (2020). Hypoxia-driven intratumor heterogeneity and immune evasion. *Cancer Lett.* 492, 1–10. doi:10.1016/j.canlet.2020.07.004

Thakur, C., and Chen, F. (2019). Connections between metabolism and epigenetics in cancers. *Semin. Cancer Biol.* 57, 52–58. doi:10.1016/j.semcancer.2019.06.006

Togo, M., Yokobori, T., Shimizu, K., Handa, T., Kaira, K., Sano, T., et al. (2020). Diagnostic value of 18F-FDG-PET to predict the tumour immune status defined by tumoural PD-L1 and CD8+tumour-infiltrating lymphocytes in oral squamous cell carcinoma. *Br. J. Cancer* 122 (11), 1686–1694. doi:10.1038/s41416-020-0820-z

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* 19 (9), 535–548. doi:10.1038/s41576-018-0017-y

Vaupel, P., Schmidberger, H., and Mayer, A. (2019). The Warburg effect: Essential part of metabolic reprogramming and central contributor to cancer progression. *Int. J. Radiat. Biol.* 95 (7), 912–919. doi:10.1080/09553002.2019.1589653

Vitale, I., Manic, G., Coussens, L., Kroemer, G., and Galluzzi, L. (2019). Macrophages and metabolism in the tumor microenvironment. *Cell Metab.* 30 (1), 36–50. doi:10.1016/j.cmet.2019.06.001

Walenta, S., Wetterling, M., Lehrke, M., Schwickert, G., Sundfør, K., Rofstad, E., et al. (2000). High lactate levels predict likelihood of metastases, tumor recurrence, and restricted patient survival in human cervical cancers. *Cancer Res.* 60 (4), 916–921.

Wang, C., Li, Y., Yan, S., Wang, H., Shao, X., Xiao, M., et al. (2020a). Interactome analysis reveals that lncRNA HULC promotes aerobic glycolysis through LDHA and PKM2. *Nat. Commun.* 11 (1), 3162. doi:10.1038/s41467-020-16966-3

Wang, X., Li, L., Zhao, K., Lin, Q., Li, H., Xue, X., et al. (2020b). A novel LncRNA HITT forms a regulatory loop with HIF-1α to modulate angiogenesis and tumor growth. *Cell Death Differ.* 27 (4), 1431–1446. doi:10.1038/s41418-019-0449-8

Wang, Y., Zhang, X., Wang, Z., Hu, Q., Wu, J., Li, Y., et al. (2018a). LncRNA-p23154 promotes the invasion-metastasis potential of oral squamous cell carcinoma by regulating Glut1-mediated glycolysis. *Cancer Lett.* 434, 172–183. doi:10.1016/j.canlet.2018.07.016

Wang, Z., Yang, B., Zhang, M., Guo, W., Wu, Z., Wang, Y., et al. (2018b). lncRNA epigenetic landscape analysis identifies EPIC1 as an oncogenic lncRNA that interacts with MYC and promotes cell-cycle progression in cancer. *Cancer Cell* 33 (4), 706–720. doi:10.1016/j.ccell.2018.03.006

Warburg, O. (1956). On the origin of cancer cells. *Sci. (New York, N.Y.)* 123 (3191), 309–314. doi:10.1126/science.123.3191.309

Wei, S., Fan, Q., Yang, L., Zhang, X., Ma, Y., Zong, Z., et al. (2017). Promotion of glycolysis by HOTAIR through GLUT1 upregulation *via* mTOR signaling. *Oncol. Rep.* 38 (3), 1902–1908. doi:10.3892/or.2017.5840

Wuest, M., Hamann, I., Bouvet, V., Glubrecht, D., Marshall, A., Trayner, B., et al. (2018). Molecular imaging of GLUT1 and GLUT5 in breast cancer: A multitracer positron emission tomography imaging study in mice. *Mol. Pharmacol.* 93 (2), 79–89. doi:10.1124/mol.117.110007

Xu, W., Zhou, G., Wang, H., Liu, Y., Chen, B., Chen, W., et al. (2020). Circulating lncRNA SNHG11 as a novel biomarker for early diagnosis and prognosis of colorectal cancer. *Int. J. Cancer* 146 (10), 2901–2912. doi:10.1002/ijc.32747

Yang, B., Zhang, L., Cao, Y., Chen, S., Cao, J., Wu, D., et al. (2017a). Overexpression of lncRNA IGFBP4-1 reprograms energy metabolism to promote lung cancer progression. *Mol. Cancer* 16 (1), 154. doi:10.1186/s12943-017-0722-8

Yang, F., Zhang, H., Mei, Y., and Wu, M. (2014). Reciprocal regulation of HIF-1α and lincRNA-p21 modulates the Warburg effect. *Mol. Cell* 53 (1), 88–100. doi:10.1016/j.molcel.2013.11.004

Yang, J., Wang, C., Zhao, F., Luo, X., Qin, M., Arunachalam, E., et al. (2017b). Loss of FBP1 facilitates aggressive features of hepatocellular carcinoma cells through the Warburg effect. *Carcinogenesis* 38 (2), 134–143. doi:10.1093/carcin/bgw109

Yang, L., Chen, Y., Liu, N., Shi, Q., Han, X., Gan, W., et al. (2021). Low expression of TRAF3IP2-AS1 promotes progression of NONO-TFE3 translocation renal cell carcinoma by stimulating N^6-methyladenosine of PARP1 mRNA and downregulating PTEN. *J. Hematol. Oncol.* 14 (1), 46. doi:10.1186/s13045-021-01059-5

Yang, L., Hou, Y., Yuan, J., Tang, S., Zhang, H., Zhu, Q., et al. (2015). Twist promotes reprogramming of glucose metabolism in breast cancer cells through PI3K/AKT and p53 signaling pathways. *Oncotarget* 6 (28), 25755–25769. doi:10.18632/oncotarget.4697

Yang, T., Shu, X., Zhang, H., Sun, L., Yu, L., Liu, J., et al. (2020). Enolase 1 regulates stem cell-like properties in gastric cancer cells by stimulating glycolysis. *Cell Death Dis.* 11 (10), 870. doi:10.1038/s41419-020-03087-4

Yin, D., Hua, L., Wang, J., Liu, Y., and Li, X. (2020). Long non-coding RNA DUXAP8 facilitates cell viability, migration, and glycolysis in non-small-cell lung cancer *via* regulating HK2 and LDHA by inhibition of miR-409-3p. *Onco. Targets. Ther.* 13, 7111–7123. doi:10.2147/ott.S243542

Yizhak, K., Le Dévédec, S., Rogkoti, V., Baenke, F., de Boer, V., Frezza, C., et al. (2014). A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration. *Mol. Syst. Biol.* 10, 744. doi:10.15252/msb.20134993

Yu, J., Li, J., Chen, Y., Cao, W., Lu, Y., Yang, J., et al. (2017). Snail enhances glycolysis in the epithelial-mesenchymal transition process by targeting FBP1 in gastric cancer. *Cell. Physiol. biochem.* 43 (1), 31–38. doi:10.1159/000480314

Yu, Z., Zhao, H., Feng, X., Li, H., Qiu, C., Yi, X., et al. (2019). Long non-coding RNA FENDRR acts as a miR-423-5p sponge to suppress the treg-mediated immune escape of hepatocellular carcinoma cells. *Mol. Ther. Nucleic Acids* 17, 516–529. doi:10.1016/j.omtn.2019.05.027

Yuan, L., Xu, Z., Ruan, S., Mo, S., Qin, J., and Cheng, X. (2020). Long non-coding RNAs towards precision medicine in gastric cancer: Early diagnosis, treatment, and drug resistance. *Mol. Cancer* 19 (1), 96. doi:10.1186/s12943-020-01219-0

Zhang, E., He, X., Zhang, C., Su, J., Lu, X., Si, X., et al. (2018). A novel long noncoding RNA HOXC-AS3 mediates tumorigenesis of gastric cancer by binding to YBX1. *Genome Biol.* 19 (1), 154. doi:10.1186/s13059-018-1523-0

Zhang, J., Chen, G., Gao, Y., and Liang, H. (2020a). HOTAIR/miR-125 axis-mediated Hexokinase 2 expression promotes chemoresistance in human glioblastoma. *J. Cell. Mol. Med.* 24 (10), 5707–5717. doi:10.1111/jcmm.15233

Zhang, L., Li, C., and Su, X. (2020b). Emerging impact of the long noncoding RNA MIR22HG on proliferation and apoptosis in multiple human cancers. *J. Exp. Clin. Cancer Res.* 39 (1), 271. doi:10.1186/s13046-020-01784-8

Zhang, Q., Lou, Y., Zhang, J., Fu, Q., Wei, T., Sun, X., et al. (2017). Hypoxia-inducible factor-2α promotes tumor progression and has crosstalk with Wnt/β-catenin signaling in pancreatic cancer. *Mol. Cancer* 16 (1), 119. doi:10.1186/s12943-017-0689-5

Zhang, X., Zhou, Y., Mehta, K., Danila, D., Scolavino, S., Johnson, S., et al. (2003). A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. *J. Clin. Endocrinol. Metab.* 88 (11), 5119–5126. doi:10.1210/jc.2003-030222

Zhang, Y., Chen, M., Liu, M., Xu, Y., and Wu, G. (2021). Glycolysis-related genes serve as potential prognostic biomarkers in clear cell renal cell carcinoma. *Oxid. Med. Cell. Longev.* 2021, 6699808. doi:10.1155/2021/6699808

Zhang, Y., Liu, Q., and Liao, Q. (2020c). Long noncoding RNA: A dazzling dancer in tumor immune microenvironment. *J. Exp. Clin. Cancer Res.* 39 (1), 231. doi:10.1186/s13046-020-01727-3

Zhang, Y., Zhao, Y., Shen, J., Sun, X., Liu, Y., Liu, H., et al. (2019). Nanoenabled modulation of acidic tumor microenvironment reverses anergy of infiltrating T cells and potentiates anti-PD-1 therapy. *Nano Lett.* 19 (5), 2774–2783. doi:10.1021/acs.nanolett.8b04296

Zhao, H., Wu, S., Li, H., Duan, Q., Zhang, Z., Shen, Q., et al. (2019). ROS/KRAS/AMPK signaling contributes to gemcitabine-induced stem-like cell properties in pancreatic cancer. *Mol. Ther. Oncolytics* 14, 299–312. doi:10.1016/j.omto.2019.07.005

Zheng, F., Chen, J., Zhang, X., Wang, Z., Chen, J., Lin, X., et al. (2021). The HIF-1α antisense long non-coding RNA drives a positive feedback loop of HIF-1α mediated transactivation and glycolysis. *Nat. Commun.* 12 (1), 1341. doi:10.1038/s41467-021-21535-3

Zheng, X., Han, H., Liu, G., Ma, Y., Pan, R., Sang, L., et al. (2017). LncRNA wires up Hippo and Hedgehog signaling to reprogramme glucose metabolism. *EMBO J.* 36 (22), 3325–3335. doi:10.15252/embj.201797609

Zheng, Y., Li, L., Jia, Y., Zhang, B., Li, J., Zhu, Y., et al. (2019a). LINC01554-Mediated glucose metabolism reprogramming suppresses tumorigenicity in hepatocellular carcinoma *via* downregulating PKM2 expression and inhibiting akt/mTOR signaling pathway. *Theranostics* 9 (3), 796–810. doi:10.7150/thno.28992

Zheng, Y., Liu, P., Wang, N., Wang, S., Yang, B., Li, M., et al. (2019b). Betulinic acid suppresses breast cancer metastasis by targeting GRP78-mediated glycolysis and ER stress apoptotic pathway. *Oxid. Med. Cell. Longev.* 2019, 8781690. doi:10.1155/2019/8781690

Zhou, Y., Huang, Y., Hu, K., Zhang, Z., Yang, J., and Wang, Z. (2020). HIF1A activates the transcription of lncRNA RAET1K to modulate hypoxia-induced glycolysis in hepatocellular carcinoma cells *via* miR-100-5p. *Cell Death Dis.* 11 (3), 176. doi:10.1038/s41419-020-2366-7

Zhou, Y., Zhong, Y., Wang, Y., Zhang, X., Batista, D., Gejman, R., et al. (2007). Activation of p53 by MEG3 non-coding RNA. *J. Biol. Chem.* 282 (34), 24731–24742. doi:10.1074/jbc.M702029200

Zhuo, W., Liu, Y., Li, S., Guo, D., Sun, Q., Jin, J., et al. (2019). Long noncoding RNA GMAN, up-regulated in gastric cancer tissues, is associated with metastasis in patients and promotes translation of ephrin A1 by competitively binding GMAN-AS. *Gastroenterology* 156 (3), 676–691. e611. doi:10.1053/j.gastro.2018.10.054

Zou, Z., Ma, C., Medoro, L., Chen, L., Wang, B., Gupta, R., et al. (2016). LncRNA ANRIL is up-regulated in nasopharyngeal carcinoma and promotes the cancer progression via increasing proliferation, reprograming cell glucose metabolism and inducing side-population stem-like cancer cells. *Oncotarget* 7 (38), 61741–61754. doi:10.18632/oncotarget.11437

Glossary

ATP Adenosine triphosphate
LncRNAs Long non-coding RNAs
NcRNAs Non-coding RNAs
RBPs RNA binding proteins
TIME Tumor immune microenvironment
TAMs Tumor associated macrophages
Tregs Regulatory T cells
EGFR Epidermal growth factor receptor
RBM5 RNA-binding motif protein 5
NK cell Nature killer cell
PPP Pentose-phosphate pathway
DAPDH Diaminopimelate dehydrogenase
ENO1 Enolase 1
AKT AKT serine/threonine kinase
HCC Hepatocellular carcinoma
PDK1 Pyruvate dehydrogenase kinase 1
GLUT1 Glucose transporter1
HK3 Hexokinase3
EMT Epithelial-mesenchymal transition
PFKFB3 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
FBP1 Fructose-bisphosphatase 1
GSH Antioxidant glutathione
Zeb1 Zinc finger E-box binding homeobox 1
Snail Snail zinc finger protein
Twist Twist protein
SIRT3 Sirtuin 3
LDHA Lactate dehydrogenase A
G6PD Glucose-6-phosphate dehydrogenase
PFKP Phosphofructokinase
CSCs Cancer stem cells
HIF-2α Hypoxia-inducible factor-2α
ARRB1 β-arrestin1
MPC1 Mitochondrial pyruvate carrier 1
GEM Gemcitabine
NANOG Nanog homeobox
OCT4 Organic cation/carnitine transporter4
SOX2 SRY-box transcription factor 2
NEAT1 Nuclear-enrich abundant transcript 1
CD44 CD44 molecule
ALDH Aldehyde dehydrogenase
ESCs Embryonic stem cells
MDSCs Myeloid derived suppressor cells
Th cells T helper cells
CTLs $CD8^+$ cytotoxic T lymphocytes
DCs Dendritic cells
TME Tumor microenvironment
PEP Phosphoenolpyruvate
CRC Colon cancer
NADH Nicotinamide adenine dinucleotide
FGFR1 Fibroblast growth factor receptor type 1
UCA1 Urothelial cancer associated 1
STAT3 Signal transducer and activator of transcription 3
mTOR Mechanistic target of rapamycin kinase
AMPK Protein kinase AMP-activated catalytic subunit alpha 1
ER^+ Estrogen receptor-positive
MB Medulloblastoma
MET Mesenchymal-epithelial transition

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 30 August 2022
DOI 10.3389/fmolb.2022.940290

OPEN ACCESS

EDITED BY
Gian Gaetano Tartaglia,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Fabrizio Ferrè,
University of Bologna, Italy
Desi Shang,
Harbin Medical University, China

*CORRESPONDENCE
Dianying Zhang,
zdy8016@126.com
Xia Li,
lixia@hrbmu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 10 May 2022
ACCEPTED 05 August 2022
PUBLISHED 30 August 2022

CITATION
Zhang Y, Chen J, He S, Xiao Y, Liu A, Zhang D and Li X (2022), Systematic identification of aberrant non-coding RNAs and their mediated modules in rotator cuff tears.
Front. Mol. Biosci. 9:940290.
doi: 10.3389/fmolb.2022.940290

Systematic identification of aberrant non-coding RNAs and their mediated modules in rotator cuff tears

Yichong Zhang[1†], Jianhai Chen[1†], Shengyuan He[2†], Yun Xiao[2], Aiyu Liu[3], Dianying Zhang[1*] and Xia Li[2*]

[1]Department of Orthopedics and Trauma, Key Laboratory of Trauma and Neural Regeneration (Ministry of Education/Peking University), Peking University People's Hospital, Beijing, China, [2]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China, [3]Central Laboratory, Peking University People's Hospital, Beijing, China

Background: Rotator cuff tears (RCT) is the most common cause of shoulder dysfunction, however, its molecular mechanisms remain unclear. Non-coding RNAs(ncRNAs), such as long ncRNA (lncRNA), microRNA (miRNA) and circular RNA (circRNA), are involved in a variety of diseases, but little is known about their roles in RCT. Therefore, the purpose of this study is to identify dysregulated ncRNAs and understand how they influence RCT.

Methods: We performed RNA sequencing and miRNA sequencing on five pairs of torn supraspinatus muscles and matched unharmed subscapularis muscles to identify RNAs dysregulated in RCT patients. To better comprehend the fundamental biological processes, we carried out enrichment analysis of these dysregulated mRNAs or the co-expressed genes of dysregulated ncRNAs. According to the competing endogenous RNA (ceRNA) theory, we finally established ceRNA networks to explore the relationship among dysregulated RNAs in RCT.

Results: A total of 151 mRNAs, 38 miRNAs, 20 lncRNAs and 90 circRNAs were differentially expressed between torn supraspinatus muscles and matched unharmed subscapularis muscles, respectively. We found that these dysregulated mRNAs, the target mRNAs of these dysregulated miRNAs or the co-expressed mRNAs of these dysregulated ncRNAs were enriched in muscle structure development, actin-mediated cell contraction and actin binding. Then we constructed and analyzed the ceRNA network and found that the largest module in the ceRNA network was associated with vasculature development. Based on the topological properties of the largest module, we identified several important ncRNAs including *hsa_circ_0000722*, hsa-miR-129-5p and hsa-miR-30c-5p, whose interacting mRNAs related to muscle diseases, fat and inflammation.

Conclusion: This study presented a systematic dissection of the expression profile of mRNAs and ncRNAs in RCT patients and revealed some important ncRNAs which may contribute to the development of RCT. Such results could provide new insights for further research on RCT.

KEYWORDS
rotator cuff tears, miRNA, lncRNA, circRNA, ceRNA network

1 Introduction

The rotator cuff consists of four muscle-tendon units: supraspinatus, infraspinatus, teres minor and subscapularis, which contribute to shoulder movement. Rotator cuff tear (RCT) is the leading cause of pain and functional disability of the shoulder and is present in about 30% of individuals in their 60 s and higher in individuals over 80 years old (Dang and Davies, 2018). Although some patients can be treated successfully with surgical repair (Schemitsch et al., 2019), not all patients' outcomes of rotator cuff repair are satisfactory, which is partly because of the poor understanding of the molecular mechanism of RCT (Connor et al., 2019). Therefore, potential factors that may contribute to RCT should be identified.

Non-coding RNAs (ncRNAs) account for approximately 97% of the human genome, including long ncRNAs (lncRNA), microRNAs (miRNA) and circular RNAs (circRNA) (Anastasiadou et al., 2018). LncRNAs are a class of ncRNAs longer than 200 bp with low coding potential, while miRNAs are a class of small ncRNAs with ~22 nucleotides. They were reported to influence various stages of tendinopathy and could be implicated in skeletal muscle differentiation (Ge et al., 2020a; Plachel et al., 2020). CircRNAs are a novel class of endogenous, non-coding RNAs with closed-loop structures, which were generated during RNA alternative splicing. CircRNAs can express in striated muscle tissues including skeletal and cardiac muscles as well (Greco et al., 2018). LncRNAs and circRNAs can act as molecular sponges of miRNAs to regulate the expression of mRNAs, which is known as "competing endogenous RNA (ceRNA)" hypothesis. Previous studies have shown that lncRNAs and circRNAs could serve as ceRNAs and play roles in rotator cuff tendinopathy (Ge et al., 2020a; Ge et al., 2020b). However, the research of lncRNA- or the circRNA-mediated ceRNA networks is not comprehensive.

In this study, we screened and identified differentially expressed mRNAs, lncRNAs, circRNAs and miRNAs between samples from torn supraspinatus and unharmed subscapularis. Based on the results of differential expression analysis and miRNA targeting information, we constructed lncRNA/circRNA-associated dysregulated ceRNA networks in RCT. Finally, we found the largest module in the ceRNA network and identified several important ncRNAs in this module, which may have roles in RCT.

2 Results

2.1 Identification of differentially expressed mRNAs

RNA sequencing was performed to obtain the mRNA expression profile of torn supraspinatus muscles (group T) and matched unharmed subscapularis muscles (group P). Principal component analysis of the mRNA expression data could distinguish samples of supraspinatus muscles and unharmed subscapularis muscles, indicating that the difference between supraspinatus muscles and unharmed subscapularis muscles (Figures 1A,B). Compared with the expression of mRNAs in unharmed subscapularis muscles, a total of 151 differentially expressed mRNAs in supraspinatus muscles (absolute fold change >1.5 and $p < 0.05$) were identified, which comprised 76 up-regulated mRNAs and 75 down-regulated mRNAs (Figures 1C,D). Some of these differentially expressed mRNAs were mentioned in previous reports of rotator cuff tears. For example, *MYL6B* has been proved that it can differentially expressed in patients with rotator cuff tears (Frich et al., 2021), while *EGR1* is conducive to the repair of rotator cuff tears (Tao et al., 2015). We further applied enrichment analysis on the differentially expressed mRNAs. The results showed that differentially expressed mRNAs enriched in muscle structure development, actin filament-based process and blood vessel development (Figures 1E,F). Previous researches also indicated that rotator cuff tears could be associated with muscle actin and angiogenesis (Fuchs et al., 2008; Noh et al., 2018). After adjusting the P value using the false discovery rate (FDR) method, only one differentially expressed mRNA, *SIM2*, was identified (FDR corrected $p < 0.05$).

2.2 Identification of differentially expressed ncRNAs

Studies have characterized the biological roles of noncoding RNAs in many diseases (Falcone et al., 2014; Li et al., 2017; Kristensen et al., 2019; Li et al., 2021), therefore, we also performed miRNA sequencing on torn supraspinatus muscles and matched unharmed subscapularis muscles. Together with the previous data of RNA sequencing, we generated the expression profiles of miRNAs, lncRNAs and circRNAs.

There were 38 differentially expressed miRNAs (17 up-regulated and 21 down-regulated, Figure 2A), 20 differentially expressed lncRNAs (8 up-regulated and 12 down-regulated, Figure 2D) and 90 differentially expressed circRNAs (39 up-regulated and 51 down-regulated, Figure 3A) were identified in supraspinatus muscles and unharmed subscapularis muscles. After adjusting the P value using the FDR method, only 1 differentially expressed miRNA (hsa-miR-618), 1 differentially expressed lncRNA (*LINC01854*) and 15 differentially expressed circRNAs (*chr2:152355811–152355904:-*, *chr2:179511211–179511286:-*, *chr2:179514280–179514358:-*, *chr2:179514280–179514621:-*, *chr2:179517184–179517463:-*, *chr2:179517574–179517658:-*, *chr2:179519171–179535022:-*, *chr2:179523431–179535022:-*, *chr2:1795276*

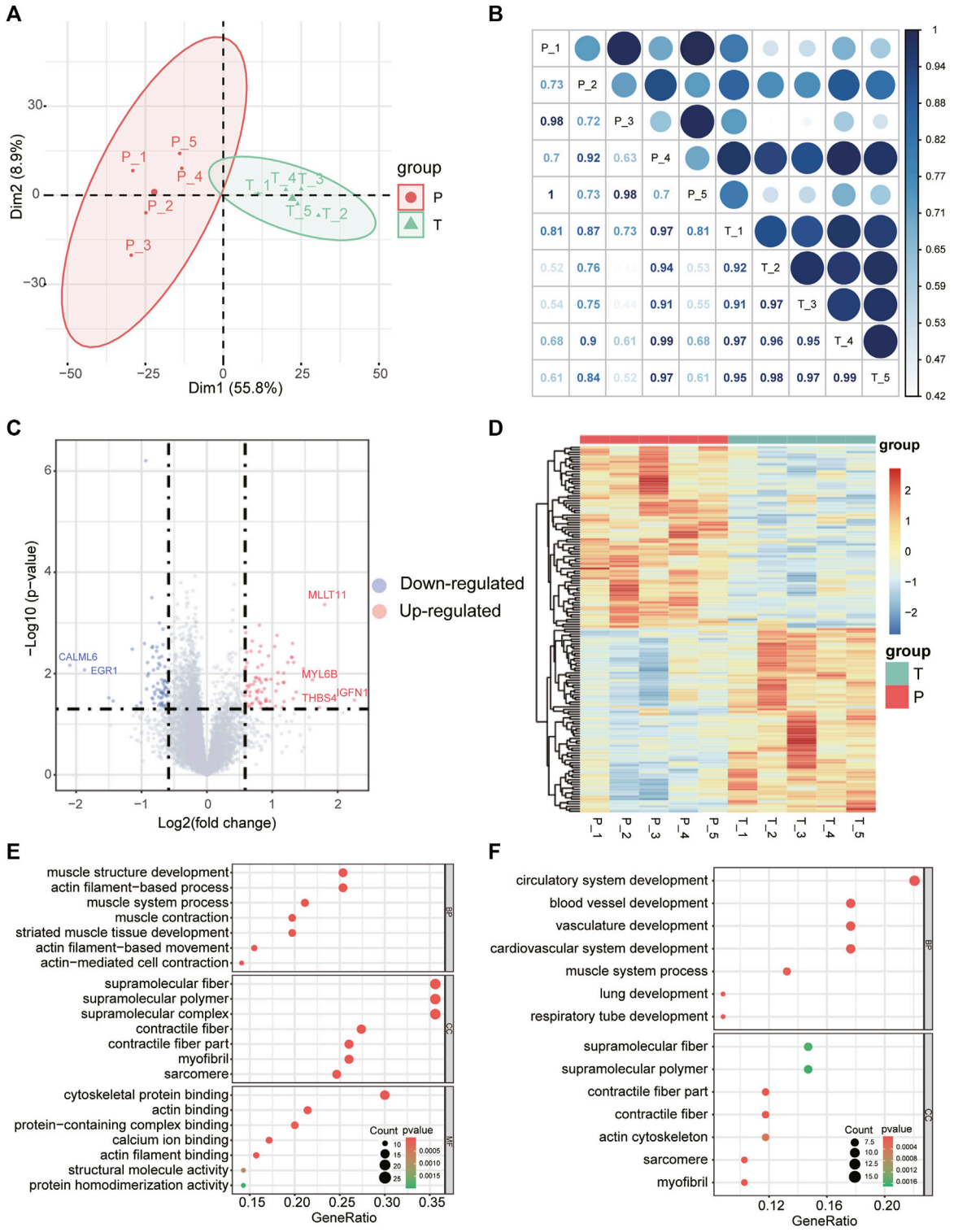


FIGURE 1
Differentially expressed mRNAs and GO enrichment analysis. **(A)** The first two principal components show distinction between torn supraspinatus muscles and matched unharmed subscapularis muscles based on the mRNA expression profile. **(B)** Heatmap of the correlation matrix of mRNA expression profile of 10 samples, the size of the node represents the correlation coefficient. **(C)** Volcano plot of differentially expressed mRNAs in RCT. Red points denote up-regulated mRNAs, and blue points denote down-regulated mRNAs. The most up- or down-regulated mRNAs are labeled. **(D)** Heatmap depicting expression levels of the differentially expressed mRNAs in RCT. **(E,F)** GO enrichment analysis for the up-regulated**(E)** and down-regulated differentially expressed mRNAs **(F)**, respectively.

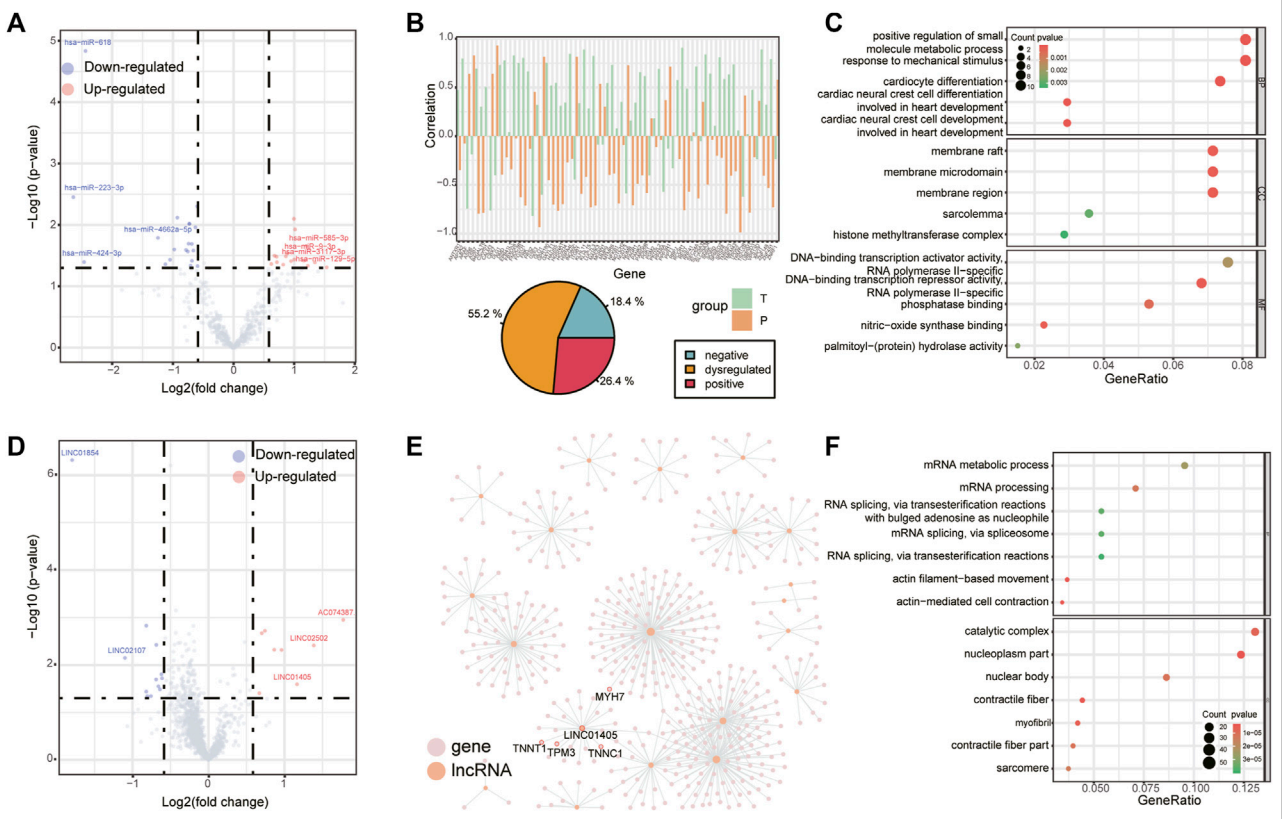


FIGURE 2
Differentially expressed miRNAs and lncRNAs. **(A)** Volcano plot of differentially expressed miRNAs in RCT. Red points denote up-regulated miRNAs, and blue points denote down-regulated miRNAs. The most up- or down-regulated miRNAs are labeled. **(B)** Changes in the correlation between hsa-miR-9-3p and its target genes, bar chart (top panel) shows the correlation of those miRNA-mRNA pairs which were positively (negatively) correlated to hsa-miR-9-3p in supraspinatus muscles but negatively (positively) correlated to hsa-miR-9-3p in unharmed subscapularis muscles (we define this correlation pattern as "dysregulated"), pie chart (bottom panel) shows the proportion of miRNA-mRNA pairs with both negative correlation, both positive correlation or "dysregulated" in supraspinatus muscles and unharmed subscapularis muscles. **(C)** GO enrichment analysis for all target mRNAs of hsa-miR-9-3p. **(D)** Volcano plot of differentially expressed lncRNAs. **(E)** Co-expression network of differentially expressed lncRNAs. Pink points denote mRNAs and orange points denote lncRNAs. **(F)** GO enrichment analysis for all co-expressed mRNAs of differentially expressed lncRNAs.

92–179535022:-, *chr2:179528353–179528437:-*, *chr2:179528545–179528629:-*, *chr2:179534318–179535022:-*, *hsa_circ_0141770*, *chr2:179542347–179544143:-*, *hsa_circ_0086735*) were identified (FDR corrected $p < 0.05$).

One of the differentially expressed miRNA, hsa-miR-9-3p (Figure 2A), may be involved in inflammation (Chakraborty et al., 2015), which is a common symptom in rotator cuff tears. We compared the correlation between the expression of hsa-miR-9-3p and its target mRNAs in supraspinatus muscles and unharmed subscapularis muscles. More than half of the target mRNAs of hsa-miR-9-3p were positively (negatively) correlated to hsa-miR-9-3p in supraspinatus muscles but negatively (positively) correlated to hsa-miR-9-3p in unharmed subscapularis muscles (Figure 2B). Such opposite correlation results in supraspinatus muscles and unharmed subscapularis muscles support the opinion that hsa-miR-9-3p was dysregulated in rotator cuff tears. We also found that target mRNAs of hsa-miR-9-3p were enriched in positive regulation of small molecule metabolic process, response to mechanical stimulus, membrane raft and sarcolemma (Figure 2C).

Next, we performed co-expression analysis to identify mRNAs correlated to differentially expressed lncRNAs (Figure 2E). Previous study had reported that one of the differentially expressed lncRNAs, *LINC01405*, was associated with muscle-related disease (Schofer et al., 2008). Some co-expressed mRNAs of *LINC01405* (*TNNT1* and *MYH7*, Figure 2E) have also been reported in a study of rotator cuff tears (Frich et al., 2021). Overall, co-expressed mRNAs of all differentially expressed lncRNAs were enriched in actin-mediated cell contraction and contractile fiber (Figure 2F).

Meanwhile, we also analyzed differentially expressed circRNAs in supraspinatus muscles and unharmed subscapularis muscles. In addition to applying edgeR, we employed the GLMM model, a newly developed approach

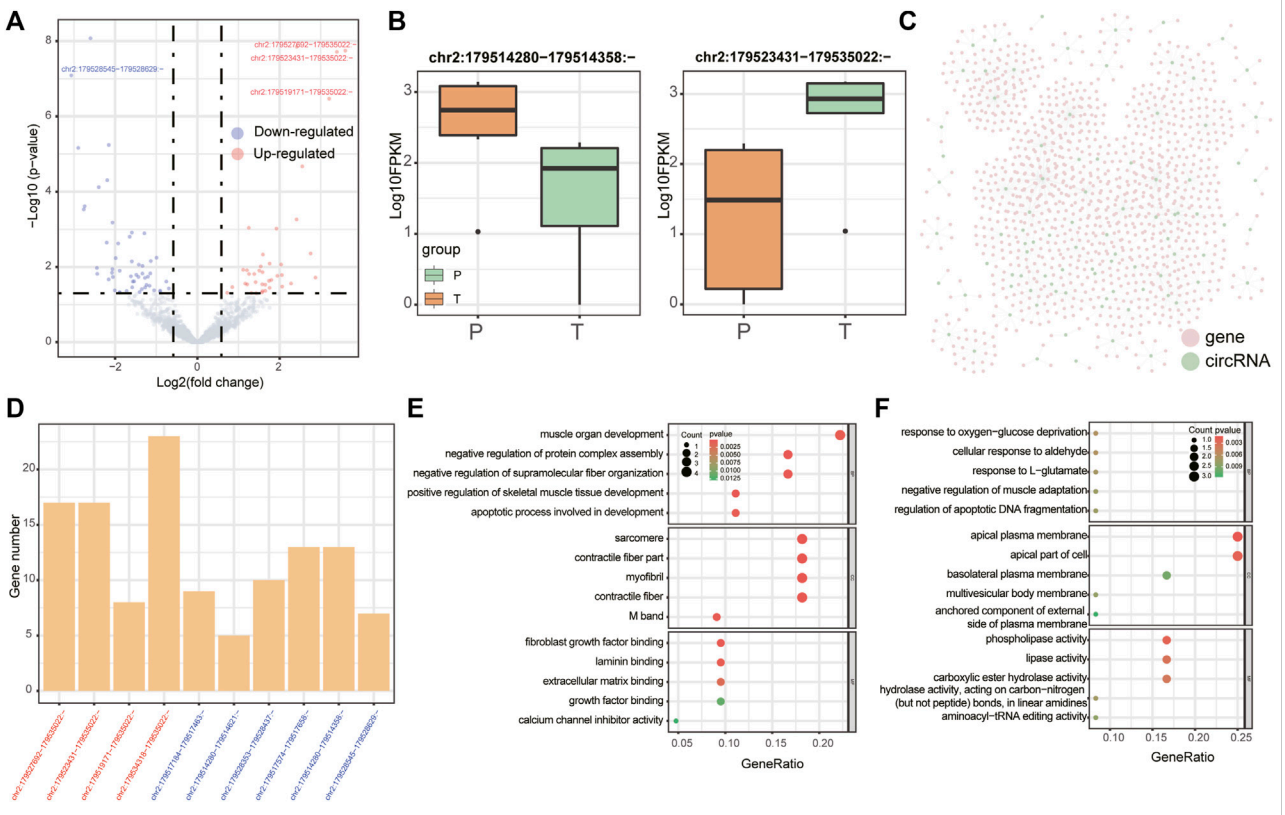


FIGURE 3
Differentially expressed circRNAs. **(A)** Volcano plot of differentially expressed circRNAs in RCT. Red points denote up-regulated circRNAs, and blue points denote down-regulated circRNAs. The most up- or down-regulated circRNAs are labeled. **(B)** Boxplots denote expression levels of 2 differentially expressed circRNAs which were derived from *TTN* (*chr2:179514280–179514358:-* (left panel) and *chr2:179523431–179535022:-* (right panel)). **(C)** Co-expression network of differentially expressed circRNAs. Pink points denote mRNAs and green points denote circRNAs. **(D)** Barplots denote the numbers of mRNAs regulated by circRNAs that were produced by *TTN* in the co-expression network. CircRNAs were ordered according to the fold changes in group T vs. group P (all circRNAs were identified both by edgeR and by GLMM). **(E,F)** GO enrichment analysis for all co-expressed mRNAs of *chr2:179534318–179535022:-* and *chr2:179514280–179514358:-*.

specifically for identifying differentially expressed circRNAs (Buratin et al., 2022). A total of 28 differentially expressed circRNAs were identified by GLMM model ($p < 0.05$ FDR corrected), and 22 of them overlapped with the results identified by edgeR ($p = 0.007$, hypergeometric test). After adjusting the p-value with FDR <0.05, there was significant overlap of 10 circRNAs between the results of edgeR and GLMM ($p = 2.92 \times 10^{-16}$, hypergeometric test).

All 10 differentially expressed circRNAs identified both by edgeR and by GLMM were derived from *TTN*(Titin) (Figures 3B,C). An earlier study has found that the molecular weight of *TTN* was changed in the injured rotator cuff (Sato et al., 2014). Although these *TTN*-derived circRNAs were derived from the same gene, we found that different functions were related to them, which also co-expressed with a various number of mRNAs (Figure 3D). For example, co-expressed mRNAs of *chr2:179534318–179535022:-* were enriched in muscle organ development and skeletal muscle tissue development (Figure 3E), while co-expressed mRNAs of *chr2:179514280–179514358:-* were associated with negative regulation of muscle adaptation, response to oxygen-glucose deprivation (Figure 3F).

2.3 Establishing dysregulated ceRNA networks of lncRNAs and circRNAs

LncRNAs can act as molecular sponges of miRNAs to regulate the expression of mRNAs and this mechanism is known as "competing endogenous RNA (ceRNA)" hypothesis. According to the ceRNA hypothesis, a large number of studies have explored the lncRNAs-miRNAs-mRNAs interactions in various diseases which include rotator cuff tears (Ge et al., 2020a; Plachel et al., 2020; Chen et al., 2021). Additional to lncRNAs, recent studies also showed that circRNAs could serve as ceRNAs, however, there has been few studies

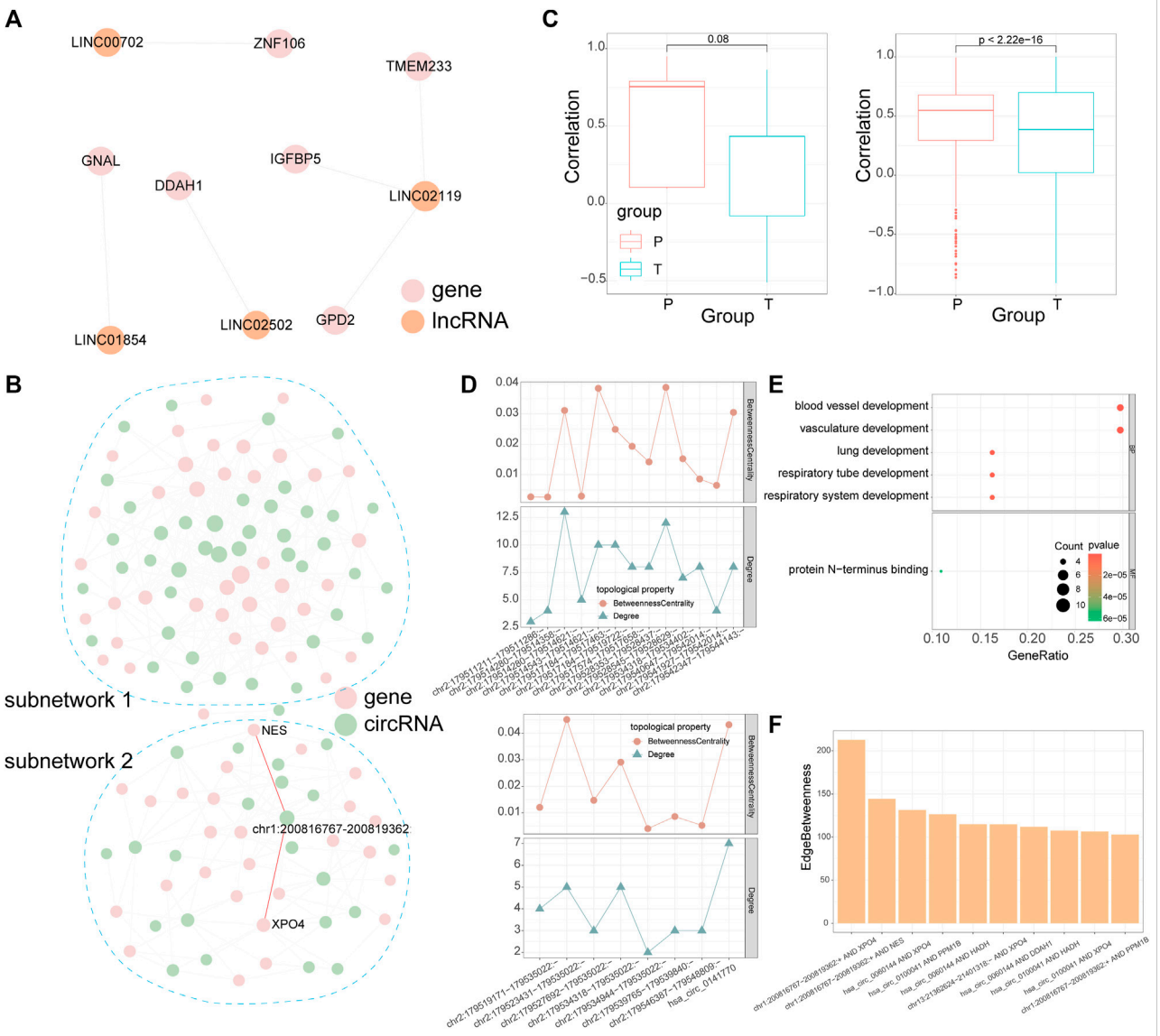


FIGURE 4
CeRNA networks of lncRNA and circRNA. **(A)** The dysregulated ceRNA network of lncRNAs. The size of the node represents the node's degree. **(B)** The dysregulated ceRNA network of circRNAs. CircRNA-mRNA pairs with the highest edge betweenness are highlighted in red. **(C)** Pearson correlation coefficient between lncRNA/circRNA-mRNA pairs in the dysregulated ceRNA networks of lncRNAs (left panel) or circRNAs (right panel). **(D)** Line chart denotes the degree and betweenness centrality of mRNAs regulated by circRNAs which were produced by *TTN* in subnetwork 1 (top panel) and subnetwork 2 (bottom panel). **(E)** GO enrichment analysis for mRNAs in subnetwork 1. **(F)** Edge betweenness centrality of circRNA-mRNA pairs in subnetwork 2, only the top 10 are shown in the diagram. In the dysregulated ceRNA network, pink points denote mRNAs, orange points denote lncRNAs and green points denote circRNAs.

characterize circRNAs-miRNAs-mRNAs interactions in rotator cuff tears. We obtained target mRNAs and lncRNAs/circRNAs from the starbase, mirTarbase, and lncbase; then we identified lncRNAs/circRNAs-mRNAs pairs which significantly share common miRNAs by the hypergeometric test based on ceRNA hypothesis (see Section 5, Figures 4A,B). As a result, the dysregulated lncRNA-mRNA ceRNA network included five lncRNAs and 8 mRNAs, while the dysregulated circRNA-mRNA ceRNA network included 65 circRNAs and 65 mRNAs. The distribution of correlation coefficients between the expression of lncRNAs/circRNAs and mRNAs in the dysregulated network was different in torn supraspinatus muscles and matched unharmed subscapularis muscles (Figure 4C), which also indicate the dysregulation of ceRNA interaction in rotator cuff tears.

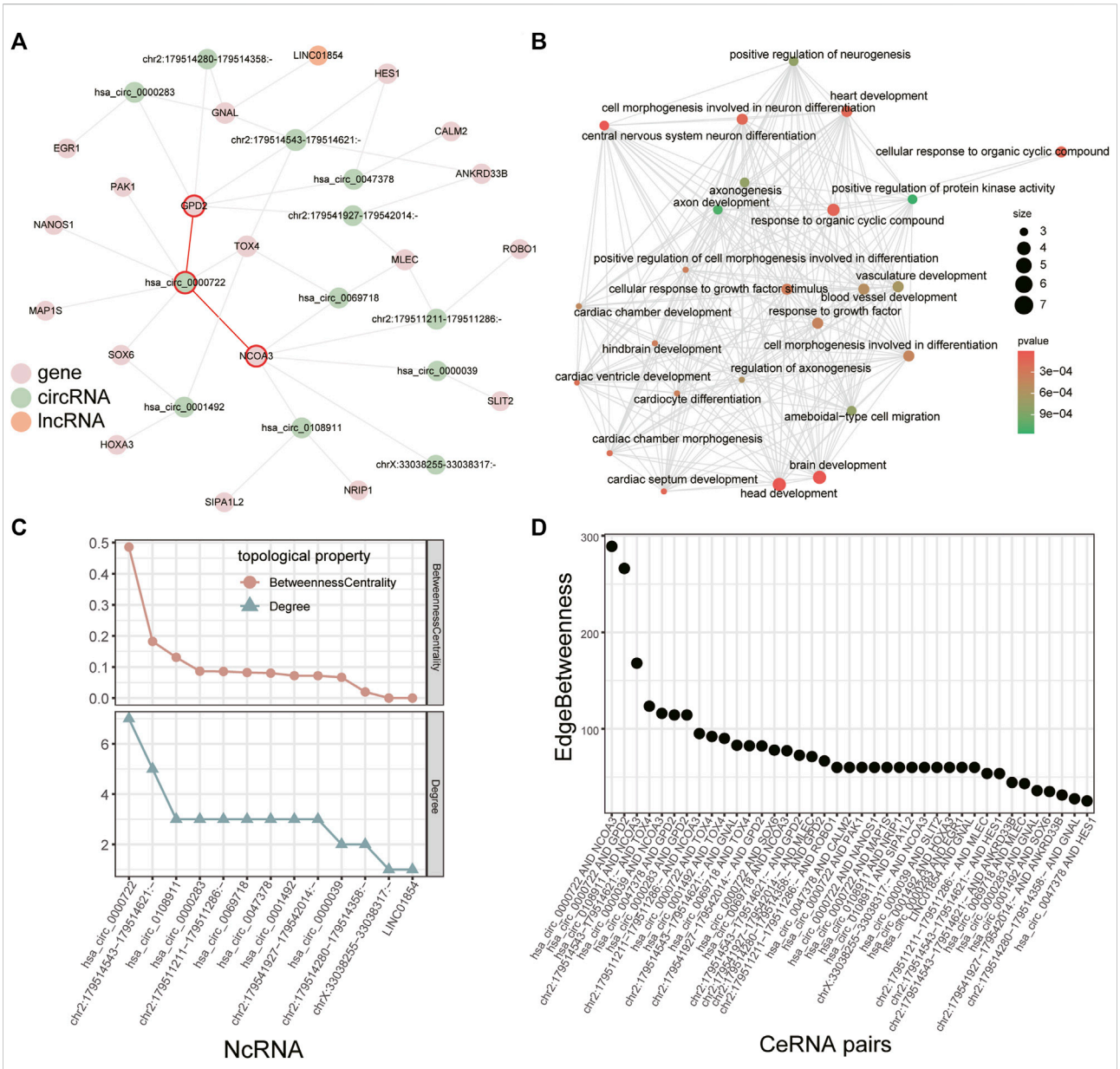


FIGURE 5
Module in the combined ceRNA network. **(A)** The largest module in the combined ceRNA network. The circRNA-mRNA pairs of *hsa_circ_0000722* are highlighted in red. **(B)** GO enrichment analysis for the mRNAs in the module. **(C)** Distribution of degree and betweenness centrality of ncRNAs in the module. **(D)** Edge betweenness centrality of lncRNA/circRNA-mRNA pairs in the module.

In the dysregulated ceRNA network of circRNAs, we detected two subnetworks (Figure 4B). These two subnetworks had a various number of *TTN*-derived circRNAs, whose degree and betweenness centrality varies in each subnetwork as well (Figure 4D). We further performed enrichment analysis on mRNAs in each subnetwork and calculated the edge betweenness centrality of each circRNA-mRNA pair in these two subnetworks. MRNAs in the subnetwork 1 were related to vasculature development and lung development by enrichment analysis (Figure 4E). Although there were no GO terms enriched in mRNAs in the subnetwork 2, the 2 circRNA-mRNA pairs with the highest edge betweenness centrality in subnetwork 2 were both mediated by a *CAMSAP2*-derived circRNA (*chr1: 200816767–200819362:+*, Figures 4B,F). Some studies indicated that *CAMSAP2* was associated with microtubule (Li et al., 2020), which could take part in muscle differentiation (Gache et al., 2017).

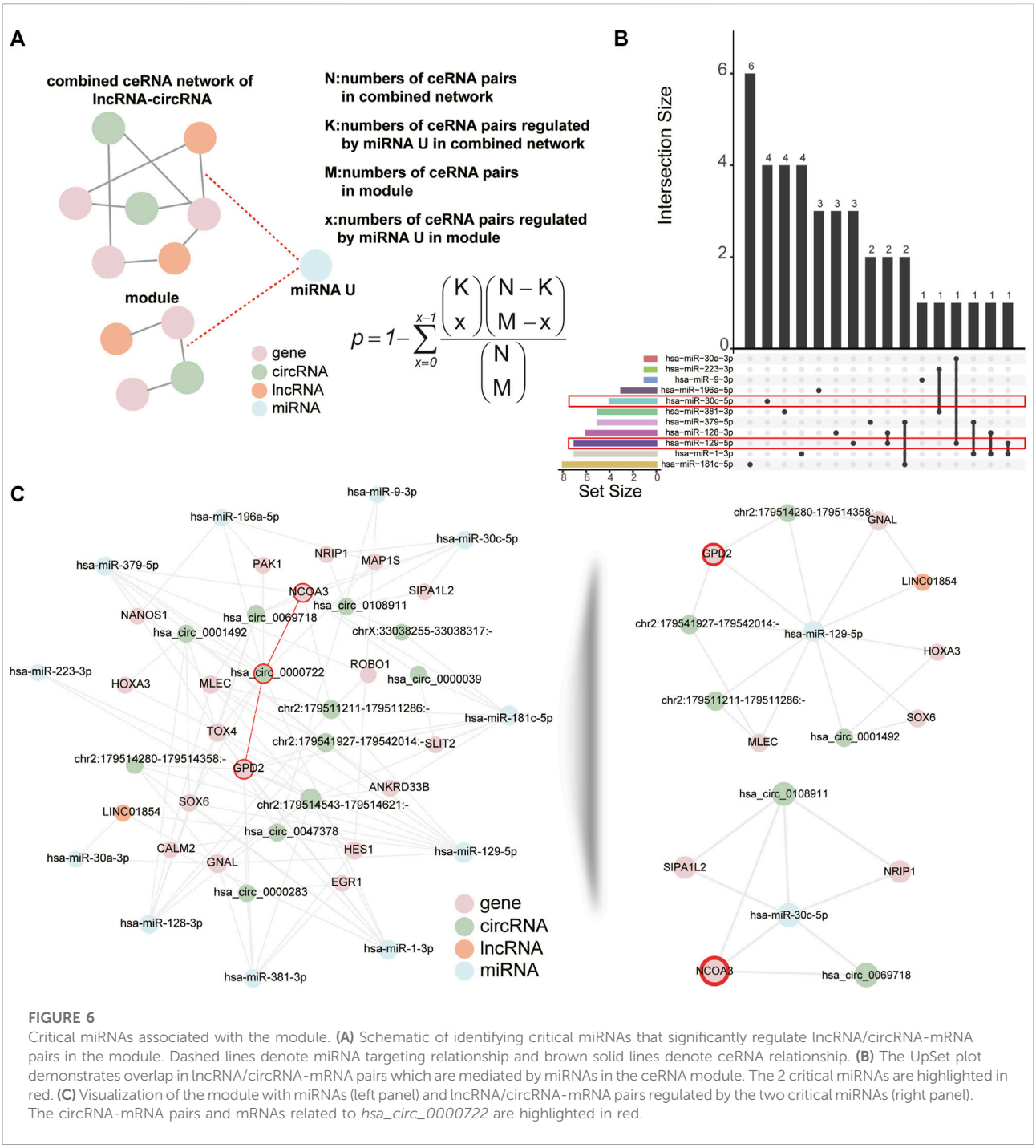


FIGURE 6
Critical miRNAs associated with the module. **(A)** Schematic of identifying critical miRNAs that significantly regulate lncRNA/circRNA-mRNA pairs in the module. Dashed lines denote miRNA targeting relationship and brown solid lines denote ceRNA relationship. **(B)** The UpSet plot demonstrates overlap in lncRNA/circRNA-mRNA pairs which are mediated by miRNAs in the ceRNA module. The 2 critical miRNAs are highlighted in red. **(C)** Visualization of the module with miRNAs (left panel) and lncRNA/circRNA-mRNA pairs regulated by the two critical miRNAs (right panel). The circRNA-mRNA pairs and mRNAs related to *hsa_circ_0000722* are highlighted in red.

2.4 Identifying the key ncRNA mediating the ceRNA module

Module analysis of the dysregulated ceRNA network can provide more information about rotator cuff tears. As lncRNAs and circRNAs both can influence mRNAs by competing for shared miRNAs, we combined the dysregulated lncRNA-mRNA ceRNA network with the dysregulated circRNA-mRNA ceRNA network according to mRNAs in both networks (see Section 5). The combined network included five lncRNAs, 65 circRNAs and 67 mRNAs, then we used "clusterMaker2" (Morris et al., 2011) to identify modules in the combined network. The largest module had 12 circRNAs, 1 lncRNA and 18 mRNAs (Figure 5A). Further analysis showed that mRNAs in the largest module were associated with vasculature development and heart development (Figure 5B).

We identified key ncRNAs by calculating the degree and betweenness centrality of each node in the largest module. *Hsa_circ_0000722* had the largest degree and betweenness centrality (Figure 5C) and recent study has revealed that *hsa_circ_0000722* has many potential binding sites for genes of MBNL family, a gene family which can cause muscle disease (Czubak et al., 2019). Based on the edge betweenness centrality of each lncRNA/circRNA-mRNA pair in this module (Figure 5D), we found that the lncRNA/circRNA-mRNA pairs with the highest edge betweenness centrality were mediated by *hsa_circ_0000722* as well (*hsa_circ_0000722-GPD2* and *hsa_circ_0000722-NCOA3*, Figures 5A,D). Studies have shown that *GPD2* and *NCOA3* relate to the fat (Mollah and Ishikawa, 2010; Han et al., 2017), which could infiltrate in injured rotator cuff muscles (Khanna et al., 2019).

2.5 Identifying the key miRNAs regulating the ceRNA module

Given that each ceRNA shares multiple miRNAs with numerous other ceRNAs, we then attempted to identify the key miRNAs in the largest ceRNA module (Figure 6A). A total of 11 differentially expressed miRNAs were associated with the ceRNA module and the lncRNA/circRNA-mRNA pairs they mediated were quite different (Figure 6B). Based on the hypergeometric test, we could identify which miRNAs mediated ceRNA pairs more significantly in the largest module than in the combined network. Finally we detected 2 important miRNAs: hsa-miR-129-5p and hsa-miR-30c-5p (p-value < 0.05, Figures 6A,C, see Section 5). The *hsa_circ_0000722-GPD2* and *hsa_circ_0000722-NCOA3* pairs that we identified above were mediated by hsa-miR-129-5p (*hsa_circ_0000722*-hsa-miR-129-5p-*GPD2*) and hsa-miR-30c-5p (*hsa_circ_0000722*-hsa-miR-30c-5p-*NCOA3*), respectively (Figure 6C). Some target mRNAs of hsa-miR-129-5p or hsa-miR-30c-5p, such as *GNAL*, *SOX6* and *NRIP1*, have been demonstrated to be associated with muscle diseases (Connor et al., 1995; Hagiwara et al., 2000; Kumar et al., 2014; De Marinis et al., 2017). *Hsa_circ_0108911* is derived from ATP9B and is also a target circRNA of hsa-miR-30c-5p. Although there were no studies about *hsa_circ_0108911*, Sun et al. (2020) have shown that another ATP9B-derived circRNA, circAtp9b, can contribute to the inflammation.

3 Discussion

In this study, we comprehensively dissected the dysregulated transcriptome of RCT, including mRNAs, miRNAs, lncRNAs, and circRNAs. According to the ceRNA hypothesis, we constructed the dysregulated ceRNA network and identified several important ncRNAs in the largest module of the ceRNA network, including *hsa_circ_0000722*, hsa-miR-129-5p and hsa-miR-30c-5p.

We have obtained transcripts that were dysregulated in the RCT by identifying differentially expressed mRNAs/ncRNAs between torn supraspinatus muscles and normal supraspinatus muscles. However, the transcripts that we identified may also show the inherent differences between supraspinatus muscles and subscapularis muscles. Recent studies have investigated the properties of different muscles by identifying differentially expressed genes between different muscles (Alto et al., 2021; Smith et al., 2022); Terry et al. revealed that an average of 13% of transcripts were differentially expressed between any two skeletal muscles (Terry et al., 2018). Therefore, further transcriptome sequencing of normal supraspinatus muscles and normal subscapularis muscles is required to confirm that the dysregulation of these transcripts is associated with RCT rather than being caused by tissue differences.

It is noted that all of differentially expressed circRNAs identified both by edgeR and by GLMM were derived from *TTN*, a gene which encodes proteins of striated muscle and contributes to muscle contraction (Hessel et al., 2017). Such a number of *TTN*-derived circRNAs may be due to the giant size and complexity of *TTN* (Chauveau et al., 2014). Because of the important role of *TTN* itself in muscle, the *TTN*-derived circRNAs are also more likely to be involved in muscle-related functions, which means further studies are needed to demonstrate the effect of these *TTN*-derived circRNAs in rotator cuff tears.

Functional enrichment analysis of the dysregulated transcriptome revealed several functions which may be associated with rotator cuff tears, such as muscle contraction, vasculature development, lung development and heart development. Studies have shown that muscle contraction of the rotator cuff plays an important role in moving and stabilizing the glenohumeral joint (Edouard et al., 2011). Since increased vascularization around the tendon-bone interface is essential for promoting rotator cuff tendon-bone healing (Randelli et al., 2016; Noh et al., 2018), angiogenesis is a key process after repairing of rotator cuff tears. In addition to these studies, some studies also explored the potential association of rotator cuff tears with lung diseases and cardiovascular system (Bachasson et al., 2015; Nam et al., 2015; Applegate et al., 2017). Taken together, these results could serve as a reference for future studies of rotator cuff tears.

We identified an important ncRNA (*hsa_circ_0000722*) by analyzing the dysregulated ceRNA networks of rotator cuff tears. Several mRNAs or ncRNAs which interact with *hsa_circ_0000722* in the largest ceRNA module, including *GPD2*, *NCOA3* and hsa-miR-30c-5p, are associated with fat (Mollah and Ishikawa, 2010; De Marinis et al., 2017; Han et al., 2017; Yaman et al., 2021). Hsa-miR-129-5p and hsa-miR-30c-5p were the other 2 key ncRNA we identified, which all interact with *hsa_circ_0000722* in the ceRNA network

(*hsa_circ_0000722*-hsa-miR-129-5p-*GPD2*/*hsa_circ_0000722*-hsa-miR-30c-5p-*NCOA3*). These 2 miRNAs have been found to be dysregulated in inflammation-related diseases, implying that they may involve in the inflammatory response in RCT (Duecker et al., 2022; Ling et al., 2022). Studies showed that inflammation in RCT contributes to fatty infiltration (Nelson et al., 2021), which is associated with poor surgical outcomes and postoperative failure of rotator cuff repair (Shen et al., 2008; Wieser et al., 2019), but the roles of *hsa_circ_0000722* and their ceRNA interactions (*hsa_circ_0000722*-hsa-miR-129-5p-*GPD2*/*hsa_circ_0000722*-hsa-miR-30c-5p-*NCOA3*) in rotator cuff tears remain to be elucidated.

However, this study is based on a small sample of five pairs of torn supraspinatus muscle samples and matched unharmed subscapularis muscle samples, which may have limited generalizability. Therefore, key ncRNAs that found in this work need further experimental verification to confirm their potential application in rotator cuff tears.

4 Conclusion

We identified some ceRNA modules and important ncRNA in RCT, which may play roles in the development of RCT or rotator cuff repair. Our findings offer a new perspective on the transcriptome analysis of RCT, while pre-clinical studies followed by clinical trials are still needed to validate our findings in the future.

5 Materials and methods

5.1 Patient information

5 female patients with unilateral shoulder pain were enrolled in this study, aging 50–60. All five patients didn't have diabetes, history of smoking, previous shoulder surgery or steroid injection. They were diagnosed with unilateral supraspinatus tears with intact subscapularis by MRI, which was then confirmed by the arthroscope. Informed consent was obtained from all subjects or their legal guardians. All methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by the Ethics Committee of Peking University People's Hospital.

5.2 Arthroscopic surgery and sample collection

All procedures were performed under general anesthesia. The patients were positioned in the beach-chair position and normal portals were established. The torn supraspinatus and intact subscapularis were confirmed by irrigation and debridement in intraarticular and subacromial space, then a small piece of muscle belly from each muscle was carefully collected by a grasper. After that, the supraspinatus tear was repaired using suture anchors. A total of 10 samples from five patients were stored in liquid nitrogen for further analysis.

5.3 Total RNA-seq library construction

Approximately 1–2 μg total RNA from each sample was used for library construction. The integrity of the total RNA was checked by agarose gel electrophoresis and the RNA concentration was quantified with a Nanodrop ND 1000 spectrophotometer (NanoDrop Technologies, United States). mRNA was enriched by NEBNext ® Poly(A) mRNA Magnetic Isolation Module (NEB, United States) and ribosomal RNA (rRNA) was depleted using Ribo-Zero Magnetic Gold Kit (Human/Mouse/Rat) (Epicentre, United States). cDNA libraries were prepared using a KAPA Stranded RNA-Seq Library Prep Kit (Illumina) according to the manufacturer's instructions. The constructed libraries were qualified by Agilent 2100 Bioanalyzer system (Agilent Technologies, CA, United States) and quantified by qPCR.

5.4 Total RNA sequencing

The DNA fragments in libraries were denatured with 0.1 M NaOH to generate single-stranded DNA molecules. Then the libraries were diluted to 8 nM and sequenced for 150 cycles on NovaSeq 6000 (Illumina Inc.) using NovaSeq 6000 S4 Reagent Kit (300 cycles) (Illumina Inc.).

5.5 RNA-seq data analysis

Quality control of RNA-seq data was carried out with FastQC (v.0.11.7) (Andrews et al., 2020). Reads were trimmed using cutadapt (v.1.17) (Martin, 2011) and aligned to human reference genome (GRCh37) using Hisat2 (v.2.1.0) (Kim et al., 2015). The differential alternative splicing events were detected by rMATS (v.4.0.1) (Shen et al., 2014). Reference-based transcriptome assembly and quantification were carried out using StringTie (v.1.3.3) (Pertea et al., 2015). The coding potential of novel transcripts was measured by CPAT software (v.1.2.4) (Wang et al., 2013). Ballgown (v.2.10.0) was applied to calculate FPKM (Fragments per kilobase of transcript per million mapped reads) values of each mRNA and lncRNA (Frazee et al., 2015). We filtered out the lowly expressed transcripts and only transcripts with average FPKM >= 0.5 in torn supraspinatus muscles or matched unharmed subscapularis muscles were selected for the subsequent analysis.

5.6 miRNA-seq library construction

The miRNA-seq library was constructed using NEB Multiplex Small RNA Library Prep Set for Illumina (NEB, United States). Briefly, the total RNA of each sample was used to prepare the miR sequencing library, which included the following steps: (1) 3′-adaptor ligation; (2) 5′-adaptor ligation; (3) the cDNA synthesis; (4) PCR amplification; and (5) size selection of 135–155 bp PCR-amplified fragments (corresponding to ~15–35 nt small RNAs). Constructed miRNA-Seq library was controlled for quality using Agilent 2100 Bioanalyzer system (Agilent Technologies, CA, United States).

5.7 miRNA sequencing

The miRNA-seq library was denatured as single-stranded DNA molecules, captured on Illumina flow cells, amplified *in situ* as clusters and finally sequenced for 51 cycles on Illumina NextSeq 500 Sequencer according to the manufacturer's instructions.

5.8 miRNA quantification

The total raw miRNA sequencing reads were filtered using a Solexa CHASTITY quality control filter. Reads were trimmed using cutadapt (v.1.14) (Martin, 2011) and aligned to the human reference genome (GRCh38) with the bowtie. Then the quantitation of miRNAs expression and novel miRNA prediction was done using miRDeep2 (v.0.0.8) (Friedlander et al., 2012). The read counts were normalized by CPM (Counts per million reads) approach. After normalization, miRNAs with average CPM >1 in torn supraspinatus muscles or matched unharmed subscapularis muscles were selected for the subsequent analysis.

5.9 CircRNA identification and quantification

For identification and quantification of circRNAs, reads that passed quality control were filtered to obtain trimmed data. Trimmed reads were then aligned to the human reference genome (GRCh37) using STAR (v.2.5.2b) (Dobin et al., 2013). Circexplorer2 (v.2.3.2) pipeline was used to identify the back-splice junction (circRNA) and quantify the back-splice junction reads (Zhang et al., 2016). Next, the read counts were normalized by CPM (Counts per million reads) approach. Following the analysis strategy of previous works (An et al., 2019; Wu et al., 2021; Liao et al., 2022), high confidence circRNAs were selected for subsequent differential expression analysis based on a stringent threshold of average CPM >100 in torn supraspinatus muscles or matched unharmed subscapularis muscles. The identified circRNAs were converted to circRNA ID with the web server circBase and other circRNAs were named according to the genomic locus (Glazar et al., 2014).

5.10 Differential expression analysis

Differential expression analysis of mRNAs and lncRNAs was performed by Ballgown (average FPKM >= 0.5 in torn supraspinatus muscles or matched unharmed subscapularis muscles, $p < 0.05$ and absolute fold change >1.5) (Frazee et al., 2014). EdgeR (v.3.20.9) (Robinson et al., 2010) was applied to identify differentially expressed miRNAs (average CPM >1 in torn supraspinatus muscles or matched unharmed subscapularis muscles, $p < 0.05$ and absolute fold change >1.5) and circRNAs (average CPM >100 in torn supraspinatus muscles or matched unharmed subscapularis muscles, $p < 0.05$ and absolute fold change >1.5). We also used GLMM model to identify differentially expressed circRNAs (Buratin et al., 2022), differential expressed circRNAs with FDR adjusted P values <0.05 were considered significant.

5.11 Co-expression analysis

Pearson correlation coefficient (PCC) was calculated for each lncRNA-mRNA pair and each circRNA-mRNA pair across all samples. For the co-expression networks, we only kept lncRNA/circRNA-mRNA pairs with absolute PCC >0.9. All the networks in this study were visualized using Cytoscape (v.3.7.0) (Shannon et al., 2003).

5.12 Functional enrichment analysis and gene set enrichment analysis

To further understand the potential functions and mechanisms of genes involved in this study, Gene Ontology (GO) enrichment analysis was performed using clusterProfiler (v.3.18.0) (Yu et al., 2012). GO terms with adjusted p-values < 0.05 were considered as significant.

5.13 Construction of the ceRNA network

Differential expressed mRNAs/circRNAs with uncorrected $p < 0.05$ were used to construct the ceRNA network. The regulatory relationship between miRNAs and mRNAs, lncRNAs or circRNAs were downloaded from the starbase, mirTarbase, and lncbase (Yang et al., 2011; Paraskevopoulou et al., 2016; Chou et al., 2018). We first constructed global ceRNA

networks for lncRNA-mRNA pairs and circRNA-mRNA pairs, respectively, based solely on the significant sharing of miRNAs. For novel circRNAs, miRNAs targeting their host genes were considered as their regulatory miRNAs. The statistical significance of each lncRNA/circRNA-mRNA pair on sharing common miRNAs can be calculated by the hypergeometric test, which was calculated as follows:

$$p = 1 - \sum_{x=0}^{x-1} \frac{\binom{K}{x}\binom{N-K}{M-x}}{\binom{N}{M}}$$

Where N is the number of all human miRNAs, K represents the total number of miRNAs regulating candidate ceRNA A, M represents the total number of miRNAs regulating candidate ceRNA B, and x is the number of shared miRNAs between A and B. False discovery rate (FDR) was employed to correct the p-values, only lncRNA/circRNA-mRNA pairs with an FDR <0.05 were selected to construct the global ceRNA networks.

Then we extracted ceRNA networks consisting of only differentially expressed mRNAs, miRNAs, lncRNAs and circRNAs and calculated PCC for each lncRNA/circRNA-mRNA pair across all samples. Only lncRNA/circRNA-mRNA pairs with significantly positive correlation were retained to construct dysregulated ceRNA networks for lncRNAs and circRNAs.

5.14 Identification of dysregulated ceRNA modules and key non-coding RNAs

We combined the dysregulated ceRNA networks of lncRNAs and circRNAs, then utilized clusterMaker2 (v.1.3.1, https://apps.cytoscape.org/apps/clustermaker2) (Morris et al., 2011) plug-in in Cytoscape to identify network modules. In the process of identifying the module, the absolute value of the PCC was calculated for each lncRNA/circRNA-mRNA pair across all samples and it was used as the weight of edge in the network. Then ncRNAs with the highest degree and betweenness centrality were chosen as key ncRNAs (circRNAs or lncRNAs). MiRNAs that significantly regulate lncRNA/circRNA-mRNA pairs in the module were defined as key miRNAs. The hypergeometric test was used to identify these key miRNAs from differentially expressed miRNAs (p-value < 0.05, Figure 6A).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE199486.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Peking University People's Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors contributed to the study conception and design. Material preparation was performed by JC and YZ, data collection was performed by SH and AL, analysis was performed by DZ and YX. The first draft of the manuscript was written by XL and YZ, all authors commented on previous versions of the manuscript. All authors reviewed and approved the final manuscript.

Funding

This work was supported by the Key Laboratory of Trauma and Neural Regeneration (Peking University) of the Ministry of Education of China, No. BMU2020XY005-03; Science and Technology Planning Project of Beijing of China, No D161100002816001; The Beijing Municipal Science and Technology Project, Z181100001718159; Peking University People's Hospital Scientific Research Development Funds (RDG2021-01, RDL2021-08).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alto, S. I., Chang, C. N., Brown, K., Kioussi, C., and Filtz, T. M. (2021). Gene expression profiling of skeletal muscles. *Genes (Basel)* 12 (11), 1718. doi:10.3390/genes12111718

An, T., Zhang, J., Ma, Y., Lian, J., Wu, Y. X., Lv, B. H., et al. (2019). Relationships of Non-coding RNA with diabetes and depression. *Sci. Rep.* 9 (1), 10707. doi:10.1038/s41598-019-47077-9

Anastasiadou, E., Jacob, L. S., and Slack, F. J. (2018). Non-coding RNA networks in cancer. *Nat. Rev. Cancer* 18 (1), 5–18. doi:10.1038/nrc.2017.99

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2020). FastQC: a quality control tool for high throughput sequence data Avilable from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 27.2020

Applegate, K. A., Thiese, M. S., Merryweather, A. S., Kapellusch, J., Drury, D. L., Wood, E., et al. (2017). Association between cardiovascular disease risk factors and rotator cuff tendinopathy: A cross-sectional study. *J. Occup. Environ. Med.* 59 (2), 154–160. doi:10.1097/JOM.0000000000000929

Bachasson, D., Singh, A., Shah, S. B., Lane, J. G., and Ward, S. R. (2015). The role of the peripheral and central nervous systems in rotator cuff disease. *J. Shoulder Elb. Surg.* 24 (8), 1322–1335. doi:10.1016/j.jse.2015.04.004

Buratin, A., Romualdi, C., Bortoluzzi, S., and Gaffo, E. (2022). Detecting differentially expressed circular RNAs from multiple quantification methods using a generalized linear mixed model. *Comput. Struct. Biotechnol. J.* 20, 2495–2502. doi:10.1016/j.csbj.2022.05.026

Chakraborty, S., Zawieja, D. C., Davis, M. J., and Muthuchamy, M. (2015). MicroRNA signature of inflamed lymphatic endothelium and role of miR-9 in lymphangiogenesis and inflammation. *Am. J. Physiol. Cell Physiol.* 309 (10), C680–C692. doi:10.1152/ajpcell.00122.2015

Chauveau, C., Rowell, J., and Ferreiro, A. (2014). A rising titan: TTN review and mutation update. *Hum. Mutat.* 35 (9), 1046–1059. doi:10.1002/humu.22611

Chen, J., Zhang, J., Gao, Y., Li, Y., Feng, C., Song, C., et al. (2021). LncSEA: a platform for long non-coding RNA related sets and enrichment analysis. *Nucleic Acids Res.* 49 (D1), D969–D980. doi:10.1093/nar/gkaa806

Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., et al. (2018). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 46 (D1), D296–D302. doi:10.1093/nar/gkx1067

Connor, D. E., Paulus, J. A., Dabestani, P. J., Thankam, F. K., Dilisio, M. F., Gross, R. M., et al. (2019). Therapeutic potential of exosomes in rotator cuff tendon healing. *J. Bone Min. Metab.* 37 (5), 759–767. doi:10.1007/s00774-019-01013-z

Connor, F., Wright, E., Denny, P., Koopman, P., and Ashworth, A. (1995). The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse. *Nucleic Acids Res.* 23 (17), 3365–3372. doi:10.1093/nar/23.17.3365

Czubak, K., Taylor, K., Piasecka, A., Sobczak, K., Kozlowska, K., Philips, A., et al. (2019). Global increase in circular RNA levels in myotonic dystrophy. *Front. Genet.* 10, 649. doi:10.3389/fgene.2019.00649

Dang, A., and Davies, M. (2018). Rotator cuff disease: Treatment options and considerations. *Sports Med. Arthrosc. Rev.* 26 (3), 129–133. doi:10.1097/JSA.0000000000000207

De Marinis, Y., Sun, J., Bompada, P., Domenech Omella, J., Luan, C., Halu, A., et al. (2017). Regulation of nuclear receptor interacting protein 1 (NRIP1) gene expression in response to weight loss and exercise in humans. *Obes. (Silver Spring)* 25 (8), 1400–1409. doi:10.1002/oby.21899

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635

Duecker, R. P., De Mir Messa, I., Jerkic, S. P., Kochems, A., Gottwald, G., Moreno-Galdo, A., et al. (2022). Epigenetic regulation of inflammation by microRNAs in post-infectious bronchiolitis obliterans. *Clin. Transl. Immunol.* 11 (2), e1376. doi:10.1002/cti2.1376

Edouard, P., Degache, F., Beguin, L., Samozino, P., Gresta, G., Fayolle-Minon, I., et al. (2011). Rotator cuff strength in recurrent anterior shoulder instability. *J. Bone Jt. Surg. Am.* 93 (8), 759–765. doi:10.2106/JBJS.I.01791

Falcone, G., Perfetti, A., Cardinali, B., and Martelli, F. (2014). Noncoding RNAs: emerging players in muscular dystrophies. *Biomed. Res. Int.* 2014, 503634. doi:10.1155/2014/503634

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* 33 (3), 243–246. doi:10.1038/nbt.3172

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. (2014). Flexible isoform-level differential expression analysis with Ballgown. Biorxiv, 003665.

Frich, L. H., Fernandes, L. R., Schroder, H. D., Hejbol, E. K., Nielsen, P. V., Jorgensen, P. H., et al. (2021). The inflammatory response of the supraspinatus muscle in rotator cuff tear conditions. *J. Shoulder Elb. Surg.* 30 (6), e261–e275. doi:10.1016/j.jse.2020.08.028

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40 (1), 37–52. doi:10.1093/nar/gkr688

Fuchs, B., Zumstein, M., Regenfelder, F., Steinmann, P., Fuchs, T., Husmann, K., et al. (2008). Upregulation of alpha-skeletal muscle actin and myosin heavy polypeptide gene products in degenerating rotator cuff muscles. *J. Orthop. Res.* 26 (7), 1007–1011. doi:10.1002/jor.20577

Gache, V., Gomes, E. R., and Cadot, B. (2017). Microtubule motors involved in nuclear movement during skeletal muscle differentiation. *Mol. Biol. Cell* 28 (7), 865–874. doi:10.1091/mbc.E16-06-0405

Ge, Z., Tang, H., Lyu, J., Zhou, B., Yang, M., Tang, K., et al. (2020a). Conjoint analysis of lncRNA and mRNA expression in rotator cuff tendinopathy. *Ann. Transl. Med.* 8 (6), 335. doi:10.21037/atm.2020.02.149

Ge, Z., Zhou, B., Zheng, X., Yang, M., Lu, J., Deng, H., et al. (2020b). Circular RNA expression pattern and competing endogenous RNA network involved in rotator cuff tendinopathy. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi* 34 (5), 608–614. doi:10.7507/1002-1892.201911094

Glazar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. *RNA* 20 (11), 1666–1670. doi:10.1261/rna.043687.113

Greco, S., Cardinali, B., Falcone, G., and Martelli, F. (2018). Circular RNAs in muscle function and disease. *Int. J. Mol. Sci.* 19 (11), E3454. doi:10.3390/ijms19113454

Hagiwara, N., Klewer, S. E., Samson, R. A., Erickson, D. T., Lyon, M. F., and Brilliant, M. H. (2000). Sox6 is a candidate gene for p100H myopathy, heart block, and sudden neonatal death. *Proc. Natl. Acad. Sci. U. S. A.* 97 (8), 4180–4185. doi:10.1073/pnas.97.8.4180

Han, H., Gu, S., Chu, W., Sun, W., Wei, W., Dang, X., et al. (2017). miR-17-5p regulates differential expression of NCOA3 in pig intramuscular and subcutaneous adipose tissue. *Lipids* 52 (11), 939–949. doi:10.1007/s11745-017-4288-4

Hessel, A. L., Lindstedt, S. L., and Nishikawa, K. C. (2017). Physiological mechanisms of eccentric contraction and its applications: A role for the giant titin protein. *Front. Physiol.* 8, 70. doi:10.3389/fphys.2017.00070

Khanna, R., Saltzman, M. D., Elliott, J. M., Hoggarth, M. A., Marra, G. M., Omar, I., et al. (2019). Development of 3D method to assess intramuscular spatial distribution of fat infiltration in patients with rotator cuff tear: reliability and concurrent validity. *BMC Musculoskelet. Disord.* 20 (1), 295. doi:10.1186/s12891-019-2631-z

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi:10.1038/nmeth.3317

Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., and Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 20 (11), 675–691. doi:10.1038/s41576-019-0158-7

Kumar, K. R., Lohmann, K., Masuho, I., Miyamoto, R., Ferbert, A., Lohnau, T., et al. (2014). Mutations in GNAL: a novel cause of craniocervical dystonia. *JAMA Neurol.* 71 (4), 490–494. doi:10.1001/jamaneurol.2013.4677

Li, C. Q., Huang, G. W., Wu, Z. Y., Xu, Y. J., Li, X. C., Xue, Y. J., et al. (2017). Integrative analyses of transcriptome sequencing identify novel functional lncRNAs in esophageal squamous cell carcinoma. *Oncogenesis* 6 (2), e297. doi:10.1038/oncsis.2017.1

Li, D., Ding, X., Xie, M., Huang, Z., Han, P., Tian, D., et al. (2020). CAMSAP2-mediated noncentrosomal microtubule acetylation drives hepatocellular carcinoma metastasis. *Theranostics* 10 (8), 3749–3766. doi:10.7150/thno.42596

Li, Y., Li, X., Yang, Y., Li, M., Qian, F., Tang, Z., et al. (2021). TRlnc: a comprehensive database for human transcriptional regulatory information of lncRNAs. *Brief. Bioinform.* 22 (2), 1929–1939. doi:10.1093/bib/bbaa011

Liao, F., Zhu, L., Yang, J., Wu, X., Zhao, Z., Xu, B., et al. (2022). Whole transcriptome sequencing identified CircRNA profiles and the related networks in schizophrenia. *J. Mol. Neurosci.* 72 (8), 1622–1635. doi:10.1007/s12031-022-02013-x

Ling, J., Xie, X., Wang, Y., Huang, W., Luo, J., Su, J., et al. (2022). Differential expression profiles of miRNA in granulomatous lobular mastitis and identification of possible biomarkers. *Exp. Ther. Med.* 24 (2), 500. doi:10.3892/etm.2022.11427

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17 (1), 10–12. doi:10.14806/ej.17.1.200

Mollah, M. B., and Ishikawa, A. (2010). A wild derived quantitative trait locus on mouse chromosome 2 prevents obesity. *BMC Genet.* 11, 84. doi:10.1186/1471-2156-11-84

Morris, J. H., Apeltsin, L., Newman, A. M., Baumbach, J., Wittkop, T., Su, G., et al. (2011). clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinforma.* 12, 436. doi:10.1186/1471-2105-12-436

Nam, S. O., Shin, D., Park, K., Kim, T. K., and Kim, H. S. (2015). Pancoast syndrome accompanied by rotator cuff tear. *Clin. Shoulder Elb.* 18 (1), 43–46. doi:10.5397/cise.2015.18.1.43

Nelson, G. B., McMellen, C. J., Kolaczko, J. G., Millett, P. J., Gillespie, R. J., and Su, C. A. (2021). Immunologic contributions following rotator cuff injury and development of cuff tear arthropathy. *JBJS Rev.* 9 (11), e21. doi:10.2106/JBJS.RVW.21.00126

Noh, K. C., Park, S. H., Yang, C. J., Lee, G. W., Kim, M. K., and Kang, Y. H. (2018). Involvement of synovial matrix degradation and angiogenesis in oxidative stress-exposed degenerative rotator cuff tears with osteoarthritis. *J. Shoulder Elb. Surg.* 27 (1), 141–150. doi:10.1016/j.jse.2017.08.007

Paraskevopoulou, M. D., Vlachos, I. S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., et al. (2016). DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.* 44 (D1), D231–D238. doi:10.1093/nar/gkv1270

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi:10.1038/nbt.3122

Plachel, F., Heuberer, P., Gehwolf, R., Frank, J., Tempfer, H., Lehner, C., et al. (2020). MicroRNA profiling reveals distinct signatures in degenerative rotator cuff pathologies. *J. Orthop. Res.* 38 (1), 202–211. doi:10.1002/jor.24473

Randelli, P., Menon, A., Ragone, V., Creo, P., Bergante, S., Randelli, F., et al. (2016). Lipogems product treatment increases the proliferation rate of human tendon stem cells without affecting their stemness and differentiation capability. *Stem Cells Int.* 2016, 4373410. doi:10.1155/2016/4373410

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi:10.1093/bioinformatics/btp616

Sato, E. J., Killian, M. L., Choi, A. J., Lin, E., Esparza, M. C., Galatz, L. M., et al. (2014). Skeletal muscle fibrosis and stiffness increase after rotator cuff tendon injury and neuromuscular compromise in a rat model. *J. Orthop. Res.* 32 (9), 1111–1116. doi:10.1002/jor.22646

Schemitsch, C., Chahal, J., Vicente, M., Nowak, L., Flurin, P. H., Lambers Heerspink, F., et al. (2019). Surgical repair versus conservative treatment and subacromial decompression for the treatment of rotator cuff tears: a meta-analysis of randomized trials. *Bone Jt. J.* 101-B (9), 1100–1106. doi:10.1302/0301-620X.101B9.BJJ-2018-1591.R1

Schofer, M. D., Patzer, T., and Quante, M. (2008). Atypical manifestation of late onset limb girdle muscular dystrophy presenting with recurrent falling and shoulder dysfunction: a case report. *Cases J.* 1 (1), 402. doi:10.1186/1757-1626-1-402

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Shen, P. H., Lien, S. B., Shen, H. C., Lee, C. H., Wu, S. S., and Lin, L. C. (2008). Long-term functional outcomes after repair of rotator cuff tears correlated with atrophy of the supraspinatus muscles on magnetic resonance images. *J. Shoulder Elb. Surg.* 17, 1S–7s. doi:10.1016/j.jse.2007.04.014

Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., et al. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* 111 (51), E5593–E5601. doi:10.1073/pnas.1419161111

Smith, L. B., Anderson, C. V., Withangage, M. H. H., Koch, A., Roberts, T. J., and Liebl, A. L. (2022). Relationship between gene expression networks and muscle contractile physiology differences in Anolis lizards. *J. Comp. Physiol. B* 192 (3-4), 489–499. doi:10.1007/s00360-022-01441-w

Sun, J., Wang, X., Wang, D., Zhao, Z., Zhang, L., and Zhang, J. (2020). circAtp9b knockdown alleviates LPS-caused inflammation provided that microRNA-27a is upregulated. *Int. Immunopharmacol.* 78, 105925. doi:10.1016/j.intimp.2019.105925

Tao, X., Liu, J., Chen, L., Zhou, Y., and Tang, K. (2015). EGR1 induces tenogenic differentiation of tendon stem cells and promotes rabbit rotator cuff repair. *Cell. Physiol. Biochem.* 35 (2), 699–709. doi:10.1159/000369730

Terry, E. E., Zhang, X., Hoffmann, C., Hughes, L. D., Lewis, S. A., Li, J., et al. (2018). Transcriptional profiling reveals extraordinary diversity among skeletal muscle tissues. *Elife* 7, e34613. doi:10.7554/eLife.34613

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 41 (6), e74. doi:10.1093/nar/gkt006

Wieser, K., Joshy, J., Filli, L., Kriechling, P., Sutter, R., Furnstahl, P., et al. (2019). Changes of supraspinatus muscle volume and fat fraction after successful or failed arthroscopic rotator cuff repair. *Am. J. Sports Med.* 47 (13), 3080–3088. doi:10.1177/0363546519876289

Wu, P., Fang, X., Liu, Y., Tang, Y., Wang, W., Li, X., et al. (2021). N6-methyladenosine modification of circCUX1 confers radioresistance of hypopharyngeal squamous cell carcinoma through caspase1 pathway. *Cell Death Dis.* 12 (4), 298. doi:10.1038/s41419-021-03558-2

Yaman, S. O., Orem, A., Yucesan, F. B., Kural, B. V., and Orem, C. (2021). Evaluation of circulating miR-122, miR-30c and miR-33a levels and their association with lipids, lipoproteins in postprandial lipemia. *Life Sci.* 264, 118585. doi:10.1016/j.lfs.2020.118585

Yang, J. H., Li, J. H., Shao, P., Zhou, H., Chen, Y. Q., and Qu, L. H. (2011). starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.* 39, D202–D209. doi:10.1093/nar/gkq1056

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, X. O., Dong, R., Zhang, Y., Zhang, J. L., Luo, Z., Zhang, J., et al. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 26 (9), 1277–1287. doi:10.1101/gr.202895.115

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 20 September 2022
DOI 10.3389/fmolb.2022.931525

OPEN ACCESS

EDITED BY
Gian Gaetano Tartaglia,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Chenchen Zhao,
Chinese Academy of Agricultural Sciences (CAAS), China
Nitika,
University of North Carolina at Charlotte, United States

*CORRESPONDENCE
Zhongshi Zhou,
zhouzhongshi@caas.cn

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 29 April 2022
ACCEPTED 30 August 2022
PUBLISHED 20 September 2022

CITATION
Zhang Y, Ma W, Ma C, Zhang Q, Tian Z, Tian Z, Chen H, Guo J, Wan F and Zhou Z (2022), The *hsp70* new functions as a regulator of reproduction both female and male in Ophraella communa. *Front. Mol. Biosci.* 9:931525. doi: 10.3389/fmolb.2022.931525

The *hsp70* new functions as a regulator of reproduction both female and male in *Ophraella communa*

Yan Zhang[1], Weihua Ma[2], Chao Ma[1], Qinglu Zhang[1], Zhenya Tian[1,3], Zhenqi Tian[1], Hongsong Chen[1,3], Jianying Guo[1], Fanghao Wan[1] and Zhongshi Zhou[1,4]*

[1]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China, [2]Hubei Insect Resources Utilization and Sustainable Pest Management Key Laboratory, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [3]Guangxi Key Laboratory of Biology for Crop Diseases and Insect Pests, Institute of Plant Protection, Guangxi Academy of Agricultural Sciences, Nanning, China, [4]National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya, China

Heat shock proteins (Hsps) function as molecular chaperones that enable organisms to withstand stress and maintain normal life activities. In this study, we identified heat shock protein 70 (encoded by *hsp70*), which exhibits a higher expression in the mature male testis than in the unmature testis of *Ophraella communa*. Tissue expression profile revealed that *Ochsp70* levels in males were highest in the testis, whereas those in females were highest in the head. Moreover, the expression of *Ochsp70* was found to be significantly induced in female bursa copulatrix after mating. Double-stranded RNA ds*Ochsp70* was injected into males to performance RNA interference, which significantly decreased the male *Ochsp70* expression levels within 20 d post-injection, whereas no effect was observed on the *Ochsp70* expression level in the females after mating with ds*Ochsp70*-injected males. However, significant downregulation of female fertility was marked simultaneously. Furthermore, knockdown of female *Ochsp70* expression also led to a significant reduction in fertility. Finally, comparative transcriptomic analysis identified glucose dehydrogenase and insulin-like growth factor binding protein as putative downstream targets of *Ochsp70*. Overall, we deduced that *Ochsp70* is an indispensable gene and a potential male mating factor in *O. communa*, which regulates reproduction.

KEYWORDS

***hsp70*, male mating factor, fertility, *Ophraella communa*, pathway**

Introduction

When originating from their internal or external environments, the heat shock response of cells is activated to respond to the protein-damaging (proteotoxic) effects of stress (Sørensen et al., 2003). Heat shock genes are a subset of genes that encode for molecular chaperones called heat shock proteins, including a stress-related groups of proteins generated or synthesized by cells under the effect of high temperature (heat shock) or other stress stimuli. Hsps commonly exist in both prokaryotes and eukaryotes. Based on molecular weight (kDa), Hsps are divided into four types, Hsp90, Hsp70, Hsp60, and small Hsps, which are involved in the transport, folding, unfolding, assembly, and disassembly of multi-structured units, and in the degradation of misfolded or aggregated proteins (Lindquist, 1985; Feder and Hofmann, 1999; Pockley et al., 2007).

Hsp70 is a predominant Hsp family, and the previous studies were mainly focused on unraveling the important roles of this family in restoring the native conformation of proteins after experiencing stress (temperature, hypoxia, oxidative stress, pesticides, radiation, etc.) (Morimoto, 1993). In addition, members of the Hsp70 family are vital for the folding and intracellular trafficking of *denovo* synthesized proteins under normal conditions (Zatsepina et al., 2021). The highly dynamic nature of Hsp70 is a key factor responsible for its chaperone function (Clark and Peck, 2009; Mayer, 2013). Normally, Hsp70 is located in the cytoplasm, however, when cells are stimulated by heat stress, Hsp70 in the cytoplasm is rapidly transferred to the nucleus. Nuclear translocation of Hsp70 protects the cells from the damaging caused by hypoxia and high temperature (Velazquez and Lindquist, 1984). Interestingly, *hsp70* gene expression has also been reported to determine the variation in fitness and geographical distribution of *Nucella* species (Sorte and Hofmann, 2005), and a similar phenomenon has been noted in marine organisms (Clark and Peck, 2009). During the evaluation of contaminated environments, the *hsp70* gene may serve as a biomarker to detect adverse circumstances (Cristina et al., 2018). In mammals, certain Hsps have been identified in the seminal fluid, which play important roles in spermatogenesis, sperm-egg recognition, and the post-testicular maturation of mammalian spermatozoa (Walsh et al., 2008; Dun et al., 2012; Redgrove et al., 2012; Nixon et al., 2015). In boars, Hsp70 is associated with semen quality, which tends to decline significantly with Hsp70 levels (Huang et al., 2000). In insects, studies regarding the functional characterization of *hsp70* are emergent. *hsp70* gene is differentially regulated in response to diapause (Macrae, 2010), and a similar change is recorded for other influencing factors (King and MacRae, 2015). *hsp70* is also involved in midgut metamorphosis in Spodoptera litura, wherein its expression is induced by hormones (Gu et al., 2012). In addition, *hsp70* is associated with reproductive diapause (Baker and Russell, 2009) and aging, and has a positive effect in prolonging the lifespan of *Drosophila melanogaster* (Bourg et al., 2001).

Ophraella communa (Coleoptera: Chrysomelidae) is used worldwide as an important biological control agent of the ragweed *Ambrosia artemisiifolia* worldwide (Zhou et al., 2011). *Ambrosia artemisiifolia* invaded China in the 1930s (Li et al., 2015) and posed a serious threat to agriculture and ecosystem (Zhou et al., 2011; Smith et al., 2013). The *O. communa* feeds on foliage at both larval and adult stages, and either restricts the ragweed can not enter the vegetative genitals or die directly (Guo et al., 2011). Ragweed is spreading rapidly in China (Guo et al., 2011), and the new areas of *A. artemisiifolia* distribution lack a natural enemy population, making it particularly dangerous. Therefore, a prompt release of *O. communa* populations is required in these areas to prevent further propagation of this weed. In previous studies, we have investigated the biology and physiology of *O. communa* (Ma et al., 2019a; Ma et al. 2019b; Ma et al. 2020; Tian et al., 2021; Zhang et al., 2021), and found that these leaf beetles are bisexual reproductive insects that can mate multiple times per day after sexual maturity.

In the present study, we identified the *hsp70* genes that were highly expressed from a cDNA library of male testes in *O. communa*. We noted that *hsp70* is preferentially expressed in mature testes compared to unmature ones, and is also significantly upregulated in the bursa copulatrix (BC) of mated females. To further elucidate the potential functions of *Ochsp70*, we examined the tissue-specific transcript abundance patterns of *Ochsp70* in males and females. Then, we used the RNA interference (RNAi) technique to further demonstrate its role in reproduction in males and females. Finally, a comparative transcriptome analysis of RNAi-treated females (dsgfp vs ds*hsp70*) was carried out, and the potential mechanisms by which *Ochsp70* regulates reproduction were discussed.

Materials and methods

Plant growth and *O.communa* rearing

The *A. artemisiifolia* plants used in the present study were grown by following a previously reported method (Zhou et al., 2010). *Ophraella communa* population had been raised on ragweed plants for 1 year in the laboratory (Chinese Academy of Agricultural Sciences, Institute of Plant Protection, Beijing, China) at 27 ± 1°C, 70 ± 5% relative humidity, and a photoperiod of 14/10 h (light/dark).

Sample collection, RNA extraction and cDNA synthesis

Diverse tissues, including head, thorax, fat body, gut, male accessory glands (MAG), testis, bursa copulatrix (BC), were collected from eight male and female *O. communa* adults at day 5 post-eclosion. The post-mating bursa copulatrix (M-BC) tissue was obtained from 15 females immediately after mating, while the unmated bursa copulatrix (U-BC) tissue was obtained from 20 unmated females of the same age. All tissue samples collected for this study were immediately frozen in liquid nitrogen and stored at −80 C. Three biological replicates were used for quantitative real-time polymerase chain reaction (qPCR) analysis. Subsequently, total RNA from all samples was extracted following the manufacturer's protocol using TRIzol™ reagent (Invitrogen, MA, United States). cDNA was synthesized using the TransScript® One-Step RT-PCR SuperMix (TransGen Biotech Co., Ltd, China) as per the recommended protocol.

Cloning and sequence analysis of *Ochsp70*

The rapid amplification of cDNA ends (RACE) approach was used to amplify the full-length cDNA sequence according to the manufacturer's guide (SMARTer® RACE 5′/3′ Kit, Clontech, TaKaRa Bio Inc, United States) based on local transcriptome data. The primer sequences are listed in Supplementary Table S1. The complete Coding sequence region was analyzed according to the smart website (https://smart.embl.de/), and the conserved site was predicted using the Prosite tool (https://prosite.expasy.org/). The full-length cDNAs of hsp70 were used as query sequences to search for hsp70 homologs in other insect genomes available in GenBank using NCBI-BLASTn (http://www.ncbi.nlm.nih.gov/). Multiple sequence alignment was performed using DNAMAN 8.0, and phylogenetic trees were constructed by the maximum-likelihood method using MAGE 6.06 and phylogenetic relationships were determined by bootstrap analysis with values of 1,000 trials.

qPCR analysis

qPCR was performed to quantify the relative *Ochsp70* expression levels in different tissues, including female and male, mating and unmating, and after double-stranded RNA (dsRNA) treatments. For this purpose, the ABI 7500 PCR detection system (Applied Biosystems, United States) was used. RPL19 was used as reference gene, as described by Zhang et al. (2020).

dsRNA synthesis and RNAi

PCR was carried out using a gene-specific primer pair containing a T7 promoter sequence (5′-TAATACGACTCA CTATAGGG-3′) at the 5′ end and a recombinant plasmid containing *OcHsp70* as template. Thereafter, the PCR product was used as a template for dsRNA synthesis using Ambion™ MEGAscript® T7 Transcription Kit (Thermo-Fisher Scientific, CA, United States) according to the recommended protocol. The double-stranded green fluorescent protein (*gfp*) RNA, ds*gfp*, was used as blank (negative) control. Finally, the quality of ds*Ochsp70* was assessed using 1% agarose gel electrophoresis and quantified to 10 μg/ul. ds*OcHsp70* and dsgfp were stored at −80 C for subsequent experiments (Jin et al., 2020).

For the RNAi experiment, newly emerged adults (males and females <12 h after eclosing) were injected with 500 ng of dsRNA in 100 nL water solution at the abdomen using the Nanoject III Programmable Nanoliter Injector (Drummond Scientific Co., Inc, PA, United States). At 5, 10, 15, and 20 d post injection (PI), the five injected adults of each biological replicate were collected for the evaluation of silencing efficiency using qPCR. The primers used in this study are listed in Supplementary Table S1.

Bioassay for *O.communa* fecundity

Fecundity was assayed using single male-female mating pairs. The dsRNA (ds*hsp70* or ds*gfp*) injected adults (male or female) were mated with virgin (unmated) adults of the opposite sex and same age without dsRNA injection at 3 d PI. Each pair of adults was grouped in a Petri dish containing robust *A. artemisiifolia* leaves with wet cotton. The number of eggs laid from per pair per day was recorded every day until 20 d PI. The egg hatching rate was calculated as the percentage of hatched larvae among the total number of the eggs laid in the first 5 days.

RNA-sequencing

To identify the potential interactors of *hsp70* particularly related to reproduction, the global transcriptome profiles of ds*Ochsp70*-treated and dsgfp-treated females were investigated and compared using high-throughput sequencing. To this end, RNA was extracted from all samples, and the *Ochsp70*-silencing efficiency of each sample was evaluated via qPCR before transcriptome sequencing.

Data analysis

Data from qPCR and bioassays were analyzed using SAS System for Windows V8. The qPCR data was analyzed using the 2−ΔΔCt method (Schmittgen, 2008). One-way ANOVA was

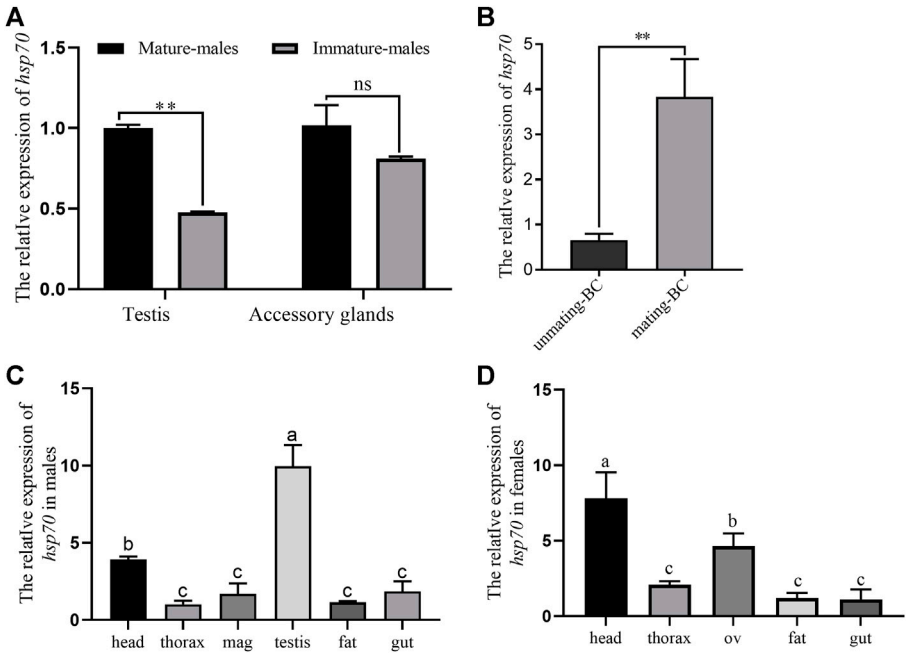


FIGURE 1
A) Expression levels of *Ochsp70* in male testicular tissues; the expression patterns validation validate of the *Ochsp70* transcriptome data in the male reproductive system. **(B)** Expression level of *Ochsp70* in mating-bursa copulatrix (mating-BC) than unmating-bursa copulatrix (unmatingm-BC). **(C)** Expression profiles of *Ochsp70* in different tissues of males beetles *O. communa*. **(D)** Expression profiles of *Ochsp70* in different tissues of females beetles *O. communa*. Values are represent means ± SD. The data were analyzed usingby one-way ANOVA followed by the least significant difference (LSD) test. *$p < 0.05$, **$p < 0.01$.

performed to compare the variation between PCR data and bioassays, followed by a least significant difference (LSD) test for multiple comparisons. Differences among mean values were determined using a LSD test at $p < 0.05$.

Results

Ochsp70 identification and sequence analysis

The full-length cDNA of *Ochsp70* was obtained by RACE-PCR and submitted to GenBank (GenBank number: OM162158), which consists of a 2,472 bp-long open reading frame encoding a polypeptide of 824 amino acids, and 186 bp long 5′ and 247 bp long 3′ untranslated regions. The molecular weight of *OcHsp70* was predicted to be 91.98 kDa and the isoelectric point was 5.63, according to the ExPasy tools. The motif VEIVGGSSRIPAIKQ was found to be highly conserved in *Ochsp70* and its homologs from other coleopteran species (Supplementary Figure S1), and *Ochsp70* shares the HSPA4_like_NDB domain with these species. Homology analysis showed that the highest sequence similarity among *OcHsp70* and other coleopteran Hsp70 proteins was 81.12% (Supplementary Figure S2). Meanwhile, phylogenetic analysis revealed that the *Hsp70* clustered with strong bootstrapping support on the basis of the insect order of origin, whereas the amino acid sequences derived from insects of different orders were clustered in one clade, indicating that these *Hsp70* are conserved within the same order of insects. The *Ochsp70* sequence displayed the highest homology with that of *Diabrotica virgifera* (Supplementary Figure S1).

Ochsp70 is highly expressed in the female ovaries and male testes and is induced by mating

The relative expression of *Ochsp70* in male testes was significantly higher in mature testes than in unmature testes (Figure 1A). We also observed that the expression level of *OcHsp70* in BC (the female organ for storage of sperm and seminal fluid protein) increased significantly after mating (Figure 1B). Furthermore, the expression domain analysis of *Ochsp70* revealed the highest expression in the testes in males and in the heads in females (Figure 1C,D).

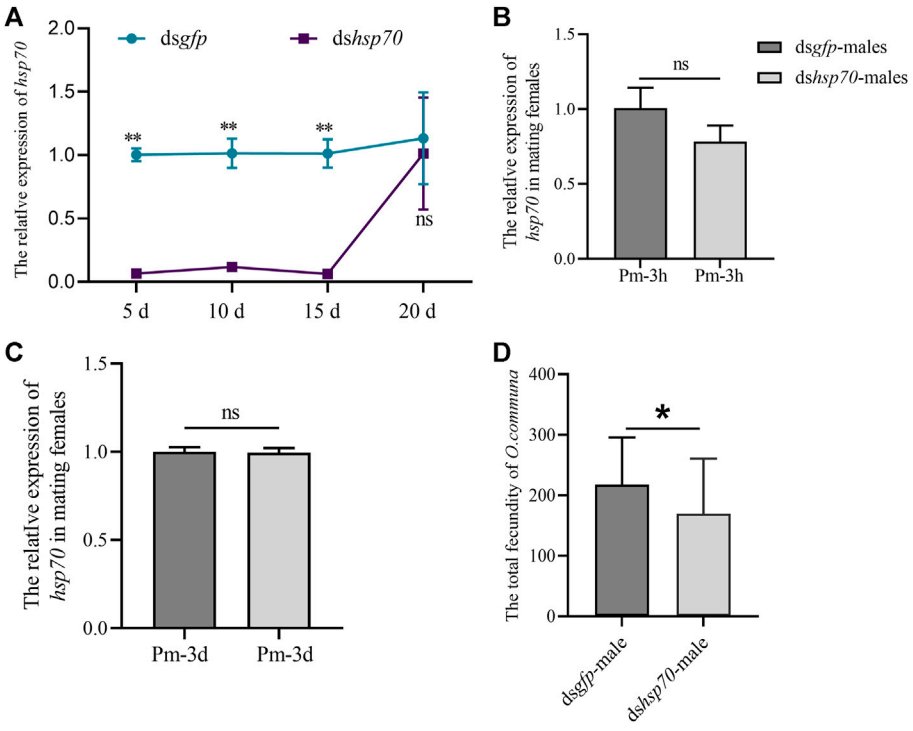


FIGURE 2
Functional characterization of *Ochsp70* and evaluation of RNA silencing efficiency in males. **(A)** Expression levels of *Ochsp70* (5, 10, 15, and 20 d) after dsRNA was injected into the males. **(B)** Expression levels of *Ochsp70* in the reproductive system of females copulated with dsRNA-injected males, 3 h post-mating (Pm-3h). **(C)** Expression levels of *Ochsp70* in the reproductive system of females copulated with dsRNA-injected males, 3 d post-mating (Pm-3d). **(D)** Effect of *Ochsp70* on *O. communa* fecundity. Bars with the same letter are not significantly different from each other at $p < 0.05$, as per the LSD test.

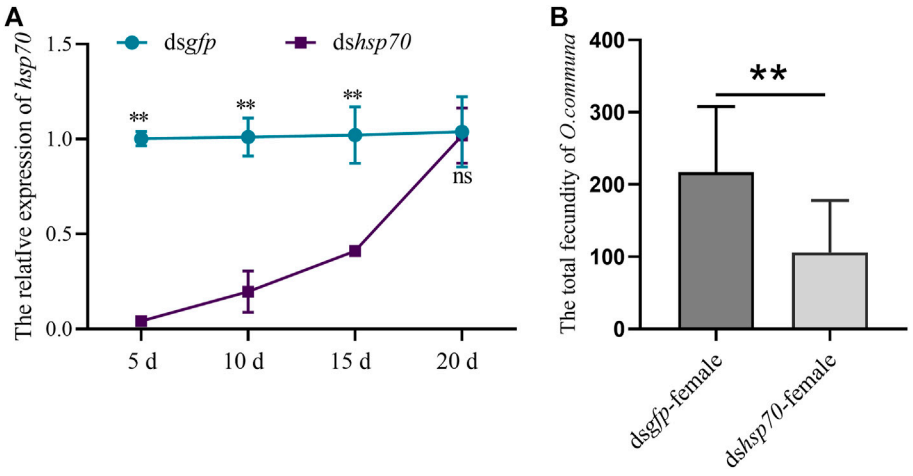


FIGURE 3
Functional characterization of *Ochsp70* and evaluation of RNA silencing efficiency in females. **(A)** Expression levels of *Ochsp70* (5, 10, 15, and 20 d) after dsRNA was injected dsRNA was injected into the females. **(B)** Effect of *Ochsp70* on female fecundity. Bars with the same letter are not significantly different from each other at $p < 0.05$ level, as per the LSD test.

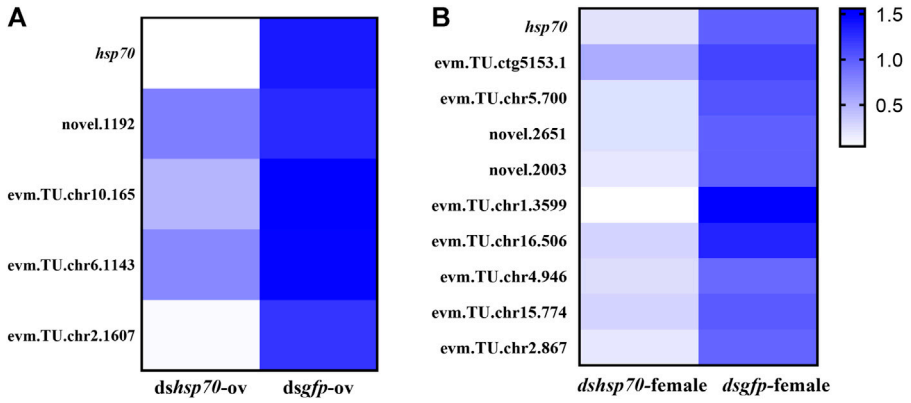


FIGURE 4
Verification of differential gene expression between two RNA interference-treated groups by quantitative real-time polymerase chain reaction represented as heat maps. **(A)** ds*hsp70*-ov vs. ds*gfp*-ov. **(B)** ds*hsp70*-female vs. ds*gfp*-female.

The knockdown of *Ochsp70* reduces the fertility of *O. communa* males

The *Ochsp70* expression was significantly reduced on the fifth day PI until 20th day PI in males (Figure 2A). In the meantime, we also tested the expression of *Ochsp70* in the reproductive system of females that mated with males injected with ds*Ochsp70*, and the results showed no significant changes (Figure 2B,C). However, the fecundity of these females was 26% lower than those administered ds*gfp* (Figure 2D). These results indicate that *Ochsp70* is a putative male mating factor that plays an important role in reproduction.

The knockdown of *Ochsp70* reduces the fertility of *O. communa* females

To illustrate whether *Ochsp70* was involved in regulating reproduction in females, ds*Ochsp70* was injected into freshly emerged females of *O. communa*. Similar to their male counterparts, the females displayed a significant reduction in *Ochsp70* expression from the fifth to the 20th day PI (Figure 3A). Furthermore, the number of eggs laid by the dshsp70-treated females decreased by 56% compared to the control (Figure 3B). These results further suggest that *Ochsp70* has a crucial role in the regulating the reproduction of *O. communa* females.

Ochsp70 knockdown impacts fertility-related pathways in *O. communa*

To elucidate the potential pathway of Ochsp70 mediated regulation of the reproduction in *O. communa*, total RNA was extracted from RNAi-treated mated female adults and was subjected to transcriptome sequencing (Supplementary Table S2). The raw data has been uploaded to NCBI (BioProject accession: PRJNA796368). Comparative transcriptomic analysis revealed significant alterations in the expression profiles of multiple genes associated with pathways involved in stress, reproductive development, and reproduction. Among the differentially expressed genes, we noticed that two fecundity-related genes, glucose dehydrogenase (evm.TU.chr5.700, GDH) and insulin-like growth factor binding protein (novel. 2003, IGF-BP) were downregulated 5.046 and 7.8136 times, respectively, after ds*Ochsp70* treatment. These results were subsequently validated by the relative expression levels quantified using qPCR (Figure 4).

Discussion

Insects produce Hsps in response to stress such as heat, cold, crowding, and anoxia. In concert with cochaperones and accessory proteins, Hsps mediate essential activities such as folding, assembly, intracellular localization, secretion, regulation, and degradation of other proteins (Hendrick and Hartl 1993). Previous studies have reported that Hsp exhibits characteristic and distinctive expression patterns during various stages of development, including gametogenesis (Dix, 1997) and embryogenesis (Heikkila, 1993; Krone et al., 2003), However, the role and significance of the high Hsp levels in the absence of stress stimuli remain unclear. In this study, the *Ochsp70* gene was successfully isolated and was found to display a constitutive and preferential expres sion profile in male testes. Interestingly, knocking down male *Ochsp70* resulted in diminished fertility in their female mates. A similar observation was made in *Tribolium castaneum*, wherein Hsp70 was found to be involved in reproductive regulation when *Tchsp70* knock-down males were examined (Xu et al., 2013). In addition,

Ochsp70 expression is highest in the female heads, which is similar with *Nilaparvata lugens* (Lu et al., 2018) and *Cydia pomonellao* (Yang et al., 2016), and some small hsp genes are also abundant in head (Sun et al., 2014; Li et al., 2019). As a chaperone, hsp might play an important role in maintaining the normal function of the insect brain, either olfaction or neuro/developmental processing (Yang et al., 2016). More importantly, *Ochsp70* expression is highly in the female ovaries, which is consistent with the ovary-specific expression of *Tchsp70* and *Dmhsp70* (Marin and Tanguay, 1996; Xie et al., 2019), and constitutive expression of Hsp70 has been confirmed in mammalian oocytes (Dix, 1997), and hsp expression in female reproduction tissue and spermatogenesis was showed to correlate with HSP reproduction function (Neuer et al., 2000). Our results also revealed that the knockdown of female *Ochsp70* expression led to reduced reproduction both males and females. This finding supports the previous prediction that the shsp and *hsp70* genes may regulate reproduction in *T. castaneum* (Xie et al., 2019) and *Agasicles hygrophila* (Jin et al., 2020). However, expression level of *hsp70* gene might be a balancer of benefits and costs. During the response of *D. melanogaster* against heat shock, the *hsp70* expression increases, whereas the fecundity decreases (Krebs and Loeschcke, 1994; Huang et al., 2007), and growth and cell division are impeded (Feder, 1997).

In general, males offer male mating factors (e.g., seminal fluid proteins or other synthesized secretions) to females to ensure successful mating or to signify paternal investment (Thornhill, 1983; Avila et al., 2011). In the Hsp family, *hsp60* was present in the upregulated gene cluster obtained from the mated females of *D. melanogaster* (Mack et al., 2006), and Hsp70 was identified as a seminal fluid protein in *T. castaneum* (Xu et al., 2013). In this study, female *Ochsp70* expression was also induced through mating, which combined with the high expression levels of male *Ochsp70* in the mature testis, suggests that *Ochsp70* may functions as a male mating factor in *O. communa*. Similarly, 32 HSPs constitute a group of most abundant proteins in the adult testis proteomics of *Bombyx mori*, a 94.4 kDa Hsp70was also included (Zhang et al., 2014), which were considered to be associated with spermatogenesis, reproduction, mitosis, and fertilization. This phenomenon is even more comprehensible in mammals (Boelens et al., 2004; Jha et al., 2013; Zhang et al., 2014), wherein several hsp70 genes are expressed specifically in male germ cells (Dix, 1997; Neuer et al., 2000; Carreira and Santos, 2020). Testicular sperms are the most diverse of all cell types, so it is not surprising that spermatogenesis is accompanied by the expression of hsp gene different expression.

However, the knockdown of female *Ochsp70* led to a reduction in egg production, suggesting that *Ochsp70* may also be related to protein transport and nutrient supply in females, as observed previously (Marin and Tanguay, 1996). Hsp70 does not function independently and is associated with a team of cochaperones. In addition, hsp expression results from the activation of various intracellular signaling pathways (Feder and Hofmann, 1999). Liu et al. (2013) has been predicted that hsp90 is involved in regulating 20E and JH-inducible gene expression in Helicoverpa armigera, which may be another possible pathway for Hsp family-mediated reproductive regulation. In the present study, several pathways were revealed via RNA-sequencing analysis as potential downstream targets of *Ochsp70* involved in the regulating the reproduction in *O. communa*, such as Foxo signaling pathway, MAPK signaling pathway and insect hormone biosynthesis. Particularly, we noticed that both GDH and IGF-BP were maximally down-regulated with decreasing expression of *hsp70*. Previous studies showed that GDH and IGF-BP are involved in reproduction-related pathways and homeostasis (Smykal and Raikhel, 2015), while GDH is also associated with lifespan regulation (Von Wyschetzki et al., 2015). Unfortunately, in our study, after we silenced *Gdh* and *Igf-bp*, respectively or combined, the female fertility was non-different (Supplementary materials). Meanwhile, when the expression of hsp70, Gdh and Igf-bp in female of *O. communa* was interfered simultaneously, the female fecundity decreased obviously, compared with the control (Supplementary materials). Hence, GDH and IGF-BP may be not directly regulate reproduction, and are not directly related to Hsp70. The process of Hsp70 involved in reproduction is multimodulated in males and females of *O. communa*, next we will contribute to explore and find out this mechanism.

Conclusion

Our study provides evidence that hsp70 is a regulator of *O. communa* reproduction. Our findings also supports the notion that Ochsp70 is a potential male mating factor. A high-throughput approach was used to analyze the potential regulatory mechanism of the function of Hsp family in reproduction. However, further studies are required to elucidate the gene regulatory network involved in Hsp-mediated regulation of reproduction.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

ZZ conceived, designed, coordinated the study, and helped with the drafting of the manuscript; YZ and QZ performed the experiments and participated in

data analysis; WM participated in designing the study and drafted the manuscript; CM, HC, ZT, and JG were involved in data analysis; ZZ, YZ, WM, ZT, and FW composed the manuscript and performed statistical analyses. All authors have approved the manuscript for publication.

Funding

This study was supported by the National Natural Science Foundation of China (Nos. 31672089 and 31972340).

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer CZ declared a shared affiliation with the author(s). ZZ, YZ, CM, QZ, TZ, ZT, HC, JG, and FW to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.931525/full#supplementary-material

References

Avila, F. W., Sirot, L. K., LaFlamme, B. A., Rubinstein, C. D., and Wolfner, M. F. (2011). Insect seminal fluid proteins: Identification and function. *Annu. Rev. Entomol.* 56, 21–40. doi:10.1146/annurev-ento-120709-144823

Baker, D. A., and Russell, S. (2009). Gene expression during *Drosophila melanogaster* egg development before and after reproductive diapause. *BMC Genomics* 10, 242. doi:10.1186/1471-2164-10-242

Boelens, W. C., Wit, N. D., Verschuure, P., King, S. M., Kappe, G., Muijen, G. V., et al. (2004). Testis-specific human small heat shock protein HSPB9 is a cancer/testis antigen, and potentially interacts with the dynein subunit TCTEL1. *Eur. J. Cell Biol.* 83, 337–345. doi:10.1078/0171-9335-00396

Bourg, R. L., Valenti, P., Lucchetta, P., and Payre, F. (2001). Effects of mild heat shocks at young age on aging and longevity in *Drosophila melanogaster*. *Biogerontology* 2, 155–164. doi:10.1023/A:1011561107055

Carreira, R. P., and Santos, D. L. D. (2020). *The role of HSP70 in sperm quality in HSP70s: Discovery, structure and functions*. New York: Nova Science Publishers, Inc

Clark, M. S., and Peck, L. S. (2009). HSP70 heat shock proteins and environmental stress in antarctic marine organisms: A mini-review. *Mar. Genomics* 1, 11–18. doi:10.1016/j.margen.2009.03.003

Cristina, M., Bast, O. D., and Silvia, F. C. (2018). HSP70 as a biomarker: An excellent tool in environmental contamination analysis—a review. *Water Air Soil Pollut.* 229, 264. doi:10.1007/s11270-018-3920-0

Dix, D. J. (1997). HSP 70 expression and function during gametogenesis. *Cell Stress Chaperones* 2, 73–77. doi:10.1379/1466-1268(1997)002<0073:heafdg>2.3.co;2

Dun, M., Aitken, R. J., and Nixon, B. (2012). The role of molecular chaperones in spermatogenesis and the post-testicular maturation of mammalian spermatozoa. *Hum. Reprod. Update* 18, 420–435. doi:10.1093/humupd/dms009

Feder, K., and Feder, M. E. (1997). Deleterious consequences of Hsp70 overexpression in *Drosophila melanogaster* larvae. *Cell Stress Chaperones* 2, 60–71. doi:10.1379/1466-1268(1997)002<0060:dcohoi>2.3.co;2

Feder, M. E., and Hofmann, G. E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. *Annu. Rev. Physiol.* 61, 243–282. doi:10.1146/annurev.physiol.61.1.243

Gu, J., Huang, L. X., Shen, Y., Huang, L. H., and Feng, Q. L. (2012). Hsp70 and small Hsps are the major heat shock protein members involved in midgut metamorphosis in the common cutworm, Spodoptera litura. *Insect Mol. Biol.* 21, 535–543. doi:10.1111/j.1365-2583.2012.01158.x

Guo, J. Y., Zhou, Z. S., Zheng, X. W., Chen, H. S., Wan, F. H., and Luo, Y. H. (2011). Control efficiency of leaf beetle *Ophraella communa*, on the invasive common ragweed, *Ambrosia artemisiifolia*, at different growing stages. *Biocontrol Sci. Technol.* 21, 1049–1063. doi:10.1080/09583157.2011.603823

Heikkila, J. J. (1993). Heat shock gene expression and development. II. An overview of mammalian and avian developmental systems. *Dev. Genet.* 14, 87–91. doi:10.1002/dvg.1020140202

Hendrick, J. P., and Hartl, F.-U. (1993). Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.* 62, 349–384. doi:10.1146/annurev.bi.62.070193.002025

Huang, L-H., Chen, B., and Kang, L. (2007). Impact of mild temperature hardening on thermotolerance, fecundity, and Hsp gene expression in Liriomyza huidobrensis. *J. Insect Physiol.* 53, 1199–1205. doi:10.1016/j.jinsphys.2007.06.011

Huang, S. Y., Kuo, Y. H., Lee, Y. P., Tsou, H. L., Lee, W. C., Ju, C. C., et al. (2000). Association of heat shock protein 70 with semen quality in boars. *Anim. Reprod. Sci.* 63, 231–240. doi:10.1016/s0378-4320(00)00175-5

Jha, K. N., Coleman, A. R., Wong, L., Salicioni, A. M., Howcroft, E., and Johnson, G. R. (2013). Heat shock protein 90 functions to stabilize and activate the testis-specific serine/threonine kinases, a family of kinases essential for male fertility. *J. Biol. Chem.* 288, 16308–16320. doi:10.1074/jbc.M112.400978

Jin, J., Zhao, M. T., Wang, Y., Zhou, Z. S., and Guo, J. Y. (2020). Induced thermotolerance and expression of three key hsp genes (Hsp70, Hsp21, and sHsp21) and their roles in the high temperature tolerance of *Agasicles hygrophila*. *Front. Physiol.* 10, 1593. doi:10.3389/fphys.2019.01593

King, A. M., and MacRae, T. H. (2015). Insect heat shock proteins during stress and diapause. *Annu. Rev. Entomol.* 60, 59–75. doi:10.1146/annurev-ento-011613-162107

Krebs, R., and Loeschcke, V. (1994). Costs and benefits of activation of the heat-shock response in *Drosophila melanogaster*. *Funct. Ecol.* 8, 730–737. doi:10.2307/2390232

Krone, P. H., Evans, T. G., and Blechinger, S. R. (2003). Heat shock gene expression and function during zebrafish embryogenesis. *Semin. Cell Dev. Biol.* 14, 267–274. doi:10.1016/j.semcdb.2003.09.018

Li, X. M., She, D. Y., Zhang, D. Y., and Liao, W. J. (2015). Life history trait differentiation and local adaptation in invasive populations of *Ambrosia artemisiifolia* in China. *Oecologia* 177, 669–677. doi:10.1007/s00442-014-3127-z

Li, Z. W., Li, X., Yu, Q. Y., Xiang, Z. H., Kishino, H., Zhang, Z., et al. (2009). The small heat shock protein (sHSP) genes in the silkworm, *Bombyx mori*, and comparative analysis with other insect sHSP genes. *BMC Evol. Biol.* 9, 215. doi:10.1186/1471-2148-9-215

Lindquist, S. L. (1985). The heat-shock response. *Annu. Rev. Biochem.* 55, 1151–1191. doi:10.1146/annurev.bi.55.070186.005443

Liu, W., Zhang, F. X., Cai, M. J., Zhao, W. L., Li, X. R., Wang, J.-X., et al. (2013). The hormone-dependent function of Hsp90 in the crosstalk between 20-hydroxyecdysone and juvenile hormone signaling pathways in insects is determined by differential phosphorylation and protein interactions. *Biochim. Biophys. Acta* 1830, 5184–5192. doi:10.1016/j.bbagen.2013.06.037

Lu, K., Chen, X., Liu, W. T., Zhang, Z. C., Wang, Y., You, K., et al. (2018). Characterization of heat shock protein 70 transcript from *Nilaparvata lugens* (Stal): Its response to temperature and insecticide stresses. *Pestic. Biochem. Physiol.* 142, 102–110. doi:10.1016/j.pestbp.2017.01.011

Ma, C., Cui, S., Bai, Q., Tian, Z., Zhang, Y., Chen, G., et al. (2020). Olfactory co-receptor is involved in host recognition and oviposition in *Ophraella communa* (Coleoptera: Chrysomelidae). *Insect Mol. Biol.* 29, 381–390. doi:10.1111/imb.12643

Ma, C., Cui, S., Tian, Z., Zhang, Y., Chen, G., Gao, X., et al. (2019a). OcomCSP12, a chemosensory protein expressed specifically by ovary, mediates reproduction in *Ophraella communa* (Coleoptera: Chrysomelidae). *Front. Physiol.* 10, 1290. doi:10.3389/fphys.2019.01290

Ma, C., Zhao, C., Cui, S., Zhang, Y., Zhou, Z., Chen, H., et al. (2019b). Identification of candidate chemosensory genes of *Ophraella communa* LeSage (Coleoptera: Chrysomelidae) based on antennal transcriptome analysis. *Sci. Rep.* 9, 15551. doi:10.1038/s41598-019-52149-x

Mack, P. D., Kapelnikov, A., Heifetz, Y., and Bender, M. (2006). Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 10358–10363. doi:10.1073/pnas.0604046103

Macrae, T. H. (2010). Gene expression, metabolic regulation and stress tolerance during diapause. *Cell. Mol. Life Sci.* 67, 2405–2424. doi:10.1007/s00018-010-0311-0

Marin, R., and Tanguay, R. (1996). Stage-specific localization of the small heat shock protein Hsp27 during oogenesis in *Drosophila melanogaster*. *Chromosoma* 105, 142–149. doi:10.1007/BF02509495

Mayer, M. P. (2013). Hsp70 chaperone dynamics and molecular mechanism. *Trends biochem. Sci.* 38, 507–514. doi:10.1016/j.tibs.2013.08.001

Morimoto, & R. (1993). Cells in stress: Transcriptional activation of heat shock genes. *Science* 259, 1409–1410. doi:10.1126/science.8451637

Neuer, A., Spandorfer, S. D., Giraldo, P., Dieterle, S., Rosenwaks, Z., and Witkin, S. S. (2000). The role of heat shock proteins in reproduction. *Hum. Reprod. Update* 6, 149–159. doi:10.1093/humupd/6.2.149

Nixon, B., Bromfield, E. G., Dun, M., Redgrove, K. A., Mclaughlin, E. A., and Aitken, R. J. (2015). The role of the molecular chaperone heat shock protein A2 (HSPA2) in regulating human sperm-egg recognition. *Asian J. Androl.* 17, 568–573. doi:10.4103/1008-682X.151395

Pockley, A. G., Fairburn, B., Mirza, S., Slack, L. K., Hopkinson, K., and Muthana, M. (2007). A non-receptor-mediated mechanism for internalization of molecular chaperones. *Methods*, 43, 238–244. doi: doi: doi:10.1016/j.ymeth.2007.06.007

Redgrove, K. A., Brett, N., Baker, M. A., Louise, H., Gordon, B., Liu, D. Y., et al. (2012). The molecular chaperone HSPA2 plays a key role in regulating the expression of sperm surface receptors that mediate sperm-egg recognition. *Plos One* 7, e50851. doi:10.1371/journal.pone.0050851

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3, 1101–1108. doi:10.1038/nprot.2008.73

Smith, M., Cecchi, L., Skjoth, C. A., Karrer, G., and Sikoparija, B. (2013). Common ragweed: A threat to environmental health in europe. *Environ. Int.* 61, 115–126. doi:10.1016/j.envint.2013.08.005

Smykal, V., and Raikhel, A. S. (2015). Nutritional control of insect reproduction. *Curr. Opin. Insect Sci.* 11, 31–38. doi:10.1016/j.cois.2015.08.003

Sørensen, J. G., Kristensen, T. N., and Loeschcke, V. (2003). The evolutionary and ecological role of heat shock proteins. *Ecol. Lett.* 6, 1025–1037. doi:10.1046/j.1461-0248.2003.00528.x

Sorte, C., and Hofmann, G. (2005). Thermotolerance and heat-shock protein expression in Northeastern Pacific Nucella species with different biogeographical ranges. *Mar. Biol.* 146, 985–993. doi:10.1007/s00227-004-1508-2

Sun, M., Lu, M. X., Tang, X. T., and Du, Y. Z. (2014). Characterization and expression of genes encoding three small heat shock proteins in sesamia inferens (Lepidoptera: Noctuidae). *Int. J. Mol. Sci.* 15, 23196–23211. doi:10.3390/ijms151223196

Thornhill, R. (1983). Cryptic female choice and its implications in the scorpionfly Harpobittacus nigriceps. *Am. Nat.* 122, 765–788. doi:10.1086/284170

Tian, Z., Chen, G., Zhang, Y., Ma, C., Tian, Z., Gao, X., et al. (2021). Rapid adaptive evolution of *Ophraella communa* in new low temperature environment. *J. Pest Sci*. doi:10.21203/rs.3.rs-490874/v1

Velazquez, J. M., and Lindquist, S. (1984). hsp70: Nuclear concentration during environmental stress and cytoplasmic storage during recovery. *Cell* 36, 655–662. doi:10.1016/0092-8674(84)90345-3

Von Wyschetzki, K., Rueppell, O., Oettler, J., and Heinze, J. (2015). Transcriptomic signatures mirror the lack of the fecundity/longevity trade-off in ant queens. *Mol. Biol. Evol.* 32, 3173–3185. doi:10.1093/molbev/msv186

Walsh, A., Whelan, D., Bielanowicz, A., Skinner, B., Aitken, R. J., O., Bryan, M. K., et al. (2008). Identification of the molecular chaperone, heat shock protein 1 (chaperonin 10), in the reproductive tract and in capacitating spermatozoa in the male mouse. *Biol. Reprod.* 78, 983–993. doi:10.1095/biolreprod.107.066860

Wan, F., Jiang, M., and Zhan, A. (2017). "Biological invasions and its management in China," in *Common ragweed* Ambrosia artemisiifolia *L.* Editors Z. S. Zhou, F. H. Wan, and J. Y. Guo (Springer Nature Singapore), 99–109.

Xie, J., Hu, X. X., Zhai, M. F., Yu, X. J., Song, X. W., Gao, S. S., et al. (2019). Characterization and functional analysis of hsp18.3 gene in the red flour beetle, *Tribolium castaneum*. *Insect Sci.* 026, 263–273. doi:10.1111/1744-7917.12543

Xu, J., Baulding, J., and Palli, S. R. (2013). Proteomics of *Tribolium castaneum* seminal fluid proteins: Identification of an angiotensin-converting enzyme as a key player in regulation of reproduction. *J. Proteomics* 78, 83–93. doi:10.1016/j.jprot.2012.11.011

Yang, X. Q., Zhang, Y. L., Wang, X. Q., Dong, H., Gao, P., and Jia, L. Y. (2016). Characterization of multiple heat-shock protein transcripts from Cydia pomonella: Their response to extreme temperature and insecticide exposure. *J. Agric. Food Chem.* 64, 4288–4298. doi:10.1021/acs.jafc.6b01914

Zatsepina, O. G., Evgen'ev, M. B., and Garbuz, D. G. (2021). Role of a heat shock transcription factor and the major heat shock protein Hsp70 in memory formation and neuroprotection. *Cells* 10, 1638. doi:10.3390/cells10071638

Zhang, Y., Chen, J., Chen, G., Ma, C., Chen, H., Gao, X., et al. (2020). Identification and validation of reference genes for quantitative gene expression analysis in *Ophraella communa*. *Front. Physiol.* 11, 355. doi:10.3389/fphys.2020.00355

Zhang, Y., Dong, Z., Gu, P., Zhang, W., Wang, D., Guo, X., et al. (2014). Proteomics analysis of adult testis from B ombyx mori. *Proteomics* 14, 2345–2349. doi:10.1002/pmic.201300507

Zhang, Y., Zhao, C., Ma, W., Cui, S., Chen, H., Ma, C., et al. (2021). Larger males facilitate population expansion in *Ophraella communa*. *J. Anim. Ecol.* 90, 2782–2792. doi:10.1111/1365-2656.13579

Zhou, Z. S., Guo, J. Y., Chen, H. S., and Wan, F. H. (2010). Effects of temperature on survival, development, longevity, and fecundity of *Ophraella communa* (Coleoptera: Chrysomelidae), a potential biological control agent against *Ambrosia artemisiifolia* (Asterales: Asteraceae). *Environ. Entomol.* 39, 1021–1027. doi:10.1603/EN09176

ZhouGuo, Z-S. J-Y., Ai, H-M., Li, M., Wan, F. H., and Wan, F-H. (2011). Rapid cold-hardening response in *Ophraella communa* LeSage (Coleoptera: Chrysomelidae), a biological control agent of *Ambrosia artemisiifolia* L. *Biocontrol Sci. Technol.* 21, 215–224. doi:10.1080/09583157.2010.534549

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 21 September 2022
DOI 10.3389/fmolb.2022.854250

OPEN ACCESS

EDITED BY
Gian Gaetano Tartaglia,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Arun Samidurai,
Virginia Commonwealth University, United States
Yongchang Chen,
Kunming University of Science and Technology, China

*CORRESPONDENCE
Shengguo Zhao,
zhaosg@gsau.edu.cn

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 13 January 2022
ACCEPTED 30 August 2022
PUBLISHED 21 September 2022

CITATION
Yang Y, Li Y, Yuan H, Liu X, Ren Y, Gao C, Jiao T, Cai Y and Zhao S (2022), Characterization of circRNA–miRNA–mRNA networks regulating oxygen utilization in type II alveolar epithelial cells of Tibetan pigs. *Front. Mol. Biosci.* 9:854250. doi: 10.3389/fmolb.2022.854250

Characterization of circRNA–miRNA–mRNA networks regulating oxygen utilization in type II alveolar epithelial cells of Tibetan pigs

Yanan Yang[1], Yongqing Li[2], Haonan Yuan[1], Xuanbo Liu[1], Yue Ren[3], Caixia Gao[4], Ting Jiao[1,5], Yuan Cai[1] and Shengguo Zhao[1]*

[1]College of Animal Science and Technology, Gansu Agricultural University, Lanzhou, China, [2]Xinjiang Academy of Animal Sciences, Ürümqi, Xinjiang, China, [3]Academy of Agriculture and Animal Husbandry Sciences, Institute of Animal Husbandry and Veterinary Medicine, Lhasa, China, [4]State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, China, [5]College of Grassland Science, Gansu Agricultural University, Lanzhou, China

Understanding the signaling pathway regulatory mechanisms in type II alveolar epithelial (ATII) cells, the progenitor cells responsible for proliferating and regenerating type I alveolar epithelial (ATI) and ATII cells, in Tibetan pigs is beneficial for exploring methods of preventing and repairing cellular damage during hypoxia. We simulated a hypoxic environment (2% O_2) for culture ATII cells of Tibetan pigs and Landrace pigs, with cells cultured under normoxic conditions (21% O_2) as a control group, and performed integrated analysis of circular RNA (circRNA)–microRNA (miRNA)–messenger RNA (mRNA) regulatory axes by whole-transcriptome sequencing. Functional enrichment analysis indicated that the source genes of the differential expressed circRNAs (DEcircRNAs) were primarily involved in cell proliferation, cellular processes, and cell killing. A series of DEcircRNAs were derived from inhibitors of apoptosis proteins and led to a key autonomous effect as modulators of cell repair in Tibetan pigs under hypoxia. The significant higher expression of *COL5A1* in TL groups may inhibited apoptosis of ATII cells in Tibetan pigs under lower oxygen concentration, and may lead their better survive in the hypoxia environment. In addition, a competing endogenous RNA (ceRNA) network of functional interactions was constructed that included novel_circ_000898-ssc-miR-199a-5p-*CAV1* and novel_circ_000898-ssc-miR-378-*BMP2*, based on the node genes ssc-miR-199a-5p and ssc-miR-378, which may regulate multiple miRNAs and mRNAs that mediate endoplasmic reticulum (ER) stress-induced apoptosis and inflammation and attenuate hypoxia-induced injury in ATII cells under hypoxic conditions. These results broaden our

Abbreviations: DEcircRNAs, Differentially expressed circRNAs; DEmiRNAs, Differentially expressed miRNAs; DEmRNAs, Differentially expressed mRNAs; TN, ATII cells of Tibetan pigs were cultured under 21% O_2; TL, ATII cells of Tibetan pigs were cultured under 2% O_2; LN, ATII cells of Landrace pigs were cultured under 21% O_2; LL, ATII cells of Landrace pigs were cultured under 2% O_2.

knowledge of circRNAs, miRNAs, and mRNAs associated with hypoxia and provide new insights into the hypoxic response of ATII cells in Tibetan pigs.

KEYWORDS

noncoding RNA, ATII cells, ceRNA, Tibetan pigs, cellular processes

Introduction

The Tibetan pig is a unique domestic highland breed in China and has well adapted to the hypoxic environment than other pigs, as the native breed live at Qinghai-Tibet Plateau; indeed, the nucleotide diversity in most regions of the mitogenome is greater in wild Tibetan pigs than in domestic pigs (Li et al., 2013; Li et al., 2016; Wu et al., 2020). Specific characteristics of Tibetan pigs, including their well-developed hearts and lungs, denser pulmonary artery networks, and miRNA–mRNA coexpression regulatory networks in the lungs, facilitate more effective oxygen transport in hypoxic environments (Yang et al., 2021a; Yang et al., 2021b). The lung is the primary respiratory organ that exchanges oxygen and has a large and vascularized epithelial surface area, which is susceptible to hypoxia-induced injury as a "rate-limiting" organ under severely hypoxic conditions (Gallagher and Hackett, 2004; Herriges and Morrisey, 2014). ATI cells and ATII cells were covered the surface of alveolar, and gas exchange relies on the integrity of the epithelium and its dynamic interaction with other cells (Aspal and Zemans, 2020). The hypoxic response of cells depends on regulatory networks at the transcriptional and post-transcriptional levels that result in variations in gene expression (Jozefczuk et al., 2010; Richter et al., 2010). Therefore, cells must be able to sense, respond, and adapt rapidly to short- or long-term hypoxic conditions for optimum survival. Cell turnover is slower in the lungs than in other organs, such as the skin and intestine (Sender and Milo, 2021), and hypoxia (Greco et al., 2019) or hypoxia-induced (Qin et al., 2018) acute lung diseases or injury, including pulmonary fibrosis (Darby and Hewitson, 2016). ATII cells functioning as progenitor cells in the alveoli can regenerate and proliferate into ATII cells or undergo repair and differentiate into ATI cells after physiological insult under hypoxia. The existing literature shows that hepatocyte growth factor (Ito et al., 2014), keratinocyte growth factor (Schmeckebier et al., 2013), and the Wnt/β-catenin (Aspal and Zemans, 2020) signaling pathway can promote the proliferation of ATII cells. In addition, the levels of genes and proteins involved in HIF-related pathways and inflammation activation were found to be significantly increased in ATII cells under hypoxic conditions (Grek et al., 2011; Sherman et al., 2018). Oxidative stress may be the primary reason for ATII cell apoptosis, which could undergo cell death and replaced by myofibroblasts in hypoxia-induced IPF to prevent repair and renewal of the alveolar wall (Dias-Freitas et al., 2016; Alvarez-Palomo et al., 2020). To date, an approach utilizing integrated analysis of protein-coding messenger RNAs (mRNAs) and many noncoding RNAs, i.e., circular RNAs (circRNAs) and microRNAs (miRNAs), to yield phase-specific gene expression or regulation patterns has been effective. Previous studies have revealed that RNA regulation mediates diverse biological processes, such as vascular remodeling (Fujiwara et al., 2021), innate immunity (Colgan et al., 2020), brain activity (Butt et al., 2021), and inflammation (Colgan et al., 2016), in plateau animals. Hence, insight into the underlying RNA regulatory mechanism in ATII cells of Tibetan pigs is of great significance for understanding hypoxic adaptation mechanisms. Here, we assumed that the expressions and functions of hypoxia-related genes are partially or entirely regulated by miRNAs or circRNAs. To validate this hypothesis, we constructed a competing endogenous RNA (ceRNA) network for ATII cell under normoxic (21% O_2) and hypoxic (2% O_2) conditions, which will allow identification of the changes in RNA regulation that occur in response to hypoxia in Tibetan pigs and Landrace pigs.

Methods

Sample collection

Lung tissue samples collected from healthy newborn male piglets (Tibetan pigs and Landrace pigs) at 7 days of age were soaked in PBS. Primary ATII cells were isolated as described previously (Wang et al., 2016) with minor modifications. Then, ATII cells were cultured in complete medium at 37°C in an environment containing either 21% O_2, 5% CO_2, and 79% N_2 (normoxic conditions) or 2% O_2, 5% CO_2, and 98% N_2 (hypoxic conditions).

We harvested ATII cells (n = 6 for each group) at 48 h under the 21% O_2 (Tibetan–normoxic (TN) and Landrace–normoxic (LN)) or 2% O_2 (Tibetan–low hypoxic (TL) and Landrace–hypoxic (LL)) conditions. Three of each group were flash-frozen in liquid nitrogen for RNA extraction, and the rest were used for analyze cell apoptosis by A BD FACSCanto II flow cytometer (BD Biosciences, San Jose, CA, USA).

RNA isolation, library preparation, and sequencing

Total RNA was extracted with a TRIzol reagent kit (Invitrogen, Carlsbad, CA, USA), and the RNA integrity

number (RIN) was confirmed to be >7.5. mRNA and noncoding RNA (ncRNA) were retained by removing ribosomal RNA (rRNA), and short fragments were reverse transcribed into complementary DNA (cDNA) with random primers. Second-strand cDNA was synthesized and purified with a QiaQuick PCR Extraction Kit (Qiagen, Venlo, Netherlands). Twelve RNA-seq libraries were generated and sequenced on the Illumina HiSeq™ 4,000 sequencing platform (Illumina Inc., San Diego, CA, USA) by Gene *Denovo* Biotechnology Co. (Guangzhou, China).

Read mapping and transcript assembly

Clean reads were obtained by filtering adapter reads and poor-quality reads from the raw data using fastp (Chen et al., 2018). HISAT 2 was used to map the clean paired-end reads to the *Sus scrofa* RefSeq genome (*Sus scrofa* 11.1). StringTie (Pertea et al., 2015) was used to reconstruct transcripts and calculate expression abundances and variations as fragments per kilobase of transcript per million mapped reads (FPKM) values.

Identification of known and novel miRNAs and circRNAs

Clean reads were obtained and mapped to the miRBase 21.0 database (http://www.mirbase.org/) to identify known porcine miRNAs. Other known miRNAs in other species were identified and mapped to the remaining miRNA sequences. Novel miRNA candidates were predicted by aligning the reference genome with unannotated tags according to hairpin structures and genomic positions using miRDeep2.

Clean reads were obtained and mapped to the porcine reference genome by discarding low-quality reads, and the retained reads were analyzed with find_circ to identify circRNAs. Clean reads that could not be annotated were defined as novel circRNAs.

Quantification of RNA abundance and evaluation of differential expression

The expression of the identified circRNAs, miRNAs, and mRNAs was quantified by calculating the reads per million mapped reads (RPM), transcripts per million (TPM), and FPKM values, respectively. Differentially expressed miRNAs (DEmiRNAs) and differentially expressed circRNAs (DEcircRNAs) were identified as those with a fold change (FC) ≥2 and p value < 0.05 by edgeR, and differentially expressed mRNAs (DEmRNAs) were identified as those with a false discovery rate (FDR) of less than 0.05 and an absolute FC ≥ 2 by DESeq.

Functional annotation and enrichment analysis of RNAs

Source genes of DEcircRNAs, candidate target genes of DEmiRNAs, and DEmRNAs were subjected to Gene Ontology (GO) term enrichment analysis (http://www.geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis (http://www.genome.jp/kegg/). FDR ≤0.05 was established as the threshold for determining significance.

Construction of the circRNA–miRNA–mRNA network

RNA pairs with Spearman rank correlation coefficients (SCCs) < −0.7 were selected as negatively coexpressed mRNA–miRNA or circRNA–miRNA pairs, and RNA pairs with Pearson correlation coefficients (PCCs) > 0.9 were selected as coexpressed circRNA–mRNA pairs. p values < 0.05 were computed to assess shared miRNA sponges between circRNAs and mRNAs as final ceRNA pairs. The circRNA–miRNA–mRNA ceRNA regulatory network was constructed and visualized using Cytoscape. We considered significant nodes at the cores of the regulated networks to be highly associated with hypoxic adaptation.

Quantitative real-time–PCR validation

To verify the RNA-seq results, we randomly selected mRNAs, miRNAs, and circRNAs to verify the data reliability. The primers used for qRT–PCR were synthesized by Qingke Biological Company (Xi'an, China) and are shown (Supplementary Table S1 in Supplementary Material S2). cDNA was synthesized from RNA samples identified by transcriptome sequencing and used as the template for gene expression analysis. qPCR was performed in a LightCycler 96 Real-Time System (Roche, Switzerland) using SYBR® Premix Ex Taq™ II (TaKaRa, China). Statistical analyses were performed with analysis of variance (ANOVA) using SPSS 20.0 (SPSS, IL, USA). p < 0.05 was considered to indicate a significant difference. Graphs were generated using GraphPad Prism 8.0 (GraphPad Software, CA, USA).

Dual-luciferase reporter assay

The putative miR-141 binding sites on the target gene *HOOK3* were predicted using Target Scan (http://www.Targetscan.org), miRBase (http://www.mirbase.org), and

TABLE 1 Apoptosis rate of ATII cells under hypoxia.

Groups	Late apoptosis rate (%)	Early apoptosis rate (%)	Total rate of apoptosis (%)	Normal viable cell rate (%)
LN	19.78	4.45	24.23	72.95
LL	29.98	6.89	36.87	60.68
TN	7.29	6.58	13.87	84.63
TL	15.40	16.50	31.90	66.47

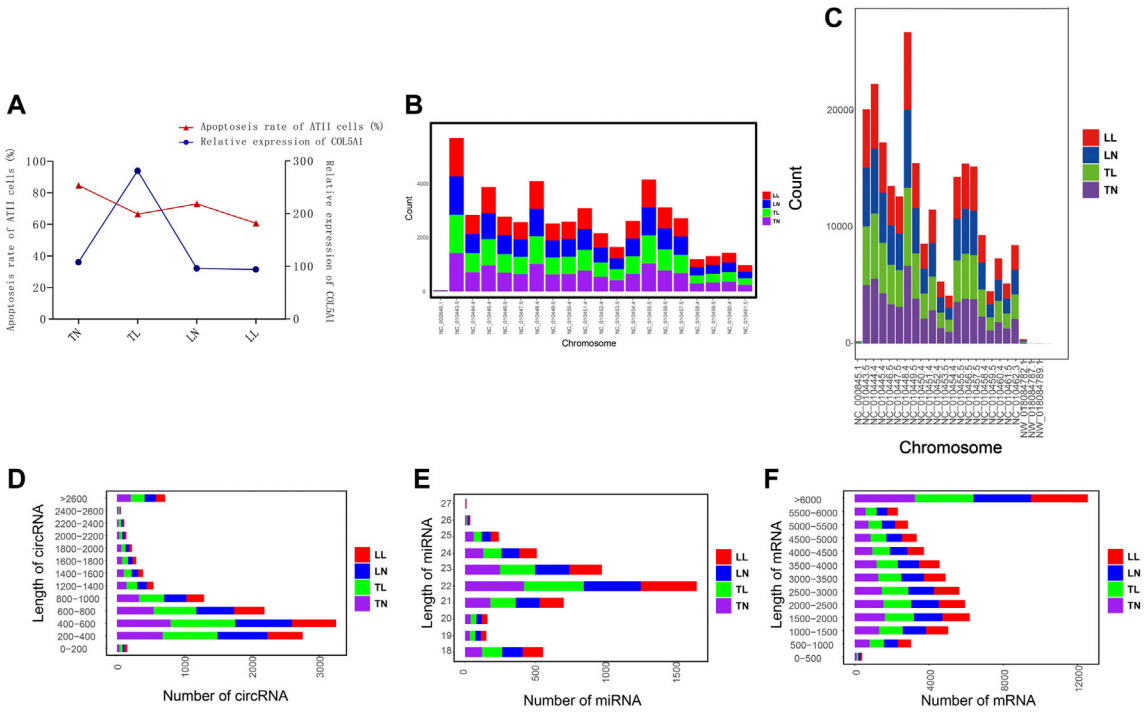


FIGURE 1
General features of circRNAs, miRNAs, and mRNAs. **(A)** Correlation analysis of COL5A1 and cell apoptosis. **(B)** Chromosome distribution map of circRNAs. **(C)** Chromosome distribution map of mRNAs. **(D–F)** Length distributions for the identified circRNAs, miRNAs, and mRNAs.

miRDB (http://www.mirdb.org). The wild-type (WT) and mutant-type (MUT) *HOOK3* containing the putative binding site of miR-141 were synthesized by GENEWIZ Biotechnology Co., Ltd. and purchased from GENEWIZ (Suzhou, China) andcloned into the pmirGLO Dual-Luciferase vector XhoI/SalI sites (Promega, United States). The mimic negative control (NC) were bought by Jima (Shanghai, China). HOOK3-MT or HOOK3-MUT and miR-141 mimic or mimic NC were co-transfected into 293 T cells using Lipofectamine 2000 transfection vehicle (Invitrogen, Carlsbad, CA, United States) for the luciferase reporter assay. After 48 h, luciferase activity were examined by the Dual-Luciferase Reporter Assay System (Promega). The ratios of firefly luciferase to renilla luciferase were identified as relative luciferase enzyme activity.

Results

Correlation analysis of *COL5A1* and cell apoptosis

Cell apoptosis was investigated by flow cytometric assays, and our results showed that cell apoptosis was higher in the hypoxic groups (TL, LL) than that in the normoxic groups (TN, LN), and the rate of normal viable cell rate were 84.63, 66.47, 72.95, and 60.68% in TN, TL, LN, LL groups, respectively (Table 1). Notably, the expression of *COL5A1* in TL were significant higher than that of TN groups, and not significant different between LN and LL groups, as the source genes of numerous circRNAs, including novel_circ_001282, novel_circ_001569, and novel_circ_003568 (Figure 1A).

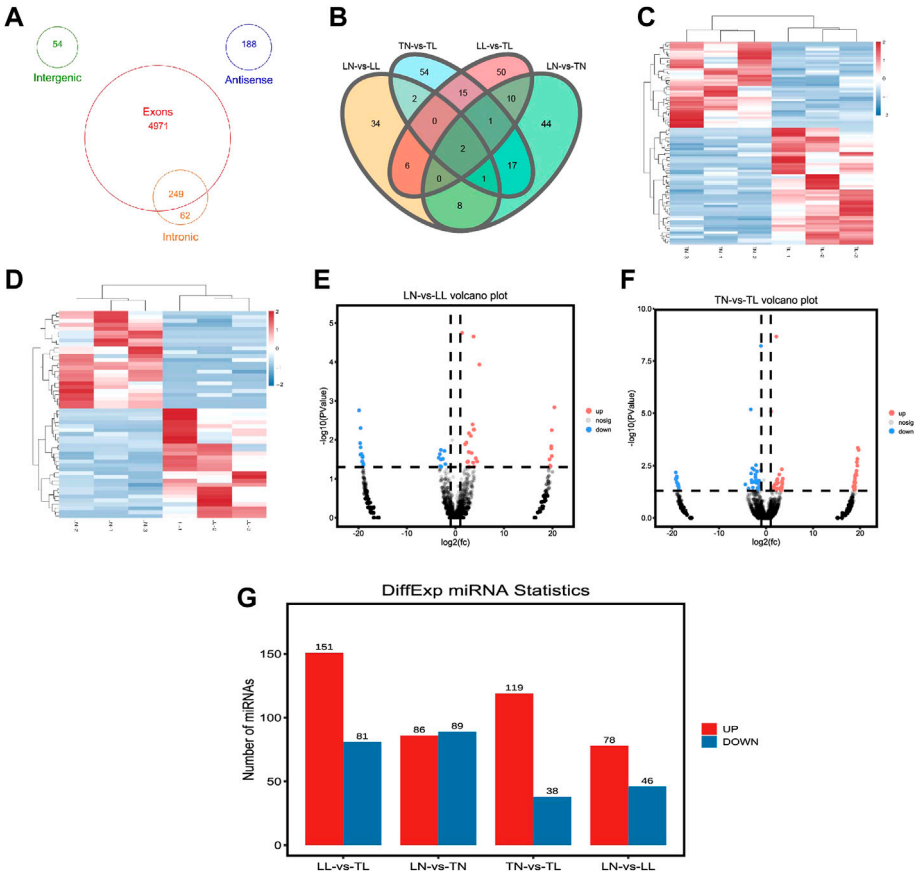


FIGURE 2
Summary of differential expression analysis of the circRNAs and miRNAs. **(A)** Venn diagram indicating that circRNAs numbers originated from different genomic sites. **(B)** Venn diagram of circRNAs interactions based on the overlapping circRNAs among the four groups. **(C)** Heatmap showing the relative expression patterns of DEcircRNAs between TN and TL. **(D)** Heatmap showing the relative expression patterns of DEcircRNAs between LN and LL. **(E)** Volcano plot showing the relative expression patterns of DEcircRNAs between LN and LL. **(F)** Volcano plot showing the relative expression patterns of DEcircRNAs between TN and TL. **(G)** Histogram showing the number of DEmiRNAs identified among the four groups.

Identification of DEcircRNAs, DEmiRNAs, and DEmRNAs

Whole-transcriptome profiling of ncRNAs (i.e., circRNAs), mRNAs, and miRNAs was performed to assess the regulation of hypoxia-related genes in ATII cells (Supplementary Figure S1 in Supplementary Material S1, Supplementary Tables S2, S3 in Supplementary Material S2). Averages of 74.97 and 11.89 million clean reads obtained for circRNAs (mRNAs) and miRNAs, respectively, after removal of redundant and low-quality reads. mRNAs and circRNAs were identified on all chromosomes; most mRNAs were located on chromosomes NC_010448.4, NC_010444.4, and NC_010443.5, and most circRNAs were located on chromosomes NC_010443.5, NC_010455.5, and NC_010448.4 (Figures 1B,C). In total, 5,524 circRNAs were identified in this study, and these were named and numbered novel_circ_000001 to novel_circ_005524. Furthermore, the vast majority of these circRNAs were shown to have originated from exonic circRNAs, and intronic, intergenic, antisense overlapping and sense overlapping circRNAs were the mainly kind of circRNAs identified (Figure 2A). Most of the circRNAs, miRNAs, and mRNAs had lengths of 200–1,000 bp, 21–24 bp, and more than 1,000 bp, respectively; 400–600 bp, 22–23 bp, and more than 6,000 bp were the most abundant respective lengths and were consistent with the RNA features (Figures 1D–F).

In total, 5,524 circRNAs, 1,332 miRNAs, and 20,720 mRNAs were obtained. A total of 92 (53 upregulated and 39 downregulated), 53 (28 upregulated and 25 downregulated), 83 (47 upregulated and 36 downregulated), and 84 (49 upregulated and 35 downregulated) DEcircRNAs were identified in ATII cells between the TN and TL, LN and LL, TN and LN, and TL and LL groups, respectively (Table 1; Figures 2B–F). Analysis identified 340 existing miRNAs, 675 known miRNAs, and 318 novel miRNAs (Supplementary Material S3). In addition,

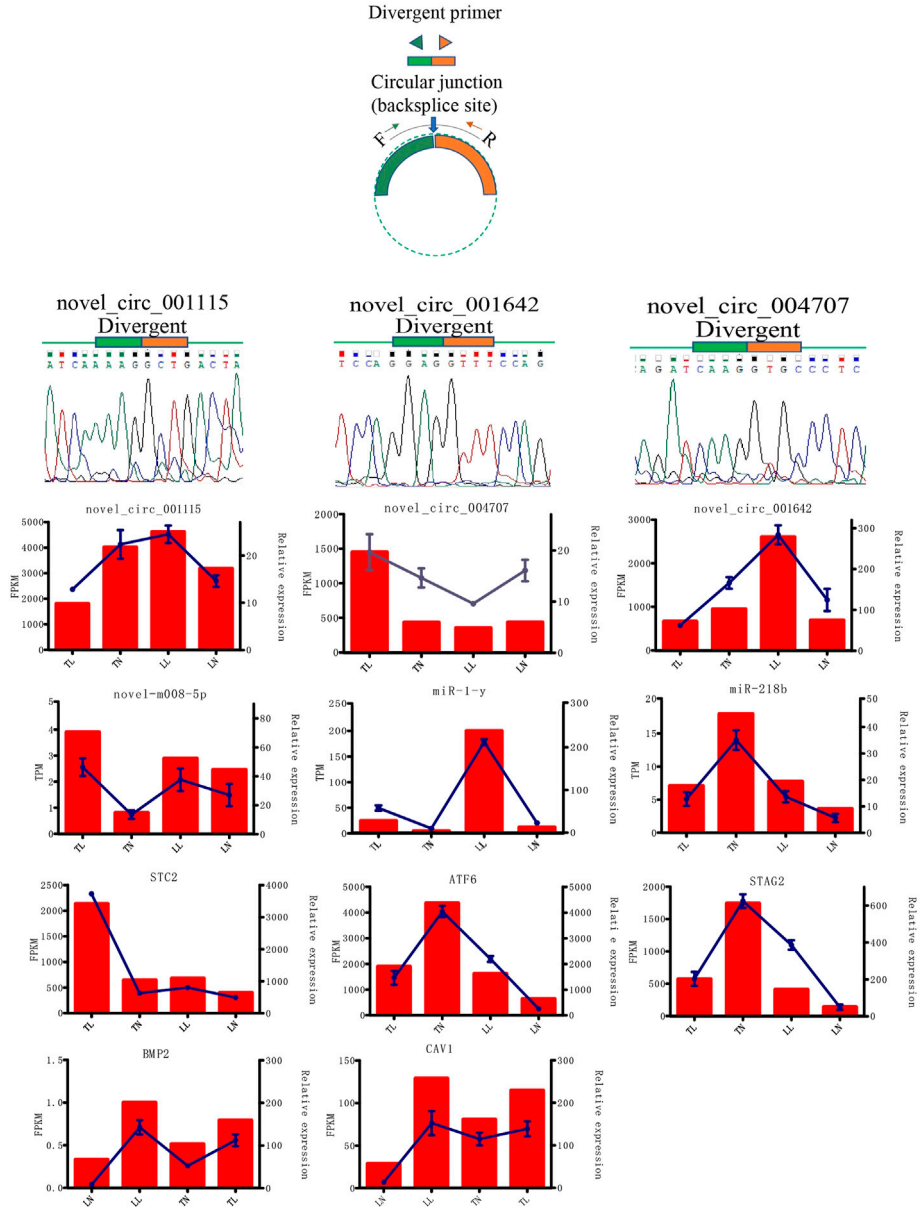


FIGURE 3
Expression patterns of randomly selected 3 DEcircRNAs, 3 DEmiRNAs and 3 DEmRNAs.

5 DEcircRNAs, 37 DEmiRNAs, and 1,470 DEmRNAs were identified between cells cultured under different oxygen conditions (21% O_2 and 2% O_2) (Supplementary Material S4). The 3 DEcircRNAs, 3 DEmiRNAs, and 3 DEmRNAs randomly selected and detected by qPCR analysis and Sanger sequencing to validate the accuracy of the sequencing data. Result suggest that the sequencing data and expression of circRNAs, miRNAs, and mRNAs identified in this study are reliable (Figure 3, Supplementary Figure S2 in Supplementary Material S1).

Identification of miRNA–mRNA and circRNA–miRNA pairs

A total of 157 and 124 hypoxia-dependent DEmiRNAs were identified in the Tibetan pig (TN and TL) groups and Landrace pig (LN and LL) groups (Figure 2G), respectively. Furthermore, we predicted the target circRNAs and mRNAs of these DEmiRNAs and identified 109 (25) circRNA–miRNA and 25,434 (18,506) mRNA-miRNA pairs between the TN (LN) and TL (LL) groups (Supplementary Material S5).

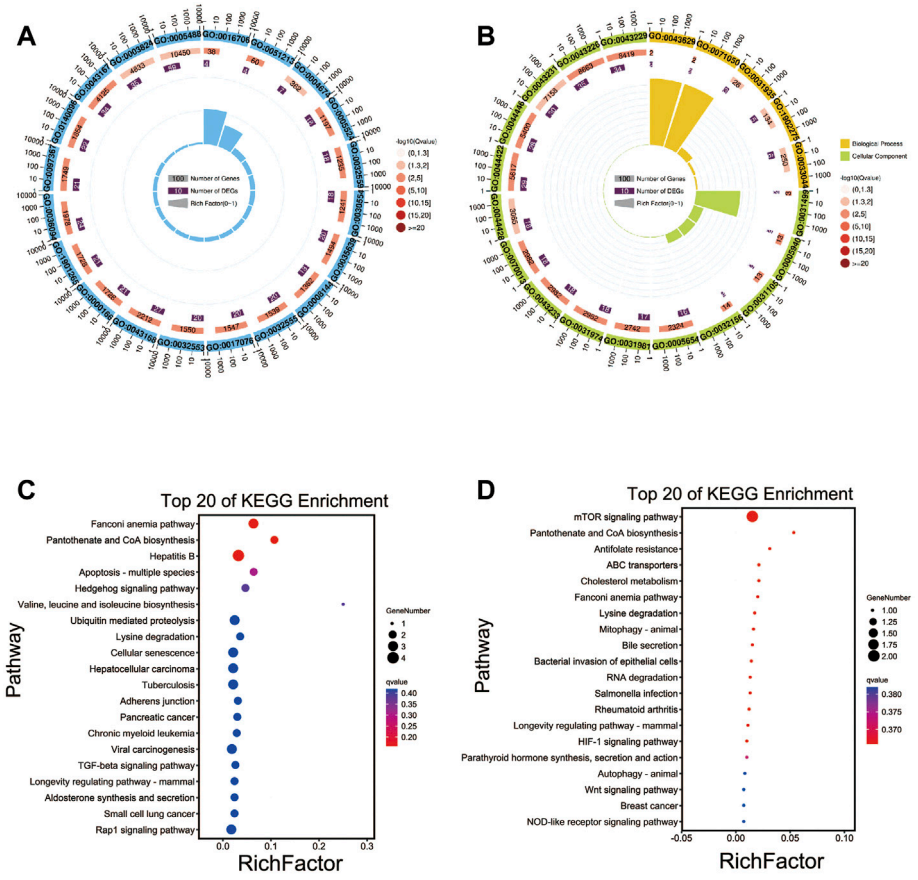


FIGURE 4
Functional annotation analysis of the source genes of DEcircRNAs. **(A)** GO annotation of the source genes of DEcircRNAs between TN and TL groups. **(B)** GO annotation of the source genes of DEcircRNAs between LN and LL groups. **(C)** KEGG enrichment of the source genes of DEcircRNAs between TN and TL groups. **(D)** KEGG enrichment of the source genes of DEcircRNAs between LN and LL groups.

GO and KEGG enrichment analyses of DEcircRNA source genes

We found that the majority of DEcircRNAs among the four groups were derived from source genes and that some were derived from intergenic regions. We also performed GO functional analysis and KEGG pathway enrichment analysis of the circRNA source genes to evaluate their biological roles. In total, 41, 38, 45, and 47 GO terms for the TN vs. TL, LN vs. LL, LN vs. TN, and LL vs. TL comparisons, respectively, were enriched in biological process, cellular component, and molecular function categories (Supplementary Figure S3A in Supplementary Material S1). Moreover, anion binding (GO: 0043168), small molecule binding (GO:0036094), and ATP binding (GO:0005524) were the most significant terms between the TN and TL groups (Figure 4A). ncRNA polyadenylation (GO:0043629), snoRNA polyadenylation (GO: 0071050), and TRAMP complex (GO:0031499) were the most significant terms between the LN and LL groups (Figure 4B).

KEGG enrichment analysis revealed that several pivotal pathways, including protein processing in the endoplasmic reticulum (ER), focal adhesion, and adherens junction, were associated with the regulatory mechanisms of the circRNAs (Supplementary Figure S3B in Supplementary Material S1). The source genes of the DEcircRNAs in the TN group compared to the TL group were significantly enriched in the Fanconi anemia pathway, pantothenate and CoA biosynthesis, and hepatitis B pathways (Figure 4C). The source genes of the DEcircRNAs in the LN group compared to the LL group were significantly enriched in the mTOR signaling pathway and pantothenate and CoA biosynthesis pathways (Figure 4D).

Enrichment analysis of target genes of the DEmiRNAs in the network

We performed GO enrichment analysis to investigate the biological roles of target genes of DEmiRNAs and found that

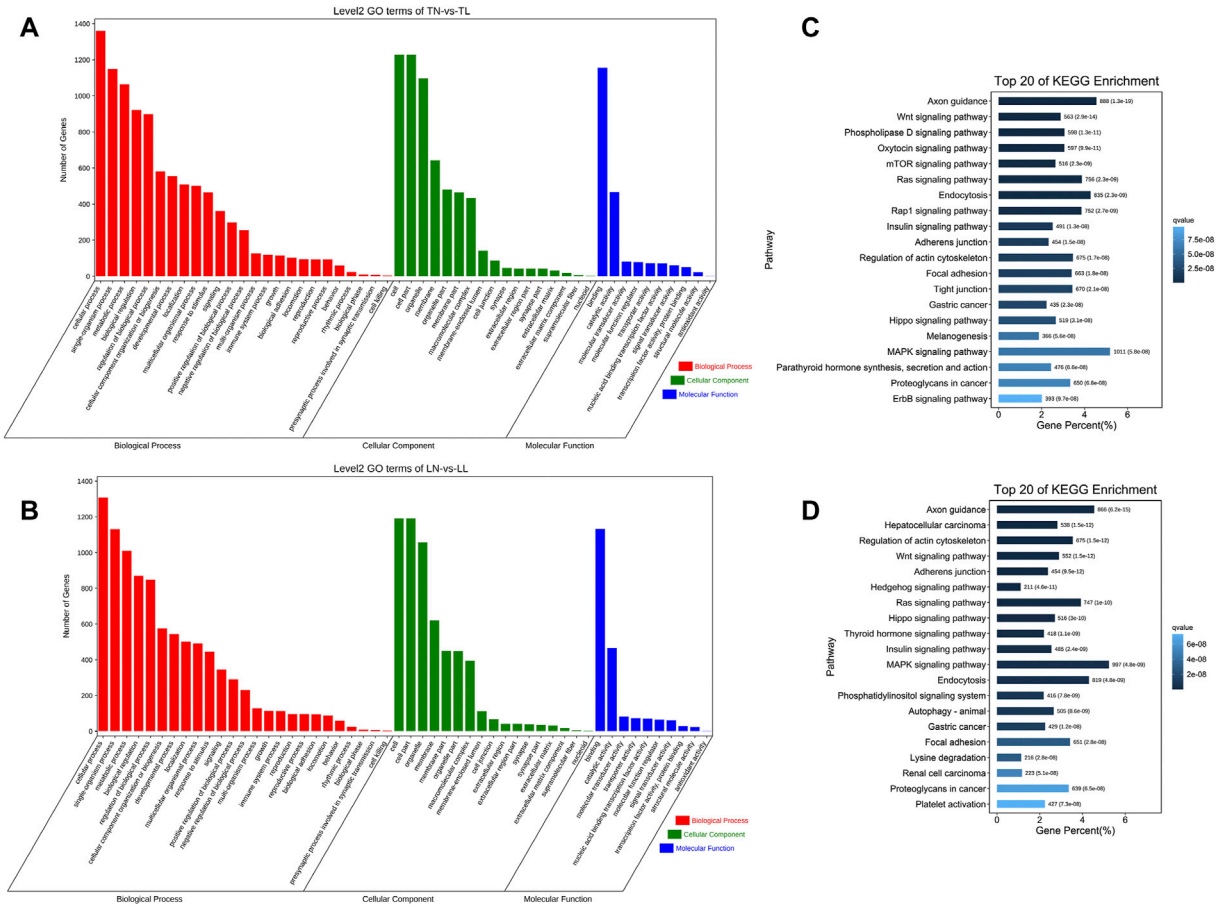


FIGURE 5
Functional annotation analysis of the target genes of DEmiRNAs. **(A)** GO annotation of the target genes of DEmiRNAs between TN and TL groups. **(B)** GO annotation of the target genes of DEmiRNAs between LN and LL groups. **(C)** KEGG enrichment of the target genes of DEmiRNAs between TN and TL groups. **(D)** KEGG enrichment of the target genes of DEmiRNAs between LN and LL groups.

these genes were significantly enriched in the terms binding, intracellular, and intracellular part under hypoxic conditions. These terms were identified by comparison of the normoxia (21% O_2) and hypoxia (2% O_2) groups and represent the common effect of oxygen concentration changes on target genes of the DEmiRNAs (Supplementary Material S6). Most target genes of the DEmiRNAs between the TN and TL groups were significantly enriched in the cell term in the cellular component category, the binding term in the molecular function category, and the cellular component organization term in the biological process category (Figure 5A). Numerous target genes of the DEmiRNAs between the LN and LL groups were significantly enriched in the cellular component organization, nucleoside binding, and membrane-bounded organelle terms in the biological process, molecular function, and cellular component categories, respectively (Figure 5B).

As demonstrated by the KEGG pathway enrichment analysis, target genes of the DEmiRNAs between Tibetan pigs and Landrace pigs at different oxygen concentrations (2% O_2 and 21% O_2) were significantly enriched in the following pathways: axon guidance, focal adhesion, and MAPK signaling pathway (Supplementary Material S6). DEmiRNAs in Tibetan pigs at different oxygen concentrations were significantly enriched in the following pathways: axon guidance, wnt signaling pathway, and phospholipase D signaling pathway (Figure 5C). DEmiRNAs in Landrace pigs at different oxygen concentrations were significantly enriched in the following pathways: axon guidance, hepatocellular carcinoma, and regulation of actin cytoskeleton (Figure 5D).

Establishment of the circRNA–miRNA–mRNA network

The intersection of circRNA–miRNA pairs and mRNA–miRNA pairs was filtered, and pairs based on the

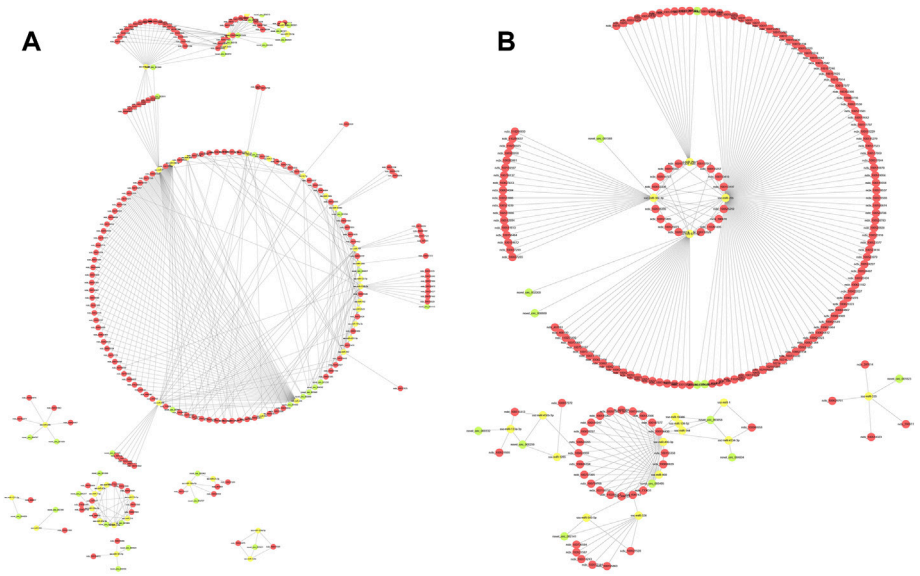


FIGURE 6
The circRNA-miRNA-mRNA interaction network. **(A)** The circRNA-miRNA-mRNA interaction network between TN and TL. **(B)** The circRNA-miRNA-mRNA interaction network between LN and LL. Green circles represent circRNAs, yellow circles represent miRNAs, and red circles represent mRNAs.

DEmiRNAs were identified in the TN vs. TL and LN vs. LL comparisons (Supplementary Material S7). To further reveal the potential regulatory networks and their biological processes in ATII cells at different oxygen concentrations, we combined the miRNA–mRNA and circRNA–miRNA pairs to construct two preliminary circRNA–miRNA–mRNA networks in Tibetan pigs and Landrace pigs. The network in Tibetan pigs presented an initial view of the associations among the 53 circRNA nodes, 44 miRNA nodes, 1,375 mRNA nodes, and 428 edges and included ssc-let-7g, ssc-miR-1, ssc-miR-10382, novel_circ_000083, novel_circ_000136, novel_circ_000480, *ENPP1*, *EPAS1*, *ITGB8*, etc. Ssc-miR-1, ssc-miR-490–3p, ssc-miR-490, and novel_circ_004392 were considered the most notable RNAs between the TN and TL groups (Figure 6A).

The network in Landrace pigs (LN vs. LL) was composed of 14 circRNA nodes (novel_circ_000256, novel_circ_000332, novel_circ_000880, etc.), 19 miRNA nodes (ssc-miR-1, ssc-miR-326, ssc-miR-335, etc.), 827 mRNA nodes (*EPAS1*, *NREP*, *P4HA1*, etc.) and 244 edges. Interestingly, ssc-miR-30c-3p, ssc-miR-20b, ssc-miR-671–5p, ssc-miR-29a-5p, and novel_circ_003405 were identified as the most important and central RNAs between the LN and LL groups (Figure 6B).

Core regulatory networks of the DEcircRNAs, DEmiRNAs, and DEmRNAs

We combined the circRNA–miRNA and miRNA–mRNA pairs among the four groups and constructed a preliminary circRNA–miRNA–mRNA network. The top 50 relationship pairs are shown in the network diagrams. The network was composed of 49 circRNA nodes (novel_circ_001145, novel_circ_001390, novel_circ_001681, etc.), 5 miRNA nodes (ssc-miR-145–5p, ssc-miR-1, ssc-miR-378, etc.), 6 mRNA nodes (*MAPK1*, *IRS1*, *CAV1*, etc.) and 170 edges. Several miRNAs, such as ssc-miR-199a-5p, ssc-miR-378, and ssc-miR-1, were identified as the most important nodes, and related circRNAs (novel_circ_000857, novel_circ_001835, and novel_circ_001145) and mRNAs (ncbi_404,693, ncbi_100153927, NewGene.10854.1, and ncbi_100157103) were identified by miRanda (Figure 7).

Validation of targeting relationships between miR-141 and *HOOK3*

Based on the inverse expression trends between miRNA and mRNAs in type II alveolar epithelial cells under different altitude. We randomly selected miR-141 to verify it targeting relationships with *HOOK3*. The dual luciferase reporter assay indicated that

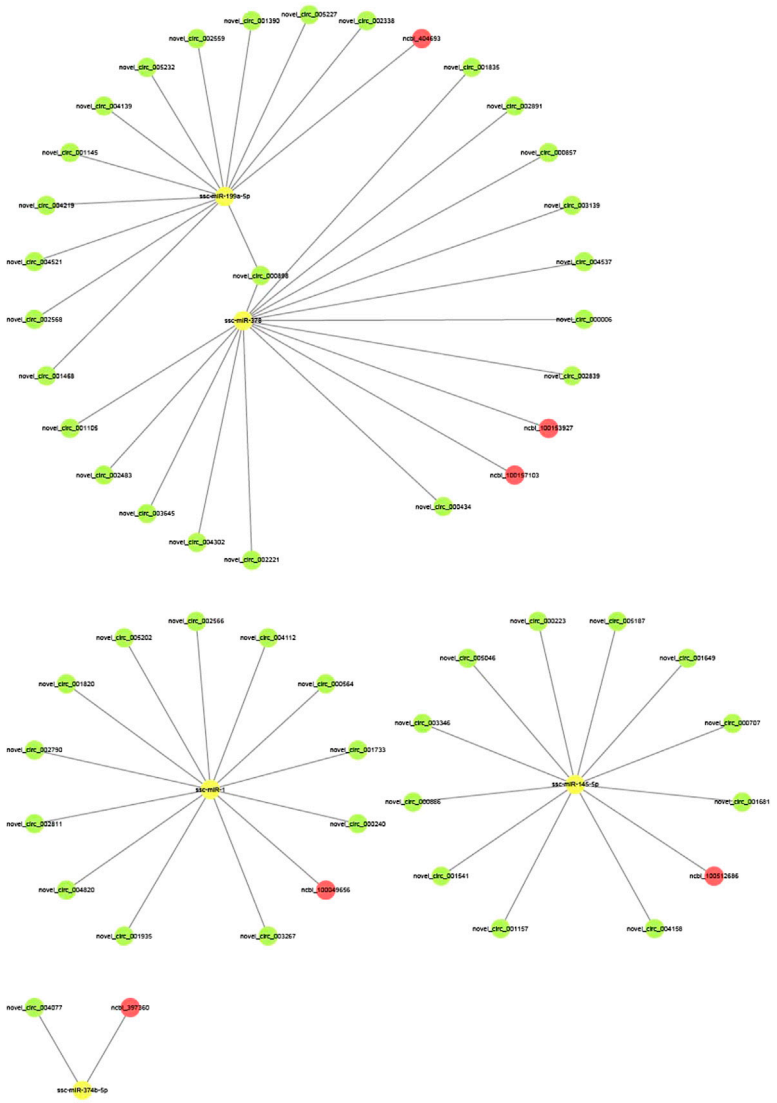


FIGURE 7
The circRNA-miRNA-mRNA interaction network. Green circles represent circRNAs, yellow circles represent miRNAs, and red circles represent mRNAs.

the luciferase activity was significantly decreased following co-transfection with the miR-141 mimic and HOOK3-WT ($p < 0.01$), while no effect on the mutant types of HOOK3-MUT (Supplementary Figure S4 in Supplementary Material S1). These results initially confirmed the direct interactions between miR-141 and *HOOK3*.

Discussion

Tibetan pigs have historically lived in hypoxic environments and have undergone strong selection, and their molecular patterns may represent regulatory mechanisms of adaptation to high-altitude hypoxia (Li et al., 2013; Yang et al., 2021b). Accumulating research indicates that circRNAs, miRNAs, and mRNAs are involved in various types of physiological responses to hypoxia, including glycolysis (Fuller et al., 2020), aging (Haque et al., 2020), and cancer (Pan et al., 2021). The critical roles of mRNAs (Sahu et al., 2019), miRNAs (Hoefel et al., 2019) and circRNAs (Xu et al., 2020) are well known for lung development and disease prevention. Since 2013, an increasing number of researchers have focused on explaining the novel functions of circRNAs based on their

tissue-specific expression and high stability as potential biomarkers (Shi et al., 2020; Khan et al., 2021). Occasionally, ATII cells can proliferate and differentiate into ATI cells after the death of them, which are responsible for gas exchange and serve as a barrier that participates in pathogen defense (Aspal and Zemans, 2020). Understanding the circRNA–miRNA–mRNA axis in ATII cells of Tibetan pigs and Landrace pigs could provide important insights into the protective mechanisms of the hypoxia response. The emerging hypoxia regulation function of circRNAs is particularly interesting, as they might thus be candidates for new therapeutic targets and biomarkers for disease. In this study, we identified and characterized the expression patterns of 5,524 circRNAs, 1,332 miRNAs, and 20,720 mRNAs in ATII cells of Tibetan pigs and Landrace pigs via high-throughput sequencing, further explored the circRNA–miRNA–mRNA regulatory axis and provided new insights into the regulatory roles of RNAs in hypoxia.

Hypoxia may activated DNA repair and damage pathways

To investigate the roles of circRNAs in the response of ATII cells to hypoxia, functional analysis of their source genes was performed. DNA repair and damage pathways may be activated during replication stress in cells under hypoxia. GO functional enrichment analysis revealed that the circRNA source genes participated mainly in biological processes, including cell proliferation, cellular processes, and cell killing. Most DEcircRNAs between the LN and LL groups (such as novel_circ_003062, novel_circ_004904, and novel_circ_003307) were generated from DNA damage-related genes, such as *ZNF*451, *TLK*1, and *TET*1, indicating the presence of DNA damage in ATII cells of Landrace pigs under hypoxic conditions, leading to the activation of numerous distinct repair mechanisms and signaling pathways. The small ubiquitin-related modifier (SUMO) ligase ZATT (*ZNF*451) can control cellular responses to topoisomerase 2-mediated damage as a multifunctional DNA repair factor (Schellenberg et al., 2017). Mammalian *TLK*1 was identified in 1999 and is involved in chromatin formation and newly replicated DNA processing as a serine/threonine kinase; in turn, it can be phosphorylated and deactivated by Chk1 in response to DNA damage under hypoxia (Silljé et al., 1999; Krause et al., 2003). Ten-eleven translocation-1 (TET1) mediates the influence of hypoxia on modifying adipocytokine DNA methylation, and its DNA demethylase activity is induced by *HIF-1α* (Ali et al., 2020). In addition, oncogenic *TNKS*2 can promote the migration and invasion of cervical cancer cells by directly upregulating miR-20a (Kang et al., 2012). *KLF*12 is a member of the Krüppel-like factor family, and its overexpression directly affects proangiogenic processes via transcriptional regulators participating in a multitude of cancer-relevant processes (Mao et al., 2020). Several DEcircRNAs between Landrace pigs in the LN and LL groups (such as novel_circ_001713, novel_circ_000707, and novel_circ_005275) were derived from cancer- and disease-related genes, such as *TNKS*2, *KLF*12, and *BTAF*1, which activate numerous signaling pathways involved in DNA damage, apoptosis promotion, and mitochondrial dysfunction caused by an insufficient oxygen supply (Adami et al., 2018; Vinchure et al., 2021).

ATII cell apoptosis may effect a modulator of cell repair in tibetan pigs under hypoxia

In the TN and TL groups, DEcircRNAs (such as novel_circ_000961, novel_circ_004860, and novel_circ_004404) were generated from inhibitors of apoptosis proteins (e.g., baculoviral IAP repeat containing 6 (*BIRC*6)), which may lead to a key autonomous effect as a modulator of cell repair in Tibetan pigs under hypoxia (Guo et al., 2019). Moreover, we suspect that *BIRC*6 could inhibit apoptosis by facilitating ubiquitination-mediated degradation of apoptotic proteins (Hao et al., 2020). In contrast, *ATK*17*A* and *CASP*2 exhibit apoptosis-inducing activity and can mediate cellular apoptosis (Okumu et al., 2017). DEcircRNAs between the TN and TL groups (such as novel_circ_005192, novel_circ_003369, and novel_circ_000923) were derived from *PLOD*2, which may modulate the migration and invasion of cells via PI3K/Akt signaling (Wan et al., 2020). DEcircRNAs between the TN and TL groups (such as novel_circ_001296, novel_circ_001418, novel_circ_000169) were derived from *PARD*3, which may be associated with asymmetrical cell division and direct polarized cell growth. *PARD*3 contains multiple postsynaptic densities and is localized to tight junctions; in addition, it is correlated with invasion in lung squamous cell carcinoma via impaired STAT3 signaling (Ohno, 2001; Ozdamar et al., 2005; Bonastre et al., 2015). As expected, we found that DEcircRNAs between the TN and TL groups (such as novel_circ_001282, novel_circ_001569, and novel_circ_003568) were derived from *COL5A1*, which encodes an alpha chain for one of the low-abundance fibrillar collagens and may regulated in a hypoxia- and SMC-NFATc3-dependent manner in Tibetan pigs during the response to hypoxia (Sheak et al., 2020). The significant higher expression of *COL5A1 i*n TL groups may inhibited apoptosis of ATII cells in Tibetan pigs under lower oxygen concentration, may lead their better survive in the hypoxia environment than Landrace pigs (Liu et al., 2018).

CircRNA-associated ceRNA networks of ATII cells under hypoxia

Ssc-miR-136–5p, ssc-miR-490–3p, ssc-miR-1, ssc-miR-490, ssc-miR-335, ssc-miR-2320–3p, and ssc-miR-10390 were subsequently identified as the most affected miRNAs and were strongly correlated with numerous mRNAs and circRNAs in the ceRNA network based on DEmiRNAs between the TN and TL groups. CircTLK1 positively regulates CBX4 expression, promoting the proliferation and metastasis of renal cell carcinoma, by sponging miR-136–5p (Li et al., 2020). The novel_circ_002672–ssc-miR-136-5p–SOCS4 axis between the TN and TL groups may have a vital role in a series of pathological changes, including thyroid carcinoma, lung squamous cell cancer, pentose and glucuronate interconversion, and ascorbate, aldarate, and retinol metabolism; it may also regulate the IKKβ/NF-κB/A20 pathway as a modulator of the inflammatory response (Deng et al., 2018; Xie et al., 2018; Gao R. Z. et al., 2019). Among these miRNAs, altered expression of ssc-miR-136–5p, ssc-miR-490–3p, and miR-10390 target genes may occur in response to cellular stressors to facilitate the survival of ATII cells in Tibetan pigs (Jeffery and Harries, 2019).

Based on the ceRNA network between the LN and LL groups, ssc-miR-30c-3p, ssc-miR-20b, ssc-miR-671–5p, and ssc-miR-29a-5 could serve as central miRNAs to regulate the expression of numerous target circRNAs and mRNAs. The novel_circ_005449-ssc–miR-20b–*ZDBF2* and novel_circ_000880–ssc-miR-671-5p–*CDC14B* axes are partial axes based on ssc-miR-20b and ssc-miR-671–5p, which are important regulators of 18 mRNAs, 1 circRNA and 21 mRNAs. Two DEcircRNAs were found in our network between the LN and LL groups, and these circRNAs may cause hypoxia injury-induced autophagy through autophagosomal degradation in Landrace pigs (Gao L. et al., 2019; Lu et al., 2020).

Multiple lines of evidence have demonstrated that circRNAs indirectly enhance gene expression by acting as miRNA sponges (Zhong et al., 2018). The overexpressed circRNA competitively binds to and inhibits the activity of a miRNA to prevent the degradation of the target mRNA. We further analyzed the circRNA–miRNA–mRNA network based on ceRNAs and identified ssc-miR-199a-5p, ssc-miR-378, ssc-miR-1, and ssc-miR-145–5p. Ssc-miR-199a-5p and ssc-miR-378 may regulate many miRNAs and mRNAs, including novel_circ_000898-ssc-miR-199a-5p-*CAV1* and novel_circ_000898-ssc-miR-378-*BMP2*, which could mediate ER stress-induced apoptosis and inflammation and attenuate hypoxia-induced injury in ATII cells under hypoxic conditions (Wen et al., 2020). These findings may explain why ATII cells from Tibetan pigs can survive better than those from Landrace pigs under hypoxic conditions.

Conclusion

In conclusion, we present comprehensive documentation elucidating the changes in and regulation of circRNA–miRNA–mRNA axes in ATII cells of Tibetan pigs and Landrace pigs under hypoxia induction. We identified several hypoxia-related genes and pathways showing adaptive changes in ATII cells, including cell proliferation, cellular processes, and cell killing. Our results also provide new insights into the response of ATII cells in Tibetan pigs to hypoxia, and the numerous identified circRNA miRNAs and mRNAs will serve as references for further investigation of their functions.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The animal study was reviewed and approved by the Laboratory Animal Welfare and Ethics Committee of Gansu Agricultural University. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

SZ was the overall project leader who provided financial support and experimental conception. YY was involved in data analyses, statistical analyses, language revisions, journal selection, and manuscript submissions and revisions. YL and HY contributed to the experimental design and implementation. CG contributed to the supervision and assistance of students in managing animals and collecting and analyzing samples. XL and YR were responsible for the trial implementation, supervision of students collecting and analyzing samples, and manuscript preparation. YC and TJ contributed to supervision of sample collection and analysis and manuscript editing. All authors contributed to the article and approved the submitted version.

Funding

The study was supported by the National Natural Science Foundation of China (32060730, 31760644).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.854250/full#supplementary-material

References

Adami, A., Hobbs, B. D., McDonald, M. N., Casaburi, R., and Rossiter, H. B. (2018). Genetic variants predicting aerobic capacity response to training are also associated with skeletal muscle oxidative capacity in moderate-to-severe COPD. *Physiol. Genomics* 50, 688–690. doi:10.1152/physiolgenomics.00140.2017

Ali, M. M., Phillips, S. A., and Mahmoud, A. M. (2020). HIF1α/TET1 pathway mediates hypoxia-induced adipocytokine promoter hypomethylation in human adipocytes. *Cells* 9, 134. doi:10.3390/cells9010134

Alvarez-Palomo, B., Sanchez-Lopez, L. I., Moodley, Y., Edel, M. J., and Serrano-Mollar, A. (2020). Induced pluripotent stem cell-derived lung alveolar epithelial type II cells reduce damage in bleomycin-induced lung fibrosis. *Stem Cell Res. Ther.* 11, 213. doi:10.1186/s13287-020-01726-3

Aspal, M., and Zemans, R. L. (2020). Mechanisms of ATII-to-ATI cell differentiation during lung regeneration. *Int. J. Mol. Sci.* 21, 3188. doi:10.3390/ijms21093188

Bonastre, E., Verdura, S., Zondervan, I., Facchinetti, F., Lantuejoul, S., Chiara, M. D., et al. (2015). PARD3 inactivation in lung squamous cell carcinomas impairs STAT3 and promotes malignant invasion. *Cancer Res.* 75, 1287–1297. doi:10.1158/0008-5472.CAN-14-2444

Butt, U. J., Steixner-Kumar, A. A., Depp, C., Sun, T., Hassouna, I., Wüstefeld, L., et al. (2021). Hippocampal neurons respond to brain activity with functional hypoxia. *Mol. Psychiatry* 26, 1790–1807. doi:10.1038/s41380-020-00988-w

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018b). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinforma. Oxf. Engl.* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Colgan, S. P., Campbell, E. L., and Kominsky, D. J. (2016). Hypoxia and mucosal inflammation. *Annu. Rev. Pathol.* 11, 77–100. doi:10.1146/annurev-pathol-012615-044231

Colgan, S. P., Furuta, G. T., and Taylor, C. T. (2020). Hypoxia and innate immunity: Keeping up with the HIFsters. *Annu. Rev. Immunol.* 38, 341–363. doi:10.1146/annurev-immunol-100819-121537

Darby, I. A., and Hewitson, T. D. (2016). Hypoxia in tissue repair and fibrosis. *Cell Tissue Res.* 365, 553–562. doi:10.1007/s00441-016-2461-3

Deng, G., Gao, Y., Cen, Z., He, J., Cao, B., Zeng, G., et al. (2018). miR-136-5p regulates the inflammatory response by targeting the IKKβ/NF-κB/A20 pathway after spinal cord injury. *Cell. Physiol. biochem.* 50, 512–524. doi:10.1159/000494165

Dias-Freitas, F., Metelo-Coimbra, C., and Roncon-Albuquerque, J, R. (2016). Molecular mechanisms underlying hyperoxia acute lung injury. *Respir. Med.* 119, 23–28. doi:10.1016/j.rmed.2016.08.010

Fujiwara, T., Takeda, N., Hara, H., Ishii, S., Numata, G., Tokiwa, H., et al. (2021). Three-dimensional visualization of hypoxia-induced pulmonary vascular remodeling in mice. *Circulation* 144, 1452–1455. doi:10.1161/CIRCULATIONAHA.121.056219

Fuller, G. G., Han, T., Freeberg, M. A., Moresco, J. J., Ghanbari Niaki, A., Roach, N. P., et al. (2020). RNA promotes phase separation of glycolysis enzymes into yeast G bodies in hypoxia. *eLife* 9, e48480. doi:10.7554/eLife.48480

Gallagher, S. A., and Hackett, P. H. (2004). High-altitude illness. *Emerg Med Clin North Am.* 22, 329–355. doi:10.1016/j.emc.2004.02.001

Gao, L., Dou, Z. C., Ren, W. H., Li, S. M., Liang, X., and Zhi, K. Q. (2019a). CircCDR1as upregulates autophagy under hypoxia to promote tumor cell survival via AKT/ERK(½)/mTOR signaling pathways in oral squamous cell carcinomas. *Cell Death Dis.* 10, 745. doi:10.1038/s41419-019-1971-9

Gao, R. Z., Que, Q., Lin, P., Pang, Y. Y., Wu, H. Y., Li, X. J., et al. (2019b). Clinical roles of miR-136-5p and its target metadherin in thyroid carcinoma. *Am. J. Transl. Res.* 11, 6754–6774.

Greco, F., Wiegert, S., Baumann, P., Wellmann, S., Pellegrini, G., and Cannizzaro, V. (2019). Hyperoxia-induced lung structure-function relation, vessel rarefaction, and cardiac hypertrophy in an infant rat model. *J. Transl. Med.* 17, 91. doi:10.1186/s12967-019-1843-1

Grek, C. L., Newton, D. A., Spyropoulos, D. D., and Baatz, J. E. (2011). Hypoxia up-regulates expression of hemoglobin in alveolar epithelial cells. *Am. J. Respir. Cell Mol. Biol.* 44, 439–447. doi:10.1165/rcmb.2009-0307OC

Guo, Y., Tan, J., Miao, Y., Sun, Z., and Zhang, Q. (2019). Effects of Microvesicles on Cell Apoptosis under Hypoxia. *Oxidative Medicine and Cellular Longevity*, 2019, , 597215211.

Hao, T., Wang, Z., Yang, J., Zhang, Y., Shang, Y., and Sun, J. (2020). MALAT1 knockdown inhibits prostate cancer progression by regulating miR-140/BIRC6 axis. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 123, 109666doi:10.1016/j.biopha.2019.109666

Haque, S., Ames, R. M., Moore, K., Pilling, L. C., Peters, L. L., Bandinelli, S., et al. (2020). circRNAs expressed in human peripheral blood are associated with human aging phenotypes, cellular senescence and mouse lifespan. *GeroScience* 42, 183–199. doi:10.1007/s11357-019-00120-z

Herriges, M., and Morrisey, E. E. (2014). Lung development: Orchestrating the generation and regeneration of a complex organ. *Dev. Camb. Engl.* 141, 502–513. doi:10.1242/dev.098186

Hoefel, G., Tay, H., and Foster, P. (2019). MicroRNAs in lung diseases. *Chest* 156, 991–1000. doi:10.1016/j.chest.2019.06.008

Ito, Y., Correll, K., Schiel, J. A., Finigan, J. H., Prekeris, R., and Mason, R. J. (2014). Lung fibroblasts accelerate wound closure in human alveolar epithelial cells through hepatocyte growth factor/c-Met signaling. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 307, 94–105. doi:10.1152/ajplung.00233.2013

Jeffery, N., and Harries, L. W. (2019). miRNAs responsive to the diabetic microenvironment in the human beta cell line EndoC-βH1 may target genes in the FOXO, HIPPO and Lysine degradation pathways. *Exp. Cell Res.* 384, 111559doi:10.1016/j.yexcr.2019.111559

Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhauser, D., et al. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* 6, 364. doi:10.1038/msb.2010.18

Kang, H. W., Wang, F., Wei, Q., Zhao, Y. F., Liu, M., Li, X., et al. (2012). miR-20a promotes migration and invasion by regulating TNKS2 in human cervical cancer cells. *FEBS Lett.* 586, 897–904. doi:10.1016/j.febslet.2012.02.020

Khan, S., Jha, A., Panda, A. C., and Dixit, A. (2021). Cancer-associated circRNA-miRNA-mRNA regulatory networks: A meta-analysis. *Front. Mol. Biosci.* 8, 671309doi:10.3389/fmolb.2021.671309

Krause, D. R., Jonnalagadda, J. C., Gatei, M. H., Sillje, H. H., Zhou, B. B., Nigg, E. A., et al. (2003). Suppression of Tousled-like kinase activity after DNA damage or replication block requires ATM, NBS1 and Chk1. *Oncogene* 22, 5927–5937. doi:10.1038/sj.onc.1206691

Li, J., Huang, C., Zou, Y., Ye, J., Yu, J., and Gui, Y. (2020). CircTLK1 promotes the proliferation and metastasis of renal cell carcinoma by sponging miR-136-5p. *Mol. Cancer* 19, 103. doi:10.1186/s12943-020-01225-2

Li, M., Jin, L., Ma, J., Tian, S., Li, R., and Li, X. (2016). Detecting mitochondrial signatures of selection in wild Tibetan pigs and domesticated pigs. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 27, 747–752. doi:10.3109/19401736.2014.913169

Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45, 1431–1438. doi:10.1038/ng.2811

Liu, W., Wei, H., Gao, Z., Chen, G., Liu, Y., Gao, X., et al. (2018). COL5A1 may contribute the metastasis of lung adenocarcinoma. *Gene* 30, 57–66. doi:10.1016/j.gene.2018.04.066

Lu, Y., Wang, S., Cai, S., Gu, X., Wang, J., Yang, Y., et al. (2020). Propofol-induced miR-20b expression initiates endogenous cellular signal changes mitigating hypoxia/re-oxygenation-induced endothelial autophagy *in vitro*. *Cell Death Dis.* 11, 681. doi:10.1038/s41419-020-02828-9

Mao, S., Lu, Z., Zheng, S., Zhang, H., Zhang, G., Wang, F., et al. (2020). Exosomal miR-141 promotes tumor angiogenesis via KLF12 in small cell lung cancer. *J. Exp. Clin. Cancer Res.* 39, 193. doi:10.1186/s13046-020-01680-1

Ohno, S. (2001). Intercellular junctions and cellular polarity: The PAR-aPKC complex, a conserved core cassette playing fundamental roles in cell polarity. *Curr. Opin. Cell Biol.* 13, 641–648. doi:10.1016/s0955-0674(00)00264-7

Okumu, D. O., East, M. P., Levine, M., Herring, L. E., Zhang, R., Gilbert, T. S. K., et al. (2017). BIRC6 mediates imatinib resistance independently of Mcl-1. *PloS one* 12, e0177871. doi:10.1371/journal.pone.0177871

Ozdamar, B., Bose, R., Barrios-Rodiles, M., Wang, H. R., Zhang, Y., and Wrana, J. L. (2005). Regulation of the polarity protein Par6 by TGFbeta receptors controls epithelial cell plasticity. *Science* 307, 1603–1609. doi:10.1126/science.1105718

Pan, Y., Kadash-Edmondson, K. E., Wang, R., Phillips, J., Liu, S., Ribas, A., et al. (2021). RNA dysregulation: An expanding source of cancer immunotherapy targets. *Trends Pharmacol. Sci.* 42, 268–282. doi:10.1016/j.tips.2021.01.006

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122

Qin, S., Chen, M., Ji, H., Liu, G. Y., Mei, H., Li, K., et al. (2018). miR-21-5p regulates type II alveolar epithelial cell apoptosis in hyperoxic acute lung injury. *Mol. Med. Rep.* 17, 5796–5804. doi:10.3892/mmr.2018.8560

Richter, K., Haslbeck, M., and Buchner, J. (2010). The heat shock response: Life on the verge of death. *Mol. Cell* 40, 253–266. doi:10.1016/j.molcel.2010.10.006

Sahu, I., Haque, A., Weidensee, B., Weinmann, P., and Kormann, M. S. D. (2019). Recent developments in mRNA-based protein supplementation therapy to target lung diseases. *Mol. Ther.* 27, 803–823. doi:10.1016/j.ymthe.2019.02.019

Schellenberg, M. J., Lieberman, J. A., Herrero-Ruiz, A., Butler, L. R., Williams, J. G., Muñoz-Cabello, A. M., et al. (2017). ZATT (ZNF451)-mediated resolution of topoisomerase 2 DNA-protein cross-links. *Science* 357, 1412–1416. doi:10.1126/science.aam6468

Schmeckebier, S., Mauritz, C., Katsirntaki, K., Sgodda, M., Puppe, V., Duerr, J., et al. (2013). Keratinocyte growth factor and dexamethasone plus elevated cAMP levels synergistically support pluripotent stem cell differentiation into alveolar epithelial type II cells. *Tissue Eng. Part A* 19, 938–951. doi:10.1089/ten.tea.2012.0066

Sender, R., and Milo, R. (2021). The distribution of cellular turnover in the human body. *Nat. Med.* 27, 45–48. doi:10.1038/s41591-020-01182-9

Sheak, J. R., Jones, D. T., Lantz, B. J., Maston, L. D., Vigil, D., Resta, T. C., et al. (2020). NFATc3 regulation of collagen V expression contributes to cellular immunity to collagen type V and hypoxic pulmonary hypertension. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 319, L968–l980. doi:10.1152/ajplung.00184.2020

Sherman, M. A., Suresh, M. V., Dolgachev, V. A., McCandless, L. K., Xue, X., Ziru, L., et al. (2018). Molecular characterization of hypoxic alveolar epithelial cells after lung contusion indicates an important role for HIF-1α. *Ann. Surg.* 267, 382–391. doi:10.1097/SLA.0000000000002070

Shi, Y., Jia, X., and Xu, J. (2020). The new function of circRNA: translation. *Clinical & translational oncology* 22, 2162–2169. doi:10.1007/s12094-020-02371-1

Silljé, H. H., Takahashi, K., Tanaka, K., Van Houwe, G., and Nigg, E. A. (1999). Mammalian homologues of the plant Tousled gene code for cell-cycle-regulated kinases with maximal activities linked to ongoing DNA replication. *EMBO J.* 18, 5691–5702. doi:10.1093/emboj/18.20.5691

Vinchure, O, S., Whittemore, K., Kushwah, D., Blasco, M, A., and Kulshreshtha, R. (2021). miR-490 suppresses telomere maintenance program and associated hallmarks in glioblastoma. *Cell. Mol. Life Sci.* 78, 2299–2314. doi:10.1007/s00018-020-03644-2

Wan, J., Qin, J., Cao, Q., Hu, P., Zhong, C., and Tu, C. (2020). Hypoxia-induced PLOD2 regulates invasion and epithelial-mesenchymal transition in endometrial carcinoma cells. *Genes Genomics* 42, 317–324. doi:10.1007/s13258-019-00901-y

Wang, X., Zhang, L., and Sun, B. (2016). Neonatal type II alveolar epithelial cell transplant facilitates lung reparation in piglets with acute lung injury and extracorporeal life support. *Pediatr. Crit. Care Med.* 17, e182–192. doi:10.1097/PCC.0000000000000667

Wen, J., Ma, L., Xu, Y., Wu, J., Yu, Y., Peng, J., et al. (2020, 10943Effects of probiotic litchi juice on immunomodulatory function and gut microbiota in mice. *Food Res. Int.* 137. doi:10.1016/j.foodres.2020.109433

Wu, D. D., Yang, C. P., Wang, M. S., Dong, K. Z., Yan, D. W., Hao, Z. Q., et al. (2020). Convergent genomic signatures of high-altitude adaptation among domestic mammals. *Natl. Sci. Rev.* 7, 952–963. doi:10.1093/nsr/nwz213

Xie, Z. C., Li, T. T., Gan, B. L., Gao, X., Gao, L., Chen, G., et al. (2018). Investigation of miR-136-5p key target genes and pathways in lung squamous cell cancer based on TCGA database and bioinformatics analysis. *Pathol. Res. Pract.* 214, 644–654. doi:10.1016/j.prp.2018.03.028

Xu, Y., Yu, J., Huang, Z., Fu, B., Tao, Y., Qi, X., et al. (2020). Circular RNA hsa_circ_0000326 acts as a miR-338-3p sponge to facilitate lung adenocarcinoma progression. *J. Exp. Clin. Cancer Res.* 39, 57. doi:10.1186/s13046-020-01556-4

Yang, Y., Gao, C., Yang, T., Sha, Y., Cai, Y., Wang, X., et al. (2021a). Vascular characteristics and expression of hypoxia genes in Tibetan pigs' hearts. *Vet. Med. Sci.* 8, 177–186. doi:10.1002/vms3.639

Yang, Y., Yuan, H., Yang, T., Li, Y., Gao, C., Jiao, T., et al. (2021b). The expression regulatory network in the lung tissue of Tibetan pigs provides insight into hypoxia-sensitive pathways in high-altitude hypoxia. *Front. Genet.* 12, 691592doi:10.3389/fgene.2021.691592

Zhong, Y., Du, Y., Yang, X., Mo, Y., Fan, C., Xiong, F., et al. (2018). Circular RNAs function as ceRNAs to regulate and control human cancer progression. *Mol. Cancer* 17, 79. doi:10.1186/s12943-018-0827-8

frontiers | Frontiers in Molecular Biosciences

TYPE Brief Research Report
PUBLISHED 27 September 2022
DOI 10.3389/fmolb.2022.954087

OPEN ACCESS

EDITED BY
Gian Gaetano Tartaglia,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Elias Georges Bechara,
Italian Institute of Technology (IIT), Italy
S. Hossein Fatemi,
University of Minnesota Twin Cities, United States
Stephanie Byrum,
University of Arkansas for Medical Sciences, United States
Kirsty Sawicka,
University of Cambridge, United Kingdom

*CORRESPONDENCE
Carole Gwizdek,
gwizdek@ipmc.cnrs.fr

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 26 May 2022
ACCEPTED 05 August 2022
PUBLISHED 27 September 2022

CITATION
Kieffer F, Hilal F, Gay A-S, Debayle D, Pronot M, Poupon G, Lacagne I, Bardoni B, Martin S and Gwizdek C (2022), Combining affinity purification and mass spectrometry to define the network of the nuclear proteins interacting with the N-terminal region of FMRP.
Front. Mol. Biosci. 9:954087.
doi: 10.3389/fmolb.2022.954087

Combining affinity purification and mass spectrometry to define the network of the nuclear proteins interacting with the N-terminal region of FMRP

Félicie Kieffer[1], Fahd Hilal[1], Anne-Sophie Gay[1], Delphine Debayle[1], Marie Pronot[1], Gwénola Poupon[1], Iliona Lacagne[1], Barbara Bardoni[2], Stéphane Martin[2] and Carole Gwizdek[1]*

[1]Université Côte d'Azur, Centre National de la Recherche Scientifique, Institut de Pharmacologie Moléculaire et Cellulaire, Valbonne, France, [2]Université Côte d'Azur, Institut National de la Santé Et de la Recherche Médicale, Centre National de la Recherche Scientifique, Institut de Pharmacologie Moléculaire et Cellulaire, Valbonne, France

Fragile X-Syndrome (FXS) represents the most common inherited form of intellectual disability and the leading monogenic cause of Autism Spectrum Disorders. In most cases, this disease results from the absence of expression of the protein FMRP encoded by the *FMR1* gene (Fragile X messenger ribonucleoprotein 1). FMRP is mainly defined as a cytoplasmic RNA-binding protein regulating the local translation of thousands of target mRNAs. Interestingly, FMRP is also able to shuttle between the nucleus and the cytoplasm. However, to date, its roles in the nucleus of mammalian neurons are just emerging. To broaden our insight into the contribution of nuclear FMRP in mammalian neuronal physiology, we identified here a nuclear interactome of the protein by combining subcellular fractionation of rat forebrains with pull-down affinity purification and mass spectrometry analysis. By this approach, we listed 55 candidate nuclear partners. This interactome includes known nuclear FMRP-binding proteins as Adar or Rbm14 as well as several novel candidates, notably Ddx41, Poldip3, or Hnrnpa3 that we further validated by target-specific approaches. Through our approach, we identified factors involved in different steps of mRNA biogenesis, as transcription, splicing, editing or nuclear export, revealing a potential central regulatory function of FMRP in the biogenesis of its target mRNAs. Therefore, our work considerably enlarges the nuclear proteins interaction network of FMRP in mammalian neurons and lays the basis for exciting future mechanistic studies deepening the roles of nuclear FMRP in neuronal physiology and the etiology of the FXS.

KEYWORDS

nuclear fractionation, FMRP, mRNA metabolism, nuclear protein network, proteomics

Introduction

The Fragile X-Syndrome (FXS) represents the most common inherited form of intellectual disability and the first monogenic cause of Autism Spectrum Disorders, affecting 1/4,000 males and 1/7,000 females (Dahlhaus 2018; Maurin and Bardoni 2018; Richter and Zhao 2021). This neurodevelopmental disorder is characterized by a broad range of neurologic/psychiatric phenotypes including mental impairment, autism, attention deficit, hyperactivity, social anxiety, and epilepsy. In the majority of cases, FXS is due to the silencing of the *FMR1* gene, recently renamed as Fragile X messenger ribonucleoprotein 1, encoding FMRP (or FXP, Fragile X Protein) (Herring et al., 2022).

FMRP is an RNA binding protein, mostly cytoplasmic, able to interact with thousands of messenger RNAs (mRNAs) (Darnell et al., 2011; Ascano et al., 2012; Maurin et al., 2015; Maurin et al., 2018; Sawicka et al., 2019; Tran et al., 2019; Hale et al., 2021). These lists comprise mRNAs encoding proteins with large array of roles in cell processes, including proteins essential for the development and the function of synapses. Canonically, FMRP is defined as a translational suppressor through interactions with the translational machinery and the miRNA pathway. However, evidence exist about its capacity to enhance translation (Bechara et al., 2009; Tabet et al., 2016; Richter and Zhao 2021). It may also participate in the transport of its target mRNAs along the dendrites, within ribonucleoprotein complexes called "transport granules", and represses their translation until they arrive at the synapses (Maurin et al., 2014; Richter and Zhao 2021). Thus, the cognitive deficiencies observed in patients with FXS are thought to result, at least in part, from the deregulation in protein translation of mRNAs bound by FMRP. In addition, FMRP may be involved in the storage and the stability of some of its mRNA targets (Davis and Broadie 2017; Richter and Zhao 2021). Lastly, FMRP directly binds different ion channels to regulate their gating, thus impacting neuronal excitability (Davis and Broadie 2017).

To accomplish its functions, FMRP interacts with numerous proteins in addition to its target mRNAs. In this context, the N-terminal domain of FMRP plays a central role (Ramos et al., 2006). Indeed, it presents a combination of Tudor and pseudo-KH patterns that promotes many of the known protein interactions of FMRP, including its homomerisation (Ramos, et al., 2006; Hu et al., 2015; Myrick et al., 2015). Intriguingly, the N-terminal domain of FMRP also contains a nuclear localization signal while its central region bears a nuclear export signal and its C-terminus two nucleolar localization sequences (Eberhart et al., 1996; Sittler et al., 1996; Bardoni et al., 1997; Taha et al., 2014), allowing the protein to enter in and exit from the nucleus and the nucleolus. Very recently, mutations within the nuclear export signal of FMRP have been found in some FXS patients, suggesting that the nucleocytoplasmic shuttling of FMRP may be important to neuronal physiology (Zeidler et al., 2021; Mangano et al., 2022). Consistently, a growing number of studies associates FMRP with nuclear functions, such as DNA damage response (Alpatov et al., 2014; Zhang et al., 2014; Chakraborty et al., 2020), certain steps in mRNA biogenesis (Didiot et al., 2009; Bhogal et al., 2011; Shamay-Ramot et al., 2015; Filippini et al., 2017; Zhou et al., 2017; Tran, et al., 2019) or their export (Kim et al., 2009; Edens et al., 2019; Hsu et al., 2019; Westmark et al., 2020; Kim et al., 2021), ribosomal RNA methylation (D'Souza et al., 2018) or nuclear pore assembly (Agote-Aran et al., 2020). However, the roles of FMRP in the nucleus of mammalian neurons remain insufficiently documented and their relevance a physiological context are still poorly understood. Notably, few information is available about the molecular mechanisms by which FMRP is involved in these different nuclear processes and whether the protein may play additional roles in this compartment. To acquire a more comprehensive representation of the functions of FMRP in the nucleus of mammalian neurons, we identified the nuclear protein partners of the N-terminal protein/protein interaction domain of FMRP in rat forebrain, using affinity pull down on nuclear fraction isolated from rat forebrain coupled to quantitative mass spectrometry analysis.

Materials and methods

Rat strain

Wistar rats were purchased exclusively from a commercial source (Janvier, St Berthevin, France). All animals were handled and treated in accordance with the ARRIVE Guidelines. Animals had free access to water and food. Lightning was controlled as a 12 h light and dark cycle and the temperature maintained at 23°C ± 1°C. The protocols for PND 14 pups euthanasia by decapitation and the preparation of primary neuronal cultures from rat embryos at E17 were approved by the Animal Care and Ethics Committee (APAFIS #18647-2019011110552947v3). For biochemical analyses, forebrains were immediately excised, frozen in liquid nitrogen and stored at −80°C until use.

Nuclear preparation

Forebrains of PND 14 rats were homogenized in ice-cold hypotonic buffer at 1.5 mM MgCl2 (see composition of all buffers in Supplementary Material) using a glass-Teflon homogenizer. Igepal (MP) was added at a final concentration of 0.3% and the homogenate was filtrated on nylon Cell Stainer 70 µm (Falcon) (Total fraction). The filtrate was centrifuged at 800 g for 5 min at 4°C. The supernatant was removed (Cytoplasmic fraction). After resuspension of the pellet in hypotonic buffer at 0.5 mM MgCl2 and filtration, 1.1 volume of Optiprep (StemCell) was added. After a gentle

homogenization, the mix was subjected to centrifugation at 5,000 g for 15 min at 4°C. The pellet, corresponding to the purified nuclei, was resuspended in native lysis buffer, sonicated 7 times at 15% of the power (Sonic ruptor 400, Omni International) and centrifuge at 4,000 g for 15 min at 4°C (Nuclear fraction). Protein concentration was determined using standard Bradford assay. Buffer compositions are available in the Supplementary Material.

Glutathione S-transferase-pull down

pGEX-4T1 plasmids encoding the Glutathione S-transferase (GST) or the N-terminal fragment (amino acids 1–213) of human FMRP fused to the GST (GST-FNT) were transfected to in *E. coli* BL21 (DE3) bacteria (Invitrogen) and recombinant GST or GST-FNT were produced and purified as previously described (Khayachi et al., 2018). 50 µg of GST or 100 µg of GST-FNT purified recombinant proteins were incubated 1 h at 4°C under soft rotation with 25 µl of Glutathione Sepharose 4B beads (GE HealthCare). Beads were washed twice with PBS and bound recombinant proteins were cross-linked to the beads using 30 mM dimethyl pimelimidate (Sigma) according to the previously published protocol (Pronot et al., 2021). 6 mg of proteins from nuclear lysates were incubated with 25 µl of GST-FNT or GST cross-linked beads overnight at 4°C under soft rotation. Beads were washed three times for 5 min at 4°C in wash buffer. To decrease unspecific or indirect bindings, the beads were further washed in high stringency wash buffer containing 500 mM NaCl. Proteins bound to the beads were then eluted in 30 µl of Laemmli buffer for 10 min at 95°C. Buffer compositions are available in the Supplementary Material.

Mass spectrometry analysis

After separation by short SDS-PAGE, gel slicing in two bands per lane and in gel Trypsin digestion, proteins from PND14 rat forebrain nuclear extract or isolated from GST and GST-FNT pull down were identified by liquid nano-chromatography coupled to tandem mass spectrometry as described in the Supplementary Material and Methods.

Bioinformatics

Uniprot protein identifiers were converted into Entrez Gene identifiers using the Uniprot or the db2db conversion tools. When rat datasets were compared to human datasets, homologs of our dataset were identified using DIOPT or Blast. Enrichment analysis for Gene Ontology terms or REACTOME pathways against the Rattus Norvegicus Proteome and annotations for Uniprot Keywords were performed using DAVID. Network analysis were conducted using STRING (v11.5) with physical subnetwork mode (excluding Text mining sources) and a medium confidence score. Clusters prediction and annotation were performed using Cytoscape (v3.8.0) implemented by the StringApp.

Proximity ligation assay on primary neuronal culture

Hippocampal neurons were prepared as previously described (Schorova et al., 2019). The Duo-link® using PLA Technology kit (Sigma-Aldrich) was used for the proximity ligation assay, accordingly to the manufacturer instructions. The primary antibodies incubation was performed overnight at 4°C as indicated in the Supplementary Material. Neuronal cells in the cultures were identified upon their MAP2 labelling. Confocal images were acquired with a ×63 oil immersion lens (numerical aperture NA 1.4) on an inverted TCS-SP5 confocal microscope (Leica Microsystems, Nanterre, France).

Results

At the steady state, FMRP is mainly cytoplasmic and barely detectable in the nucleus. Indeed, nuclear accumulation of FMRP was essentially detected in cell lines exogenously over-expressing full length, mutated or truncated proteins (Eberhart, et al., 1996; Fridell et al., 1996; Sittler, et al., 1996; Willemsen et al., 1996; Bardoni, et al., 1997; Feng et al., 1997; Tamanini et al., 1999a; Taha, et al., 2014). It has been estimated from subcellular fractionation of human lymphoblastoid cells that approximatively 2%–4% of the endogenous FMRP are present in the nuclear compartment (Feng, et al., 1997). In mammalian neurons, only immuno-electron microscopy or PLA approaches provided sufficient sensitivity to detect endogenous FMRP-labeled particles in the nucleus (Feng, et al., 1997; Bakker et al., 2000; Filippini, et al., 2017), suggesting that only a small proportion of the protein goes to the nucleus and/or that its passage is very transient. In this context, the identification of the nuclear protein interactome of FMRP appears particularly challenging. To overcome this limitation, we chose to use a GST pull-down co-purification approach on an enriched nuclear fraction from rat forebrain followed by mass spectrometry analysis.

Preparation of an enriched nuclear fraction from developing rat brain

In both human and rodent, FMRP, essential for proper neuronal development, is highly expressed in neonatal brain and declines to reach low levels of expression in adults

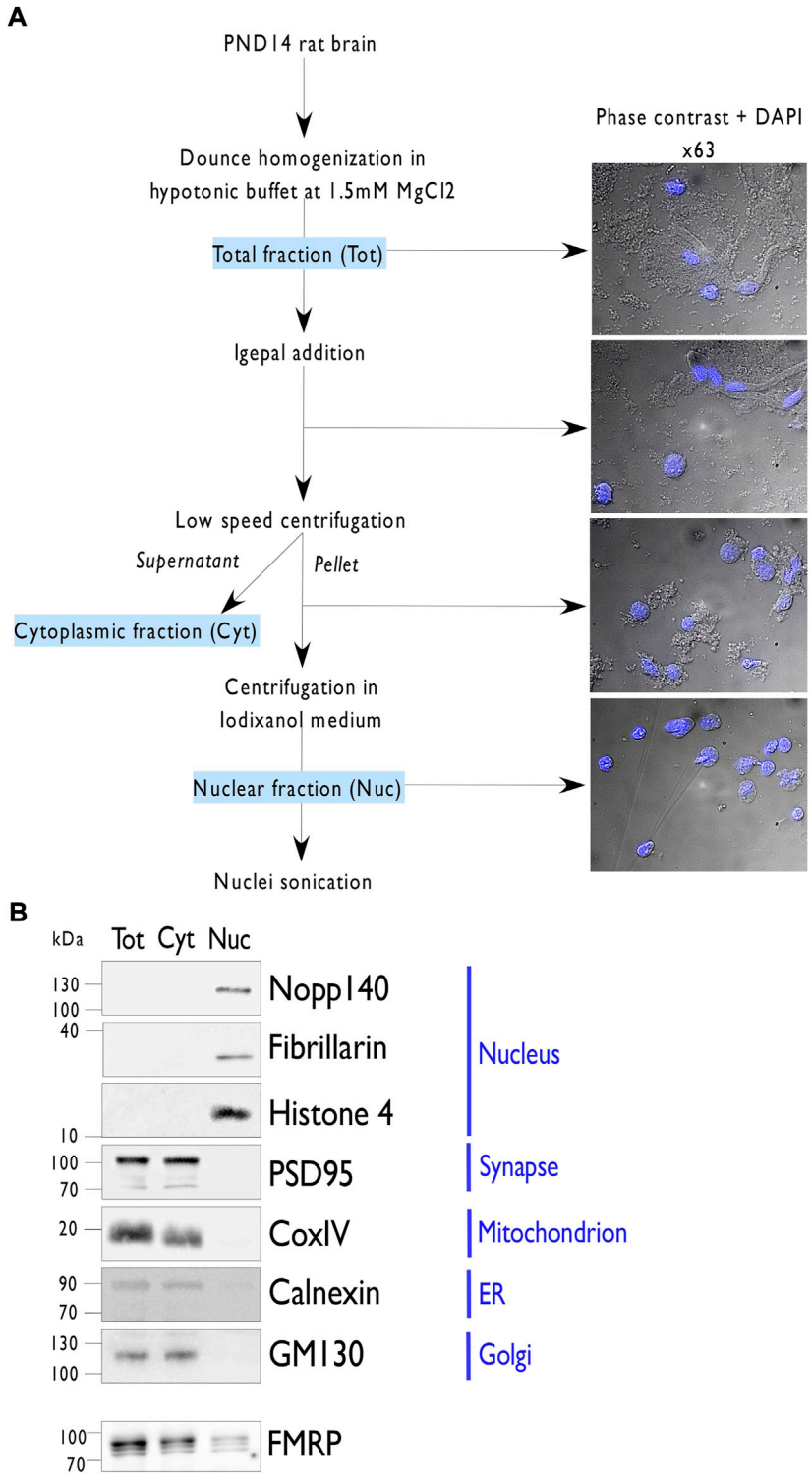


FIGURE 1
Nuclei isolation from PND14 rat forebrains. **(A)**. Representation of the protocol used to purify nuclei from postnatal day 14 (PND) rat forebrains. Briefly, forebrains were dounced in hypotonic buffer and then supplemented with 0.3% of Igepal detergent. After filtration, the total forebrain lysate (Tot) was subjected to low-speed centrifugation to separate the nuclei in the pellet from the cytoplasmic material in the supernatant (Cyt). After resuspension of the pellet in hypotonic buffer, 1.1 volume of Optiprep was added. After gentle homogenization, the mix was subjected to high-speed centrifugation and the purified nuclei (Nuc) were recovered in the pellet. To extract nuclear proteins, purified nuclei were resuspended in a native lysis buffer, sonicated, and clarified by high-speed centrifugation. To follow the purity fraction, images are acquired by phase contrast and superposition with DAPI staining at the indicated steps. **(B)**. Immunoblot analysis of 10 µg of proteins from the indicated fractions (Tot, Cyt, Nuc) per lane and detected by western blotting using antibodies against different subcellular markers or FMRP.

(Bonaccorso et al., 2015; Prieto et al., 2021). In this context, postnatal day (PND) 14 represents an interesting period to identify the nuclear interactome of FMRP in rat as it combines intense synaptogenesis (Semple et al., 2013) and high levels of the protein (Bonaccorso, et al., 2015). To prepare the nuclear protein fraction, whole nuclei were isolated from the forebrain of PND14 rats *via* a series of differential centrifugations (Figure 1). The quality of the fractionation was verified by phase contrast microscopy (Figure 1A) as well as by western blotting (Figure 1B; Supplementary Figure S1). As observed by light microscopy, each step of the workflow increased the purity of the preparation, until a fraction highly enriched in pure and intact nuclei is obtained (Figure 1A). To address the relative purity of the nuclear extract, protein samples from the total (Tot), cytoplasmic (Cyt), and nuclear (Nuc) lysates were analyzed by western blotting for specific markers of various sub-cellular compartments (Figure 1B; Supplementary Figure S1). The nuclear fraction results to be highly enriched in the nuclear markers Histone H4, Fibrillarin, and Nopp140 compared to the other fractions and devoid of markers for synapses (Synaptotagmin), mitochondrion (CoxIV), endoplasmic reticulum (Calnexin), and Golgi apparatus (GM130). Interestingly, FMRP is also found in the nuclear fraction (Figure 1B). We estimated that around 2% of the endogenous protein is localized in the nucleus, comparable to what has been previously described for human lymphoiblastoid cells (Feng, et al., 1997).

Lastly, to evaluate the quality of the purified fraction, two independent nuclear lysates from PND14 rat forebrains were subjected to proteomics analysis after protein separation by short SDS-PAGE, gel slicing in two bands per lane and in gel Trypsin digestion. A total of 1,196 distinct proteins was identified and 945 of them (79%) were present in both sets of nuclear preparation (Figure 2A; Supplementary Table S1). Gene Ontology (GO) enrichment analysis showed that the top 10 for enriched GO Cellular Components terms is clearly associated to the nuclear compartment whereas the top 10 for enriched GO Molecular Function terms refers essentially to RNA binding activities. Consistently, the top 10 for enriched GO Biological Processes or REACTOME pathways revealed the involvement of the identified proteins in different steps of RNA metabolic processes (Figure 2B; Supplementary Table S2). Altogether, these data confirm the enrichment of the samples in nuclear components thus highlighting the quality of the nuclei preparation.

Identification of a nuclear interactome of FMRP by pull-down purification and mass spectrometry analysis

The nuclear lysates from PND14 rat forebrains were used to investigate the nuclear interactors of FMRP by stringent GST pull-down assays using the recombinant N-terminal protein/protein interaction domain of FMRP fused to the GST (GST-FNT) as a bait (Supplementary Figure S2), or the GST alone as negative control, as detailed in the Material and Methods section. Three independent GST-FNT co-purifications with their respective GST controls were analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). MS data were processed as described in the Material and Methods section and the differential statistical analysis for each identified prey was performed using SAINTexpress, the upgraded implementation of the Significance Analysis of INTeractome Tool (Mellacheruvu et al., 2013; Teo et al., 2014). As previously reported (Guard et al., 2019), proteins presenting a fold-change enrichment (FC) cutoff of >3.00 and a score probability (SP) cutoff >0.7 were classified as "high-confidence" interactor whereas we referred to all other proteins with a FC cutoff of >2.00 and a SP cutoff >0.5 as "medium-confidence" interactors. This workflow led to a list of 55 FMRP-interacting nuclear candidates, with 20 proteins satisfying the "high-confidence" requirements and 35 passing the "medium-confidence" standards (Figure 3A; Supplementary Table S3). Noteworthy, our dataset presents five reported FMRP partners previously identified by different approaches: FMRP itself (Ramos, et al., 2006; Hu, et al., 2015; Myrick, et al., 2015), FXR1P (Fragile X Related Protein 1) (Zhang et al., 1995), the zinc finger RNA-binding protein ZFR (Worringer et al., 2009), the splicing factor Rbm14 (Zhou, et al., 2017), and the mRNA editing enzyme Adar (Shamay-Ramot, et al., 2015; Filippini, et al., 2017). Besides, two proteomic screenings by affinity pull-down with the N-terminal domain of FMRP as bait were previously conducted using total cell extracts from HEK293 cells as source of preys (He and Ge 2017; Taha et al., 2021). The comparison of our list with these dataset brought out five proteins in common: FMRP, FXR1P, the chromatin-remodeling factor CHD4 and the transcriptional factors TCF20 and ZNF638.

Validation of novel nuclear FMRP interacting proteins

To go further with the validation of the proteomic screen, we then verified the interaction between FMRP and FXR1P or three novel candidates, Ddx41, Poldip3, and Hnrnpa3 by target specific approaches.

As FMRP, FXR1P is a predominantly cytoplasmic RNA binding protein, playing a role in the local translation or the stability of certain mRNAs (Khlghatyan and Beaulieu 2018; George et al., 2021). Interestingly, FXR1P forms heterodimers with FMRP and these proteins have common mRNA targets (Tamanini et al., 1999b; Darnell et al., 2009). FXR1P is also able to shuttle between the nucleus and the cytoplasm (Tamanini et al., 1999a; Bakker, et al., 2000). Tran et al. recently showed that

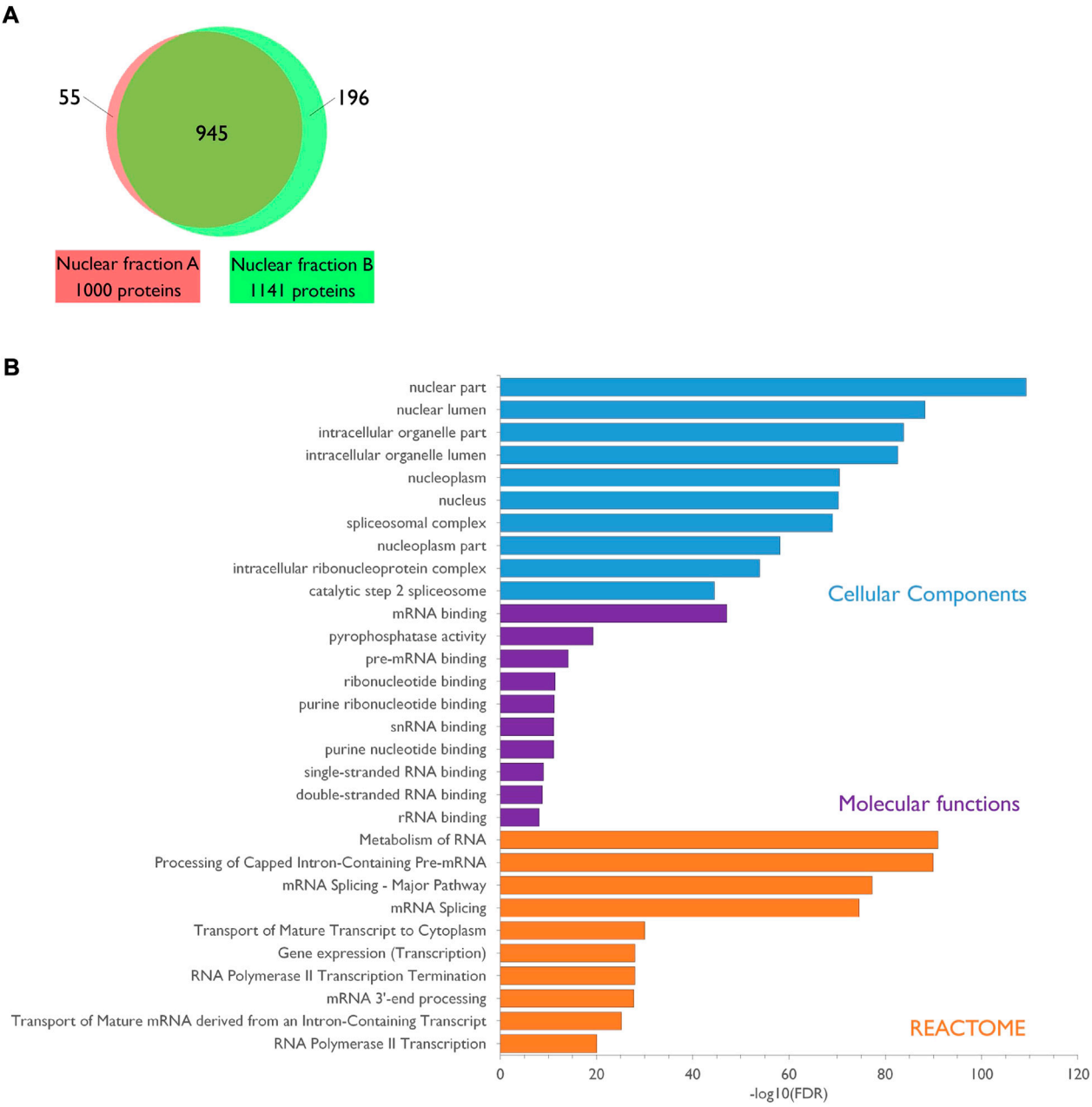
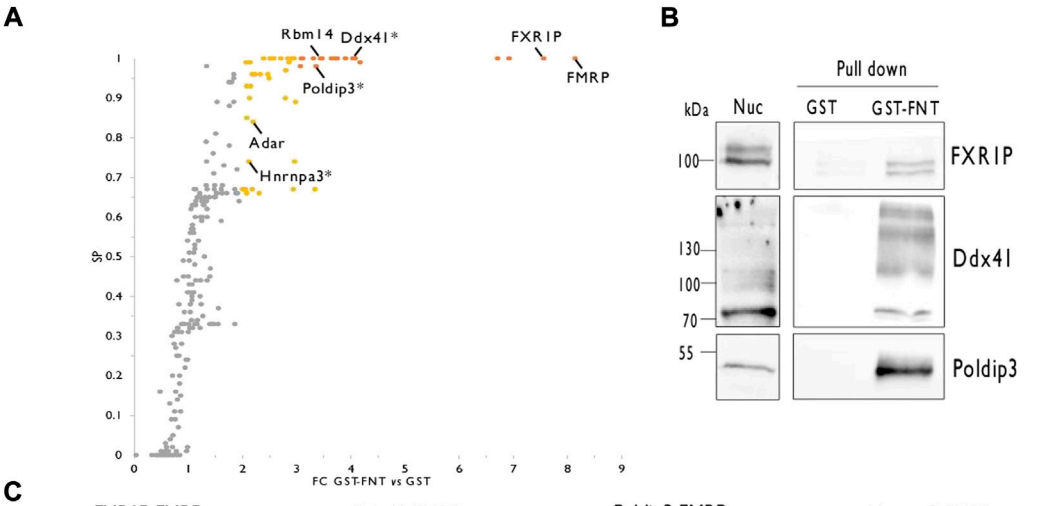


FIGURE 2
Characterization of the PND14 rat forebrain nuclear proteome. **(A)**. Venn diagram showing the overlap between the two protein sets identified by mass spectrometry analysis of two independent preparations of nuclear extracts from PND14 rat forebrains. Details are available in Supplementary Table S1. **(B)**. The nuclear proteome dataset was subjected to enrichment analysis for GO Cellular Components, GO Molecular Functions and Reactome pathways against the rat proteome. Categories were classified according to the−Log10(FDR). Details are available in Supplementary Table S2.

both FMRP and FXR1P interacts with ADAR1 in HeLa cells to positively and negatively regulate A to I RNA editing, respectively. They proposed that the proteins may contribute to the alterations of RNA editing that they observed in the post-mortem brain of ASD patients (Tran, et al., 2019). In cancer cells, FXR1P was involved in the recruitment of transcription factors to gene promotors (Fan et al., 2017). We first performed GST-FNT pull down experiments on nuclear fraction from PND14 rat forebrain followed by FXR1P detection by western blotting and found a specific binding of the protein with GST-FNT while it was barely detected in the GST control lane (Figure 3B). Next, to assess the interaction of FXR1P with endogenous FMRP in the



D

		FMRP/FXR1P	FMRP/Ddx41	FMRP/Poldip3	FMRP/Hnrnpa3
Pourcentage with IC95%	No staining	3.7 ± 1.8%	6.8 ± 2.5%	8.6 ± 2.9%	2.0 ± 1.4%
	Cytoplasmic staining	48.6 ± 4.8%	11.7± 3.2%	10.8± 3.2%	5.9 ± 2.3%
	Nuclear staining	3.7 ± 1.8%	30.1 ± 4.5%	34.4 ± 4.9%	39.2 ± 4.8%
	Nuclear and cytoplasmic staining	43.9 ± 4.8%	51.5 ± 1.8%	46.2 ± 5.2%	52.9 ± 4.9%

FIGURE 3
Identification of rat forebrain nuclear FMRP-interacting proteins. **(A)**. Volcano plot highlighting proteins differentially co-purified in triplicate pull-down experiments using the GST-FNT recombinant protein versus the GST alone. Statistically significant differences were assessed using the Saint-Express tool. Unenriched proteins are displayed in gray, while enriched proteins presenting a FC > 3.00 and SP > 0.7 are displayed in orange, and were classified as "high-confidence" interactor. Yellow dots corresponding to all other significant proteins with FC > 2.0 and SP > 0.5, and were referred as "medium-confidence" interactor. Know and novel (*) FMRP interacting proteins are noted in the scatter plot. **(B)**. Immunoblot analysis of identified FMRP-interacting proteins. Pull down experiments were conducted with GST-FNT and GST (negative control) immobilized on Glutathione Sepharose with nuclear fractions. Protein retained on the beads were resolved by SDS-PAGE and processed for western blotting using antibodies against FXR1P, Poldip3 or Ddx41. **(C)**. Representative confocal images of the interaction between endogenous FXR1P, Ddx41, Poldip3 or
(*Continued*)

FIGURE 3 (Continued)
Hnrnpa3, and FMRP detected by Proximity Ligation Assay (PLA) in rat primary hippocampal neuron cultures. Neuronal cells were identified upon their MAP2 labelling (not shown). **(D)**. Percentage of neurons presenting no PLA dots (No staining), PLA dots exclusively in the cytoplasm (Cytoplasmic staining), exclusively in the nucleus (Nuclear staining) or in both the cytoplasm, and the nucleus (Cytoplasmic and nuclear staining) for the indicated interactions. The percentages with the 95% Confidence Interval were calculated from 107 (FMRP/FXR1P), 103 (FMRP/Ddx41), 93 (FMRP/Poldip3), and 103 (FMRP/Hnrnpa3) neurons in culture processed from three independent experiments.

nucleus of neurons, we conducted Proximity Ligation Assays (PLA) in cultured hippocampal rat neurons at 21 days *in vitro* (DIV). Consistent with the known interaction between both proteins in the cytoplasm, the PLA for FMRP and FXR1P showed many cytoplasmic dots in neurons (Figures 3C,D). No PLA dots were detected as background in the absence of primary antibodies (Supplementary Figure S3). Besides, some dots were also detected in the nucleus, thus confirming the association of the two proteins in the nuclear compartment of neurons. It is very interesting to underline that the profile of the PLA labelling is heterogeneous, with almost 44% of the neurons presenting dots in both the nucleus and the cytoplasm, 4% in the nucleus only and about 48% in the cytoplasm only (Figure 3D). This diversity suggests that the interaction between FMRP and FXR1P in the nucleus may depend on the neuronal cell types or their levels of activity for example.

Ddx41 is a multi-functional DEAD box helicase both nuclear and cytoplasmic. It is involved in pre-mRNA splicing and counteracts the accumulation of R-loops in promoter regions of active genes thus participating in genome stability. The protein also plays a role as a cytosolic DNA-sensor in DNA-mediated innate immunity (Andreou 2021; Mosler et al., 2021). GST-FNT pull down on nuclear fractions from rat forebrain followed by western blotting showed a specific binding between the N-terminal domain of FMRP and Ddx41 (Figure 3B) whereas PLA reveals an interaction between the endogenous proteins in the nucleus and the cytoplasm of neurons (Figures 3C,D). Nonetheless, as in the case for the FMRP-FXR1P interaction, the profile of the labelling is heterogeneous (Figure 3D) suggesting that the association Ddx41-FMRP may depends on the neuronal status.

Poldip3 (Polymerase δ-interacting protein 3), also called PDIP46 (46 kDa Polymerase δ -interacting protein) or SKAR (S6K1 Aly/REF-like target) is distributed in both the nucleus and the cytoplasm. Poldip3 is involved in cellular DNA replication and genome stability (Wang et al., 2016; Bjorkman et al., 2020). In addition, the protein is a member of the exon-exon junction complex and recruits the SK6 kinase on newly synthesized mRNA, which will later enhance the pioneer round of translation of spliced mRNA (Ma et al., 2008). Lastly, by interacting with the TREX complex, Poldip3 would participate in the nuclear export of mRNA (Folco et al., 2012). As illustrated Figures 3B,C, an interaction between FMRP and Poldip3 is detected by pull down assay and PLA. Images from PLA show an association of the endogenous proteins either in the nucleus or in both the nucleus and the cytoplasm of neurons (Figures 3C,D), with pattern consisting in 2–4 dots in the nucleus and also few spots in the cytoplasm (Figure 3C).

Lastly, we analyzed by PLA the interaction of FMRP with a third novel interactor, Hnrnpa3 (Heterogeneous nuclear ribonucleoprotein A3). In the nucleus, Hnrnpa3 recognizes single-stranded telomeric DNA and is involved in telomere maintenance (Huang et al., 2010). Besides, this protein is also implicated in the splicing, the stability and the cytoplasmic trafficking of mRNAs (Ma et al., 2002; Papadopoulou et al., 2012; Kwon et al., 2021). We showed here that consistent with the pull down screen, endogenous FMRP and Hnrnpa3 interact in neurons, mainly in the nucleus only or in both the nucleus and the cytoplasm (Figures 3C,D).

Altogether, these target specific analyses further validated the pull down approach on nuclear-enriched fractions from PND14 rat forebrains to identify a nuclear interactome of FMRP.

Functional categorization of the nuclear interactome of FMRP

We identified 55 FMRP-interacting protein candidates. To assess the biological meaning of this nuclear interactome, we performed annotation analyses for UniProt Key Words (UP-KW) on the reviewed human homologues of our rat dataset (Figure 4A; Supplementary Table S4). Consistent with our strategy based on affinity purification on nuclear enriched extracts, 47 proteins are associated with the UP-KW for Cellular Component “nucleus”, i.e., 85% of the list, with 17 proteins (30%) presenting both a nuclear and a cytoplasmic localization (Supplementary Table S4). Moreover, “RNA-binding”, “DNA-binding,” and “Chromatin regulator” appeared among the highest UP-KW for Molecular Functions. The annotation for UP-KW linked to Biological Processes highlighted among the most representative terms the involvement of the FMRP partners in several steps of mRNA biogenesis as “transcription”, “mRNA splicing” or “transport”, besides to “DNA damage”. Consistently, a network analysis based on protein association to physical complexes revealed clusters associated to histone acetylation, transcriptional regulation, mRNA metabolism, mRNA export, translational regulation as well as protein de-phosphorylation and ribosome biogenesis (Figure 4B). Lastly, UP-KW annotation for Diseases respectively associated 10 and 3 proteins to the “Mental

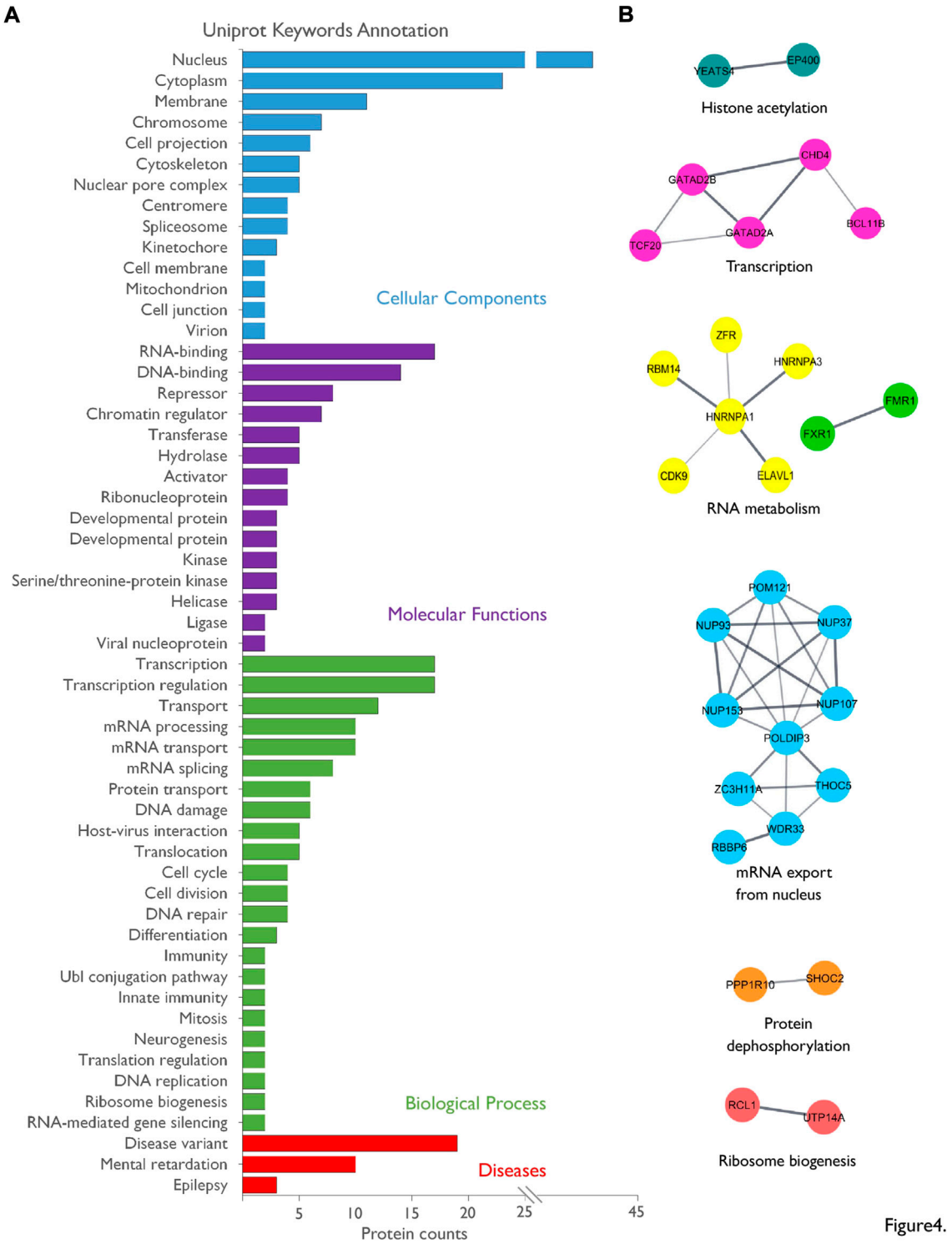


FIGURE 4
Bio-informatics analysis of the nuclear interactome of FMRP. **(A)**. The list of the nuclear FMRP-interacting proteins was subjected to Uniprot Keywords (UP-KW) annotation analysis using the reviewed human homologues as data source: UP-KW for Cellular Components, UP-KW for Molecular functions, UP-KW for Biological process, and UP-KW for Diseases. Terms were ranked according to the number of proteins in the category. Details are available in Supplementary Table S4. **(B)** A network was built on physical interaction to complexes predicted by STRING using all identified FMRP-interacting proteins candidates as source, with an interaction confidence of 0.4 or greater and based on databases and experiments sources. The STRING network was imported into the Cytoscape application and clusters were created and annotated through the StringApp plug-in.

retardation" and "Epilepsy" categories, which fully correlates with phenotypes associated with the FXS (Figure 4A).

Discussion

While, in the past, the nuclear presence of FMRP was mainly associated with the its role in RNA export from nucleus to cytoplasm, now increasing evidence strongly suggest the implication of FMRP in nuclear processes. However, the molecular mechanisms underlying these nuclear functions notably in neurons are currently missing. In the present study, we identified 55 protein candidates interacting with of the N-terminal domain of FMRP from nuclear PND14 rat forebrain extracts. A bibliographic analysis revealed that five proteins identified in this dataset, FMRP itself (Ramos, et al., 2006; Hu, et al., 2015; Myrick, et al., 2015), FXR1P (Fragile X Related Protein 1) (Zhang, et al., 1995), the zinc finger RNA-binding protein ZFR (Worringer, et al., 2009), the splicing factor Rbm14 (Zhou, et al., 2017), and the mRNA editing enzyme Adar (Shamay-Ramot, et al., 2015; Filippini, et al., 2017), were previously reported to associate with FMRP in target specific studies. Besides, comparing our dataset with two published proteomic screenings using affinity pull-down with the N-terminal domain of FMRP revealed five proteins in common: FMRP, FXR1P, the chromatin-remodeling factor CHD4 and the transcriptional factors TCF20 and ZNF638. This narrow overlap may largely be explained by distinct experimental conditions, notably the use of HEK 293 cell line extracts and/or the absence of a nuclear enrichment in the earlier proteomic approaches (He and Ge 2017; Taha, et al., 2021). Besides, we could not find the Nuclear FMRP Interacting Protein 1 (NuFIP1), known to interact with N-terminal domain of FMRP (Bardoni et al., 1999), neither in the list of candidate nuclear partners, nor in the nuclear proteome. Yet, NuFIP1 is a nucleocytoplasmic shuttling protein (Bardoni et al., 2003) and its expression in time and space in the brain is not known. Its absence in our datasets may be explain by a low level of expression and/or a predominantly cytoplasmic localization in the PND14 rat forebrain. In the future, it will be very informative to complement the present results with pull-down experiments performed on nuclear extracts prepared from the forebrain or brain sub-structures of rats of different ages, in order to access the variation of the FMRP nuclear interactome across brain regions and over the life span.

In the present study, we found that FXR1P, known to interact with FMRP in the cytoplasm, also binds the protein in the nucleus. In addition, we experimentally validated three novel interactors, Ddx41, Poldip3, and Hnrnpa3. It should be noted here the identification of protein–protein interactions by pull down affinity purification may, as co-immunoprecipitation, result from direct physical interactions but also from indirect interactions. Besides, in the case of GST-pull down experiments, unspecific binding may result from the association of some proteins to the beads or with the GST tag. To identify and discard these background proteins, we used a pull down condition with GST alone cross-linked to the glutathione beads as negative control. In addition, to minimize indirect binding, we introduced a washing step with 500 mM NaCl prior to elution. However, some of the interactions we detected may still arise from indirect protein/protein or protein/nucleic acid binding with direct and specific preys. Nonetheless, whether FMRP interacts directly with the identified candidates or in the context of ribonucleoprotein complexes, our dataset clearly reflects an association of FMRP with nuclear machineries, notably with those involved in pre-mRNAs biogenesis. Moreover, our data indicate that the profile of the endogenous interaction between FMRP and its nuclear partners in neurons may vary, likely reflecting differences in cell characteristics such as neuronal cell type or activity.

The identification of the biological pathways associated with our dataset as well as the network analysis match with the nuclear functions reported to be modulated by FMRP, including mRNA editing (Bhogal, et al., 2011; Shamay-Ramot, et al., 2015; Filippini, et al., 2017; Tran, et al., 2019), mRNA nuclear export (Kim, et al., 2009; Edens, et al., 2019; Hsu, et al., 2019; Westmark, et al., 2020; Kim, et al., 2021) or DNA damage response (Alpatov, et al., 2014; Zhang, et al., 2014; Chakraborty, et al., 2020). Our work thus provides valuable insights and a foundation for future mechanistic investigations of the overlooked nuclear functions of FMRP in neurons. Interestingly, we detected an interaction in the nucleus of neurons between FMRP and FXR1P, which positively and negatively regulate ADAR-mediated RNA editing, respectively. Whether and how this interaction may modulate the modification of their respective or common RNA targets will need to be further explored.

Transcription and splicing recently emerged as a key nuclear process regulated by FMRP. Indeed, large-scale screens identified splice site variations between mouse or drosophila models of FXS and their wild-type controls (Brooks et al., 2015; Shah et al., 2020). It has been proposed that these alterations result from the dysregulated translation of FMRP target mRNAs encoding chromatin modifying enzymes (Shah, et al., 2020; Hale, et al., 2021) that would alter the profile of histones post-translational modifications and in turn, mRNA splicing (Shah, et al., 2020). Similarly, other epigenetic changes would affect transcriptional activation (Korb et al., 2017). Interestingly, the current dataset presents several factors involved in chromatin remodeling, transcriptional regulation or mRNA splicing. Consistent with these results, Kim, et al. (2009) detected an association between FMRP and nascent transcripts on lampbrush chromosomes in amphibian oocytes. In another context, Alpatov and colleagues showed in mouse embryonic fibroblasts an association of FMRP with chromatin fraction, enhanced by stress replication (Alpatov, et al., 2014). Our list includes the reported FMRP-interacting

protein Rbm14 (Zhou, et al., 2017), the helicase Ddx41 (Figure 3) and Hnrnpa3 (Figure 3), all connected to mRNA splicing. These results are in agreement with studies describing a direct role of FMRP in the splicing of some mRNAs including its own mRNA (Didiot et al., 2008; Zhou, et al., 2017) and complement to the indirect effects outlined above. In the same way, FMRP could directly participate in gene transcription and/or mRNA export, for example through interaction with Poldip3 (Figure 3).

Many studies have shown that the processes involved in mRNA metabolism, from transcription to nuclear export and cytoplasmic trafficking, are tightly coupled (Vitaliano-Prunier et al., 2012; Bjork and Wieslander 2017; Woodward et al., 2017; Garland and Jensen 2020). This functional coordination results from physical interactions between members the different mRNA biogenesis machineries, which synchronize or cross-stimulate the connected processes. In this line of view, FMRP may act as a hub protein that could follow, partially or completely, the fate of its target mRNAs. Interestingly, some studies indicate that other neuronal mRNA-binding protein such as ELAV or SMN proteins, involved in mRNA dendritic transport in neurons, transit through the nucleus and participate in early post-transcriptional regulatory events such as pre-mRNA splicing or poly-adenylation (Colombrita et al., 2013; Raimer et al., 2017; Ravanidis et al., 2018; Jung and Lee 2021; Wei and Lai 2022). We propose here that the workflow we set up, combining an efficient nuclear fractionation from rat forebrain coupled to affinity pull down followed by identification using tandem mass spectrometry, could be adapted to assess the nuclear interactome of these dual-distributed but predominantly cytoplasmic neuronal factors.

Lastly, annotation analysis indicates that twenty-one of the FMRP-interacting proteins identified in this work present one or more genetic variants involved in a disease and/or are connected to "mental retardation" or "epilepsy". As an example, identification of *de novo* mutations in patients linked the FMRP partner candidates Tbl1xr1 and Chd4 to autism spectrum disorders (Coe et al., 2019; O'Roak et al., 2012; Quan et al., 2020). Strikingly, about 50% of the FXS males and 20% of the FXS females meet the criteria for ASD (Kaufmann et al., 2017). More generally, several studies have highlighted links between RNA metabolism and neurodevelopmental and neurological diseases (Chatterjee et al., 2021), with shared molecular pathways between various disorders. We believe that our proteomic screen, clearly associating FMRP with mRNA biogenesis, may provide leads to further explore these molecular links.

Data availability statement

The mass spectrometry proteomics data presented in this study have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository (https://www.ebi.ac.uk/pride/archive/) with the dataset identifier PXD034157.

Ethics statement

The animal study was reviewed and approved by Animal Care and Ethics Committee (APAFIS #18647-2019011110552947v3).

Author contributions

FK and FH performed the biochemical experiments with the help of MP, IL, and GP. FK and CG performed the bioinformatics. A-SG and DD performed the Mass Spectrometry analysis. FK, BB, and CG contributed to hypothesis development and data interpretation. CG provided the overall supervision. CG and SM provided human resources. CG and SM provided the funding. CG wrote the original draft. FK, SM, BB, and CG edited the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

We gratefully acknowledge the 'Fondation pour la Recherche Médicale' for the PhD fellowship (#ECO201906008982) and the and 4th year PhD fellowship to MP (#FDT202012010480) and the Academy of Excellence 4 "Complexity and Diversity of the Living Systems" of IDEX UCA-Jedi (ANR-15-IDEX-01) for the Master fellowship to FK. We also thank the "Fondation Jérôme Lejeune" (Session 2017a – Project #1643) for financial support to CG and the 'Agence Nationale de la Recherche' (ANR-20-CE16-0006-01) to SM. This work was also supported in part by a grant from FRAXA Research Foundation to SM and grants from the French government *via* the "Investments for the Future" LabEx "SIGNALIFE" (ANR-11-LABX-0028-01) as well as the CG06 (AAP santé), the GIS IBiSA and the CSI funding program of Université Côte d'Azur for the IPMC Proteomic/Lipidomic platform.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.954087/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Detection of Nopp140, Fibrillarin, and Histone H4 in the total lysate from PND14 rat forebrain. To appreciate the levels of the nuclear markers Nopp140, Fibrillarin, and Histone H4 in the total lysate of PND14 rat forebrain compared to the nuclear fraction, 10, 20, and 50 µg of proteins from the total lysate (Tot) or 10 µg of proteins from the nuclear fraction (Nuc) were subjected to immunoblotting using the indicated antibodies and then the appropriate horseradish peroxidase (HRP)-conjugated secondary antibodies. Proteins were visualized using the Enhanced chemiluminescence western blot kit from Millipore (ECL) or the SuperSignal West Femto Maximum Sensitivity substrate from Thermo Scientific (Ultra sensitive ECL) at the indicated time of acquisition. Our results indicate that the visualization of the nuclear markers in the total fraction compared to their detection in the nuclear lysate requires higher amount of proteins, associated with longer time of exposure and/or ultra sensitive ECL substrate, thus reflecting a clear enrichment of the nuclear fraction in nuclear proteins.

SUPPLEMENTARY FIGURE S2
Description of the GST-FNT bait. High panel: Schematic diagram of the human N-terminal fragment (1-213aa) fused to the GST tag used in this study (GST-FNT). Major domains (Ag1, Ag2, KH0) and the nuclear localization signal (NLS) of FMRP are highlighted. Low panel: Coomassie stained gel of the purified GST and GST-FNT recombinant proteins. Black arrow indicates GST and red arrow indicates GST-FNT.

SUPPLEMENTARY FIGURE S3
Negative control for Proximity Ligation Assays (PLA). Representative confocal images of primary hippocampal neurons in culture processed with the PLA Duo-link® kit (Sigma-Aldrich) without primary antibodies (Scale bar = 50 µm). Enlargement of delimited area is also shown (Scale bar = 10 µm). No PLA dots could be detected as background signal.

SUPPLEMENTARY TABLE 1
PND14 rat forebrain nuclear proteome. 10 µg of proteins from two independent nuclear preparations were separated on short gradient SDS-PAGE. Gels were then sliced into two bands per lane, subjected to in gel Trypsin digestion and analyzed by LC-MS/MS. Protein identification from raw data was performed using Max Quant v1.5.5.1 with 1% FDR for both peptides and proteins. Proteins detected in the two replicates were selected to establish the reference PND14 rat forebrain nuclear proteome.

SUPPLEMENTARY TABLE 2
Enrichment analysis of the PND14 rat forebrain nuclear proteome. Details for enrichment analysis of the nuclear protein dataset against Rattus norvegicus proteome. for GO Cellular Component, GO Molecular Function terms or GO Biological Process terms and REACTOME pathways using the DAVID webtool.

SUPPLEMENTARY TABLE 3
Nuclear FMRP interactome. Proteins eluted from three independent pull down assays using as GST-FNT recombinant protein as the bait or GST as the negative control, were separated on gradient SDS-PAGE, subjected to in gel Trypsin digestion and analyzed by LC-MS/MS. Protein identification from raw data was performed using MaxQuant v1.5.5.1 with 1% FDR for both peptides and proteins. Proteins with no missing value (MS/MS≠0) in GST-FNT were selected and used for differential statistical analysis using the SAINTexpress tool. Proteins with a cut off of FC >2 and SP>0.5 were significant and protein with a cut off of FC>3 and SP>0.7 were considered as high confidence interactor.

SUPPLEMENTARY TABLE 4
Annotation analysis of the nuclear FMRP interacting protein dataset. Details for UniProt Keyword (UP-KW) annotation analysis of the human homologs of the nuclear FMRP interacting protein dataset using the DAVID webtool for UP-KW Cellular Component, UP KW Molecular Function, UP-KW Biological Process terms or UP-KW Disease.

References

Agote-Aran, A., Schmucker, S., Jerabkova, K., Jmel Boyer, I., Berto, A., Pacini, L., et al. (2020). Spatial control of nucleoporin condensation by fragile X-related proteins. *EMBO J.* 39, e104467. doi:10.15252/embj.2020104467

Alpatov, R., Lesch, B. J., Nakamoto-Kinoshita, M., Blanco, A., Chen, S., Stutzer, A., et al. (2014). A chromatin-dependent role of the fragile X mental retardation protein FMRP in the DNA damage response. *Cell* 157, 869–881. doi:10.1016/j.cell.2014.03.040

Andreou, A. Z. (2021). DDX41: A multifunctional DEAD-box protein involved in pre-mRNA splicing and innate immunity. *Biol. Chem.* 27 (402), 645–651. doi:10.1515/hsz-2020-0367

Ascano, M., Jr., Mukherjee, N., Bandaru, P., Miller, J. B., Nusbaum, J. D., Corcoran, D. L., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 492, 382–386. doi:10.1038/nature11737

Bakker, C. E., de Diego Otero, Y., Bontekoe, C., Raghoe, P., Luteijn, T., Hoogeveen, A. T., et al. (2000). Immunocytochemical and biochemical characterization of FMRP, FXR1P, and FXR2P in the mouse. *Exp. Cell Res.* 10, 162–170. doi:10.1006/excr.2000.4932

Bardoni, B., Schenck, A., and Mandel, J. L. (1999). A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. *Hum. Mol. Genet.* 8, 2557–2566. doi:10.1093/hmg/8.13.2557

Bardoni, B., Sittler, A., Shen, Y., and Mandel, J. L. (1997). Analysis of domains affecting intracellular localization of the FMRP protein. *Neurobiol. Dis.* 4, 329–336. doi:10.1006/nbdi.1997.0142

Bardoni, B., Willemsen, R., Weiler, I. J., Schenck, A., Severijnen, L. A., Hindelang, C., et al. (2003). NUFIP1 (nuclear FMRP interacting protein 1) is a nucleocytoplasmic shuttling protein associated with active synaptoneurosomes. *Exp. Cell Res.* 289, 95–107. doi:10.1016/s0014-4827(03)00222-2

Bechara, E. G., Didiot, M. C., Melko, M., Davidovic, L., Bensaid, M., Martin, P., et al. (2009). A novel function for fragile X mental retardation protein in translational activation. *PLoS Biol.* 7, e16. doi:10.1371/journal.pbio.1000016

Bhogal, B., Jepson, J. E., Savva, Y. A., Pepper, A. S., Reenan, R. A., and Jongens, T. A. (2011). Modulation of dADAR-dependent RNA editing by the Drosophila fragile X mental retardation protein. *Nat. Neurosci.* 14, 1517–1524. doi:10.1038/nn.2950

Bjork, P., and Wieslander, L. (2017). Integration of mRNP formation and export. *Cell. Mol. Life Sci.* 74, 2875–2897. doi:10.1007/s00018-017-2503-3

Bjorkman, A., Johansen, S. L., Lin, L., Schertzer, M., Kanellis, D. C., Katsori, A. M., et al. (2020). Human RTEL1 associates with Poldip3 to facilitate responses to replication stress and R-loop resolution. *Genes Dev.* 34, 1065–1074. doi:10.1101/gad.330050.119

Bonaccorso, C. M., Spatuzza, M., Di Marco, B., Gloria, A., Barrancotto, G., Cupo, A., et al. (2015). Fragile X mental retardation protein (FMRP) interacting proteins exhibit different expression patterns during development. *Int. J. Dev. Neurosci.* 42, 15–23. doi:10.1016/j.ijdevneu.2015.02.004

Brooks, A. N., Duff, M. O., May, G., Yang, L., Bolisetty, M., Landolin, J., et al. (2015). Regulation of alternative splicing in Drosophila by 56 RNA binding proteins. *Genome Res.* 25, 1771–1780. doi:10.1101/gr.192518.115

Chakraborty, A., Jenjaroenpun, P., Li, J., El Hilali, S., McCulley, A., Haarer, B., et al. (2020). Replication stress induces global chromosome breakage in the fragile X genome. *Cell Rep.* 32, 108179. doi:10.1016/j.celrep.2020.108179

Chatterjee, B., Shen, C. J., and Majumder, P. (2021). RNA modifications and RNA metabolism in neurological disease pathogenesis. *Int. J. Mol. Sci.* 1, 11870. doi:10.3390/ijms222111870

Coe, B. P., Stessman, H. A. F., Sulovari, A., Geisheker, M. R., Bakken, T. E., Lake, A. M., et al. (2019). Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat. Genet.* 51, 106–116. doi:10.1038/s41588-018-0288-4

Colombrita, C., Silani, V., and Ratti, A. (2013). ELAV proteins along evolution: Back to the nucleus? *Mol. Cell. Neurosci.* 56, 447–455. doi:10.1016/j.mcn.2013.02.003

D'Souza, M. N., Gowda, N. K. C., Tiwari, V., Babu, R. O., Anand, P., Dastidar, S. G., et al. (2018). FMRP interacts with C/D box snoRNA in the nucleus and regulates ribosomal RNA methylation. *iScience* 30 (9), 399–411. doi:10.1016/j.isci.2018.11.007

Dahlhaus, R. (2018). Of men and mice: Modeling the fragile X syndrome. *Front. Mol. Neurosci.* 11, 41. doi:10.3389/fnmol.2018.00041

Darnell, J. C., Fraser, C. E., Mostovetsky, O., and Darnell, R. B. (2009). Discrimination of common and unique RNA-binding activities among Fragile X mental retardation protein paralogs. *Hum. Mol. Genet.* 18, 3164–3177. doi:10.1093/hmg/ddp255

Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y., Mele, A., Fraser, C. E., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 22 (146), 247–261. doi:10.1016/j.cell.2011.06.013

Davis, J. K., and Broadie, K. (2017). Multifarious functions of the fragile X mental retardation protein. *Trends Genet.* 33, 703–714. doi:10.1016/j.tig.2017.07.008

Didiot, M. C., Subramanian, M., Flatter, E., Mandel, J. L., and Moine, H. (2009). Cells lacking the fragile X mental retardation protein (FMRP) have normal RISC activity but exhibit altered stress granule assembly. *Mol. Biol. Cell* 20, 428–437. doi:10.1091/mbc.e08-07-0737

Didiot, M. C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J. L., and Moine, H. (2008). The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.* 36, 4902–4912. doi:10.1093/nar/gkn472

Eberhart, D. E., Malter, H. E., Feng, Y., and Warren, S. T. (1996). The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum. Mol. Genet.* 5, 1083–1091. doi:10.1093/hmg/5.8.1083

Edens, B. M., Vissers, C., Su, J., Arumugam, S., Xu, Z., Shi, H., et al. (2019). FMRP modulates neural differentiation through m(6)a-dependent mRNA nuclear export. *Cell Rep.* 23 (28), 845–854. e845. doi:10.1016/j.celrep.2019.06.072

Fan, Y., Yue, J., Xiao, M., Han-Zhang, H., Wang, Y. V., Ma, C., et al. (2017). FXR1 regulates transcription and is required for growth of human cancer cells with TP53/FXR2 homozygous deletion. *Elife* 6, e26129. doi:10.7554/eLife.26129

Feng, Y., Gutekunst, C. A., Eberhart, D. E., Yi, H., Warren, S. T., and Hersch, S. M. (1997). Fragile X mental retardation protein: Nucleocytoplasmic shuttling and association with somatodendritic ribosomes. *J. Neurosci.* 1 (17), 1539–1547. doi:10.1523/jneurosci.17-05-01539.1997

Filippini, A., Bonini, D., Lacoux, C., Pacini, L., Zingariello, M., Sancillo, L., et al. (2017). Absence of the Fragile X Mental Retardation Protein results in defects of RNA editing of neuronal mRNAs in mouse. *RNA Biol.* 2 (14), 1580–1591. doi:10.1080/15476286.2017.1338232

Folco, E. G., Lee, C. S., Dufu, K., Yamazaki, T., and Reed, R. (2012). The proteins PDIP3 and ZC11A associate with the human TREX complex in an ATP-dependent manner and function in mRNA export. *PLoS One* 7, e43804. doi:10.1371/journal.pone.0043804

Fridell, R. A., Benson, R. E., Hua, J., Bogerd, H. P., and Cullen, B. R. (1996). A nuclear role for the Fragile X mental retardation protein. *EMBO J.* 15, 5408–5414. doi:10.1002/j.1460-2075.1996.tb00924.x

Garland, W., and Jensen, T. H. (2020). Nuclear sorting of RNA. *Wiley Interdiscip. Rev. RNA* 11, e1572. doi:10.1002/wrna.1572

George, J., Li, Y., Kadamberi, I. P., Parashar, D., Tsaih, S. W., Gupta, P., et al. (2021). RNA-binding protein FXR1 drives cMYC translation by recruiting eIF4F complex to the translation start site. *Cell Rep.* 37, 109934. doi:10.1016/j.celrep.2021.109934

Guard, S. E., Ebmeier, C. C., and Old, W. M. (2019). Label-free immunoprecipitation mass spectrometry workflow for large-scale nuclear interactome profiling. *J. Vis. Exp.* 17, 1. doi:10.3791/60432

Hale, C. R., Sawicka, K., Mora, K., Fak, J. J., Kang, J. J., Cutrim, P., et al. (2021). FMRP regulates mRNAs encoding distinct functions in the cell body and dendrites of CA1 pyramidal neurons. *Elife* 23, e71892. doi:10.7554/eLife.71892

He, Q., and Ge, W. (2017). The tandem Agenet domain of fragile X mental retardation protein interacts with FUS. *Sci. Rep.* 19 (7), 962. doi:10.1038/s41598-017-01175-8

Herring, J., Johnson, K., and Richstein, J. (2022). The use of "retardation" in FRAXA, FMRP, FMR1 and other designations. *Cells* 11, 1044. doi:10.3390/cells11061044

Hsu, P. J., Shi, H., Zhu, A. C., Lu, Z., Miller, N., Edens, B. M., et al. (2019). The RNA-binding protein FMRP facilitates the nuclear export of N (6)-methyladenosine-containing mRNAs. *J. Biol. Chem.* 294, 19889–19895. doi:10.1074/jbc.AC119.010078

Hu, Y., Chen, Z., Fu, Y., He, Q., Jiang, L., Zheng, J., et al. (2015). The amino-terminal structure of human fragile X mental retardation protein obtained using precipitant-immobilized imprinted polymers. *Nat. Commun.* 23 (6), 6634. doi:10.1038/ncomms7634

Huang, P. R., Hung, S. C., and Wang, T. C. (2010). Telomeric DNA-binding activities of heterogeneous nuclear ribonucleoprotein A3 *in vitro* and *in vivo*. *Biochim. Biophys. Acta* 1803, 1164–1174. doi:10.1016/j.bbamcr.2010.06.003

Jung, M., and Lee, E. K. (2021). RNA-binding protein HuD as a versatile factor in neuronal and non-neuronal systems. *Biol. (Basel)* 10, 361. doi:10.3390/biology10050361

Kaufmann, W. E., Kidd, S. A., Andrews, H. F., Budimirovic, D. B., Esler, A., Haas-Givler, B., et al. (2017). Autism spectrum disorder in fragile X syndrome: Cooccurring conditions and current treatment. *Pediatrics* 139, S194–S206. doi:10.1542/peds.2016-1159F

Khayachi, A., Gwizdek, C., Poupon, G., Alcor, D., Chafai, M., Casse, F., et al. (2018). Sumoylation regulates FMRP-mediated dendritic spine elimination and maturation. *Nat. Commun.* 9, 757. doi:10.1038/s41467-018-03222-y

Khlghatyan, J., and Beaulieu, J. M. (2018). Are FXR family proteins integrators of dopamine signaling and glutamatergic neurotransmission in mental illnesses? *Front. Synaptic Neurosci.* 10, 22. doi:10.3389/fnsyn.2018.00022

Kim, G. W., Imam, H., and Siddiqui, A. (2021). The RNA binding proteins YTHDC1 and FMRP regulate the nuclear export of N(6)-methyladenosine-modified hepatitis B virus transcripts and affect the viral life cycle. *J. Virol.* 5, e0009721. doi:10.1128/JVI.00097-21

Kim, M., Bellini, M., and Ceman, S. (2009). Fragile X mental retardation protein FMRP binds mRNAs in the nucleus. *Mol. Cell. Biol.* 29, 214–228. doi:10.1128/MCB.01377-08

Korb, E., Herre, M., Zucker-Scharff, I., Gresack, J., Allis, C. D., and Darnell, R. B. (2017). Excess translation of epigenetic regulators contributes to fragile X syndrome and is alleviated by Brd4 inhibition. *Cell* 170, 1209–1223. doi:10.1016/j.cell.2017.07.033

Kwon, S. M., Min, S., Jeoun, U. W., Sim, M. S., Jung, G. H., Hong, S. M., et al. (2021). Global spliceosome activity regulates entry into cellular senescence. *FASEB J.* 35, e21204. doi:10.1096/fj.202000395RR

Ma, A. S., Moran-Jones, K., Shan, J., Munro, T. P., Snee, M. J., Hoek, K. S., et al. (2002). Heterogeneous nuclear ribonucleoprotein A3, a novel RNA trafficking response element-binding protein. *J. Biol. Chem.* 277, 18010–18020. doi:10.1074/jbc.M200050200

Ma, X. M., Yoon, S. O., Richardson, C. J., Julich, K., and Blenis, J. (2008). SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* 133, 303–313. doi:10.1016/j.cell.2008.02.031

Mangano, G. D., Fontana, A., Salpietro, V., Antona, V., Mangano, G. R., and Nardello, R. (2022). Recurrent missense variant in the nuclear export signal of FMR1 associated with FXS-like phenotype including intellectual disability, ASD, facial abnormalities. *Eur. J. Med. Genet.* 65, 104441. doi:10.1016/j.ejmg.2022.104441

Maurin, T., and Bardoni, B. (2018). Fragile X mental retardation protein: To Be or not to Be a translational enhancer. *Front. Mol. Biosci.* 5, 113. doi:10.3389/fmolb.2018.00113

Maurin, T., Lebrigand, K., Castagnola, S., Paquet, A., Jarjat, M., Popa, A., et al. (2018). HITS-CLIP in various brain areas reveals new targets and new modalities of RNA binding by fragile X mental retardation protein. *Nucleic Acids Res.* 46, 6344–6355. doi:10.1093/nar/gky267

Maurin, T., Melko, M., Abekhoukh, S., Khalfallah, O., Davidovic, L., Jarjat, M., et al. (2015). The FMRP/GRK4 mRNA interaction uncovers a new mode of binding of the Fragile X mental retardation protein in cerebellum. *Nucleic Acids Res.* 43, 8540–8550. doi:10.1093/nar/gkv801

Maurin, T., Zongaro, S., and Bardoni, B. (2014). Fragile X Syndrome: From molecular pathology to therapy. *Neurosci. Biobehav Rev.* 46, 242–255. doi:10.1016/j.neubiorev.2014.01.006

Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J. P., St-Denis, N. A., Li, T., et al. (2013). The CRAPome: A contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* 10, 730–736. doi:10.1038/nmeth.2557

Mosler, T., Conte, F., Longo, G. M. C., Mikicic, I., Kreim, N., Mockel, M. M., et al. (2021). R-loop proximity proteomics identifies a role of DDX41 in transcription-associated genomic instability. *Nat. Commun.* 12, 7314. doi:10.1038/s41467-021-27530-y

Myrick, L. K., Hashimoto, H., Cheng, X., and Warren, S. T. (2015). Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. *Hum. Mol. Genet.* 24, 1733–1740. doi:10.1093/hmg/ddu586

O'Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338, 1619–1622. doi:10.1126/science.1227764

Papadopoulou, C., Boukakis, G., Ganou, V., Patrinou-Georgoula, M., and Guialis, A. (2012). Expression profile and interactions of hnRNP A3 within hnRNP/mRNP complexes in mammals. *Arch. Biochem. Biophys.* 523, 151–160. doi:10.1016/j.abb.2012.04.012

Prieto, M., Folci, A., Poupon, G., Schiavi, S., Buzzelli, V., Pronot, M., et al. (2021). Missense mutation of Fmr1 results in impaired AMPAR-mediated plasticity and socio-cognitive deficits in mice. *Nat. Commun.* 12, 1557. doi:10.1038/s41467-021-21820-1

Pronot, M., Kieffer, F., Gay, A. S., Debayle, D., Forquet, R., Poupon, G., et al. (2021). Proteomic identification of an endogenous synaptic SUMOylome in the developing rat brain. *Front. Mol. Neurosci.* 14, 780535. doi:10.3389/fnmol.2021.780535

Quan, Y., Zhang, Q., Chen, M., Wu, H., Ou, J., Shen, Y., et al. (2020). Genotype and phenotype correlations for TBL1XR1 in neurodevelopmental disorders. *J. Mol. Neurosci.* 70, 2085–2092. doi:10.1007/s12031-020-01615-7

Raimer, A. C., Gray, K. M., and Matera, A. G. (2017). Smn - a chaperone for nuclear RNP social occasions? *RNA Biol.* 14, 701–711. doi:10.1080/15476286.2016.1236168

Ramos, A., Hollingworth, D., Adinolfi, S., Castets, M., Kelly, G., Frenkiel, T. A., et al. (2006). The structure of the N-terminal domain of the fragile X mental retardation protein: A platform for protein-protein interaction. *Structure* 14, 21–31. doi:10.1016/j.str.2005.09.018

Ravanidis, S., Kattan, F. G., and Doxakis, E. (2018). Unraveling the pathways to neuronal homeostasis and disease: Mechanistic insights into the role of RNA-binding proteins and associated factors. *Int. J. Mol. Sci.* 3, E2280. doi:10.3390/ijms19082280

Richter, J. D., and Zhao, X. (2021). The molecular biology of FMRP: New insights into fragile X syndrome. *Nat. Rev. Neurosci.* 22, 209–222. doi:10.1038/s41583-021-00432-0

Sawicka, K., Hale, C. R., Park, C. Y., Fak, J. J., Gresack, J. E., Van Driesche, S. J., et al. (2019). FMRP has a cell-type-specific role in CA1 pyramidal neurons to regulate autism-related transcripts and circadian memory. *Elife* 1, e46919. doi:10.7554/eLife.46919

Schorova, L., Pronot, M., Poupon, G., Prieto, M., Folci, A., Khayachi, A., et al. (2019). The synaptic balance between sumoylation and desumoylation is maintained by the activation of metabotropic mGlu5 receptors. *Cell. Mol. Life Sci.* 76, 3019–3031. doi:10.1007/s00018-019-03075-8

Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M., and Noble-Haeusslein, L. J. (2013). Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog. Neurobiol.* 106-107, 1–16. doi:10.1016/j.pneurobio.2013.04.001

Shah, S., Molinaro, G., Liu, B., Wang, R., Huber, K. M., and Richter, J. D. (2020). FMRP control of ribosome translocation promotes chromatin modifications and alternative splicing of neuronal genes linked to autism. *Cell Rep.* 30, 4459–4472. doi:10.1016/j.celrep.2020.02.076

Shamay-Ramot, A., Khermesh, K., Porath, H. T., Barak, M., Pinto, Y., Wachtel, C., et al. (2015). Fmrp interacts with adar and regulates RNA editing, synaptic density and locomotor activity in zebrafish. *PLoS Genet.* 11, e1005702. doi:10.1371/journal.pgen.1005702

Sittler, A., Devys, D., Weber, C., and Mandel, J. L. (1996). Alternative splicing of exon 14 determines nuclear or cytoplasmic localisation of fmr1 protein isoforms. *Hum. Mol. Genet.* 5, 95–102. doi:10.1093/hmg/5.1.95

Tabet, R., Moutin, E., Becker, J. A., Heintz, D., Fouillen, L., Flatter, E., et al. (2016). Fragile X Mental Retardation Protein (FMRP) controls diacylglycerol kinase activity in neurons. *Proc. Natl. Acad. Sci. U. S. A.* 113, E3619–E3628. doi:10.1073/pnas.1522631113

Taha, M. S., Haghighi, F., Stefanski, A., Nakhaei-Rad, S., Kazemein Jasemi, N. S., Al Kabbani, M. A., et al. (2021). Novel FMRP interaction networks linked to cellular stress. *FEBS J. Feb* 288, 837–860. doi:10.1111/febs.15443

Taha, M. S., Nouri, K., Milroy, L. G., Moll, J. M., Herrmann, C., Brunsveld, L., et al. (2014). Subcellular fractionation and localization studies reveal a direct interaction of the fragile X mental retardation protein (FMRP) with nucleolin. *PLoS One* 9, e91465. doi:10.1371/journal.pone.0091465

Tamanini, F., Bontekoe, C., Bakker, C. E., van Unen, L., Anar, B., Willemsen, R., et al. (1999a). Different targets for the fragile X-related proteins revealed by their distinct nuclear localizations. *Hum. Mol. Genet.* 8, 863–869. doi:10.1093/hmg/8.5.863

Tamanini, F., Van Unen, L., Bakker, C., Sacchi, N., Galjaard, H., Oostra, B. A., et al. (1999b). Oligomerization properties of fragile-X mental-retardation protein (FMRP) and the fragile-X-related proteins FXR1P and FXR2P. *Biochem. J.* 343 (3), 517–523. doi:10.1042/bj3430517

Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., Gingras, A. C., and Choi, H. (2014). SAINTexpress: Improvements and additional features in significance analysis of INTeractome software. *J. Proteomics* 100, 37–43. doi:10.1016/j.jprot.2013.10.023

Tran, S. S., Jun, H. I., Bahn, J. H., Azghadi, A., Ramaswami, G., Van Nostrand, E. L., et al. (2019). Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* 22, 25–36. doi:10.1038/s41593-018-0287-x

Vitaliano-Prunier, A., Babour, A., Herissant, L., Apponi, L., Margaritis, T., Holstege, F. C., et al. (2012). H2B ubiquitylation controls the formation of export-competent mRNP. *Mol. Cell* 45, 132–139. doi:10.1016/j.molcel.2011.12.011

Wang, X., Zhang, S., Zheng, R., Yue, F., Lin, S. H., Rahmeh, A. A., et al. (2016). PDIP46 (DNA polymerase δ interacting protein 46) is an activating factor for human DNA polymerase δ.. *Oncotarget* 7, 6294–6313. doi:10.18632/oncotarget.7034

Wei, L., and Lai, E. C. (2022). Regulation of the alternative neural transcriptome by ELAV/Hu RNA binding proteins. *Front. Genet.* 13, 848626. doi:10.3389/fgene.2022.848626

Westmark, C. J., Maloney, B., Alisch, R. S., Sokol, D. K., and Lahiri, D. K. (2020). FMRP regulates the nuclear export of Adam9 and Psen1 mRNAs: Secondary analysis of an N(6)-methyladenosine dataset. *Sci. Rep.* 10, 10781. doi:10.1038/s41598-020-66394-y

Willemsen, R., Bontekoe, C., Tamanini, F., Galjaard, H., Hoogeveen, A., and Oostra, B. (1996). Association of FMRP with ribosomal precursor particles in the nucleolus. *Biochem. Biophys. Res. Commun.* 25, 27–33. doi:10.1006/bbrc.1996.1126

Woodward, L. A., Mabin, J. W., Gangras, P., and Singh, G. (2017). The exon junction complex: A lifelong guardian of mRNA fate. *Wiley Interdiscip. Rev. RNA* 8, 1. doi:10.1002/wrna.1411

Worringer, K. A., Chu, F., and Panning, B. (2009). The zinc finger protein Zn72D and DEAD box helicase Belle interact and control maleless mRNA and protein levels. *BMC Mol. Biol.* 10, 33. doi:10.1186/1471-2199-10-33

Zeidler, S., Severijnen, L. A., de Boer, H., van der Toorn, E. C., Ruivenkamp, C. A. L., Bijlsma, E. K., et al. (2021). A missense variant in the nuclear export signal of the FMR1 gene causes intellectual disability. *Gene* 768, 145298. doi:10.1016/j.gene.2020.145298

Zhang, W., Cheng, Y., Li, Y., Chen, Z., Jin, P., and Chen, D. (2014). A feed-forward mechanism involving Drosophila fragile X mental retardation protein triggers a replication stress-induced DNA damage response. *Hum. Mol. Genet.* 23, 5188–5196. doi:10.1093/hmg/ddu241

Zhang, Y., O'Connor, J. P., Siomi, M. C., Srinivasan, S., Dutra, A., Nussbaum, R. L., et al. (1995). The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. *EMBO J. Nov.* 14, 5358–5366. doi:10.1002/j.1460-2075.1995.tb00220.x

Zhou, L. T., Ye, S. H., Yang, H. X., Zhou, Y. T., Zhao, Q. H., Sun, W. W., et al. (2017). A novel role of fragile X mental retardation protein in pre-mRNA alternative splicing through RNA-binding protein 14. *Neuroscience* 349, 64–75. doi:10.1016/j.neuroscience.2017.02.044

frontiers | Frontiers in Molecular Biosciences

TYPE Perspective
PUBLISHED 06 October 2022
DOI 10.3389/fmolb.2022.1008921

OPEN ACCESS

EDITED BY
Barbara Bardoni,
UMR7275 Institut de Pharmacologie Moléculaire et Cellulaire (IPMC), France

REVIEWED BY
Shilpi Minocha,
Indian Institute of Technology Delhi, India
Nikolay Shirokikh,
Australian National University, Australia
Sara J Aton,
University of Michigan, United States

*CORRESPONDENCE
André P. Gerber,
a.gerber@surrey.ac.uk
Julie Seibt,
j.seibt@surrey.ac.uk

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 01 August 2022
ACCEPTED 20 September 2022
PUBLISHED 06 October 2022

CITATION
Buchanan IM, Smith TM, Gerber AP and Seibt J (2022), Are there roles for heterogeneous ribosomes during sleep in the rodent brain?
Front. Mol. Biosci. 9:1008921.
doi: 10.3389/fmolb.2022.1008921

Are there roles for heterogeneous ribosomes during sleep in the rodent brain?

Isla M. Buchanan[1], Trevor M. Smith[2,3], André P. Gerber[2]* and Julie Seibt[3]*

[1]Integrated Master Programme in Biochemistry, University of Surrey, Guildford, United Kingdom, [2]Department of Microbial Sciences, School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom, [3]Surrey Sleep Research Centre, University of Surrey, Guildford, United Kingdom

The regulation of mRNA translation plays an essential role in neurons, contributing to important brain functions, such as brain plasticity and memory formation. Translation is conducted by ribosomes, which at their core consist of ribosomal proteins (RPs) and ribosomal RNAs. While translation can be regulated at diverse levels through global or mRNA-specific means, recent evidence suggests that ribosomes with distinct configurations are involved in the translation of different subsets of mRNAs. However, whether and how such proclaimed ribosome heterogeneity could be connected to neuronal functions remains largely unresolved. Here, we postulate that the existence of heterologous ribosomes within neurons, especially at discrete synapses, subserve brain plasticity. This hypothesis is supported by recent studies in rodents showing that heterogeneous RP expression occurs in dendrites, the compartment of neurons where synapses are made. We further propose that sleep, which is fundamental for brain plasticity and memory formation, has a particular role in the formation of heterologous ribosomes, specialised in the translation of mRNAs specific for synaptic plasticity. This aspect of our hypothesis is supported by recent studies showing increased translation and changes in RP expression during sleep after learning. Thus, certain RPs are regulated by sleep, and could support different sleep functions, in particular brain plasticity. Future experiments investigating cell-specific heterogeneity in RPs across the sleep-wake cycle and in response to different behaviour would help address this question.

KEYWORDS

ribosomal protein, ribosome heterogeneity, neuron, neurites, sleep, brain plasticity, synapse

Local translation in neurons: evidence and functions

Neurons are the primary cell type in the nervous system (the other being glia, which provide structural and metabolic support), responsible for transmission of electrochemical signals in the form of action potentials (APs). On average, the volume of a neuron is more than 10,000 times greater than most mammalian cells.

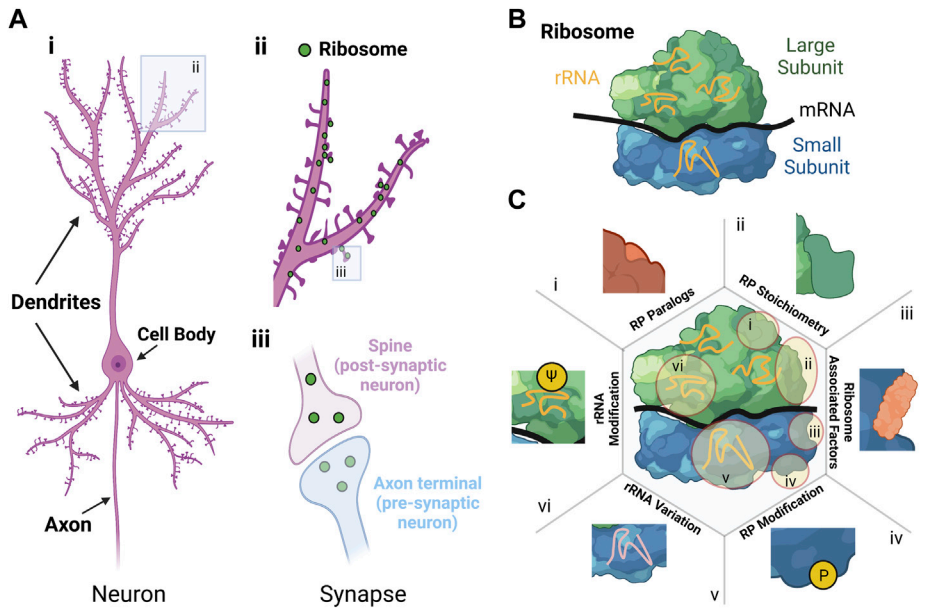


FIGURE 1
Sources and effect of ribosomal heterogeneity in neurons. **(A)** Morphology of a neuron. i) Schematic of a neuron showing the cell body, dendrites and axon. ii) Close-up schematic of dendrites with ribosomes shown as green dots located both in and adjacent to synapses. iii) Close-up schematic of the synapse with ribosomes shown as green dots in the cytoplasm in both the post-synaptic neuron (i.e., spine in purple) and pre-synaptic neuron (i.e., axon terminal in blue). **(B)** Structure of the ribosome with the large subunit (green), small subunit (blue), ribosomal RNA (rRNA in orange), and messenger RNA (mRNA in black). For the large and small subunit, the different colours indicate different ribosomal proteins. **(C)** Schematic showing six sources of ribosomal variation, with illustrative examples of how these changes manifest: i) RP paralogs, ii) RP stoichiometry, iii) Ribosomal associated factors, iv) RP modification, v) rRNA variation, and vi) rRNA modification (adapted from (Norris et al., 2021)). Created with BioRender.com.

Neurons have a complex and polarised morphology with processes (i.e., neurites) which are divided into different compartments called dendrites and axons, that can extend millimetres to meters away from the cell body containing the nucleus. Axons transmit APs to other neurons and dendrites receive APs and release neurotransmitters onto little protrusions called spines (Donato et al., 2019) (Figure 1A). The site of connection between axon termini and dendritic spines is called the synapse (Figure 1A). One neuron can host up to hundred thousand spines on its dendrites and thus receive thousands of contacts from different axons, maximising the number of possible synapses between neurons (Chidambaram et al., 2019; Nishiyama, 2019; Megías et al., 2001; Spruston, 2008).

In neurons, timely production of proteins in individual neurites and synapses is critical for spatial control of cellular function (Perez et al., 2021b). Indeed, using high resolution imaging of neurons *in vivo* and *in vitro*, it has been shown that most pre-synaptic axon termini and post-synaptic spines contain ribosomes, translation factors, and mRNAs (Cajigas et al., 2012; Kitamura et al., 2015; Holt et al., 2019; Sun et al., 2021) (Figure 1A). Local protein synthesis in neurons is required for growth and remodelling of synapses (i.e., synaptic plasticity) (Sutton and Schuman, 2006), which enables critical processes such as brain development, brain plasticity, learning and memory (Banko and Klann, 2008) and recovery from brain injury by promoting synapse-specific production of functionally distinct groups of proteins.

For example, strengthening and weakening of individual synaptic contacts relies on the synthesis of different pools of proteins which ultimately increase or decrease efficiency of communication (Zukin et al., 2009). Local translation of proteins critical for long-term increases in synaptic transmission includes α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors which increases the sensitivity of dendrites to incoming stimuli and strengthens individual synapses (Perez et al., 2021b). Similarly, local translation of proteins with key roles in synaptic structure, such as Arc, actin, PSD95 and α/βCaMKII, participates in the structural modification of dendritic spines during plasticity processes (Nakahata and Yasuda, 2018; Newpher et al., 2018).

As modulation of synaptic communication is naturally linked to memory formation, translation regulation is also important for memory (Sutton and Schuman, 2006; Banko and Klann, 2008), as shown in *Drosophila* (Ashraf et al., 2006) and rodents. Inhibition of translation initiation

(i.e., mTOR signalling) in the rodent hippocampus, a brain structure important for memory formation, impairs performance in a widely used single trial inhibitory avoidance memory task in rodents (Bekinschtein et al., 2007). While *in vivo* manipulation of local translation at specific synapses remains challenging, experimental and computational studies support the specific involvement of molecular changes in dendrites (Kastellakis and Poirazi, 2019) and spines (Hayashi-Takagi et al., 2015) for memory formation. Altogether, localised mRNA translation within neurons plays a critical role in many brain functions involving the growth and remodelling of synapses. However, a complete picture of the forms and functions of various translation regulatory mechanisms contributing to brain function in neurons remains largely unexplored.

Translation regulation through heterogeneous ribosomes

The eukaryotic ribosome is a large ribonucleoprotein complex that consists of a large (60S) and a small (40S) subunit that assemble during the initiation step of translation to form complete 80S ribosomes (Figure 1B). The two subunits share four ribosomal RNAs (rRNAs) (Ford, 2020) and 80 ribosomal proteins RPs, 33 of them allocated to the small (RPS) and 47 to the large (RPL) subunit (De La Cruz et al., 2015). Translation can be regulated at diverse levels (e.g., initiation, elongation, or termination) but major impact is given at the initiation step. Thereby, translation can be controlled at a global level or for specific mRNAs, for example through modification of translation initiation factors and specific RNA binding proteins (RBPs) or microRNAs, respectively (Sonenberg and Hinnebusch, 2009; King and Gerber, 2016).

Beside the regulation of translation mediated through those accessory factors, the view that ribosomes are all identical prevailed for decades. However, increasing evidence over the last decade suggests that cell-specific heterogeneous populations of ribosomes could exist and may result in different preferences of individual ribosomes (or "specialised ribosomes") for the translation of diverse subsets of mRNAs or in the modulation of translation elongation rates (Genuth and Barna, 2018; Gay et al., 2022). Such heterogenous populations of ribosomes can be formed by (Figure 1C, (Norris et al., 2021)) 1) the exchange or substitution of RP paralogues (e.g., RPL39 and RPL39L) (Komili et al., 2007), 2) altered RP stoichiometry based on differences in RP expression (Emmott et al., 2019); 3) different ribosome-associated proteins (Simsek et al., 2017), 4) post-translational modifications of RPs (Carroll et al., 2008); 5) rRNA composition through variation of rRNA gene sequences (Locati et al., 2017); and 6) rRNA modifications, such as ribose-methylation or pseudouridylation (ψ) (Natchiar et al., 2017). Consequently, these factors could lead to a cell-specific array of ribosome variants, some of them having specialised functions in translation (Li and Wang, 2020; Gay et al., 2022).

A well-described example of ribosome specialisation concerns RPL38 (Xue et al., 2015). Ribosomes containing RPL38 are required for translation of certain *Homeobox (Hox)* mRNAs that code for proteins defining the body axis and structures (Kondrashov et al., 2011). In developing mice embryos, *Rpl38* transcripts were enriched in certain regions including the face, eye, neural tube (brain and spinal cord precursor) and importantly somites, which are precursors of the axial skeleton. These locations tended to overlap with regions where tissue patterning defects were observed in mice with the "Tail short" (Ts) mutation (which display a short and curled tail, an anteroposterior skeletal patterning defect and several other skeletal abnormalities), a phenotype thought to be caused by *Rpl38* gene mutation (Kondrashov et al., 2011). RPL38 has been reported to control cap-independent translation of *Hox* mRNAs *via* specific internal ribosome entry sites (IRES) (Xue et al., 2015), although results from a recent study suggest that transcriptional promoters or splice sites may instead be responsible for the putative IRES activity in *Hox* genes (Akirtava et al., 2022). However, selective translation through IRES specificity is likely replicated with other RPs (Hertz et al., 2013; Shi et al., 2017; Kampen et al., 2019). Selective capture of ribosomes containing two specific RPs, RPS25, and RPL10A, combined with ribosome profiling, showed that those RPs preferentially translated unique sets of transcripts (Shi et al., 2017). Specifically, RPS25-containing ribosomes were preferentially associated with transcripts coding for proteins involved in the cell cycle and vitamin B12 pathway, while RPL10A-enriched transcripts were associated with extracellular matrix organisation, system development and steroid metabolism (Shi et al., 2017). In the same way that RPL38 is thought to be required for IRES-dependent translation of *Hox* mRNAs (Xue et al., 2015), Shi et al. (2017) found that RPL10A is required for the translation of three mRNAs known to contain IRES elements (*Igf2, App,* and *Chmp2a*). IRES-dependent translation of these transcripts was significantly reduced upon knockdown of *Rpl10a*, but not the control *Rpl29*, indicating that translation of *Igf2*, *App* and *Chmp2a* is attributed to RPL10A.

Further support for the model of ribosome specialisation is provided by examples of human diseases such as ribosomopathies and fragile-X syndrome. Ribosomopathies are pathologies which result from mutations in certain RP genes (Orgebin et al., 2020), with as many as 20 RP genes, including *RPS19,* involved in Diamond-Blackfan anaemia (DBA) (Leger-Silvestre et al., 2005; Da Costa et al., 2020). Besides a deficiency of erythroblasts, about 50% of DBA patients experience other congenital anomalies such as growth retardation, cardiac and urogenital abnormalities, increased risk of cancer, cephalic malformations and learning difficulties (Da Costa et al., 2020; Panda et al., 2020). Fragile-X syndrome is another well-characterised neurodevelopmental disorder associated with

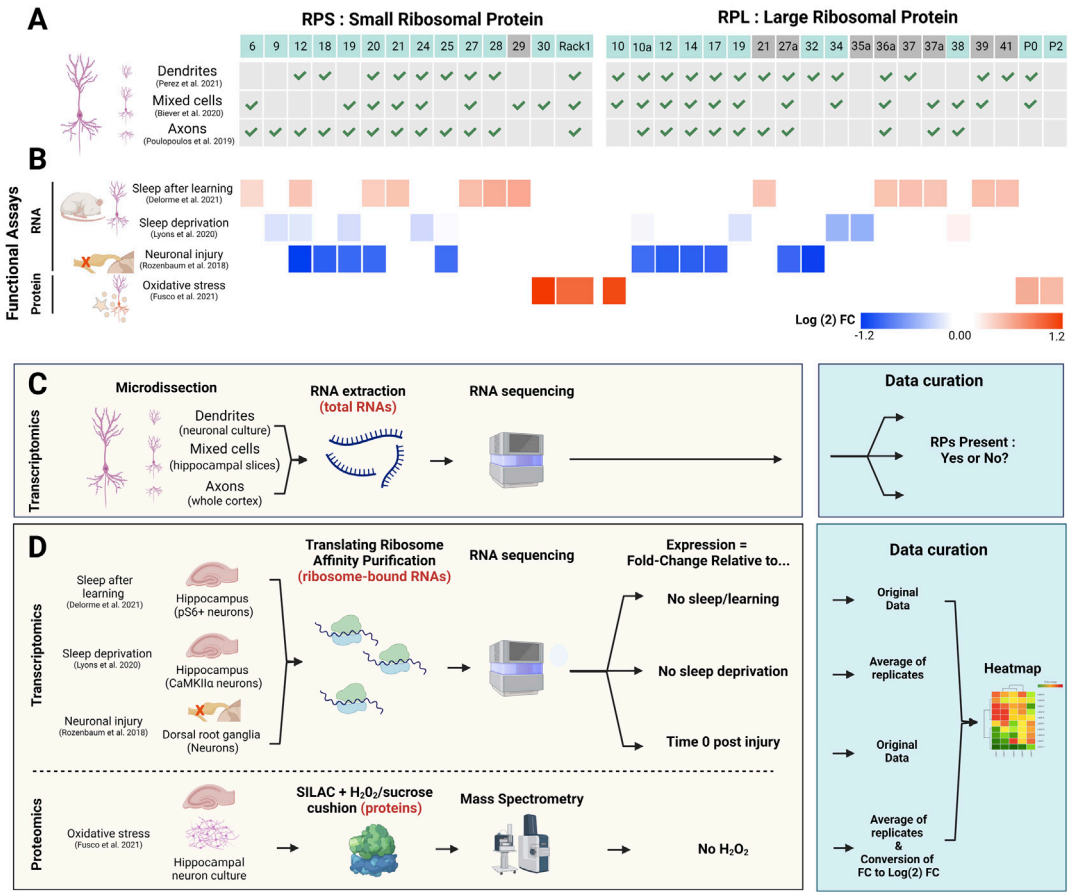


FIGURE 2
Contribution of RPs to ribosome heterogeneity in neurons and functional significance. **(A)** Table indicating the presence (green check mark) or absence of select RP (top line) mRNAs in whole neurons vs. dendrites or axons, according to previous studies (Poulopoulos et al., 2019; Biever et al., 2020; Perez et al., 2021a). RP location (non-surface: grey boxes; surface: blue boxes) was determined using human ribosome structure (PDB: 4V6X). **(B)** Heatmap (Morpheus, https://software.broadinstitute.org/morpheus) representing heterogeneous expression of selected RPs in neurons from *in vivo* studies in rodents or *in vitro* studies. The colour bar depicts log (2) fold changes (FC) of the particular conditions (indicated on the left) against control condition values (see D and (Rozenbaum et al., 2018; Lyons et al., 2020; Delorme et al., 2021; Fusco et al., 2021) for experimental details). From top to bottom: 1) 13 RP transcripts on membrane-bound ribosomes in activated (pS6+) neurons in the hippocampus following contextual fear conditioning and subsequent sleep in mice (Delorme et al., 2021). 2) 7 RP mRNAs affected by acute (5 h) sleep deprivation in excitatory neurons (CamKIIα+) in the mouse hippocampus (Lyons et al., 2020). 3) The five largest changes for small and large ribosomal subunits (10 RPs) in dorsal root ganglion (DRG) neurons following sciatic nerve injury, 4 h after injury (Rozenbaum et al., 2018). 4) 5 RPs that are significantly upregulated after 0.1 mM H_2O_2 treatment of primary neuronal culture; values are average of three biological replicates (Fusco et al., 2021). Additionally, log (2)FC values for RPs previously mentioned in the text and associated with ribosome specialisation (RPS19, RPS25, RPL10A and RPL38) were included for each study if they were present in the original data. **(C,D)** Schematic experimental approaches of the studies performed in neuronal compartments **(C)** relates to **(A)** and functional assays **(D)** relates to **(B)**. The blue boxes specify the type of analysis performed for each study for data visualisation. Created with BioRender.com.

intellectual disability and learning difficulties. Excessive translation due to the loss of the RBP fragile-X mental retardation protein (FMRP) caused by CGG triplet repeat expansions in the FMR1 gene is a key mechanism in the disease (Richter et al., 2015). In this regard, a recent study suggests that excessive translation of RPs in neurons reduces the translation of longer-length transcripts coding for proteins contributing to synaptic stability (Seo et al., 2022), supporting the established link between the loss of FMRP and aberrant synaptic plasticity (Sidorov et al., 2013).

While research is needed to determine the biological functions of ribosome heterogeneity, some studies have highlighted its importance in specific cellular processes, in particular during development (Li and Wang, 2020; Norris et al., 2021). Whether ribosome heterogeneity exists and has functions in the brain remains under-investigated.

Ribosome heterogeneity in neurons

Although specific RPs (e.g., RPL38, RPS25, and RPL10A) have functional roles in different cell types (see above) (Li and Wang, 2020), still very little is known about whether RPs are differentially regulated or expressed in different areas of the brain or in specific brain cells. For example, RP heterogeneity between brain regions has been reported in normal and brain cancer samples (Panda et al., 2020). Conversely, ageing does apparently not drastically influence RP stoichiometry in the cortex, cerebellum and hippocampus when assessed in a mixed cell preparation (Amirbeigarab et al., 2019); this does not, however, exclude the possibility for heterogeneity associated with specific cell types forming those tissues, including neurons and glial cells. In the following sections, we further focus our considerations on ribosome heterogeneity and functions in neurons, which are the brain cells responsible for activity in the brain and are central for information processing supporting cognitive functions.

The recent implementation of various transcriptomics and proteomics approaches applied *in vitro* and *in vivo* (Figure 2C) revealed that neurons show differential expression of RPs within the dendritic and axonal sub-compartment compared to a mixed cell population (Poulopoulos et al., 2019; Biever et al., 2020; Perez et al., 2021a) (Figure 2A) and suggest a compartment-specific heterogeneity in RP expression within neurons (Figure 2B). Additionally, an influential study into ribosome specialisation was recently carried out in *Xenopus laevis* showing that remodelling of ribosomes occurred in axons by exchange of locally synthesised RPs (Shigeoka et al., 2019). While ribosome biogenesis typically occurs in the nucleus, this study provides further evidence that ribosomes are dynamic structures.

Several studies have been using the translating ribosome affinity purification (TRAP) method to study neuron-specific changes in the translatome (i.e., pool of translated mRNA). Those studies have provided evidence for differential expression of ribosomal components and RPs in neurons, in particular upon changing conditions/neuronal activation (Figures 2B,D). For instance, more than 1,600 differentially expressed transcripts were identified in dorsal root ganglion (DRG) neurons following sciatic nerve injury (Rozenbaum et al., 2018). Importantly, in lumbar DRG neurons, mRNAs for several RPs were significantly decreased 4 h after injury, with the greatest decrease in expression seen for RPS12 and RPL32 (Figure 2B), and recovery of RP expression after 12 h. Another study in rats utilised single-molecule fluorescence *in situ* hybridisation to show variable expression of 29 different RP mRNAs in dendrites in both hippocampal slices and primary neuronal culture (Fusco et al., 2021). Using an elegant approach combining heavy amino acid labelling (dynamic SILAC) followed by mass spectrometry to label and identify newly synthesised RPs in translating ribosomes, the authors revealed that a subset of 12 RPs showed a dynamic profile of association/dissociation with ribosomes in dendrites, including RPL38 described above. The majority of those 12 RPs are located at the surface of the ribosome, making those proteins susceptible to exchange or post-translational modifications (PTMs) that could—in principle—modulate the function/activity of ribosomes. In fact, many of the RPs showing compartment-specific differences and functional changes are located on the surface of the ribosome (Figure 2). An example of such PTMs of RPs is the phosphorylation of RPS6 in response to a variety of neuronal stimuli, adding the possibility for introduction of compartment specific and activity specific PTM of RPs, propagating functional impact for translation of mRNA subsets (Knight et al., 2012). Finally, changes in the environment (i.e., oxidative stress induced by hydrogen peroxide exposure) elicit five of those 12 RPs to increase their association with the ribosome (Fusco et al., 2021) (Figure 2B), suggesting that physiological stress modifies ribosome composition and function, which could directly affect translation of specific mRNAs. Therefore, other changes in physiological states could also have a significant impact on RP ribosomal protein heterogeneity in neurons.

Altogether, we propose that the existence of heterologous ribosomes within neurons, especially at discrete synapses, may contribute to specific functions, including synaptic plasticity. Dynamic incorporation of RPs to alter ribosome stoichiometry could facilitate rapid formation of specialised ribosomes and enable the translation of subsets of mRNAs involved in the remodelling of synapses. We further postulate that sleep, which is accompanied by major physiological changes in the brain and is an important regulator of the synaptic proteome (Noya et al., 2019), translation and brain plasticity (Seibt and Frank, 2019), could use changes in ribosome heterogeneity and specialised ribosomes to regulate the translation of mRNAs important for synaptic remodelling.

Sleep, plasticity, and ribosomal proteins

The role of sleep in brain plasticity and memory is well established; mounting evidence shows that sleep enhances the physiological and behavioural changes associated with new experiences (Abel et al., 2013; Rasch and Born, 2013; Raven et al., 2018; Puentes-Mestril et al., 2019). For example, performance in various types of memory tasks in humans and animals is significantly increased when sleep occurs right after learning (Alger et al., 2015; Schmid et al., 2020). Similarly, new sensory experience changes perception in a sleep-dependent manner during development (Frank et al., 2001) and adulthood (Durkin and Aton, 2019). Furthermore, work in the last decade, using high-resolution imaging techniques, has provided strong evidence that sleep influences changes in dendritic spine structure, linked to brain plasticity and

memory (Yang et al., 2014; Li et al., 2017; Seibt et al., 2017; Zhou et al., 2020; Aime et al., 2022). Finally, the establishment of our basic sensorimotor system in the central nervous system is thought to largely depend on sleep during early development (Blumberg, 2015). Since long-term changes associated with synaptic plasticity and memory require protein synthesis to persist over time (Davis and Squire, 1984; Costa-Mattioli et al., 2009), sleep may support brain plasticity *via* translation regulation, including specialised ribosomes, for translation of particular subsets of mRNAs.

The current view suggests that experience-dependent transcription (e.g., immediate early genes) (Yap and Greenberg, 2018) occurs preferentially during wakefulness in the nucleus, while mRNA translation occurs mostly during sleep in a distributed manner across neurites (Seibt and Frank, 2019). This is supported by evidence showing that translation rates are increased during sleep in the brain in various species (Ramm and Smith, 1990; Nakanishi et al., 1997) and the expression of regulators of translation initiation and elongation occurs preferentially during sleep (Cirelli et al., 2004; Mackiewicz et al., 2007; Seibt et al., 2012). Moreover, sleep deprivation leads to a decrease in translation initiation with associated memory deficits (Tudor et al., 2016) and pharmacological disruption of translation initiation during sleep impairs experience-dependent synaptic plasticity *in vivo* (Seibt et al., 2012), further suggesting the importance of translation during sleep for brain plasticity and memory consolidation. The underlying pathways for translational control during sleep are still not well-characterised, but global control of translation initiation, *via* the mTORC-1 signalling pathway, seems to be specifically activated during sleep (Seibt et al., 2012; Tudor et al., 2016). (Gingras et al., 1999; Seibt et al., 2012; Tudor et al., 2016) Other factors such as RBPs, microRNAs or non-coding RNAs (ncRNAs) are all potential mechanisms involved in translational control of specific mRNAs during sleep. There is some evidence that sleep deprivation differentially impacts the expression of groups of microRNAs in different parts of the brain, with a trend toward decreased expression in the cortex and increased expression in the hippocampus (Davis et al., 2007). Although this supports a region-specific regulation of translation by the sleep-wake cycle *via* microRNAs, the data on this remain isolated.

Besides global and specific control through signalling pathways and RBPs/ncRNAs, respectively, whether the formation of specialised ribosomes could also contribute to selected translation during sleep remains unclear. However, sleep-dependent translational changes were examined within neurons in the hippocampus after learning in mice (Delorme et al., 2021) (Figure 2C). Using TRAP, ribosome-associated transcripts were identified from different subcellular fractions of neurons (i.e., cytosolic vs. membrane-associated ribosomes). Sleep deprivation primarily affected mRNA translated in the cytosol, while learning mainly altered transcripts on membrane-bound ribosomes, suggesting a first level of translational specificity within neurons. Importantly, sleep after learning showed increased translation of membrane-bound transcripts, including mRNAs of 13 RPs, such as *Rps27* and *Rps28,* with ~50% of the RPs located at the ribosome surface (Figure 2B) (Delorme et al., 2021). These changes were specifically allocated to sleep as RP mRNA expression did not increase if sleep was prevented after learning. Differential RP translation during sleep may thus support compartmentalised heterogenous populations of ribosomes occurring through exchange and incorporation of RPs at the surface of ribosomes. Another, indirect, evidence supporting a role for sleep in increased translation of RPs was provided by a study investigating the impact of sleep deprivation (SD) on hippocampal neurons using TRAP in mice (Lyons et al., 2020). Following 5 h of SD, 198 mRNAs showed differential association with ribosomes compared to sleep control mice (Lyons et al., 2020). Certain RPs showed a trend toward decreased expression; transcripts for RPs previously associated with ribosome specialisation such as *Rps25*, *Rpl10a* and *Rpl38* displayed little to no change, but other RP mRNAs like *Rpl34* and *Rpl35a* were more affected (Figure 2B). Although none of these changes were found to be significant, the data nevertheless suggests that sleep may promote the expression of particular RPs as short sleep deprivation tends to reduce their translation (Lyons et al., 2020).

Although the data are still sparse, they do support a specific effect of sleep on differential RP expression in brain regions important for memory. The variability of changes observed in the various physiological and behavioural paradigms align with the idea that different neuronal functions are accompanied by expression of different ribosomes, which may favour formation of specialised ribosomes.

Discussion and conclusions

Several studies revealed differences in RP expression in neurons, some of them specifically during sleep. Furthermore, many RPs are located on the ribosomal surface, adding the possibility for alternative integration or exchange with other RPs. Changes in RP expression and stoichiometry could contribute to the remodelling of neuronal networks and other processes that benefit from sleep, such as general metabolism and membrane repair (Mackiewicz et al., 2007; Anafi et al., 2019), energy conservation (Roth et al., 2010), mood and stress restoration (Goldstein and Walker, 2014), or the clearance of toxins (Xie et al., 2013; van Alphen et al., 2021). While the incorporation and presence of specialised ribosomes in neurons needs to be shown, it may allow for translation of different subsets of mRNAs across individual neuronal compartments and sleep stages.

Experience-dependent plasticity, including memory, leads to the formation or strengthening of certain synapses, whereas

others are weakened or even removed. (Yang et al., 2009; Sanders et al., 2012; El-Boustani et al., 2018). Consequently, within the same dendrites, some spines grow while others retract. These dynamic processes involve different mechanisms and proteins, which are - at least in part - instructed by the synthesis of process-relevant proteins. Furthermore, sleep is composed of two different stages, rapid-eye movement (REM) and non-REM sleep, which alternate within minutes in rodents (i.e., one NREM-REM cycle ~5–10 min) (Trachsel et al., 1991). Interestingly, those phases are coupled with the formation and removal of synapses (Bellesi and De Vivo, 2020; Sun et al., 2020), occurring preferentially during NREM and REM sleep phases, respectively (Yang et al., 2014; Zhou et al., 2020). Hence, assembly of specialised ribosomes during NREM and/or REM sleep could be well-suited to quickly adjust mRNA translation, thereby promoting and/or consolidating the bi-directional plasticity at synapses. How specialised ribosomes are established, controlled, and could become selectively activated during particular sleep phases remains to be uncovered and may be linked to specific brain waves during NREM and REM sleep, known to reactivate selected circuits and enhance memory (Poe, 2017; Ngo et al., 2020; Zhou et al., 2020; Skelin et al., 2021).

The currently available data relate mostly to differential mRNA expression of RPs (Figure 2). However, the data does not necessarily show that the RPs are also synthesised and incorporated into active ribosomes, possibly contributing to ribosome heterogeneity. Thus, besides establishing and monitoring specific RP synthesis in neurons, several questions need to be addressed in the future: 1) Does RP heterogeneity take place in neuronal sub-compartments? 2) How does RP heterogeneity in specific neuronal compartments impact the translation of subsets of mRNAs? 3) How do brain states modulate ribosome heterogeneity to generate specialised ribosomes? Finally, besides neurons there are other important brain cells, such as glia (e.g., astrocytes, microglia) and the vasculature, where ribosome heterogeneity may apply and entail specialised functions (Knight et al., 2012; Bellesi et al., 2015).

In the future, we expect that fundamental questions will be addressed *in vitro* using either brain slices or primary neuronal cultures. For instance, the application of proximity-based labelling techniques could allow the isolation of proteins/RNAs in the vicinity of a target molecule (e.g., RPs and mRNAs) (Padron et al., 2019; Ramanathan et al., 2019), which should identify the spatial partners present under different physiological conditions to help understand function. Furthermore, we have shown evidence of translating RPs in different physiological states (Figure 2). Combining puromycylation (Tom Dieck et al., 2015), to tag newly synthesised proteins, with specific antibodies (i.e. RPs) could help understand the spatial location of these newly translated RPs, to further understand RNA/protein interactions (Weissinger et al., 2021) in different brain cells and neuronal compartments *in vitro*. Obtaining functional *in vivo* data remains the gold standard for understanding molecular mechanisms linked to behaviour, including sleep. However, due to technical challenges of *in vivo* pharmacology (e.g., diffusion, dilution, biochemical reactions), the application of the above-mentioned methods remains difficult (Uezu et al., 2016) and further advances for the relevant techniques are required. For example, *in vivo* imaging of small RPs and RNA (Tatavarty et al., 2012; Tutucci et al., 2018; Cawte et al., 2020) is currently difficult, but not impossible (Grimm et al., 2017; Wegner et al., 2017; Hrabetova et al., 2018). Technical improvements such as brighter and more photostable fluorophores, higher resolution imaging, and better access to the tissue of interest (Tutucci et al., 2018; Das et al., 2021) may be key developments to advance the field, allowing better monitoring of RP localisation in cells. Furthermore, improved methods for genetic manipulations (Raguram et al., 2022), to target RPs with fewer off-target side effects, may allow tagging of different ribosomal components at the same time, facilitating isolation of different ribosome complexes from the same cell with TRAP. At the end, those approaches could be combined with behavioural and sleep manipulations, opening the paths towards fundamental understanding of the functional impact of ribosome heterogeneity in complex physiological processes.

Data availability statement

The original contributions presented in the study are included in the article; further enquiries can be directed to the corresponding authors.

Author contributions

IB performed data analysis. TS and JS made the figures. All authors made intellectual contributions and wrote the manuscript.

Funding

This work was supported by a grant from the Leverhulme Trust to JS and APG (RPG-2020-340), and a Wolfson Research Merit Award from the Royal Society to APG (WM170026). We thank the University of Surrey for defraying publication costs.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abel, T., Havekes, R., Saletin, J. M., and Walker, M. P. (2013). Sleep, plasticity and memory from molecules to whole-brain networks. *Curr. Biol.* 23, R774–R788. doi:10.1016/j.cub.2013.07.025

Aime, M., Calcini, N., Borsa, M., Campelo, T., Rusterholz, T., Sattin, A., et al. (2022). Paradoxical somatodendritic decoupling supports cortical plasticity during REM sleep. *Science* 376, 724–730. doi:10.1126/science.abk2734

Akirtava, C., May, G. E., and Mcmanus, C. J. (2022). False-positive IRESes from Hoxa9 and other genes resulting from errors in mammalian 5' UTR annotations. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2122170119. doi:10.1073/pnas.2122170119

Alger, S. E., Chambers, A. M., Cunningham, T., and Payne, J. D. (2015). The role of sleep in human declarative memory consolidation. *Curr. Top. Behav. Neurosci.* 25, 269–306. doi:10.1007/7854_2014_341

Amirbeigarab, S., Kiani, P., Sanchez, A. V., Krisp, C., Kazantsev, A., Fester, L., et al. (2019). Invariable stoichiometry of ribosomal proteins in mouse brain tissues with aging. *Proc. Natl. Acad. Sci. U. S. A.* 116, 22567–22572. doi:10.1073/pnas.1912060116

Anafi, R. C., Kayser, M. S., and Raizen, D. M. (2019). Exploring phylogeny to find the function of sleep. *Nat. Rev. Neurosci.* 20, 109–116. doi:10.1038/s41583-018-0098-9

Ashraf, S. I., Mcloon, A. L., Sclarsic, S. M., and Kunes, S. (2006). Synaptic protein synthesis associated with memory is regulated by the RISC pathway in Drosophila. *Cell* 124, 191–205. doi:10.1016/j.cell.2005.12.017

Banko, J. L., and Klann, E. (2008). Cap-dependent translation initiation and memory. *Prog. Brain Res.* 169, 59–80. doi:10.1016/S0079-6123(07)00004-0

Bekinschtein, P., Katche, C., Slipczuk, L. N., Igaz, L. M., Cammarota, M., Izquierdo, I., et al. (2007). mTOR signaling in the hippocampus is necessary for memory formation. *Neurobiol. Learn. Mem.* 87, 303–307. doi:10.1016/j.nlm.2006.08.007

Bellesi, M., and De Vivo, L. (2020). Structural synaptic plasticity across sleep and wake. *Curr. Opin. Physiol.* 15, 74–81. doi:10.1016/j.cophys.2019.12.007

Bellesi, M., De Vivo, L., Tononi, G., and Cirelli, C. (2015). Transcriptome profiling of sleeping, waking, and sleep deprived adult heterozygous Aldh1L1 - eGFP-L10a mice. *Genom. Data* 6, 114–117. doi:10.1016/j.gdata.2015.08.031

Biever, A., Glock, C., Tushev, G., Ciirdaeva, E., Dalmay, T., Langer, J. D., et al. (2020). Monosomes actively translate synaptic mRNAs in neuronal processes. *Science* 367, eaay4991. doi:10.1126/science.aay4991

Blumberg, M. S. (2015). Developing sensorimotor systems in our sleep. *Curr. Dir. Psychol. Sci.* 24, 32–37. doi:10.1177/0963721414551362

Cajigas, I. J., Tushev, G., Will, T. J., Tom Dieck, S., Fuerst, N., and Schuman, E. M. (2012). The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron* 74, 453–466. doi:10.1016/j.neuron.2012.02.036

Carroll, A. J., Heazlewood, J. L., Ito, J., and Millar, A. H. (2008). Analysis of the Arabidopsis cytosolic ribosome proteome provides detailed insights into its components and their post-translational modification. *Mol. Cell. Proteomics* 7, 347–369. doi:10.1074/mcp.M700052-MCP200

Cawte, A. D., Unrau, P. J., and Rueda, D. S. (2020). Live cell imaging of single RNA molecules with fluorogenic Mango II arrays. *Nat. Commun.* 11, 1283. doi:10.1038/s41467-020-14932-7

Chidambaram, S. B., Rathipriya, A. G., Bolla, S. R., Bhat, A., Ray, B., Mahalakshmi, A. M., et al. (2019). Dendritic spines: Revisiting the physiological role. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 92, 161–193. doi:10.1016/j.pnpbp.2019.01.005

Cirelli, C., Gutierrez, C. M., and Tononi, G. (2004). Extensive and divergent effects of sleep and wakefulness on brain gene expression. *Neuron* 41, 35–43. doi:10.1016/s0896-6273(03)00814-6

Costa-Mattioli, M., Sossin, W. S., Klann, E., and Sonenberg, N. (2009). Translational control of long-lasting synaptic plasticity and memory. *Neuron* 61, 10–26. doi:10.1016/j.neuron.2008.10.055

Da Costa, L., Leblanc, T., and Mohandas, N. (2020). Diamond-Blackfan anemia. *Blood* 136, 1262–1273. doi:10.1182/blood.2019000947

Das, S., Vera, M., Gandin, V., Singer, R. H., and Tutucci, E. (2021). Intracellular mRNA transport and localized translation. *Nat. Rev. Mol. Cell Biol.* 22, 483–504. doi:10.1038/s41580-021-00356-8

Davis, C. J., Bohnet, S. G., Meyerson, J. M., and Krueger, J. M. (2007). Sleep loss changes microRNA levels in the brain: A possible mechanism for state-dependent translational regulation. *Neurosci. Lett.* 422, 68–73. doi:10.1016/j.neulet.2007.06.005

Davis, H. P., and Squire, L. R. (1984). Protein synthesis and memory: A review. *Psychol. Bull.* 96, 518–559. doi:10.1037/0033-2909.96.3.518

De La Cruz, J., Karbstein, K., and Woolford, J. L., Jr. (2015). Functions of ribosomal proteins in assembly of eukaryotic ribosomes *in vivo*. *Annu. Rev. Biochem.* 84, 93–129. doi:10.1146/annurev-biochem-060614-033917

Delorme, J., Wang, L., Kodoth, V., Wang, Y., Ma, J., Jiang, S., et al. (2021). Hippocampal neurons' cytosolic and membrane-bound ribosomal transcript profiles are differentially regulated by learning and subsequent sleep. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2108534118. doi:10.1073/pnas.2108534118

Donato, A., Kagias, K., Zhang, Y., and Hilliard, M. A. (2019). Neuronal sub-compartmentalization: A strategy to optimize neuronal function. *Biol. Rev. Camb. Philos. Soc.* 94, 1023–1037. doi:10.1111/brv.12487

Durkin, J. M., and Aton, S. J. (2019). How sleep shapes thalamocortical circuit function in the visual system. *Annu. Rev. Vis. Sci.* 5, 295–315. doi:10.1146/annurev-vision-091718-014715

El-Boustani, S., Ip, J. P. K., Breton-Provencher, V., Knott, G. W., Okuno, H., Bito, H., et al. (2018). Locally coordinated synaptic plasticity of visual cortex neurons *in vivo*. *Science* 360, 1349–1354. doi:10.1126/science.aao0862

Emmott, E., Jovanovic, M., and Slavov, N. (2019). Ribosome stoichiometry: From form to function. *Trends biochem. Sci.* 44, 95–109. doi:10.1016/j.tibs.2018.10.009

Ford, D. (2020). Ribosomal heterogeneity - a new inroad for pharmacological innovation. *Biochem. Pharmacol.* 175, 113874. doi:10.1016/j.bcp.2020.113874

Frank, M. G., Issa, N. P., and Stryker, M. P. (2001). Sleep enhances plasticity in the developing visual cortex. *Neuron* 30, 275–287. doi:10.1016/s0896-6273(01)00279-3

Fusco, C. M., Desch, K., Dorrbaum, A. R., Wang, M., Staab, A., Chan, I. C. W., et al. (2021). Neuronal ribosomes exhibit dynamic and context-dependent exchange of ribosomal proteins. *Nat. Commun.* 12, 6127. doi:10.1038/s41467-021-26365-x

Gay, D. M., Lund, A. H., and Jansson, M. D. (2022). Translational control through ribosome heterogeneity and functional specialization. *Trends biochem. Sci.* 47, 66–81. doi:10.1016/j.tibs.2021.07.001

Genuth, N. R., and Barna, M. (2018). The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol. Cell* 71, 364–374. doi:10.1016/j.molcel.2018.07.018

Gingras, A. C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., et al. (1999). Regulation of 4E-BP1 phosphorylation: A novel two-step mechanism. *Genes Dev.* 13, 1422–1437. doi:10.1101/gad.13.11.1422

Goldstein, A. N., and Walker, M. P. (2014). The role of sleep in emotional brain function. *Annu. Rev. Clin. Psychol.* 10, 679–708. doi:10.1146/annurev-clinpsy-032813-153716

Grimm, J. B., Muthusamy, A. K., Liang, Y., Brown, T. A., Lemon, W. C., Patel, R., et al. (2017). A general method to fine-tune fluorophores for live-cell and *in vivo* imaging. *Nat. Methods* 14, 987–994. doi:10.1038/nmeth.4403

Hayashi-Takagi, A., Yagishita, S., Nakamura, M., Shirai, F., Wu, Y. I., Loshbaugh, A. L., et al. (2015). Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature* 525, 333–338. doi:10.1038/nature15257

Hertz, M. I., Landry, D. M., Willis, A. E., Luo, G., and Thompson, S. R. (2013). Ribosomal protein S25 dependency reveals a common mechanism for diverse internal ribosome entry sites and ribosome shunting. *Mol. Cell. Biol.* 33, 1016–1026. doi:10.1128/MCB.00879-12

Holt, C. E., Martin, K. C., and Schuman, E. M. (2019). Local translation in neurons: Visualization and function. *Nat. Struct. Mol. Biol.* 26, 557–566. doi:10.1038/s41594-019-0263-5

Hrabetova, S., Cognet, L., Rusakov, D. A., and Nagerl, U. V. (2018). Unveiling the extracellular space of the brain: From super-resolved microstructure to *in vivo* function. *J. Neurosci.* 38, 9355–9363. doi:10.1523/JNEUROSCI.1664-18.2018

Kampen, K. R., Sulima, S. O., Verbelen, B., Girardi, T., Vereecke, S., Rinaldi, G., et al. (2019). The ribosomal RPL10 R98S mutation drives IRES-dependent BCL-2 translation in T-ALL. *Leukemia* 33, 319–332. doi:10.1038/s41375-018-0176-z

Kastellakis, G., and Poirazi, P. (2019). Synaptic clustering and memory formation. *Front. Mol. Neurosci.* 12, 300. doi:10.3389/fnmol.2019.00300

King, H. A., and Gerber, A. P. (2016). Translatome profiling: Methods for genome-scale analysis of mRNA translation. *Brief. Funct. Genomics* 15, 22–31. doi:10.1093/bfgp/elu045

Knight, Z. A., Tan, K., Birsoy, K., Schmidt, S., Garrison, J. L., Wysocki, R. W., et al. (2012). Molecular profiling of activated neurons by phosphorylated ribosome capture. *Cell* 151, 1126–1137. doi:10.1016/j.cell.2012.10.039

Komili, S., Farny, N. G., Roth, F. P., and Silver, P. A. (2007). Functional specificity among ribosomal proteins regulates gene expression. *Cell* 131, 557–571. doi:10.1016/j.cell.2007.08.037

Kondrashov, N., Pusic, A., Stumpf, C. R., Shimizu, K., Hsieh, A. C., Ishijima, J., et al. (2011). Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* 145, 383–397. doi:10.1016/j.cell.2011.03.028

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361, 13–37. doi:10.1016/j.gene.2005.06.037

Leger-Silvestre, I., Caffrey, J. M., Dawaliby, R., Alvarez-Arias, D. A., Gas, N., Bertolone, S. J., et al. (2005). Specific role for yeast homologs of the Diamond blackfan anemia-associated Rps19 protein in ribosome synthesis. *J. Biol. Chem.* 280, 38177–38185. doi:10.1074/jbc.M506916200

Li, D., and Wang, J. (2020). Ribosome heterogeneity in stem cells and development. *J. Cell Biol.* 219, e202001108. doi:10.1083/jcb.202001108

Li, W., Ma, L., Yang, G., and Gan, W. B. (2017). REM sleep selectively prunes and maintains new synapses in development and learning. *Nat. Neurosci.* 20, 427–437. doi:10.1038/nn.4479

Locati, M. D., Pagano, J. F. B., Girard, G., Ensink, W. A., Van Olst, M., Van Leeuwen, S., et al. (2017). Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. *RNA* 23, 1188–1199. doi:10.1261/rna.061515.117

Lyons, L. C., Chatterjee, S., Vanrobaeys, Y., Gaine, M. E., and Abel, T. (2020). Translational changes induced by acute sleep deprivation uncovered by TRAP-Seq. *Mol. Brain* 13, 165. doi:10.1186/s13041-020-00702-5

Mackiewicz, M., Shockley, K. R., Romer, M. A., Galante, R. J., Zimmerman, J. E., Naidoo, N., et al. (2007). Macromolecule biosynthesis: A key function of sleep. *Physiol. Genomics* 31, 441–457. doi:10.1152/physiolgenomics.00275.2006

Megías, M., Emri, Z., Freund, T. F., and Gulyás, A. I. (2001). Total number and distribution of inhibitory and excitatory synapses on hippocampal CA1 pyramidal cells. *Neuroscience* 102, 527–540.

Nakahata, Y., and Yasuda, R. (2018). Plasticity of spine structure: Local signaling, translation and cytoskeletal reorganization. *Front. Synaptic Neurosci.* 10, 29. doi:10.3389/fnsyn.2018.00029

Nakanishi, H., Sun, Y., Nakamura, R. K., Mori, K., Ito, M., Suda, S., et al. (1997). Positive correlations between cerebral protein synthesis rates and deep sleep in *Macaca mulatta*. *Eur. J. Neurosci.* 9, 271–279. doi:10.1111/j.1460-9568.1997.tb01397.x

Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I., and Klaholz, B. P. (2017). Visualization of chemical modifications in the human 80S ribosome structure. *Nature* 551, 472–477. doi:10.1038/nature24482

Newpher, T. M., Harris, S., Pringle, J., Hamilton, C., and Soderling, S. (2018). Regulation of spine structural plasticity by Arc/Arg3.1. *Semin. Cell Dev. Biol.* 77, 25–32. doi:10.1016/j.semcdb.2017.09.022

Ngo, H. V., Fell, J., and Staresina, B. (2020). Sleep spindles mediate hippocampal-neocortical coupling during long-duration ripples. *Elife* 9, e57011. doi:10.7554/eLife.57011

Nishiyama, J. (2019). Plasticity of dendritic spines: Molecular function and dysfunction in neurodevelopmental disorders. *Psychiatry Clin. Neurosci.* 73, 541–550. doi:10.1111/pcn.12899

Norris, K., Hopes, T., and Aspden, J. L. (2021). Ribosome heterogeneity and specialization in development, *Wiley Interdiscip. Rev. RNA*, 12. e1644. doi:10.1002/wrna.1644

Noya, S. B., Colameo, D., Bruning, F., Spinnler, A., Mircsof, D., Opitz, L., et al. (2019). The forebrain synaptic transcriptome is organized by clocks but its proteome is driven by sleep. *Science* 366, eaav2642. doi:10.1126/science.aav2642

Orgebin, E., Lamoureux, F., Isidor, B., Charrier, C., Ory, B., Lezot, F., et al. (2020)., 9. Cells, E2080. doi:10.3390/cells9092080Ribosomopathies: New therapeutic perspectives*Cells*

Padron, A., Iwasaki, S., and Ingolia, N. T. (2019). Proximity RNA labeling by APEX-seq reveals the organization of translation initiation complexes and repressive RNA granules. *Mol. Cell* 75, 875–887. doi:10.1016/j.molcel.2019.07.030

Panda, A., Yadav, A., Yeerna, H., Singh, A., Biehl, M., Lux, M., et al. (2020). Tissue- and development-stage-specific mRNA and heterogeneous CNV signatures of human ribosomal proteins in normal and cancer samples. *Nucleic Acids Res.* 48, 7079–7098. doi:10.1093/nar/gkaa485

Perez, J. D., Dieck, S. T., Alvarez-Castelao, B., Tushev, G., Chan, I. C., and Schuman, E. M. (2021a). Subcellular sequencing of single neurons reveals the dendritic transcriptome of GABAergic interneurons. *Elife* 10, e63092. doi:10.7554/eLife.63092

Perez, J. D., Fusco, C. M., and Schuman, E. M. (2021b). A functional dissection of the mRNA and locally synthesized protein population in neuronal dendrites and axons. *Annu. Rev. Genet.* 55, 183–207. doi:10.1146/annurev-genet-030321-054851

Poe, G. R. (2017). Sleep is for forgetting. *J. Neurosci.* 37, 464–473. doi:10.1523/JNEUROSCI.0820-16.2017

Poulopoulos, A., Murphy, A. J., Ozkan, A., Davis, P., Hatch, J., Kirchner, R., et al. (2019). Subcellular transcriptomes and proteomes of developing axon projections in the cerebral cortex. *Nature* 565, 356–360. doi:10.1038/s41586-018-0847-y

Puentes-Mestril, C., Roach, J., Niethard, N., Zochowski, M., and Aton, S. J. (2019). How rhythms of the sleeping brain tune memory and synaptic plasticity. *Sleep* 42, zsz095. doi:10.1093/sleep/zsz095

Raguram, A., Banskota, S., and Liu, D. R. (2022). Therapeutic *in vivo* delivery of gene editing agents. *Cell* 185, 2806–2827. doi:10.1016/j.cell.2022.03.045

Ramanathan, M., Porter, D. F., and Khavari, P. A. (2019). Methods to study RNA-protein interactions. *Nat. Methods* 16, 225–234. doi:10.1038/s41592-019-0330-1

Ramm, P., and Smith, C. T. (1990). Rates of cerebral protein synthesis are linked to slow wave sleep in the rat. *Physiol. Behav.* 48, 749–753. doi:10.1016/0031-9384(90)90220-x

Rasch, B., and Born, J. (2013). About sleep's role in memory. *Physiol. Rev.* 93, 681–766. doi:10.1152/physrev.00032.2012

Raven, F., Van Der Zee, E. A., Meerlo, P., and Havekes, R. (2018). The role of sleep in regulating structural plasticity and synaptic strength: Implications for memory and cognitive function. *Sleep. Med. Rev.* 39, 3–11. doi:10.1016/j.smrv.2017.05.002

Richter, J. D., Bassell, G. J., and Klann, E. (2015). Dysregulation and restoration of translational homeostasis in fragile X syndrome. *Nat. Rev. Neurosci.* 16, 595–605. doi:10.1038/nrn4001

Roth, T. C., Rattenborg, N. C., and Pravosudov, V. V. (2010). The ecological relevance of sleep: The trade-off between sleep, memory and energy conservation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 945–959. doi:10.1098/rstb.2009.0209

Rozenbaum, M., Rajman, M., Rishal, I., Koppel, I., Koley, S., Medzihradszky, K. F., et al. (2018). Translatome regulation in neuronal injury and axon regrowth. *eNeuro* 5, ENEURO.0276–17.2018. doi:10.1523/ENEURO.0276-17.2018

Sanders, J., Cowansage, K., Baumgartel, K., and Mayford, M. (2012). Elimination of dendritic spines with long-term memory is specific to active circuits. *J. Neurosci.* 32, 12570–12578. doi:10.1523/JNEUROSCI.1131-12.2012

Schmid, D., Erlacher, D., Klostermann, A., Kredel, R., and Hossner, E. J. (2020). Sleep-dependent motor memory consolidation in healthy adults: A meta-analysis. *Neurosci. Biobehav. Rev.* 118, 270–281. doi:10.1016/j.neubiorev.2020.07.028

Seibt, J., Dumoulin, M. C., Aton, S. J., Coleman, T., Watson, A., Naidoo, N., et al. (2012). Protein synthesis during sleep consolidates cortical plasticity *in vivo*. *Curr. Biol.* 22, 676–682. doi:10.1016/j.cub.2012.02.016

Seibt, J., and Frank, M. G. (2019). Primed to sleep: The dynamics of synaptic plasticity across brain states. *Front. Syst. Neurosci.* 13, 2. doi:10.3389/fnsys.2019.00002

Seibt, J., Richard, C. J., Sigl-Glockner, J., Takahashi, N., Kaplan, D. I., Doron, G., et al. (2017). Publisher Correction: Cortical dendritic activity correlates with spindle-rich oscillations during sleep in rodents. *Nat. Commun.* 8, 1838. doi:10.1038/s41467-017-01652-8

Seo, S. S., Louros, S. R., Anstey, N., Gonzalez-Lozano, M. A., Harper, C. B., Verity, N. C., et al. (2022). Excess ribosomal protein production unbalances translation in a model of Fragile X Syndrome. *Nat. Commun.* 13, 3236. doi:10.1038/s41467-022-30979-0

Shi, Z., Fujii, K., Kovary, K. M., Genuth, N. R., Rost, H. L., Teruel, M. N., et al. (2017). Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell* 67, 71–83. doi:10.1016/j.molcel.2017.05.021

Shigeoka, T., Koppers, M., Wong, H. H., Lin, J. Q., Cagnetta, R., Dwivedy, A., et al. (2019). On-site ribosome remodeling by locally synthesized ribosomal proteins in axons. *Cell Rep.* 29, 3605–3619. doi:10.1016/j.celrep.2019.11.025

Sidorov, M. S., Auerbach, B. D., and Bear, M. F. (2013). Fragile X mental retardation protein and synaptic plasticity. *Mol. Brain* 6, 15. doi:10.1186/1756-6606-6-15

Simsek, D., Tiu, G. C., Flynn, R. A., Byeon, G. W., Leppek, K., Xu, A. F., et al. (2017). The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity. *Cell* 169, 1051–1065. doi:10.1016/j.cell.2017.05.022

Skelin, I., Zhang, H., Zheng, J., Ma, S., Mander, B. A., Kim Mcmanus, O., et al. (2021). Coupling between slow waves and sharp-wave ripples engages distributed neural activity during sleep in humans. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2012075118. doi:10.1073/pnas.2012075118

Sonenberg, N., and Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: Mechanisms and biological targets. *Cell* 136, 731–745. doi:10.1016/j.cell.2009.01.042

Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* 9, 206–21.

Sun, C., Nold, A., Fusco, C. M., Rangaraju, V., Tchumatchenko, T., Heilemann, M., et al. (2021). The prevalence and specificity of local protein synthesis during neuronal synaptic plasticity. *Sci. Adv.* 7, eabj0790. doi:10.1126/sciadv.abj0790

Sun, L., Zhou, H., Cichon, J., and Yang, G. (2020). Experience and sleep-dependent synaptic plasticity: From structure to activity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375, 20190234. doi:10.1098/rstb.2019.0234

Sutton, M. A., and Schuman, E. M. (2006). Dendritic protein synthesis, synaptic plasticity, and memory. *Cell* 127, 49–58. doi:10.1016/j.cell.2006.09.014

Tatavarty, V., Ifrim, M. F., Levin, M., Korza, G., Barbarese, E., Yu, J., et al. (2012). Single-molecule imaging of translational output from individual RNA granules in neurons. *Mol. Biol. Cell* 23, 918–929. doi:10.1091/mbc.E11-07-0622

Tom Dieck, S., Kochen, L., Hanus, C., Heumuller, M., Bartnik, I., Nassim-Assir, B., et al. (2015). Direct visualization of newly synthesized target proteins *in situ*. *Nat. Methods* 12, 411–414. doi:10.1038/nmeth.3319

Trachsel, L., Tobler, I., Achermann, P., and Borbély, A. A. (1991). Sleep continuity and the REM-nonREM cycle in the rat under baseline conditions and after sleep deprivation. *Physiol. Behav.* 49, 575–580. doi:10.1016/0031-9384(91)90283-t

Tudor, J. C., Davis, E. J., Peixoto, L., Wimmer, M. E., Van Tilborg, E., Park, A. J., et al. (2016). Sleep deprivation impairs memory by attenuating mTORC1-dependent protein synthesis. *Sci. Signal.* 9, ra41. doi:10.1126/scisignal.aad4949

Tutucci, E., Livingston, N. M., Singer, R. H., and Wu, B. (2018). Imaging mRNA *in vivo*, from birth to death. *Annu. Rev. Biophys.* 47, 85–106. doi:10.1146/annurev-biophys-070317-033037

Uezu, A., Kanak, D. J., Bradshaw, T. W., Soderblom, E. J., Catavero, C. M., Burette, A. C., et al. (2016). Identification of an elaborate complex mediating postsynaptic inhibition. *Science* 353, 1123–1129. doi:10.1126/science.aag0821

Van Alphen, B., Semenza, E. R., Yap, M., Van, S. B., and Allada, R. (2021). A deep sleep stage in Drosophila with a functional role in waste clearance. *Sci. Adv.* 7, eabc2999. doi:10.1126/sciadv.abc2999

Wegner, W., Ilgen, P., Gregor, C., Van Dort, J., Mott, A. C., Steffens, H., et al. (2017). *In vivo* mouse and live cell STED microscopy of neuronal actin plasticity using far-red emitting fluorescent proteins. *Sci. Rep.* 7, 11781. doi:10.1038/s41598-017-11827-4

Weissinger, R., Heinold, L., Akram, S., Jansen, R. P., and Hermesh, O. (2021). RNA proximity labeling: A new detection tool for RNA-protein interactions. *Molecules* 26, 2270. doi:10.3390/molecules26082270

Xie, L., Kang, H., Xu, Q., Chen, M. J., Liao, Y., Thiyagarajan, M., et al. (2013). Sleep drives metabolite clearance from the adult brain. *Science* 342, 373–377. doi:10.1126/science.1241224

Xue, S., Tian, S., Fujii, K., Kladwang, W., Das, R., and Barna, M. (2015). RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature* 517, 33–38. doi:10.1038/nature14010

Yang, G., Lai, C. S., Cichon, J., Ma, L., Li, W., and Gan, W. B. (2014). Sleep promotes branch-specific formation of dendritic spines after learning. *Science* 344, 1173–1178. doi:10.1126/science.1249098

Yang, G., Pan, F., and Gan, W. B. (2009). Stably maintained dendritic spines are associated with lifelong memories. *Nature* 462, 920–924. doi:10.1038/nature08577

Yap, E. L., and Greenberg, M. E. (2018). Activity-regulated transcription: Bridging the gap between neural activity and behavior. *Neuron* 100, 330–348. doi:10.1016/j.neuron.2018.10.013

Zhou, Y., Lai, C. S. W., Bai, Y., Li, W., Zhao, R., Yang, G., et al. (2020). REM sleep promotes experience-dependent dendritic spine elimination in the mouse cortex. *Nat. Commun.* 11, 4819. doi:10.1038/s41467-020-18592-5

Zukin, R. S., Richeter, J. D., and Bagni, C. (2009). Signals, synapses, and synthesis: how new proteins control plasticity. *Front. Neural. Circuits* 3, 14.

frontiers | Frontiers in Molecular Biosciences

TYPE Review
PUBLISHED 07 October 2022
DOI 10.3389/fmolb.2022.1000205

OPEN ACCESS

EDITED BY
Gian Gaetano Tartaglia,
Italian Institute of Technology, Italy

REVIEWED BY
Zhi-Ping Liu,
Shandong University, China
Zilong Zhang,
Hainan University, China
Alessio Colantoni,
Italian Institute of Technology, Italy

*CORRESPONDENCE
G Pepe,
gerardo.pepe@uniroma2.it
M Helmer-Citterich,
citterich@uniroma2.it

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 21 July 2022
ACCEPTED 20 September 2022
PUBLISHED 07 October 2022

CITATION
Pepe G, Appierdo R, Carrino C, Ballesio F, Helmer-Citterich M and Gherardini PF (2022), Artificial intelligence methods enhance the discovery of RNA interactions. *Front. Mol. Biosci.* 9:1000205. doi: 10.3389/fmolb.2022.1000205

Artificial intelligence methods enhance the discovery of RNA interactions

G Pepe[1]*, R Appierdo[1], C Carrino[2], F Ballesio[2], M Helmer-Citterich[1]* and PF Gherardini[1]

[1]Department of Biology, University of Rome "Tor Vergata", Rome, Italy, [2]PhD Program in Cellular and Molecular Biology, Department of Biology, University of Rome "Tor Vergata", Rome, Italy

Understanding how RNAs interact with proteins, RNAs, or other molecules remains a challenge of main interest in biology, given the importance of these complexes in both normal and pathological cellular processes. Since experimental datasets are starting to be available for hundreds of functional interactions between RNAs and other biomolecules, several machine learning and deep learning algorithms have been proposed for predicting RNA-RNA or RNA-protein interactions. However, most of these approaches were evaluated on a single dataset, making performance comparisons difficult. With this review, we aim to summarize recent computational methods, developed in this broad research area, highlighting feature encoding and machine learning strategies adopted. Given the magnitude of the effect that dataset size and quality have on performance, we explored the characteristics of these datasets. Additionally, we discuss multiple approaches to generate datasets of negative examples for training. Finally, we describe the best-performing methods to predict interactions between proteins and specific classes of RNA molecules, such as circular RNAs (circRNAs) and long non-coding RNAs (lncRNAs), and methods to predict RNA-RNA or RNA-RBP interactions independently of the RNA type.

KEYWORDS

RNA, RNA interaction predictors, natural language processing, deep learning, machine learning, embedding, RNA sequence, RNA secondary structure

Introduction

The involvement of RNAs in a wide range of biological processes, such as transcription, translation, neurogenesis, and the biogenesis and function of non-coding RNAs (ncRNAs) has been discussed in multiple studies (Newman et al., 2015; Turner and Díaz-Muñoz, 2018; Peng et al., 2022). Basic cellular physiology is critically dependant on RNA-Protein interactions (RPIs), as exemplified by their role in RNA splicing, transcription efficiency, stabilization and termination (Kelaini et al., 2021), in triggering RNA release from the transcription complex (Van Assche et al., 2015), and in regulating RNA degradation (Gilbertson et al., 2018). RNAs interact with RNA-Binding Proteins (RBPs) through sequence and structural motifs (Dominguez et al., 2018). Adinolfi et al. (Adinolfi et al., 2019) identified several RNA binding motifs by

analyzing PAR-CLIP, eCLIP and HITS-CLIP experiments. Starting from these motifs, Guarracino et al. developed a web server for the identification of enriched structure or sequence motifs in a pool of RNAs which returns putative interacting RBPs (Guarracino et al., 2021). Altered functionality of RBPs and subsequent disruption of RNA-RBPs regulatory networks are commonly observed in human genetic diseases, neurodegenerative diseases and multiple cancer types (Pereira et al., 2017; Gebauer et al., 2021; Schieweck et al., 2021). Besides interacting with proteins, RNAs can also interact with each other, giving rise to complex regulatory networks that control cellular physiology in health and disease (e.g. mRNA regulation exerted by miRNA) (Chen et al., 2019; Pepe et al., 2022b; Wang et al., 2022). Moreover, RNAs influence each others' expression level by competing for a limited pool of microRNAs (miRNAs) (Seitz, 2009; Poliseno et al., 2010), as postulated by the "competitive endogenous RNA" (ceRNA) theory (Salmena et al., 2011). The interaction between viral DNA or RNA genomes and host miRNAs is involved in immune system evasion and viral replication (Qiao et al., 2019). Accordingly, the role of exogenous DNA or RNA in viral infection has been extensively studied, highlighting how viral genomes can act as "sponges" for specific host miRNAs. This mechanism has been described for Hepatitis C Virus (Luna et al., 2015) and Epstein-Barr Virus (Riley et al., 2012) and it has also been suggested for SARS-CoV-2 (Pepe et al., 2022a).

Given the importance that RNA interactions play in fundamental cellular processes, cancer, and other diseases, several methods for studying the physical interactions between RNA and proteins have been developed (Ferrè et al., 2015). These *in vitro* or *in vivo* methods can be classified into two main categories: i) RNA-centric methods used to study proteins associated with a specific RNA; ii) protein-centric methods used to identify RNAs interacting with a specific protein (Ramanathan et al., 2019). Despite the large number of RNA interactions identified thanks to these methods, experimental validation is still expensive and time-consuming and computational approaches remain an active area of research.

In this review, we aim to elucidate recent advances in RNA interaction predictions, focusing on state-of-the-art methods currently used for the prediction of RNA-RNA or RNA-RBP interactions. The development of these methods is critically dependent on the quality and characteristics of datasets of known interactions. Accordingly, we will also review publicly available sources of RNA interaction data.

Overview of databases

A crucial element in the development of RNA-protein interaction prediction models is the retrieval of datasets containing known interacting pairs to be used for ML models' training. The present section will therefore survey two fundamental aspects in this respect. Firstly, we describe the main features of the most widely employed datasets for RNA-protein interaction prediction. Indeed, during the last decade, various datasets have been constructed and released to pursue this task. Such datasets, typically, rely on information maintained in databases or obtained through literature-mining operations and they involve interactions supported by experimental evidence. Subsequently in this section, a second crucial aspect is pointed out. Since machine learning methods for binary classification need to be trained on datasets containing a balanced number of samples from both classes to be predicted, in the case of RPI prediction this translates into disposing of datasets containing RNA-protein pairs that are known to interact (which will henceforth be referred to as "positive dataset") as well as non-interacting RNA-protein pairs (which will henceforth be referred to as "negative dataset"). We reported an overview of the major methods employed for the construction of RPI negative sets as well as a summary table reporting assumptions and outlines of such strategies.

Publicly available datasets of RNA interactions

Datasets currently considered as benchmarks for training, cross-validating or testing RPI prediction models include RPI369 and RPI2241 (Muppirala et al., 2011), RPI488 (Pan et al., 2016), and RPI1807 (Suresh et al., 2015). These are structure-based datasets which incorporate interaction pairs obtained from RNA-protein complexes whose structures have been deposited in the PDB (Velankar et al., 2021). Another commonly used dataset is NPInter2.0 (Yuan et al., 2014), which contains interactions derived from literature-mining and other databases.

The RPI2241 and RPI369 datasets were obtained from PRIDB (Lewis et al., 2011), a database of protein-RNA interfaces derived from PDB complexes (Burley et al., 2021). A total of 943 complexes from PRIDB (9,689 protein chains and 2,074 RNA chains) were initially selected. A final dataset consisting of 2241 experimentally validated RNA-protein interacting pairs (952 protein chains and 443 RNA chains) was derived, by redundancy reduction (discarding similar interaction on the basis of sequence identity) and sequence length filtering. When the RPI2241 dataset was constructed, a sizable fraction of all the RNA-protein complexes in the PDB corresponded to ribosomal structures, leading to a strong bias towards ribosomal RPIs. Accordingly, a second dataset, RPI369, was generated from RPI2241 by removing all RPIs that contained ribosomal proteins or ribosomal RNAs. Moreover, to generate a balanced dataset of non-interacting RNA-protein pairs, the RNAs and proteins from the original 943 complexes were

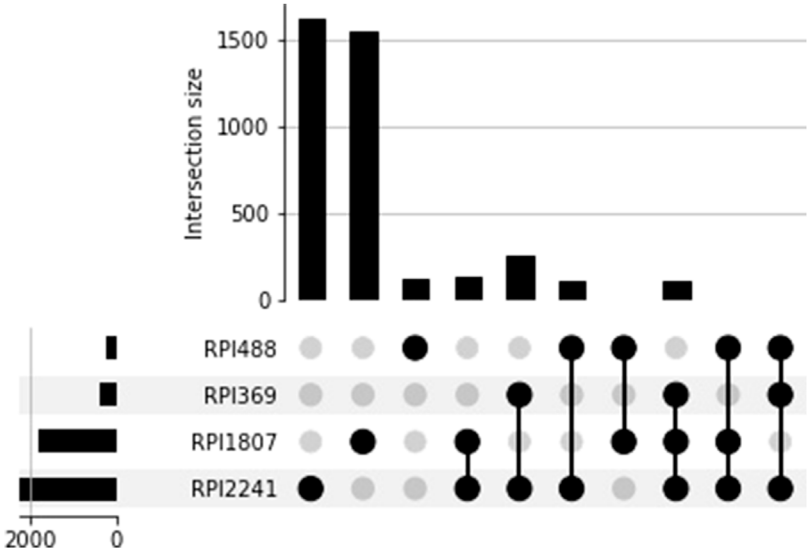


FIGURE 1
Overlap between the four RPI datasets generated from PDB RNA-protein complexes.

randomly paired and pairs similar to known interactions were further discarded.

The RPI488 dataset is a structure-based dataset, derived from PDB complexes and specifically incorporating lncRNA-protein interactions. In order to generate the dataset, 18 ncRNA-protein complexes were downloaded from the PDB and 726 lncRNA-protein pairs were collected from them. In order to derive both a positive and a negative dataset, a distance cutoff of 5Å was used. Also, redundant sequences (sequence identity greater than 90% for both protein and lncRNA sequences) were excluded by using CD-HIT (Fu et al., 2012). Following redundancy reduction, the final RPI488 dataset contains 488 protein-lncRNA pairs (243 interacting pairs and 245 non-interacting ones).

The RPI1807 dataset was derived by integrating the Nucleic Acid Database (NDB) (Coimbatore Narayanan et al., 2014) and the PRIDB. A total of 1560 RPI complexes available in NDB were selected and, for 1336 of them, atomic interactions were extracted from PRIDB, thus obtaining 13,163 protein and 2715 RNA chains. The procedure for constructing the dataset included sequence length filtering and redundancy removal according to sequence similarity. In order to obtain both positive and negative sets, the selected non-redundant pairs were further analyzed for atomic interactions with a distance threshold (3.40 Å). This threshold was used to distinguish strongly interacting protein-RNA pairs (positive set) from weakly interacting protein-RNA pairs (negative set). The final RPI1807 dataset consists of 1807 positive pairs and 1436 negative pairs.

The overlap between RPI datasets is reported in Figure 1. This overlap could be greater than that obtained by simply intersecting the RNA-protein pairs since a redundancy reduction was applied to each one of the RPI datasets. In each of the RPI datasets, RNA-protein pairs were clustered and only one pair was chosen as representative; this could influence the overlap between the four datasets.

NPInter2.0 is a database that integrates experimentally-validated functional interactions between ncRNAs and other biomolecules (RNAs, proteins and DNAs), collected both from literature mining and from multiple databases. Although newer releases of the database exist (up to NPInter v4.0), NPInter v2.0 is the most widely used dataset for the development of prediction models. The dataset contains a total of 201,107 ncRNA interactions from 18 organisms, excluding interactions involving tRNAs and rRNAs.

Interactions were derived from manual annotation of articles published between 2008 and 2013 and include both experimentally-validated interactions as well as binding sites identified by genome-wide techniques (Yuan et al., 2014). The authors also integrated data from other resources, mainly the LncRNADisease database (Chen et al., 2013), and finally performed a redundancy reduction procedure within the dataset.

Several datasets have been derived from NPInter2.0 by selecting subsets of interactions with characteristics of interest. More specifically, the most widely used non-structure-based dataset for the development and testing of RPI prediction models is a subset of this database (namely NPInter10412) first assembled by Suresh et al, (2015), and subsequently used in numerous other works (Li et al., 2021; Wang et al., 2021; Zhao et al., 2021). NPInter10412 contains 10,412 ncRNA-protein

TABLE 1 Number of RNA-protein interactions by species in the NPInter10412 dataset.

Species	# of RNA-protein interactions
H. sapiens	6975
M. musculus	2198
D. melanogaster	91
C. elegans	36
S. cerevisiae	905
S. cerevisiae S288c	5
E. coli	202

TABLE 2 Number of interactions of each type in the RNAInter database.

Interaction type	# of interactions
RNA-RNA	9,483,936
RNA-Protein	37,060,698
DNA-RNA	138,552
RNA-Histone modification	1,060,684
RNA-Compound	10,889

TABLE 3 Number of interactions in the 8 taxa in the RNAInter database. An overview of the main publicly available datasets for RNA-protein interaction prediction is given in Table 4.

Taxon	# of interactions
Actiniaria	872
Arthropoda	538,643
Bacteria	72,132
Fungi	622,927
Nematoda	883,131
Vertebrata	45,584,924
Viridiplantia	58,875
Virus	712,704

interactions, distributed among the different species as illustrated in Table 1.

A fourth release of NPInter was published in 2019 that increases the amount of high-throughput interactomes available data. NPInter4.0 (Teng et al., 2019) includes 600,000 new ncRNA interactions, particularly ncRNA–DNA interactions obtained *via* the ChIRP-seq technique, as well as interactions involving circular RNAs. Additionally, disease associations were added to the database.

Lastly, RNAInter4.0 is a recent resource that integrates experimentally validated and computationally predicted RNA interactions from literature-mining and databases (Kang et al., 2022). It provides information about different types of interactions in different *taxa*. Tables 2 and 3 summarize RNAInter's content.

Ultimately, despite remarkable advances in experimental techniques, the development of large and reliable RPI datasets is still the main bottleneck for training ML models. Hence, we would also like to stress the importance of redundancy control within data, since its presence may cause a leakage of information between training and test set during model training, resulting in untruthful prediction performance.

Strategies for the construction of a negative dataset

The lack of reliable datasets of non-interacting RNA-protein pairs is a major concern in the development of computational methods for RPI prediction. Indeed, it is not trivial to conclusively state that a given protein does not interact with a given RNA molecule (absence of evidence does not constitute evidence of absence). Indeed, various papers have demonstrated the critical effect of negative dataset composition on the performance of Machine Learning and Deep Learning models (Muppirala et al., 2011; Pan et al., 2016; Peng et al., 2019). Additionally, having balanced positive and negative sets is crucial to avoid overfitting on one class.

The most often used (Muppirala et al., 2011; Pan et al., 2016; Yi et al., 2020) method to construct a dataset of non-interacting pairs is to randomly pair RNAs and proteins in the positive set, followed by discarding the thus obtained pairs that showed high sequence similarity to the interacting ones, while retaining the others.

An interesting, albeit not widely used, technique to construct negative samples is the FIRE (FInding Reliable nEgative samples) method (Cheng et al., 2017). The core idea of this method relies on the following observation: given an experimentally-validated interaction between protein p1 and RNA r, and given another protein p2, the more similar p2 is to p1, the higher the likelihood that r interacts with p2. Thus, for each positive RPI (p1, r) the p2 protein that is most dissimilar to p1 is selected; if (p2, r) is not an experimentally-validated RPI, then it is selected as a negative RPI. The innovation introduced in this work lies in the way the similarity between each pair of proteins was computed, by taking into account functional annotations and protein domains information in addition to sequence similarity.

An additional approach that circumvents the requirement to create a negative dataset is PU learning, a binary classification method that can be applied when only positive (P) and unlabeled (U) data are available. For example, PRIPU trains a biased SVM on only positive and unlabelled examples (Cheng et al., 2015, 2017).

TABLE 4 Publicly available datasets for RNA-protein interaction prediction.

Dataset	# Of positive interactions	# Of negative interactions	Description	Negative set strategy	References
RPI2241	2241	2241	Structure-based dataset, containing RNA-protein interactions enriched in ribosomal RPIs	Random Pairing	Muppirala et al. (2011)
RPI369	369	369	Structure-based dataset, obtained from RPI2241 after removal of interactions derived from ribosomal complexes	Random Pairing	Muppirala et al. (2011)
RPI488	243	245	Structure-based dataset, comprising interactions between proteins and different classes of RNAs	Least atom distance	Pan et al. (2016)
RPI1807	1807	1436	Structure-based dataset, comprising interactions between proteins and different classes of RNAs	Least atom distance	Suresh et al. (2015)
NPInter10412	10,412	-	Non structure-based dataset, comprising RNA-protein interactions integrated from literature mining and other databases	-	Yuan et al. (2014); Suresh et al. (2015)

TABLE 5 Strategies for the construction of a negative dataset for RNA-protein interaction prediction.

Strategy	Assumption	Description
Random pairing	The likelihood of interaction occurring between randomly paired RNAs and proteins is low	By using known interacting pairs as starting point, the same number of non-interacting pairs are generated by randomly pairing RNAs and proteins from the positive set, followed by discarding pairs that are similar to interactions already present in the positive set
FIRE method	Given a known RNA-protein interacting pair (p1, r), and given a second protein p2, the smaller the sequence similarity between p1 and p2, the lower the likelihood that r interacts with p2	For each positive RNA-protein interaction (p1, r) the p2 protein that is most dissimilar to p1 is selected, similarity between each pair of proteins was computed by taking into account functional annotations and protein domain information in addition to sequence similarity
Subcellular localization method	RNAs and proteins that are not in the same subcellular compartment do not interact with each other	This method requires subcellular localization data
Least atom distance criterion	Only applicable to interactions derived from known-structure complexes	Given a multimolecular RNA-protein complex, for each pairwise combination of its constituent RNA and protein molecules, if there is at least one atom of the RNA located closer than a threshold to at least one protein atom, the pair is considered to be interacting otherwise it is included in the negative dataset

Some of the most often employed strategies for the construction of a negative dataset are listed in Table 5.

Computational methods for RNA-protein interaction discovery

If on the one hand, the choice of the right training dataset is critical, on the other hand the choice of the right algorithm for RNA-RBP interaction prediction is also important, considering that some predictors were developed for a specific class of RNAs, such as lncRNA or circRNA. We will therefore review the latest methods for RNA-RBP interaction prediction. Some methods predict interactions between proteins and a specific class of RNA molecules, such as circRNAs (Yang et al., 2021; Niu et al., 2022) and lncRNAs (Ge et al., 2016; Zhao et al., 2018; Xie et al., 2019; Zhou et al., 2021). Others were developed to predict RNA-RBP interactions independently from the RNA type (Akbaripour-Elahabad et al., 2016; Yi et al., 2018; Zhan et al., 2018; Wang et al., 2019, 2021; Zhang et al., 2020).

LPI-deepGBDT: An artificial intelligence algorithm for the prediction of long non-coding RNA-protein interactions

Long non-coding RNAs (lncRNAs) are a class of RNA molecules that have attracted strong interest in recent years due to their abundance and their role in many physiological and pathological processes (Kornienko et al., 2016). Since many of the functions performed by lncRNAs require their interaction with proteins (LPIs), and most of lncRNAs are of unknown function, identifying new LPIs is a very important task. Most of the methods developed for this task are based on hand-crafted features, which is a process that requires time, domain knowledge and is based on strong assumptions. We describe the LPI-

TABLE 6 Description of the train/test datasets, feature encoding and machine learning strategy for each of the described methods.

Method	Interacting molecules	Train/test dataset	Feature encoding	Machine learning strategy	References
LPI-deepGBDT	lncRNA-RBP	Derived from NPInter	Sequence features extracted using Pyfeat (Muhammod et al., 2019) and BioProt (Márquez and Castro Amaya, 2019)	Gradient boosting decision trees	Zhou et al. (2021)
LncPNet	lncRNA-RBP	Derived from NPInter v2.0	Heterogeneous network embedding of lncRNAs and proteins similarity networks and of the known lncRNA-protein interaction network	Support-vector machine	Zhao et al. (2021)
CRBPDL	circRNA-RBP	CLIP-seq experiments	k-nucleotide frequency (KNF), Doc2vec, electron-ion interaction pseudopotential (EIIP), chemical characteristics of nucleotides (CCN) and accumulated nucleotide frequency (ANF)	Deep multi-scale residual network (ResNet) and bidirectional gated recurrent unit with a self-attention mechanism (BiGRUs)	Niu et al. (2022)
EDLMFC	ncRNA-RBP	RPI1807 NPInter v2.0 RPI488	k-mer frequencies of the sequence and structure representations	Ensemble deep learning framework including convolutional neural networks (CNN) and bi-directional long short-term memory net-work (BLSTM)	Wang et al. (2021)
preMLI	miRNA-mRNA	Plants lncRNA-miRNA interaction dataset constructed using RNAHybrid 2.1.2	word2vec based sequence embedding	CNN and bidirectional gated recurrent unit (Bi-GRU)	Yu et al. (2022)
PrismNet	RNA-RBP	CLIP-seq experiments	One-hot-encoded sequence vectors and icSHAPE structure scores	Convolutional layers, squeeze-and-excitation networks (SE) and residual blocks	Sun et al. (2021)
PRNA	RNA-RBP	RsiteDB	Number of atoms, electrostatic charge, potential hydrogen bonds, hydrophobicity and relative accessible surface area were used as sequence features. Secondary structure of amino acid residues, conservation score (PSI-BLAST), side-chain environment were used as structure features. A sliding window was used to encode amino acid residues and create feature vectors	Random Forest	Liu et al. (2010)

deepGBDT algorithm (Table 6), which uses a feed-forward deep architecture based on gradient boosting decision trees (Zhou et al., 2021). In this work three human and two plant LPI datasets, derived from the NPInter database, were used as training for the classifier. These datasets were processed using several filters, similar to previous works (Li et al., 2015; Zheng et al., 2017; Zheng et al., 2017; Zhang et al., 2018; Bai et al., 2019). Multiple features of lncRNAs and proteins were calculated from their sequences using Pyfeat (Muhammod et al., 2019) and BioProt (Márquez and Castro Amaya, 2019). The dimensionality of the feature space was then reduced using PCA, and protein and RNA features were concatenated to obtain a matrix of features representing the interaction pairs. This matrix was used as input to the classifier, which consisted of a multi-layered deep framework based on a gradient boosting model. The authors compared their model with five state-of-the-art LPI prediction methods, namely LPI-BLS, LPI-CatBoost, PLIPCOM, LPI-SKF and LPI-HNM (Yang et al., 2016; Deng et al., 2018; Fan and Zhang, 2019; Wekesa et al., 2020; Zhou et al., 2020), using six measurements: precision, recall, accuracy, F1-score, AUC and AUPR, and obtaining better average performances. Furthermore, the LPI-deepGBDT algorithm was successfully applied to the identification of potential protein partners for a specific lncRNA and, given a specific protein, to infer its potential interacting lncRNAs. The authors highlight that one of the main drivers of performance improvement for this method is the integration of biological features.

LncPNet: A human long non-coding RNA-protein interactions predictor

Most models are developed to predict lncRNA-protein interactions irrespective of the species, which can result in the introduction of noise and negatively affect performance.

To address this and other limitations, Zhao et al, (2021) introduced a new predictor model called LncPNet (Table 6). This method is designed to exclusively predict human lncRNA-protein interactions. Moreover, protein and lncRNA features are automatically generated using a network embedding. For this study, human lncRNA-protein interactions were selected from NPInter v2.0 resulting in 7523 experimentally validated pairs,

including 3052 lncRNAs and 212 proteins. LncRNAs and proteins lacking sequence information were removed, thus obtaining a dataset of 4578 interactions between 2009 lncRNAs and 78 proteins. The negative dataset was built using the subcellular localization method (see Table 5). This method is based on a heterogeneous network of lncRNA-protein which is constructed using: i) lncRNA-lncRNA and protein-protein similarity; ii) known lncRNA-protein association. The similarity between lncRNAs and proteins is both calculated by Jaccard similarity and BLAST similarity. Subsequently the metapath2vec (Dong et al., 2017) method is used for network embedding and dimensionality reduction. LncRNA-protein interactions are represented as vectors of dimensionality 1 x 256 and those vectors are used to train a Support Vector Machine in order to predict whether an lncRNA interacts with a protein. Comparison with other state-of-the-art methods shows that LncPNet achieves better performances in terms of accuracy, F1-score and MCC.

CRBPDL: A deep learning approach for the prediction of circular RNA-RBP interactions

Circular RNAs or circRNAs are non-coding RNA molecules which can bind RBPs and are involved in multiple regulatory processes (Zang et al., 2020). CRBPDL (Table 6) (Niu et al., 2022) is a recently developed method that uses a deep learning approach (also used in other studies, e.g. Pan and Shen, 2018; Zhang et al., 2019; Yang et al., 2021) to predict interactions between circRNAs and proteins. The main improvement of CRBPDL is in the feature encoding step, which is critical for prediction performance. CRBPDL uses five different coding schemes (k-nucleotide frequency, Doc2vec, electron-ion interaction pseudopotential, nucleotide chemical properties, and cumulative nucleotide frequency) for the construction of a feature matrix. The method then uses a deep neural network architecture in order to extract local and global context information and subsequently train the model with a self-attention mechanism checking the robustness of the method. The deep neural network architecture is composed by a ResNet (a deep multi-scale residual network) and a BiGRUs (bidirectional gated recurrent unit) with the final integration of AdaBoost algorithm in order to improve the prediction performances. The authors trained and benchmarked CRBPDL using a circRNAs-RBPs interaction dataset derived from the CircInteractome database (Dudekula et al., 2016), consisting of interactions from 37 CLIP-seq experiments, consistently obtaining better performances when compared with existing methods. CRBPDL encodes different types of information about the sequence of circRNA: the dinucleotide and trinucleotide composition frequency (KNF), the free electron energy (EEIP), and also chemical informations about the nucleotides that compose circRNA sequences. For long-term context dependencies Doc2vec, used as encoding scheme, demonstrated to give a great contribution to the feature representation. CRBPDL was also tested on 31 datasets of linear RNA-RBP interactions, obtaining an average AUC of 0.91, which is significantly higher than the AUCs of other methods (ICIRCRBP-DHN (Yang et al., 2021), CRIP (Zhang et al., 2019), iDeepS (Pan et al., 2018), and CIRCSLNN (Ju et al., 2019)). CRBPDL is available on Github (https://github.com/nmt315320/CRBPDL).

EDLMFC: An ensemble deep learning framework for the prediction of non-coding RNA-RBP interactions

In this section, we discuss a class of ncRNA-RBP interaction predictors not designed for a specific RNA type. A recent computational method developed in this field, called EDLMFC, uses an Ensemble Deep Learning framework with Multi-scale Features Combination (Table 6) (Wang et al., 2021). EDLMFC was trained on ncRNA-RBP interaction pairs derived from the RPI1807, NPInter v2.0, and RPI488 datasets and uses different types of features as input such as the primary sequence and the secondary and tertiary structure of ncRNAs and proteins. Using a greater number of features was shown to increase prediction performance compared with single features. This method combines two different techniques: i) a convolutional neural network (CNN); ii) a bi-directional long short-term memory network (BLSTM). The first one is a deep learning-based method which is used to extract high-level information from the features and the second one is a recurrent neural network method which learns long-range dependencies between features, mainly on sequential data. Finally, a three-layer, fully connected, layer is able to predict ncRNA-protein interactions. In a five-fold cross-validation experiment, EDLMFC obtained better performance than RPITER (Peng et al., 2019), IPMiner (Pan et al., 2016), and CFRP (Dai et al., 2019). Moreover, independent tests demonstrated that EDLMFC can be effectively used to predict potential ncRNA-protein interactions in different organisms.

PRNA: Binding site features enable improvement RNA-protein interaction prediction

For the prediction of RNA-RBP interactions, several methods have been developed in order to find the potential binding sites in RNA or in RBP sequences. One of them is from Liu et al, (2010) (Table 6). In this work the authors highlighted the importance of both sequence and structure features in RNA-binding proteins, that simultaneously contribute towards the recognition of a

specific RNA sequence site. In order to determine in a more comprehensive way the interacting sites in protein sequences, the authors suggested a parameter to consider interaction propensity of an amino acid. This variable represents a measure of mutual dependence of a triplet of amino acids in proteins where the central amino acid binds a nucleotide on the RNA sequence. Then this feature is encoded in a vector of other hybrid features to describe exhaustively the amino acids in the protein sequence. The method was trained using a dataset of protein-RNA complexes obtained from RsiteDB and used to predict RNA binding residues in proteins given the previous set of features, using Random Forest (RF), that with a sliding window of 5 amino acids on the protein sequence predicts the possible site of a binding event. The result in terms of AUC is of 0.905 with a ACC of 81.4% indicating a good performance if compared to other methods (RNABindR (Terribilini et al., 2007), BindN (Wang and Brown, 2006), RNAProB (Cheng et al., 2008), PPRint (Kumar et al., 2008)). In this paper the idea emerges that by integrating the information carried by the neighborhood of an amino acid with other features of the protein sequence and structure analyzed, we can substantially improve the prediction of RNA-RBP interactions by finding the binding sites. A concept well developed also in a recent work of Niu et al. in which instead of focusing on the binding protein sequence, the RNA sequence is fundamental.

PrismNet: A deep learning algorithm to predict RPIs that uses *in vivo* RNA structures

One of the most important factors determining the interaction between RNAs and proteins is the RNA secondary structure (Taliaferro et al., 2016). Therefore, leveraging this feature in prediction models can significantly increase their performance. Although there are different methods for the prediction of RNA secondary structure (Seetin and Mathews, 2012), computational methods based exclusively on the primary sequence do not take into account the dynamic nature of these structures. Indeed, RNA secondary structures are extremely dynamic and can change depending on various factors such as the interaction with chaperones and other RBPs. All these factors, ultimately, vary depending on the cellular conditions *in vivo* (Lewis et al., 2017). PrismNet is an RNA-protein prediction method that leverages experimental data on RNA secondary structures, being capable, in this way, to take into account their dynamism (Table 6). This method is based on secondary structure information obtained *via in vivo* click selective 2′-hydroxyl acylation and profiling experiments (icSHAPE) (Flynn et al., 2016) that were carried out in 7 cell types (i.e. K562, HepG2, HEK293, HEK 293T, HeLa, H9, and mES) in which RNA structures were profiled transcriptome-wide. This data was integrated with RBPs binding sites data from CLIP experiments in the same cell types. To construct the model input, the structure scores derived from the icSHAPE experiments were encoded as a one-dimensional vector and the sequence was represented as a four-dimensional one-hot-encoded vector. The deep learning model consists of a series of convolutional layers, while squeeze-and-excitation networks were used to recalibrate the convolutional channels and residual blocks to capture the joint sequence and structural determinants of RBP binding. The authors compared their model with other computational methods including RCK (Orenstein et al., 2016), GraphProt (Maticzka et al., 2014; Orenstein et al., 2016) and DeepBind (Alipanahi et al., 2015), using the binding sites obtained from the CLIP-seq datasets for each RBP, and obtaining better performance in terms of AUC and AUPRC. Furthermore, by training their model using different combinations of inputs, they observed that the model trained using both the sequence and the experimentally determined RNA secondary structures outperformed other models, demonstrating that experimental information on the RNA secondary structure *in vivo* is critical to the performance improvement.

Computational methods for RNA-RNA interactions prediction

RNAs can also interact with other RNAs and several studies have shown these interactions to be crucially involved in the regulation of gene transcription, cell metabolism, and other key cellular functions (Deogharia and Gurha, 2021; Singh et al., 2022; Wang et al., 2022). Despite the fact that a large number of RNA-RNA interactions have been experimentally validated, many more have yet to be identified. Therefore, several computational methods have been developed for the prediction of RNA-RNA interaction, many of which are based on sequence complementarity (Kang et al., 2020, 2021; Yang et al., 2020). In the last 5 years, these methods have been gradually revolutionized by the introduction of deep learning approaches borrowed from the field of natural language processing. PreMLI is one of the latest methods in this field, it was published in early 2022 by Yu and collaborators (Table 6) (Yu et al., 2022), and, currently, it achieves better overall performance compared with other existing methods. This method was specifically built to predict miRNA-lncRNA interactions and relies exclusively on RNA sequence information. PreMLI was trained using a plant lncRNA-miRNA interaction dataset, constructed using RNAHybrid 2.1.2. The approach consists of three steps: i) in the pre-training phase the RNA sequences are used as input for rna2vec training in order to obtain a weight matrix that better describes the RNA sequence and can be used as the input to the next step; ii) deep feature mining approaches, based on Convolutional Neural Network, Bidirectional Gated Recurrent Unit, and attention layers are used to obtain additional potential

features; iii) in the last step the two feature vectors are connected as input to the prediction layer. The authors demonstrate how the pre-training and the deep feature mining phases improve prediction performance and, furthermore, they show how this method performs better than already existing advanced RNA-RNA interaction predictors in terms of sensitivity, specificity, and AUC. Although the pre-training step improves the model performance, it also increases the computational time required for the entire prediction process. Moreover, this method is optimized for the prediction of miRNA-lncRNA interactions in plants. In order to extend its use to other types of RNA-RNA interactions or other organisms the model needs to be trained on an appropriate specific dataset and the hyperparameters need to be adjusted.

Conclusion

In the last few years several studies have explored the RNA interactions landscape, given the crucial role that RNA-RBPs and RNA-RNA networks play in cell biology. Despite the advances made so far, novel experimental methods for the identification of binding sites (such as HITS-CLIP and PAR-CLIP) are still time-consuming and cost-intensive. That is why computational approaches represent a complementary strategy to guide experimental work. In this review, we provide an overview of the most recent prediction methods. We summarize recent advances in the algorithms developed to solve specific tasks, such as circRNA- or lncRNA-RBPs interaction predictions or, more generally RNA-RBPs interactions. Besides, we highlight how the development of a larger dataset of interactions is crucial to increase performance. Lastly, despite the fact that many methods rely only on sequence information, among the ones analyzed, those that obtain the best performances tend to include a variety of different biological features. Performance comparison of the described methods shows how the inclusion of structure information contributes to improving the accuracy and efficiency of the models. Only one of the described methods uses both RNAs and proteins structural information as input features for the predictive model because if, on the one hand, a large number of reliable protein structures is available, on the other hand, RNA structures are mainly obtained through computational prediction. RNA structure uncertainty could add noise to the model, resulting in untruthful prediction performances. The prediction of protein structure has reached satisfactory levels of performance thanks to the development of AlphaFold (Jumper et al., 2021). Conversely, RNA structure prediction still lags far behind. One of the main limitations is the paucity of known RNA structures that can be used for model training. To address this issue a new deep learning model called Atomic Rotationally Equivariant Scorer (ARES) has been developed (Townshend et al., 2021). ARES achieves good performances in the prediction of RNA structures, based on a training dataset of only 18 experimentally determined RNA structures. While this is a useful development, further work is needed in this area. Ultimately, as demonstrated by the methods described in this review, the availability of high-quality RNA structure predictions could greatly improve the inference of RNA-RBP and RNA-RNA interactions. Moreover, the advances in RNA secondary structure determination methods, that takes into account the information from biochemical assay like icSHAPE-seq (Flynn et al., 2016), could improve the confidence of such information as a feature for prediction models, likely leading to an improvement of their performance.

Author contributions

GP, RA, CC, FB performed the literature research and wrote the paper; PFG, MH-C reviewed the manuscript writing; GP, PFG, MH-C supervised the study.

Funding

AIRC project [to MHC] (grant number IG 23539) funded by the European Union – NextGenerationEU: National Center for Gene Therapy and Drugs based on RNA Technology.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adinolfi, M., Pietrosanto, M., Parca, L., Ausiello, G., Ferrè, F., and Helmer-Citterich, M. (2019). Discovering sequence and structure landscapes in RNA interaction motifs. *Nucleic Acids Res.* 47, 4958–4969. doi:10.1093/nar/gkz250

Akbaripour-Elahabad, M., Zahiri, J., Rafeh, R., Eslami, M., and Azari, M. (2016). rpiCOOL: A tool for *in silico* RNA–protein interaction detection using random forest. *J. Theor. Biol.* 402, 1–8. doi:10.1016/j.jtbi.2016.04.025

Alipanahi, B., Delong, A., Weirauch, M. T., and Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33, 831–838. doi:10.1038/nbt.3300

Bai, Y., Dai, X., Ye, T., Zhang, P., Yan, X., Gong, X., et al. (2019). PlncRNADB: A repository of plant lncRNAs and lncRNA-RBP protein interactions. *Curr. Bioinform.* 14, 621–627. doi:10.2174/1574893614666190131161002

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., et al. (2021). RCSB protein data bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* 49, D437–D451. doi:10.1093/nar/gkaa1038

Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., et al. (2013). LncRNADisease: A database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* 41, D983–D986. doi:10.1093/nar/gks1099

Chen, Q., Meng, X., Liao, Q., and Chen, M. (2019). Versatile interactions and bioinformatics analysis of noncoding RNAs. *Brief. Bioinform.* 20, 1781–1794. doi:10.1093/bib/bby050

Cheng, C.-W., Su, E. C.-Y., Hwang, J.-K., Sung, T.-Y., and Hsu, W.-L. (2008). Predicting RNA-binding sites of proteins using support vector machines and evolutionary information. *BMC Bioinforma.* 9, S6. doi:10.1186/1471-2105-9-S12-S6

Cheng, Z., Huang, K., Wang, Y., Liu, H., Guan, J., and Zhou, S. (2017). Selecting high-quality negative samples for effectively predicting protein-RNA interactions. *BMC Syst. Biol.* 11, 9. doi:10.1186/s12918-017-0390-8

Cheng, Z., Zhou, S., and Guan, J. (2015). Computationally predicting protein-RNA interactions using only positive and unlabeled examples. *J. Bioinform. Comput. Biol.* 13, 1541005. doi:10.1142/s021972001541005x

Coimbatore Narayanan, B., Westbrook, J., Ghosh, S., Petrov, A. I., Sweeney, B., Zirbel, C. L., et al. (2014). The nucleic acid database: New features and capabilities. *Nucleic Acids Res.* 42, D114–D122. doi:10.1093/nar/gkt980

Dai, Q., Guo, M., Duan, X., Teng, Z., and Fu, Y. (2019). Construction of complex features for computational predicting ncRNA-protein interaction. *Front. Genet.* 10, 18. doi:10.3389/fgene.2019.00018

Deng, L., Wang, J., Xiao, Y., Wang, Z., and Liu, H. (2018). Accurate prediction of protein-lncRNA interactions by diffusion and HeteSim features across heterogeneous network. *BMC Bioinforma.* 19, 370. doi:10.1186/s12859-018-2390-0

Deogharia, M., and Gurha, P. (2021). *The "guiding" principles of noncoding RNA function*. New Jersey, United States: Wiley Interdiscip. Rev. RNA, e1704.

Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., et al. (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell.* 70, 854–867. e9. doi:10.1016/j.molcel.2018.05.001

Dong, Y., Chawla, N. V., and Swami, A. (2017). metapath2vec. *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* doi:10.1145/3097983.3098036

Dudekula, D. B., Panda, A. C., Grammatikakis, I., De, S., Abdelmohsen, K., and Gorospe, M. (2016). CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol.* 13, 34–42. doi:10.1080/15476286.2015.1128065

Fan, X.-N., and Zhang, S.-W. (2019). LPI-BLS: Predicting lncRNA–protein interactions with a broad learning system-based stacked ensemble classifier. *Neurocomputing* 370, 88–93. doi:10.1016/j.neucom.2019.08.084

Ferrè, F., Colantoni, A., and Helmer-Citterich, M. (2015). Revealing protein–lncRNA interaction. *Brief. Bioinform.* 17, 106–116. doi:10.1093/bib/bbv031

Flynn, R. A., Zhang, Q. C., Spitale, R. C., Lee, B., Mumbach, M. R., and Chang, H. Y. (2016). Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc.* 11, 273–290. doi:10.1038/nprot.2016.011

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi:10.1093/bioinformatics/bts565

Ge, M., Li, A., and Wang, M. (2016). A bipartite network-based method for prediction of long non-coding RNA-protein interactions. *Genomics Proteomics Bioinforma.* 14, 62–71. doi:10.1016/j.gpb.2016.01.004

Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* 22, 185–198. doi:10.1038/s41576-020-00302-y

Gilbertson, S., Federspiel, J. D., Hartenian, E., Cristea, I. M., and Glaunsinger, B. (2018). Changes in mRNA abundance drive shuttling of RNA binding proteins, linking cytoplasmic RNA degradation to transcription. *Elife* 7, e37663. doi:10.7554/eLife.37663

Guarracino, A., Pepe, G., Ballesio, F., Adinolfi, M., Pietrosanto, M., Sangiovanni, E., et al. (2021). Brio: A web server for RNA sequence and structure motif scan. *Nucleic Acids Res.* 49, W67–W71. doi:10.1093/nar/gkab400

Ju, Y., Yuan, L., Yang, Y., and Zhao, H. (2019). CircSLNN: Identifying RBP-binding sites on circRNAs via sequence labeling neural networks. *Front. Genet.* 10, 1184. doi:10.3389/fgene.2019.01184

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Kang, J., Tang, Q., He, J., Li, L., Yang, N., Yu, S., et al. (2022). RNAInter v4.0: RNA interactome repository with redefined confidence scoring system and improved accessibility. *Nucleic Acids Res.* 50, D326–D332. doi:10.1093/nar/gkab997

Kang, Q., Meng, J., Cui, J., Luan, Y., and Chen, M. (2020). PmliPred: A method based on hybrid model and fuzzy decision for plant miRNA–lncRNA interaction prediction. *Bioinformatics* 36, 2986–2992. doi:10.1093/bioinformatics/btaa074

Kang, Q., Meng, J., Shi, W., and Luan, Y. (2021). Ensemble deep learning based on multi-level information enhancement and greedy fuzzy decision for plant miRNA-lncRNA interaction prediction. *Interdiscip. Sci.* 13, 603–614. doi:10.1007/s12539-021-00434-7

Kelaini, S., Chan, C., Cornelius, V. A., and Margariti, A. (2021). RNA-binding proteins hold key roles in function, dysfunction, and disease. *Biology* 10, 366. doi:10.3390/biology10050366

Kornienko, A. E., Dotter, C. P., Guenzl, P. M., Gisslinger, H., Gisslinger, B., Cleary, C., et al. (2016). Long non-coding RNAs display higher natural expression variation than protein-coding genes in healthy humans. *Genome Biol.* 17, 14. doi:10.1186/s13059-016-0873-8

Kumar, M., Gromiha, M. M., and Raghava, G. P. S. (2008). Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins* 71, 189–194. doi:10.1002/prot.21677

Lewis, B. A., Walia, R. R., Terribilini, M., Ferguson, J., Zheng, C., Honavar, V., et al. (2011). Pridb: A protein-RNA interface database. *Nucleic Acids Res.* 39, D277–D282. doi:10.1093/nar/gkq1108

Lewis, C. J. T., Pan, T., and Kalsotra, A. (2017). RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat. Rev. Mol. Cell. Biol.* 18, 202–210. doi:10.1038/nrm.2016.163

Li, A., Ge, M., Zhang, Y., Peng, C., and Wang, M. (2015). Predicting long noncoding RNA and protein interactions using heterogeneous network model. *Biomed. Res. Int.* 2015, 671950. doi:10.1155/2015/671950

Li, Y., Sun, H., Feng, S., Zhang, Q., Han, S., and Du, W. (2021). Capsule-LPI: A LncRNA-protein interaction predicting tool based on a capsule network. *BMC Bioinforma.* 22, 246. doi:10.1186/s12859-021-04171-y

Liu, Z.-P., Wu, L.-Y., Wang, Y., Zhang, X.-S., and Chen, L. (2010). Prediction of protein-RNA binding sites by a random forest method with combined features. *Bioinformatics* 26, 1616–1622. doi:10.1093/bioinformatics/btq253

Luna, J. M., Scheel, T. K. H., Danino, T., Shaw, K. S., Mele, A., Fak, J. J., et al. (2015). Hepatitis C virus RNA functionally sequesters miR-122. *Cell.* 160, 1099–1110. doi:10.1016/j.cell.2015.02.025

Márquez, B. G., and Castro Amaya, J. (2019). BIOPROT contenedor autónomo de Residuos biológicos. *Rev. Colomb. De. Tecnol. De. Av. (RCTA)* 1, 33. doi:10.24054/16927257.v33.n33.2019.3330

Maticzka, D., Lange, S. J., Costa, F., and Backofen, R. (2014). GraphProt: Modeling binding preferences of RNA-binding proteins. *Genome Biol.* 15, R17. doi:10.1186/gb-2014-15-1-r17

Muhammod, R., Ahmed, S., Md Farid, D., Shatabda, S., Sharma, A., and Dehzangi, A. (2019). PyFeat: A python-based effective feature generation tool for DNA, RNA and protein sequences. *Bioinformatics* 35, 3831–3833. doi:10.1093/bioinformatics/btz165

Muppirala, U. K., Honavar, V. G., and Dobbs, D. (2011). Predicting RNA-protein interactions using only sequence information. *BMC Bioinforma.* 12, 489. doi:10.1186/1471-2105-12-489

Newman, R., McHugh, J., and Turner, M. (2015). RNA binding proteins as regulators of immune cell biology. *Clin. Exp. Immunol.* 183, 37–49. doi:10.1111/cei.12684

Niu, M., Zou, Q., and Lin, C. (2022). Crbpdl: Identification of circRNA-RBP interaction sites using an ensemble neural network approach. *PLoS Comput. Biol.* 18, e1009798. doi:10.1371/journal.pcbi.1009798

Orenstein, Y., Wang, Y., and Berger, B. (2016). Rck: Accurate and efficient inference of sequence- and structure-based protein-RNA binding models from RNAcompete data. *Bioinformatics* 32, i351–i359. doi:10.1093/bioinformatics/btw259

Pan, X., Fan, Y.-X., Yan, J., and Shen, H.-B. (2016). IPMiner: Hidden ncRNA-protein interaction sequential pattern mining with stacked autoencoder for accurate computational prediction. *BMC Genomics* 17, 582. doi:10.1186/s12864-016-2931-8

Pan, X., Rijnbeek, P., Yan, J., and Shen, H.-B. (2018). Prediction of RNA-protein sequence and structure binding preferences using deep convolutional and recurrent neural networks. *BMC Genomics* 19, 511. doi:10.1186/s12864-018-4889-1

Pan, X., and Shen, H.-B. (2018). Predicting RNA-protein binding sites and motifs through combining local and global deep convolutional neural networks. *Bioinformatics* 34, 3427–3436. doi:10.1093/bioinformatics/bty364

Peng, C., Han, S., Zhang, H., and Li, Y. (2019). Rpiter: A hierarchical deep learning framework for ncRNA–Protein interaction prediction. *Int. J. Mol. Sci.* 20, E1070. doi:10.3390/ijms20051070

Peng, S., Guo, D., Guo, Y., Zhao, H., Mei, J., Han, Y., et al. (2022). CONSTITUTIVE EXPRESSER OF PATHOGENESIS-RELATED GENES 5 is an RNA-binding protein controlling plant immunity via an RNA processing complex. *Plant Cell.* 34, 1724–1744. doi:10.1093/plcell/koac037

Pepe, G., Guarracino, A., Ballesio, F., Parca, L., Ausiello, G., and Helmer-Citterich, M. (2022a). Evaluation of potential miRNA sponge effects of SARS genomes in human. *Noncoding. RNA Res.* 7, 48–53. doi:10.1016/j.ncrna.2022.01.003

Pepe, G., Parca, L., Viviani, L., Ausiello, G., and Helmer-Citterich, M. (2022b). Variation in the co-expression profile highlights a loss of miRNA-mRNA regulation in multiple cancer types. *Noncoding. RNA Res.* 7, 98–105. doi:10.1016/j.ncrna.2022.03.003

Pereira, B., Billaud, M., and Almeida, R. (2017). RNA-binding proteins in cancer: Old players and new actors. *Trends Cancer* 3, 506–528. doi:10.1016/j.trecan.2017.05.003

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038. doi:10.1038/nature09144

Qiao, Y., Zhao, X., Liu, J., and Yang, W. (2019). Epstein-Barr virus circRNAome as host miRNA sponge regulates virus infection, cell cycle, and oncogenesis. *Bioengineered* 10, 593–603. doi:10.1080/21655979.2019.1679698

Ramanathan, M., Porter, D. F., and Khavari, P. A. (2019). Methods to study RNA-protein interactions. *Nat. Methods* 16, 225–234. doi:10.1038/s41592-019-0330-1

Riley, K. J., Rabinowitz, G. S., Yario, T. A., Luna, J. M., Darnell, R. B., and Steitz, J. A. (2012). EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J.* 31, 2207–2221. doi:10.1038/emboj.2012.63

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell.* 146, 353–358. doi:10.1016/j.cell.2011.07.014

Schieweck, R., Ninkovic, J., and Kiebler, M. A. (2021). RNA-binding proteins balance brain function in health and disease. *Physiol. Rev.* 101, 1309–1370. doi:10.1152/physrev.00047.2019

Seetin, M. G., and Mathews, D. H. (2012). RNA structure prediction: An overview of methods. *Methods Mol. Biol.* 905, 99–122. doi:10.1007/978-1-61779-949-5_8

Seitz, H. (2009). Redefining microRNA targets. *Curr. Biol.* 19, 870–873. doi:10.1016/j.cub.2009.03.059

Singh, S., Shyamal, S., and Panda, A. C. (2022). *Detecting RNA-RNA interactome*. New Jersey, United States: Wiley Interdiscip. Rev. RNA, e1715.

Sun, L., Xu, K., Huang, W., Yang, Y. T., Li, P., Tang, L., et al. (2021). Predicting dynamic cellular protein-RNA interactions by deep learning using *in vivo* RNA structures. *Cell Res.* 31, 495–516.

Suresh, V., Liu, L., Adjeroh, D., and Zhou, X. (2015). RPI-pred: Predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res.* 43, 1370–1379. doi:10.1093/nar/gkv020

Taliaferro, J. M., Lambert, N. J., Sudmant, P. H., Dominguez, D., Merkin, J. J., Alexis, M. S., et al. (2016). RNA sequence context effects measured *in vitro* predict *in vivo* protein binding and regulation. *Mol. Cell.* 64, 294–306. doi:10.1016/j.molcel.2016.08.035

Teng, X., Chen, X., Xue, H., Tang, Y., Zhang, P., Kang, Q., et al. (2019). NPInter v4.0: An integrated database of ncRNA interactions. *Nucleic Acids Res.* 48, D160–D165. doi:10.1093/nar/gkz969

Terribilini, M., Sander, J. D., Lee, J.-H., Zaback, P., Jernigan, R. L., Honavar, V., et al. (2007). RNABindR: A server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res.* 35, W578–W584. doi:10.1093/nar/gkm294

Townshend, R. J. L., Eismann, S., Watkins, A. M., Rangan, R., Karelina, M., Das, R., et al. (2021). Geometric deep learning of RNA structure. *Science* 373, 1047–1051. doi:10.1126/science.abe5650

Turner, M., and Díaz-Muñoz, M. D. (2018). RNA-binding proteins control gene expression and cell fate in the immune system. *Nat. Immunol.* 19, 120–129. doi:10.1038/s41590-017-0028-4

Van Assche, E., Van Puyvelde, S., Vanderleyden, J., and Steenackers, H. P. (2015). RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front. Microbiol.* 6, 141 doi:10.3389/fmicb.2015.00141

Velankar, S., Burley, S. K., Kurisu, G., Hoch, J. C., and Markley, J. L. (2021). The protein data bank archive. *Methods Mol. Biol.* 2305, 3–21. doi:10.1007/978-1-0716-1406-8_1

Wang, D., Ye, R., Cai, Z., and Xue, Y. (2022). *Emerging roles of RNA-RNA interactions in transcriptional regulation*. New Jersey, United States: Wiley Interdiscip. Rev. RNA, e1712.

Wang, J., Zhao, Y., Gong, W., Liu, Y., Wang, M., Huang, X., et al. (2021). Edlmfc: An ensemble deep learning framework with multi-scale features combination for ncRNA–protein interaction prediction. *BMC Bioinforma.* 22, 133. doi:10.1186/s12859-021-04069-9

Wang, L., and Brown, S. J. (2006). BindN: A web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Res.* 34, W243–W248. doi:10.1093/nar/gkl298

Wang, L., Yan, X., Liu, M.-L., Song, K.-J., Sun, X.-F., and Pan, W.-W. (2019). Prediction of RNA-protein interactions by combining deep convolutional neural network with feature selection ensemble method. *J. Theor. Biol.* 461, 230–238. doi:10.1016/j.jtbi.2018.10.029

Wekesa, J. S., Meng, J., and Luan, Y. (2020). Multi-feature fusion for deep learning to predict plant lncRNA-protein interaction. *Genomics* 112, 2928–2936. doi:10.1016/j.ygeno.2020.05.005

Xie, G., Wu, C., Sun, Y., Fan, Z., and Liu, J. (2019). LPI-IBNRA: Long non-coding RNA-protein interaction prediction based on improved bipartite network recommender algorithm. *Front. Genet.* 10, 343. doi:10.3389/fgene.2019.00343

Yang, J., Li, A., Ge, M., and Wang, M. (2016). Relevance search for predicting lncRNA–protein interactions based on heterogeneous network. *Neurocomputing* 206, 81–88. doi:10.1016/j.neucom.2015.11.109

Yang, S., Wang, Y., Lin, Y., Shao, D., He, K., and Huang, L. (2020). LncMirNet: Predicting LncRNA–miRNA interaction based on deep learning of ribonucleic acid sequences. *Molecules* 25, 4372. doi:10.3390/molecules25194372

Yang, Y., Hou, Z., Ma, Z., Li, X., and Wong, K.-C. (2021). iCircRBP-DHN: identification of circRNA-RBP interaction sites using deep hierarchical network. *Brief. Bioinform.* 22, bbaa274. doi:10.1093/bib/bbaa274

Yi, H.-C., You, Z.-H., Huang, D.-S., Li, X., Jiang, T.-H., and Li, L.-P. (2018). A deep learning framework for robust and accurate prediction of ncRNA-protein interactions using evolutionary information. *Mol. Ther. Nucleic Acids* 11, 337–344. doi:10.1016/j.omtn.2018.03.001

Yi, H.-C., You, Z.-H., Cheng, L., Zhou, X., Jiang, T.-H., Li, X., et al. (2020). Learning distributed representations of RNA and protein sequences and its application for predicting lncRNA-protein interactions. *Comput. Struct. Biotechnol. J.* 18, 20–26. doi:10.1016/j.csbj.2019.11.004

Yu, X., Jiang, L., Jin, S., Zeng, X., and Liu, X. (2022). preMLI: a pre-trained method to uncover microRNA–lncRNA potential interactions. *Brief. Bioinform.* 23, bbab470. doi:10.1093/bib/bbab470

Yuan, J., Wu, W., Xie, C., Zhao, G., Zhao, Y., and Chen, R. (2014). NPInter v2.0: An updated database of ncRNA interactions. *Nucleic Acids Res.* 42, D104–D108. doi:10.1093/nar/gkt1057

Zang, J., Lu, D., and Xu, A. (2020). The interaction of circRNAs and RNA binding proteins: An important part of circRNA maintenance and function. *J. Neurosci. Res.* 98, 87–97. doi:10.1002/jnr.24356

Zhan, Z.-H., You, Z.-H., Li, L.-P., Zhou, Y., and Yi, H.-C. (2018). Accurate prediction of ncRNA-protein interactions from the integration of sequence and evolutionary information. *Front. Genet.* 9, 458. doi:10.3389/fgene.2018.00458

Zhang, K., Pan, X., Yang, Y., and Shen, H.-B. (2019). Crip: Predicting circRNA-RBP-binding sites using a codon-based encoding and hybrid deep neural networks. *RNA* 25, 1604–1615. doi:10.1261/rna.070565.119

Zhang, S.-W., Zhang, X.-X., Fan, X.-N., and Li, W.-N. (2020). LPI-CNNCP: Prediction of lncRNA-protein interactions by using convolutional neural network with the copy-padding trick. *Anal. Biochem.* 601, 113767. doi:10.1016/j.ab.2020.113767

Zhang, W., Qu, Q., Zhang, Y., and Wang, W. (2018). The linear neighborhood propagation method for predicting long non-coding RNA–protein interactions. *Neurocomputing* 273, 526–534. doi:10.1016/j.neucom.2017.07.065

Zhao, G., Li, P., Qiao, X., Han, X., and Liu, Z.-P. (2021). Predicting lncRNA-protein interactions by heterogenous network embedding. *Front. Genet.* 12, 814073. doi:10.3389/fgene.2021.814073

Zhao, Q., Yu, H., Ming, Z., Hu, H., Ren, G., and Liu, H. (2018). The bipartite network projection-recommended algorithm for predicting long non-coding RNA-protein interactions. *Mol. Ther. Nucleic Acids* 13, 464–471. doi:10.1016/j.omtn.2018.09.020

Zheng, X., Wang, Y., Tian, K., Zhou, J., Guan, J., Luo, L., et al. (2017). Fusing multiple protein-protein similarity networks to effectively predict lncRNA-protein interactions. *BMC Bioinforma.* 18, 420. doi:10.1186/s12859-017-1819-1

Zhou, L., Wang, Z., Tian, X., and Peng, L. (2021). LPI-deepGBDT: A multiple-layer deep framework based on gradient boosting decision trees for lncRNA–protein interaction identification. *BMC Bioinforma.* 22, 479. doi:10.1186/s12859-021-04399-8

Zhou, Y.-K., Hu, J., Shen, Z.-A., Zhang, W.-Y., and Du, P.-F. (2020). LPI-SKF: Predicting lncRNA-protein interactions using similarity kernel fusions. *Front. Genet.* 11, 615144. doi:10.3389/fgene.2020.615144

frontiers | Frontiers in Molecular Biosciences

TYPE Mini Review
PUBLISHED 02 November 2022
DOI 10.3389/fmolb.2022.950408

OPEN ACCESS

EDITED BY
Gian Gaetano Tartaglia,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Patrick O'Donoghue,
Western University, Canada

*CORRESPONDENCE
Tao Sun,
suntao_nxmu@163.com
Feng Wang,
nxwwang@163.com

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 22 May 2022
ACCEPTED 18 October 2022
PUBLISHED 02 November 2022

CITATION
Li W, Wang YY, Xiao L, Ding J, Wang L, Wang F and Sun T (2022), Mysterious long noncoding RNAs and their relationships to human disease. *Front. Mol. Biosci.* 9:950408. doi: 10.3389/fmolb.2022.950408

Mysterious long noncoding RNAs and their relationships to human disease

Wenchao Li[1,2], Yang Yang Wang[1], Lifei Xiao[1], Jiangwei Ding[1], Lei Wang[1], Feng Wang[3]* and Tao Sun[1]*

[1]Ningxia Key Laboratory of Craniocerebral Disease, Incubation Base of National Key Laboratory, Ningxia Medical University, Yinchuan, China, [2]The First Affiliated Hospital of Xinxiang Medical University, Xinxiang, China, [3]Zhejiang University School of Medicine, Hangzhou, China

Increasingly studies have shown that the formation mechanism of many human diseases is very complex, which is determined by environmental factors and genetic factors rather than fully following Mendel's genetic law of inheritance. Long non-coding RNA (lncRNA) is a class of endogenous non-protein coding RNA with a length greater than 200 nt, which has attracted much attention in recent years. Studies have shown that lncRNAs have a wide range of biological functions, such as roles in gene imprinting, cell cycle progression, apoptosis, senescence, cell differentiation, and stress responses, and that they regulate the life processes of mammals at various levels, such as epigenetic transcription, processing, modification, transport, translation and degradation. Analyzing the characteristics of lncRNAs and revealing their internal roles can not only deepen our understanding of human physiological and pathological processes, but also provide new ideas and solutions for the diagnosis, prevention and treatment of some diseases. This article mainly reviews the biological characteristics of lncRNAs and their relationship with some diseases, so as to provide references for the related research of lncRNAs.

KEYWORDS

noncoding RNA, tumors, nervous system, disease, epigenetics

Introduction

Ribose nucleic acid (RNA) is a single strand formed by transcribing a deoxyribonucleic acid (DNA) strand as a template according to the principle of complementary base pairing. In recent years, with the rapid development and improvement of next-generation high-throughput sequencing technologies, our understanding of the field of transcription has been expanded. An important finding is that although nearly 80% of the genome can be transcribed, only approximately 2% of mammalian genomes consist of protein-coding genes, most of the rest of the genome sequence only produces a large amount of noncoding RNA (ncRNA). These RNAs can only be successfully transcribed and lack the ability to directly encode proteins, only a few lncRNAs can encode small molecular peptides. Previously, researchers believed that non-coding RNA, as a kind of gene "garbage",

had no special biological effect (Wang and Chang, 2011). These non-coding RNAs are roughly divided into two classes, depending on the length of the transcript. That is, small noncoding RNA (snRNA) with a length less than 200 nt and long non-coding RNA (lncRNA) with a length greater than 200 nt. Interesting, lncRNAs are the largest component of non-coding transcripts in mammals, accounting for about 80% (Vennin and Adriaenssens, 2018).

With further research, scholars found that lncRNAs can play the role of "regulatory factors" to participate in various levels of transcription, post-transcription and translation, and also participate in various biological processes, such as embryo development invasion, apoptosis, metastasis and angiogenesis (Renoux and Todd, 2012; Mazar et al., 2014). In cancer, metabolic diseases and nervous system diseases, lncRNA expression profile is also changed, and dysregulation of lncRNA expression is closely related to the occurrence and development of diseases. Moreover, studies have revealed that the expression of lncRNAs is more cell-specific than that of protein-coding genes. A considerable number of lncRNAs showed 3′ polyadenylation and 5′ caprization, with multiple exons, and showed transcriptional activation activity similar to mRNA (Cabili et al., 2011). The biological characteristics, functions, and their relationship with diseases of lncRNAs have become a research hotspot.

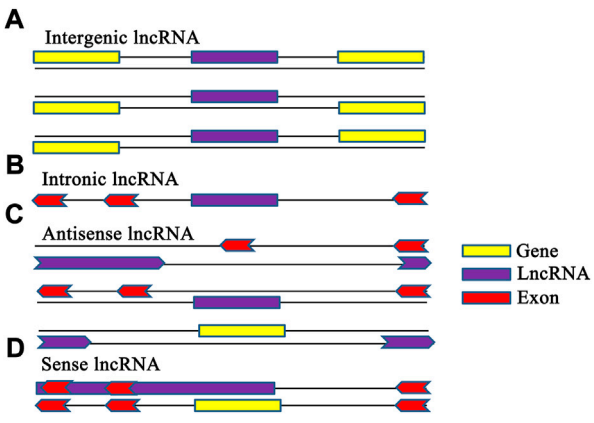


FIGURE 1
Classification of lncRNAs. Panel **(A)**: Intergenic lncRNA was transcribed intergenically from both strands. Panel **(B)**: intronic lncRNA, transcribed entirely from introns of protein-coding genes. Panel **(C)**: antisense lncRNA, transcribed from the antisense strand of protein-coding genes, overlapping with exonic or intronic regions or covering the entire protein-coding sequence through an intron. Panel **(D)**: Sense lncRNA was transcribed from the sense strand of protein-coding genes and contain exons from protein-coding genes, overlapping with part of protein-coding genes or covering the entire sequence of a protein-coding gene through an intron.

Biological characteristics of long non-coding RNAs

Classification of long non-coding RNAs

In order to carry out more in-depth research on lncRNAs and understand their functions and mechanisms, researchers divided lncRNAs into intronic lncRNAs, intergenic lncRNAs, sense lncRNAs and antisense lncRNAs four categories according to their location and background in the genome (Ma et al., 2013). Intergenic lncRNAs may be located more than 10 kb from any nearby protein-coding locus. But, intronic lncRNAs originate from long introns that are transcribed from the same strand as the associated protein-coding genes. Studies suggest that intergenic lncRNAs and intronic lncRNAs may play regulatory roles through different transcriptional activation mechanisms, and may have different poly (A) modification and expression activities at different locations of cells (Cheng et al., 2005; Prensner et al., 2011). Antisense lncRNAs are lncRNAs that are transcribed from the antisense strand of a gene locus and overlap with the RNA transcribed from the sense strand. In contrast, sense lncRNAs are lncRNAs that contain a protein coding gene and are transcribed in the same direction as that gene (Figure 1).

The mode of action of long non-coding RNAs

As another new field in molecular biology, with the in-depth study of lncRNAs, researchers have found that the main function of lncRNA may be to combine with DNA, RNA and protein to play a regulatory role (Zhu et al., 2013). As a regulatory molecule of gene expression, the mode of action of lncRNAs can be roughly divided into three aspects. One approach is that lncRNAs affect gene expression by regulating chromatin epigenetic modifications (Figure 2B). In 2013, Di ruscio et al. (2013) showed that the lncRNA derived from the CCAAT-enhancer binding protein-alpha (CEBPA) gene can interact with the DNA methyltransferase DNMT1 to regulate the methylation level at the CEBPA gene locus, thereby promoting gene expression.

As a molecular scaffold or bridge, regulating the related chromatin modifying enzymes in the transcription process, thereby changing the level of gene expression is another mode of action of lncRNAs (Figures 2A,D,E). The non-coding RNA snRNA7SK, the guardian of transcriptional termination and bidirectional transcription in embryonic stem cells, acts as a scaffold for the protein complex formed by HEXIM1, HEXIM2, LARP7 and P-TEFb. The formation of this complex causes the loss of P-TEFb kinase activity and affects the transcriptional activity of related genes (Castelo-branco et al., 2013).

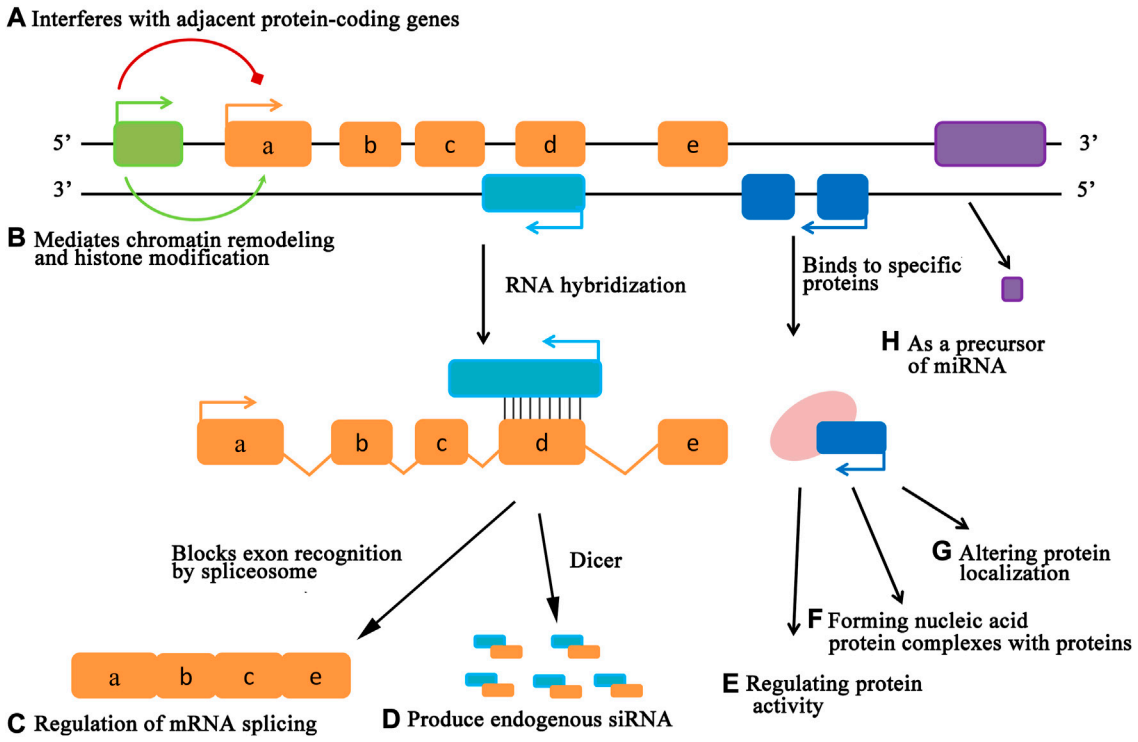


FIGURE 2
The mode of action of lncRNAs. LncRNAs can regulate gene expression levels at various levels, including epigenetic regulation, transcriptional regulation, and post-transcriptional regulation.

The regulation of gene expression by lncRNA in the post-transcription process is mainly achieved by the level regulation of mRNA and miRNA (Figures 2C,F–H). Cesana et al. (2011) found in their study that LINCMD1 could bind to corresponding miRNA, thus blocking the binding of target gene and miRNA, so as to lose its transcriptional inhibition of target gene.

Relationship between long non-coding RNA and disease

Tumors

At present, the most in-depth research on the function of lncRNAs was their role in tumor diseases. A large number of studies have shown that lncRNAs play an extremely important role in the occurrence and development of tumors (Harries, 2012; Kitagawa et al., 2012). Interestingly, lncRNAs not only promote tumor formation as a proto-oncogene, but also inhibit tumor cell proliferation and migration as a tumor suppressor gene. The occurrence and development of many cancers are accompanied by abnormal expression of lncRNAs. Zhu et al. analyzed the bladder cancer pathological tissues by microarray showed that 3324 lncRNAs were abnormally express, including 22 lncRNAs were significantly up-regulated and 88 lncRNAs were significantly down-regulated (≥8-fold change) in the tumor group compared with the controls. Further verification experiments found that the expression changes of lncRNA TNXA, CTA-134P22.2, CTC-276P9.1 and KRT19P3 were highly consistent with microarray data. They also observed that down-regulated lncRNAs were more common than up-regulated lncRNAs (Zhu et al., 2014a). In addition, studies have confirmed that lncRNA H19 is highly expressed in bladder cancer patients, and the expression level was 3 times higher than that in normal patients. Then *in vivo* and *in vitro* experiments have further confirmed that up-regulation of lncRNA H19 can accelerate the metastasis of bladder cancer cells. The researchers believe that lncRNA H19 can first bind to enhancer of zeste homolog 2 (EZH2), then activate Wnt/β-catenin, and down-regulate E-cadherin (E-cad). Therefore, lncRNA H19 may promote bladder cancer metastasis by combining with EZH2 and inhibiting the expression of E-cad (Luo et al., 2013).

It is worth noting that, in addition to bladder cancer, the researchers found that lncRNA H19 also plays a regulatory function in prostate cancer. LncRNA H19-mir675 axis may inhibit the metastasis of prostate cancer by transforming growth factor beta induced protein (TGFBI) (Zhu et al.,

2014b). Among them, TGFBI is closely associated with cancer metastasis, and miR675 can directly bind to the 3′UTR of TGFBI-mRNA to inhibit the normal translation of TGFBI. HOTAIRM1, as lincRNA (long intergenic noncoding RNA), was found to be highly specifically expressed in mature myeloid cells. The barretts esophagus (BE) is considered the precancerous lesion of the esophageal adenocarcinoma (EAC), which results in 80 percent of esophageal adenocarcinoma. Previous studies have reported that lncRNA AFAP1-AS1 is highly expressed in esophageal adenocarcinoma and Barrett esophageal tissues. Targeted silencing of lncRNA AFAP1-AS1 can inhibit the differentiation of EAC, promote its apoptosis, and inhibit the invasion and metastasis of tumor cells without affecting the expression of AFAP1 protein (Wu et al., 2013).

Studies related to urological tumors have shown that lncRNA urothelialcarcinorna associated 1 (UCA1) can promote the proliferation of bladder cancer cells BLS-211 by regulating genes, and can enhance its drug resistance (Wang et al., 2008). A study conducted by Silva et al. (2011) shown long stress induced non-coding transcripts 5 (LSINCT5), is highly expressed in breast and ovarian cancer. LSINCT5 has greater than ten-fold increased expression in all cancer cell lines tested as compared to normal cell lines from the same tissues. Several breast cancer cell lines had 30-fold higher expression of LSINCT5 than in HMEC, including MDA468, T47D and BT474 cells. Targeted knockout of lncRNA LSINCT5 significantly inhibited the proliferation of cancer cells. Other studies have found that in breast cancer, the expression level of growth rest gene transcript 5 (Gas5) tends to reduce to 34% of those of adjacent normal tissue, while the Gas5 can induce cell apoptosis by regulating the target gene and play the function of tumor suppressor gene (Mourtada-maarabouni et al., 2009). In addition, Askarian-amiri et al. (2011) confirmed zinc finger antisense 1 (ZFAS1) was also significantly down-regulated in breast cancer. The result shows ZFAS1 expression is decreased (2.0-fold) in ductal carcinoma relative to normal epithelial cells. Taken together with the effects of ZFAS1 knockdown on mammary epithelial cell proliferation and differentiation, their results suggest ZFAS1 as a novel human tumor suppressor gene in breast cancer and that its dysregulation may be useful as a marker for breast cancer.

Neurodegenerative and psychiatric diseases

With the deepening of research, it has been gradually recognized that there are a large number of specifically expressed lncRNAs in the mammalian brain, and the occurrence and development of many neurological diseases are often accompanied by abnormal expression of some lncRNAs (Knauss and Sun, 2013). Alzheimer's disease (AD) is one of the most common neurological diseases. Studies demonstrate that AD is caused by the abnormally high expression of amyloid-β (Aβ) in the brain, and then form senile plaque (SP), while Aβ is the product after the splicing processing of amyloidprecursor protein (APP) by secretase. B-secretase1 (BACE1) is one of the secretase enzymes that splicing APP, which plays a key role not only in the production of Aβ and but also in the aggregation of Aβ.

It has been reported that the antisense lncRNA BACE1AS of BACE1 plays a very important role in the occurrence and development of AD. Under stress conditions, BACE1AS can form complex with BACE1-mRNA to increase the stability of the latter and prevent its degradation, thus promoting the aggregation of β -amyloid protein. Subsequent experimental studies also found that BACE1AS was highly expressed in AD patients and BACE1 transgenic mice. In addition, studies have found that amyloid beta protein 142 (Aβ142) can inhibit the differentiation of SH-SY5Y cell, induce the expression of APP-related factors and the formation of senile plaques. In AD model group, the expression of Aβ142 and Aβ140 protein and mRNA were up-regulated, accompanied by the down-regulation of Ki67 expression. It was also confirmed that exogenous Aβ142 not only promoted the expression of BACE1, but also promoted the expression of lncRNA BACE1AS. LncRNA BACE1AS can increase the stability of BACE1mRNA. In the subsequent reverse validation test, targeted down-regulation of lncRNA BACE1AS expression in SH-SY5Y cells attenuated the ability of BACE1 to cleave APP and slowed the formation of senile plaques in SP AD SH-SY5Y model (Faghihi and Wahlestedt, 2009; Liu et al., 2014). Recent studies have reported that the occurrence and development of AD is also related to the variable expression and abnormal location of BC200 RNA. Mus et al. (2007) found that BC200 was significantly up-regulated in AD brains, and this up-regulation in AD was specific to brain areas that are involved, BC200 levels are specifically elevated in area 9, which is involved in the AD, but not in area 17, which is generally not. Not only that, the researchers found relative BC200 levels in those areas increased in parallel with the progression of AD.

LncRNAs are also involved in the regulation of psychiatric diseases. The occurrence and development of related diseases such as major depression, autism spectrum disorder, schizophrenia, affective schizophrenia and bipolar disorder are closely related to the abnormal expression of disorder in schizophrenia-1 (DISC1) (Chubb et al., 2008). LncRNA DISC2 regulates DISC1, and lncRNA DISC2 may be a potential target for the treatment of psychiatric disorders. In addition, Tamura et al. (2007) have reported the possibility that epigenetic aberration from the normal DNA methylation status of RELN may confer susceptibility to psychiatric disorders.

Endocrine disease

Diabetes is a metabolic disease caused by a variety of factors. Its main clinical features are chronic hyperglycemia and metabolic disorders of sugar, fat and protein caused by impaired insulin secretion or defective insulin action. The World Health Organization classifies diabetes into four broad categories: I-diabetes, II-diabetes, gestational diabetes and other types of diabetes.

The pathogenesis of different types of diabetes is different, but most of them are pancreatic islet β-cell dysfunction, unable to secrete enough insulin, resulting in hyperglycemia. At present, it has been confirmed that miRNA plays an important role in the occurrence and development of diabetes. However, only a few studies have reported the role of lncRNAs in the occurrence and development of diabetes. For example, the antisense transcript IGF2AS of Insulin likegrowth factor 2 (IGF2) and the antisense transcript lncRNA PINK1 of PTENinduced putative kinase 1 (PINK1) lost on chromosome X. High concentration of glucose can stimulate up-regulation of IGF2AS expression in pancreatic islet β-cell, suggesting that IGF2AS expression may be correlated with blood glucose concentration (Mutskov and FELSENFELD, 2009). PINK1 can be activated by PTEN, which is an important inhibitor of insulin signaling pathway. In addition, array analysis performed by Scheele et al. (2007a); Scheele et al. (2007b) demonstrated a reduction in muscle PINK1 expression due to reduced activity of participants in the experiment; then their qRT-PCR data confirmed that PINK1 was 40% reduced following 5 weeks of inactivity in healthy volunteers. In contrast, natural antisense PINK1 (naPINK1) tended to be up-regulated (50%). In general, lncRNA-related studies are still in the initial stage, and there are fewer reports on the relationship between lncRNA and diabetes, which requires a lot of research work by various research teams in the later stage.

Substance use disorder

Substance use disorder is an uncontrollable, chronic and recurrent brain disease characterized by compulsive drug seeking and continuous craving, which causes serious physical and psychological harm to users. In addition, drug use disorder will lead to an increase in social crime rate and the spread of HIV, hepatitis and other related infectious diseases, which has become a global public health problem (Merz, 2018; Heikkilä et al., 2021). The formation mechanism of substance use disorder is very complex. In the process of continuous exploration and research, the epigenetic mechanism has attracted more and more attention from experts in different fields. Epigenetics refers to changes in gene expression through DNA methylation, histone modification, chromatin remodeling, and non-coding RNA regulation without changes in genetic information and DNA sequence (Goldman et al., 2005; Godino et al., 2015). Previous studies have shown that lncRNAs can regulate gene expression through epigenetic mechanisms, play an important role in the formation of synaptic plasticity, and then promote the formation of drug use disorder. In addition, researchers believe that addictive drugs can alter gene expression in brain tissues. In the hippocampus of cocaine-induced conditioned place preference (CPP), 214 transcripts were altered, and 151 genes were increased significantly. These genes belong to several functional classes including transcription, translation and protein synthesis, signal transduction, protein kinases/phosphatases, metabolic enzymes and cytoskeleton organization. Subsequently, 39 genes were found to be significantly altered in the prefrontal cortex (PFC), of which 22 genes were transcriptionally increased and 17 genes were transcriptionally decreased. Their data support the possibility that genes changes in the hippocampus and cortex might participate in the formation and of memory patterns induced by cocaine. In addition, CM156, as α -receptor antagonist, can reduce cocaine-induced CPP formation. Meanwhile, CM156 can also reverse the expression trend of some cocaine-induced transcripts in rat brain, including MATAL1, suggesting that MATAL1 may play a certain role in regulating gene expression under cocaine exposure and act through α-receptor (Kalivas and Brien, 2008; Krasnova et al., 2008; Xu et al., 2012).

Nucleus accumbens were extracted from victims of cocaine and methamphetamine use disorders. Transcriptome study found that the two addictive drugs had effects on the transcription of related genes, and the affected genes rarely overlapped. Even about half of the overlapping genes are regulated in opposite trend (Albertson et al., 2006), which suggests the possibility that cocaine and methamphetamine have different mechanisms of action in the brain. In subsequent studies, it was found that this sequence transcribed lncRNA MIAT (myocardial infarction associated transcript) (Ishii et al., 2006). At present, MIAT has been confirmed to be involved in oligodendrocyte formation and nuclear matrix formation (Mercer et al., 2010). Subsequently, Michelhaugh et al. (2011) used affymetrix microarray to study the expression changes of lncRNA in heroin abusers' brain tissues, and found that the expressions of MIAT, MEG3 and NEAT1/2 were up-regulated in nucleus accumbens. A genome-wide scan study in Caucasians also showed that MEG3 may be associated with heroin addiction (Nielsen et al., 2008). Moreover, MEG3 is expressed in the nucleus as a maternal imprint and exists in different subsets of neuronal cells, which is related to early neurogenesis. Knockout of MEG3 in the mouse brain resulted in altered expression of angiogenic genes and increased microvascular formation. The expression changes and functions of these lncRNAs indicate that relevant lncRNAs may play an important role in post-transcriptional regulation and neural adaptation after drug use disorder.

Recently, the effects of alcohol on lncRNA expression in the nervous system have also attracted attention.

MALAT1 transcription level was significantly up-regulated in the cerebellum of patients with chronic alcohol use disorder, as well as in the hippocampus (1.8-fold) and brain stem (1.5-fold), but not in the frontal lobe and motor cortex (Kryger et al., 2012). In later animal experiments, the researchers found that the expression of lncRNAs in the brain did not change significantly during acute exposure to alcohol in rats. However, MALAT1 was significantly up-regulated in rat cortex during alcohol withdrawal. MALAT1 can bind to SR splicing proteins to regulate alternative splicing of precursor mRNA (Tripathi et al., 2010). These results suggest that alcohol-mediated up-regulation of MALAT1 may play an important role in the expression of different isoforms or variants of neurotransmitter receptors and ion channel-related proteins.

Summary and prospect

LncRNAs are a very important part of eukaryotic transcripts. As the main body of ncRNAs, lncRNAs can regulate more than 70% of gene expression and play a very important role in physiological and pathological processes of the body. At present, the study of lncRNAs has become a new direction in the field of RNA, but it is still in the primary phase. The natures, structures and functions of many lncRNAs have not been clarified. Although existing researches suggest that lncRNAs play an important role in a variety of systemic diseases, there are still many basic and application problems to be solved, and their characteristics and regulation mechanisms need to be further clarified.

Human life activities depend on complex regulatory networks among DNA, RNA and proteins in the body. Therefore, future research will tend to study the human life process from a holistic and systemic perspective. Studying the expression of proteins, DNA and RNA in specific cells in different time and space and their interrelationships is delicate. Accurately studying the regulatory network of proteins, DNA and RNA in specific tissues or even in specific space-time is one of the core issues to clearly reveal human life activities. In current review, we summarize the research progress of lncRNAs in tumors, neurodegenerative and psychiatric diseases, endocrine disease and substance use disorder. Among them, some lncRNAs may well serve as biomarkers for diagnosing disease, but their accuracy and stability need to be carefully evaluated in preclinical research stage. In addition, the potential role of lncRNA in other diseases also needs to be explored and summarized. It is believed that with the in-depth study of lncRNAs by scholars, it will not only deepen human's further understanding of the regulatory network of disease occurrence and development, but also provide a basis for disease diagnosis and new therapeutic targets and approaches.

Author contributions

WL and YW wrote the first draft of the manuscript. TS and FW have substantial contributions to the conception and design of the work. LX, JD, and LW assist to organize literature and draw figures.

Funding

This study was supported by the Ningxia Hui Autonomous Region "13th Five-Year Plan" Major Science and Technology Projects (Ningxia Brain Project) (No. 2016BZ07), National Natural Science Foundation of China (NSFC) (No. 82060261) and the Key Research Project of Ningxia (No. 2018YBZD04917).

Acknowledgments

The authors would like to thank the Ningxia Key Laboratory of Cerebrocranial Disease for technical support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Albertson, D. N., Schmidt, C. J., Kapatos, G., and Bannon, M. J. (2006). Distinctive profiles of gene expression in the human nucleus accumbens associated with cocaine and heroin abuse. *Neuropsychopharmacology* 31 (10), 2304–2312. doi:10.1038/sj.npp.1301089

Askarian-amiri, M. E., Crawford, J., French, J. D., Smart, C. E., Smith, M. A., Clark, M. B., et al. (2011). SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *RNA* 17 (5), 878–891. doi:10.1261/rna.2528811

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25 (18), 1915–1927. doi:10.1101/gad.17446611

Castelo-branco, G., Amaral, P. P., EngströM, P. G., Robson, S. C., Marques, S. C., Bertone, P., et al. (2013). The non-coding snRNA 7SK controls transcriptional termination, poising, and bidirectionality in embryonic stem cells. *Genome Biol.* 14 (9), R98. doi:10.1186/gb-2013-14-9-r98

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., et al. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147 (2), 358–369. doi:10.1016/j.cell.2011.09.028

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308 (5725), 1149–1154. doi:10.1126/science.1108625

Chubb, J. E., Bradshaw, N. J., Soares, D. C., Porteous, D. J., and Millar, J. K. (2008). The DISC locus in psychiatric illness. *Mol. Psychiatry* 13 (1), 36–64. doi:10.1038/sj.mp.4002106

Di ruscio, A., Ebralidze, A. K., Benoukraf, T., Amabile, G., Goff, L. A., Terragni, J., et al. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503 (7476), 371–376. doi:10.1038/nature12598

Faghihi, M. A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.* 10 (9), 637–643. doi:10.1038/nrm2738

Godino, A., Jayanthi, S., and Cadet, J. L. (2015). Epigenetic landscape of amphetamine and methamphetamine addiction in rodents. *Epigenetics* 10 (7), 574–580. doi:10.1080/15592294.2015.1055441

Goldman, D., Oroszi, G., and Ducci, F. (2005). The genetics of addictions: Uncovering the genes. *Nat. Rev. Genet.* 6 (7), 521–532. doi:10.1038/nrg1635

Harries, L. W. (2012). Long non-coding RNAs and human disease. *Biochem. Soc. Trans.* 40 (4), 902–906. doi:10.1042/BST20120020

Heikkilä, H., Maalouf, W., and Campello, G. (2021). The united nations office on drugs and crime's efforts to strengthen a culture of prevention in low- and middle-income countries. *Prev. Sci.* 22 (1), 18–28. doi:10.1007/s11121-020-01088-5

Ishii, N., Ozaki, K., Sato, H., Mizuno, H., Takahashi, A., Miyamoto, Y, et al. (2006). Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J. Hum. Genet.* 51 (12), 1087–1099. doi:10.1007/s10038-006-0070-9

Kalivas, P. W., and Brien, C. (2008). Drug addiction as a pathology of staged neuroplasticity. *Neuropsychopharmacology* 33 (1), 166–180. doi:10.1038/sj.npp.1301564

Kitagawa, M., Kotake, Y., and Ohhata, T. (2012). Long non-coding RNAs involved in cancer development and cell fate determination. *Curr. Drug Targets* 13 (13), 1616–1621. doi:10.2174/138945012803530026

Knauss, J. L., and Sun, T. (2013). Regulatory mechanisms of long noncoding RNAs in vertebrate central nervous system development and function. *Neuroscience* 235, 200–214. doi:10.1016/j.neuroscience.2013.01.022

Krasnova, I. N., Li, S. M., Wood, W. H., McCoy, M. T., Prabhu, V. V., Becker, K. G., et al. (2008). Transcriptional responses to reinforcing effects of cocaine in the rat hippocampus and cortex. *Genes Brain Behav.* 7 (2), 193–202. doi:10.1111/j.1601-183X.2007.00338.x

Kryger, R., Fan, L., Wilce, P. A., and Jaquet, V. (2012). MALAT-1, a non protein-coding RNA is upregulated in the cerebellum, hippocampus and brain stem of human alcoholics. *Alcohol* 46 (7), 629–634. doi:10.1016/j.alcohol.2012.04.002

Liu, T., Huang, Y., Chen, J., Chi, H., Yu, Z., Wang, J., et al. (2014). Attenuated ability of BACE1 to cleave the amyloid precursor protein via silencing long noncoding RNA BACE1-AS expression. *Mol. Med. Rep.* 10 (3), 1275–1281. doi:10.3892/mmr.2014.2351

Luo, M., Li, Z., Wang, W., Zeng, Y., Liu, Z., and Qiu, J. (2013). Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. *Cancer Lett.* 333 (2), 213–221. doi:10.1016/j.canlet.2013.01.033

Ma, L., Bajic, V. B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biol.* 10 (6), 925–933. doi:10.4161/rna.24604

Mazar, J., Zhao, W., Khalil, A. M., Lee, B., Shelley, J., Govindarajan, S. S., et al. (2014). The functional characterization of long noncoding RNA SPRY4-IT1 in human melanoma cells. *Oncotarget* 5 (19), 8959–8969. doi:10.18632/oncotarget.1863

Mercer, T. R., Qureshi, I. A., Gokhan, S., Dinger, M. E., Mattick, J. S., Mehler, M. F., et al. (2010). Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. *BMC Neurosci.* 11, 14. doi:10.1186/1471-2202-11-14

Merz, F. (2018). United nations office on drugs and crime: World drug report 2017. 2017. *SIRIUS - Z. fur Strateg. Anal.* 2, 85–86. doi:10.1515/sirius-2018-0016

Michelhaugh, S. K., Lipovich, L., Blythe, J., Jia, H., Kapatos, G., and Bannon, M. J. (2011). Mining affymetrix microarray data for long non-coding RNAs: Altered expression in the nucleus accumbens of heroin abusers. *J. Neurochem.* 116 (3), 459–466. doi:10.1111/j.1471-4159.2010.07126.x

Mourtada-maarabouni, M., Pickard, M. R., Hedge, V. L., FarzanehF.and Williams, G. T. (2009). GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* 28 (2), 195–208. doi:10.1038/onc.2008.373

Mus, E., Hof, P. R., and Tiedge, H. (2007). Dendritic BC200 RNA in aging and in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* 104 (25), 10679–10684. doi:10.1073/pnas.0701532104

Mutskov, V., and Felsenfeld, G. (2009). The human insulin gene is part of a large open chromatin domain specific for human islets. *Proc. Natl. Acad. Sci. U. S. A.* 106 (41), 17419–17424. doi:10.1073/pnas.0909288106

Nielsen, D. A., Ji, F., Yuferov, V., Ho, A., Chen, A., Levran, O., et al. (2008). Genotype patterns that contribute to increased risk for or protection from developing heroin addiction. *Mol. Psychiatry* 13 (4), 417–428. doi:10.1038/sj.mp.4002147

Prensner, J. R., Iyer, M. K., Balbin, O. A., Dhanasekaran, S. M., Cao, Q., Brenner, J. C., et al. (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* 29 (8), 742–749. doi:10.1038/nbt.1914

Renoux, A. J., and Todd, P. K. (2012). Neurodegeneration the RNA way. *Prog. Neurobiol.* 97 (2), 173–189. doi:10.1016/j.pneurobio.2011.10.006

Scheele, C., Nielsen, A. R., Walden, T. B., Sewell, D. A., Fischer, C. P., Brogan, R. J., et al. (2007). Altered regulation of the PINK1 locus: A link between type 2 diabetes and neurodegeneration?[J]. *FASEB J.* 21 (13), 3653–3665. doi:10.1096/fj.07-8520com

Scheele, C., Petrovic, N., Faghihi, M. A., Lassmann, T., Fredriksson, K., Rooyackers, O., et al. (2007). The human PINK1 locus is regulated *in vivo* by a non-coding natural antisense RNA during modulation of mitochondrial function. *BMC Genomics* 8, 74. doi:10.1186/1471-2164-8-74

Silva, J. M., Boczek, N. J., Berres, M. W., Ma, X., and Smith, D. I. (2011). LSINCT5 is over expressed in breast and ovarian cancer and affects cellular proliferation. *RNA Biol.* 8 (3), 496–505. doi:10.4161/rna.8.3.14800

Tamura, Y., Kunugi, H., Ohashi, J., and HoHjoH, H. (2007). Epigenetic aberration of the human REELIN gene in psychiatric disorders. *Mol. Psychiatry* 12 (6519), 593–600. doi:10.1038/sj.mp.4002014

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39 (6), 925–938. doi:10.1016/j.molcel.2010.08.011

Vennin, C., and Adriaenssens, E. (2018). Long non-coding RNA and messenger RNA-the meeting of two worlds. *J. Thorac. Dis.* 10 (2), 544–546. doi:10.21037/jtd.2018.01.103

Wang, F., Li, X., Xie, X., Zhao, L., and Chen, W. (2008). UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. *FEBS Lett.* 582 (13), 1919–1927. doi:10.1016/j.febslet.2008.05.012

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43 (6), 904–914. doi:10.1016/j.molcel.2011.08.018

Wu, W., Bhagat, T. D., Yang, X., Song, J. H., Cheng, Y., Agarwal, R., et al. (2013). Hypomethylation of noncoding DNA regions and overexpression of the long noncoding RNA, AFAP1-AS1, in Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterology* 144 (5), 956–966. e4. doi:10.1053/j.gastro.2013.01.019

Xu, Y. T., Robson, M. J., Szeszel-Fedorowicz, W., Patel, D., Rooney, R., McCurdy, C. R., et al. (2012). CM156, a sigma receptor ligand, reverses cocaine-induced place conditioning and transcriptional responses in the brain. *Pharmacol. Biochem. Behav.* 101 (1), 174–180. doi:10.1016/j.pbb.2011.12.016

Zhu, J., Fu, H., Wu, Y., and Zheng, X. (2013). Function of lncRNAs and approaches to lncRNA-protein interactions. *Sci. China. Life Sci.* 56 (10), 876–885. doi:10.1007/s11427-013-4553-6

Zhu, M., Chen, Q., Liu, X., Sun, Q., Zhao, X., Deng, R., et al. (2014). lncRNA H19/miR-675 axis represses prostate cancer metastasis by targeting TGFBI. *FEBS J.* 281 (16), 3766–3775. doi:10.1111/febs.12902

Zhu, Y. P., Bian, X. J., Ye, D. W., Yao, X. D., Zhang, S. L., Dai, B., et al. (2014). Long noncoding RNA expression signatures of bladder cancer revealed by microarray. *Oncol. Lett.* 7 (4), 1197–1202. doi:10.3892/ol.2014.1843

frontiers | Frontiers in Molecular Biosciences

TYPE Mini Review
PUBLISHED 14 November 2022
DOI 10.3389/fmolb.2022.1062448

OPEN ACCESS

EDITED BY
André P. Gerber,
University of Surrey, United Kingdom

REVIEWED BY
Olivier Bensaude,
École Normale Supérieure, France

*CORRESPONDENCE
Roberto Giambruno,
roberto.giambruno@cnr.it
Francesco Nicassio,
francesco.nicassio@iit.it

SPECIALTY SECTION
This article was submitted to RNA Networks and Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 05 October 2022
ACCEPTED 25 October 2022
PUBLISHED 14 November 2022

CITATION
Giambruno R and Nicassio F (2022), Proximity-dependent biotinylation technologies for mapping RNA-protein interactions in live cells.
Front. Mol. Biosci. 9:1062448.
doi: 10.3389/fmolb.2022.1062448

Proximity-dependent biotinylation technologies for mapping RNA-protein interactions in live cells

Roberto Giambruno[1,2]* and Francesco Nicassio[1]*

[1]Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia, Milano, Italy, [2]Institute of Biomedical Technologies, National Research Council, Segrate, Italy

Proximity ligation technologies are extremely powerful tools for unveiling RNA-protein interactions occurring at different stages in living cells. These approaches mainly rely on the inducible activity of enzymes (biotin ligases or peroxidases) that promiscuously biotinylate macromolecules within a 20 nm range. These enzymes can be either fused to an RNA binding protein or tethered to any RNA of interest and expressed in living cells to biotinylate the amino acids and nucleic acids of binding partners in proximity. The biotinylated molecules can then be easily affinity purified under denaturing conditions and analyzed by mass spectrometry or next generation sequencing. These approaches have been widely used in recent years, providing a potent instrument to map the molecular interactions of specific RNA-binding proteins as well as RNA transcripts occurring in mammalian cells. In addition, they permit the identification of transient interactions as well as interactions among low expressed molecules that are often missed by standard affinity purification strategies. This review will provide a brief overview of the currently available proximity ligation methods, highlighting both their strengths and shortcomings. Furthermore, it will bring further insights to the way these technologies could be further used to characterize post-transcriptional modifications that are known to regulate RNA-protein interactions.

KEYWORDS

protein-RNA interactions, proximity biotinylation, APEX2, affinity purification, affinity purification coupled to mass spectrometry, APEX-seq, dCas13

Introduction

RNA and proteins are constantly interacting in living cells at multiple stages and their interaction is fundamental for their biological functions (Yeo, 2014; Hentze et al., 2018). RNA molecules can transiently interact with RNA binding proteins (RBPs) or lying in complexes with proteins, forming ribonucleoprotein complexes. For RNAs, the interaction with proteins is required for their biogenesis and functions, regulating several aspects of cell biology including transcription, splicing and translation. For proteins, interaction with RNAs is necessary to maintain the folding or the integrity of a multi-subunit complex, to direct the catalytic moiety to specific targets or

compartments, and to modulate the protein biological activity. Physical interaction is mediated by RNA-binding domains and intrinsically disordered regions of RBPs, which can recognize specific RNA structures (hairpins, stems, or loops), RNA sequence motifs or simply have a high affinity to bind RNA molecules. Importantly, RNA-protein interactions are not usually stable and are frequently regulated by post-transcriptional and post-translational events that modulate the binding affinity (Lewis et al., 2017; Spadotto et al., 2020). Since RNA-protein interactions play a key role in molecular, cellular and developmental biology it is not surprising that alterations can affect cellular homeostasis and have been linked to many human diseases, including neurodegenerative disorders and cancer (Gebauer et al., 2021; Kelaini et al., 2021).

Several methods have been developed to characterize RNA-protein interactions at global level, distinguished into RNA-centric or protein-centric approaches. Usually, the RNA or protein of interest is isolated by affinity purification from cells and binding partners are identified by high-throughput mass spectrometry and RNA sequencing approaches, respectively (McHugh et al., 2014; Giambruno et al., 2018; Gerber, 2021). Despite being powerful, these methodologies often miss interactions that are transient or involve low abundant molecules. The use of crosslinking agents can sensibly increase the number of detected interactions, although introducing biases towards pyrimidine-reach RNA regions as well as increasing the risk of false positives and non-direct interactions (Ramanathan et al., 2019). An alternative strategy is provided by Proximity Dependent Biotinylation (PDB) approaches, in which bacterial enzymes are used to biotinylate functional groups in proximity of the protein of interest (<20 nm) and used to detect protein-protein and protein-nucleic acid interactions (Ramanathan et al., 2018; Fazal et al., 2019; Padrón et al., 2019). The biotinylated molecules are isolated through biotin-streptavidin interaction, which is efficient and specific even under denaturing conditions, a condition which allows the enrichment of true interactors and removal of non-specific binders. The main advantages are: 1) the biotinylation occurs in living cells, preserving the biological cellular environment and avoiding any artificial interactions that might occur during cell lysis; 2) it tracks stable as well transient interactions, even at picomolar scale, without the need of any crosslinking step; 3) it is compatible with OMICS technologies allowing the global identification of the interactions occurring between the RNA or protein of interest in cells.

The use of PDB approaches to assess protein-protein interactions, protein cellular localization and compartmentalization, including the regulation mediated by post-translational modifications have been recently reviewed (Dionne and Gingras, 2022). The focus of this mini review is exclusively related to the application of these approaches to unveil RNA-protein interactions and their dynamics in living cells, avoiding the use of crosslinking agents to stabilize molecular interactions.

PDB enzymes

Biotin ligases are enzymes able to convert, in the presence of ATP, biotin into an active biotin-5-AMP intermediate that is covalently linked to primary amines (epsilon group of lysine residues and protein N-termini) of proximal proteins. The first used biotin ligase is the *E. Coli* Bifunctional ligase/repressor enzyme carrying the R118G mutation (BirA*) that promiscuously biotinylates any proximal protein (Roux et al., 2012). Several enzymes have been purified from different bacteria and engineered to enhance their catalytic activity towards protein substrates. Among them, the BASU protein derived from *Bacillus subtilis* that biotinylates protein within 30 min (Ramanathan et al., 2018); and the TurboID protein which is a mutated form of BirA* able to biotinylate proteins in less than 10 min (Branon et al., 2018).

Alternatively, the activity of the mutated form of the heme ascorbate peroxidase enzyme APEX2 (Apurinic/Apyrimidinic Endodeoxyribonuclease 2), has been exploited for PDB approaches. APEX2 is the more active variant of the initial APEX enzyme derived from soybean (Lam et al., 2015). Upon hydrogen peroxide (H_2O_2) treatment, APEX2 converts phenol substrates into short-lived phenoxyl radicals with a half-life of less than one millisecond that covalently attach electron-rich amino acids and nucleotides with a special preference for tyrosine and guanine, respectively (Qin et al., 2021a).

The biotinylation process mediated by PDB enzymes is strongly dependent on the abundance, length, composition and structure of the targeted molecules. Moreover, substrate regions have to be exposed and freely available to the labeling of the PDB enzyme. Hence, the labeling intensity is not directly correlated with the strength of the interaction (Mair and Bergmann, 2022). A detailed summary of the various PDB enzymes currently used in molecular biology and their mechanism of actions can be found elsewhere (Samavarchi-Tehrani et al., 2020; Qin et al., 2021b).

RNA centric methods

RNA centric PDB methods assess which proteins are interacting with a selected RNA transcript, in living cells. The PDB enzyme is recruited to the RNA of interest and, once activated, it starts biotinylating any protein present in proximity over time. Even proteins that transiently interact with the targeted RNA transcript can be covalently labeled with one or multiple biotin molecules. The resulting biotinylated proteins are then affinity purified under denaturing conditions and identified by MS-based proteomics

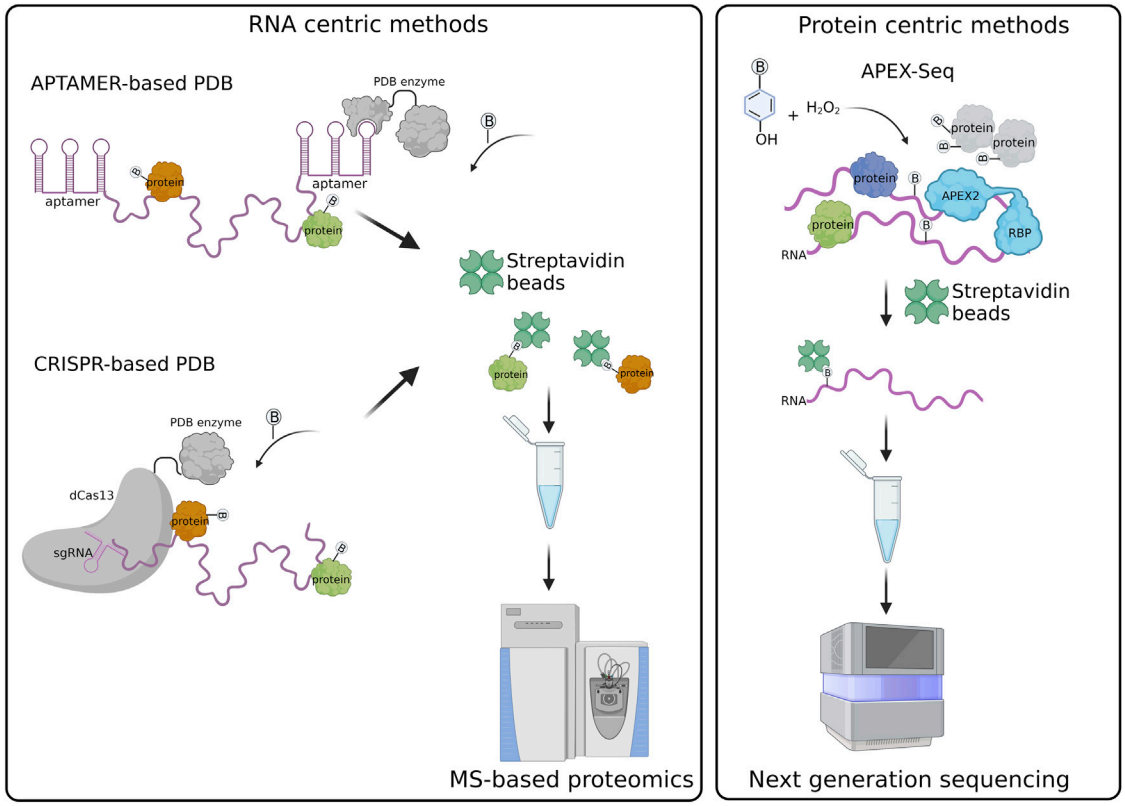


FIGURE 1
Schematic representation of RNA-centric and protein centric PDB created with BioRender.com.

through either label free quantification or stable isotope labeling methods, as previously summarized (Lindemann et al., 2017; Giambruno et al., 2018). Differently from standard RNA centric methods, RNA centric PDB strategies are highly sensitive especially for the detection of transient interactions and do not require any step to preserve RNA-protein interactions prior cell harvesting. Thus, these methods maximize the sensitivity without increasing false positive interactions (Ramanathan et al., 2019). However, they cannot provide information about: 1) whether the protein is a direct RNA-binder and to which portion of the RNA is bound; 2) if the identified proteins simultaneously bind the RNA or at different stages; 3) whether the identified interactors belong to multiprotein complexes; 4) if the detected RNA-protein interaction is mediated by RNA post-transcriptional modifications.

Currently, the main approaches are based on 1) Aptamer; 2) Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR). They are represented in Figure 1.

Aptamer-based PDB exploits the tagging of the RNA of interest with the MS2 or BoxB aptamers that are specifically recognized and bound by the MS2 coat protein (MCP) and λN peptide fused in frame with the labeling enzyme, respectively (Weissinger et al., 2021). An example is represented by the RNA-protein interaction detection—mass spectrometry (RaPID-MS) strategy (Ramanathan et al., 2018), where the RNA of interest is expressed in living cells tagged with three BoxB aptamers located both at the 5′ and 3′ ends. The aptamers are bound by the co-expressed enzyme BASU carrying at its N-terminus the λN-peptide that allows the tethering of the PDB enzyme to the BoxB-tagged RNA. BASU activity is then promoted by the administration of exogenous biotin in the cell culture medium. Biotinylated proteins are affinity purified under denaturing conditions and analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). In this approach the RNA transcripts are overexpressed, therefore it is better suitable for the identification of proteins bound to specific RNA motifs or to compare the interactome of a wild-type versus mutated RNA sequences (Ramanathan et al., 2018). It can be also exploited to identify the host-protein interactions of exogenous transcripts, such as viral RNA transcripts, which are usually expressed at high levels in infected cells, as reported for the Zika and SARS-CoV-2 viruses (Ramanathan et al., 2018; Giambruno et al., 2022).

An alternative strategy is RNA-BioID (Mukherjee et al., 2019) that has been used to analyze the protein interactome of the endogenous β-Actin RNA through the insertion of 24 repeats of the MS2 aptamer at the 3′UTR of the gene. The MS2-tagged RNA is bound by a stably expressed MCP-BirA*, which biotinylates proteins associated with the RNA. The labeled proteins are then purified and identified by LC-MS/MS. However, the BirA* proximity labeling time was conducted for 24 h, a relatively long time during which multiple RBPs can enter in proximity with the targeted RNA, therefore being biotinylated. Thus, strongly reducing the signal-to-noise ratio of the true interactors identified by this strategy.

The time for the biotinylation has been sensibly minimized through the development of the MS2-based APEX method (Han et al., 2020). The authors co-expressed in living cells the MCP-APEX2 enzyme and the human telomerase RNA (hTR) carrying a tag of 4x MS2 repeats fused to its 5′ RNA. The main advantage is represented by the fact that APEX2 has an extremely fast kinetic promoting proximal protein biotinylation in less than a minute. Thus, APEX2 allows the detection of transient and dynamic RNA-protein interactions to the same extent of those interactions that are more stable and, hence, can be detected more easily by standard biochemical approaches. In addition, the shorter number of aptamer repeats as compared to the one used in RNA-BioID better preserves the biological properties of the tagged RNA (i.e., MS2-tagged hTR) (Laprade et al., 2020). However, this strategy works well only for overexpressed RNAs (Han et al., 2020).

CRISPR PDB approach exploits the activity of the dCas13 enzyme, which specifically binds RNA sequences under the guidance of single guide RNAs (sgRNAs), without cleaving the RNA or targeting DNA sequences (Abudayyeh et al., 2017). The PDB enzyme is fused in frame to the dCas13 and therefore recruited to the endogenous RNA target.

Four similar CRISPR PDB tools have been developed: 1) CARPID (CRISPR assisted RNA-protein interaction detection method) (Yi et al., 2020), 2) Cas13-based APEX method (Han et al., 2020), 3) CBRPP (CRISPR-based RNA proximity proteomics) (Li et al., 2021), and 4) RPL (RNA proximity labeling) (Lin et al., 2021). They have in common the use of a fusion protein composed of catalytically inactive Cas13 variants (dCas13 or dRfxCas13d) and a PDB enzyme (APEX2, BASU and BioID2) (Han et al., 2020; Yi et al., 2020; Lin et al., 2021). Differently from aptamer-based strategies, CRISPR PDB approaches directly target endogenous RNA transcripts in living cells, without the need of a pre-labelling step. The fusion protein dCas13-PDB enzyme is tethered to the RNA of interest by a single or multiple sgRNAs. The number of sgRNAs is chosen according to the length of the targeted RNA. In the case of lncRNA, such as XIST and MALAT1, a set of different sgRNAs has been used to probe the different regions of the RNA (Yi et al., 2020). As the secondary structures of the targeted RNA can influence sgRNA pairing, multiple sgRNAs should be tested to select those that are effective (Han et al., 2020). The recruitment of the dCas13-APEX2 to the RNA target can be improved by different strategies, such as: 1) the insertion of a double strand RNA binding domain (dsRBD) at the C-terminus of the fusion protein, which stabilize the protein-RNA complex (Han et al., 2020); 2) the adoption of inducible expression systems that regulate the expression of the dCas13-APEX2 in cells and enhance the signal-to-noise labeling ratio (Han et al., 2020; Li et al., 2021); 3) the addition of a nuclear export sequence (NES) or nuclear localization signal (NLS) to concentrate the fusion protein in the same cellular compartment of the targeted RNA (Lin et al., 2021).

RNA centric PDB methods require the use of appropriate experimental controls. It is advisable to include in the analysis an unrelated RNA, with length and GC-content similar to the RNA of interest, whose results can be used to measure the experimental background. Moreover, an RNA with known interacting partners can be used as a positive control, assessing the efficacy of the strategy and the sensitivity of the assay and the related instrumentation (Table 1).

Protein centric methods

Protein centric PDB methods allow the characterization of the RNA transcripts that are bound to or in proximity of a protein of interest through next generation sequencing (NGS) approaches. Differently from standard protein centric methods, this strategy does not require the use of an antibody for the protein target and, hence, can be applied to any protein or protein isoform (Qin et al., 2021b). The protein bait is expressed in living cells as a fusion protein having either at its N- or C-terminus a spacer containing a tag used for detection (i.e. FLAG/HA epitope) and followed by a PDB enzyme. So far, researchers have preferred the use of APEX enzymes, which guarantee a faster labeling time compared to biotin ligase enzymes (Bosch et al., 2021). Once expressed and activated by the administration of phenol-biotin into the cell culture medium followed by H_2O_2 treatment, the APEX enzyme starts biotinylating any macromolecule present in its proximity. The cells are then harvested and biotinylated proteins, together with their associated RNA transcripts, pulled down through streptavidin beads. The RNA is extracted from the beads, purified and analyzed by standard NGS. This strategy has been extremely helpful for the characterization of the RNA transcripts associated with subcellular compartments, such as stress granules (SGs) and the nuclear lamina, using protein markers such G3BP1 and LAMIN B1, as baits (Somasekharan et al., 2020; Tran et al., 2021).

In the APEX-Seq approach, APEX2 is used to directly biotinylate nucleic acids (Figure 1). Thus, biotinylated RNA transcripts are affinity purified through streptavidin beads and analyzed by NGS (Fazal et al., 2019; Padrón et al., 2019; Zhou et al., 2019). APEX-Seq can be used to efficiently map both the proteins and RNAs interacting with a protein of interest, used as

TABLE 1 Summary of strengths, limitations and available tools of the current PDB strategies.

	RaPID-MS	RNA BioID	MS2-based APEX method	CRISPR proximity biotinylation tools	Purification of protein/RNA complexes	APEX-Seq
BAIT	• BoxB-tagged RNA	• MS2-tagged RNA	• MS2-tagged RNA	• Endogenous RNA	• APEX- tagged protein	• APEX- tagged protein
PREYS	• Endogenous proteins	• Endogenous proteins	• Endogenous proteins	• Endogenous proteins	• Endogenous RNAs	• Endogenous RNAs
PURIFICATION APPROACH	• Streptavidin pull-down of biotinylated proteins	• Streptavidin pull-down of biotinylated proteins	• Streptavidin pull-down of biotinylated proteins	• Streptavidin pull-down of biotinylated proteins	• Streptavidin pull-down of biotinylated proteins	• Streptavidin pull-down of biotinylated RNAs
STRENGHTS AND LIMITATIONS	✓ No crosslinking agents	✓ No crosslinking agents	✓ No crosslinking agents	✓ No crosslinking agents	✓ No crosslinking agents	✓ No crosslinking agents
	✓ Low amount of material	* High amount of material	✓ Low amount of material	* High amount of material	✓ Low amount of material	✓ Low amount of material
	✓ Fast	* Slow	✓ Fast	* Slow	✓ Fast	✓ Fast
	✓ Not expensive	* Expensive	✓ Not expensive	* Expensive	✓ Not expensive	* Expensive
	✓ Easy to use technology	* Laborious technology	✓ Easy to use technology	* Laborious technology	✓ Easy to use technology	* Laborious technology
	* Not applicable to endogenous RNAs	✓ Applicable to endogenous RNAs	* Not applicable to endogenous RNAs	✓ Applicable to endogenous RNAs	* Not applicable to endogenous proteins	* Not applicable to endogenous proteins
	* Presence of not direct interactors	* Presence of not direct interactors	* Presence of not direct interactors	* Presence of not direct interactors	* The identified RNAs cannot be address to a single RBP but rather to the whole protein complex	* Presence of not direct interactors
	* The tag can alter the localization and interaction profile of the bait	* The tag can alter the localization and interaction profile of the bait	* The tag can alter the localization and interaction profile of the bait	* The off-target binding of the sgRNAs and the presence of the dCas13 can alter the interaction profile of the bait	* The tag can alter the localization and interaction profile of the bait	* The tag can alter the localization and interaction profile of the bait
AVAILABLE TOOLS	• Plasmid to clone BoxB-tagged RNAs (Addgene #107253) • plasmid for mammalian expression of lN-HA-BASU (Addgene #107250) • Plasmid expressing the positive control BoxB-EDEN15 (Addgene #107252		• Plasmid encoding for MCP-APEX2 (Addgene #154936)		• Plasmid encoding for CARPID BASU-dCasRx (Addgene #153209), CARPID dCasRx-BASU (Addgene #153303), dCas13d-dsRBD-APEX2 (Addgene #154939)	• Addgene plasmids encoding for APEX2-OMM (#79056), ERM-APEX2 (#79055), mito-APEX2 (#72480), APEX2-SENP (#129276), APEX2-eIF41 (#129645), APEX2-eIFE1 (#129644), C1-APEX2 (#129641), APEX2-FBL (#187577), APEX2-SRSF7 (#187582), APEX2-SRSF1 (#187575), APEX2-PML (#187583), APEX2-SP100 (#187584), APEX2-NPAT (#187585), APEX2-LMNA (#187576)
PROPER CONTROLS	• Scramble RNA of the same length and similar GC-content	• Scramble RNA of the same length and similar GC-content	Scramble RNA of the same length and similar GC-content	• Non targeting sgRNAs	• APEX2-GFP	• APEX2-GFP
	• Positive control RNA with	• Positive control RNA with	• Positive control RNA with	• Positive control RNA with known interactors	• APEX2-tagged protein located on a	• APEX2-tagged protein located on a different cellular compartment

(Continued on following page)

TABLE 1 (*Continued*) Summary of strengths, limitations and available tools of the current PDB strategies.

	RaPID-MS	RNA BioID	MS2-based APEX method	CRISPR proximity biotinylation tools	Purification of protein/RNA complexes	APEX-Seq
	known interactors	known interactors	known interactors		different cellular compartment	
References (DOI)	- 10.1038/nmeth.4601	- 10.1073/pnas.1820737116	- 10.1073/pnas.2006617117	- 10.1073/pnas.2006617117 - 10.1038/s41592-020-0866-0	- 10.1093/nar/gkaa376	- 10.1016/j.cell.2019.05.027
				- 10.1080/15476286.2021.1873620	- 10.1083/jcb.202002129	- 10.1016/j.molcel.2019.07.030
				- 10.1080/15476286.2021.1917215		

bait. This approach has proven extremely useful for the definition of protein-RNA and protein-protein interactions occurring within cellular structures or membrane-less organelles without the need of isolation approaches (Fazal et al., 2019; Padrón et al., 2019). APEX-Seq was applied for the characterization of the RNA-protein interaction patterns of different subcellular organelles and compartments including the inner (Mito-APEX2) (Fazal et al., 2019; Zhou et al., 2019) and outer mitochondrial membrane (APEX2-OMM) (Fazal et al., 2019), the endoplasmic reticulum membrane [ERM-APEX2 or C1(1-29)-APEX2] (Fazal et al., 2019; Padrón et al., 2019) and the nuclear pore (APEX2-SENP) (Fazal et al., 2019). Furthermore, APEX-Seq was employed to studying RNA-protein interactions in macromolecular complexes, such as SGs (APEX2-eIF4A1) or the translation initiation complex (APEX2-eIF4A1 and APEX2-eIF4E1) (Padrón et al., 2019). This approach has been exploited also to analyze the dynamics of RNA-protein interactions upon perturbations, as shown by the analysis of the pattern of RNAs recruited by eIF4A1 to SGs in response to different types of stress (Padrón et al., 2019) and the analysis of RNA interaction patterns of the outer mitochondrial membrane in response to drug perturbations (Fazal et al., 2019).

Recently, APEX-Seq has been used to investigate even membraneless domains, using as bait their specific markers. Relevant examples are APEX2-FBL for the nucleolus; APEX2-SRSF7, APEX2-SRSF1, and APEX2-RNPS1 for nuclear speckles; APEX2-SMN2 for Cajal bodies; APEX2-SAM68 for the SAM68 bodies; APEX2-PML and APEX2-SP100 for the PML bodies; APEX2-NPAT for the histone locus bodies; and APEX2-LMNA for the nuclear lamina (Barutcu et al., 2022).

The labeling of RNA transcripts can be improved by using biotin-aniline as peroxidase substrate. Conversely to proteins, APEX2 biotinylates RNA transcripts approximately 3-fold times more in the presence of biotin-aniline compared to biotin-phenol (Zhou et al., 2019). Nevertheless, APEX2 can also biotinylate DNA hence, it is necessary to remove any traces of DNA before analyzing biotinylated RNA transcripts (Matěju and Chao, 2022).

To correctly define the RNA interactome of a given bait, it is preferable to conduct the same purification strategy on multiple baits that preferably localize to different cellular compartments. This can help to properly assess the experimental background and identify those frequently recurring RNA-protein interactions.

Discussion

The characterization of RNA-protein interactions in living cells has to take into consideration four aspects: 1) the bait has to be soluble to be affinity purified; 2) the preys must interact with the bait for a sufficient time and amount to be detected; 3) RNA-protein interactions have to be preserved during the whole affinity purification procedure; 4) the signal to noise ratio must be high to permit the correct identification of true interactors and minimize number of false positives. To this extent, PDB approaches have sensibly favored the biochemical workflow applied to assess the interactome of an RNA or a protein of interest. The biotinylation process occurs in living cells where the cellular environment is preserved. This eliminates the formation of false positive RNA-protein interactions that can occur during cell lysis, when the cellular membranes are disrupted. At the same time, any bait can be purified thanks to the possibility of adopting even denaturing cell lysis buffers.

The strong affinity between streptavidin and biotin (Kd 10^{-14} M) (Michael Green, 1990) allows purification of biotinylated molecules under denaturing conditions, including the presence of high salt and detergents in the purification buffer. Thus, only biotinylated molecules are purified for the subsequent identification. However, PDB methods cannot discriminate between direct and proximally located interactors of a given bait. In addition, many RBPs have a promiscuous association with RNA transcripts, especially when they are abundantly expressed (Nielsen et al., 2016; Protter et al., 2018; Corley et al., 2020). Therefore, if a protein is directly and specifically interacting with an RNA, it should be verified by orthogonal techniques.

PDB technologies rely on the activity of a bacterial enzyme fused or tethered to the molecule of interest and, hence, poses important limitations. The enzyme may alter the biological properties of the targeted molecules. For instance, the

paraspeckle proteins NONO, PSPC1, and EWSR1 tagged with APEX2 at their N-terminus showed a non-physiological localization compared to the respective endogenous proteins (Barutcu et al., 2022). Thus, it is necessary to assess that the tagged molecule maintains its proper localization and, if possible, biological function. In addition, the bacterial enzyme can interfere with the binding of protein or cellular RNAs to one or multiple portions of the tagged bait. The dCas13 protein, for instance, is a 130 KDa protein that once tethered to the targeted RNA can sterically outcompete RBPs that transiently bind to the RNA regions where the dCas13 is present (Han et al., 2020). The same issue may occur when fusing a PDB enzyme to a protein of interest, thus impacting its interactions (Qin et al., 2021a).

One of the current limits for the characterization of RNA-protein interactions is the difficulties in performing an unbiased identification and characterization of the post-transcriptional modifications present in the RNA transcripts that are interacting with cellular proteins. RNA post-transcriptional events have been recognized with important regulatory functions and are known to regulate RNA folding into secondary structures and the propensity of the RNA to interact with partners (Li and Mason, 2014). Protein centric PDB technologies may potentially be exploited to detect post-transcriptional modifications that decorate RNA transcripts while interacting with proteins. However, the current NGS protocols require the conversion of RNA into cDNA prior to sequencing. Thus, eliminating any information related to the presence of post-transcriptionally modified nucleotides. Nanopore direct RNA-Sequencing (DRS) has emerged as a new technology that offers for the first time the possibility to sequence full-length native RNA molecules, allowing the study of RNA modifications in an unbiased way and at single nucleotide resolution. In the next future, the coupling of nanopore DRS protocols downstream of protein centric PDB technologies will allow for a protein of interest the identification of the associated RNAs and their post-transcriptional modifications.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

Associazione Italiana per la Ricerca sul Cancro (AIRC) (IG18774, IG22851 to FN); Fondazione Cariplo (2015-0590 to FN). Funding for open access charge: Core funding from IIT.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA targeting with CRISPR–Cas13. *Nature* 550, 280–284. doi:10.1038/nature24049

Barutcu, A. R., Wu, M., Braunschweig, U., Dyakov, B. J. A., Luo, Z., Turner, K. M., et al. (2022). Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol. Cell* 82, 1035e9–1052.e9. doi:10.1016/j.molcel.2021.12.010

Bosch, J. A., Chen, C., and Perrimon, N. (2021). Proximity-dependent labeling methods for proteomic profiling in living cells: An update. *Wiley Interdiscip. Rev. Dev. Biol.* 10, e392. doi:10.1002/wdev.392

Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., et al. (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* 36, 880–887. doi:10.1038/nbt.4201

Corley, M., Burns, M. C., and Yeo, G. W. (2020). How RNA-binding proteins interact with RNA: Molecules and mechanisms. *Mol. Cell* 78, 9–29. doi:10.1016/j.molcel.2020.03.011

Dionne, U., and Gingras, A.-C. (2022). Proximity-dependent biotinylation approaches to explore the dynamic compartmentalized proteome. *Front. Mol. Biosci.* 9, 852911. doi:10.3389/fmolb.2022.852911

Fazal, F. M., Han, S., Parker, K. R., Kaewsapsak, P., Xu, J., Boettiger, A. N., et al. (2019). Atlas of subcellular RNA localization revealed by APEX-seq. *Cell* 178, 473e26–490. doi:10.1016/j.cell.2019.05.027

Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* 22, 185–198. doi:10.1038/s41576-020-00302-y

Gerber, A. P. (2021). RNA-centric approaches to profile the RNA–protein interaction landscape on selected RNAs. *Noncoding. RNA* 7, 11. doi:10.3390/ncrna7010011

Giambruno, R., Zacco, E., Ugolini, C., Vandelli, A., Mulroney, L., D'Onghia, M., et al. (2022). *Discovering host protein interactions specific for SARS-CoV-2 RNA genome*. Cold Spring Harbor Laboratory: bioRxiv. doi:10.1101/2022.07.18.499583

Giambruno, R., Mihailovich, M., and Bonaldi, T. (2018). Mass spectrometry-based proteomics to unveil the non-coding RNA world. *Front. Mol. Biosci.* 5, 90. doi:10.3389/fmolb.2018.00090

Han, S., Zhao, B. S., Myers, S. A., Carr, S. A., He, C., and Ting, A. Y. (2020). RNA–protein interaction mapping via MS2- or Cas13-based APEX targeting. *Proc. Natl. Acad. Sci. U. S. A.* 117, 22068–22079. doi:10.1073/pnas.2006617117

Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341. doi:10.1038/nrm.2017.130

Kelaini, S., Chan, C., Cornelius, V. A., and Margariti, A. (2021). RNA-binding proteins hold key roles in function, dysfunction, and disease. *Biology* 10, 366. doi:10.3390/biology10050366

Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K., et al. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* 12, 51–54. doi:10.1038/nmeth.3179

Laprade, H., Querido, E., Smith, M. J., Guerit, D., Crimmins, H., Conomos, D., et al. (2020). Single-molecule imaging of telomerase RNA reveals a recruitment-retention model for telomere elongation. *Mol. Cell* 79, 115–126. doi:10.1016/j.molcel.2020.05.005

Lewis, C. J. T., Pan, T., and Kalsotra, A. (2017). RNA modifications and structures cooperate to guide RNA–protein interactions. *Nat. Rev. Mol. Cell Biol.* 18, 202–210. doi:10.1038/nrm.2016.163

Li, S., and Mason, C. E. (2014). The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genomics Hum. Genet.* 15, 127–150. doi:10.1146/annurev-genom-090413-025405

Li, Y., Liu, S., Cao, L., Luo, Y., Du, H., Li, S., et al. (2021). Cbrpp: A new RNA-centric method to study RNA–protein interactions. *RNA Biol.* 18, 1608–1621. doi:10.1080/15476286.2021.1873620

Lin, X., Fonseca, M. A. S., Breunig, J. J., Corona, R. I., and Lawrenson, K. (2021). *In vivo* discovery of RNA proximal proteins via proximity-dependent biotinylation. *RNA Biol.* 18, 2203–2217. doi:10.1080/15476286.2021.1917215

Lindemann, C., Thomanek, N., Hundt, F., Lerari, T., Meyer, H. E., Wolters, D., et al. (2017). Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biol. Chem.* 398, 687–699. doi:10.1515/hsz-2017-0104

Mair, A., and Bergmann, D. C. (2022). Advances in enzyme-mediated proximity labeling and its potential for plant research. *Plant Physiol.* 188, 756–768. doi:10.1093/plphys/kiab479

Matěju, Daniel, and Chao, Jeffrey A. (Editors) (2022). *The Integrated stress response: Methods and protocols. Vol. 2428* (Berlin, Germany: Springer US).

McHugh, C. A., Russell, P., and Guttman, M. (2014). Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol.* 15, 203. doi:10.1186/gb4152

Michael Green, N. (1990). [5] Avidin and streptavidin. *Methods Enzym.* 184, 51–67.

Mukherjee, J., Hermesh, O., Eliscovich, C., Nalpas, N., Franz-Wachtel, M., Macek, B., et al. (2019). β-Actin mRNA interactome mapping by proximity biotinylation. *Proc. Natl. Acad. Sci. U. S. A.* 116, 12863–12872. doi:10.1073/pnas.1820737116

Nielsen, F. C., Hansen, H. T., and Christiansen, J. (2016). RNA assemblages orchestrate complex cellular processes. *BioEssays* 38, 674–681. doi:10.1002/bies.201500175

Padrón, A., Iwasaki, S., and Ingolia, N. T. (2019). Proximity RNA labeling by APEX-seq reveals the organization of translation initiation complexes and repressive RNA granules. *Mol. Cell* 75, 875e5–887. doi:10.1016/j.molcel.2019.07.030

Protter, D. S. W., Rao, B. S., Van Treeck, B., Lin, Y., Mizoue, L., Rosen, M. K., et al. (2018). Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly. *Cell Rep.* 22, 1401–1412. doi:10.1016/j.celrep.2018.01.036

Qin, W., Myers, S. A., Carey, D. K., Carr, S. A., and Ting, A. Y. (2021b). Spatiotemporally-resolved mapping of RNA binding proteins via functional proximity labeling reveals a mitochondrial mRNA anchor promoting stress recovery. *Nat. Commun.* 12, 4980. doi:10.1038/s41467-021-25259-2

Qin, W., Cho, K. F., Cavanagh, P. E., and Ting, A. Y. (2021a). Deciphering molecular interactions by proximity labeling. *Nat. Methods* 18, 133–143. doi:10.1038/s41592-020-01010-5

Ramanathan, M., Majzoub, K., Rao, D. S., Neela, P. H., Zarnegar, B. J., Mondal, S., et al. (2018). RNA–protein interaction detection in living cells. *Nat. Methods* 15, 207–212. doi:10.1038/nmeth.4601

Ramanathan, M., Porter, D. F., and Khavari, P. A. (2019). Methods to study RNA–protein interactions. *Nat. Methods* 16, 225–234. doi:10.1038/s41592-019-0330-1

Roux, K. J., Kim, D. I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* 196, 801–810. doi:10.1083/jcb.201112098

Samavarchi-Tehrani, P., Samson, R., and Gingras, A.-C. (2020). Proximity dependent biotinylation: Key enzymes and adaptation to proteomics approaches. *Mol. Cell. Proteomics* 19, 757–773. doi:10.1074/mcp.R120.001941

Somasekharan, S. P., Zhang, F., Saxena, N., Huang, J. N., Kuo, I. C., Low, C., et al. (2020). G3BP1-linked mRNA partitioning supports selective protein synthesis in response to oxidative stress. *Nucleic Acids Res.* 48, 6855–6873. doi:10.1093/nar/gkaa376

Spadotto, V., Giambruno, R., Massignani, E., Mihailovich, M., Maniaci, M., Patuzzo, F., et al. (2020). PRMT1-mediated methylation of the microprocessor-associated proteins regulates microRNA biogenesis. *Nucleic Acids Res.* 48, 96–115. doi:10.1093/nar/gkz1051

Tran, J. R., Paulson, D. I., Moresco, J. J., Adam, S. A., Yates, J. R., Goldman, R. D., et al. (2021). An APEX2 proximity ligation method for mapping interactions with the nuclear lamina. *J. Cell Biol.* 220, e202002129. doi:10.1083/jcb.202002129

Weissinger, R., Heinold, L., Akram, S., Jansen, R.-P., and Hermesh, O. (2021). RNA proximity labeling: A new detection tool for RNA–protein interactions. *Molecules* 26, 2270. doi:10.3390/molecules26082270

Yeo, Gene W. (Editor) (2014). *Systems biology of RNA binding proteins. Vol. 825* (New york, USA: Springer New York).

Yi, W., Li, J., Zhu, X., Wang, X., Fan, L., Sun, W., et al. (2020). CRISPR-assisted detection of RNA–protein interactions in living cells. *Nat. Methods* 17, 685–688. doi:10.1038/s41592-020-0866-0

Zhou, Y., Wang, G., Wang, P., Li, Z., Yue, T., Wang, J., et al. (2019). Expanding APEX2 substrates for proximity-dependent labeling of nucleic acids and proteins in living cells. *Angew. Chem. Int. Ed. Engl.* 58, 11763–11767. doi:10.1002/anie.201905949

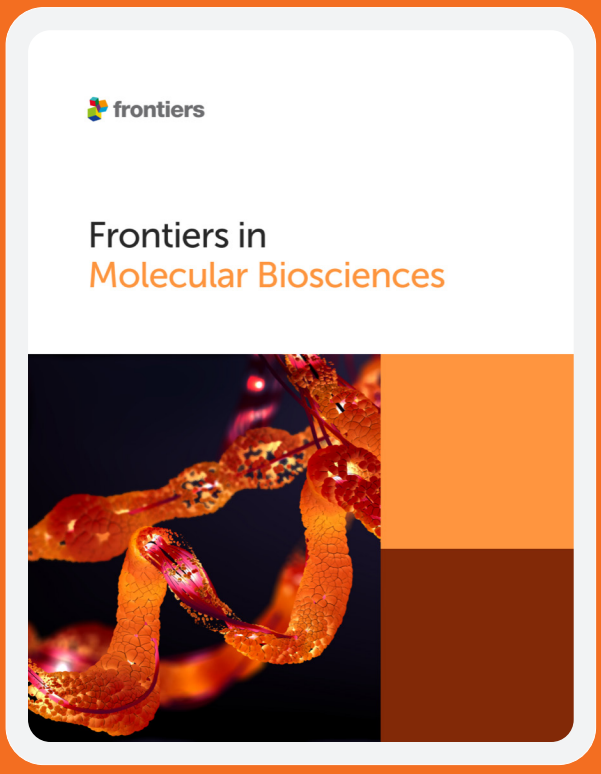
Frontiers in
Molecular Biosciences

Explores biological processes in living organisms on a molecular scale

Focuses on the molecular mechanisms underpinning and regulating biological processes in organisms across all branches of life.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact