frontiers | Research Topics

Insights in RNA 2022

Edited by
William C. Cho and Yadong Zheng

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-4801-1
DOI 10.3389/978-2-8325-4801-1

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: 2022

Topic editors

William C. Cho — QEH, Kowloon, SAR China

Yadong Zheng — Zhejiang Agriculture and Forestry University, China

Citation

Cho, W. C., Zheng, Y., eds. (2024). *Insights in RNA: 2022*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4801-1

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 22 February 2024
DOI 10.3389/fgene.2024.1382435

OPEN ACCESS

EDITED AND REVIEWED BY
Fangqing Zhao,
Beijing Institutes of Life Science (CAS), China

*CORRESPONDENCE
William C. Cho,
williamcscho@gmail.com
Yadong Zheng,
zhengyadong@zafu.edu.cn

RECEIVED 05 February 2024
ACCEPTED 16 February 2024
PUBLISHED 22 February 2024

CITATION
Li R, Zheng Y and Cho WC (2024), Editorial: Insights in RNA: 2022.
Front. Genet. 15:1382435.
doi: 10.3389/fgene.2024.1382435

Editorial: Insights in RNA: 2022

Rui Li[1], Yadong Zheng[1]* and William C. Cho[2]*

[1]Key Laboratory of Applied Technology on Green-Eco-Healthy Animal Husbandry of Zhejiang Province, Zhejiang Provincial Engineering Laboratory for Animal Health Inspection & Internet Technology, Zhejiang International Science and Technology Cooperation Base for Veterinary Medicine and Health Management, China-Australia Joint Laboratory for Animal Health Big Data Analytics, College of Animal Science and Technology and College of Veterinary Medicine of Zhejiang A&F University, Hangzhou, China, [2]Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong, China

KEYWORDS

message RNA, noncoding RNA, microRNA, RNA folding, RNA-seq

Editorial on the Research Topic
Insights in RNA: 2022

RNA is a captivating realm that consists of coding RNA and noncoding RNA, with the latter experiencing significant growth in recent years. The advent of next-generation sequencing technologies has shed light on the functional importance of noncoding RNA, dispelling the notion that it is merely "junk" in the genome. RNA plays a fundamental role in various biological processes, including genetic inheritance, mediating interactions between DNA and proteins, catalyzing biochemical reactions, and regulating gene expression. Given the rapid advancements in this field, we have launched the Research Topic "Insights in RNA: 2022" to provide an overview of the latest technologies, discoveries, and theories in the RNA world, inspiring future research endeavors.

This Research Topic includes a total of 9 published papers, comprising 3 research articles and 6 reviews (https://www.frontiersin.org/research-topics/46100/insights-in-rna-2022/articles). In one of the research articles by Zahid et al., they identified five potential targets of brain-specific microRNA-153 in Alzheimer's disease (AD). These targets, including ortilin-related receptor 1 (SORL1), amyloid precursor protein (APP), phosphatidylinositol binding clathrin assembly protein (PICALM), upstream stimulatory factor 1 (USF1), and presenilin-1 (PSEN1), are part of a protein interaction network implicated in AD. Previous studies have shown that APP, the precursor of the β-amyloid (Aβ) peptide, is downregulated in AD patients and negatively correlated with miR-153 expression (Liang et al., 2012; Long et al., 2012). AD is characterized by the accumulation of Aβ in senile plaques, and chronic brain hypoperfusion (CBH) has been implicated in Aβ deposition and synaptic plasticity reduction (de la Torre, 2021). In a recent study using a rat model of CBH, miR-153 was found to be upregulated and associated with impaired presynaptic vesicle release. Overexpression of miR-153 led to the suppression of several proteins involved in presynaptic vesicle release. Conversely, knockdown of miR-153 attenuated the decrease in presynaptic vesicle release and cognitive decline in the rat model, suggesting that miR-153 plays a role in impaired presynaptic plasticity in CBH (Yan et al., 2020). Notably, the expression levels of miR-153 in AD patients and the CBH rat model show opposite trends. This discrepancy may be due to differential expression at different stages of AD, with increased expression during synaptic dysfunction, which is implicated in the initiation of AD (Chakroborty et al., 2019). In addition to its role in presynaptic vesicle release, miR-153 has been shown to inhibit the differentiation and proliferation of neural stem cells, which have potential as disease-modifying biologics for AD treatment (Dong et al., 2023). Although the exact role of miR-153 in AD is still being elucidated, it is considered a

promising therapeutic target for combating this disease. Furthermore, the paper discusses the therapeutic potential of other miRNAs in AD (Zainal Abidin et al.).

Understanding the intricate structures of RNA is crucial for unraveling its functions, and the field of RNA structure prediction has garnered considerable interest. Machine learning (ML) algorithms have emerged as a potential approach for predicting potential structures of RNA sequences. In this Research Topic, Chasles and Major evaluated the effectiveness of ML algorithms with different parameters in predicting RNA folding, highlighting the need to optimize models for specific data. Various ML methods with different model architectures and output predictions have been developed, such as RNA3DCNN, trRosettaRNA, and DRfold. ML has also been successfully employed in identifying binding sites of metal ions, including Mg^{2+}, Na^{+}, and K^{+} (Zhao et al., 2023). However, when it comes to truly generalizing ML methods to unseen, structurally distinct RNA families (not just unseen sequences), they do not appear to have an advantage over traditional non-learning techniques (Wu et al., 2023). To further advance the application of ML in RNA structure prediction, it is necessary to establish standardized benchmark training examples/datasets, possibly using a cluster-based k-fold cross-validation approach (Wu et al., 2023).

Sequencing technologies, particularly RNA sequencing (RNA-seq), have revolutionized our understanding of cellular and tissue physiology and pathology. By providing genome-wide RNA expression profiles, transcriptomics enables us to examine the transcriptional landscape and identify differentially expressed molecules relevant to the biology and pathogenesis of interest. However, reliable transcriptomic data necessitates the extraction of high-quality total RNA. In this Research Topic, He et al. conducted a comparative evaluation of the performance of various commercial RNA extraction kits and examined the factors influencing RNA quality in sera used in clinical settings. They observed significant variations in the quality of extracted total RNA when different commercial kits were employed, and identified storage time and temperature of sera as negative factors. Furthermore, they found that all preanalytical processes introduced a bias to the transcriptomes, highlighting the importance of RNA quality control prior to RNA-seq. These findings emphasize the need for ensuring high-quality RNA for accurate and reliable downstream analyses.

The full review papers included in this Research Topic cover a range of important areas in RNA research. These include intron biology, RNA sequencing technologies for T cell receptors, RNA-based therapeutics for the treatment of lung and central nervous diseases, and the mechanisms underlying mRNA deadenylation. These reviews provide readers some of the latest advancements and hot topics in of RNA research, offering valuable insights for future studies and guiding researchers towards new directions.

Author contributions

RL: Writing–original draft. YZ: Writing–review and editing. WC: Writing–review and editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We are indebted to the authors for their valuable contributions to this Research Topic and for the reviewers for their constructive suggestions.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Chakroborty, S., Hill, E. S., Christian, D. T., Helfrich, R., Riley, S., Schneider, C., et al. (2019). Reduced presynaptic vesicle stores mediate cellular and network plasticity defects in an early-stage mouse model of Alzheimer's disease. *Mol. Neurodegener.* 14 (1), 7. doi:10.1186/s13024-019-0307-7

de la Torre, J. C. (2021). Deciphering alzheimer's disease pathogenic pathway: role of chronic brain hypoperfusion on p-tau and mTOR. *J. Alzheimers Dis.* 79 (4), 1381–1396. doi:10.3233/JAD-201165

Dong, X., Wang, H., Zhan, L., Li, Q., Li, Y., Wu, G., et al. (2023). miR-153-3p suppresses the differentiation and proliferation of neural stem cells via targeting GPR55. *Aging (Albany NY)* 15 (16), 8518–8527. doi:10.18632/aging.204002

Liang, C., Zhu, H., Xu, Y., Huang, L., Ma, C., Deng, W., et al. (2012). MicroRNA-153 negatively regulates the expression of amyloid precursor protein and amyloid precursor-like protein 2. *Brain Res.* 1455, 103–113. doi:10.1016/j.brainres.2011.10.051

Long, J. M., Ray, B., and Lahiri, D. K. (2012). MicroRNA-153 physiologically inhibits expression of amyloid-beta precursor protein in cultured human fetal brain cells and is dysregulated in a subset of Alzheimer disease patients. *J. Biol. Chem.* 287 (37), 31298–31310. doi:10.1074/jbc.M112.366336

Wu, K. E., Zou, J. Y., and Chang, H. (2023). Machine learning modeling of RNA structures: methods, challenges and future perspectives. *Brief. Bioinform* 24 (4). doi:10.1093/bib/bbad210

Yan, M. L., Zhang, S., Zhao, H. M., Xia, S. N., Jin, Z., Xu, Y., et al. (2020). MicroRNA-153 impairs presynaptic plasticity by blocking vesicle release following chronic brain hypoperfusion. *Cell Commun. Signal* 18 (1), 57. doi:10.1186/s12964-020-00551-8

Zhao, Y., Wang, J., Chang, F., Gong, W., Liu, Y., and Li, C. (2023). Identification of metal ion-binding sites in RNA structures using deep learning method. *Brief. Bioinform* 24 (2). doi:10.1093/bib/bbad049

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1091685

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Haixin Li,
Tianjin Medical University Cancer Institute and Hospital, China
Fei Yu,
Guangxi University, China

*CORRESPONDENCE
Xiaojiao Li,
lixiaojiao@xjtuhf.edu.cn
Qinying Luo,
Luoqy2009@163.com
Yawen Wang,
wangyw1269@xjtu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to RNA, a section of the journal Frontiers in Genetics

RECEIVED 07 November 2022
ACCEPTED 05 December 2022
PUBLISHED 04 January 2023

CITATION
He Y, Dong L, Yi H, Zhang L, Shi X, Su L, Gan B, Guo R, Wang Y, Luo Q and Li X (2023), Improper preanalytical processes on peripheral blood compromise RNA quality and skew the transcriptional readouts of mRNA and LncRNA.
Front. Genet. 13:1091685.
doi: 10.3389/fgene.2022.1091685

Improper preanalytical processes on peripheral blood compromise RNA quality and skew the transcriptional readouts of mRNA and LncRNA

Yinli He[1†], Lele Dong[2†], Hongyang Yi[3], Linpei Zhang[1], Xue Shi[1], Lin Su[1], Baoyu Gan[1], Ruirui Guo[1], Yawen Wang[1*], Qinying Luo[1*] and Xiaojiao Li[1*]

[1]BioBank, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China, [2]Department of Pharmacy, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China, [3]National Clinical Research Centre for Infectious Diseases, The Third People's Hospital of Shenzhen, The Second Affiliated Hospital of Southern University of Science and Technology, Shenzhen, Guangdong, China

Genetic and epigenetic reprogramming caused by disease states in other tissues is always systemically reflected in peripheral blood leukocytes (PBLs). Accurate transcriptional readouts of Messenger RNA (mRNA) and Long non-coding RNA (lncRNA) in peripheral blood leukocytes are fundamental for disease-related study, diagnosis and treatment. However, little is known about the impact of preanalytical variables on RNA quality and downstream messenger RNA and Long non-coding RNA readouts. In this study, we explored the impact of RNA extraction kits and timing of blood placement on peripheral blood leukocyte-derived RNA quality. A novel enhanced evaluation system including RNA yields, purity, RNA integrity number (RIN) values and β-actin copies was employed to more sensitively identify RNA quality differences. The expression levels of informative mRNAs and Long non-coding RNAs in patients with chronic obstructive pulmonary disease (COPD) or triple-negative breast cancer (TNBC) were measured by Quantitative reverse transcription polymerase chain reaction (qRT–PCR) to investigate the impact of RNA quality on transcriptional readouts. Our results showed that the quality of RNA extracted by different kits varies greatly, and commercial kits should be evaluated and managed before batch RNA extraction. In addition, the quality of extracted RNA was highly correlated with the timing of blood placement, and the copy number of β-actin was significantly decreased after leaving blood at RT over 12 h. More importantly, compromised RNA leads to skewed transcriptional readouts of informative mRNAs and Long non-coding RNAs in patients with chronic obstructive pulmonary disease or triple-negative breast cancer. These findings have significant implications for peripheral blood leukocyte-derived RNA quality management and suggest that quality control is necessary prior to the analysis of patient messenger RNA and Long non-coding RNA expression.

KEYWORDS

peripheral blood leukocytes (PBLs), preanalytical processes, RNA quality, transcriptional readouts, COPD, TNBC

1 Introduction

Peripheral blood leukocytes (PBLs) often act as proxies for disease states since they can systemically reflect genetic and epigenetic reprogramming in other tissues (Glade et al., 1968; Vorup-Jensen et al., 2005; Delobel et al., 2010; Javierre et al., 2010; Kiltschewskij and Cairns, 2020; Valero et al., 2020; Marczyk et al., 2021). RNA, which is the most important fraction of leukocytes, plays a pivotal role in understanding the transcriptomic profile of diseases (Raeymaekers, 1993). Messenger RNA (mRNA) is a pivotal molecule of life and is involved in almost all aspects of cell biology (Sahin et al., 2014). Long noncoding RNAs (lncRNAs) are a class of RNA transcripts with a length greater than 200 nucleotides that exert their functions by regulating gene expression and functions at transcriptional, translational, and posttranslational levels (Chi et al., 2019; Volovat et al., 2020). Any changes in mRNA and lncRNA expression levels may lead to inflammation and malignant disease, and aberrantly expressed mRNAs and lncRNAs are considered strong biomarkers in diseases. Hence, accurate transcriptional readouts of informative mRNAs and lncRNAs are fundamental for disease-related studies, diagnosis and treatment. RNA is frangible and prone to degradation (Auer et al., 2003; Zychowska et al., 2020), and the process of RNA extraction is susceptible to the variability of many factors, such as blood storage temperature and collection tubes (Delobel et al., 2010; Tang et al., 2019) (De Cecco et al., 2009; Rudloff et al., 2010; Hatzis et al., 2011). However, little is known about the impact of RNA extraction kits and timing of blood placement on RNA quality and downstream mRNA and lncRNA analysis.

RNA yields, purity and RNA integrity number (RIN) are routinely used as standard indicators to estimate RNA quality (Imbeaud et al., 2005; Schroeder et al., 2006). Recently, some studies suggested that these conventional indicators were not sensitive enough to discern between low- and high-quality RNA (Ribeiro-Silva et al., 2007; Greytak et al., 2015; Webster et al., 2015; Hester et al., 2016), implying that a novel evaluation system should be constructed to more accurately identify RNA quality inconsistencies. Fragmented RNA always leads to lower effective RNA inputs and further compromises the copy number of housekeeping genes in cDNA, which can be sensitively detected by digital droplet PCR (ddPCR) (Hindson et al., 2011; Pinheiro et al., 2012; Millier et al., 2017; Poh et al., 2020). Therefore, housekeeping gene copies may provide an enhanced indicator for RNA quality evaluation.

Quantitative reverse transcription polymerase chain reaction (qRT–PCR) is the most commonly used technique for detecting mRNA and lncRNA expression and verifying the candidates of RNA-seq (Nolan et al., 2006). Some reports have demonstrated that the cycle threshold (Ct) value had an opposite trend compared to the RIN (Wang et al., 2015), and RNA integrity significantly affected the 18 s, 28 s and IL-1β crossover point values (Fleige and Pfaffl, 2006). However, little is known about the impact of RNA quality on ΔΔCt values, which is the most commonly used indicator for representing the relative expression alteration in mRNA and lncRNA (Livak and Schmittgen, 2001). In this study, we first explored the impact of different RNA extraction kits and the timing of blood placement on the quality of PBL-derived RNA. A novel evaluation system including RNA yields, purity, RIN values and β-actin copies was employed to identify RNA quality inconsistencies. The expression levels of informative mRNAs and lncRNAs in patients with chronic obstructive pulmonary disease (COPD) or triple-negative breast cancer (TNBC) were detected by qRT–PCR to investigate the impact of RNA quality on mRNA and lncRNA readouts. Our results found that the quality of PBL-derived RNA extracted by different commercial kits varies greatly, suggesting that kit evaluation and management should be performed before batch RNA extraction. Additionally, the timing of blood placement should be limited to 12 h to obtain high-quality RNA. More importantly, the relative transcriptional readouts of the aberrantly expressed mRNAs and LncRNAs in patients with COPD or TNBC were heavily dependent on the quality of extracted RNA, and improper preanalytical processes resulted in skewed qRT–PCR results. This study thus comprehensively evaluated the preanalytical processes on PBL-derived RNA quality and downstream mRNA and lncRNA readouts.

2 Results

2.1 The quality of PBL-derived RNA extracted by different commercial kits varies greatly

Various methods have been developed to purify PBL-derived RNA, and the red blood cell (RBC) lysis method has been primarily recommended for biobanks for further research (Heng et al., 2018). In recent years, several commercial kits have been successfully developed based on the RBC lysis method. However, due to the diversity of detection methods of various manufacturers, quality inconsistences may exist in RNA processed with different extraction kits. Therefore, it is necessary to evaluate the impact of different extraction kits on the quality of PBL-derived RNA. Thirty volunteers were recruited, and PBL-derived RNA was extracted immediately using three commercial RNA extraction kits: kit 1, kit 2, and kit 3. TRIzol reagent was employed as a standard control group to compare

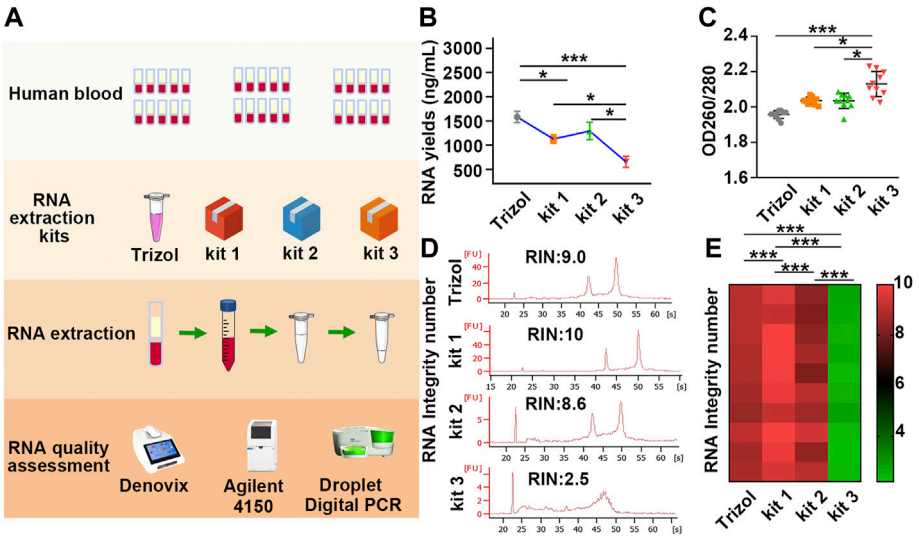


FIGURE 1
The yields, purity and RIN values of RNA extracted by different kits vary greatly. **(A)** Schematic illustration of the experimental schedule. **(B)** RNA yields in each group were assessed by Denovix. $n = 40$. **(C)** The 260/280 values were quantified using a DeNovix spectrophotometer. $n = 40$. **(D)** Representative image of RNA analysis by an Agilent bioanalyzer in each group. The first peak is a 20 bp molecular marker. The second and third peaks are 18 s and 28 s rRNA. **(E)** A heatmap based on RIN values in each sample. Red and green represent high and low RIN values, respectively. Data are represented as the mean ± SD ($n = 10$). Tamhane's T2 test **(B,C)**, LSD test **(E)**, $*p < .05$, $**p < .01$, $***p < .001$.

the quality and efficiency of different commercial kits. A novel enhanced evaluation system including RNA yields, purity, RIN values and β-actin copies was employed to identify the RNA quality differences (Figure 1A).

Our results showed that the yields of RNA extracted by TRIzol were slightly higher than those of RNA purified by kit 1 and kit 2, while the mean yields of RNA extracted by kit 3 were half as low as those of RNA purified by TRIzol, kit 1 and kit 2 (Figure 1B; Table 1). In addition, the 260/280 values of RNA extracted by kit 3 were considerably higher than 2 (Figure 1C; Table 1), suggesting the presence of protein contamination in RNA extracted by kit 3 (Manchester, 1996; Okamoto and Okabe, 2000; Wahlberg et al., 2012). The RIN value in each group was assessed using an Agilent 4150 Bioanalyzer, which can provide a separate RIN value as well as the correlating electrophoretic gel-like image for each sample (Imbeaud et al., 2005; Schroeder et al., 2006). As the results showed, RNA purified by TRIzol, kit 1 and kit 2 tended to be intact, and TRIzol and kit 1 exhibited better performance than kit 2. However, RNA extracted by kit 3 suffered severe degradation, with RIN values of only 2.5 ± 0.2 (Figures 1D, E; Supplementary Figure S1; Table 1). Gaps in RNA integrity may depend primarily on the composition of the three kits. In kit 1, β-mercaptoethanol (β-ME), a well-known reducing agent that irreversibly denatures RNase by reducing disulfide bonds and destroying the native conformation (van der Poel-van de Luytgaarde et al., 2013), was added to the leukocyte lysis buffer to eliminate the RNase. In kit 2, cells were lysed by guanidinium thiocyanate–phenol, which is also used in the TRIzol kit and can prevent the activity of RNA enzymes by denaturing them to yield undegraded RNA (Chomczynski and Sacchi, 1987). However, no RNase inactivator is labelled in Kit 3, which may be the main reason for the severe degradation of RNA. In support of this notion, we added 1% β-ME to the leukocyte lysis buffer. As our results showed, the presence of β-ME significantly increased the mean RIN value from 2.5 ± 0.2 to 7.0 ± 1.4 (Supplementary Figure S2; Supplementary Table S1), indicating that the poor performance of kit 3 was mainly due to the lack of RNase inactivator.

DdPCR was performed to detect the variance of β-actin copies to more accurately evaluate the inconsistencies of RNA in each group. In ddPCR, the reaction is divided into at least 10,000 partitions, and 40 cycles of classical PCR are carried out in each of these impervious nanocompartments. Those partitions with the amplified product are designated positive, and those with no amplified product are designated negative. Quantification was then achieved using Poisson statistics by counting fluorescence-positive and total droplet numbers (Passby et al., 2019) (Figure 2A). As the results showed, samples processed with kit 1 exhibited the highest β-actin copies. In particular, β-actin copies in samples processed with kit 1 or kit 2 were more than 400 times higher than those processed with kit 3 (Figures 2B–D; Supplementary Table S2), yet the variation in RIN values was only 4-fold across different groups, suggesting that the copy number of β-actin was more sensitive for identifying RNA integrity than RIN values. These results demonstrated that the quality of PBL-derived RNA

TABLE 1 The quality of PBL-derived RNA extracted by different kits varies widely[a].

Kits	RNA yields (ng/ml)	OD260/280	RIN
TRIzol	1,580.25 ± 371.10	1.96 ± 0.02	8.9 ± 0.2
kit 1	1,125.37 ± 288.19	2.04 ± 0.02	9.7 ± 0.3
kit 2	1,290.80 ± 565.84	2.03 ± 0.04	8.6 ± 0.4
kit 3	652.43 ± 359.49	2.13 ± 0.07	2.5 ± 0.2

[a]Results are means and standard deviations of ten independent extractions.

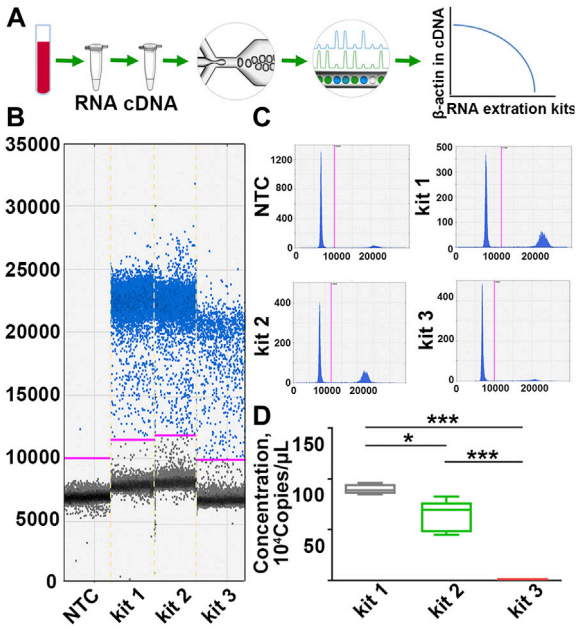


FIGURE 2
The copy number of β-actin is a more sensitive indicator to identify RNA quality inconsistencies. **(A)** A scheme for the detection principle of ddPCR. **(B,C)** Representative one-dimensional plots of droplets measured for fluorescence signal (ampli-tude indicated on *y*-axis) emitted from β-actin. Blue dots are positive signals, and black dots represent non-amplification signals. **(D)** Quantification of β-actin copies in samples processed with different RNA extraction kits. $n = 30$. Data are represented as the mean ± SD ($n = 10$). Tamhane's T2 test (D), $*p < .05$, $***p < .001$.

extracted by different commercial kits varies widely, and β-actin copies can be a more sensitive indicator to assess RNA quality.

2.2 Blood placed at RT over 12 h significantly compromise PBL-derived RNA quality

Although the fresher the better is the golden rule when dealing with clinical samples, in most cases, blood collection and further RNA extraction are always conducted at different times and in different spaces, so it is necessary to find a balance between the storage conditions and RNA quality. Whereas previous studies have demonstrated that the storage temperature and duration are critical to RNA quality, the conclusion needs more verification due to the limited samples and evaluation system (Zhang et al., 2019).

To explore the impact of the timing of blood placement at RT on the quality of PBL-derived RNA, sixty tubes of blood were placed on the lab bench at RT for 0, 30, 60, and 90 min before RNA extraction with kit 1 (Figure 3A). Our results showed that timing of blood at RT within 12 h displayed undetectable lesions on RNA quality. Prolonging the blood placement time to 24 h significantly compromised β-actin copies but had little effect on the yields, purity and RIN values of extracted RNA, confirming that the copy number of β-actin is a more sensitive indicator for RNA quality evaluation (Figures 3B–G; Supplementary Figure S3; Table 2). In particular, the longer the timing of blood

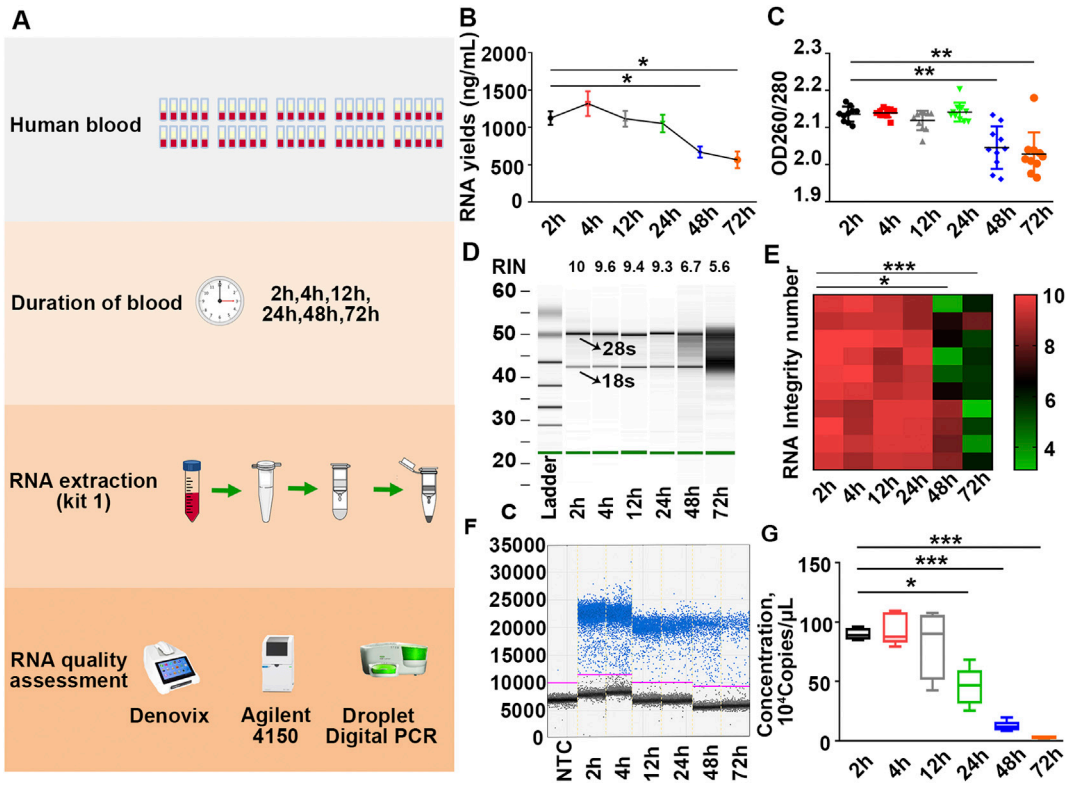


FIGURE 3
Placing blood at RT over 12 h significantly compromises RNA quality. **(A)** Schematic illustration of the experimental schedule. **(B,C)** Yields **(B)** and purity **(C)** of RNA in each group were measured by Denovix spectrophotometry. *n* = 60. **(D)** Representative electrophoretic gel-like image of RNA in each group. The arrows indicate 18 s and 28 s ribosomal bands. **(E)** Heatmap of RIN values of samples. *n* = 60. **(F)** Representative one-dimensional plots of droplets measured for fluorescence signal emitted from β-actin in each group. **(G)** Number of β-actin copies in each group. *n* = 60. Tamhane's T2 test **(B,C,E,G)**, **p* < .05, ***p* < .01, ****p* < .001.

placement, the poorer the RNA quality. Placing blood at RT for 48 h markedly compromised RNA yields, purity, RIN values and β-actin copies (Figures 3B–G; Supplementary Figure S3; Table 2). These results implied that the timing of blood placement at RT should be performed within 12 h to obtain desirable RNA.

2.3 Compromised RNA leads to skewed mRNA and lncRNA readouts in patients with COPD or TNBC

Having found the influence of extraction kits and the timing of blood placement on PBL-derived RNA quality, we further investigated the impact of RNA quality on downstream transcriptional readouts. Inflammation plays a pivotal role in the pathogenesis of COPD, where $CD8^+$ T lymphocytes, neutrophils, and macrophages are the main types of immune cells of the local inflammatory milieu (Cosio et al., 2002; Stankiewicz et al., 2002). Previous studies have identified informative mRNAs (CXCL16, HMOX1, SLA2, *etc.*) and lncRNAs (ENST00000502883.1, HIT000648516, XR_429541.1, *etc.*) in peripheral blood mononuclear cells (PBMCs) from COPD patients *versus* smokers (Sui et al., 2013; Song et al., 2015; Dang et al., 2017; Qu et al., 2018). Here, a smoker and a COPD patient were recruited, and the blood of each individual was equally divided into three aliquots for RNA isolation by the three RNA extraction kits. The expression of informative mRNAs (CXCL16, HMOX1, SLA2) and lncRNAs (ENST00000502883.1, HIT000648516, XR_429541.1) was detected in each sample (Figure 4A).

As the results showed, the quality of RNA extracted by different kits varied greatly even for the same individual (Figures 4B,C; Supplementary Figure S4; Supplementary Table S3). To minimize the interference of RNA inputs on transcriptional readouts, we performed relative quantification according to the ΔΔCt method, and the expression level of β-actin was used as an internal control (Nolan et al., 2006). As the results showed, when PBL-derived RNA was extracted by kit 1, the mean relative transcriptional readout of CXCL16 in COPD patient was 3 times higher than that in healthy control (HC) (Figure 4D). However, the

TABLE 2 Blood placed at RT over 12 h significantly compromised PBL-derived RNA quality[a].

Time(h)	RNA yields (ng/ml)	OD260/280	RIN	β-actin copies (Copies/μl)
2	1,125.37 ± 288.19	2.04 ± 0.02	9.7 ± 0.3	897,200 ± 51,621
4	1,319.51 ± 524.09	2.04 ± 0.01	9.4 ± 0.5	938,560 ± 135,087
12	1,116.02 ± 336.48	2.02 ± 0.03	9.4 ± 0.4	826,400 ± 294,752
24	1,053.13 ± 372.91	2.04 ± 0.03	9.4 ± 0.4	457,280 ± 159,569
48	670.14 ± 231.77	1.95 ± 0.06	6.6 ± 2.0	123,872 ± 45,041
72	567.16 ± 352.47	1.93 ± 0.06	5.5 ± 1.3	29,600 ± 6,556

[a]Results are means and standard deviations of ten independent extractions.

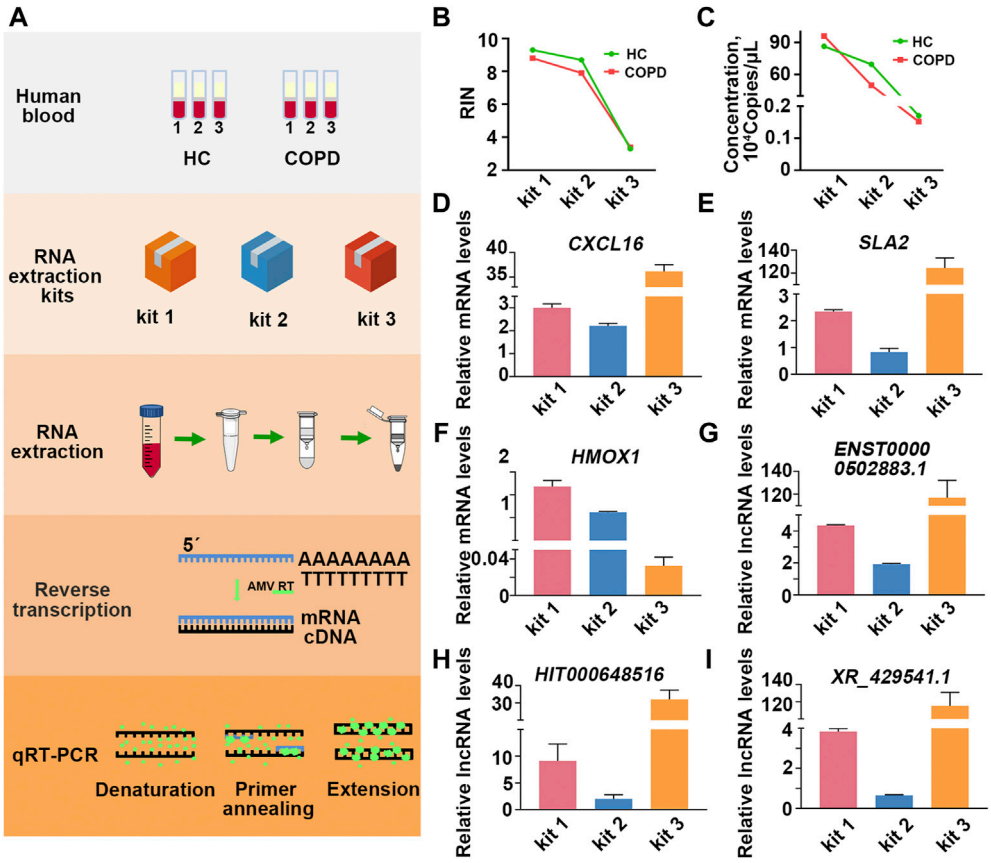


FIGURE 4
Comprised RNA leads to skewed transcriptional readouts in COPD patient. **(A)** Schematic illustration of the experimental schedule. **(B)** The RIN values in each group. **(C)** The quantification of β-actin copies in each group. **(D–F)** The relative transcriptional readouts of the informative mRNAs in COPD patient in each group. **(G–I)** The relative informative lncRNAs readouts in COPD patient in each group.

alteration of CXCL16 mRNA levels between the two individuals was approximately 2-fold and 40-fold when samples were processed with kit 2 and kit 3, respectively (Figure 4D; Table 3). Similar results were observed in another informative COPD marker, SLA2 (Figure 4E; Table 3). Even more striking, the relative transcriptional readout of HMOX1 in COPD patient was markedly increased in RNA extracted by kit 1 but significantly decreased in RNA treated with kit 3 (Figure 4F; Table 3). Moreover, the relative transcriptional readouts of informative lncRNAs in COPD (ENST00000502883.1, HIT000648516, XR_429541.1) were also affected by PBL-derived RNA

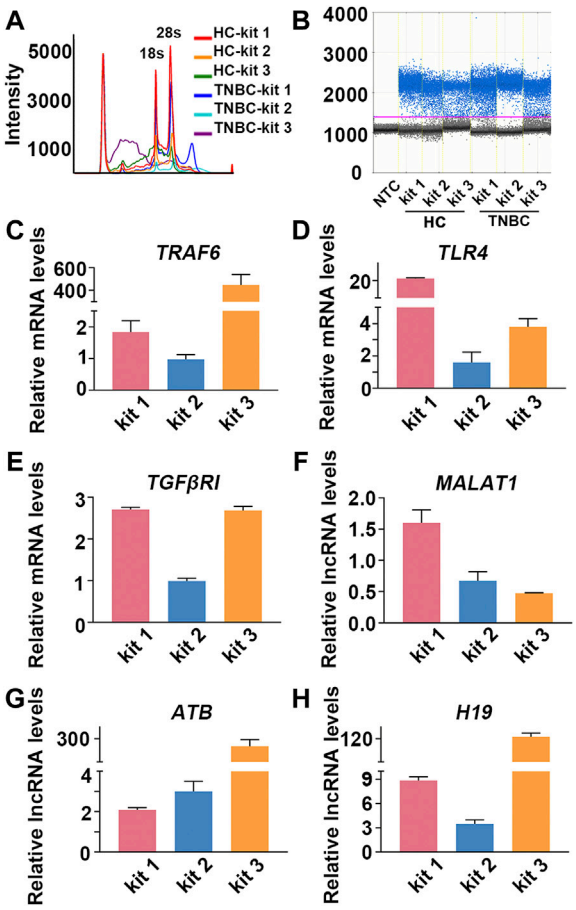


FIGURE 5
Compromised RNA results in skewed mRNA and lncRNA readouts in TNBC patient. **(A)** Representative image of RNA analysis by an Agilent bioanalyzer in each group. **(B)** Representative one-dimensional plots of droplets measured for fluorescence signal emitted from β-actin in each group. **(C–H)** The relative transcriptional readouts of the informative mRNAs and lncRNAs in TNBC patient in each group.

quality, and compromised RNA significantly resulted in skewed results (Figures 4G–I; Table 3).

In addition to inflammatory disease, interactions between the immune system and tumors are highly reciprocal in nature, and the presence of cancer cells causes immune cells to undergo various phenotypic and functional changes (Dirkx et al., 2006; Talmadge and Gabrilovich, 2013; Noy and Pollard, 2014; Suzuki et al., 2019). Based on these concepts, attempts have been made to detect the presence of cancer cells by analyzing the gene expression profile of PBMCs in patients with lung cancer, pancreatic cancer and breast cancer (Saeed et al., 2003; Showe et al., 2009). TNBC is a very aggressive subtype of normal breast cancer, and great efforts have been made to identify informative mRNAs and lncRNAs in TNBC. A previous study reported that TLR4, TNF receptor-associated factor 6 (TRAF6), and TGF-beta receptor type I (TGFβRI) were considerably upregulated in TNBC patients. In addition, lncRNA-ATB (Li et al., 2018), metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) and lncRNA H19 (Jin et al., 2016; Li et al., 2020; Huang et al., 2021), which are abnormally expressed in TNBC, may provide a less invasive diagnostic procedure to reveal immunological insight of breast cancer. Thus, we explored the influence of PBL-derived RNA quality on the transcriptional readouts of informative mRNAs and lncRNAs in TNBC.

As the results showed, the quality of RNA extracted by different kits was quite different in each individual. The performance of kit 1 was better than that of kit 2, and RNA extracted by kit 3 suffered severe degradation (Figures 5A,B; Supplementary Figure S5; Supplementary Table S4). Importantly, the relative transcriptional readouts of TRAF6 in TNBC patient were 1.8 times higher than that detected in HC when RNA was extracted by kit 1 (Figure 5C; Table 4). However, there was no difference in the expression level of TRAF6 when samples were treated with kit 2, and processing samples with kit 3 resulted in a suspicious increase in the mRNA content of TRAF6 (Figure 5C; Table 4). The expression level of TLR4 in TNBC patient was 21-fold higher than that in HC when RNA was isolated by kit 1, while a significant decrease in TLR4 expression in patient were observed when samples were processed with kit 2 and kit 3 (Figure 5D; Table 4). Interestingly, for TGFβRI readouts, there was no difference between the samples processed with kit 1 and kit 3 (Figure 5E; Table 4). In accordance with the mRNA results, the relative transcriptional readouts of the same lncRNA in each sample processed with distinct extraction kits were quite different and irregular (Figures 5F–H; Table 4). These results demonstrated that the transcriptional readouts of mRNAs and LncRNAs were heavily dependent on the quality of extracted RNA, and improper preanalytical handling may lead to skewed results.

TABLE 3 The relative informative mRNA and lncRNA readouts in COPD in each group.

Kits/RNA	*CXCL16*	*HMOX1*	*SLA2*	*ENST00000 502,883.1*	*HIT000 648,516*	*XR_429541.1*
kit 1	3.00 ± 0.14	1.69 ± 0.10	2.35 ± 0.05	4.34 ± 0.04	9.13 ± 2.56	3.83 ± 0.12
kit 2	2.22 ± 0.09	1.11 ± 0.02	0.83 ± 0.11	1.91 ± 0.04	2.02 ± 0.63	0.65 ± 0.03
kit 3	36.20 ± 1.10	0.03 ± 0.01	124.60 ± 7.11	117.15 ± 12.31	32.11 ± 4.34	135.17 ± 12.87

TABLE 4 The relative transcriptional readouts of informative mRNAs and lncRNAs in TNBC in each group.

Kits/RNA	*TLR4*	*TGFβ1*	*TRAF6*	*MALAT-1*	*ATB*	*H19*
kit 1	21.25 ± 0.34	2.71 ± 0.04	1.84 ± 0.29	1.60 ± 0.17	2.09 ± 0.09	8.86 ± 0.37
kit 2	1.59 ± 0.52	0.99 ± 0.05	0.97 ± 0.12	0.67 ± 0.12	3.00 ± 0.41	3.48 ± 0.41
kit 3	3.80 ± 0.41	2.69 ± 0.08	449.18 ± 74.91	0.48 ± 0.01	268.23 ± 23.23	121.88 ± 2.51

3 Conclusion and discussion

Genetic and epigenetic reprogramming caused by disease states in other tissues are systemically reflected in peripheral blood leukocytes (Javierre et al., 2010; Wang et al., 2010; Smith et al., 2014). As the most important fraction of leukocytes, RNA has been commonly used to study the response of transcriptome to disease-related stress. A large number of studies have demonstrated that mRNAs and lncRNAs are involved in various diseases, such as inflammation and malignancy, and aberrantly expressed mRNAs and lncRNAs are considered possible strong biomarkers. An accurate transcriptional readout of mRNA and lncRNA is fundamental for disease-related study, diagnosis and treatment. However, studies of preanalytical variables on RNA quality and downstream mRNA and lncRNA readouts are still lacking.

Numerous studies have demonstrated that the process of PBL-derived RNA extraction is susceptible to the variability of many factors (De Cecco et al., 2009; Rudloff et al., 2010; Hatzis et al., 2011), while the potential impact of different extraction kits on extracted RNA quality remains poorly understood. In addition, although some experiments suggested that storage time was critical for RNA quality (Zhang et al., 2019), the impact of timing of blood placement at RT on RNA quality needs more verification due to the limited samples and evaluation system. Moreover, some reports have pointed out that conventional indicators, such as RNA yields, purity and RIN values, might not be sensitive enough to discern between low- and high-quality RNA (Ribeiro-Silva et al., 2007; Greytak et al., 2015; Webster et al., 2015; Hester et al., 2016), suggesting that an enhanced evaluation system should be constructed to assess RNA quality inconsistencies.

qRT–PCR is the most commonly used technique to detect the expression of mRNA and lncRNA. Some researchers have explored the correlation between RIN values and qRT–PCR results and found that the expression values of 4 housekeeping genes (GAPDH, KYNF, NEFL, β2M) were heavily reliant on RNA integrity (Schroeder et al., 2006). Additionally, Ct values had an opposite trend compared to the RIN (Wang et al., 2015). A few studies have assessed the influence of RNA integrity on CP and delta CP. Their findings showed that CP decreased with increasing RIN, while delta CP was slightly affected (Fleige and Pfaffl, 2006). However, little is known about the effect of RNA quality on relative gene expression calculated by delta-delta Ct, which has been the most preferred method for qRT–PCR data analysis.

In this study, we first assessed the effects of extraction kits and timing of blood placement on PBL-derived RNA quality. A novel enhanced evaluation system including RNA yields, purity, RIN values and β-actin copies was employed to identify RNA differences. We found that the quality of RNA extracted by kit 1 and kit 2 was comparable to that of RNA purified by TRIzol, and kit 1 had a better performance than kit 2 in RIN values and β-actin copies. However, RNA extracted by kit 3 was subjected to severe degradation due to the lack of RNase inactivator, suggesting that kit evaluation and management should be performed before RNA extraction in batches. In addition, placement of blood at RT over 12 h significantly compromised the copy number of β-actin, indicating that the timing of blood placement at RT should be within 12 h to obtain desirable RNA. More importantly, the relative transcriptional readouts of the informative mRNAs and lncRNAs in patients with TNBC or COPD were heavily dependent on the quality of extracted RNA, and improper preanalytical handling may lead to skewed results. Our findings thus exhibited significant implications for PBL-derived RNA quality assessment and downstream qRT–PCR analysis.

4 Materials and methods

4.1 Blood collection

In brief, 5 ml blood in each sample was centrifuged at 1,500 g for 10 min at 4°C. After centrifugation, three different fractions are distinguishable: The upper layer is plasma; the intermediate layer is buffy coat, which concentrates leukocytes; and the bottom layer contains concentrated erythrocytes. The buffy coat was then harvested, and RBC lysis buffer was added to lyse the remaining erythrocytes. After centrifugation, total WBCs were harvested from the pellets. The following procedure was carried out according to the manufacturer's instructions for the evaluated RNA extraction kits (kit 1, kit 2 and kit 3).

A COPD patient was eligible for this study if he met the following criteria: smoking history (≥20 pack years); postbronchodilator FEV1 ≥ 25% of the predicted value and postbronchodilator FEV1/forced vital capacity (FVC) ≤ 0.70; and no history of asthma, atopy (as assessed by an allergy skin prick test during screening) or any other active lung disease.

A TNBC patient was eligible for this study if she met the following criteria: grade 2 or 3 infiltrating ductal carcinoma with

negative expression of ER, PR and HER2 proteins or accompanied by medullary features, infiltrating micropapillary carcinoma, or occasional vascularized thrombus.

4.2 RNA quantification

The yields and purity of extracted RNA were assessed by using a denovix spectrophotometer. The absorption at UV 260 nm was used to assess the RNA yields, and the ratio of 260 nm and 280 nm was used to evaluate RNA purity. An Agilent 4150 Bioanalyzer and the RNA 6000 Nano LabChip kit were employed to calculate the RNA Integrity Number (RIN). The RIN index ranges from 1 to 10, with 1 indicating the greatest degradation and 10 being the most intact RNA (Schroeder et al., 2006). The copy number of β-actin was detected on a digital droplet PCR (ddPCR) platform. Briefly, RNA in each group was reverse-transcribed into cDNA with the ReverTraAce qRT–PCR RT Kit (TOYOBO, Cat: FSQ-101). The aqueous ddPCR mixture containing 10 μl of ddPCR™ EvaGreen Supermix (BIO-RAD, Cat: #1864033), 3 μl of β-actin primers (3.75 μM) and 7 μl of cDNA was emulsified into picoliter droplets of thermostable oil in a QX200™ Droplet Generator. Subsequently, β-actin was amplified on a QX200 PCR system (Bio-Rad) at 95°C (30 s) and 60°C (60 s) for 40 PCR cycles. The ramp rate between any two consecutive steps was set to 2°C to ensure reliable thermal control. Next, the positive *versus* negative droplets were read by a QX200™ Droplet Reader, and the absolute quantification of β-actin was calculated using QuantaSoft software (Bio-Rad). Primer sequences for ddPCR are listed in Supplementary Table S5.

4.3 Real-time PCR

The relative transcriptional readouts of the informative mRNAs and LncRNAs in patients with COPD or TNBC were measured by qRT–PCR. Quantitative PCR was performed with SYBR Green Real-Time PCR Master Mix (TOYOBO, Cat: QPK-201) in the Real-Time PCR System (Bio-Rad, CFX96). The cycling profile was as follows: initial denaturation at 95°C for 5°min, followed by 40 cycles with 20 s at 95°C, 20 s at 60°C and 30 s at 72°C. The expression level of β-actin was used as an internal control. Primer sequences for qRT–PCR are listed in Supplementary Table S5.

4.4 Statistical analysis

Statistical analysis was performed using SPSS version 21 (IBM Corp.). All statistical graphs were constructed using Prism 8.0 (GraphPad Software, Inc.). All the results are presented as the mean ± standard deviation (SD). The comparison between groups was drawn through *t*-test and analysis of variance. Multiple comparison correction was conducted by Tamhane's T2 (or T2') and LSD(L) test for normally distributed data with unequal or equal variances, respectively. $P < .05$ was considered to indicate a statistically significant difference.

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 authors.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Xi'an Jiaotong University . The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization, XL, QL, and YW; methodology, YH, LD; validation, YH, LZ, LS, BG, and RG; formal analysis, HY; investigation, YH and LD; resources, YH; data curation, YH and XS; writing—original draft preparation, YH and LD; writing—review and editing, XL; visualization, YW; supervision, XL and QL; project administration, XL and YW.

Funding

This research was funded by the Science and Technology Planning Project of Shaanxi Province, China (No. 2020KRM-202), the National Natural Science Foundation of China (No. 81903320), the Science and Technology Planning Project of Shaanxi Province, China (No. 2021JM-280), the institutional foundation of the first affiliated hospital of Xi'an Jiaotong University (No. 2019QN-18), and the Fundamental Research Funds for the Central Universities (xzy012022112).

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/fgene.2022.1091685/full#supplementary-material

References

Auer, H., Lyianarachchi, S., Newsom, D., Klisovic, M. I., Marcucci, G., Kornacker, K., et al. (2003). Chipping away at the chip bias: RNA degradation in microarray analysis. *Nat. Genet.* 35 (4), 292–293. doi:10.1038/ng1203-292

Chi, Y., Wang, D., Wang, J., Yu, W., and Yang, J. (2019). Long non-coding RNA in the pathogenesis of cancers. *Cells* 8 (9), 1015. doi:10.3390/cells8091015

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162 (1), 156–159. doi:10.1006/abio.1987.9999

Cosio, M. G., Majo, J., and Cosio, M. G. (2002). Inflammation of the airways and lung parenchyma in COPD: role of T cells. *Chest* 121 (5), 160S–5S. doi:10.1378/chest.121.5_suppl.160s

Dang, X., Qu, X., Wang, W., Liao, C., Li, Y., Zhang, X., et al. (2017). Bioinformatic analysis of microRNA and mRNA Regulation in peripheral blood mononuclear cells of patients with chronic obstructive pulmonary disease. *Respir. Res.* 18 (1), 4. doi:10.1186/s12931-016-0486-5

De Cecco, L., Musella, V., Veneroni, S., Cappelletti, V., Bongarzone, I., Callari, M., et al. (2009). Impact of biospecimens handling on biomarker research in breast cancer. *BMC Cancer* 9, 409. doi:10.1186/1471-2407-9-409

Delobel, J., Rubin, O., Prudent, M., Crettaz, D., Tissot, J. D., and Lion, N. (2010). Biomarker analysis of stored blood products: Emphasis on pre-analytical issues. *Int. J. Mol. Sci.* 11 (11), 4601–4617. doi:10.3390/ijms11114601

Dirkx, A. E., Oude Egbrink, M. G., Wagstaff, J., and Griffioen, A. W. (2006). Monocyte/macrophage infiltration in tumors: Modulators of angiogenesis. *J. Leukoc. Biol.* 80 (6), 1183–1196. doi:10.1189/jlb.0905495

Fleige, S., and Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Asp. Med.* 27 (2-3), 126–139. doi:10.1016/j.mam.2005.12.003

Glade, P. R., Kasel, J. A., Moses, H. L., Whang-Peng, J., Hoffman, P. F., Kammermeyer, J. K., et al. (1968). Infectious mononucleosis: Continuous suspension culture of peripheral blood leucocytes. *Nature* 217 (5128), 564–565. doi:10.1038/217564a0

Greytak, S. R., Engel, K. B., Bass, B. P., and Moore, H. M. (2015). Accuracy of molecular data generated with FFPE biospecimens: Lessons from the literature. *Cancer Res.* 75 (8), 1541–1547. doi:10.1158/0008-5472.CAN-14-2378

Hatzis, C., Sun, H., Yao, H., Hubbard, R. E., Meric-Bernstam, F., Babiera, G. V., et al. (2011). Effects of tissue handling on RNA integrity and microarray measurements from resected breast cancers. *J. Natl. Cancer Inst.* 103 (24), 1871–1883. doi:10.1093/jnci/djr438

Heng, Z., Ruan, L., and Gan, R. (2018). Three methods to purify leukocytes and RNA quality assessment. *Biopreserv. Biobank.* 16 (6), 434–438. doi:10.1089/bio.2018.0058

Hester, S. D., Bhat, V., Chorley, B. N., Carswell, G., Jones, W., Wehmas, L. C., et al. (2016). Editor's highlight: Dose-response analysis of RNA-seq profiles in archival formalin-fixed paraffin-embedded samples. *Toxicol. Sci.* 154 (2), 202–213. doi:10.1093/toxsci/kfw161

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., et al. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* 83 (22), 8604–8610. doi:10.1021/ac202028g

Huang, Y., Zhou, Z., Zhang, J., Hao, Z., He, Y., Wu, Z., et al. (2021). lncRNA MALAT1 participates in metformin inhibiting the proliferation of breast cancer cell. *J. Cell. Mol. Med.* 25 (15), 7135–7145. doi:10.1111/jcmm.16742

Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., et al. (2005). Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res.* 33 (6), e56. doi:10.1093/nar/gni054

Javierre, B. M., Fernandez, A. F., Richter, J., Al-Shahrour, F., Martin-Subero, J. I., Rodriguez-Ubreva, J., et al. (2010). Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.* 20 (2), 170–179. doi:10.1101/gr.100289.109

Jin, C., Yan, B., Lu, Q., Lin, Y., and Ma, L. (2016). Reciprocal regulation of Hsa-miR-1 and long noncoding RNA MALAT1 promotes triple-negative breast cancer development. *Tumour Biol.* 37 (6), 7383–7394. doi:10.1007/s13277-015-4605-6

Kiltschewskij, D. J., and Cairns, M. J. (2020). Transcriptome-Wide analysis of interplay between mRNA stability, translation and small RNAs in response to neuronal membrane depolarization. *Int. J. Mol. Sci.* 21 (19), 7086. doi:10.3390/ijms21197086

Li, R. H., Chen, M., Liu, J., Shao, C. C., Guo, C. P., Wei, X. L., et al. (2018). Long noncoding RNA ATB promotes the epithelial-mesenchymal transition by upregulating the miR-200c/Twist1 axe and predicts poor prognosis in breast cancer. *Cell Death Dis.* 9 (12), 1171. doi:10.1038/s41419-018-1210-9

Li, Y., Ma, H. Y., Hu, X. W., Qu, Y. Y., Wen, X., Zhang, Y., et al. (2020). LncRNA H19 promotes triple-negative breast cancer cells invasion and metastasis through the p53/TNFAIP8 pathway. *Cancer Cell Int.* 20, 200. doi:10.1186/s12935-020-01261-4

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25 (4), 402–408. doi:10.1006/meth.2001.1262

Manchester, K. L. (1996). Use of UV methods for measurement of protein and nucleic acid concentrations. *Biotechniques* 20 (6), 968–970. doi:10.2144/96206bm05

Marczyk, M., Polanska, J., Wojcik, A., and Lundholm, L. (2021). Analysis of the applicability of microRNAs in peripheral blood leukocytes as biomarkers of sensitivity and exposure to fractionated radiotherapy towards breast cancer. *Int. J. Mol. Sci.* 22 (16), 8705. doi:10.3390/ijms22168705

Millier, M. J., Stamp, L. K., and Hessian, P. A. (2017). Digital-PCR for gene expression: Impact from inherent tissue RNA degradation. *Sci. Rep.* 7 (1), 17235. doi:10.1038/s41598-017-17619-0

Nolan, T., Hands, R. E., and Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* 1 (3), 1559–1582. doi:10.1038/nprot.2006.236

Noy, R., and Pollard, J. W. (2014). Tumor-associated macrophages: From mechanisms to therapy. *Immunity* 41 (1), 49–61. doi:10.1016/j.immuni.2014.06.010

Okamoto, T., and Okabe, S. (2000). Ultraviolet absorbance at 260 and 280 nm in RNA measurement is dependent on measurement solution. *Int. J. Mol. Med.* 5 (6), 657–659. doi:10.3892/ijmm.5.6.657

Passby, L. S. S., Brock, I., Wells, G., Cox, A., and Danson, S. (2019). Assessing melanoma BRAF status through ddPCR of cfDNA. *Ann. Oncol.* 30, vii15. vii15. doi:10.1093/annonc/mdz413.055

Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S., et al. (2012). Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal. Chem.* 84 (2), 1003–1011. doi:10.1021/ac202578x

Poh, T. Y., Ali, N., Chan, L. L. Y., Tiew, P. Y., and Chotirmall, S. H. (2020). Evaluation of droplet digital polymerase chain reaction (ddPCR) for the absolute quantification of Aspergillus species in the human airway. *Int. J. Mol. Sci.* 21 (9), 3043. doi:10.3390/ijms21093043

Qu, X., Dang, X., Wang, W., Li, Y., Xu, D., Shang, D., et al. (2018). Long noncoding RNAs and mRNA regulation in peripheral blood mononuclear cells of patients with chronic obstructive pulmonary disease. *Mediat. Inflamm.* 2018, 7501851. doi:10.1155/2018/7501851

Raeymaekers, L. (1993). Quantitative PCR: Theoretical considerations with practical implications. *Anal. Biochem.* 214 (2), 582–585. doi:10.1006/abio.1993.1542

Ribeiro-Silva, A., Zhang, H., and Jeffrey, S. S. (2007). RNA extraction from ten year old formalin-fixed paraffin-embedded breast cancer samples: A comparison of column purification and magnetic bead-based technologies. *BMC Mol. Biol.* 8, 118. doi:10.1186/1471-2199-8-118

Rudloff, U., Bhanot, U., Gerald, W., Klimstra, D. S., Jarnagin, W. R., Brennan, M. F., et al. (2010). Biobanking of human pancreas cancer tissue: Impact of *ex-vivo* procurement times on RNA quality. *Ann. Surg. Oncol.* 17 (8), 2229–2236. doi:10.1245/s10434-010-0959-6

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., et al. (2003). TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34 (2), 374–378. doi:10.2144/03342mt01

Sahin, U., Kariko, K., and Tureci, O. (2014). mRNA-based therapeutics--developing a new class of drugs. *Nat. Rev. Drug Discov.* 13 (10), 759–780. doi:10.1038/nrd4278

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., et al. (2006). The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* 7, 3. doi:10.1186/1471-2199-7-3

Showe, M. K., Vachani, A., Kossenkov, A. V., Yousef, M., Nichols, C., Nikonova, E. V., et al. (2009). Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. *Cancer Res.* 69 (24), 9202–9210. doi:10.1158/0008-5472.CAN-09-1378

Smith, A. K., Conneely, K. N., Pace, T. W., Mister, D., Felger, J. C., Kilaru, V., et al. (2014). Epigenetic changes associated with inflammation in breast cancer patients treated with chemotherapy. *Brain Behav. Immun.* 38, 227–236. doi:10.1016/j.bbi.2014.02.010

Song, J., Kim, D., Han, J., Kim, Y., Lee, M., and Jin, E. J. (2015). PBMC and exosome-derived Hotair is a critical regulator and potent marker for rheumatoid arthritis. *Clin. Exp. Med.* 15 (1), 121–126. doi:10.1007/s10238-013-0271-4

Stankiewicz, W., Dabrowski, M. P., Chcialowski, A., and Plusa, T. (2002). Cellular and cytokine immunoregulation in patients with chronic obstructive pulmonary disease and bronchial asthma. *Mediat. Inflamm.* 11 (5), 307–312. doi:10.1080/0962935021000015719

Sui, W., Yan, Q., Li, H., Liu, J., Chen, J., Li, L., et al. (2013). Genome-wide analysis of long noncoding RNA expression in peripheral blood mononuclear cells of uremia patients. *J. Nephrol.* 26 (4), 731–738. doi:10.5301/jn.5000201

Suzuki, E., Sugimoto, M., Kawaguchi, K., Pu, F., Uozumi, R., Yamaguchi, A., et al. (2019). Gene expression profile of peripheral blood mononuclear cells may contribute to the identification and immunological classification of breast cancer patients. *Breast Cancer* 26 (3), 282–289. doi:10.1007/s12282-018-0920-2

Talmadge, J. E., and Gabrilovich, D. I. (2013). History of myeloid-derived suppressor cells. *Nat. Rev. Cancer* 13 (10), 739–752. doi:10.1038/nrc3581

Tang, R., She, Q., Lu, Y., Yin, R., Zhu, P., Zhu, L., et al. (2019). Quality control of RNA extracted from PAXgene blood RNA tubes after different storage periods. *Biopreserv. Biobank.* 17 (5), 477–482. doi:10.1089/bio.2019.0029

Valero, C., Zanoni, D. K., McGill, M. R., Ganly, I., Morris, L. G. T., Quer, M., et al. (2020). Pretreatment peripheral blood leukocytes are independent predictors of survival in oral cavity cancer. *Cancer* 126 (5), 994–1003. doi:10.1002/cncr.32591

van der Poel-van de Luytgaarde, S. C., Geertsma-Kleinekoort, W. M., Goudswaard, C. S., Hogenbirk-Hupkes, P. E., van Hoven-Beijen, M. A., van de Werf, M., et al. (2013). Addition of beta-mercaptoethanol is a prerequisite for high-quality RNA isolation using QIAsymphony technology as demonstrated by detection of molecular aberrations in hematologic malignancies. *Genet. Test. Mol. Biomarkers* 17 (6), 475–480. doi:10.1089/gtmb.2012.0448

Volovat, S. R., Volovat, C., Hordila, I., Hordila, D. A., Mirestean, C. C., Miron, O. T., et al. (2020). MiRNA and LncRNA as potential biomarkers in triple-negative breast cancer: A review. *Front. Oncol.* 10, 526850. doi:10.3389/fonc.2020.526850

Vorup-Jensen, T., Carman, C. V., Shimaoka, M., Schuck, P., Svitel, J., and Springer, T. A. (2005). Exposure of acidic residues as a danger signal for recognition of fibrinogen and other macromolecules by integrin alphaXbeta2. *Proc. Natl. Acad. Sci. U. S. A.* 102 (5), 1614–1619. doi:10.1073/pnas.0409057102

Wahlberg, K., Huggett, J., Sanders, R., Whale, A. S., Bushell, C., Elaswarapu, R., et al. (2012). Quality assessment of biobanked nucleic acid extracts for downstream molecular analysis. *Biopreserv. Biobank.* 10 (3), 266–275. doi:10.1089/bio.2012.0004

Wang, X., Zhu, H., Snieder, H., Su, S., Munn, D., Harshfield, G., et al. (2010). Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC Med.* 8, 87. doi:10.1186/1741-7015-8-87

Wang, Y., Zheng, H., Chen, J., Zhong, X., Wang, Y., Wang, Z., et al. (2015). The impact of different preservation conditions and freezing-thawing cycles on quality of RNA, DNA, and proteins in cancer tissue. *Biopreserv. Biobank.* 13 (5), 335–347. doi:10.1089/bio.2015.0029

Webster, A. F., Zumbo, P., Fostel, J., Gandara, J., Hester, S. D., Recio, L., et al. (2015). Mining the archives: A cross-platform analysis of gene expression profiles in archival formalin-fixed paraffin-embedded tissues. *Toxicol. Sci.* 148 (2), 460–472. doi:10.1093/toxsci/kfv195

Zhang, X., Han, Q. Y., Zhao, Z. S., Zhang, J. G., Zhou, W. J., and Lin, A. (2019). Biobanking of fresh-frozen gastric cancer tissues: Impact of long-term storage and clinicopathological variables on RNA quality. *Biopreserv. Biobank.* 17 (1), 58–63. doi:10.1089/bio.2018.0038

Zychowska, M., Grzybkowska, A., Wiech, M., Urbanski, R., Pilch, W., Piotrowska, A., et al. (2020). Exercise training and vitamin C supplementation affects ferritin mRNA in leukocytes without affecting prooxidative/antioxidative balance in elderly women. *Int. J. Mol. Sci.* 21 (18), 6469. doi:10.3390/ijms21186469

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 16 May 2023
DOI 10.3389/fgene.2023.1150212

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Hari Krishna Yalamanchili,
Baylor College of Medicine, United States
Bruce McKay,
Carleton University, Canada

*CORRESPONDENCE
Rahul N. Kanadia,
rahul.kanadia@uconn.edu

RECEIVED 23 January 2023
ACCEPTED 28 April 2023
PUBLISHED 16 May 2023

CITATION
Girardini KN, Olthof AM and Kanadia RN (2023), Introns: the "dark matter" of the eukaryotic genome.
Front. Genet. 14:1150212.
doi: 10.3389/fgene.2023.1150212

Introns: the "dark matter" of the eukaryotic genome

Kaitlin N. Girardini[1], Anouk M. Olthof[1,2] and Rahul N. Kanadia[1,3]*

[1]Physiology and Neurobiology Department, University of Connecticut, Storrs, CT, United States, [2]Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark, [3]Institute for Systems Genomics, University of Connecticut, Storrs, CT, United States

The emergence of introns was a significant evolutionary leap that is a major distinguishing feature between prokaryotic and eukaryotic genomes. While historically introns were regarded merely as the sequences that are removed to produce spliced transcripts encoding functional products, increasingly data suggests that introns play important roles in the regulation of gene expression. Here, we use an intron-centric lens to review the role of introns in eukaryotic gene expression. First, we focus on intron architecture and how it may influence mechanisms of splicing. Second, we focus on the implications of spliceosomal snRNAs and their variants on intron splicing. Finally, we discuss how the presence of introns and the need to splice them influences transcription regulation. Despite the abundance of introns in the eukaryotic genome and their emerging role regulating gene expression, a lot remains unexplored. Therefore, here we refer to introns as the "dark matter" of the eukaryotic genome and discuss some of the outstanding questions in the field.

KEYWORDS

intron, evolution, splicing, snRNA, spliceosome, eukaryotes, gene expression

Introduction

Historically, introns were considered the non-coding, non-functional sequence elements which disrupt those that are protein-coding, called exons (Gilbert, 1978). While this protein-centric definition of introns (Figure 1, left) has served its purpose, their presence in long non-coding RNA reveals that introns are not specific to protein-coding genes but instead serve a broader role in eukaryotic gene expression (Krchňáková et al., 2019; Abou Alezz et al., 2020). Moreover, introns have been found to host other lariat-derived RNAs, including microRNAs, long noncoding RNAs, small nucleolar RNAs, small nuclear RNAs, and circular RNAs that are crucial for gene regulation (Liu and Maxwell, 1990; Hesselberth, 2013; Seal et al., 2020; Kumari et al., 2022; Vakirlis et al., 2022). Introns can also house enhancer elements that drive tissue-specific expression kinetics during complex vertebrate development and embryogenesis (Emera et al., 2016; Blankvoort et al., 2018; Meng et al., 2021; Shiau et al., 2022). These intervening sequences necessitated co-evolution of splicing machinery to facilitate production of a contiguous transcript capable of encoding a functional unit (Grabowski et al., 1985; Nilsen, 2003). Inhibition of splicing results in retention of introns in the mature transcript, which often disrupts the open reading frame and ultimately dictates the fate of the final transcript (Kaida et al., 2007; Effenberger et al., 2017; Olthof et al., 2021). Since the discovery of splicing, introns have been extensively investigated and the significance of splicing in regulating gene expression is well documented (Singh and Padgett, 2009; Tellier et al., 2020; Zhu et al., 2020; Agirre et al., 2021; Reimer et al., 2021). Taken together, the presence of introns has a significant impact on eukaryotic gene

expression and underpins many of the complexities required to build higher eukaryotes. Therefore, here we present an intron-centric perspective (Figure 1, right) towards understanding regulation of eukaryotic gene expression.

Function and evolution of intronic elements

Introns date back to the last eukaryotic common ancestor, after invasion into the early eukaryotic genome (Russell et al., 2006; Carmel et al., 2007; Csuros et al., 2011). While an endogenous model has been proposed to explain the emergence of introns (Catania et al., 2009), there is a general consensus that prokaryotic group II self-splicing introns underwent invasion and mutational degeneration during early eukaryogenesis, resulting in inert introns and trans-acting splicing machinery (Michel et al., 1989; Sharp, 1991; Sontheimer et al., 1999; Shukla and Padgett, 2002). As the origin of eukaryotic introns has been extensively described (Koonin, 2006; Rogozin et al., 2012; Vosseberg and Snel, 2017; Baumgartner et al., 2019; Smathers and Robart, 2019), here, we focus on the continued maintenance and diversification of introns in eukaryotic genomes.

Prokaryotic group II self-splicing introns behaved largely as transposable elements, which may have facilitated their invasion of the eukaryotic genome (Figure 2) (Lambowitz and Zimmerly, 2011). Initially characterized in the maize genome, transposable elements are repetitive sequences found across eukaryotes and are critically known for their ability to relocate in the genome and alter gene expression (McClintock, 1950; SanMiguel et al., 1996; Elliott et al., 2005; Wells and Feschotte, 2020). Short and long interspersed retro-transposable elements (SINEs/LINEs) belong to the non-long terminal repeat class of elements which have retained transposable activity and are highly represented in the human genome as Alu and L1 elements, respectively (Kazazian and Moran, 1998; Lander et al., 2001; Balachandran et al., 2022). When carrying splice sites, these transposable elements can create novel exon/intron boundaries, which hold the potential to alter expression of that gene (Figure 2); a detailed description of exon/intron boundaries and their recognition by splicing machinery is discussed in the following sections. For example, a recent study queried pathogenic mutations that were associated with novel intron-exon boundaries in humans and identified those which aligned with transposable elements. They found that clusters of transposable elements are more liable to exonization, likely due to the combined effort of LTR and Alu elements in potentiating all necessary splice sites (Alvarez et al., 2021). In another computational investigation of the human genome, mutagenesis of Alu elements into weak splice sites was found to be well-tolerated if not retained long-term and was often associated with exon skipping events (Sorek et al., 2002). Exon skipping is a frequently observed form of alternative splicing, which more broadly serves as an important regulatory node for gene expression in developing systems (Baralle and Giudice, 2017). One can then speculate that Alu elements in this manner allow for transient sampling of novel functions of proteins encoded by these alternatively spliced transcripts. This idea is an extension of the already known role of Alu elements in tissue-specific transcription regulation (Franchini et al., 2011). Notably, weak splice sites in Alu elements can eventually become constitutively spliced exons, losing their capacity for transposition and become exons used in regulating tissue-specific gene expression, as is observed in the human *NARF* gene (Lev-Maor et al., 2007).

Inherent to the jumping nature of transposable elements is the impartiality of transposon landing. Transposon insertion would likely be deleterious in the protein-coding region of a gene, leading to evolutionary selection against that gene configuration. However, in a heterozygote, transposon-induced activation of a novel splice site within an intron could allow for a low-cost trial of differentially spliced isoforms, while still maintaining a functionally expressed copy. A susceptibility of spliceosomal introns to genomic recombination was demonstrated in two *Saccharomyces cerevisiae* genes, RPL8B and ADH2. Truncated versions of these genes were used in a splicing reporter construct, such that the second exon was expressed in frame with a fused EGFP cassette. Additionally, each construct carried an embedded *S. pombe his5*$^{+}$ gene within the first intron, encoded for in the opposite direction as EGFP. Here, the *his5*$^{+}$ gene contains an artificial intron lacking a catalytic branch point, and containing splice sites in such an orientation that they are only capable of splicing from the EGFP transcript. Thus, splicing of the artificial intron followed by transposition of the EGFP intron into the genomic loci was required to confer a positive result (Lee and Stevens, 2016). Meanwhile, Gozashti et al. (2022), has attributed rapid, lineage-specific intron gains to Intron elements derived from transposable elements. Through analysis of 1,700 species, these “intron-generating transposable elements families” were identified in approximately 5% of genomes and significantly overrepresented in aquatic lineages. Based on statistical association models and a consideration of likely propagation mechanisms, they concluded that Intron elements may facilitate recent intron gain, particularly through horizontal gene transfer in aquatic lineages. The activity of Intron elements is particularly interesting, as mechanisms of Introns in *Micromonas pusilla* and *Aureococcus anophagefferens* exhibit seemingly preferential insertion between pre-existing nucleosomes (Huff et al., 2016). The rationale here is such that the linker sequence between nucleosomes is often open and available for insertion events. Further support for this idea is seen in the unequal distribution and position of nucleosomes observed between protein-coding exons, pseudo exons, and introns in human and *Caenorhabditis elegans* (Andersson et al., 2009). Using transcriptomic and genomic sequencing data, Huff et al. (2016), reported that Introns are largely capable of co-opting splice sites and inserting by DNA transposition in both orientations, though with biases consistent with species-specific patterns in genome organization. Outside of splice site generation, transposons have also been implicated in regulation of splicing-competent snRNAs, such as those L1 transposons which are associated with formation of U6 pseudogenic snRNA (Doucet et al., 2015). Pseudogenes can encode variations of spliceosomal snRNAs, the implications of which are discussed further below. In all, transposable elements further expand gene structure by modifying intronic elements, thus revealing a critical role of non-coding intronic elements in eukaryotic genome evolution.

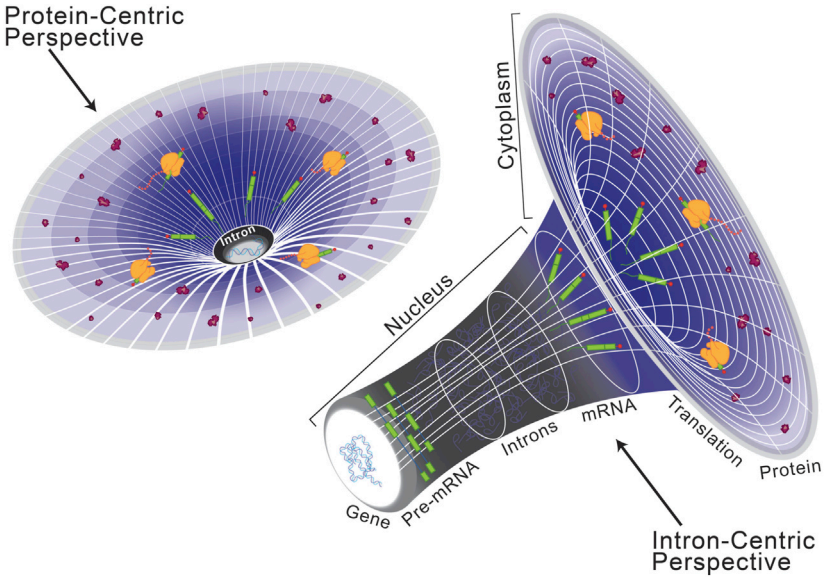


FIGURE 1
Schematization of a protein-centric versus intron-centric perspective on gene expression. Here, we model the role of introns in the genome after that of dark matter in astronomy, as both are difficult to characterize but critical organizing principles. From a protein-centric perspective (left), whereby the transcriptome and genome are interpreted in reference to a protein-coding sequence, it is easy to oversee the role of introns in eukaryotic gene expression. However, as depicted on the right, when the same model is viewed from an intron-centric perspective it becomes clear to see a regulatory mechanism by which introns are critical for expression of the eukaryotic genome.

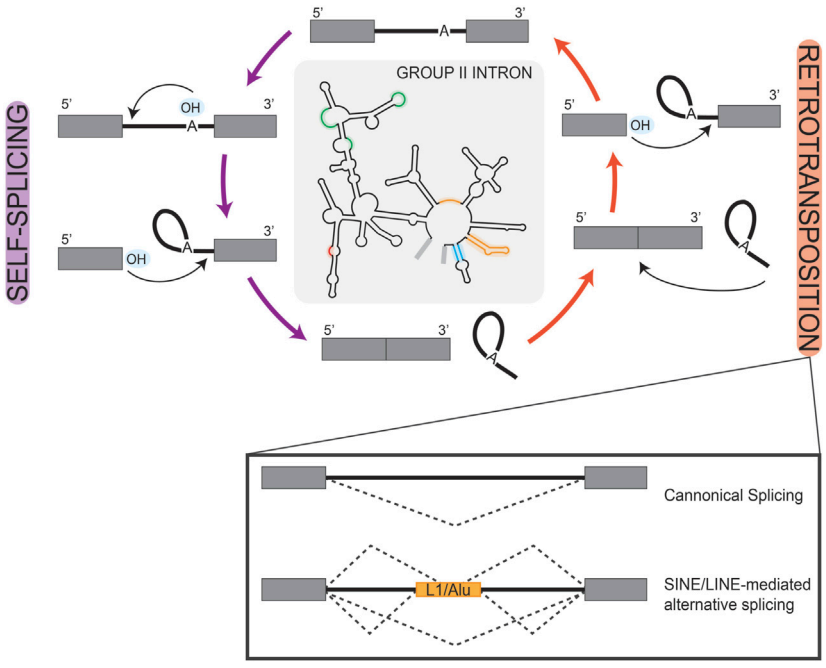


FIGURE 2
Retrotransposition of Introns. Simplified schematic of the reciprocal self-splicing (left, purple arrows) and retro-transposition (right, orange arrows) mechanisms that underlie the processing and mobility of group II self-splicing prokaryotic introns. These mechanisms are depicted as cyclic to highlight the parallel reactions that underlay each process. In the center, we show a group II self-splicing intron, with highlighted regions to represent loci that are analogous to eukaryotic snRNAs. In the box inset under retrotransposition, we show splicing schematics depicting the consequences of transposon-mediated alternative splicing in a eukaryotic gene. Here, boxes are used to represent exons and solid lines represent introns; splice patterns are represented by dashed lines.

Classification and splicing of introns

After the discovery of splicing, identified introns appeared to show a pattern of conserved terminal di-nucleotides at the exon-intron and intron-exon boundaries, and this feature became a defining characteristic of spliced introns (Breathnach et al., 1978; Crick, 1979; Breathnach and Chambon, 1981). As sequencing techniques have progressed and data now includes more diverse eukaryotic genomes, it is increasingly clear that introns are defined by several extended consensus sequences. These include the 5′ splice site (5′SS), the branch point sequence (BPS), and the 3′ splice site (3′SS) (Dietrich et al., 1997; Mercer et al., 2015). Not long after their discovery, it was determined that most introns are processed by five Uridylyl-rich snRNAs—U1, U2, U4, U5, and U6—that are highly conserved between eukaryotes and assemble into a ribonucleoprotein complex, the spliceosome (Bringmann et al., 1983; 1984; Bringmann and Lührmann, 1986; Nilsen, 2003; Wahl et al., 2009). Specifically, U1 snRNA has complementarity at the 5′ splice site, marking the exon-intron boundary, while U2 snRNA base pairs around a conserved adenosine toward the 3′ end, at what has become known as the branch point sequence (Yan and Ares, 1996; Malca et al., 2003). The direct base pairing of these snRNAs with splice site consensus sequences helps to recognize and remodel the intron during splicing, conferring the core function of the spliceosome.

As this mechanism was coming into focus, Jackson (1991), discovered spliced transcripts, that when mapped to the genome, showed intronic splice site sequences that were incompatible with the identified snRNAs. The fact that these introns were nonetheless spliced suggested the existence of a separate mechanism for their removal. This discordant finding led to sequence-based investigations for U snRNAs with complementary to non-consensus splice sites. This included an exploratory genomics investigation by Hall and Padgett (1994), and ultimately led to the hypothesis that newly identified U11 and U12 snRNAs serve in roles analogous to U1 and U2 during splicing (Montzka and Steitz, 1988). A role for U11 and U12 was confirmed *in vitro* (Tarn and Steitz, 1996a) and *in vivo* (Hall and Padgett, 1996; Kolossova and Padgett, 1997), and bolstered by the additional identification of snRNAs analogous to U4/U6, U4atac and U6atac (Tarn and Steitz, 1996b; Incorvaia and Padgett, 1998). Based on their relative abundance in analyzed genomes, the intron types and their respective spliceosomes were henceforward labeled major (U2-type) and minor (U12-type) in those eukaryotes that maintain them in parallel (Burge et al., 1998; Lynch and Richardson, 2002; Lin et al., 2010). Of note, major introns and the major spliceosome are ubiquitous in the eukaryotic lineage, while minor introns and the minor spliceosome are reportedly absent in some lineages, such as *Caenorhabditis elegans* (Burge et al., 1998).

Both the major and minor spliceosomes employ U5 snRNA, and each snRNA further associates with specific proteins in their splicing-competent forms (Tarn and Steitz, 1996a; Tarn and Steitz, 1997). Though the individual snRNAs have specific proteins associated with their regulation and maturation, many of the remaining proteins that comprise the spliceosome are shared between both the major and minor molecular machineries (Will et al., 1999); for a more comprehensive presentation of individual spliceosome components, see Olthof et al. (2022). Worth noting, the same protein can carry out different roles in each spliceosome, as is observed by URP (also called ZRSR2) (Tronchère et al., 1997; Shen et al., 2010). While the size and dynamic composition of the spliceosome can make it difficult to fully resolve, identifying the proteins involved in splicing regulation remains an area of active investigation. Recent biochemical and cryogenic electron microscopy investigations to this end have significantly enhanced our understanding of minor spliceosome-specific proteins. For example, the protein compositions of U4.U6/U5 and U4atac.U6atac/U5 tri-snRNP complexes were previously thought to be identical. However, co-immunoprecipitation and co-migration analyses have suggested that CENATAC may aid in 5′SS recognition for a subclass of minor introns characterized by AT-AN terminal di-nucleotides. Previously known as *CCDC84*, *CENATAC* was renamed following its mutagenic link to intron retention in human genes that contribute to chromosome stability and segregation (de Wolf et al., 2021). Interestingly, phylogenetic profiling of CENATAC across 90 eukaryotic species showed that it co-enriched with other components of the minor spliceosome, including the newly characterized SCNM1 protein (de Wolf et al., 2021). The U12 snRNA is flanked by the N-terminal C_2H_2 zinc fingers of SCNM1, which interacts with the U12/BPS duplex and the U12 Sm ring (Bai et al., 2021). The N-terminus of SCNM1 also functions to stabilize U6atac and RNF113A at the 5′SS, maintenance of which is required for spliceosome activation *in vivo* (Incorvaia and Padgett, 1998; Bai et al., 2021). Structural insights were also important in identifying the novel minor spliceosome protein, RBM48, which is now known to bind ARMC7 and interact with terminal ends of U6atac snRNA via conserved RNA binding residues (Bai et al., 2021; Siebert et al., 2022). Structural analyses of the minor spliceosome are a recent advancement and do not yet cover all phases of splicing, notably excluding the U11/U12 di-snRNP. As such, there remains the possibility for other unidentified components regulating the nuances of minor intron splicing.

A delineation between major versus minor intron splicing is often based on the quantitative analysis of splice site conservation, and thus relative splice site strength. Intron splice sites are generally scored based on the degree of similarity to the major versus minor intron consensus sequences found in Figure 3, using position weight matrices (Sheth et al., 2006; Alioto, 2007; Olthof et al., 2019; Moyer et al., 2020). The resultant major or minor intron classification inherently dictates how we interpret its processing, such that bioinformatically classified minor introns are predicted to be spliced by the minor spliceosome, and *vice versa*. However, RNA sequencing data has shown that, upon inhibition of the minor spliceosome, not all bioinformatically classified minor introns show a splicing defect (Olthof et al., 2019). Thus, parallel existence of major and minor spliceosomes, combined with diverging intron consensus sequences, reveal an added complexity in the relationship between a given intron and its recruited spliceosome. Akin to how the concept of a single intron type was disrupted by the discovery of minor introns; it seems increasingly likely that the binary classification or major versus minor itself is insufficient to fully resolve all introns. Rather, evidence has begun to suggest that the stringency of the classification schema fails to consider the fluidity of exons and introns. For example, use of novel splice sites within exonic regions in the unicellular Paramecium is evidence of intronization activity in

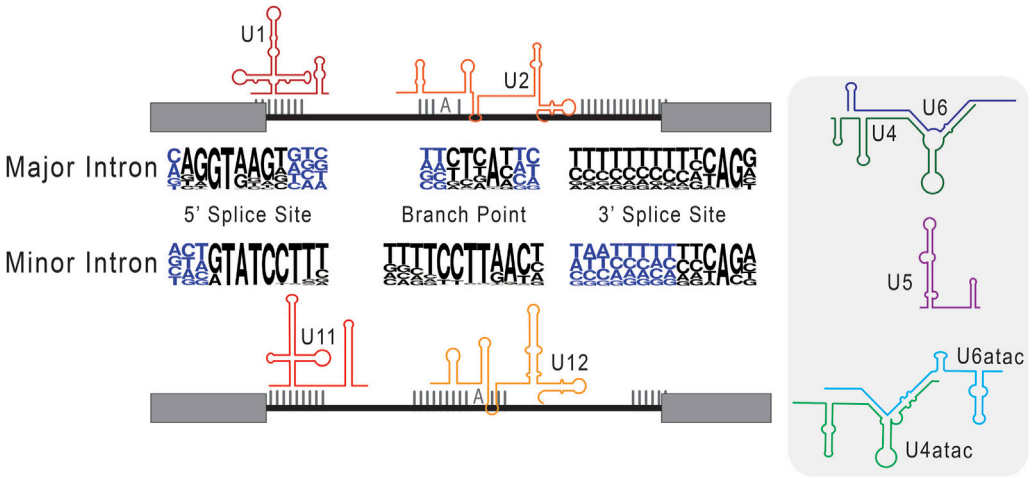


FIGURE 3
Consensus sequences used in the classification of major versus minor introns. Here, we schematize splice site selection by the respective components of the major and minor spliceosomes. The snRNAs of the major (U1 and U1) and minor (U11 and U12) spliceosome are shown base pairing to their cognate consensus sequences. In the center, next to the respective major intron and minor intron labels, we depict consensus sequences as nucleotide frequency plots. Here, the relative size of the nucleotide represents how frequently it is observed in that genomic position. Right of this schematic, we include the remaining core snRNAs that are unique to major (U4 and U6) and minor (U4atac and U6atac) intron splicing, as well as the shared U5 snRNA.

eukaryotes (Ryll et al., 2022). In essence, these findings increasingly suggest that the current approach to intron classification is too reductive to fully capture the complexities and dynamic regulation of eukaryotic introns. Towards this end, an examination of minor-type splice sites in Physarum polycephalum has suggested that minor introns may exist in divergent, if not degenerative, types (Larue et al., 2021) and this idea is currently being refined in other studies that combine principles of speciation and comparative genomics.

How gene architecture informs splice site selection

Spliceosomal introns are known to range from tens of base pairs in length to hundreds of kilobases in length, with a mean length that is smaller in lower eukaryotes and larger in higher eukaryotes (Sakharkar et al., 2004; Piovesan et al., 2015; Abebrese et al., 2017; Li et al., 2017; Jakt et al., 2022). The size of an intron has an inherent impact on gene expression, as it will take longer for transcription machinery to create nascent transcripts. In turn, this will impact the kinetics of co-transcriptional intron splicing; these ideas have been explored in depth (Herzel et al., 2017; Wallace and Beggs, 2017; Neugebauer, 2019). It is long since established that relative intron and exon lengths can differentially affect splicing efficiency due to a presence or absence of regulatory elements and differing requirements for catalysis (Fox-Walsh et al., 2005; Kandul and Noor, 2009; Pai et al., 2017). Splicing efficiency refers to the proportion of spliced versus un-spliced transcripts relative to the number of total transcripts. This is commonly assessed using computational strategies that characterize splice events in the transcriptome (de Melo Costa et al., 2021; Jiang et al., 2023), followed by a validation of observed changes in expression using techniques such as RT-PCR. In one assessment of how splicing efficiency and gene expression patterns may be coupled, intron length was found to contribute to the temporal coordination that is required for co-expression of genes with interdependent biochemical functions (Keane and Seoighe, 2016). This idea is further reflected by distinct differences in splice site strength relative to intron length, and by differences in splicing efficiency and mRNA abundance relative to gene length (Gelfman et al., 2011; Sánchez-Escabias et al., 2022). Vertebrates are known to increase splicing efficiency around longer introns via cell-specific recursive splicing and transposable elements that form stems with intronic RNA loops to juxtapose splice sites (Shepard et al., 2009; Zhang et al., 2018). For details on recursive splicing, please see published reviews (Georgomanolis et al., 2016; Gehring and Roignant, 2021; Joseph et al., 2022; Pitolli et al., 2022).

Separate from this, longer introns may also have a propensity to contain multiple splice sites within one intronic feature, leading to alternative splicing from competing splice site use (Sun and Chasin, 2000; Roca et al., 2003; Kapustin et al., 2011). Meaning it becomes increasingly likely that multiple splice sites be present, in addition to exonic splicing enhancers and silencer elements, which themselves can act as determinants of splice site usage (Black, 2003; Wang et al., 2006). It thus follows that the sequence content of the intron to be excised can drive splicing progression. Splice site selection is thought to occur by competing intron- and exon-definition models, which describe how the spliceosome assembles either through cross-bridging interactions across the intron itself or across the flanking exon. Specifically, the intron-definition model refers to the mechanism whereby 3′ SS selection is informed by recognition of the upstream 5′SS, such that the spliceosome assembles across the intron. For exon-definition interactions, 3′ SS recognition depends instead on recognition of the downstream 5′SS (Robberson et al., 1990; Berget, 1995; Romfo et al., 2000; De Conti et al., 2013; Olthof

et al., 2021). For example, most genes in *Saccharomyces cerevisiae,* contain only one, short intron. With this gene architecture, it is not surprising that intron-definition interactions predominate. Surprisingly, cryo-electron microscopy structures of the pre-catalytic spliceosome demonstrated that the same splicing machinery can perform exon-definition interactions in multi-intronic genes (Li et al., 2019). This finding brings to bear uncertainty as to how and when an intron- versus exon-centric model is utilized. This becomes especially important in higher vertebrates which have a larger intronic burden.

Reconciliation between the intron- and exon-definition models is coupled with new insight into how proximity rules inform splice site selection. Based on the length of an intron, the intron-centric proximity rule dictates a preference for the spliceosome to assemble over a splice site pair that minimizes the distance between 5′ and 3′ end selection (Reed and Maniatis, 1986). More recently, computational analyses by Carranza et al. (2022) refined the exon-centric proximity rule, by which splice sites are selected to minimize the exon-spanning distance. Meaning if one were to imagine an intron with two adjacent sets of 5′ and 3′ splice sites, the intron-centric proximity rule would employ the innermost splice site pair, maximizing the resultant exons. Meanwhile, the exon-centric proximity rule would, in contrast, use the exon-proximal splice sites to maximize the size of the intron being excised. In either case, commitment to the intron-centric or exon-centric proximity rule has commensurate intronization/exonization consequences as molecular machinery decides whether to select for the smaller or larger exonic sequences. In addition to intron size, studies suggest that GC content of the intron may also be a determinative factor in the mechanism employed for splice site selection. In one study, (Tammer et al. (2022), examined the nucleotide composition of exons versus introns and subsequently identified genes they refer to as "differential" or "leveled". In "leveled" genes, GC content is found to be similarly high in exons and introns, while "differential" genes are ones wherein GC content is low in exons, and even lower in introns. Notably, Tammer et al. (2022), describe a partiality for intron-definition interactions across "leveled" genes, while exon-definition interactions predominate over "differential" genes. This finding is in line with previously reported links between differential GC content and splice site selection (Amit et al., 2012).

Spliceosomal sRNAs

As described above, snRNAs confer the primary function of the spliceosome through formation of specific base pair interactions with consensus sequences in the intron. The presence and function of snRNAs is essential for recognition and restructuring of the nascent mRNA transcript in the sequential, exothermic transesterification reactions that constitute splicing.

In general, most snRNAs (U1, U2, U4, U5, U11, and U12) are transcribed by RNA polymerase II, while U6 and U6atac expression are largely dependent on RNA polymerase III (Reddy et al., 1987; Jawdekar and Henry, 2008; Younis et al., 2013). Initiation of transcription of these snRNAs is highly reliant on the proximal and distal sequence elements located upstream of the snRNA-encoding region. Specifically, because they serve as promoter and enhancer elements for recruitment of transcription machinery through interactions with the SNAPc transcription factor complex and stabilizing co-activators (Sadowski et al., 1993; Henry et al., 1998; Mittal et al., 1999; Dergai et al., 2018). Structural insights by cryogenic electron microscopy of SNAPc during the transcription of U6 has revealed the importance of conserved subunits which recognize and bind the proximal sequence element (Sun et al., 2022). One unique exception to this rule is for the expression of human U4atac snRNA, which is embedded into an intron of *CLASP1* (Edery et al., 2011). Therefore, U4atac expression relies on RNA polymerase II mediated transcription of this gene, as well as successful splicing of this intron.

Within the genome, spliceosomal snRNAs often exist both as gene copies and gene families, whereby divergent genes can encode for variant snRNAs with nucleotide polymorphisms (Denison et al., 1981; Abel et al., 1989). There are both productive and unproductive variants of the snRNAs annotated; productive snRNAs are capable of splicing, while those that are not are termed pseudogenic (Mabin et al., 2021). For example, the U6 snRNA has many pseudogenes and fewer productive copies that are dispersed throughout the genome, whereas U1 and U2 snRNAs are encoded by many functional copies that are organized in homogenous repeats (Van Arsdell and Weiner, 1984; Theissen et al., 1985; Tichelaar et al., 1998; Domitrovich and Kunkel, 2003; O'Reilly et al., 2013; Anjos et al., 2015). The presence of multiple gene copies may in part explain the splicing-independent roles of U1 and U2 in regulating transcription termination and 3′ end processing (Friend et al., 2007; Di et al., 2019; So et al., 2019). Moreover, the idea that multiple gene copies exist for minor spliceosomal snRNAs, including U4atac and U6atac, warrants further investigation. Even if multiple gene copies do exist, it must be noted that U6atac expression is maintained at a lower level through rapid post-transcriptional turnover (Younis et al., 2013).

Perhaps counterintuitively, U5 snRNA has the smallest gene family, yet it is the only shared snRNA between the major and minor spliceosomes. Investigations by Mabin et al. (2021) into the relevance of snRNA variants in splicing led to the discovery of high sequence identity between U5 variants. In fact, they report several U5 variants with a conserved stem consensus sequence (CUUUU) that can be incorporated into catalytic spliceosomes. Based on these observations, it has been suggested that U5 may not have a canonical snRNA; rather, specific variants may be optimal for use in one spliceosome type over the other (Mabin et al., 2021). While mechanistically unvalidated, this logic is consistent with the analogous nature of the other major versus minor snRNAs. Yet, it also remains possible that these U5 variants are regulated in a context-dependent way, as is observed for U1 snRNA variants during human stem cell programming (Vazquez-Arango et al., 2016). Additionally, U5 snRNA variants have been identified in regulating development in humans, *Drosophila,* and *Lytechinus variegatus* (Sontheimer and Steitz, 1992; Morales et al., 1997; Chen et al., 2005). The expression of snRNA variants to specify a differentiating transcriptome is not unique to U5 snRNA, but more broadly detected for other snRNAs and across species (Lo and Mount, 1990; Cáceres et al., 1992; Sierra-Montes et al., 2005; O'Reilly et al., 2013; Lu and Matera, 2014).

Functional sequence variants of the snRNAs have the potential to contact cryptic or degenerating splice sites, make novel protein

interactions, and adopt secondary structures that alter spliceosome conformation. It is thus possible, given our evolving understanding of consensus sequences, that these variant snRNAs do confer complementarity to specific intron splice sites. Accordingly, from an intron-centric perspective, we must allow for the possibility that seemingly unproductive snRNAs are leveraged to splice a specific subset of introns. A role for non-consensus intron classes was voiced by Hudson et al. (2019), whose bioinformatics analyses of diplomonad and parabasalid lineage eukaryotes revealed splice site sequences that diverged from both the major and minor consensus sequences. They similarly identify divergent snRNAs, though they maintained key functional structures including stem loops and putative Sm binding sites. Perhaps more compelling, the discovered snRNAs showed aggregate features of both the major- and minor-type snRNAs, suggesting a propensity for the spliceosome to adopt complementarity to *trans*-spliced introns.

It remains to be established if variant snRNAs are evolutionarily selected for use in differential splicing or if they arise stochastically. Though, one could imagine that selective use of a variant splicing component would provide an opportunity to splice novel or divergent splice site sequences. It is known that mutations in the snRNAs can have pathogenic effects, as demonstrated by *RNU12* which is causal to early onset cerebellar ataxia (Elsaid et al., 2017). Additionally, snRNA secondary structure is important for splicing as it dictates the RNA-protein interactions necessary for spliceosome assembly. For example, U11/U12-65K binds the 3′ stem loop II (SLII) of U12 snRNA based on the integrity of this structure and its RNA binding motif. Further, 3' truncation mutants that disrupt the U12 SLIII are targeted for degradation by the nuclear exosome targeting complex upon reimport to the nucleus (Norppa and Frilander, 2021). In another example, the U2/U6 and U12/U6atac complexes are remodeled and stabilized prior to the first catalytic step in splicing by intramolecular base pairing with RBM22 (Ciavarella et al., 2020). Regardless, developmentally regulated snRNA variants demonstrate that mutations outside of critical structures may maintain, albeit differential, functionality. Thus, it stands to reason that variant snRNAs without disease-causing consequences to splicing may have a context-dependent role in the regulation of introns with divergent consensus sequences.

The evolutionary advantages of introns

Introns have served a valuable evolutionary role for eukaryotes in that they are more prone to genetic drift compared to exons. Introns appear to be under weaker selection than exons in somatic cells, which may be due to a mismatch repair system employed for exons that is notably lacking for introns (Hoffman and Birney, 2006; Resch et al., 2007; Frigola et al., 2017; Rodriguez-Galindo et al., 2020). Using a combinatorial multi-omics approach, Huang et al. (2018), has attributed the selective protection and mismatch repair of actively transcribed genes to an enrichment of H3K36me markers, which help regulate molecular responses to DNA damages induced by prolonged euchromatic conformation. More broad analyses of the differentiating human methylome reveal distinct differences in methylation pattern between genomics features, such that methylation is generally more common to exons than introns (Laurent et al., 2010). This unequal distribution may explain the higher frequency of mismatch repair observed for exons versus introns. In this capacity, introns can essentially act as a sponge to harbor mutations that would be otherwise detrimental in exonic sequences. Nevertheless, many mutations in intronic elements are linked to diseases, suggesting that there are limits to the number of mutations an intron can withstand. Mutations at splice sites and within introns are known to underscore an array of genetic and developmental disorders, including muscular dystrophy (Dominov et al., 2019) and inherited retinal diseases (Qian et al., 2021). Pathogenic disorders due to mutation of the spliceosome, i.e., spliceosomopathies, include but are not limited to craniofacial defects, myelodysplastic syndromes, and retinitis pigmentosa (Griffin and Saint-Jeannet, 2020). For review of major and minor splicing-associated diseases, see (Anna and Monika, 2018; Olthof et al., 2022).

While introns are seemingly advantageous, prokaryotes show that the absence of introns is not prohibitive to life. This begs the question, to what extent do eukaryotic cells really require introns? In one study, Parenteau et al. (2008), investigated the consequences of intron depletion in *Saccharomyces cerevisiae* (Figure 4A). Introns are far less abundant in *S. cerevisiae* compared to other species, such as vertebrates and land plants, making the yeast genome a strong model for intron depletion studies (Csuros et al., 2011). Indeed, *S. cerevisiae* could survive without introns, however, intron-depleted strains fared variably when subjected to drug-induced and carbon source stresses. However, transcription machinery was found to be capable of responding to expression deficits following intron-depletion by using alternative promoter selections, highlighting the role introns play in expanding the eukaryotic transcriptome (Parenteau et al., 2008). Should one suppose that introns can be leveraged to induce stress-related patterns of gene expression, it then follows that the splicing efficiency of an intron is responsive to stress application. This idea was recently explored by Frumkin et al. (2019), who employed YFP reporter constructs containing known introns with high and low splicing efficiencies embedded and fused to a kanamycin resistance cassette (Figure 4B). To test the capacity of introns and the spliceosome to respond to metabolic pressure, the constructs were expressed in *S. cerevisiae* cells under antibiotic selection and subjected to a lab-evolution paradigm. Growth and transcriptomic analyses of derived cell generations revealed independent, adaptive mutations occurring both cis- and trans-to improve splicing efficiency and thus antibiotic resistance and cell survivability. The cis-mutations were proposed to increase accessibility of splice site sequences, while trans-mutations might increase the cellular abundance of splicing machinery. Importantly, cis-fitness-inducing mutations could alleviate selection-independent splicing inefficiencies, however, mutations in trans-were particularly advantageous during periods of active selection (Frumkin et al., 2019). Though these experiments were performed in *S. cerevisiae*, one can imagine that similar mechanisms may be employed for evolutionary adaptation. For example, in ecotypic Cichlid fish, alternative splicing is a dominant mechanism for rapid changes in gene expression. Specifically, alternative splicing underpins the diversification of jaw morphology as it relates to the food they have evolved to consume in different ecological niches (Singh et al., 2017).

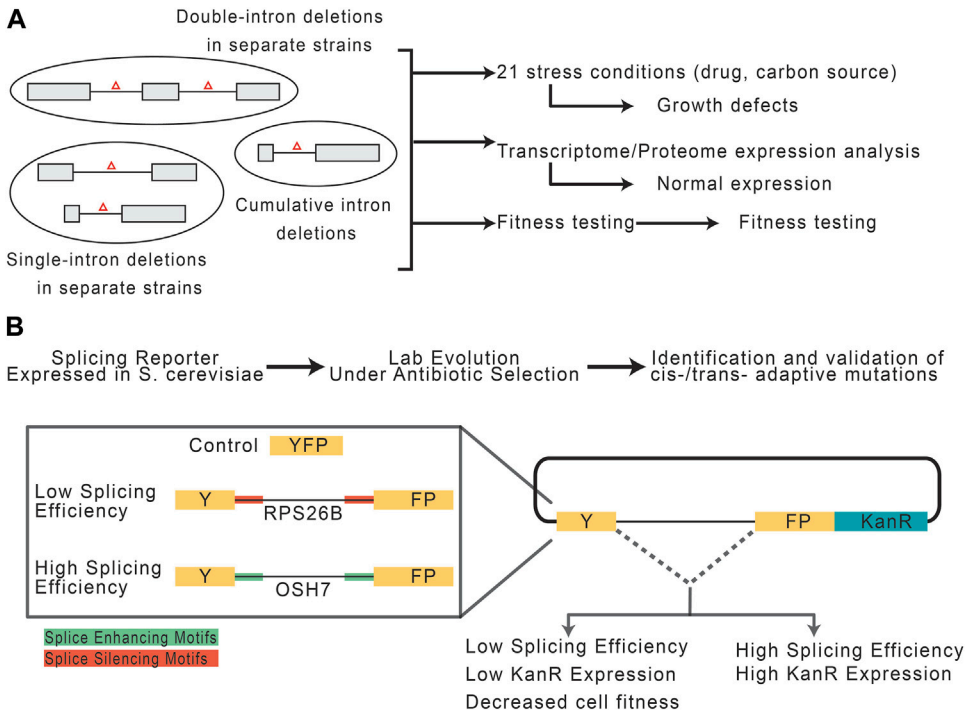


FIGURE 4
Potential role of introns and spliceosomal snRNAs in stress response. **(A)** Experimental paradigm, adapted from Parenteau et al. (2008), to assess the consequences of intron depletion in *S. cerevisiae*. Yeast with sets of removed intron(s) were grown under normal or stress conditions and assessed for fitness. **(B)** Experimental paradigm, adapted from Frumkin et al. (2019), to assess the capacity of introns and the splicing machinery to adapt to selective pressures.

The influence of introns on gene expression

In both mammals and plants, the presence of introns is known to enhance gene expression in a phenomenon sometimes referred to as intron-mediated enhancement (Brinster et al., 1988; Furger et al., 2002; Samadder et al., 2008). The recent development of sequencing techniques such as GRO-seq, mNET-seq and long read sequencing have revealed that splicing of neither major nor minor introns occurs in isolation, but rather in a highly active genomic context where splicing and transcription are coupled both kinetically and physically (Nojima et al., 2015; Herzel et al., 2017; Sheridan et al., 2019; Drexler et al., 2020; Reimer et al., 2021; Zhang et al., 2021). In the context of splicing informing transcription, the position of the intron matters, as promoter-proximal introns are especially known to enhance transcription (Furger et al., 2002; Rose et al., 2008). The knowledge that introns may enhance transcriptional output was leveraged to modify the generally used CMV promoter for expression plasmids, whereby introduction of an intron significantly upregulated transcription of downstream coding sequences (Simari et al., 1998).

The mechanism by which 5' introns regulate transcription involves, at least in part, control of the open chromatin signatures H3K4me3 and H3K9ac, which facilitate recruitment of RNA polymerase II and general transcription factors to promoters. These marks are deposited at the first exon-intron boundary of genes, explaining how the distance between transcription start site and the first intron can influence the expression level of a gene (Bieberstein et al., 2012, Lister, 2009). Interestingly, differential methylation patterns are not unique to protein-coding genes, as revealed through a bioinformatics model which considered the modified human nucleosome library and analysis of splicing efficiency. For example, high nucleosome density was observed in the internal exons of long non-coding RNAs, while high H3K4me3 signals were observed in upstream introns. Importantly, these signatures were often associated with exon skipping and intron retention, particularly around the first intron (Dey and Mattick, 2021). While a tissue-independent model likely obscures some of the nuanced features regulating splicing-dependent gene expression, a genome-wide comparative analysis by Anastasiadi et al. (2018) revealed that correlation between CpG methylation and gene expression is unique to the first exon and intron. As CpG markers of DNA methylation tend to decrease across exons and increase across introns, it is possible that methylation may inform gene expression by mediating intron splice site recognition (Laurent et al., 2010). In fact, removal of promoter-proximal introns altogether reduces levels of H3K4me3 and chromatin-bound RNA polymerase II, reducing transcriptional output (Bieberstein et al., 2012; Laxa et al., 2016). Similarly, reduction in chromatin accessibility was observed when formation of the active spliceosome was inhibited by spliceostatin A. This finding highlights an important role for the spliceosome in regulating transcriptional output. Notably, this effect was not intrinsic to the presence of introns, but dependent on their splicing (Bieberstein et al., 2012).

One caveat to the spliceostatin A experiment is that it inhibits the entire splicing machinery, without revealing the specific interactions between the spliceosome and intron consensus sequences that enhance transcription. In fact, it is not the entire spliceosome that needs to be activated for transcription enhancement, as the formation of stable interactions between U1 snRNA and the promoter-proximal 5′SS can enhance transcription (Engreitz et al., 2014). Recruitment of the U1 snRNP to the first intron enhances transcription initiation through recruitment of general transcription factors, such as TFIIH, and stabilization the first formed phosphodiester bond by RNA polymerase II (Kwek et al., 2002; Damgaard et al., 2008). Notably, this effect is independent of its role in major intron splicing, as mere introduction of a 5′SS sequence is sufficient to enhance transcription (Damgaard et al., 2008). This splicing-independent function of U1 might help explain its constitutive association with the elongating RNA polymerase II and why it is likewise recruited to intronless genes (Spiluttini et al., 2010; Leader et al., 2021).

Beside the role of U1 in transcription initiation, U1 snRNA is also independently involved in preventing pre-mature transcription termination, which can occur if RNA polymerase II encounters a polyadenylation site within an intron. Surmounting 3′ end sequencing data has revealed that introns often contain cryptic or pre-mature polyadenylation sites that result in the destabilization of RNA polymerase II, thereby producing truncated transcripts incapable of encoding a protein (Di Giammartino et al., 2011). Remarkably, the production of these truncated transcripts can be blocked by the U1 snRNA in a process called telescripting. In this capacity, U1 is capable of complexing with 3′ processing factors to protect the mRNA from premature cleavage and termination (Kaida et al., 2010; Berg et al., 2012). This mechanism occurs alongside the elongating polymerase to allow for U1-mediated suppression of cryptic polyadenylation sites in the intron or 3′ UTR (Di et al., 2019). Proper transcription termination is important in regulating the length and structure of the 3’ UTR, which in turn promotes formation of the export-competent messenger ribonucleoprotein. Similar to U1, U11 is expressed more highly than is necessary for its function in splicing (Baumgartner et al., 2015). Given that U11 is more abundant than U12 though they present at the same stoichiometric ratio within the minor spliceosome, U11 may similarly have splicing-independent functions. We speculate that U11 may either function in a mechanism like telescripting or participate in an alternative function, such as the subnuclear clustering of expressed minor intron-containing genes.

Localization of spliceosome components

Genes, chromatin, and RNA polymerase II have a subnuclear organization around topologically-associated domains to phase-separate euchromatic regions of active transcription (Szentirmay and Sawadogo, 2000; Ulianov et al., 2016; Szabo et al., 2020). Alongside this, it would be reasonable to hypothesize that splicing machinery is also organized to support efficient gene expression. In fact, major and minor spliceosome snRNPs display similar partiality for nuclear localization, except for U6 and U6atac snRNPs (Spiller et al., 2007; Pessa et al., 2008; Steitz et al., 2008). In the nucleus, matured snRNPs of the major spliceosome accumulate in phase-separated speckles that serve to organize spliceosome components adjacent to perichromatin regions of active transcription. This was concluded following nonradioactive and fluorescence *in situ* hybridization analyses, as well as RNA and protein blotting of subcellular compartment extracts (Pessa et al., 2008). While this model is an enticing way to interpret speckles as a regulatory mechanism over major intron splicing, it does not necessarily extend to that of minor introns. Given that the major and minor spliceosomes are known to interact with each other in the splicing of minor intron-containing genes, the model does not encompass all mechanisms of splicing (Akinyi and Frilander, 2021; Olthof et al., 2021). Punctate subcellular localization of spliceosome machinery is not specific to core snRNP components, but also includes some of the auxiliary splicing factors that contribute to spliceosome stability, conformational changes, and catalytic activity during splicing. These non-snRNP factors are integral to spliceosome assembly and the coordinated action of snRNPs during splicing (Bindereif and Green, 1990). For example, a new model supposes that the unequal phasic separation of SR proteins and heterogenous nuclear ribonucleoproteins proteins (hnRNP) at nuclear speckles can contribute to splice site selection. Specifically, the positional distribution of SR proteins and hnRNPs around a splice site generally determines the positive or negative regulatory effect of their binding, and taken with their distinct subnuclear distributions, can dictate the use of splice sites (Liao and Regev, 2021).

In all, here through an intron-centric lens, we focus our attention on the myriad of regulatory and functional consequences that have emerged by the presence of introns in the genome. Thus, we hope that future studies will begin to shed light on this “dark matter” of the eukaryotic genome to uncover the secrets buried within. Importantly, the advent of next-generation sequencing and computational analysis will invariably play a critical role in uncovering some of these mysteries. Throughout this article, we have described several of these methods, and here we point readers to other reviews (Halperin et al., 2021; Lorenzi et al., 2021; Gondane and Itkonen, 2023).

Author contributions

KG was responsible for curation of literature, organizing, writing text, and generating figures. AO was responsible for editing, structural organization, and help with literature curation. RK was responsible for the vision, writing, figures, and structure of the document. All authors contributed to the article and approved the submitted version.

Funding

Funding for this study comes from the National Institute of Neurological Disorders and Stroke (R01NS102538 to RK).

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

Abebrese, E. L., Ali, S. H., Arnold, Z. R., Andrews, V. M., Armstrong, K., Burns, L., et al. (2017). Identification of human short introns. *PLoS ONE* 12, e0175393. doi:10.1371/journal.pone.0175393

Abel, S., Kiss, T., and Solymosy, F. (1989). Molecular analysis of eight U1 RNA gene candidates from tomato that could potentially be transcribed into U1 RNA sequence variants differing from each other in similar regions of secondary structure. *Nucleic Acids Res.* 17, 6319–6337. doi:10.1093/nar/17.15.6319

Abou Alezz, M., Celli, L., Belotti, G., Lisa, A., and Bione, S. (2020). GC-AG introns features in long non-coding and protein-coding genes suggest their role in gene expression regulation. *Front. Genet.* 11. doi:10.3389/fgene.2020.00488

Agirre, E., Oldfield, A. J., Bellora, N., Segelle, A., and Luco, R. F. (2021). Splicing-associated chromatin signatures: A combinatorial and position-dependent role for histone marks in splicing definition. *Nat. Commun.* 12, 682. doi:10.1038/s41467-021-20979-x

Akinyi, M. V., and Frilander, M. J. (2021). At the intersection of major and minor spliceosomes: Crosstalk mechanisms and their impact on gene expression. *Front. Genet.* 12, 700744. doi:10.3389/fgene.2021.700744

Alioto, T. S. (2007). U12DB: A database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res.* 35, D110–D115. doi:10.1093/nar/gkl796

Alvarez, M. E. V., Chivers, M., Borovska, I., Monger, S., Giannoulatou, E., Kralovicova, J., et al. (2021). Transposon clusters as substrates for aberrant splice-site activation. *RNA Biol.* 18, 354–367. doi:10.1080/15476286.2020.1805909

Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., et al. (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* 1, 543–556. doi:10.1016/j.celrep.2012.03.013

Anastasiadi, D., Esteve-Codina, A., and Piferrer, F. (2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* 11, 37. doi:10.1186/s13072-018-0205-1

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., and Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* 19, 1732–1741. doi:10.1101/gr.092353.109

Anjos, A., Ruiz-Ruano, F. J., Camacho, J. P. M., Loreto, V., Cabrero, J., de Souza, M. J., et al. (2015). U1 snDNA clusters in grasshoppers: Chromosomal dynamics and genomic organization. *Heredity* 114, 207–219. doi:10.1038/hdy.2014.87

Anna, A., and Monika, G. (2018). Splicing mutations in human genetic disorders: Examples, detection, and confirmation. *J. Appl. Genet.* 59, 253–268. doi:10.1007/s13353-018-0444-7

Bai, R., Wan, R., Wang, L., Xu, K., Zhang, Q., Lei, J., et al. (2021). Structure of the activated human minor spliceosome. *Science* 371, eabg0879. doi:10.1126/science.abg0879

Balachandran, P., Walawalkar, I. A., Flores, J. I., Dayton, J. N., Audano, P. A., and Beck, C. R. (2022). Transposable element-mediated rearrangements are prevalent in human genomes. *Nat. Commun.* 13, 7115. doi:10.1038/s41467-022-34810-8

Baralle, F. E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi:10.1038/nrm.2017.27

Baumgartner, M., Drake, K., and Kanadia, R. N. (2019). An integrated model of minor intron emergence and conservation. *Front. Genet.* 10, 1113. doi:10.3389/fgene.2019.01113

Baumgartner, M., Lemoine, C., Al Seesi, S., Karunakaran, D. K. P., Sturrock, N., Banday, A. R., et al. (2015). Minor splicing snRNAs are enriched in the developing mouse CNS and are crucial for survival of differentiating retinal neurons. *Dev. Neurobiol.* 75, 895–907. doi:10.1002/dneu.22257

Berg, M. G., Singh, L. N., Younis, I., Liu, Q., Pinto, A. M., Kaida, D., et al. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64. doi:10.1016/j.cell.2012.05.029

Berget, S. M. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414. doi:10.1074/jbc.270.6.2411

Bieberstein, N. I., Carrillo Oesterreich, F., Straube, K., and Neugebauer, K. M. (2012). First exon length controls active chromatin signatures and transcription. *Cell Rep.* 2, 62–68. doi:10.1016/j.celrep.2012.05.019

Bindereif, A., and Green, M. R. (1990). Identification and functional analysis of mammalian splicing factors. *Genet. Eng. (N. Y.)* 12, 201–224. doi:10.1007/978-1-4613-0641-2_11

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336. doi:10.1146/annurev.biochem.72.121801.161720

Blankvoort, S., Witter, M. P., Noonan, J., Cotney, J., and Kentros, C. (2018). Marked diversity of unique cortical enhancers enables neuron-specific tools by enhancer-driven gene expression. *Curr. Biol. CB* 28, 2103–2114. doi:10.1016/j.cub.2018.05.015

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. U. S. A.* 75, 4853–4857. doi:10.1073/pnas.75.10.4853

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50, 349–383. doi:10.1146/annurev.bi.50.070181.002025

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lührmann, R. (1984). Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J.* 3, 1357–1363. doi:10.1002/j.1460-2075.1984.tb01977.x

Bringmann, P., and Lührmann, R. (1986). Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* 5, 3509–3516. doi:10.1002/j.1460-2075.1986.tb04676.x

Bringmann, P., Rinke, J., Appel, B., Reuter, R., and Lührmann, R. (1983). Purification of snRNPs U1, U2, U4, U5 and U6 with 2,2,7-trimethylguanosine-specific antibody and definition of their constituent proteins reacting with anti-Sm and anti-(U1)RNP antisera. *EMBO J.* 2, 1129–1135. doi:10.1002/j.1460-2075.1983.tb01557.x

Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E., and Palmiter, R. D. (1988). Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* 85, 836–840. doi:10.1073/pnas.85.3.836

Burge, C. B., Padgett, R. A., and Sharp, P. A. (1998). Evolutionary fates and origins of U12-type introns. *Mol. Cell* 2, 773–785. doi:10.1016/s1097-2765(00)80292-0

Cáceres, J. F., McKenzie, D., Thimmapaya, R., Lund, E., and Dahlberg, J. E. (1992). Control of mouse U1a and U1b snRNA gene expression by differential transcription. *Nucleic Acids Res.* 20, 4247–4254. doi:10.1093/nar/20.16.4247

Carmel, L., Wolf, Y. I., Rogozin, I. B., and Koonin, E. V. (2007). Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res.* 17, 1034–1044. doi:10.1101/gr.6438607

Carranza, F., Shenasa, H., and Hertel, K. J. (2022). Splice site proximity influences alternative exon definition. *RNA Biol.* 19, 829–840. doi:10.1080/15476286.2022.2089478

Catania, F., Gao, X., and Scofield, D. G. (2009). Endogenous mechanisms for the origins of spliceosomal introns. *J. Hered.* 100, 591–596. doi:10.1093/jhered/esp062

Chen, L., Luillo, D. J., Ma, E., Celniker, S. E., Rio, D. C., and Doudna, J. A. (2005). Identification and analysis of U5 snRNA variants in Drosophila. *RNA* 11, 1473–1477. doi:10.1261/rna.2141505

Ciavarella, J., Perea, W., and Greenbaum, N. L. (2020). Topology of the U12–U6 $_{atac}$ snRNA complex of the minor spliceosome and binding by NTC-related protein RBM22. *ACS Omega* 5, 23549–23558. doi:10.1021/acsomega.0c01674

Crick, F. (1979). Split genes and RNA splicing. *Science* 204, 264–271. doi:10.1126/science.373120

Csuros, M., Rogozin, I. B., and Koonin, E. V. (2011). A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLOS Comput. Biol.* 7, e1002150. doi:10.1371/journal.pcbi.1002150

Damgaard, C. K., Kahns, S., Lykke-Andersen, S., Nielsen, A. L., Jensen, T. H., and Kjems, J. (2008). A 5' splice site enhances the recruitment of basal transcription initiation factors *in vivo*. *Mol. Cell* 29, 271–278. doi:10.1016/j.molcel.2007.11.035

De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *WIREs RNA* 4, 49–60. doi:10.1002/wrna.1140

de Melo Costa, V. R., Pfeuffer, J., Louloupi, A., Ørom, U. A. V., and Piro, R. M. (2021). SPLICE-Q: A Python tool for genome-wide quantification of splicing efficiency. *BMC Bioinforma.* 22, 368. doi:10.1186/s12859-021-04282-6

de Wolf, B., Oghabian, A., Akinyi, M. V., Hanks, S., Tromer, E. C., van Hooff, J. J. E., et al. (2021). Chromosomal instability by mutations in the novel minor spliceosome component CENATAC. *EMBO J.* 40, e106536. doi:10.15252/embj.2020106536

Denison, R. A., Van Arsdell, S. W., Bernstein, L. B., and Weiner, A. M. (1981). Abundant pseudogenes for small nuclear RNAs are dispersed in the human genome. *Proc. Natl. Acad. Sci.* 78, 810–814. doi:10.1073/pnas.78.2.810

Dergai, O., Cousin, P., Gouge, J., Satia, K., Praz, V., Kuhlman, T., et al. (2018). Mechanism of selective recruitment of RNA polymerases II and III to snRNA gene promoters. *Genes Dev.* 32, 711–722. doi:10.1101/gad.314245.118

Dey, P., and Mattick, J. S. (2021). High frequency of intron retention and clustered H3K4me3-marked nucleosomes in short first introns of human long non-coding RNAs. *Epigenetics Chromatin* 14, 45. doi:10.1186/s13072-021-00419-2

Di, C., So, B. R., Cai, Z., Arai, C., Duan, J., and Dreyfuss, G. (2019). U1 snRNP telescripting roles in transcription and its mechanism. *Cold Spring Harb. Symp. Quant. Biol.* 84, 115–122. doi:10.1101/sqb.2019.84.040451

Di Giammartino, D. C., Nishida, K., and Manley, J. L. (2011). Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* 43, 853–866. doi:10.1016/j.molcel.2011.08.017

Dietrich, R. C., Incorvaia, R., and Padgett, R. A. (1997). Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol. Cell* 1, 151–160. doi:10.1016/s1097-2765(00)80016-7

Dominov, J. A., Uyan, Ö., McKenna-Yasek, D., Nallamilli, B. R. R., Kergourlay, V., Bartoli, M., et al. (2019). Correction of pseudoexon splicing caused by a novel intronic dysferlin mutation. *Ann. Clin. Transl. Neurol.* 6, 642–654. doi:10.1002/acn3.738

Domitrovich, A. M., and Kunkel, G. R. (2003). Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. *Nucleic Acids Res.* 31, 2344–2352. doi:10.1093/nar/gkg331

Doucet, A. J., Droc, G., Siol, O., Audoux, J., and Gilbert, N. (2015). U6 snRNA pseudogenes: Markers of retrotransposition dynamics in mammals. *Mol. Biol. Evol.* 32, 1815–1832. doi:10.1093/molbev/msv062

Drexler, H. L., Choquet, K., and Churchman, L. S. (2020). Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* 77, 985–998. doi:10.1016/j.molcel.2019.11.017

Edery, P., Marcaillou, C., Sahbatou, M., Labalme, A., Chastang, J., Touraine, R., et al. (2011). Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science* 332, 240–243. doi:10.1126/science.1202205

Effenberger, K. A., Urabe, V. K., and Jurica, M. S. (2017). Modulating splicing with small molecular inhibitors of the spliceosome: Modulating splicing with small molecular inhibitors. *Wiley Interdiscip. Rev. RNA* 8, e1381. doi:10.1002/wrna.1381

Elliott, B., Richardson, C., and Jasin, M. (2005). Chromosomal translocation mechanisms at intronic Alu elements in mammalian cells. *Mol. Cell* 17, 885–894. doi:10.1016/j.molcel.2005.02.028

Elsaid, M. F., Chalhoub, N., Ben-Omran, T., Kumar, P., Kamel, H., Ibrahim, K., et al. (2017). Mutation in noncoding RNA RNU12 causes early onset cerebellar ataxia. *Ann. Neurol.* 81, 68–78. doi:10.1002/ana.24826

Emera, D., Yin, J., Reilly, S. K., Gockley, J., and Noonan, J. P. (2016). Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc. Natl. Acad. Sci.* 113, E2617–E2626. doi:10.1073/pnas.1603718113

Engreitz, J. M., Sirokman, K., McDonel, P., Shishkin, A. A., Surka, C., Russell, P., et al. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* 159, 188–199. doi:10.1016/j.cell.2014.08.018

Fox-Walsh, K. L., Dou, Y., Lam, B. J., Hung, S., Baldi, P. F., and Hertel, K. J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl. Acad. Sci.* 102, 16176–16181. doi:10.1073/pnas.0508489102

Franchini, L. F., López-Leal, R., Nasif, S., Beati, P., Gelman, D. M., Low, M. J., et al. (2011). Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. *Proc. Natl. Acad. Sci. USA* 108, 15270–15275. doi:10.1073/pnas.1104997108

Friend, K., Lovejoy, A. F., and Steitz, J. A. (2007). U2 snRNP binds intronless histone pre-mRNAs to facilitate U7-snRNP-Dependent 3′-end formation. *Mol. Cell* 28, 240–252. doi:10.1016/j.molcel.2007.09.026

Frigola, J., Sabarinathan, R., Mularoni, L., Muiños, F., Gonzalez-Perez, A., and López-Bigas, N. (2017). Reduced mutation rate in exons due to differential mismatch repair. *Nat. Genet.* 49, 1684–1692. doi:10.1038/ng.3991

Frumkin, I., Yofe, I., Bar-Ziv, R., Gurvich, Y., Lu, Y.-Y., Voichek, Y., et al. (2019). Evolution of intron splicing towards optimized gene expression is based on various Cis- and Trans-molecular mechanisms. *PLoS Biol.* 17, e3000423. doi:10.1371/journal.pbio.3000423

Furger, A., O'Sullivan, J. M., Binnie, A., Lee, B. A., and Proudfoot, N. J. (2002). Promoter proximal splice sites enhance transcription. *Genes Dev.* 16, 2792–2799. doi:10.1101/gad.983602

Gehring, N. H., and Roignant, J.-Y. (2021). Anything but ordinary – emerging splicing mechanisms in eukaryotic gene regulation. *Trends Genet.* 37, 355–372. doi:10.1016/j.tig.2020.10.008

Gelfman, S., Burstein, D., Penn, O., Savchenko, A., Amit, M., Schwartz, S., et al. (2011). Changes in exon–intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res.* 22, 35–50. doi:10.1101/gr.119834.110

Georgomanolis, T., Sofiadis, K., and Papantonis, A. (2016). Cutting a long intron short: Recursive splicing and its implications. *Front. Physiol.* 7, 598. doi:10.3389/fphys.2016.00598

Gilbert, W. (1978). Why genes in pieces? *Nature* 271, 501. doi:10.1038/271501a0

Gondane, A., and Itkonen, H. M. (2023). Revealing the history and mystery of RNA-seq. *Curr. Issues Mol. Biol.* 45, 1860–1874. doi:10.3390/cimb45030120

Gozashti, L., Roy, S. W., Thornlow, B., Kramer, A., Ares, M., and Corbett-Detig, R. (2022). Transposable elements drive intron gain in diverse eukaryotes. *Proc. Natl. Acad. Sci.* 119, e2209766119. doi:10.1073/pnas.2209766119

Grabowski, P. J., Seiler, S. R., and Sharp, P. A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* 42, 345–353. doi:10.1016/S0092-8674(85)80130-6

Griffin, C., and Saint-Jeannet, J.-P. (2020). Spliceosomopathies: Diseases and mechanisms. *Dev. Dyn.* 249, 1038–1046. doi:10.1002/dvdy.214

Hall, S. L., and Padgett, R. A. (1994). Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* 239, 357–365. doi:10.1006/jmbi.1994.1377

Hall, S. L., and Padgett, R. A. (1996). Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* 271, 1716–1718. doi:10.1126/science.271.5256.1716

Halperin, R. F., Hegde, A., Lang, J. D., Raupach, E. A., Legendre, C., Liang, W. S., et al. (2021). Improved methods for RNAseq-based alternative splicing analysis. *Sci. Rep.* 11, 10740. doi:10.1038/s41598-021-89938-2

Henry, R. W., Mittal, V., Ma, B., Kobayashi, R., and Hernandez, N. (1998). SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. *Genes Dev.* 12, 2664–2672. doi:10.1101/gad.12.17.2664

Herzel, L., Ottoz, D. S. M., Alpert, T., and Neugebauer, K. M. (2017). Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* 18, 637–650. doi:10.1038/nrm.2017.63

Hesselberth, J. R. (2013). Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA* 4, 677–691. doi:10.1002/wrna.1187

Hoffman, M. M., and Birney, E. (2006). Estimating the neutral rate of nucleotide substitution using introns. *Mol. Biol. Evol.* 24, 522–531. doi:10.1093/molbev/msl179

Huang, Y., Gu, L., and Li, G.-M. (2018). H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J. Biol. Chem.* 293, 7811–7823. doi:10.1074/jbc.RA118.002839

Hudson, A. J., McWatters, D. C., Bowser, B. A., Moore, A. N., Larue, G. E., Roy, S. W., et al. (2019). Patterns of conservation of spliceosomal intron structures and spliceosome divergence in representatives of the diplomonad and parabasalid lineages. *BMC Evol. Biol.* 19, 162. doi:10.1186/s12862-019-1488-y

Huff, J. T., Zilberman, D., and Roy, S. W. (2016). Mechanism for DNA transposons to generate introns on genomic scales. *Nature* 538, 533–536. doi:10.1038/nature20110

Incorvaia, R., and Padgett, R. A. (1998). Base pairing with U6atac snRNA is required for 5' splice site activation of U12-dependent introns *in vivo*. *RNA* 4, 709–718. doi:10.1017/s1355838298980207

Jackson, I. J. (1991). A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* 19, 3795–3798. doi:10.1093/nar/19.14.3795

Jakt, L. M., Dubin, A., and Johansen, S. D. (2022). Intron size minimisation in teleosts. *BMC Genomics* 23, 628. doi:10.1186/s12864-022-08760-w

Jawdekar, G., and Henry, R. (2008). Transcriptional regulation of human small nuclear RNA genes. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1779, 295–305. doi:10.1016/j.bbagrm.2008.04.001

Jiang, M., Zhang, S., Yin, H., Zhuo, Z., and Meng, G. (2023). A comprehensive benchmarking of differential splicing tools for RNA-seq analysis at the event level. *Brief. Bioinform.*, bbad121. bbad121. doi:10.1093/bib/bbad121

Joseph, B., Scala, C., Kondo, S., and Lai, E. C. (2022). Molecular and genetic dissection of recursive splicing. *Life Sci. Alliance* 5, e202101063. doi:10.26508/lsa.202101063

Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., et al. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668. doi:10.1038/nature09479

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., et al. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.* 3, 576–583. doi:10.1038/nchembio.2007.18

Kandul, N. P., and Noor, M. A. (2009). Large introns in relation to alternative splicing and gene evolution: A case study of Drosophila bruno-3. *BMC Genet.* 10, 67. doi:10.1186/1471-2156-10-67

Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R. M., et al. (2011). Cryptic splice sites and split genes. *Nucleic Acids Res.* 39, 5837–5844. doi:10.1093/nar/gkr203

Kazazian, H. H., and Moran, J. V. (1998). The impact of L1 retrotransposons on the human genome. *Nat. Genet.* 19, 19–24. doi:10.1038/ng0598-19

Keane, P. A., and Seoighe, C. (2016). Intron length coevolution across mammalian genomes. *Mol. Biol. Evol.* 33, 2682–2691. doi:10.1093/molbev/msw151

Kolossova, I., and Padgett, R. A. (1997). U11 snRNA interacts *in vivo* with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA* 3, 227–233.

Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: A compromise solution to the introns-early versus introns-late debate? *Biol. Direct* 1, 22. doi:10.1186/1745-6150-1-22

Krchňáková, Z., Thakur, P. K., Krausová, M., Bieberstein, N., Haberman, N., Müller-McNicoll, M., et al. (2019). Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5′ splice-site sequences due to weak interactions with SR proteins. *Nucleic Acids Res.* 47, 911–928. doi:10.1093/nar/gky1147

Kumari, A., Sedehizadeh, S., Brook, J. D., Kozlowski, P., and Wojciechowska, M. (2022). Differential fates of introns in gene expression due to global alternative splicing. *Hum. Genet.* 141, 31–47. doi:10.1007/s00439-021-02409-6

Kwek, K. Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., et al. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat. Struct. Biol.* 9, 800–805. doi:10.1038/nsb862

Lambowitz, A. M., and Zimmerly, S. (2011). Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3, a003616. doi:10.1101/cshperspect.a003616

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi:10.1038/35057062

Larue, G. E., Eliáš, M., and Roy, S. W. (2021). Expansion and transformation of the minor spliceosomal system in the slime mold Physarum polycephalum. *Curr. Biol.* 31, 3125–3131.e4. doi:10.1016/j.cub.2021.04.050

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res.* 20, 320–331. doi:10.1101/gr.101907.109

Laxa, M., Müller, K., Lange, N., Doering, L., Pruscha, J. T., and Peterhänsel, C. (2016). The 5'UTR intron of arabidopsis GGT1 aminotransferase enhances promoter activity by recruiting RNA polymerase II. *Plant Physiol.* 172, 313–327. doi:10.1104/pp.16.00881

Leader, Y., Lev Maor, G., Sorek, M., Shayevitch, R., Hussein, M., Hameiri, O., et al. (2021). The upstream 5′ splice site remains associated to the transcription machinery during intron synthesis. *Nat. Commun.* 12, 4545. doi:10.1038/s41467-021-24774-6

Lee, S., and Stevens, S. W. (2016). Spliceosomal intronogenesis. *Proc. Natl. Acad. Sci.* 113, 6514–6519. doi:10.1073/pnas.1605113113

Lev-Maor, G., Sorek, R., Levanon, E. Y., Paz, N., Eisenberg, E., and Ast, G. (2007). RNA-editing-mediated exon evolution. *Genome Biol.* 8, R29. doi:10.1186/gb-2007-8-2-r29

Li, X., Liu, S., Zhang, L., Issaian, A., Hill, R. C., Espinosa, S., et al. (2019). A unified mechanism for intron and exon definition and back-splicing. *Nature* 573, 375–380. doi:10.1038/s41586-019-1523-6

Li, Y., Xu, Y., and Ma, Z. (2017). Comparative analysis of the exon-intron structure in eukaryotic genomes. *Yangtze Med.* 01, 50–64. doi:10.4236/ym.2017.11006

Liao, S. E., and Regev, O. (2021). Splicing at the phase-separated nuclear speckle interface: A model. *Nucleic Acids Res.* 49, 636–645. doi:10.1093/nar/gkaa1209

Lin, C.-F., Mount, S. M., Jarmołowski, A., and Makałowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol. Biol.* 10, 47. doi:10.1186/1471-2148-10-47

Liu, J., and Maxwell, E. S. (1990). Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. *Nucleic Acids Res.* 18, 6565–6571. doi:10.1093/nar/18.22.6565

Lo, P. C., and Mount, S. M. (1990). *Drosophila melanogaster* genes for U1 snRNA variants and their expression during development. *Nucleic Acids Res.* 18, 6971–6979. doi:10.1093/nar/18.23.6971

Lorenzi, C., Barriere, S., Arnold, K., Luco, R. F., Oldfield, A. J., and Ritchie, W. (2021). IRFinder-S: A comprehensive suite to discover and explore intron retention. *Genome Biol.* 22, 307. doi:10.1186/s13059-021-02515-8

Lu, Z., and Matera, A. G. (2014). Developmental analysis of spliceosomal snRNA isoform expression. *G3 Bethesda Md* 5, 103–110. doi:10.1534/g3.114.015735

Lynch, M., and Richardson, A. O. (2002). The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* 12, 701–710. doi:10.1016/s0959-437x(02)00360-x

Mabin, J. W., Lewis, P. W., Brow, D. A., and Dvinge, H. (2021). Human spliceosomal snRNA sequence variants generate variant spliceosomes. *RNA* 27, 1186–1203. doi:10.1261/rna.078768.121

Malca, H., Shomron, N., and Ast, G. (2003). The U1 snRNP base pairs with the 5′ splice site within a penta-snRNP complex. *Am. Soc. Microbiol.* 3, 3442–3455. doi:10.1128/MCB.23.10.3442–3455.2003

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U. S. A.* 36, 344–355. doi:10.1073/pnas.36.6.344

Meng, F., Zhao, H., Zhu, B., Zhang, T., Yang, M., Li, Y., et al. (2021). Genomic editing of intronic enhancers unveils their role in fine-tuning tissue-specific gene expression in *Arabidopsis thaliana*. *Plant Cell* 33, 1997–2014. doi:10.1093/plcell/koab093

Mercer, T. R., Clark, M. B., Andersen, S. B., Brunck, M. E., Haerty, W., Crawford, J., et al. (2015). Genome-wide discovery of human splicing branchpoints. *Genome Res.* 25, 290–303. doi:10.1101/gr.182899.114

Michel, F., Umesono, K., and Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns-a review. *Gene* 82, 5–30. doi:10.1016/0378-1119(89)90026-7

Mittal, V., Ma, B., and Hernandez, N. (1999). SNAP(c): A core promoter factor with a built-in DNA-binding damper that is deactivated by the oct-1 POU domain. *Genes Dev.* 13, 1807–1821. doi:10.1101/gad.13.14.1807

Montzka, K. A., and Steitz, J. A. (1988). Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc. Natl. Acad. Sci. U A* 85, 8885–8889. doi:10.1073/pnas.85.23.8885

Morales, J., Borrero, M., Sumerel, J., and Santiago, C. (1997). Identification of developmentally regulated sea urchin U5 snRNA genes. *DNA Seq. J. DNA Seq. Mapp.* 7, 243–259. doi:10.3109/10425179709034044

Moyer, D. C., Larue, G. E., Hershberger, C. E., Roy, S. W., and Padgett, R. A. (2020). Comprehensive database and evolutionary dynamics of U12-type introns. *Nucleic Acids Res.* 48, 7066–7078. doi:10.1093/nar/gkaa464

Neugebauer, K. M. (2019). Nascent RNA and the coordination of splicing with transcription. *Cold Spring Harb. Perspect. Biol.* 11, a032227. doi:10.1101/cshperspect.a032227

Nilsen, T. W. (2003). The spliceosome: The most complex macromolecular machine in the cell? *Bioessays* 25, 1147–1149. doi:10.1002/bies.10394

Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., et al. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* 161, 526–540. doi:10.1016/j.cell.2015.03.027

Norppa, A. J., and Frilander, M. J. (2021). The integrity of the U12 snRNA 3′ stem–loop is necessary for its overall stability. *Nucleic Acids Res.* 49, 2835–2847. doi:10.1093/nar/gkab048

Olthof, A. M., Hyatt, K. C., and Kanadia, R. N. (2019). Minor intron splicing revisited: Identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. *BMC Genomics* 20, 686. doi:10.1186/s12864-019-6046-x

Olthof, A. M., White, A. K., and Kanadia, R. N. (2022). The emerging significance of splicing in vertebrate development. *Development* 149, dev200373. doi:10.1242/dev.200373

Olthof, A. M., White, A. K., Mieruszynski, S., Doggett, K., Lee, M. F., Chakroun, A., et al. (2021). Disruption of exon-bridging interactions between the minor and major spliceosomes results in alternative splicing around minor introns. *Nucleic Acids Res.* 49, 3524–3545. doi:10.1093/nar/gkab118

O'Reilly, D., Dienstbier, M., Cowley, S. A., Vazquez, P., Drozdz, M., Taylor, S., et al. (2013). Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Res.* 23, 281–291. doi:10.1101/gr.142968.112

Pai, A. A., Henriques, T., McCue, K., Burkholder, A., Adelman, K., and Burge, C. B. (2017). The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. *eLife* 6, e32537. doi:10.7554/eLife.32537

Parenteau, J., Durand, M., Véronneau, S., Lacombe, A.-A., Morin, G., Guérin, V., et al. (2008). Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol. Biol. Cell* 19, 1932–1941. doi:10.1091/mbc.E07-12-1254

Pessa, H. K. J., Will, C. L., Meng, X., Schneider, C., Watkins, N. J., Perälä, N., et al. (2008). Minor spliceosome components are predominantly localized in the nucleus. *Proc. Natl. Acad. Sci.* 105, 8655–8660. doi:10.1073/pnas.0803646105

Piovesan, A., Caracausi, M., Ricci, M., Strippoli, P., Vitale, L., and Pelleri, M. C. (2015). Identification of minimal eukaryotic introns through GeneBase, a user-friendly tool for parsing the NCBI Gene databank. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 22, 495–503. doi:10.1093/dnares/dsv028

Pitolli, C., Marini, A., Sette, C., and Pagliarini, V. (2022). Non-canonical splicing and its implications in brain physiology and cancer. *Int. J. Mol. Sci.* 23, 2811. doi:10.3390/ijms23052811

Qian, X., Wang, J., Wang, M., Igelman, A. D., Jones, K. D., Li, Y., et al. (2021). Identification of deep-intronic splice mutations in a large cohort of patients with inherited retinal diseases. *Front. Genet.* 12. doi:10.3389/fgene.2021.647400

Reddy, R., Henning, D., Das, G., Harless, M., and Wright, D. (1987). The capped U6 small nuclear RNA is transcribed by RNA polymerase III. *J. Biol. Chem.* 262, 75–81. doi:10.1016/s0021-9258(19)75890-6

Reed, R., and Maniatis, T. (1986). A role for exon sequences and splice-site proximity in splice-site selection. *Cell* 46, 681–690. doi:10.1016/0092-8674(86)90343-0

Reimer, K. A., Mimoso, C. A., Adelman, K., and Neugebauer, K. M. (2021). Co-transcriptional splicing regulates 3′ end cleavage during mammalian erythropoiesis. *Mol. Cell* 81, 998–1012.e7. doi:10.1016/j.molcel.2020.12.018

Resch, A. M., Carmel, L., Mariño-Ramírez, L., Ogurtsov, A. Y., Shabalina, S. A., Rogozin, I. B., et al. (2007). Widespread positive selection in synonymous sites of mammalian genes. *Mol. Biol. Evol.* 24, 1821–1831. doi:10.1093/molbev/msm100

Robberson, B. L., Cote, G. J., and Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10, 84–94. doi:10.1128/mcb.10.1.84

Roca, X., Sachidanandam, R., and Krainer, A. R. (2003). Intrinsic differences between authentic and cryptic 5′ splice sites. *Nucleic Acids Res.* 31, 6321–6333. doi:10.1093/nar/gkg830

Rodriguez-Galindo, M., Casillas, S., Weghorn, D., and Barbadilla, A. (2020). Germline de novo mutation rates on exons versus introns in humans. *Nat. Commun.* 11, 3304. doi:10.1038/s41467-020-17162-z

Rogozin, I. B., Carmel, L., Csuros, M., and Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biol. Direct* 7, 11. doi:10.1186/1745-6150-7-11

Romfo, C. M., Alvarez, C. J., van Heeckeren, W. J., Webb, C. J., and Wise, J. A. (2000). Evidence for splice site pairing via intron definition in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* 20, 7955–7970. doi:10.1128/mcb.20.21.7955-7970.2000

Rose, A. B., Elfersi, T., Parra, G., and Korf, I. (2008). Promoter-proximal introns in *Arabidopsis thaliana* are enriched in dispersed signals that elevate gene expression. *Plant Cell* 20, 543–551. doi:10.1105/tpc.107.057190

Russell, A. G., Charette, J. M., Spencer, D. F., and Gray, M. W. (2006). An early evolutionary origin for the minor spliceosome. *Nature* 443, 863–866. doi:10.1038/nature05228

Ryll, J., Rothering, R., and Catania, F. (2022). Intronization signatures in coding exons reveal the evolutionary fluidity of eukaryotic gene architecture. *Microorganisms* 10, 1901. doi:10.3390/microorganisms10101901

Sadowski, C. L., Henry, R. W., Lobo, S. M., and Hernandez, N. (1993). Targeting TBP to a non-TATA box cis-regulatory element: A TBP-containing complex activates transcription from snRNA promoters through the PSE. *Genes Dev.* 7, 1535–1548. doi:10.1101/gad.7.8.1535

Sakharkar, M. K., Chow, V. T. K., and Kangueane, P. (2004). Distributions of exons and introns in the human genome. *Silico Biol.* 4, 387–393.

Samadder, P., Sivamani, E., Lu, J., Li, X., and Qu, R. (2008). Transcriptional and post-transcriptional enhancement of gene expression by the 5' UTR intron of rice rubi3 gene in transgenic rice cells. *Mol. Genet. Genomics MGG* 279, 429–439. doi:10.1007/s00438-008-0323-8

Sánchez-Escabias, E., Guerrero-Martínez, J. A., and Reyes, J. C. (2022). Co-transcriptional splicing efficiency is a gene-specific feature that can be regulated by TGFβ. *Nat. Commun.* 5, 277. doi:10.1038/s42003-022-03224-z

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., et al. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274, 765–768. doi:10.1126/science.274.5288.765

Seal, R. L., Chen, L.-L., Griffiths-Jones, S., Lowe, T. M., Mathews, M. B., O'Reilly, D., et al. (2020). A guide to naming human non-coding RNA genes. *EMBO J.* 39, e103777. doi:10.15252/embj.2019103777

Sharp, P. A. (1991). Five easy pieces. *Science* 254, 663. doi:10.1126/science.1948046

Shen, H., Zheng, X., Luecke, S., and Green, M. R. (2010). The U2AF35-related protein Urp contacts the 3′ splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev.* 24, 2389–2394. doi:10.1101/gad.1974810

Shepard, S., McCreary, M., and Fedorov, A. (2009). The peculiarities of large intron splicing in animals. *PLOS ONE* 4, e7853. doi:10.1371/journal.pone.0007853

Sheridan, R. M., Fong, N., D'Alessandro, A., and Bentley, D. L. (2019). Widespread backtracking by RNA pol II is a major effector of gene activation, 5' pause release, termination, and transcription elongation rate. *Mol. Cell* 73, 107–118. doi:10.1016/j.molcel.2018.10.031

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R., and Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* 34, 3955–3967. doi:10.1093/nar/gkl556

Shiau, C.-K., Huang, J.-H., Liu, Y.-T., and Tsai, H.-K. (2022). Genome-wide identification of associations between enhancer and alternative splicing in human and mouse. *BMC Genomics* 22, 919. doi:10.1186/s12864-022-08537-1

Shukla, G. C., and Padgett, R. A. (2002). A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol. Cell* 9, 1145–1150. doi:10.1016/S1097-2765(02)00505-1

Siebert, A. E., Corll, J., Paige Gronevelt, J., Levine, L., Hobbs, L. M., Kenney, C., et al. (2022). Genetic analysis of human RNA binding motif protein 48 (RBM48) reveals an essential role in U12-type intron splicing. *Genetics* 222, iyac129. doi:10.1093/genetics/iyac129

Sierra-Montes, J. M., Pereira-Simon, S., Smail, S. S., and Herrera, R. J. (2005). The silk moth *Bombyx mori* U1 and U2 snRNA variants are differentially expressed. *Gene* 352, 127–136. doi:10.1016/j.gene.2005.02.013

Simari, R. D., Yang, Z.-Y., Ling, X., Stephan, D., Perkins, N. D., Nabel, G. J., et al. (1998). Requirements for enhanced transgene expression by untranslated sequences from the human cytomegalovirus immediate-early gene. *Mol. Med.* 4, 700–706. doi:10.1007/BF03401764

Singh, J., and Padgett, R. A. (2009). Rates of *in situ* transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16, 1128–1133. doi:10.1038/nsmb.1666

Singh, P., Börger, C., More, H., and Sturmbauer, C. (2017). The role of alternative splicing and differential gene expression in Cichlid adaptive radiation. *Genome Biol. Evol.* 9, 2764–2781. doi:10.1093/gbe/evx204

Smathers, C. M., and Robart, A. R. (2019). The mechanism of splicing as told by group II introns: Ancestors of the spliceosome. *Biochim. Biophys. Acta Gene Regul. Mech.* 1862, 194390. doi:10.1016/j.bbagrm.2019.06.001

So, B. R., Di, C., Cai, Z., Venters, C. C., Guo, J., Oh, J.-M., et al. (2019). A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells. *Mol. Cell* 76, 590–599. doi:10.1016/j.molcel.2019.08.007

Sontheimer, E. J., Gordon, P. M., and Piccirilli, J. A. (1999). Metal ion catalysis during group II intron self-splicing: Parallels with the spliceosome. *Genes Dev.* 13, 1729–1741. doi:10.1101/gad.13.13.1729

Sontheimer, E. J., and Steitz, J. A. (1992). Three novel functional variants of human U5 small nuclear RNA. *Mol. Cell Biol.* 12, 734–746. doi:10.1128/mcb.12.2.734

Sorek, R., Ast, G., and Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* 12, 1060–1067. doi:10.1101/gr.229302

Spiller, M. P., Boon, K.-L., Reijns, M. A. M., and Beggs, J. D. (2007). The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Res.* 35, 923–929. doi:10.1093/nar/gkl1130

Spiluttini, B., Gu, B., Belagal, P., Smirnova, A. S., Nguyen, V. T., Hebert, C., et al. (2010). Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *J. Cell Sci.* 123, 2085–2093. doi:10.1242/jcs.061358

Steitz, J. A., Dreyfuss, G., Krainer, A. R., Lamond, A. I., Matera, A. G., and Padgett, R. A. (2008). Where in the cell is the minor spliceosome? *Proc. Natl. Acad. Sci. U. S. A.* 105, 8485–8486. doi:10.1073/pnas.0804024105

Sun, H., and Chasin, L. A. (2000). Multiple splicing defects in an intronic false exon. *Mol. Cell. Biol.* 20, 6414–6425. doi:10.1128/MCB.20.17.6414-6425.2000

Sun, J., Li, X., Hou, X., Cao, S., Cao, W., Zhang, Y., et al. (2022). Structural basis of human SNAPc recognizing proximal sequence element of snRNA promoter. *Nat. Commun.* 13, 6871. doi:10.1038/s41467-022-34639-1

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G. L., Cheutin, T., Bonev, B., et al. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* 52, 1151–1157. doi:10.1038/s41588-020-00716-8

Szentirmay, M. N., and Sawadogo, M. (2000). Spatial organization of RNA polymerase II transcription in the nucleus. *Nucleic Acids Res.* 28, 2019–2025. doi:10.1093/nar/28.10.2019

Tammer, L., Hameiri, O., Keydar, I., Roy, V. R., Ashkenazy-Titelman, A., Custódio, N., et al. (2022). Gene architecture directs splicing outcome in separate nuclear spatial regions. *Mol. Cell* 82, 1021–1034.e8. doi:10.1016/j.molcel.2022.02.001

Tarn, W. Y., and Steitz, J. A. (1996a). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell* 84, 801–811. doi:10.1016/s0092-8674(00)81057-0

Tarn, W. Y., and Steitz, J. A. (1996b). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* 273, 1824–1832. doi:10.1126/science.273.5283.1824

Tarn, W. Y., and Steitz, J. A. (1997). Pre-mRNA splicing: The discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* 22, 132–137. doi:10.1016/s0968-0004(97)01018-9

Tellier, M., Maudlin, I., and Murphy, S. (2020). Transcription and splicing: A two-way street. *WIREs RNA* 11, e1593. doi:10.1002/wrna.1593

Theissen, H., Rinke, J., Traver, C. N., Lührmann, R., and Appel, B. (1985). Novel structure of a human U6 snRNA pseudogene. *Gene* 36, 195–199. doi:10.1016/0378-1119(85)90086-1

Thompson, P. J., Macfarlan, T. S., and Lorincz, M. C. (2016). Long terminal repeats: From parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* 62, 766–776. doi:10.1016/j.molcel.2016.03.029

Tichelaar, J. W., Wieben, E. D., Reddy, R., Vrabel, A., and Camacho, P. (1998). *In vivo* expression of a variant human U6 RNA from a unique, internal promoter. *Biochemistry* 37, 12943–12951. doi:10.1021/bi9811361

Tronchère, H., Wang, J., and Fu, X. D. (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* 388, 397–400. doi:10.1038/41137

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., et al. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 26, 70–84. doi:10.1101/gr.196006.115

Vakirlis, N., Vance, Z., Duggan, K. M., and McLysaght, A. (2022). De novo birth of functional microproteins in the human lineage. *Cell Rep.* 41, 111808. doi:10.1016/j.celrep.2022.111808

Van Arsdell, S. W., and Weiner, A. M. (1984). Human genes for U2 small nuclear RNA are tandemly repeated. *Mol. Cell. Biol.* 4, 492–499. doi:10.1128/mcb.4.3.492

Vazquez-Arango, P., Vowles, J., Browne, C., Hartfield, E., Fernandes, H. J. R., Mandefro, B., et al. (2016). Variant U1 snRNAs are implicated in human pluripotent stem cell maintenance and neuromuscular disease. *Nucleic Acids Res.* 44, 10960–10973. doi:10.1093/nar/gkw711

Vosseberg, J., and Snel, B. (2017). Domestication of self-splicing introns during eukaryogenesis: The rise of the complex spliceosomal machinery. *Biol. Direct* 12, 30. doi:10.1186/s13062-017-0201-6

Wahl, M. C., Will, C. L., and Lührmann, R. (2009). The spliceosome: Design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi:10.1016/j.cell.2009.02.009

Wallace, E. W. J., and Beggs, J. D. (2017). Extremely fast and incredibly close: Cotranscriptional splicing in budding yeast. *RNA* 23, 601–610. doi:10.1261/rna.060830.117

Wang, Z., Xiao, X., Van Nostrand, E., and Burge, C. B. (2006). General and specific functions of exonic splicing silencers in splicing control. *Mol. Cell* 23, 61–70. doi:10.1016/j.molcel.2006.05.018

Wells, J. N., and Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* 54, 539–561. doi:10.1146/annurev-genet-040620-022145

Will, C. L., Schneider, C., Reed, R., and Lührmann, R. (1999). Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* 284, 2003–2005. doi:10.1126/science.284.5422.2003

Yan, D., and Ares, M. (1996). Invariant U2 RNA sequences bordering the branchpoint recognition region are essential for interaction with yeast SF3a and SF3b subunits. *Mol. Cell. Biol.* 16, 818–828. doi:10.1128/mcb.16.3.818

Younis, I., Dittmar, K., Wang, W., Foley, S. W., Berg, M. G., Hu, K. Y., et al. (2013). Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *eLife* 2, e00780. doi:10.7554/eLife.00780

Zhang, S., Aibara, S., Vos, S. M., Agafonov, D. E., Lührmann, R., and Cramer, P. (2021). Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science* 371, 305–309. doi:10.1126/science.abf1870

Zhang, X.-O., Fu, Y., Mou, H., Xue, W., and Weng, Z. (2018). The temporal landscape of recursive splicing during Pol II transcription elongation in human cells. *PLOS Genet.* 14, e1007579. doi:10.1371/journal.pgen.1007579

Zhu, D., Mao, F., Tian, Y., Lin, X., Gu, L., Gu, H., et al. (2020). The features and regulation of Co-transcriptional splicing in arabidopsis. *Mol. Plant* 13, 278–294. doi:10.1016/j.molp.2019.11.004

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 20 June 2023
DOI 10.3389/fgene.2023.1159109

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Jingang Gui,
Beijing Children Hospital Affiliated to Capital Medical University, China
Mark Izraelson,
Institute of Bioorganic Chemistry (RAS), Russia

*CORRESPONDENCE
Cuong Q. Nguyen,
nguyenc@ufl.edu

RECEIVED 05 February 2023
ACCEPTED 12 June 2023
PUBLISHED 20 June 2023

CITATION
Shen Y, Voigt A, Leng X, Rodriguez AA and Nguyen CQ (2023), A current and future perspective on T cell receptor repertoire profiling.
Front. Genet. 14:1159109.
doi: 10.3389/fgene.2023.1159109

A current and future perspective on T cell receptor repertoire profiling

Yiran Shen[1], Alexandria Voigt[1], Xuebing Leng[2], Amy A. Rodriguez[1] and Cuong Q. Nguyen[1,3,4]*

[1]Department of Infectious Diseases and Immunology, College of Veterinary Medicine, University of Florida, Gainesville, FL, United States, [2]Department of Microbiology and Immunology, Miller School of Medicine, University of Miami, Miami, FL, United States, [3]Department of Oral Biology, College of Dentistry, University of Florida, Gainesville, FL, United States, [4]Center of Orphaned Autoimmune Diseases, University of Florida, Gainesville, FL, United States

T cell receptors (TCR) play a vital role in the immune system's ability to recognize and respond to foreign antigens, relying on the highly polymorphic rearrangement of TCR genes. The recognition of autologous peptides by adaptive immunity may lead to the development and progression of autoimmune diseases. Understanding the specific TCR involved in this process can provide insights into the autoimmune process. RNA-seq (RNA sequencing) is a valuable tool for studying TCR repertoires by providing a comprehensive and quantitative analysis of the RNA transcripts. With the development of RNA technology, transcriptomic data must provide valuable information to model and predict TCR and antigen interaction and, more importantly, identify or predict neoantigens. This review provides an overview of the application and development of bulk RNA-seq and single-cell (SC) RNA-seq to examine the TCR repertoires. Furthermore, discussed here are bioinformatic tools that can be applied to study the structural biology of peptide/TCR/MHC (major histocompatibility complex) and predict antigenic epitopes using advanced artificial intelligence tools.

KEYWORDS

TCR, RNA, autoimmune, HLA, sequencing

1 Introduction

T cell function is initiated by recognition of a peptide antigen in a specific interaction via the T cell receptor (TCR) in the context of the major histocompatibility complex (MHC) expressed on antigen-presenting cells (APC). TCRs are heterodimeric membrane proteins that are composed of two chains, αβ or γδ. The α chain is made up of the variable (V), joining (J) and constant (C) segments, and the β chain contains the V, D (diversity), J, and C segments. The gene segment organization of the TCRγ and TCRδ chains is similar to that of the αβ TCR. TCR development in the thymus is critical for development of a functional immune system. The gene rearrangement of a TCR involves the selection of immature T cells in the thymus maturing to become functional T cells that recognize foreign molecules and respond to them appropriately. The mature T cells undergo positive and negative selection, in which they are presented with self-antigens from the thymus for affinity selection to prevent autoreactive TCR repertoires. This process helps to ensure that only mature T cells respond to foreign antigens exclusively are allowed to survive and develop (Schatz and Ji, 2011). The rearrangement leads to a vast diversity of TCR repertoires capable of recognizing almost any peptide presented by MHC molecules (Mitchell and Michels, 2020). The diversity

of the αβ TCR from the unique pairing of various gene segments or loci generates on the order of 10^{18} or more possible combinations (Murphy and Weaver, 2022). Once naïve T cells encounter the peptide-MHC complex (pMHC) presented by an APC, these T cells will start to undergo clonal expansion while retaining the initial TCR sequence (Huang et al., 2019).

T cells and their receptors are crucial in autoimmunity. Recognition of autoantigens by T cells with self-reactive TCRs can result in tissue-specific damage of systemic autoimmune diseases (Seiringer et al., 2022). A fitting model for this process is Sjögren's disease (SjD), which is a debilitating disease affecting as many as 3.1 million individuals in the United States(Kassan and Moutsopoulos, 2004; Helmick et al., 2008; Nguyen and Peck, 2009). In addition to secretory dysfunction resulting in dry mouth (xerostomia) and dry eyes (keratoconjunctivitis sicca), symptoms can manifest systemically to the skin, gastrointestinal tract, lungs, blood vessels, liver, pancreas, kidneys, vagina, and peripheral and central nervous system (Cornec et al., 2014; Voigt et al., 2014; Nocturne and Mariette, 2015; Voigt and Nguyen, 2015). The TCR usage of individual αβ T cells showed that the TCR-Vα repertoire of infiltrating T cells is restricted with limited heterogeneity. Specifically, Vα usage of TCR genes, including Vα17.1, Vα2, and Vα11.1, were found dominantly in salivary glands (SG) and not peripheral blood mononuclear cells (PBMCs) (Sumida et al., 1994a). A study (Joachims et al., 2016) demonstrated that glandular memory T cells showed a number of TCRs, specifically TRAV8-2, 12-3, 12-2, 16, and TRBV30, 20-1, 19, 7-6, 14, 20-1, 3-1, and 24-1. In the non-obese diabetic (NOD) mouse model, it has been shown that 15% of the TRBV gene is Vβ8.1.2, followed by Vβ6, Vβ10b, Vβ11, Vβ2, and Vβ7 (Sumida et al., 1994b; Skarstein et al., 1995). During autoimmune sialadenitis or early stages of the disease, the predominant expression of the Vβ8 gene increased over time in the MRL/lpr strain. Although the self-antigen was not identified, the usage of TCR-Vβ elements being restricted according to the stage of the disease indicates a clonal selection of antigen-specific TCR in the SG, suggesting that the diversity of TCR repertoires is disease- and stage-dependent (Hayashi et al., 1995).

The studies, as mentioned earlier, applied various techniques to study TCR and cell types based on transcriptomic data. To advance beyond the transcriptome, one must be able to decipher the antigen or autoantigens presented to the T cells, which will further our understanding of the immunological mechanism underpinning the onset and progression as well as improve clinical diagnostics and therapeutics. The overall objective of this review is to describe the latest technological advances that have had a significant impact on profiling TCR repertoires and concomitantly linking them to the cellular transcriptomic profiles of the target cells. In addition, we discuss predictive modeling based on particular antigenic epitopes and TCR repertoires.

2 Development of RNA sequencing (RNA-seq) technology to identify TCR repertoire

2.1 Single-stranded RNA-seq

Molecular cloning and Sanger sequencing were the first methods to study immune repertoires at the nucleotide sequence level (Figure 1). Early work by Sant'Angelo et al. (1998) showed that the complementarity determining region 3 (CDR3) can be obtained by designing primers for the paired V- and C-region's primary and restriction fragment length polymorphism (RFLP) with nested PCR amplification. They sequenced CDR3 regions, analyzed TCRα chains from different TCRβ chain-transgenic mice, and constructed a molecular map of T cell development; they identified the precise stage of positive selection that occurs early in thymocyte differentiation. Later, Correia-Neves et al. (2001) designed a mouse line by combining the TCRβ transgene with the TCRα minilocus consisting of a single V and two J gene segments. They also performed nested PCR by paired primers designed with a similar concept to determine the diversity of CDR3α. This approach allowed them to follow the fate of T cells with different TCR sequences, thus enabling them to study the selection and evolution of the T cell repertoires. A widely used method is multiplex PCR, wherein multiple primers are designed to amplify all possible V segments using degenerate primers and conserved region primers. Primer bias can occur with this approach which distorts the resulting TCR repertoires, therefore sequencing the final cDNA must be done to confirm the identify the targeted receptors (Liu et al., 2016). Unbiased 5′-Rapid Amplification of cDNA Ends (5′ RACE) is alternative method, as it amplifies TCR genes using only one primer targeting a constant region and a universal primer attached to the 5′ end (Mamedov et al., 2013). Recently, Cook et al. (2020) used 5′ RACE PCR to amplify TCRβ chain and Sanger sequencing to analyze the TCR repertoire of the regulatory $CD4^+$ T cell (Treg) population and found that the TCR repertoire of gluten-specific $CD39^+$ Tregs in celiac disease patients was oligoclonal compared to healthy controls, suggesting that the repertoire of gluten-specific $CD39^+$ Tregs may be driven by the specific antigen and the corresponding human leukocyte antigen (HLA) restriction. Unlike multiplex PCR, which can use both genomic DNA and RNA as input, 5′-RACE can only be applied to RNA samples, and the presence of short DNA fragments in the 5′-RACE library may result in sequencing results that do not effectively present regularly recombinant TCR sequences (Lin et al., 2020).

Even though there are many options and optimizations in the methods of molecular cloning to generate sequencing materials, Sanger sequencing is limited due to low throughput and can only sequence a small number of TCRs in a given batch. In particular, during bulk analysis, PCR technologies can only amplify and sequence one strand at a time, thus losing the pairwise information of intact immune repertoires, making it impossible to determine the antigenic specificity of pathogenic TCR information subsequently. Meanwhile, most early TCR profiling studies were based on analysis of the CDR3 region; however, full-length sequencing includes other regions, e.g., CDR1 and CDR2, involves in antigen receptor binding affinity and/or downstream signaling, and allows direct cloning and expression of identified and selected receptors for further experiments (Mazzotti et al., 2022). The widespread use of next-generation sequencing (NGS) based RNA-seq has shaped our understanding of many aspects of biology. Unlike genomic DNA-based applications, RNA-seq provides comprehensive gene expression information from mRNA in addition to the immune repertoire. Short-read RNA-seq is cheaper and easier to perform than microarrays and produces

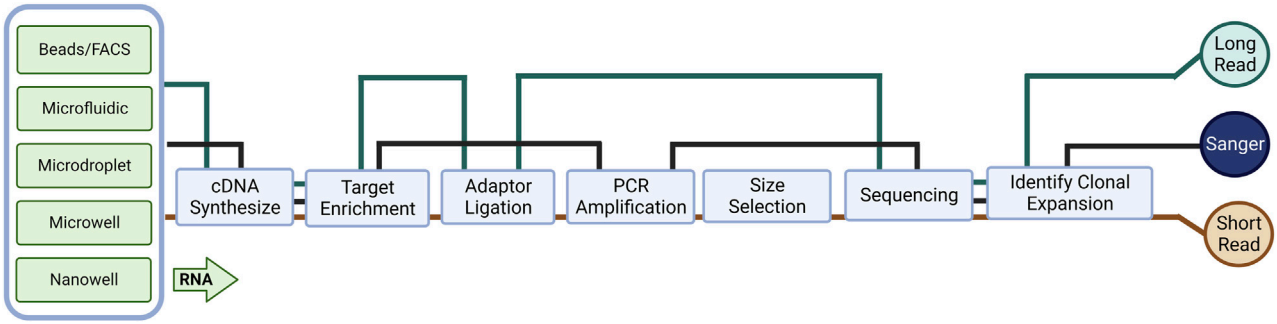


FIGURE 1
An overview of library preparation methods for different cell preparation and RNA sequencing (RNA-seq) methods. The complexity and bias of library preparation vary depending on the specific method used.

comprehensive, high-quality, less biased data in a shorter time, thus allowing easy determination of clonal expansion in TCR replication. More importantly, in TCR sequencing, the choice of RNA rather than DNA as starting material avoids small sample size of genomic DNA (gDNA), as well as reducing background interference and primer bias from amplification of V and J fragments that are not involved in recombination but remained in gDNA.

In autoimmune diseases, current commercial services can identify the most frequent single-strand used TCR from patients, typically starting with multiplex PCR to amplify all Vα or Vβ regions followed by short-read RNA-seq to confirm the clonal expansion of immune cells. Muraro et al. (2014) used high-throughput deep TCRβ sequencing provided by ImmunoSEQ to assess millions of individual TCRs in multiple sclerosis (MS) patients with poor prognosis per hematopoietic stem cell transplantation (HSCT) treatment and found that the combination of high-dose immunosuppressive therapy (HDIT) and autologous HSCT had a unique and independent effect on reconstituting $CD4^+$ and $CD8^+$ repertoires, which supports the view that TCR repertoire diversity is critical for reestablishing immune tolerance. However, ImmunoSEQ is a gDNA-based service, which is accomplished by pre-running a synthetic immune repertoire that represents all combinations of V-J genes, before selecting and adjusting primer concentrations to reduce bias during amplification. The most widely used commercial RNA-based kit is iRepertoire, Chang et al. (2019) used iRepertoire to sequence the TCRβ CDR3 region to determine the role of T cell profiles in rheumatoid arthritis patients receiving different biologic disease-modifying antirheumatic drugs (bDMARDs). An index of clonality of the TCRβ repertoires in RA patients was found to be negatively correlated with age, while a trend toward increased disease activity was observed with reduced TCRβ repertoire diversity following bDMARDs treatment. Using the same technique, Amoriello et al. (2020) tracked peripheral T cell subsets in 15 relapsing-remitting multiple sclerosis (RRMS) patients before and after 2 years of continuous treatment with Natalizumab (NTZ) and a single course of therapy with autologous hematopoietic stem cell transplantation (AHSCT) by high-throughput TCRβ sequencing, they found that both treatments left treatment-specific multidimensional traces in patient TCRβ repertoire dynamics related to clonal amplification, clonal diversity, and repertoire structure. A comparison of iRepertoire with other commercially available kits (MiLaboratories, Takara, NEB) is shown in Table 1. Amplification can also be performed by adding adaptor sequences into TCR multiplex PCR primers, Wang et al. (2021) first used scRNA-seq to reveal a novel Graves' orbitopathy (GO)-specific cell type, $CD4^+$ cytotoxic T lymphocytes (CTL), and to understand the clonal expansion of this $CD4^+$ CTL population, they performed TCRβ CDR3 sequencing, revealing the significant clonal expansion of $CD4^+$ $KLRG1^+$ CTL from GO patients.

It is possible to sequence large cell populations in this manner, but the fact that it can only be based on single-strand RNA-seq is likewise a limiting point. Due to the presence of the D loci, the TCRβ chain has a higher combinatorial potential than the TCRα chain. Also, due to allelic exclusion (Khor and Sleckman, 2002) and the possibility of two α chains being expressed by the same cell (Padovan et al., 1993), the single β chain expressed per αβ T cell has become the main target for single-strand sequencing studies, but this introduces a sample bias.

2.2 Paired-stranded sequencing based on short-read single-cell RNA-seq (scRNA-seq)

Developments in wet lab technology and computing drive the adaptation and evolution of RNA-seq. In this context, single-cell-based experimental techniques can overcome the limitations of single-strand sequencing in TCRs (Hou et al., 2016). Paired TCR αβ or γδ sequences can provide additional information on p (peptide) MHC binding specificity, which is essential for the study of autoimmune disease etiology and progression. Low-put through scRNA-seq involves manually sorting and isolating individual cells by magnetic bead sorting or fluorescence-activated cell sorting (FACS) into multi-well plates. Switching Mechanism at the end of the 5′-end of the RNA Transcript (SMART)-seq (Goetz and Trimarchi, 2012), Smart-seq2 (Picelli et al., 2013), MATQ-seq (Sheng et al., 2017), CEL-seq (Hashimshony et al., 2012) and other protocols can rely on FACS sorting. After first strand cDNA synthesis, unlike non-linear PCR, platforms, for example, CEL-seq utilize *in vitro* transcription (IVT) technology, it requires an additional round of

TABLE 1 Major commercially available kits for TCR profiling.

	Milaboratories	NEBNext® immune sequencing	SMARTer TCR a/b profiling	iRepertoire
Species	Mouse, human, and monkey	Mouse and human	Mouse and human	Mouse and human
Protocol	Multiplex PCR	5′ RACE	SMART technology	arm-PCR
UMI	Yes	Yes	No	No
Input material	Up to 500 ng	10 ng–1 μg RNA or RNA- contained cells	10 ng–3 μg of RNA or 50–10,000 cells	50 ng–1 μg RNA
Sequencing	Illumina	Illumina Miseq®	Illumina Miseq®	Illumina
Analysis	MIXCR and MIGEC	Presto (Galaxy)	Any softwares	iRepertoire

SMART, Switching Mechanism at 5′ End of RNA template; arm-PCR, amplicon rescued multiplex PCR.

TABLE 2 Current automate platform for single cell processing.

Platform	Compatible protocol	Transcript coverage	UMI
Fluidigm C1	SMART-Seq	Full-length	No
10X Genomics	Chromium	5′-/3′-	Yes
Clontech	SMART-Seq	Full-length	No
ICELL8	SMART-Seq	Full-length	No
BD Rhapsody	Whole transcriptome analysis (WTA)	3′-	Yes

reverse transcription of the amplified RNA which results in a 3′-bias.

To process a large number of single cells simultaneously, several commercial services have introduced either microfluidic (Fluidigm C1), microdroplet (10X Genomics), microwell (Clontech, BD Rhapsody), or nanowell (ICELL8)-based platforms that allow for automated isolation, lysis, and cDNA synthesis for each cell (Figure 1). These automated platforms rely on in-house developed instrumentation, which reduces the batch effect of samples but increases costs. These platforms utilize a variety of different cell isolation techniques while differing in cell lysis, reverse transcription, amplification, transcript coverage, strand specificity, or UMI (Unique Molecular Identifier) availability (Table 2). To estimate technical differences between cells, correct the errors, and normalize data, the use of UMIs can offset differences in mRNA amplification efficiency, which can detect and quantify molecular labels of unique transcripts. Another option is the use of external RNA control consortium (ERCC) introduced into the samples to calibrate measurements and account for technical variation, which was applied in SMART-seq2 protocol but is not compatible with droplet-based platforms (Svensson et al., 2017; Baran-Gale et al., 2018).

Still, automatic single-cell processing reduces intracellular RNA degradation and library preparation time, and scTCR-seq facilitates the exploration of the immune repertoire with great diversity. These factors together allow us to further explore key cell subpopulations and differentiation states through transcriptome analysis and to infer cell developmental trajectories at the single-cell level while providing additional information related to the TCR repertoire. We recently utilized Chromium Single Cell Immune Profiling (10X Genomics) to identify the specific immune cell subsets and the expressed TCR repertoire of single T cells. The technology combines single-cell sequencing and molecular barcoding to measure the TCR sequences expressed by individual T cells, allowing us to make detailed inferences about the composition and diversity of the immune system. In here we present a mockup figure to illustrate the whole workflow (Figure 2), we found different populations of immune cells present in the salivary glands of SjD-susceptible mice (Figure 2A). When we examined the TCR repertoire expressed by the effector $CD4^+$ T cells, we were able to identify the dominant receptors (Figure 2B). We further demonstrated that males and females of the same SjD background exhibited different TCR repertoires (Figure 2C).

The use of scRNA-seq for identifying autoimmune disease-related immune repertoires has only recently emerged, including type 1 diabetes (T1D) (Linsley et al., 2021; Kasmani et al., 2022), autoimmune hepatitis (AIH) (Renand et al., 2020), primary SjD (pSjD) (Hong et al., 2020; Hou et al., 2022), and systemic lupus erythematosus (SLE) (Smita et al., 2022) (Table 3). Not only is it a direct study of the disease itself, but scTCR-seq has also been used to study T cell populations and/or related mechanisms closely associated with autoimmune disorders, allowing us to visualize the immune repertoire expressed by several cell subpopulations. In spondyloarthritis (SpA) patients, arthritogenic peptides are presented by the risk allele HLA-B*27 to antigen-specific $CD8^+$ T cells to initiate or maintain an autoimmune response, Deschler et al. (2022) used scTCR-seq to analyze $CD8^+$ T cells in the patient's synovial fluid (SF) and revealed a preferential expansion of the TCR TRAV- and TRBV- families, common motifs in the CDR3 loop and identical TCR chains across patients. Follicular helper T cells are central regulators of germinal centers and contribute to the formation of pathogenic autoantibodies, Akama-Garren et al.

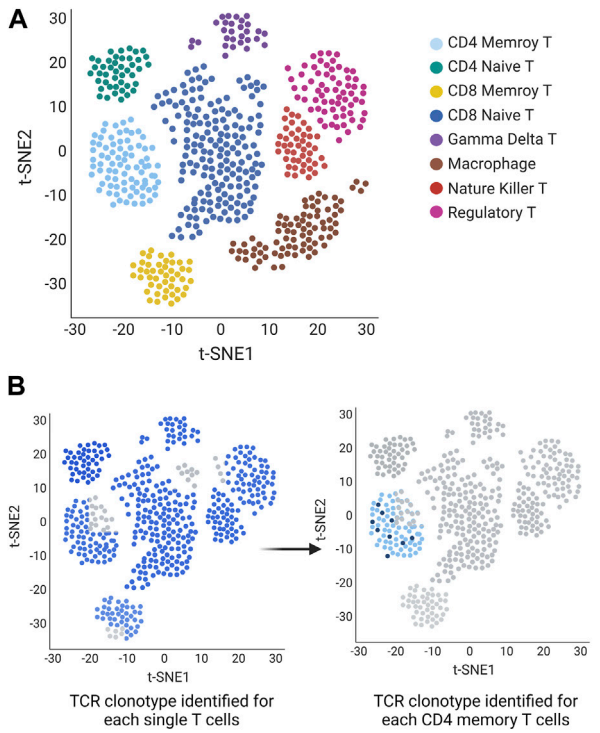


FIGURE 2
Mockup Chromium Single Cell Immune Profiling (10X Genomics) to identify the specific immune cell subsets and the expressed TCR repertoire of single T cells. **(A)** UMAP of the different cell types is classified based on different colors. **(B)** TCR repertoires in each T cell (left) and selected memory T cell (right). Each cell identified with a TCR clonotype was labeled blue, otherwise was grey (no identified TCRα/β or TCRγ/δ present). The same clonotype is marked in the same dark blue color (right). **(C)** The different clonotypes expansion between sexes showed as a percentage of unique clonotypes, and share clonotypes were presented in Venn diagram.

(2021) performed scRNA-seq and scTCR-seq cells of follicular helper T cells in a mouse model of autoantibody-mediated disease, they found that a few TCR clonotypes were preferentially shared among autoimmune follicular helper T cells and the amplification correlated with differential genetic signatures in autoimmune disease. These studies have yielded paired TCR information that complements and confirms previous studies, combining transcriptome analysis with corresponding single cells provides a comprehensive definition of the immune cell population that can provide a more accurate basis for downstream functional experiments. In the recent COVID-19 outbreak, we also observed the link of autoimmune phenotypes to SARS-CoV-2 infection in children using of scTCR-seq. Multisystemic inflammatory syndrome in children (MIS-C) is a life-threatening post-infection complication that occurs unpredictably weeks after mild or asymptomatic SARS-CoV-2 infection. Patients with clinically severe MIS-C exhibit a skewed memory T cell TCR repertoire and endothelial-reactive IgG autoantibodies. Using scRNA-seq, Ramaswamy et al. (2021) analyzed PBMC from patients and found that $CD4^+$ and $CD8^+$ memory T cells expressing TRBV11-2 were amplified in severe MIS-C.

The read length of RNA-seq is much shorter than that of first-generation sequencing (e.g., Sanger sequencing), and scRNA-seq data often contain many missing values or dropouts due to the failure to amplify the original RNA input, this frequency depends mainly on the protocol. Thus, it is crucial to use appropriate methods to overcome this problem when analyzing samples. Even so, the current scRNA-seq technology allows combining RNA-seq of the same cells with paired TCR-seq, and the great advantage of simultaneously processing cell numbers is essential for identifying the characteristics of rare T cell populations. These studies have generated new insights into disease biology and demonstrated the potential of scTCR-seq for clinical applications. Furthermore, in autoimmune diseases, identifying auto-reactive T cells by scTCR-seq may provide an indirect method to identify autoantigens.

2.3 TCR-pMHC sequencing potential based on long-read scRNA-seq

Long-read sequencing platforms, namely, long-read cDNA and long-read RNA sequencing, can capture many full-length transcripts (1–50 kb), unlike short-read sequencing, which requires fragmentation and amplification as well as introduced the previously discussed bias; additionally, assembly with bioinformatic tools relies on an existing genomic database (Salzberg and Yorke, 2005) (Figure 1), the computational approach for *de novo* transcriptome analysis utilized by long-read sequencing is easier and more unbiased (Stark et al., 2019) Processing the whole sample eliminates the amplification bias and has the ability to detect large insertions/deletions and duplicate regions. The two most widely used commercial technologies are Pacific Biosciences' single molecule real-time (SMRT) sequencing (average read length of HiFi reads ~20 kb, accuracy >99.9%) and Oxford Nanopore Technologies' nanopore sequencing (average read length of ultra-long reads ~100 kb, accuracy of R10.4 ~99%) (Method of the Year 2022: long-read sequencing., 2023). Specifically, no reverse transcription step is required in long-read RNA-seq, the PCR-free library-building protocol avoids guanine-cytosine (GC) bias and makes long-read sequencing platforms well-suited for studies of immune repertoires, as well as HLAs.

Predictably, scRNA-seq combined with long-read RNA-seq can provide higher sensitivity and accurate full-length paired TCR sequences. Singh et al. (2019) combined targeted capture and long-read TCR and BCR mRNA transcription with short-read scRNA-seq to track the transcriptomic signature of expanded clonotypes from primary tumors and draining lymph nodes of breast cancer patients. Understanding gene regulation and function requires the ability to capture gene expression levels and

TABLE 3 scRNA-seq for identifying autoimmune disease-related TCR.

Disease	Finding	References
T1D	Islet-specific glucose-6-phosphatase catalytic subunit-related protein $(IGRP)_{206-214}$-reactive $CD8^+$ T (self-reactive T cells) in T1D may inherently have a restricted TCR library as well as a substantial TCR motif overlap	Kasmani et al. (2022)
T1D	It has identified a class of autoreactive TCRs from human IAR (islet antigen reactive) $CD4^+$ T cells in patients with T1D that share the feature of germline alpha chains	Linsley et al. (2021)
AIH	Identified the central memory $CD45RA^-CD27^+PD\text{-}1^+CXCR5^-CCR6^-$ $CD4^+$ T cell population as the significant self-reactive $CD4^+$ T cell pool in AIH	Renand et al. (2020)
pSjD	T cell receptor alpha and beta chain variable genes of TRAV13-2 and TRBV7-9 were significantly expanded in patients with pSjD	Hong et al. (2020)
pSjD	The degree of TCR clonal expansion did not differ significantly between the pSjD patients and healthy individuals. Still, the frequencies of T cells with dual TCR β-chain expression were reduced considerably in pSjD patients	Hou et al. (2022)
SLE	Found $CD8^+$ kidney-infiltrating T cells (KIT) first existed in a transitional state, then clonally expanded and evolved to depletion in the kidney	Smita et al. (2022)

T1D, Type 1 diabetes; AIH, Autoimmune hepatitis; pSjD, Primary Sjögren's disease; SLE, Systemic lupus erythematosus.

TABLE 4 TCR (-pMHC) modeling platforms and capabilities.

Platform	TCRα	TCRβ	CDR3	MHC present	Output	Limitations
TCRModel	Yes	Yes	Yes	Yes	Unbound TCR, as well as TCR-pMHC complex modeling	NA
NetTCR	No	No	Yes	HLA-A*02:01 Only	List of predicted epitope binding	Only select from three peptide sequences
SCALOP-TCR	Yes	Yes	Yes	No	Predicts the structure of five CDRs (B1-2 and A1-3)	Does not include side chains
TCRBuilder	Yes	Yes	Yes	No	Multiple predicted conformations and an ensemble conformation would be returned	NA
TCRex	No	Yes	Yes	No	List of reactive epitopes	Only select from 93 viral and 5 cancer epitopes
TCRpMHCmodels	Yes	Yes	Yes	Class I only	TCR-pMHC complex modeling	Class I Only

isoform diversity at the single-cell level, in which short-read RNA-seq is limited in its capacity. Using Oxford Nanopore MinION sequencer to analyze individual murine B1a cells, Byrne et al. (2017) analyzed and identified different uses of complex isoforms in over a hundred genes, including surface receptors that determine B cell identity-determining surface receptors (e.g., CD19, CD20, and IGH). Multiple studies to date have shown that certain TCR clonotypes were expanded in the PBMCs or tissues of patients with autoimmune diseases. Still, the link between these TCRs and their functional relevance in the disease onset and development has not been identified, which requires refined studies of the gene transcriptome and the isoforms of TCR-expressing T cells. Thus, although there is no current application of long-read RNA-seq in autoimmune diseases, its future help in identifying complex etiologies can be foreseen.

Another promising application of long-read RNA-seq is in the field of HLAs. HLAs are a group of related proteins encoded by the MHC gene on human chromosome 6 and plays an essential role in autoimmune diseases. Previous *in silico* studies in our lab have shown that peptides with similar amino acid patterns may be presented to the same HLA due to structural similarities, thus initiating the autoimmune cascade (Gupta et al., 2022). Even though several analysis tools were developed to perform HLA typing from short RNA-seq reads using whole transcriptome data (Boegel et al., 2012; Kim and Pourmand, 2013; Buchkovich et al., 2017; Orenbuch et al., 2020; Chelysheva et al., 2021; Johansson et al., 2021), the large HLA genes (more than 5 kb) and the high degree of polymorphism within the class I (HLA-A, -B, and -C) and class II HLA (HLA-DR, -DQ, and -DP) often leads to ambiguous results in allele assignment. To this end, Cornaby et al. (2022) used long-read long sequencing of UMI-based high-resolution HLA typing and transcript quantification with a 99.68% overall HLA typing accuracy. Determining the profile of autoimmune-associated T cells requires deciphering the TCR and the HLA linkage. Thus, the introduction of long-read RNA-seq with the currently available scRNA-seq technology should allow a more in-depth study of innate, humoral, and T cell-mediated immunity in the future and will help provide a roadmap linking the pathogenesis of autoimmune diseases to the host immune response.

3 Structure study based on RNA sequencing results in TCR

3.1 Analysis of TCR-seq data

Retrieval of transcriptomic data enables the interrogation of multiple parameters simultaneously. More importantly, it allows for

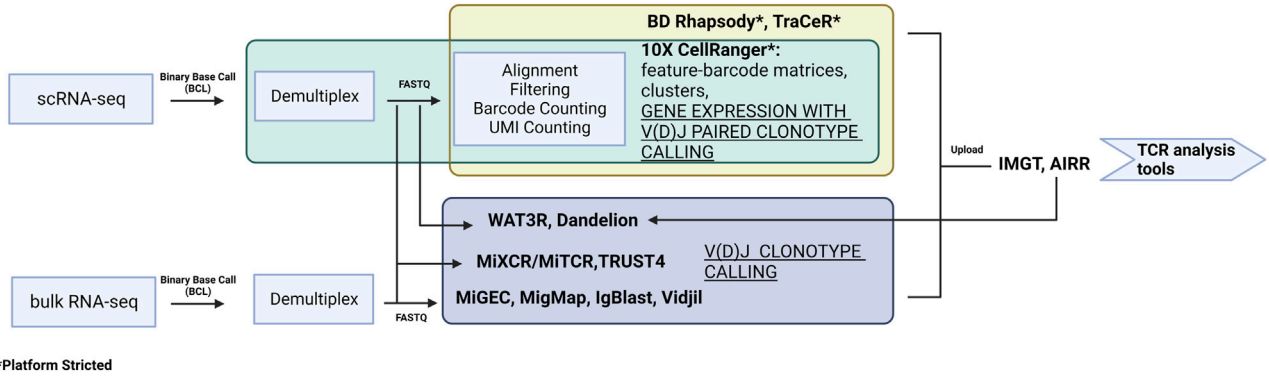


FIGURE 3
A summary of the pre-processing software for data from both scRNA-seq and bulk RNA-seq. The preprocessing relates to experimental design and library construction procedure and therefore can only follow a specific pipeline.

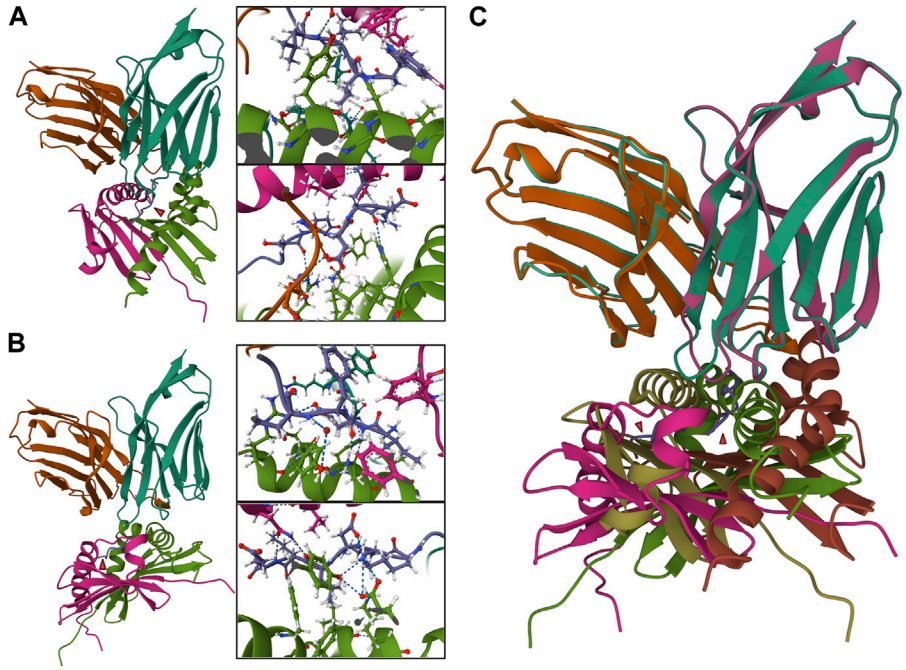


FIGURE 4
TCR-pMHC model containing autoantigens presented by SjD-susceptible HLA to the TCRs of selected patients. Two different peptides were selected to be presented by a SjD-susceptible HLA (DRA*01/DRB1*0301) to a selected paired TCR from SjD patient (VA 12-2, JA 13, CDR3A: CAVRIGGYQKVTF; VB 3-1, JB 2-3, CDR3B: CASSQEGREGRNTQYF). In molecular docking, a nine amino acid peptide is predicted to bind. The intermolecular contacts of this peptide are presented in the corresponding predictive model diagrams (aa1–4, upper, in order to clearly present the side chain amino acid order from right to left; aa5–9, lower, order from left to right). **(A)** NPWLILSEDRRQVRL, "WLILSEDRR" is predicted to bind. **(B)** FTFIQFKKDLKESMK, "IQFKKDLKE" is predicted to bind. **(C)** Superposition of the two TCR-pMHC models **(A,B)** showed a divergent presentation pattern. In **(A,B)**, Pink and green: HLA-DRA*01/DRB1*0301. Turquoise and orange: TCRα and ß chains. Purple: peptide.

the examination of a targeted objective, e.g., the expression of novel TCR repertories in specific T cell subsets that are clinically detrimental in an autoimmune disease. Recent technology and platforms enable users to follow the analytical pipelines to generate meaningful results from transcriptomics to predictive structural modeling. Raw data needs to be pre-processed before it can be applied to the downstream TCR analysis (Figure 3). Depending on the platform used [e.g., 10X CellRanger for 10X Genomics, BD Rhapsody, TraCeR (Stubbington et al., 2016) for Fluidigm C1], the raw datasets are processed slightly differently but all generate expression matrix with TCR output files. There are also tools that specialize in extracting only repertoire information from

FASTQ files. For example, MiXCR/MiTCR (Bolotin et al., 2015) and TRUST4 (Song et al., 2021) can process data from bulk RNA-seq and scRNA-seq data with and without V(D)J enrichment. MiGEC (Shugay et al., 2014), MigMap (Shugay and Davenport, 2018), IgBlast (Ye et al., 2013), and Vidjil (Giraud et al., 2014) can only work on bulk RNA-seq. Dandelion (Suo et al., 2023) is designed to work with Adaptive Immune Receptor Repertoire (AIRR) (Rubelt et al., 2017) -formatted input or 10X CellRanger VDJ output. WAT3R (Ainciburu et al., 2022) can process on 3′ single-cell RNA-seq data without V(D)J enrichment. Due to the high cost of library preparation and sequencing, there are also public databases containing V(D)J sequence information available for use, such as international ImMunoGeneTics information system (IMGT) (Lefranc et al., 2015) and AIRR. There is growing number of bioinformatic tools for TCR analysis. The output formats from pre-processing are different and the available downstream software varies, but most TCR analysis tools can recognize multiple formats.

Scirpy (Sturm et al., 2020) is a Python package that is an extension of Scanpy, which enables the user to visualize single-cell immune libraries and integrate them with transcriptomic data to characterize the TCR of single T cells. Scirpy supports multiple data formats, including 10X CellRanger, BD Rhapsody, TraCeR, Dandelion, or AIRR-compatible data. Scirpy enables the study of TCR chain configurations and explores clonotypes' abundance, diversity, expansion, and overlap across samples, patients, or cell clusters. This software also allows analysis of CDR3 sequence length and the distribution of V(D)J gene usage. Specifically, Scirpy implements a sequence-alignment-based network that enables the clustering of cells into clonotypes based on having identical/similar CDR3 amino acid sequences, which offers the opportunities to identify cells that might recognize the same antigens.

Immunarch (Samokhina et al., 2022) is an R package which accepts all standard immuno-sequencing formats. It also automatically detects and parses uploaded data in formats including ImmunoSEQ, IMGT, MiXCR/MiTCR, MiGEC, MigMap, VDJtools, AIRR, and 10X CellRanger. Immunarch can annotate clonotypes using an external immune receptor database. The exclusive features include basic statistics such as CDR3 length distribution and clonotype abundance-more specifically, it can calculate the distribution of clonotypes per CDR3 length or clonotype spectratype. It can complete the analysis of repertoires dynamics, diversity, clonality, and overlap as well as compute V/J gene usage, and the distributions of kmers and sequence profiles.

The Loupe V(D)J Browser [10x Genomics Loupe V(D)J Browser 3.0.0] is a desktop application for Windows and macOS that allows users to analyze, search, and visualize V(D)J sequences and clonotypes. The Loupe V(D)J Browser identifies a sample's most common paired αβ TCR chains. It filters clonotypes based on their antigen specificity or UMI number per antigen andexamines full-length V, D, and J amino acid and nucleotide sequences to detect variants in V(D)J transcripts, motifs within CDR3 regions, and compares clonotype frequencies between samples. It can be integrated with the Loupe Browser (formerly Loupe Cell Browser) to analyze data from different 10X genomics solutions. However, this tool has drawbacks since it is specifically designed to analyze 10X Genomics Single Cell Immune Profiling dataset.

ImmunoSEQ Analyzer (Adaptive Biotechnologies ImmunoSEQ Analyzer 3.0) is an online web-based tool for data exploration. Since the platform was developed only for ImmunoSEQ, it directly identifies V, D, and J genes and whole nucleotide sequences; non-productive sequences can be filtered out, and specific data values for immune sequencings, such as clonality, can be precomputed and visualized directly on the dashboard. Like Loupe V(D)J Browser, it provides basic statistics of clonotypes. In addition, the analyzer has tools for performing additional statistical tests and metrics on immune sequencing data. These include tools for clonotype diversity and tracking among samples. The main advantage of using this analyzer is that it contains an extensive database of TCR sequences, integrating millions of public data sequences and control samples.

VDJtools (Shugay et al., 2015) is an open-source software framework for TCR analysis based on Java. It is mainly used for post-analysis of clonotypes containing VDJ junction output for the following platforms: MiXCR/MiTCR, MiGEC, IgBlast, IMGT, ImmunoSEQ, VDJdb, Vidjil, MiXCR, ImmunoSEQ, and 10X CellRanger. VDJtools enables visualization of basic and advanced immune repertoires by applying different methods and strategies, including basic segment and segment usage, repertoire overlap, diversity analysis, data joining and clonotype tracking, and repertoire clustering.

scRepertoire (Borcherding et al., 2020) is an R package compatible and integrated with the R packages Trex for deep-learning-based autoencoding of TCR, which supports 10X CellRanger, AIRR, WAT3R, and TRUST4. scRepertoire is designed to obtain filter contig output from the pipeline, assign clonotypes according to the two TCR chains, and analyze the dynamics of clonotypes. It can be used for clonotype visualization, analysis of unique clonotypes, or clonal space quantification. Further features include clonal proportion analysis, sample similarity measures (scatter comparison between two samples), and overlap analysis for two or more samples. A unique feature is that the output data can be integrated with transcriptomic data [using Seurat (Satija et al., 2015), SingleCellExperiment (Amezquita et al., 2020), or Monocle 3 (Trapnell et al., 2014)].

There are also interactive databases available with known TCR sequences and clonotypes that can identify shared clones in multiple samples and explore the specificity of the immune response. An example is VDJdb (Shugay et al., 2018), a TCR sequence database with known antigenic specificity. The main goal of VDJdb is to facilitate access to information on the antigenic specificity of existing TCRs, i.e., the ability to identify certain epitopes in a specific MHC context. This database, which has been collecting and managing publicly available sequencing data obtained from TCRs with well-defined antigenic specificity, as well as data voluntarily shared by researchers, has been extended to a web interface that allows bulk querying of the AIRR dataset and identification of TCR sequence motifs associated with specific epitopes. There is also tcrdist3 (Mayer-Blackwell et al., 2021), an open-source python package based on distance-based TCR repertoire analysis capable of performing extensive TCR sequence analysis, including diversity analysis. The software utilizes meta-cloning concepts to group TCRs, i.e., a set of TCRs that are biochemically similar and likely to recognize the same antigen. The package has extended this to include support for gamma-delta TCRs.

Given these innovative tools, the challenges persist. Whether it is a pre-processing platform or a TCR analysis software, it is difficult for users to reach a uniform standard due to the many options, especially since most platforms can perform the same functions. Generally, the pre-processing relates to experimental design and library construction procedure following a specific pipeline (i.e., single-cell or bulk, sequencing platform, UMI integration). Therefore, it is imperative to develop a standard or "universal" pipeline that could support and simplify the process. Furthermore, most of the software is programming-based, which makes it necessary for users to have basic programming skills to operate and manipulate. A few available web or application-based platforms, which can meet the basic research needs, limit the ability to customize, and are not open-sourced or strictly product-based. Hence, these challenges are some of the major impediments that may discourage researchers interested in applying these tools for their research.

3.2 3D structural modeling

There is an array of TCR modeling platforms and capabilities including, but not limited to: Structural T Cell Receptor Modelling Tool (STCRPred) (Leem et al., 2018; Wong et al., 2019; Wong et al., 2020), TCRModel (Gowthaman and Pierce, 2018), and NetTCR (Montemurro et al., 2021) (Table 4). The former is a platform connected to SAbPred (Dunbar et al., 2016), initially constructed for 3D modeling and optimization of the B cell receptor (BCR), which also provides many of the same capabilities through SCALOP-TCR and TCRBuilder. SCALOP (Sequence-based Prediction of TCR CDR Canonical Form)-TCR is a sequence-based canonical form predictor for five of the six complementarity-determining regions (B1, B2, A1, A2, and A3) on a TCR. This provides an essential framework loop structure omitting side chains, compared to TCRBuilder, which may be more practical and include those interactions. TCRModel uses two modes: TCR-pMHC complex modeling (further discussed below) and unbound TCR modeling. The latter allows a simple model of the TCR, complete with any mutations, or by simply inputting the CDR3 sequences into the germline genes. Rosati et al. (2022) recently utilized this technology to model Crohn-associated invariant T (CAIT) cells with the paired TCR chain, which had been identified as an NKT type II population in Crohn's Disease patients. NetTCR is a very limited platform. However, it may be helpful if the following criteria are met: known CDR3 sequence, satisfied with the provided three peptide sequences, and MHC-1 prediction will be exclusively for HLA-A*02:01 (Reynisson et al., 2020); while not strictly within the scope of TCR sequencing, MHC modeling can predict peptides to be presented to the TCR. This may be a useful tool within autoimmunity if the HLA is well known, as it is in diabetes. Notably, there also exist customized programs; for example, Jokinen et al. (2021) created TCRGP with which they were able to identify an exhausted, low functional T cell cluster that was enriched with Hepatitis B virus-targeting clonotypes, which they theorized could be pathogenic in causing hepatocellular carcinoma. Likewise, pipelines like this may be helpful in autoimmune disorders, especially those with a proposed viral or bacterial etiology.

3.3 Epitope prediction

Several programs have been written to predict what TCR will react against a given antigen. Programs predicting how epitopes dock in a TCR are limited but growing significantly recently (Table 4). The aforementioned TCRex is a platform that allows for selection from 93 viral and five cancer epitopes (Gielis et al., 2019). This platform enables users to train their custom model with machine learning, which is dependent on a manually curated catalog of pathology-associated TCR sequences (McPAS-TCR) (Tickotsky et al., 2017), VDJ database (VDJdb) (Shugay et al., 2018), and the ImmuneCODE-database (Nolan et al., 2020). For this platform and those to follow, splitting known autoantigens into shorter peptides and artificially docking those peptides may be the most useful. However, if the approach is to understand the etiology of the pathogenesis of the disease and potential triggers, viral/bacterial epitope mapping may also be useful. In this case nucleotide Basic Local Alignment Search Tool (BLASTn) may be a useful tool (Ladunga, 2002). TCRpMHCmodels is a tool for 3D modeling TCRs bound to peptides presented by a MCH class I (Jensen et al., 2019). Likewise, TCRmodel's TCR-pMHC complex modeling is a very useful tool to either look at the interaction with a user-supplied peptide docked on a chosen MHC (either Class I or II) for both humans and mice (Gowthaman and Pierce, 2018). Our group has used COOT and PHENIX to predict pathogenic autoantigens presented by SjD-susceptible HLA, which has previously relied on superimposing chains on the crystal structure of solved peptide/HLA complexes on a LINUX system (Gupta et al., 2022). Now with TCRmodel we could further analyze the TCR-pMHC complex of autoantigens presented by SjD-susceptible HLA to selected patient's TCR. With a web-based platform, this allows us to predict intermolecular contacts between peptide and HLA and cognate interactions between the TCR and peptide/HLA complex (Figure 4). While this technology has yet to be widely utilized in autoimmunity, Kasmani et al. (2023) used this program to show that $CD8^+$ TCR avidity correlates with an exhausted fate during persistent infection by lymphocytic choriomeningitis virus in mice, where TCR sequences were paired with the peptide KAVYNFATC and the mouse class I MHC $H\text{-}2D^b$.

Artificial intelligence (AI) has recently gained traction within the scientific community, and the epitope mapping field is no exception. Within the last 2 years, four new programs have been created: DECODE, TITAN, DeepTCR, and pMTnet. It should be noted that all of these programs utilize known biochemical reactivities (e.g., an amino acid present at specific residues as well as their interactions with the TCR and MHC/HLA). DECODE (DEcoding t Cell receptOr binDing rulEs) is a machine learning, customizable program that can allow users to select for specific reactivities (e.g., an amino acid at a particular residue) to further specify and customize the dataset for the end user (Papadopoulou et al., 2022). TITAN (Tcr epITope bimodal Attention Networks) is a bimodal neural network that explicitly encodes both TCR sequences and epitopes, which, interestingly, was able to identify previously unseen TCRs (Weber et al., 2021). The remaining two are more based on deep learning. DeepTCR analysis provides noise-depleted scRNA-Seq and *ex vivo*

T cell assay results, which enables the user to identify rare subsets of TCRs and novel epitopes (Sidhom et al., 2022). And pMTnet (pMHC-TCR binding prediction network) was built to predict TCR-binding to neoantigens in human tumor genomics datasets. Notably, this program only utilized the CDR3β sequence of the TCR, epitope sequence, and class I MHC allele (Lu et al., 2021). While these technologies have been restricted to oncogenic research, AI is becoming more available both within research environments and from private companies. Utilization of this technology may lead to the identification of novel pathogenic T cells with specific TCRs or novel autoantigens driving autoimmune disease pathology.

4 Discussion

The rapid advances in RNA-seq technology have enabled the analysis of the transcriptome in various ways, both serving to further the understanding of genome function and crucially for studying mRNA splicing and rearrangements. Many alternative sequencing platforms are currently available, and short-read RNA-seq combined with single-cell technology is currently the mainstay. However, the future of autoimmune disease research lies in efficient long-read RNA-seq. The sequence and rearrangement of TCR are closely related to the pathogenesis of autoimmune diseases, and HLA genes are well-documented genetic risk factors for the development of certain autoimmune diseases. While current studies focus on HLA typing, the clonal expansion of the immune repertoire, or CDR3 motifs in patients (differentiating from healthy individuals), in the future, by sequencing individual T cells, we will not only be able to obtain the sequence of TCRs, but we will also be able to obtain transcriptomic data of T cells expressing TCRs, from which we can analyze the subtypes of cells. Combined with accurate HLA typing and artificial intelligence (AI)-based structural analysis, we can predict autoimmune TCR-pMHC complexes even before the onset of the disease. Identifying the autoantigen and TCR repertoire and generating a predictive autoimmune response will have a significant potential for clinical applications and also advances our knowledge of autoimmune diseases. More importantly, the approach will bring tremendous potential in infectious diseases, from which we can optimize vaccine development to target individual antigen-specific TCR enhancements. The main issues currently hindering the adoption of long-read sequencing are the increased cost per base and the higher error rate compared to short-read sequencing. Unlike short-read sequencing where errors are usually clustered at both ends of the read, long-read sequencing errors are random and can be effectively corrected by multiple sequencing events. Still, these issues will gradually be overcome as technology advances. With the vigorous development of the RNA field, multidisciplinary research can bring breakthroughs in studying autoimmune diseases.

Author contributions

XL and AR wrote the introduction. YS wrote Section 2 and 3.1. AV wrote Section 3.2 and 3.3. YS and CQN conceptualized and edited the final version of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

CQN was supported financially in part by the National Institutes of Health (NIH), National Institute of Dental and Craniofacial Research (NIDCR), Division of Extramural Research (DE028544, DE028544-02S1).

Acknowledgments

Figures 1–3 were created with BioRender.com. Figure 4 was create with Mol* Viewer under RCSB PDB.

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

Ainciburu, M., Morgan, D. M., DePasquale, E. A. K., Love, J. C., Prósper, F., and van Galen, P. (2022). WAT3R: Recovery of T-cell receptor variable regions from 3' single-cell RNA-sequencing. *Bioinformatics* 38, 3645–3647. doi:10.1093/bioinformatics/btac382

Akama-Garren, E. H., van den Broek, T., Simoni, L., Castrillon, C., van der Poel, C. E., and Carroll, M. C. (2021). Follicular T cells are clonally and transcriptionally distinct in B cell-driven mouse autoimmune disease. *Nat. Commun.* 12, 6687. doi:10.1038/s41467-021-27035-8

Amezquita, R. A., Lun, A. T. L., Becht, E., Carey, V. J., Carpp, L. N., Geistlinger, L., et al. (2020). Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* 17, 137–145. doi:10.1038/s41592-019-0654-x

Amoriello, R., Greiff, V., Aldinucci, A., Bonechi, E., Carnasciali, A., Peruzzi, B., et al. (2020). The TCR repertoire reconstitution in multiple sclerosis: Comparing one-shot and continuous immunosuppressive therapies. *Front. Immunol.* 11, 559. doi:10.3389/fimmu.2020.00559

Baran-Gale, J., Chandra, T., and Kirschner, K. (2018). Experimental design for single-cell RNA sequencing. *Brief. Funct. Genomics* 17, 233–239. doi:10.1093/bfgp/elx035

Boegel, S., Löwer, M., Schäfer, M., Bukur, T., de Graaf, J., Boisguérin, V., et al. (2012). HLA typing from RNA-Seq sequence reads. *Genome Med.* 4, 102. doi:10.1186/gm403

Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., et al. (2015). MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* 12, 380–381. doi:10.1038/nmeth.3364

Borcherding, N., Bormann, N. L., and Kraus, G. (2020). scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Res* 9, 47. doi:10.12688/f1000research.22139.2

Buchkovich, M. L., Brown, C. C., Robasky, K., Chai, S., Westfall, S., Vincent, B. G., et al. (2017). HLAProfiler utilizes k-mer profiles to improve HLA calling accuracy for rare and common alleles in RNA-seq data. *Genome Med.* 9, 86. doi:10.1186/s13073-017-0473-6

Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., et al. (2017). Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* 8, 16027. doi:10.1038/ncomms16027

Chang, C.-M., Hsu, Y.-W., Wong, H. S.-C., Wei, J. C.-C., Liu, X., Liao, H.-T., et al. (2019). Characterization of T-cell receptor repertoire in patients with rheumatoid arthritis receiving biologic therapies. *Dis. Markers* 2019, 2364943. doi:10.1155/2019/2364943

Chelysheva, I., Pollard, A. J., and O'Connor, D. (2021). RNA2HLA: HLA-based quality control of RNA-seq datasets. *Brief. Bioinforma.* 22, bbab055. doi:10.1093/bib/bbab055

Cook, L., Munier, C. M. L., Seddiki, N., Hardy, M. Y., Anderson, R. P., Zaunders, J., et al. (2020). Circulating gluten-specific, but not CMV-specific, CD39+ regulatory T cells have an oligoclonal TCR repertoire. *Clin. Transl. Immunol.* 9, e1096. doi:10.1002/cti2.1096

Cornaby, C., Montgomery, M. C., Liu, C., and Weimer, E. T. (2022). Unique molecular identifier-based high-resolution HLA typing and transcript quantitation using long-read sequencing. *Front. Genet.* 13, 901377. doi:10.3389/fgene.2022.901377

Cornec, D., Jamin, C., and Pers, J.-O. (2014). Sjögren's syndrome: Where do we stand, and where shall we go? *J. Autoimmun.* 51, 109–114. doi:10.1016/j.jaut.2014.02.006

Correia-Neves, M., Waltzinger, C., Mathis, D., and Benoist, C. (2001). The shaping of the T cell repertoire. *Immunity* 14, 21–32. doi:10.1016/s1074-7613(01)00086-3

Deschler, K., Rademacher, J., Lacher, S. M., Huth, A., Utzt, M., Krebs, S., et al. (2022). Antigen-specific immune reactions by expanded CD8+ T cell clones from HLA-B*27-positive patients with spondyloarthritis. *J. Autoimmun.* 133, 102901. doi:10.1016/j.jaut.2022.102901

Dunbar, J., Krawczyk, K., Leem, J., Marks, C., Nowak, J., Regep, C., et al. (2016). SAbPred: A structure-based antibody prediction server. *Nucleic Acids Res.* 44, W474–W478. doi:10.1093/nar/gkw361

Gielis, S., Moris, P., Bittremieux, W., De Neuter, N., Ogunjimi, B., Laukens, K., et al. (2019). Detection of enriched T cell epitope specificity in full T cell receptor sequence repertoires. *Front. Immunol.* 10, 2820. doi:10.3389/fimmu.2019.02820

Giraud, M., Salson, M., Duez, M., Villenet, C., Quief, S., Caillault, A., et al. (2014). Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. *BMC Genomics* 15, 409. doi:10.1186/1471-2164-15-409

Goetz, J. J., and Trimarchi, J. M. (2012). Transcriptome sequencing of single cells with Smart-Seq. *Nat. Biotechnol.* 30, 763–765. doi:10.1038/nbt.2325

Gowthaman, R., and Pierce, B. G. (2018). TCRmodel: High resolution modeling of T cell receptors from sequence. *Nucleic Acids Res.* 46, W396–W401. doi:10.1093/nar/gky432

Gupta, S., Li, D., Ostrov, D. A., and Nguyen, C. Q. (2022). Epitope mapping of pathogenic autoantigens on Sjögren's syndrome-susceptible human leukocyte antigens using in silico techniques. *J. Clin. Med.* 11, 1690. doi:10.3390/jcm11061690

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-seq: Single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673. doi:10.1016/j.celrep.2012.08.003

Hayashi, Y., Hamano, H., Haneji, N., Ishimaru, N., and Yanagi, K. (1995). Biased T cell receptor V beta gene usage during specific stages of the development of autoimmune sialadenitis in the MRL/lpr mouse model of Sjögren's syndrome. *Arthritis Rheum.* 38, 1077–1084. doi:10.1002/art.1780380809

Helmick, C. G., Felson, D. T., Lawrence, R. C., Gabriel, S., Hirsch, R., Kwoh, C. K., et al. (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis Rheum.* 58, 15–25. doi:10.1002/art.23177

Hong, X., Meng, S., Tang, D., Wang, T., Ding, L., Yu, H., et al. (2020). Single-cell RNA sequencing reveals the expansion of cytotoxic CD4+ T lymphocytes and a landscape of immune cells in primary Sjögren's syndrome. *Front. Immunol.* 11, 594658. doi:10.3389/fimmu.2020.594658

Hou, X. L., Wang, L., Ding, Y. L., Xie, Q., and Diao, H. Y. (2016). Current status and recent advances of next generation sequencing techniques in immunological repertoire. *Genes Immun.* 17, 153–164. doi:10.1038/gene.2016.9

Hou, X., Hong, X., Ou, M., Meng, S., Wang, T., Liao, S., et al. (2022). Analysis of gene expression and TCR/B cell receptor profiling of immune cells in primary Sjögren's syndrome by single-cell sequencing. *J. Immunol.* 209, 238–249. doi:10.4049/jimmunol.2100803

Huang, H., Sikora, M. J., Islam, S., Chowdhury, R. R., Chien, Y.-H., Scriba, T. J., et al. (2019). Select sequencing of clonally expanded CD8+ T cells reveals limits to clonal expansion. *Proc. Natl. Acad. Sci. U. S. A.* 116, 8995–9001. doi:10.1073/pnas.1902649116

Jensen, K. K., Rantos, V., Jappe, E. C., Olsen, T. H., Jespersen, M. C., Jurtz, V., et al. (2019). TCRpMHCmodels: Structural modelling of TCR-pMHC class I complexes. *Sci. Rep.* 9, 14530. doi:10.1038/s41598-019-50932-4

Joachims, M. L., Leehan, K. M., Lawrence, C., Pelikan, R. C., Moore, J. S., Pan, Z., et al. (2016). Single-cell analysis of glandular T cell receptors in Sjögren's syndrome. *JCI Insight* 1, e85609. doi:10.1172/jci.insight.85609

Johansson, T., Yohannes, D. A., Koskela, S., Partanen, J., and Saavalainen, P. (2021). HLA RNA sequencing with unique molecular identifiers reveals high allele-specific variability in mRNA expression. *Front. Immunol.* 12, 629059. doi:10.3389/fimmu.2021.629059

Jokinen, E., Huuhtanen, J., Mustjoki, S., Heinonen, M., and Lähdesmäki, H. (2021). Predicting recognition between T cell receptors and epitopes with TCRGP. *PLoS Comput. Biol.* 17, e1008814. doi:10.1371/journal.pcbi.1008814

Kasmani, M. Y., Ciecko, A. E., Brown, A. K., Petrova, G., Gorski, J., Chen, Y.-G., et al. (2022). Autoreactive CD8 T cells in NOD mice exhibit phenotypic heterogeneity but restricted TCR gene usage. *Life Sci. Alliance* 5, e202201503. doi:10.26508/lsa.202201503

Kasmani, M. Y., Zander, R., Chung, H. K., Chen, Y., Khatun, A., Damo, M., et al. (2023). Clonal lineage tracing reveals mechanisms skewing CD8+ T cell fate decisions in chronic infection. *J. Exp. Med.* 220, e20220679. doi:10.1084/jem.20220679

Kassan, S. S., and Moutsopoulos, H. M. (2004). Clinical manifestations and early diagnosis of Sjögren syndrome. *Arch. Intern. Med.* 164, 1275–1284. doi:10.1001/archinte.164.12.1275

Khor, B., and Sleckman, B. P. (2002). Allelic exclusion at the TCRβ locus. *Curr. Opin. Immunol.* 14, 230–234. doi:10.1016/S0952-7915(02)00326-6

Kim, H. J., and Pourmand, N. (2013). HLA typing from RNA-seq data using hierarchical read weighting [corrected]. *PLoS One* 8, e67885. doi:10.1371/journal.pone.0067885

Ladunga, I. (2002). Finding homologs to nucleotide sequences using network BLAST searches. *Curr. Protoc. Bioinforma.* Chapter 3, Unit 3.3. doi:10.1002/0471250953.bi0303s00

Leem, J., de Oliveira, S. H. P., Krawczyk, K., and Deane, C. M. (2018). STCRDab: The structural T-cell receptor database. *Nucleic Acids Res.* 46, D406–D412. doi:10.1093/nar/gkx971

Lefranc, M.-P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., et al. (2015). IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res.* 43, D413–D422. doi:10.1093/nar/gku1056

Lin, Y.-H., Hung, S.-J., Chen, Y.-L., Lin, C.-H., Kung, T.-F., Yeh, Y.-C., et al. (2020). Dissecting efficiency of a 5' rapid amplification of cDNA ends (5'-RACE) approach for profiling T-cell receptor beta repertoire. *PLoS One* 15, e0236366. doi:10.1371/journal.pone.0236366

Linsley, P. S., Barahmand-pour-Whitman, F., Balmas, E., DeBerg, H. A., Flynn, K. J., Hu, A. K., et al. (2021). Autoreactive T cell receptors with shared germline-like α chains in type 1 diabetes. *JCI Insight* 6, e151349. doi:10.1172/jci.insight.151349

Liu, X., Zhang, W., Zeng, X., Zhang, R., Du, Y., Hong, X., et al. (2016). Systematic comparative evaluation of methods for investigating the tcrβ repertoire. *PLoS One* 11, e0152464. doi:10.1371/journal.pone.0152464

Lu, T., Zhang, Z., Zhu, J., Wang, Y., Jiang, P., Xiao, X., et al. (2021). Deep learning-based prediction of the T cell receptor-antigen binding specificity. *Nat. Mach. Intell.* 3, 864–875. doi:10.1038/s42256-021-00383-2

Mamedov, I. Z., Britanova, O. V., Zvyagin, I. V., Turchaninova, M. A., Bolotin, D. A., Putintseva, E. V., et al. (2013). Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front. Immunol.* 4, 456. doi:10.3389/fimmu.2013.00456

Mayer-Blackwell, K., Schattgen, S., Cohen-Lavi, L., Crawford, J. C., Souquette, A., Gaevert, J. A., et al. (2021). TCR meta-clonotypes for biomarker discovery with tcrdist3 enabled identification of public, HLA-restricted clusters of SARS-CoV-2 TCRs. *Elife* 10, e68605. doi:10.7554/eLife.68605

Mazzotti, L., Gaimari, A., Bravaccini, S., Maltoni, R., Cerchione, C., Juan, M., et al. (2022). T-cell receptor repertoire sequencing and its applications: Focus on infectious diseases and cancer. *Int. J. Mol. Sci.* 23, 8590. doi:10.3390/ijms23158590

Method of the Year 2022: long-read sequencing (2023). Method of the year 2022: Long-read sequencing. *Nat. Methods* 20, 1. doi:10.1038/s41592-022-01759-x

Mitchell, A. M., and Michels, A. W. (2020). T cell receptor sequencing in autoimmunity. *J. Life Sci. (Westlake Village)* 2, 38–58. doi:10.36069/jols/20201203

Montemurro, A., Schuster, V., Povlsen, H. R., Bentzen, A. K., Jurtz, V., Chronister, W. D., et al. (2021). NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCRα and β sequence data. *Commun. Biol.* 4, 1060. doi:10.1038/s42003-021-02610-3

Muraro, P. A., Robins, H., Malhotra, S., Howell, M., Phippard, D., Desmarais, C., et al. (2014). T cell repertoire following autologous stem cell transplantation for multiple sclerosis. *J. Clin. Invest.* 124, 1168–1172. doi:10.1172/JCI71691

Murphy, K. M., and Weaver, C. (2022). *Janeway's immunobiology*. 10th Edition. W.W. Norton & Company.

Nguyen, C. Q., and Peck, A. B. (2009). Unraveling the pathophysiology of Sjogren syndrome-associated dry eye disease. *Ocul. Surf.* 7, 11–27. doi:10.1016/S1542-0124(12)70289-6

Nocturne, G., and Mariette, X. (2015). Sjögren syndrome-associated lymphomas: An update on pathogenesis and management. *Br. J. Haematol.* 168, 317–327. doi:10.1111/bjh.13192

Nolan, S., Vignali, M., Klinger, M., Dines, J. N., Kaplan, I. M., Svejnoha, E., et al. (2020). A large-scale database of T-cell receptor beta (TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. *Res. Sq.* doi:10.21203/rs.3.rs-51964/v1

Orenbuch, R., Filip, I., Comito, D., Shaman, J., Pe'er, I., and Rabadan, R. (2020). arcasHLA: high-resolution HLA typing from RNAseq. *Bioinformatics* 36, 33–40. doi:10.1093/bioinformatics/btz474

Padovan, E., Casorati, G., Dellabona, P., Meyer, S., Brockhaus, M., and Lanzavecchia, A. (1993). Expression of two T cell receptor alpha chains: dual receptor T cells. *Science* 262, 422–424. doi:10.1126/science.8211163

Papadopoulou, I., Nguyen, A.-P., Weber, A., and Martínez, M. R. (2022). Decode: A computational pipeline to discover T cell receptor binding rules. *Bioinformatics* 38, i246–i254. doi:10.1093/bioinformatics/btac257

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098. doi:10.1038/nmeth.2639

Ramaswamy, A., Brodsky, N. N., Sumida, T. S., Comi, M., Asashima, H., Hoehn, K. B., et al. (2021). Immune dysregulation and autoreactivity correlate with disease severity in SARS-CoV-2-associated multisystem inflammatory syndrome in children. *Immunity* 54, 1083–1095.e7. doi:10.1016/j.immuni.2021.04.003

Renand, A., Cervera-Marzal, I., Gil, L., Dong, C., Garcia, A., Kervagoret, E., et al. (2020). Integrative molecular profiling of autoreactive CD4 T cells in autoimmune hepatitis. *J. Hepatol.* 73, 1379–1390. doi:10.1016/j.jhep.2020.05.053

Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 48, W449–W454. doi:10.1093/nar/gkaa379

Rosati, E., Rios Martini, G., Pogorelyy, M. V., Minervina, A. A., Degenhardt, F., Wendorff, M., et al. (2022). A novel unconventional T cell population enriched in Crohn's disease. *Gut* 71, 2194–2204. doi:10.1136/gutjnl-2021-325373

Rubelt, F., Busse, C. E., Bukhari, S. A. C., Bürckert, J.-P., Mariotti-Ferrandiz, E., Cowell, L. G., et al. (2017). Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data. *Nat. Immunol.* 18, 1274–1278. doi:10.1038/ni.3873

Salzberg, S. L., and Yorke, J. A. (2005). Beware of mis-assembled genomes. *Bioinformatics* 21, 4320–4321. doi:10.1093/bioinformatics/bti769

Samokhina, M., Popov, A., Nazarov, V. I., Rumynskiy, E., et al.Ivan-Immunomind; Immunarch.Bot (2022). *immunomind/immunarch: Immunarch 0.9.0*. Zenodo. doi:10.5281/zenodo.7446955

Sant'Angelo, D. B., Lucas, B., Waterbury, P. G., Cohen, B., Brabb, T., Goverman, J., et al. (1998). A molecular map of T cell development. *Immunity* 9, 179–186. doi:10.1016/s1074-7613(00)80600-7

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502. doi:10.1038/nbt.3192

Schatz, D. G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* 11, 251–263. doi:10.1038/nri2941

Seiringer, P., Garzorz-Stark, N., and Eyerich, K. (2022). T-Cell–Mediated autoimmunity: Mechanisms and future directions. *J. Invest. Dermatol.* 142, 804–810. doi:10.1016/j.jid.2021.04.032

Sheng, K., Cao, W., Niu, Y., Deng, Q., and Zong, C. (2017). Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat. Methods* 14, 267–270. doi:10.1038/nmeth.4145

Shugay, M., and Davenport, C. (2018). MiGMAP: Mapper for full-length T- and B-cell repertoire sequencing. Available at: https://github.com/mikessh/migmap (Accessed May 4, 2023).

Shugay, M., Britanova, O. V., Merzlyak, E. M., Turchaninova, M. A., Mamedov, I. Z., Tuganbaev, T. R., et al. (2014). Towards error-free profiling of immune repertoires. *Nat. Methods* 11, 653–655. doi:10.1038/nmeth.2960

Shugay, M., Bagaev, D. V., Turchaninova, M. A., Bolotin, D. A., Britanova, O. V., Putintseva, E. V., et al. (2015). VDJtools: Unifying post-analysis of T cell receptor repertoires. *PLoS Comput. Biol.* 11, e1004503. doi:10.1371/journal.pcbi.1004503

Shugay, M., Bagaev, D. V., Zvyagin, I. V., Vroomans, R. M., Crawford, J. C., Dolton, G., et al. (2018). VDJdb: A curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* 46, D419–D427. doi:10.1093/nar/gkx760

Sidhom, J.-W., Oliveira, G., Ross-MacDonald, P., Wind-Rotolo, M., Wu, C. J., Pardoll, D. M., et al. (2022). Deep learning reveals predictive sequence concepts within immune repertoires to immunotherapy. *Sci. Adv.* 8, eabq5089. doi:10.1126/sciadv.abq5089

Singh, M., Al-Eryani, G., Carswell, S., Ferguson, J. M., Blackburn, J., Barton, K., et al. (2019). High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* 10, 3120. doi:10.1038/s41467-019-11049-4

Skarstein, K., Wahren, M., Zaura, E., Hattori, M., and Jonsson, R. (1995). Characterization of T cell receptor repertoire and anti-Ro/SSA autoantibodies in relation to sialadenitis of NOD mice. *Autoimmunity* 22, 9–16. doi:10.3109/08916939508995294

Smita, S., Chikina, M., Shlomchik, M. J., and Tilstra, J. S. (2022). Heterogeneity and clonality of kidney-infiltrating T cells in murine lupus nephritis. *JCI Insight* 7, e156048. doi:10.1172/jci.insight.156048

Song, L., Cohen, D., Ouyang, Z., Cao, Y., Hu, X., and Liu, X. S. (2021). TRUST4: Immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* 18, 627–630. doi:10.1038/s41592-021-01142-2

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: The teenage years. *Nat. Rev. Genet.* 20, 631–656. doi:10.1038/s41576-019-0150-2

Stubbington, M. J. T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A. O., Dougan, G., et al. (2016). T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* 13, 329–332. doi:10.1038/nmeth.3800

Sturm, G., Szabo, T., Fotakis, G., Haider, M., Rieder, D., Trajanoski, Z., et al. (2020). Scirpy: A Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* 36, 4817–4818. doi:10.1093/bioinformatics/btaa611

Sumida, T., Kita, Y., Yonaha, F., Maeda, T., Iwamoto, I., and Yoshida, S. (1994a). T cell receptor V alpha repertoire of infiltrating T cells in labial salivary glands from patients with Sjögren's syndrome. *J. Rheumatol.* 21, 1655–1661.

Sumida, T., Sakamaki, T., Yonaha, F., Maeda, T., Namekawa, T., Nawata, Y., et al. (1994b). HLA-DR alleles in patients with Sjögren's syndrome over-representing V beta 2 and V beta 13 genes in the labial salivary glands. *Br. J. Rheumatol.* 33, 420–424. doi:10.1093/rheumatology/33.5.420

Suo, C., Polanski, K., Dann, E., Lindeboom, R. G. H., Vilarrasa-Blasi, R., Vento-Tormo, R., et al. (2023). Dandelion uses the single-cell adaptive immune receptor repertoire to explore lymphocyte developmental originsDandelion uses the single-cell adaptive immune receptor repertoire to explore lymphocyte developmental origins. *Nat. Biotechnol.* doi:10.1038/s41587-023-01734-7

Svensson, V., Natarajan, K. N., Ly, L.-H., Miragaia, R. J., Labalette, C., Macaulay, I. C., et al. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14, 381–387. doi:10.1038/nmeth.4220

Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E., and Friedman, N. (2017). McPAS-TCR: A manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* 33, 2924–2929. doi:10.1093/bioinformatics/btx286

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386. doi:10.1038/nbt.2859

Voigt, A., and Nguyen, C. Q. (2015). Human T-lymphotrophic virus type-I: A unique association with myelopathy in Sjögren's syndrome. *Clin. Microbiol.* 4, e123. doi:10.4172/2327-5073.1000e123

Voigt, A., Sukumaran, S., and Nguyen, C. Q. (2014). Beyond the glands: An in-depth perspective of neurological manifestations in Sjögren's syndrome. *Rheumatol. (Sunnyvale)* 2014, S4–S010. doi:10.4172/2161-1149.S4-010

Wang, Y., Chen, Z., Wang, T., Guo, H., Liu, Y., Dang, N., et al. (2021). A novel CD4+ CTL subtype characterized by chemotaxis and inflammation is involved in the pathogenesis of Graves' orbitopathy. *Cell Mol. Immunol.* 18, 735–745. doi:10.1038/s41423-020-00615-2

Weber, A., Born, J., and Rodriguez Martínez, M. (2021). TITAN: T-Cell receptor specificity prediction with bimodal attention networks. *Bioinformatics* 37, i237–i244. doi:10.1093/bioinformatics/btab294

Wong, W. K., Leem, J., and Deane, C. M. (2019). Comparative analysis of the CDR loops of antigen receptors. *Front. Immunol.* 10, 2454. doi:10.3389/fimmu.2019.02454

Wong, W. K., Marks, C., Leem, J., Lewis, A. P., Shi, J., and Deane, C. M. (2020). TCRBuilder: Multi-state T-cell receptor structure prediction. *Bioinformatics* 36, 3580–3581. doi:10.1093/bioinformatics/btaa194

Ye, J., Ma, N., Madden, T. L., and Ostell, J. M. (2013). IgBLAST: An immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* 41, W34–W40. doi:10.1093/nar/gkt382

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 16 August 2023
DOI 10.3389/fgene.2023.1250276

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Shahidee Zainal Abidin,
University of Malaysia Terengganu, Malaysia
Selene Ingusci,
University of Pittsburgh, United States

*CORRESPONDENCE
Claes Wahlestedt,
cwahlestedt@med.miami.edu

RECEIVED 29 June 2023
ACCEPTED 07 August 2023
PUBLISHED 16 August 2023

CITATION
McCartan R, Khorkova O, Volmar C-H and Wahlestedt C (2023), Nucleic acid-based therapeutics for the treatment of central nervous system disorders. *Front. Genet.* 14:1250276. doi: 10.3389/fgene.2023.1250276

Nucleic acid-based therapeutics for the treatment of central nervous system disorders

Robyn McCartan[1,2], Olga Khorkova[1,3], Claude-Henry Volmar[1,2] and Claes Wahlestedt[1,2]*

[1]Center for Therapeutic Innovation, University of Miami Miller School of Medicine, Miami, Florida, United States, [2]Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, Florida, United States, [3]OPKO Health, Miami, Florida, United States

Nucleic acid-based therapeutics (NBTs) are an emerging class of drugs with potential for the treatment of a wide range of central nervous system conditions. To date, pertaining to CNS indications, there are two commercially available NBTs and a large number of ongoing clinical trials. However, these NBTs are applied directly to the brain due to very low blood brain barrier permeability. In this review, we outline recent advances in chemical modifications of NBTs and NBT delivery techniques intended to promote brain exposure, efficacy, and possible future systemic application.

KEYWORDS

siRNA, ASO, oligonucleotide, blood brain barrier, drug discovery

1 Introduction

CNS disorders comprise conditions with diverse etiologies ranging from neurodegenerative diseases and mental health disorders to epilepsy and stroke. Despite recent advancements in pharmacology and neuroscience, the therapeutic options for many CNS disorders remain limited. Factors contributing to the limited availability of treatment options include the complexity of the brain, the diverse genetic factors contributing to disease, and the presence of the blood-brain barrier. Of these factors, the presence of the blood brain barrier remains the main challenge of NBT delivery to the brain.

An emerging class of therapeutics that holds significant potential for the treatment of CNS conditions is nucleic acid-based therapeutics (NBTs). Thirty years have passed since the first successful *in vivo* application of NBTs, specifically antisense oligonucleotides (ASOs), in the brains of rats and mice (Wahlestedt et al., 1993a; Wahlestedt et al., 1993b; Standifer et al., 1994). Today, the constantly expanding range of NBT modalities includes not only ASOs but also short interfering RNAs (siRNAs), full-length therapeutic mRNAs, gene therapy constructs, nucleic acid-programmable genome and RNA-editing enzymes (CRISPR-Cas13, ADARs) as well as various vectorized constructs delivering single or multiple NBTs (Zogg et al., 2022). NBTs allow for highly specific and efficient targeting of disease-relevant genes and have benefits over traditional small molecule drugs, such as limited off-target effects and high speed of pharmaceutical development due to rapid design and synthesis of NBTs (Kim, 2022). These qualities along with the ability of NBTs to target specific alleles, isoforms, and point mutations creates new possibilities for personalized medicine in the field of rare diseases (Khorkova et al., 2021). Common chemical structures of different NBT types favors class-wide application of newly discovered chemistries and delivery techniques. Furthermore, NBTs represent an efficient modality for accessing a recently discovered set of novel therapeutic targets related to regulatory long non-coding

RNA (lncRNA), which vastly increases the range of treatable diseases. In particular, natural antisense transcripts (NATs), a subclass of lncRNA, prove as promising therapeutic targets due to their role in gene expression and RNA-based regulatory networks (Khorkova et al., 2023).

In rodents and non-human primates *in vivo* administration of NBTs has proven successful in ameliorating pathological features of various CNS disorders including focal ischemia, leptomeningeal amyloidosis, MECP2 duplication syndrome, Dravet syndrome, and Angelman syndrome (Khorkova and Wahlestedt, 2017). Given the success of these *in vivo* studies, NBTs have great promise for efficacy in a clinical setting. This review will focus on the potential of ASOs and siRNAs, two NBT modalities with several drugs already in the clinic, as treatment options for CNS disorders. Topics highlighted in this review include advancements in overcoming obstacles such as nuclease degradation and entrapment in endosomal compartments, developments in modifications and nanotechnology allowing for BBB penetration, progress in understanding the pharmacokinetics and pharmacodynamics of NBTs, and developments in routes of administration.

2 NBTs: ASOs and siRNA

2.1 ASOs

ASOs typically consist of a single strand of approximately 20 nucleotides that are complementary to a target RNA transcript or genomic sequence. ASOs interact with their target strands via Watson-Crick base pairing, leading to gene expression changes (Dhuri et al., 2020). One of the mechanisms by which ASOs can change gene expression is through the recruitment of the enzyme RNAse H, which leads to the degradation of mRNA transcripts and thus a decrease in gene expression (Lai et al., 2020). Other mechanisms of action of ASOs include inducing steric hindrance that interferes with the function of transcriptional or translational machinery. Modulation of splicing machinery by ASOs can be used to correct the reading frame of mutated transcripts leading to either exon skipping or exon inclusion (Amanat et al., 2022).

2.2 siRNAs

siRNAs are synthetic double-stranded RNA molecules of approximately 20–25 nucleotides that function by exploiting the RNA interference (RNAi) pathway, a cellular pathway that regulates gene expression through the use of endogenous double-stranded RNA molecules (microRNAs or miRNAs) to degrade specific mRNA or lncRNA transcripts (Traber and Yu, 2023). The mechanism of action of siRNA is highly specific and efficient, making it a powerful tool for modulating gene expression.

2.3 NBT chemistry: Internal chemical modifications

Unmodified DNA and especially RNA molecules are highly vulnerable to degradation by endogenous nucleases (Gagliardi and Ashizawa, 2021). To maintain structural integrity and functionality *in vivo*, therapeutic ASOs and siRNAs must undergo chemical modifications that promote stability and cellular uptake. Internal chemical modifications of NBTs can be divided into three categories: sugar modifications, backbone modifications, and base modifications.

2.4 Sugar modifications

Common sugar modifications include 2′O-methyl (2′-MeO), 2′-fluoro (2F), 2′-O-methoxy ethyl (2MOE), locked nucleic acids (LNA) and constrained 2′-O ethyl (cEt) (Figure 1) (Prakash, 2011). Sugar modifications are known to increase NBT binding affinity to target RNA/DNA, stability against nucleases, and limit immune response activation (Shen et al., 2019).

2.5 Backbone modifications

Backbone modifications range from single atom substitutions, as in phosphorothioate (PS) backbones, to complete replacement of natural phosphodiester backbone with phosphorodiamidate morpholino (PMO), thiomorpholino, tricyclo/bicyclo DNA, peptide nucleic acid (PNA) and other groups (Figure 1) (Crooke et al., 2020b). Most modified backbones increase hydrophobicity of NBTs, which improves their pharmacokinetic (PK) properties by facilitating interactions with proteins and lipids. Incorporation of neutral backbone linkages, such as methoxylpropyl, isopropyl and isobutyl phosphonates and O-isopropyl and O-tetrahydrofuranosyl phosphotriesters, formacetal and C3-amide, was shown to improve therapeutic index, with final ASO activity depending on the size, hydrophobicity, and RNA-binding affinity of the linkage (Vasquez et al., 2022). In rare cases ASOs activate TLR9 signaling in a sequence dependent-manner. This effect can be alleviated by using two mesyl phosphoramidate linkages within the PS ASO gap (Pollak et al., 2023).

2.6 Base modifications

Base modifications, such as 5′-methylcytidine, 5′-methyluridine, 2-thiothymine, 8-bromoguanine and 5-hydroxycytosine are known to reduce off-target cleavage by ASOs and inhibit immune system activation (Figure 1) (Song et al., 2022).

Although many of these modifications do not support the mechanisms of action of ASOs and siRNA through interfering with RNAseH and RNAi activity, respectively, Although many of these modifications do not support the mechanisms of action of ASOs and siRNA through interfering with RNAseH or RNAi activity, if modified oligonucleotides are positioned to form so called gapmers, with a modification-free 8–12 nucleotide gap in the center flanked by modified nucleotide “wings”, they can be used to engage these mechanisms (Brooke et al., 2021). Non-gapmer modified oligonucleotides (mixmers) can also be used in steric hindrance applications (Yoshida et al., 2023).

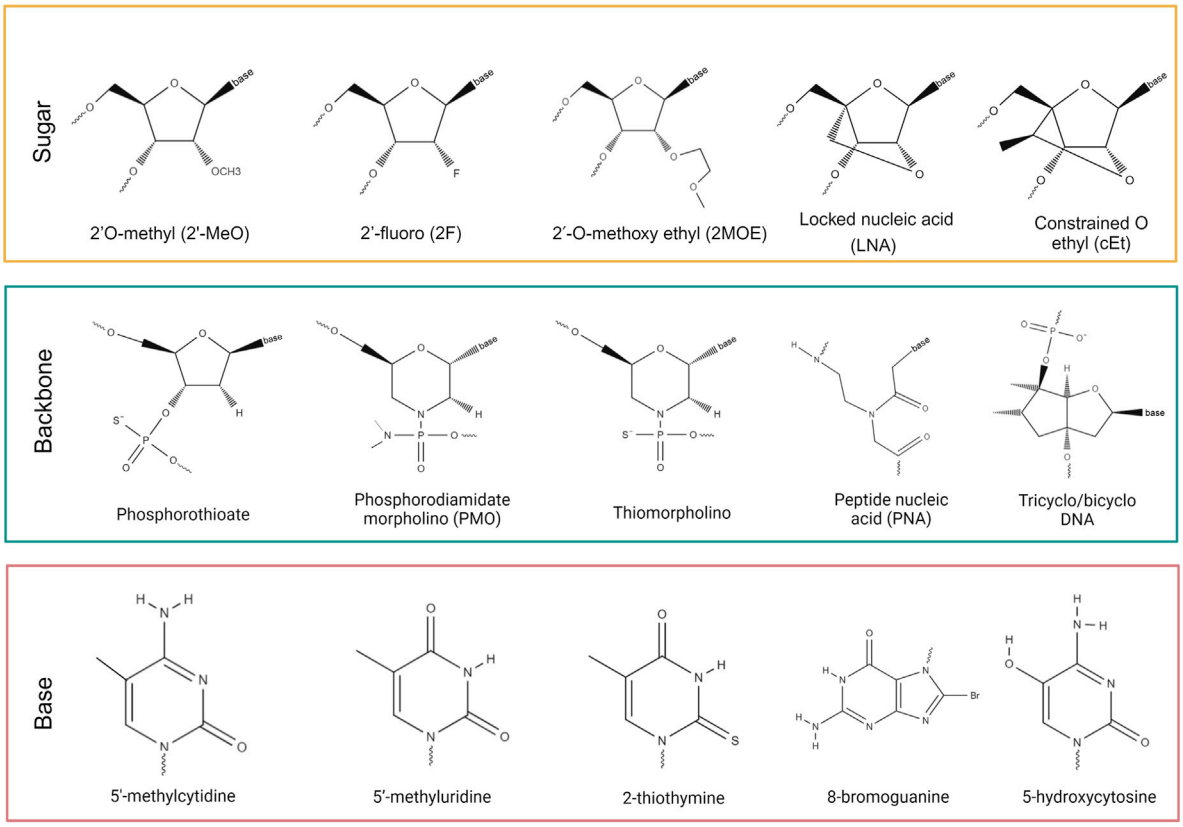


FIGURE 1
Chemical modifications of NBTs.

3 External modifications of NBTs

3.1 Di-valent siRNA scaffolds

The number of PS backbone modifications helps promote systemic and CNS distribution of NBTs, with more PS modifications resulting in better distribution (Crooke et al., 2020a). Specifically, ASOs containing a PS content of 80% have proven especially successful in regard to distribution and efficacy (Sewing et al., 2017; Valenzuela et al., 2023). However, the presence of PS modifications above 40% in siRNA can negatively impact the efficacy of RNA induced silencing complex (RISC) loading and promote off target effects and toxicity (Biscans et al., 2020; Halloy et al., 2022).

A developing method to increase PS content while retaining efficacy and reducing toxicity of siRNA are di-valent siRNA scaffolds (di-siRNA). A di-siRNA consists of two siRNAs connected by a linker. The administration of di-valent siRNAs *in vivo* has resulted in increased brain distribution, target silencing, and duration of silencing (Alterman et al., 2019; Conroy et al., 2022; O'Reilly et al., 2023). Specifically, a single high dose of 475ug of di-siRNA targeting the HTT gene was administered to BACHD-ΔN17 mice via intracerebroventricular (ICV) injection and resulted in significant HTT silencing and guide strand accumulation in key brain regions such as the cortex and striatum for up to 6 months (Alterman et al., 2019). Similarly in non-human primates (NHP), a single 25 mg dose of the same di-siRNA administered via ICV injection resulted in significant HTT silencing and guide strand accumulation throughout the brain at 1 month. Importantly, there were limited off target effects and no discernible toxicity (Alterman et al., 2019). An additional study found that when Msh3 was targeted with a di-siRNA, CAG-repeat expansion was blocked in the striatum of two different Huntington's disease mouse models for up to 4 months and no off target effects were observed (O'Reilly et al., 2023).

Given the demonstrated success of di-siRNAs in distributing throughout the brain, knocking down targets for sustained periods, and the lack of off target effects and toxicity, di-siRNAs hold great promise in the field of NBTs. Furthermore, the proven capability of linking two NBTs while preserving their effectiveness has paved the way for linking NBTs with distinct targets. This approach can be beneficial for conditions with multiple therapeutic targets, as it enables the development of an NBT that can simultaneously address two distinct targets.

3.2 Bioconjugation

Bioconjugation of siRNAs or ASOs to different moieties such as peptides, antibodies, lipids, or sugars can be done to promote targeted delivery, bioavailability, and cellular uptake (Craig et al., 2018). Bioconjugation can also help address *in vivo* delivery

challenges such as promoting endosomal escape with cell-penetrating peptides (CPPs). For example, a study examining the effectiveness of the CPP HA2-ApoE (130-150) showed promising results in increasing the levels of functional SMN2 mRNA in patient-derived fibroblasts and in the SMN2 transgenic mouse model of spinal muscular atrophy (SMA) when administered via tail vein injection at a dose of 8 mg/kg (Dastpeyman et al., 2021). The HA2-ApoE residues (130-150) are known to interact with low density lipoprotein receptors (LDLRs), which are present on brain endothelial cells and support the transport of lipoproteins across the BBB.

3.2.1 Cholesterol conjugation

Cholesterol conjugation is an approach that can increase lipophilicity to enhance BBB penetration. Heteroduplex oligonucleotides conjugated to cholesterol can reach the CNS via intravenous (IV) injection in mice and rats and suppress Malat1 expression by up to 90% (Nagata et al., 2021). Additionally, the 2′-O-hexadecyl (C16) lipophilic modification of siRNA has shown great efficacy preclinically in mouse and NHP models. In CVN-AD mice, a 120 μg dose of C16 conjugated siRNA targeting APP administered via ICV injection led to a 75% reduction of APP mRNA at 30 days, reduced associated pathological makers, and improved behavioral deficits. Intrathecal (IT) injection of 60 mg of C16 conjugated siRNA also targeting APP in NHPs resulted in an 80% knockdown of APP in the brain, with a 75% knockdown sustained for 2.5 months (Brown et al., 2022). This preclinical success has led Alnylam and Regeneron to run a phase I study testing the C16 conjugated siRNA in 20 early onset Alzheimer's disease patients. Preliminary results from this study has shown that IT injection of the C16 conjugate leads to a reduction of APP and its metabolite by 70% in the CSF of patients. Cognitive function was not evaluated (Alnylam Pharmaceuticals, 2023).

3.2.2 Conjugations targeting transferrin receptor family

Conjugation of oligonucleotides to antibodies has also shown success regarding brain administration. When an ASO targeting the SMN2 gene was conjugated to an antibody targeting the murine transferrin receptor and administered systemically at a dose of 50 mg/kg in a hSMN2-transgenic mouse model, therapeutic levels of SMN2 splicing demonstrated by a 2-fold increase of FLSMN2 expression were reached in the mouse brain (Hammond et al., 2022).

Given the widespread presence of TfR1 on brain endothelial cells, facilitating drug uptake into the brain via TfR1-mediated transcytosis has been a mechanism of interest. A recent study in preprint has demonstrated successful ASO brain administration through the targeting of TfR1 in mice and NHPs (Barker et al., 2023). In this study, ASOs were conjugated to a transport vehicle (TV) that was engineered to have a binding affinity for TfR1. The TV used was a huIgG containing an engineered Fc domain. When administered systemically at a dose of 2.4 mg/kg, the ASO-TV complex targeting Malat1 proved successful in penetrating the BBB and distributing ASO throughout the brain of $TfR^{mu/hu}$ KI mice (Barker et al., 2023). A significant and sustained level of Malat1 knockdown was demonstrated in the brain of both mice and NHPs.

siRNAs can also utilize the presence of the transferrin receptor family on the BBB to aid in brain administration. Specifically, when administered in BALB/c mice three times via IV injection at a dose of 30 mg/kg, siRNA targeting the NOX4 gene was able to bypass the BBB when conjugated to a peptide that interacts with the melanotransferrin receptor (Eyford et al., 2021). The peptide, MTfp, was demonstrated to act as a "nanomule" allowing the siRNA to reach the brain and therapeutically knock down levels of NOX4.

3.2.3 Conjugations targeting GLUT1 transporter

Through targeting the GLUT1 transporter, carrier-mediated transcytosis can be exploited to facilitate drug transfer across the BBB. GLUT1 is present on the surface of the plasma membrane of brain capillary endothelial cells (BCEC). When blood glucose levels rise after fasting, GLUT1 undergoes transcytosis and travels from the apical to the basal side of BCEC membranes (Koepsell, 2020). Nanoparticles modified with a glucose moiety can interact with GLUT1 when it is located on the apical side, and once transcytosis occurs due to changes in blood glucose levels, the nanoparticle can travel along with GLUT1 to the basal side and be released into the brain. Taking advantage of this mechanism, a study showed that after just a single IV injection of 100 μg of an ASO contained within a glucose-coated polymeric nanocarrier, a 40% knockdown of the target Malat1 was obtained in both the cortex and hippocampus (Min et al., 2020). Furthermore, a study that developed a glycosylated polymeric siRNA nanomedicine known as Gal-NP@siRNA showed that systemic administration of 1 mg/kg of the glycosylated siRNA led to lower expression of the target gene and improvement of cognitive decline in the APP/PS1 mouse model of Alzheimer's disease (Zhou et al., 2020).

3.3 Nanoparticles

Success in improving the bioavailability of NBTs in the brain after systemic administration via chemical modification is further advanced through the use of nanotechnology and nanoparticles. To effectively deliver NBTs to the brain as well as other tissues, nanoparticles must protect against degradation in the blood before reaching the BBB and must facilitate the crossing of BBB.

3.3.1 Polymeric nanoparticles

Polymeric nanoparticles are an attractive option as their composition and surface properties can be modified to control the drug release rate and promote cell type-specific delivery (Amulya et al., 2023). Polymeric nanoparticles comprise a lipid with dissolved drug cargo surrounded by a biodegradable polymeric shell, which can be engineered to control cargo release. Alternatively, they can be formed by a continuous polymeric network containing the drug (Marangoni and Menezes, 2022). However, a disadvantage of polymer-based delivery systems is that they have poor NBT transfer efficiency and are rapidly cleared. A polyethylene glycol (PEG) coating can increase circulation time and reduce nonspecific cellular uptake (Shi et al., 2021). A recent study has demonstrated that modulating the surface coating and surface density of polymeric nanoparticles can augment the active transport of the nanoparticle across the BBB (Li et al., 2021). The study showed that 50 nmol/kg of

a systemically administered poly-lactic-co-glycolic acid (PLGA)-based nanoparticle coated in polysorbate 80 (PS 80) increased brain penetration threefold compared to traditional PEG-coated nanoparticles in the mouse brain. Additionally, the PS 80 nanoparticle containing tau-targeting siRNA led to a 50% decrease in tau expression in the brain after systemic administration (Li et al., 2021). Use of 3D printing in polymeric nanoparticle production can improve their mechanical properties, shapes and release profiles as well as improve the quality and scalability of the nanoparticle manufacturing process (Kutlehria et al., 2022).

3.3.2 Lipid-based nanoparticles

Lipid-based nanoparticles (LNPs) have a lipid bilayer shell and lipid core matrix that is stabilized by emulsifiers. The lipid bilayer encapsulates the NBT and protects it from degradation, ultimately allowing for entrance into cells. Although LNPs are too large to cross the BBB, they are capable of targeting the brain when administered via ICV or IT injection (Bost et al., 2021). One way that LNPs can be modified to cross the BBB is through the creation of neurotransmitter-derived lipidoids (NT-lipidoids). The hybrid structure of NT-lipidoids can interact with receptors on the BBB and facilitate transport. NT-lipidoids containing an ASO targeting tau successfully bypassed the BBB when administered intravenously at a dose of 1 mg/kg and led to the knockdown of tau mRNA and protein (Ma et al., 2020).

3.3.3 Inorganic nanoparticles

Another category is inorganic nanoparticles, which includes gold nanoparticles (AuNPs). Although highly charged AuNPs can activate an immune response in the brain, surface modifications can be added to help alter the charge (Kus-Liskiewicz et al., 2021). AuNPs usually consist of an inner inorganic layer that is surrounded by a nonpolar organic layer that helps stabilize the core and attaches to the NBT (Kasina et al., 2022). Surface modifications can be added to aid with targeting, while the small size of AuNPs helps with crossing the BBB. Coating of an AuNP with bioengineered brain-targeting exosomes has aided in brain uptake. A study utilizing this surface modification showed accumulation of the AuNP in the brain after systemic administration of 200 µL in mice (Khongkow et al., 2019).

3.3.4 Exosomes

Lastly, exosomes have also been explored as a potential delivery system to the CNS. They do not stimulate an immune response and have an advantage in that they can carry more than one drug simultaneously (Cecchin et al., 2023). Since exosomes are surrounded by a lipid bilayer, they are membrane-permeable and can cross the BBB. Drug-containing exosomes have been demonstrated to reach the CNS via IV administration when conjugated with the rabies virus glycoprotein (Sun et al., 2022). However, loading drugs into exosomes remains a challenge (Bost et al., 2021). One way this limitation has been overcome is through the reprogramming of hepatocytes to transcribe siRNA and package it into exosomes (Zhang et al., 2021). The siRNA-containing exosomes formed by the hepatocytes are tagged with the rabies virus glycoprotein and exploit the endogenous exosome-circulating system to bypass the BBB. This technology has proven effective in reducing levels of expression of the mutated Huntington's disease (HD) gene in the cortex and striatum in three different mouse models of HD including N171-82Q, BACHD, and YAC128 (Zhang et al., 2021).

4 NBT uptake

4.1 BBB uptake mechanisms

There are several mechanisms by which NBTs can cross the BBB, including paracellular transport, carrier-mediated transport, receptor-mediated transport, and absorptive-mediated transport.

Paracellular transport occurs through the junctions between the brain endothelial cells or in areas with weak BBB integrity which allows for the passage of small hydrophilic compounds. In regard to NBT delivery, paracellular transport across the BBB is relatively inefficient. However, hyperosmotic solutions, such as those containing mannitol, can be administered systemically to alter the osmotic pressure of endothelial cells at the BBB, which allows the tight junctions to open, and facilitates paracellular transport of NBTs (Chu et al., 2022).

Carrier-mediated transport is another mechanism that allows NBTs to cross the BBB using solute carriers. Specifically, the molecules bind to carrier proteins present on the membrane, allowing the molecule to become internalized and move from one side of the BBB to the other due to a concentration gradient (Yazdani et al., 2019).

Receptor-mediated transport is another possible mechanism, in which an NBT is conjugated with a ligand specific to a receptor present in brain endothelial cells. One method by which receptor-mediated transport is exploited for drug delivery to the brain is targeting the transferrin receptor (Tfr1) abundant on brain endothelial cells (Thomsen et al., 2022).

Lastly, absorptive-mediated transport is a mechanism by which cationic molecules can bind to the luminal surface of endothelial cells and be taken up by the cell and transported across the BBB (Knox et al., 2022). An example of this is nanoparticles that interact with lipid rafts of the plasma membrane. Lipid rafts are negatively charged, which promotes interaction with positively charged nanoparticles or nanoparticles being carried by positively charged serum molecules (Hald Albertsen et al., 2022).

4.2 Cellular uptake mechanisms

It is proposed that cellular uptake of ASOs predominately occurs through receptor-mediated endocytosis via clathrin and caveolin-dependent mechanisms (Migliorati et al., 2022). Multiple pathways for ASO internalization exist, and they often involve surface receptors such as integrins, GPCRs, and scavenger receptors (D'Souza et al., 2023). Once internalized, the ASOs can be trafficked from early endosomes to late endosomes and then to lysosomes (Holm et al., 2022).

Common uptake mechanisms of siRNAs are similar to that of ASOs and include clathrin- and caveolin-mediated endocytosis (Juliano et al., 2012). Intracellular trafficking of siRNA follows a similar pathway to ASOs and involves traveling through the early

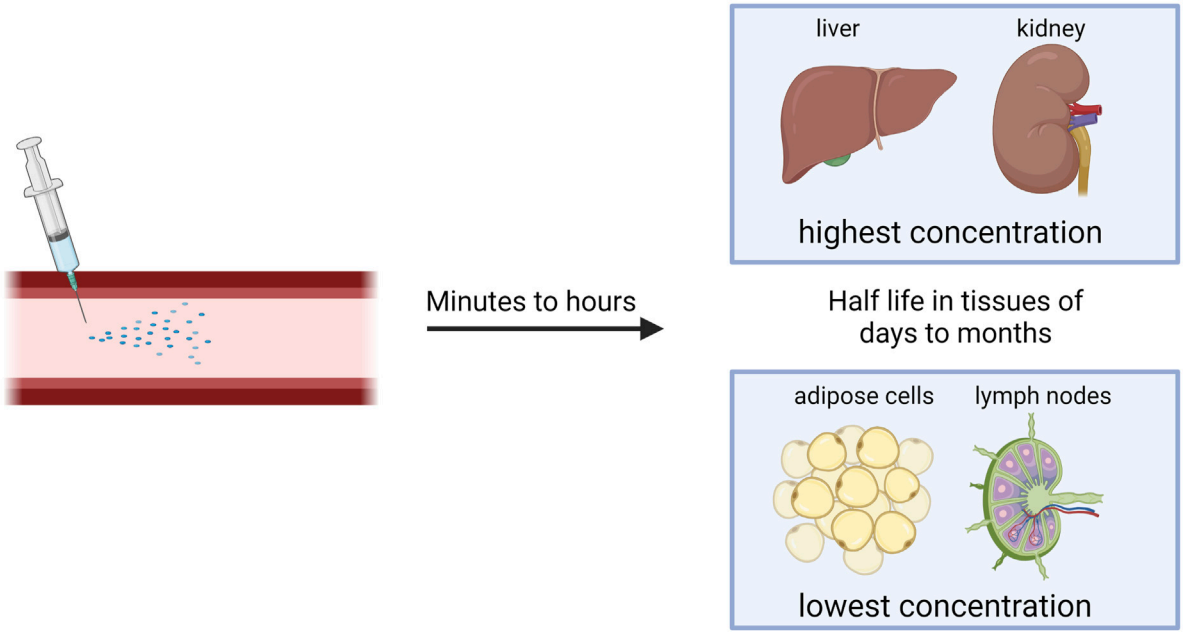


FIGURE 2
Distribution of NBTs after systemic injection.

and late endosomes. Both ASOs and siRNA must escape the late endosome to ensure entry to the cytoplasm to interact with their target pathways (Holm et al., 2022; Dowdy, 2023).

As a result, although oligonucleotides are readily taken up and retained by the cells, they are largely sequestered in these subcellular membranous structures, with only a small proportion available for productive regulation of gene expression. NBTs were shown to escape sequestration mainly from pre-lysosomal compartments, during multiple vesicle budding and fusion events (Gokirmak et al., 2021). The exact mechanisms of endosomal escape have not been elucidated yet and methods to improve this process are under development (Roberts et al., 2020; Gokirmak et al., 2021).

Multiple methods to increase the probability of endosomal escape, and thus the bioavailability of NBTs, have been proposed. For example, a non-toxic synthetic sphingosine analog SH-BC-893 inactivated the small GTPase ARF6, which inhibited lysosomal fusion and trapped endocytosed oligonucleotides in pre-lysosomal compartments. This resulted in an up to 200-fold increase in activity of ASOs and siRNAs (Finicle et al., 2023).

Furthermore, conjugation of a cell-penetrating peptide HA2-ApoE (131-150) containing several endosomal escape domains to nusinersen significantly increased the level of full-length SMN2 in the brain and spinal cord of SMN2 transgenic mice after systemic administration (Dastpeyman et al., 2021).

Identifying genes involved in productive ASO release may help in designing therapeutic interventions to increase it. Overexpression of GOLGA8, a novel protein localized predominantly to the trans-Golgi and plasma membrane increased uptake and activity of splice modulating and RNase H1-dependent ASOs (McMahon et al., 2023).

However, cellular uptake is only a small part of the challenges that complicate NBT delivery *in vivo*. After IV or subcutaneous injection, NBTs are subject to nuclease degradation in blood and interstitial tissues, sequestration by plasma proteins, phagocytosis by specialized blood, liver and kidney cells and renal clearance. Furthermore, to reach the target cells they must pass through the endothelial barrier and/or BBB to reach the CNS (Roberts et al., 2020; Gokirmak et al., 2021).

5 NBT distribution

After systemic or subcutaneous administration, tissues with highest ASO concentration are liver and kidney, but ASOs also accumulate in bone marrow, the cell body of adipocytes, and lymph nodes (Figure 2) (Takakusa et al., 2023). The pharmacokinetic properties of naked ASOs are enhanced by the presence of a phosphorothioate backbone, which allows for the binding of serum proteins that serve as molecular shuttles for the drug (Vasquez et al., 2022).

The transfer of ASOs or siRNA from blood to tissues occurs in minutes to hours after administration (Figure 2), and 80% of the plasma concentration can be excreted through kidneys within 24 h. After initial quick clearance, low levels of ASOs persist in plasma with a half-life of 2–8 weeks, possibly due to slow release from interstitial space and intracellular depots. siRNA is eliminated from the plasma within minutes to hours, but once within tissues, it has a half-life of days to months (Figure 2) (Jeon et al., 2022). To increase stability in the plasma, albumin-binding aptamer chimeras have shown success (Rosch et al., 2022). This is due to the long half-life and high concentrations of albumin within the plasma.

After direct CNS administration, ASOs quickly disperse over all brain regions with a gradient of concentration starting from the point of injection. Single nucleus transcriptomics have

demonstrated that in mice and NHPs treated with ASOs against Prnp and Malat1 target knockdown was observed in every cell type, and in most individual cells, as opposed to large effect in a small number of cells. The knockdown duration was shorter in microglia than in neurons and lasted for up to 12 weeks after injection. In NHPs, the level of PRNP knockdown in CNS adequately reflected the level of effects in CNS cells (Mortberg et al., 2023).

5.1 NBT pharmacokinetics and pharmacodynamics

Novel PK and pharmacodynamic (PD) properties of NBTs complicate toxicology studies and clinical trial design. Predictive power of animal models for different NBT modalities is also poorly studied at this time. As the number of clinical studies of NBTs and the numbers of patients dosed in each study increase, PK models can be developed. Population PK models are useful in determining therapeutic windows and dosage recommendations for NBTs. These models integrate available PK data to identify factors that can affect individual drug exposure. Based on data collected from 750 participants from five clinical studies of an IT-administered ASO tominersen (Table 2), CSF PK could be modeled using a three-compartment design with first-order transfer from CSF to plasma, while a three-compartment model with first-order elimination from plasma was needed for plasma PK. Significant covariates for CSF clearance included baseline total CSF protein, age, and levels of antidrug antibodies (ADAs). Clearance from plasma covaried with body weight (Yamamoto et al., 2023).

A whole-body mechanistic model of ASO distribution in after IT injection based on NHP data has been proposed. Physiological parameters and kinetic uptake rates were set using PK data for 2′-MOE PS gapmers in cynomolgus monkeys. The model adequately described spinal transport of an IT-administered ASO due to pulsation-enhanced mixing, advection and diffusion, as well as ASO transfer to systemic circulation, CSF and peripheral tissue compartments. The model considered six anatomical regions: spinal CSF, spinal cord tissue, cranial CSF, cranial tissue, blood and peripheral tissues. Cranial tissue was subdivided into compartments corresponding to pons, hippocampus, cerebellum, and cortex. Spatiotemporal distribution of ASO in these compartments as a function of injection volume and duration revealed that higher infusion volume can enhance drug dispersion along the spinal axis, while shorter duration of injection results in better cranial delivery compared to a long infusion of the same volume. The model uses physics-based scaling laws to assess subject-specific variations (e.g., adult-child) or enable translation of data across species (Linninger et al., 2023).

Furthermore, in order to better understand NBT distribution, there is a need for microscopic approaches. Techniques such as fluorescence imaging, mass imaging, and immunohistochemistry are emerging as promising methods for tracking these therapeutics within biological systems. Additionally, there is a need for highly sensitive bioanalytical methodologies that are capable of detecting the low plasma concentrations of NBTs (Takakusa et al., 2023). Overall, the understanding of NBT delivery is expanding with recent developments focusing on increasing endosomal escape and improving bioavailability. However, challenges such as nuclease degradation, sequestration, and delivery to target cells remain, and understanding the PK and PD of NBTs is imperative for NBT development and the optimization of clinical trials.

6 CNS administration techniques

6.1 Intrathecal administration

CNS targeting by both ASOs and siRNAs can be achieved via direct brain delivery through IT (intrathecal) injection (Goto et al., 2023). IT administration involves injecting the drug into the CFS within the spine and can be performed in a variety of ways, including intralaminar injection, transforaminal injection, or using a subcutaneously implanted catheter for repeat injections (Figure 3) (Lee et al., 2022; Alkosha, 2023). IT injection is the indicated route of administration for 16 oligonucleotide therapeutics that are under development as of 2022 (Goto et al., 2023) (Table 2). However, IT injections are invasive procedures and carry risks of side effects such as headaches and back pain, which does not encourage wide clinical acceptance of NBTs. Around 10% of spinal muscular atrophy patients who receive nusinersen IT injections report experiencing negative side effects after the injection (Veerapandiyan et al., 2020).

6.2 Intracerebroventricular administration

Another method of NBT administration that bypasses the BBB is ICV injection. While extremely invasive, ICV injections allow for drugs to directly penetrate the brain, leading to increased bioavailability and effectiveness (Figure 3) (Atkinson, 2017). In cases where repeat injections are necessary, an Ommaya reservoir can be utilized. An Ommaya reservoir allows for direct ventricular access via an intraventricular catheter system (Figure 3) (Zubair and De Jesus, 2022). A study investigating the safety and efficacy of repeat ICV injections via an Ommaya reservoir for the treatment of epilepsy has shown this administration method to be safe and effective (Cook et al., 2020). While Ommaya reservoirs have been most commonly utilized in the administration of chemotherapeutic agents, treatments for meningitis, and more currently, epilepsy, a recent study has proven the use of the reservoir to be successful in the administration of the ASO nusinersen (Iannaccone et al., 2021). In addition to an Ommaya reservoir, ventriculoperitoneal shunts can be utilized for ICV injections in patients with pre-existing shunts. By connecting an external pump or reservoir to the shunt system, drugs can be precisely and repeatedly administered into the ventricles and distributed throughout the brain (Duma et al., 2019). However, a recent study revealed that out of 111 ICV injections, approximately 11% resulted in transient meningismus and mild temperature increase, with 1.8% requiring hospitalization (Duma et al., 2019). These complications can be minimized by treatment with prophylactic dexamethasone and the use of a ventriculoperitoneal shunt in place of an Ommaya reservoir (Duma et al., 2019). Given the invasiveness and risk of side effects, ICV injections have not yet proven to be a common route of administration for NBTs.

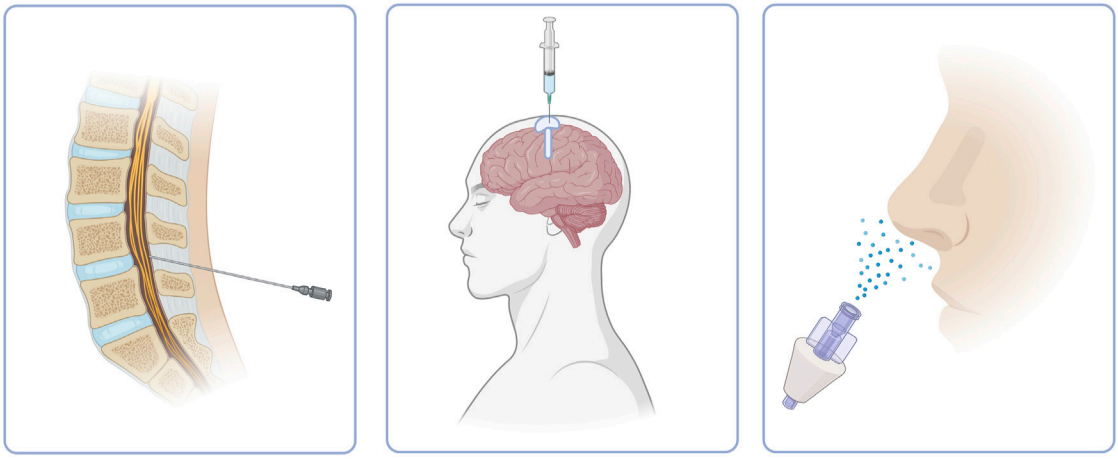


FIGURE 3
Routes of administration of NBT. Intracerebroventricular route features use of an Ommaya reservoir.

6.3 Intravenous administration

A less invasive, albeit currently limited form of drug administration, is systemic administration via intravenous injection. When administered systemically, drugs must be modified to penetrate the BBB and limit off-target effects. As previously discussed, nanotechnology has been under development to promote the bypassing of the BBB. Current nanotechnologies that have shown efficacy in delivering oligonucleotide therapeutics to the brain in animal models include PLGA-coated polymeric nanoparticles, NT-lipidoids, AuNPs, and exosome delivery systems (Khongkow et al., 2019; Ma et al., 2020; Li et al., 2021; Zhang et al., 2021; Sun et al., 2022).

To augment advancements in nanotechnology that have somewhat mitigated this problem by increasing the percentage of NBT dose available to the brain after systemic delivery, novel methods of NBT administration can also play a role in bypassing the BBB.

6.4 Intranasal administration

Intranasal administration has recently gained attention as an attractive method for drug delivery due to its numerous advantages (Figure 3). These advantages include rapid drug delivery to the brain, reduced potential for off-target effects due to low systemic exposure, a non-invasive procedure and the feasibility of self-administration (Lochhead and Thorne, 2012). When administered intranasally, drugs reside in the nasal cavity before entering the brain through one of two pathways, the olfactory nerve pathway, or the trigeminal nerve pathway. However, drug absorption via the olfactory nerve is the most effective and best understood due to the presence of extensive epithelia and vasculature and the short length of the nerve in comparison to the trigeminal nerve (Lochhead and Thorne, 2012). The olfactory nerve pathway begins at the olfactory epithelium which is located at the upper aspect of the nasal cavity. The olfactory epithelium houses olfactory receptor neurons which absorb the drug. The axons of these neurons cross the cribriform plate, a porous bone located at the base of the skull, before reaching the olfactory bulb. From the olfactory bulb, the drug can disperse to various brain regions through neuronal connections, allowing it to reach specific target areas to exert its intended therapeutic effects (Jeong et al., 2023). Drugs can also travel through the extracellular fluid surrounding the olfactory neurons after passing through the tight junctions of epithelial cells, toward the lamina propria, a noncellular layer of connective tissue. From the lamina propria the drugs can translocate through the perineural space until the subarachnoid space is reached and further brain distribution can ensue (Crowe and Hsu, 2022).

While PK studies of intranasal administration are limited, there is evidence to suggest that NBTs administered via this route can reach the olfactory bulb as soon as 5 min post-administration and more distal regions of the brain as soon as 30 min in the mouse brain (Renner et al., 2012; Falcone et al., 2014). Nanotechnological modifications including polymeric nanoparticles and solid lipid nanoparticles can make the process of uptake more efficient and successful (Maher et al., 2023). As previously discussed, the size, surface qualities, and biocompatibility of these nanocarriers allow for efficient uptake and bypassing of the BBB (Begines et al., 2020). To this effect, a cationic lipid nanoparticle encapsulating an ASO targeting the HuR gene was shown to effectively knock down levels of HuR in the mouse brain when administered intranasally at a dose of 1 nmol (Borgonetti and Galeotti, 2021). In an additional study, the same anti-HuR ASO encapsulated in the cationic lipid nanoparticle demonstrated protection against demyelination in the mouse brain when administered intranasally (Borgonetti and Galeotti, 2023). Similar to drugs administered systemically, drugs administered intranasally can also be modified to bind to receptors present on the BBB, such as T7 siRNA modifications allowing for uptake by the transferrin receptor. In a recent study, a T7-modified

TABLE 1 NBTs approved by FDA for CNS disorders *Not commercially available.

Condition	Drug name	Target	Delivery method	Manufacturer	Year approved
Spinal muscular atrophy	Nusinersen (Spinraza)	SMN2	IT	Ionis/Biogen	2016
Batten's disease	Milasen*	ATXN2	IT	Boston Children's Hospital	2018
Amyotrophic lateral sclerosis	Tofersem (Qalsody)	SOD1	IT	Ionis/Biogen	2023

siRNA targeting gliomas showed efficacy in positively altering the tumor microenvironment when administered intranasally (Tang et al., 2023).

Drug uptake via the intranasal route is most successful when the drug resides for longer periods within the nasal cavity. Considering this, modifications can be made to the drug to allow for longer residence time. One such modification is the conjugation of chitosan, a mucoadhesive agent which functions by electrostatically interacting with the negatively charged epithelial surfaces and thus enhancing the drug's duration of stay in the nasal cavity (Maher et al., 2023). A study exploring the intranasal administration of a chitosan-based nanoparticle containing HTT targeting siRNA demonstrated therapeutically significant lowering of HTT mRNA in the mouse brain (Sava et al., 2021). Another method to increase drug residence time in the nasal cavity is through the use of biogels, which are liquids that transition to gels at body temperature (Thakkar et al., 2021). When a drug was administered intranasally in mice along with a biogel, the transport efficiency of the drug from the nasal cavity to the brain was significantly enhanced (Xie et al., 2019). The use of biogels for the intranasal administration of NBTs remains to be investigated, however.

One of the drawbacks of intranasal formulations is that the drug dose cannot be delivered accurately. A recently proposed minimally invasive intranasal depot (MIND) technique circumvents this problem by depositing ASO-containing gels under the olfactory epithelium through a simple ambulatory procedure (Padmakumar et al., 2022). MIND-mediated delivery of an ASO capable of de-repressing BDNF expression resulted in its wide CNS distribution followed by sustained upregulation of BDNF to approximately 40% of levels achieved with ICV delivery (Padmakumar et al., 2021).

Other additives that have been shown to improve intranasal delivery *in vivo* include vasoconstrictors which reduce the likelihood of the drug being absorbed into circulation, enzyme modulators which can reduce the function of enzymes that may degrade the drug, and nasal permeability enhancers which increase absorption through the nasal epithelium (Dhamankar and Donovan, 2017; Rabinowicz et al., 2021; Moradi and Dashti, 2022).

7 Clinical development of CNS-targeting NBTs

To date, there are two commercially available FDA-approved NBTs for CNS disorders (Table 1). Nusinersen (Spinraza), approved in 2016 for the treatment of SMA, was the first NBT to gain FDA approval and has since shown success in improving functional outcomes and slowing disease progression (Pechmann et al., 2022). SMA is caused by a mutation in the SMN1 gene which leads to a deficiency in the SMN protein that is vital for motor neuron function. Nusinersen functions by modifying the splicing process of an inactive duplicate gene, SMN2, resulting in a transcript that codes for the functional SMA protein (Neil and Bisaccia, 2019).

After the initial loading phase, nusinersen is administered to patients via IT injection every 4 months (Zingariello et al., 2019). Adverse effects of nusinersen treatment are rare and usually limited to side effects from the IT injection (Neil and Bisaccia, 2019). Despite being a very effective treatment option, treatment with nusinersen remains inaccessible to many patients due to the high price. A recent study demonstrated that of 324 patients, 50% discontinued treatment after 12 months. Non-adherence was correlated with greater frequency of comorbidities and increased costs for patients (Gauthier-Loiselle et al., 2021).

The second, and most recently approved NBT for a CNS disorder, is Qalsody (tofersen). Tofersen was approved in April 2023 for the treatment of amyotrophic lateral sclerosis (ALS) associated with mutations in the SOD1 gene (Hara Prasad, 2023). Mutations in the SOD1 gene can lead to both toxic gain of function and toxic loss of function in mutant SOD1 proteins, leading to the damage of motor neurons (Berdynski et al., 2022). Tofersen functions by mediating the RNAse-H degradation of the pathogenic SOD1 mRNA transcript which reduces levels of the toxic SOD1 mutant protein (Miller et al., 2020). While the functional outcomes of patients treated with tofersen did not differ from patients treated with a placebo, 28 weeks of once-monthly IT injections did lead to a significant decrease in the total plasma concentration of SOD1 (Miller et al., 2022).

Another NBT, Milasen, was approved by the FDA in 2018 for the treatment of Batten's disease in a single patient. Milasen functions by modulating the splicing of the gene MFSD8 and was administered to the patient via IT injection (Kim et al., 2019). While specific to only one patient and not commercially available, the development and subsequent FDA approval of milasen serves as a proof of concept for the potential of "n of 1" NBT therapeutics.

Several other CNS targeting NBTs are now in late stages of development (Table 2). In April 2023 Ionis announced positive results of eplontersen study in hereditary ATTR polyneuropathy. At 66 weeks, the study met all endpoints, including reduction in serum TTR concentration and neuropathy impairments, as well as improvements in quality of life (Ionis Pharmaceuticals, 2023).

At the same time Alnylam and Regeneron reported positive results from the trial of ALN-APP, an siRNA NBT against the amyloid precursor protein (APP). CSF concentrations of both soluble APPα and its metabolite Abeta were reduced by up to 70% for at least 3 months at the highest dose tested (Alnylam Pharmaceuticals, 2023).

Treatment of AD patients with another ASO, IONIS-MAPTRx, resulted in a dose-dependent reduction in the CSF concentration of protein tau of more than 50% from baseline at 24 weeks after 2 or 4 doses of IONIS-MAPTRx (Mummery et al., 2023).

TABLE 2 NBTs in clinical trials for CNS disorders.

Condition	Drug name	Target	Delivery method	Sponsor name	Phase	ClinicalTrials.gov ID
Amyotrophic lateral sclerosis	Ulefnersen (ION363)	FUS	IT	Ionis	III	NCT04768972
Amyotrophic lateral sclerosis	ION541 (BIIB105)	ATXN2	IT	Ionis/Biogen	I/II	NCT04494256
Huntington's disease	AMT-130	HTT	Intra-striatal injection	UniQure Biopharma	I/II	NCT04120493
Huntington's disease	Tominersen (IONIS-HTTRx, RG6042)	HTT	IT	Ionis/Roche	II	NCT05686551, NCT04000594, NCT03842969
Huntington's disease	WVE-003	HTT	IT	Wave Life Sciences	I/II	NCT05032196
Hereditary ATTR polyneuropathy	Eplontersen (IONIS-TTR-LRx, AKCEA-TTR-LRx)	HTT	Self-administered SC	Ionis/Astra Zeneca	III	NCT05667493, NCT05071300, NCT04136171
Alexander disease	Zilganersen (ION373)	GFAP	IT	Ionis	III	NCT04849741
Alzheimer's disease	IONIS-MAPTRx (BIIB080)	MAPT	IT	Ionis/Biogen	II	NCT03186989, NCT02831517, NCT02957617
Alzheimer's disease, early onset	ALN-APP	APP	IT	Alnylam/ Regeneron	I/II	NCT05231785
Dravet syndrome	STK-001	SCN1A	IT	Stoke Therapeutics	II	NCT04740476
Parkinson's disease	ION859 (IONIS-BIIB7Rx, BIIB094)	LRRK2	IT	Ionis/Biogen	I/II	NCT03976349
Angelman syndrome	ION582	UBE3A-ATS	IT	Ionis/Biogen	I/II	NCT05127226
Multiple system atrophy	ION464 (BIIB101)	SNCA	IT	Ionis	I/II	NCT04165486

IT, intrathecal injection; SC, subcutaneous injection.

Wave Life Sciences is developing WVE-003, a stereopure ASO that targets a specific single nucleotide polymorphism (SNP3) within the huntingtin gene, allowing for specific destruction of transcripts with expanded repeat sequence. Preliminary results of the clinical trial announced in September 2022 showed media mutant HTT reduction of 30% from baseline at 85 days post single IT injection (Wave Life Sciences USA, 2022).

Other trials are on-going (Table 2), indicating a significant interest in developing NBTs for CNS disorders.

8 Current challenges and future perspectives in CNS-targeting NBTs

While NBTs have proven to be a promising treatment option for CNS diseases, there remain challenges that need to be addressed. Major challenges include effectively bypassing the BBB, improving tissue and cell-specific targeting, and enhancing cellular uptake and endosomal escape.

The development of nanocarriers has helped address the challenge of bypassing the BBB due to their small size, biocompatibility, and ability for surface functionalization (Amulya et al., 2023). Targeting extrahepatic ligands such as the transferrin receptor may be a promising avenue for brain-specific uptake. NBTs conjugated to Tfr1 binding moieties and anti-TfR antibodies have shown preclinical success (Barker et al., 2023; Gabold et al., 2023).

In regard to cell-specific targeting, methods to specifically target glial cells or neurons remain to be elucidated. This is especially important as glial cells, such as microglia, play a pivotal role in the pathology of many CNS diseases.

Cellular uptake of NBTs has been enhanced through chemical modifications and the development of nanotechnologies (Crooke et al., 2017). Endosomal escape, however, remains a pertinent issue as endosomes may retain as much as 99% of NBTs, leaving only 1% of the drug to exert therapeutic effects (Dowdy, 2023). Technologies that promote endosomal escape remain under development, but the incorporation of cell-penetrating peptides or sphingolipid treatment may be promising (Dastpeyman et al., 2021).

Furthermore, the understanding of brain distribution of NBTs remains limited and requires further elucidation. Future research into this area should include investigation of the molecular mechanisms governing the brain distribution of NBTs, physiological and anatomical differences affecting brain NBT distribution, and quantification of intracellular NBT distribution (D'Souza et al., 2023). Understanding these principles will aid in the development of future CNS-targeted NBTs. For example, in regard to the treatment of Parkinson's disease, the target dopaminergic neurons of the substantia nigra and ventral tegmental area comprise only 600,000 cells (Holm et al., 2022). Precise targeting of this small

subset of neurons will reduce the required dose and possibility off-target effects of NBTs.

Lastly, the administration methods of NBTs targeting the CNS remain a challenge, with a need for a non-invasive and effective method for bypassing the BBB. The majority of brain targeting NBTs are currently administered via IT injection (Goto et al., 2023). While effective, IT administration proves invasive and is not easily accessible to all patients. Alternative administration routes such as intranasal administration are currently being explored and hold significant potential (Maher et al., 2023).

Author contributions

RM, OK, C-HV, and CW contributed to researching, critically analysing, editing, and writing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Related ongoing research at the Center for Therapeutic Innovation (CTI) receives funding from the NIH (R01AA029924 and R01AG079373) as well as from the State of Florida.

Acknowledgments

Thank you to the members of the CTI Oligo Meeting Group for fruitful discussions regarding advancements in the field of NBTs. Figures created with Biorender.com.

Conflict of interest

Author OK was employed by OPKO Health.

The remaining 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

Alkosha, H. M. (2023). Interventional non-operative management of low back and neck pain. *Egypt. J. Neurosurg.* 38, 13. doi:10.1186/s41984-023-00189-x

Alnylam Pharmaceuticals, I. (2023). *Alnylam and Regeneron report positive interim phase 1 clinical data on ALN-APP, an investigational RNAi therapeutic for Alzheimer's disease and cerebral amyloid angiopathy*.

Alterman, J. F., Godinho, B. M. D. C., Hassler, M. R., Ferguson, C. M., Echeverria, D., Sapp, E., et al. (2019). A divalent siRNA chemical scaffold for potent and sustained modulation of gene expression throughout the central nervous system. *Nat. Biotechnol.* 37, 884–894. doi:10.1038/s41587-019-0205-0

Amanat, M., Nemeth, C. L., Fine, A. S., Leung, D. G., and Fatemi, A. (2022). Antisense oligonucleotide therapy for the nervous system: from bench to bedside with emphasis on pediatric neurology. *Pharmaceutics* 14, 2389. doi:10.3390/pharmaceutics14112389

Amulya, E., Sikder, A., Vambhurkar, G., Shah, S., Khatri, D. K., Raghuvanshi, R. S., et al. (2023). Nanomedicine based strategies for oligonucleotide traversion across the blood-brain barrier. *J. Control Release* 354, 554–571. doi:10.1016/j.jconrel.2023.01.031

Atkinson, A. J. (2017). Intracerebroventricular drug administration. *Transl. Clin. Pharmacol.* 25, 117–124. doi:10.12793/tcp.2017.25.3.117

Barker, S. J., Thayer, M. B., Kim, C., Tatarakis, D., Simon, M., Dial, R. L., et al. (2023). Targeting transferrin receptor to transport antisense oligonucleotides across the blood-brain barrier. *bioRxiv*.

Begines, B., Ortiz, T., Perez-Aranda, M., Martinez, G., Merinero, M., Arguelles-Arias, F., et al. (2020). Polymeric nanoparticles for drug delivery: recent developments and future prospects. *Nanomater. (Basel)* 10, 1403. doi:10.3390/nano10071403

Berdynski, M., Miszta, P., Safranow, K., Andersen, P. M., Morita, M., Filipek, S., et al. (2022). SOD1 mutations associated with amyotrophic lateral sclerosis analysis of variant severity. *Sci. Rep.* 12, 103. doi:10.1038/s41598-021-03891-8

Biscans, A., Caiazzi, J., Davis, S., Mchugh, N., Sousa, J., and Khvorova, A. (2020). The chemical structure and phosphorothioate content of hydrophobically modified siRNAs impact extrahepatic distribution and efficacy. *Nucleic Acids Res.* 48, 7665–7680. doi:10.1093/nar/gkaa595

Borgonetti, V., and Galeotti, N. (2021). Intranasal delivery of an antisense oligonucleotide to the RNA-binding protein HuR relieves nerve injury-induced neuropathic pain. *Pain* 162, 1500–1510. doi:10.1097/j.pain.0000000000002154

Borgonetti, V., and Galeotti, N. (2023). Posttranscriptional regulation of gene expression participates in the myelin restoration in mouse models of multiple sclerosis: antisense modulation of HuR and HuD ELAV RNA binding protein. *Mol. Neurobiol.* 60, 2661–2677. doi:10.1007/s12035-023-03236-8

Bost, J. P., Barriga, H., Holme, M. N., Gallud, A., Maugeri, M., Gupta, D., et al. (2021). Delivery of oligonucleotide therapeutics: chemical modifications, lipid nanoparticles, and extracellular vesicles. *ACS Nano* 15, 13993–14021. doi:10.1021/acsnano.1c05099

Brooke, G., Low, A., Cheryl, W., Michael, M., Fazio, M., Berdeja, A., et al. (2021). Towards next generation antisense oligonucleotides: mesylphosphoramidate modification improves therapeutic index and duration of effect of gapmer antisense oligonucleotides. *Nucleic Acids Res.* 49, 9026–9041. doi:10.1093/nar/gkab718

Brown, K. M., Nair, J. K., Janas, M. M., Anglero-Rodriguez, Y. I., Dang, L. T. H., Peng, H., et al. (2022). Expanding RNAi therapeutics to extrahepatic tissues with lipophilic conjugates. *Nat. Biotechnol.* 40, 1500–1508. doi:10.1038/s41587-022-01334-x

Cecchin, R., Troyer, Z., Witwer, K., and Morris, K. V. (2023). Extracellular vesicles: the next generation in gene therapy delivery. *Mol. Ther.* 31, 1225–1230. doi:10.1016/j.ymthe.2023.01.021

Chu, C., Jablonska, A., Gao, Y., Lan, X., Lesniak, W. G., Liang, Y., et al. (2022). Hyperosmolar blood–brain barrier opening using intra-arterial injection of hyperosmotic mannitol in mice under real-time MRI guidance. *Nat. Protoc.* 17, 76–94. doi:10.1038/s41596-021-00634-x

Conroy, F., Miller, R., Alterman, J. F., Hassler, M. R., Echeverria, D., Godinho, B., et al. (2022). Chemical engineering of therapeutic siRNAs for allele-specific gene silencing in Huntington's disease models. *Nat. Commun.* 13, 5802. doi:10.1038/s41467-022-33061-x

Cook, M., Murphy, M., Bulluss, K., D'souza, W., Plummer, C., Priest, E., et al. (2020). Anti-seizure therapy with a long-term, implanted intra-cerebroventricular delivery system for drug-resistant epilepsy: a first-in-man study. *EClinicalMedicine* 22, 100326. doi:10.1016/j.eclinm.2020.100326

Craig, K., Abrams, M., and Amiji, M. (2018). Recent preclinical and clinical advances in oligonucleotide conjugates. *Expert Opin. Drug Deliv.* 15, 629–640. doi:10.1080/17425247.2018.1473375

Crooke, S. T., Seth, P. P., Vickers, T. A., and Liang, X. H. (2020a). The interaction of phosphorothioate-containing RNA targeted drugs with proteins is a critical determinant of the therapeutic effects of these agents. *J. Am. Chem. Soc.* 142, 14754–14771. doi:10.1021/jacs.0c04928

Crooke, S. T., Vickers, T. A., and Liang, X. H. (2020b). Phosphorothioate modified oligonucleotide-protein interactions. *Nucleic Acids Res.* 48, 5235–5253. doi:10.1093/nar/gkaa299

Crooke, S. T., Wang, S., Vickers, T. A., Shen, W., and Liang, X.-H. (2017). Cellular uptake and trafficking of antisense oligonucleotides. *Nat. Biotechnol.* 35, 230–237. doi:10.1038/nbt.3779

Crowe, T. P., and Hsu, W. H. (2022). Evaluation of recent intranasal drug delivery systems to the central nervous system. *Pharmaceutics* 14, 629. doi:10.3390/pharmaceutics14030629

Dastpeyman, M., Sharifi, R., Amin, A., Karas, J. A., Cuic, B., Pan, Y., et al. (2021). Endosomal escape cell-penetrating peptides significantly enhance pharmacological effectiveness and CNS activity of systemically administered antisense oligonucleotides. *Int. J. Pharm.* 599, 120398. doi:10.1016/j.ijpharm.2021.120398

Dhamankar, V., and Donovan, M. D. (2017). Modulating nasal mucosal permeation using metabolic saturation and enzyme inhibition techniques. *J. Pharm. Pharmacol.* 69, 1075–1083. doi:10.1111/jphp.12749

Dhuri, K., Bechtold, C., Quijano, E., Pham, H., Gupta, A., Vikram, A., et al. (2020). Antisense oligonucleotides: an emerging area in drug discovery and development. *J. Clin. Med.* 9, 2004. doi:10.3390/jcm9062004

Dowdy, S. F. (2023). Endosomal escape of RNA therapeutics: how do we solve this rate-limiting problem? *RNA* 29, 396–401. doi:10.1261/rna.079507.122

D'souza, A., Nozohouri, S., Bleier, B. S., and Amiji, M. M. (2023). CNS delivery of nucleic acid therapeutics: beyond the blood–brain barrier and towards specific cellular targeting. *Pharm. Res.* 40, 77–105. doi:10.1007/s11095-022-03433-5

Duma, C., Kopyov, O., Kopyov, A., Berman, M., Lander, E., Elam, M., et al. (2019). Human intracerebroventricular (ICV) injection of autologous, non-engineered, adipose-derived stromal vascular fraction (ADSVF) for neurodegenerative disorders: results of a 3-year phase 1 study of 113 injections in 31 patients. *Mol. Biol. Rep.* 46, 5257–5272. doi:10.1007/s11033-019-04983-5

Eyford, B. A., Singh, C. S. B., Abraham, T., Munro, L., Choi, K. B., Hill, T., et al. (2021). A nanomule peptide carrier delivers siRNA across the intact blood-brain barrier to attenuate ischemic stroke. *Front. Mol. Biosci.* 8, 611367. doi:10.3389/fmolb.2021.611367

Falcone, J. A., Salameh, T. S., Yi, X., Cordy, B. J., Mortell, W. G., Kabanov, A. V., et al. (2014). Intranasal administration as a route for drug delivery to the brain: evidence for a unique pathway for albumin. *J. Pharmacol. Exp. Ther.* 351, 54–60. doi:10.1124/jpet.114.216705

Finicle, B. T., Eckenstein, K. H., Revenko, A. S., Anderson, B. A., Wan, W. B., Mccracken, A. N., et al. (2023). Simultaneous inhibition of endocytic recycling and lysosomal fusion sensitizes cells and tissues to oligonucleotide therapeutics. *Nucleic Acids Res.* 51, 1583–1599. doi:10.1093/nar/gkad023

Gabold, B., Adams, F., Brameyer, S., Jung, K., Ried, C. L., Merdan, T., et al. (2023). Transferrin-modified chitosan nanoparticles for targeted nose-to-brain delivery of proteins. *Drug Deliv. Transl. Res.* 13, 822–838. doi:10.1007/s13346-022-01245-z

Gagliardi, M., and Ashizawa, A. T. (2021). The challenges and strategies of antisense oligonucleotide drug delivery. *Biomedicines* 9, 433. doi:10.3390/biomedicines9040433

Gauthier-Loiselle, M., Cloutier, M., Toro, W., Patel, A., Shi, S., Davidson, M., et al. (2021). Nusinersen for spinal muscular atrophy in the United States: findings from a retrospective claims database analysis. *Adv. Ther.* 38, 5809–5828. doi:10.1007/s12325-021-01938-w

Gokirmak, T., Nikan, M., Wiechmann, S., Prakash, T. P., Tanowitz, M., and Seth, P. P. (2021). Overcoming the challenges of tissue delivery for oligonucleotide therapeutics. *Trends Pharmacol. Sci.* 42, 588–604. doi:10.1016/j.tips.2021.04.010

Goto, A., Yamamoto, S., and Iwasaki, S. (2023). Biodistribution and delivery of oligonucleotide therapeutics to the central nervous system: advances, challenges, and future perspectives. *Biopharm. Drug Dispos.* 44, 26–47. doi:10.1002/bdd.2338

Hald Albertsen, C., Kulkarni, J. A., Witzigmann, D., Lind, M., Petersson, K., and Simonsen, J. B. (2022). The role of lipid components in lipid nanoparticles for vaccines and gene therapy. *Adv. Drug Deliv. Rev.* 188, 114416. doi:10.1016/j.addr.2022.114416

Halloy, F., Biscans, A., Bujold, K. E., Debacker, A., Hill, A. C., Lacroix, A., et al. (2022). Innovative developments and emerging technologies in RNA therapeutics. *RNA Biol.* 19, 313–332. doi:10.1080/15476286.2022.2027150

Hammond, S. M., Abendroth, F., Goli, L., Stoodley, J., Burrell, M., Thom, G., et al. (2022). Antibody-oligonucleotide conjugate achieves CNS delivery in animal models for spinal muscular atrophy. *JCI Insight* 7, e154142. doi:10.1172/jci.insight.154142

Hara Prasad, M. (2023). Tofersen - a ray of hope for ALS patients: a mini review. *J. Pharmacovigil. Drug Res.* 4, 1–3. doi:10.53411/jpadr.2023.4.2.1

Holm, A., Hansen, S. N., Klitgaard, H., and Kauppinen, S. (2022). Clinical advances of RNA therapeutics for treatment of neurological and neuromuscular diseases. *RNA Biol.* 19, 594–608. doi:10.1080/15476286.2022.2066334

Iannaccone, S. T., Paul, D., Castro, D., Weprin, B., and Swift, D. (2021). Delivery of nusinersen through an Ommaya reservoir in spinal muscular atrophy. *J. Clin. Neuromuscul. Dis.* 22, 129–134. doi:10.1097/CND.0000000000000333

Ionis Pharmaceuticals, I. (2023). *Eplontersen halted ATTRv-PN disease progression and improved neuropathy impairment and quality of life in Phase 3 study through 66 weeks*. Carlsbad, California: PRNewswire.

Jeon, J. Y., Ayyar, V. S., and Mitra, A. (2022). Pharmacokinetic and pharmacodynamic modeling of siRNA therapeutics – A minireview. *Pharm. Res.* 39, 1749–1759. doi:10.1007/s11095-022-03333-8

Jeong, S.-H., Jang, J.-H., and Lee, Y.-B. (2023). Drug delivery to the brain via the nasal route of administration: exploration of key targets and major consideration factors. *J. Pharm. Investigation* 53, 119–152. doi:10.1007/s40005-022-00589-5

Juliano, R. L., Ming, X., and Nakagawa, O. (2012). Cellular uptake and intracellular trafficking of antisense and siRNA oligonucleotides. *Bioconjug Chem.* 23, 147–157. doi:10.1021/bc200377d

Kasina, V., Mownn, R. J., Bahal, R., and Sartor, G. C. (2022). Nanoparticle delivery systems for substance use disorder. *Neuropsychopharmacology* 47, 1431–1439. doi:10.1038/s41386-022-01311-7

Khongkow, M., Yata, T., Boonrungsiman, S., Ruktanonchai, U. R., Graham, D., and Namdee, K. (2019). Surface modification of gold nanoparticles with neuron-targeted exosome for enhanced blood-brain barrier penetration. *Sci. Rep.* 9, 8278. doi:10.1038/s41598-019-44569-6

Khorkova, O., Hsiao, J., and Wahlestedt, C. (2021). Nucleic acid-based therapeutics in orphan neurological disorders: recent developments. *Front. Mol. Biosci.* 8, 643681. doi:10.3389/fmolb.2021.643681

Khorkova, O., Stahl, J., Joji, A., Volmar, C.-H., and Wahlestedt, C. (2023). Amplifying gene expression with RNA-targeted therapeutics. *Nat. Rev. Drug Discov.* 22, 539–561. doi:10.1038/s41573-023-00704-7

Khorkova, O., and Wahlestedt, C. (2017). Oligonucleotide therapies for disorders of the nervous system. *Nat. Biotechnol.* 35, 249–263. doi:10.1038/nbt.3784

Kim, J., Hu, C., Moufawad El Achkar, C., Black, L. E., Douville, J., Larson, A., et al. (2019). Patient-customized oligonucleotide therapy for a rare genetic disease. *N. Engl. J. Med.* 381, 1644–1652. doi:10.1056/NEJMoa1813279

Kim, Y.-K. (2022). RNA therapy: rich history, various applications and unlimited future prospects. *Exp. Mol. Med.* 54, 455–465. doi:10.1038/s12276-022-00757-5

Knox, E. G., Aburto, M. R., Clarke, G., Cryan, J. F., and O'driscoll, C. M. (2022). The blood-brain barrier in aging and neurodegeneration. *Mol. Psychiatry* 27, 2659–2673. doi:10.1038/s41380-022-01511-z

Koepsell, H. (2020). Glucose transporters in brain in health and disease. *Pflügers Archiv - Eur. J. Physiology* 472, 1299–1343. doi:10.1007/s00424-020-02441-x

Kus-Liskiewicz, M., Fickers, P., and Ben Tahar, I. (2021). Biocompatibility and cytotoxicity of gold nanoparticles: recent advances in methodologies and regulations. *Int. J. Mol. Sci.* 22, 10952. doi:10.3390/ijms222010952

Kutlehria, S., D'souza, A., Bleier, B. S., and Amiji, M. M. (2022). Role of 3D printing in the development of biodegradable implants for central nervous system drug delivery. *Mol. Pharm.* 19, 4411–4427. doi:10.1021/acs.molpharmaceut.2c00344

Lai, F., Damle, S. S., Ling, K. K., and Rigo, F. (2020). Directed RNase H cleavage of nascent transcripts causes transcription termination. *Mol. Cell.* 77, 1032–1043. doi:10.1016/j.molcel.2019.12.029

Lee, J. H., Lee, Y., Park, H. S., and Lee, J. H. (2022). Comparison of clinical efficacy of transforaminal and interlaminar epidural steroid injection in radicular pain due to cervical diseases: a systematic review and meta-analysis. *Pain Physician* 25, E1351–E1366.

Li, W., Qiu, J., Li, X. L., Aday, S., Zhang, J., Conley, G., et al. (2021). BBB pathophysiology-independent delivery of siRNA in traumatic brain injury. *Sci. Adv.* 7, eabd6889. doi:10.1126/sciadv.abd6889

Linninger, A. A., Barua, D., Hang, Y., Iadevaia, S., and Vakilynejad, M. (2023). A mechanistic pharmacokinetic model for intrathecal administration of antisense oligonucleotides. *Front. Physiol.* 14, 1130925. doi:10.3389/fphys.2023.1130925

Lochhead, J. J., and Thorne, R. G. (2012). Intranasal delivery of biologics to the central nervous system. *Adv. Drug Deliv. Rev.* 64, 614–628. doi:10.1016/j.addr.2011.11.002

Ma, F., Yang, L., Sun, Z., Chen, J., Rui, X., Glass, Z., et al. (2020). Neurotransmitter-derived lipidoids (NT-lipidoids) for enhanced brain delivery through intravenous injection. *Sci. Adv.* 6, eabb4429. doi:10.1126/sciadv.abb4429

Maher, R., Moreno-Borrallo, A., Jindal, D., Mai, B. T., Ruiz-Hernandez, E., and Harkin, A. (2023). Intranasal polymeric and lipid-based nanocarriers for CNS drug delivery. *Pharmaceutics* 15, 746. doi:10.3390/pharmaceutics15030746

Marangoni, K., and Menezes, R. (2022). RNA aptamer-functionalized polymeric nanoparticles in targeted delivery and cancer therapy: an up-to-date review. *Curr. Pharm. Des.* 28, 2785–2794. doi:10.2174/1381612828666220903120755

Mcmahon, M. A., Rahdar, M., Mukhopadhyay, S., Bui, H. H., Hart, C., Damle, S., et al. (2023). GOLGA8 increases bulk antisense oligonucleotide uptake and activity in mammalian cells. *Mol. Ther. Nucleic Acids* 32, 289–301. doi:10.1016/j.omtn.2023.03.017

Migliorati, J. M., Liu, S., Liu, A., Gogate, A., Nair, S., Bahal, R., et al. (2022). Absorption, distribution, metabolism, and excretion of US food and drug administration-approved antisense oligonucleotide drugs. *Drug Metab. Dispos.* 50, 888–897. doi:10.1124/dmd.121.000417

Miller, T., Cudkowicz, M., Shaw, P. J., Andersen, P. M., Atassi, N., Bucelli, R. C., et al. (2020). Phase 1-2 trial of antisense oligonucleotide tofersen for SOD1 ALS. *N. Engl. J. Med.* 383, 109–119. doi:10.1056/NEJMoa2003715

Miller, T. M., Cudkowicz, M. E., Genge, A., Shaw, P. J., Sobue, G., Bucelli, R. C., et al. (2022). Trial of antisense oligonucleotide tofersen for SOD1 ALS. *N. Engl. J. Med.* 387, 1099–1110. doi:10.1056/NEJMoa2204705

Min, H. S., Kim, H. J., Naito, M., Ogura, S., Toh, K., Hayashi, K., et al. (2020). Systemic brain delivery of antisense oligonucleotides across the blood–brain barrier with a glucose-coated polymeric nanocarrier. *Angew. Chem.* 132, 8173–8180. doi:10.1002/anie.201914751

Moradi, F., and Dashti, N. (2022). Targeting neuroinflammation by intranasal delivery of nanoparticles in neurological diseases: a comprehensive review. *Naunyn Schmiedeb. Arch. Pharmacol.* 395, 133–148. doi:10.1007/s00210-021-02196-x

Mortberg, M. A., Gentile, J. E., Nadaf, N. M., Vanderburg, C., Simmons, S., Dubinsky, D., et al. (2023). A single-cell map of antisense oligonucleotide activity in the brain. *Nucleic Acids Res.*, gkad371. doi:10.1093/nar/gkad371

Mummery, C. J., Borjesson-Hanson, A., Blackburn, D. J., Vijverberg, E. G. B., De Deyn, P. P., Ducharme, S., et al. (2023). Tau-targeting antisense oligonucleotide MAPT(rx) in mild alzheimer's disease: a phase 1b, randomized, placebo-controlled trial. *Nat. Med.* 29, 1437–1447. doi:10.1038/s41591-023-02326-3

Nagata, T., Dwyer, C. A., Yoshida-Tanaka, K., Ihara, K., Ohyagi, M., Kaburagi, H., et al. (2021). Cholesterol-functionalized DNA/RNA heteroduplexes cross the blood–brain barrier and knock down genes in the rodent CNS. *Nat. Biotechnol.* 39, 1529–1536. doi:10.1038/s41587-021-00972-x

Neil, E. E., and Bisaccia, E. K. (2019). Nusinersen: a novel antisense oligonucleotide for the treatment of spinal muscular atrophy. *J. Pediatr. Pharmacol. Ther.* 24, 194–203. doi:10.5863/1551-6776-24.3.194

O'reilly, D., Belgrad, J., Ferguson, C., Summers, A., Sapp, E., Mchugh, C., et al. (2023). Di-valent siRNA-mediated silencing of MSH3 blocks somatic repeat expansion in mouse models of Huntington's disease. *Mol. Ther.* 31, 1661–1674. doi:10.1016/j.ymthe.2023.05.006

Padmakumar, S., D'souza, A., Parayath, N. N., Bleier, B. S., and Amiji, M. M. (2022). Nucleic acid therapies for CNS diseases: pathophysiology, targets, barriers, and delivery strategies. *J. Control Release* 352, 121–145. doi:10.1016/j.jconrel.2022.10.018

Padmakumar, S., Jones, G., Pawar, G., Khorkova, O., Hsiao, J., Kim, J., et al. (2021). Minimally invasive nasal depot (MIND) technique for direct BDNF AntagoNAT delivery to the brain. *J. Control Release* 331, 176–186. doi:10.1016/j.jconrel.2021.01.027

Pechmann, A., Behrens, M., Dörnbrack, K., Tassoni, A., Stein, S., Vogt, S., et al. (2022). Effect of nusinersen on motor, respiratory and bulbar function in early-onset spinal muscular atrophy. *Brain* 146, 668–677. doi:10.1093/brain/awac252

Pollak, A. J., Zhao, L., and Crooke, S. T. (2023). Systematic analysis of chemical modifications of phosphorothioate antisense oligonucleotides that modulate their innate immune response. *Nucleic Acid. Ther.* 33, 95–107. doi:10.1089/nat.2022.0067

Prakash, T. P. (2011). An overview of sugar-modified oligonucleotides for antisense therapeutics. *Chem. Biodivers.* 8, 1616–1641. doi:10.1002/cbdv.201100081

Rabinowicz, A. L., Carrazana, E., and Maggio, E. T. (2021). Improvement of intranasal drug delivery with Intravail® alkylsaccharide excipient as a mucosal absorption enhancer aiding in the treatment of conditions of the central nervous system. *Drugs R. D.* 21, 361–369. doi:10.1007/s40268-021-00360-5

Renner, D. B., Frey, W. H., 2ndand Hanson, L. R. (2012). Intranasal delivery of siRNA to the olfactory bulbs of mice via the olfactory nerve pathway. *Neurosci. Lett.* 513, 193–197. doi:10.1016/j.neulet.2012.02.037

Roberts, T. C., Langer, R., and Wood, M. J. A. (2020). Advances in oligonucleotide drug delivery. *Nat. Rev. Drug Discov.* 19, 673–694. doi:10.1038/s41573-020-0075-7

Rosch, J. C., Hoogenboezem, E. N., Sorets, A. G., Duvall, C. L., and Lippmann, E. S. (2022). Albumin-binding aptamer chimeras for improved siRNA bioavailability. *Cell. Mol. Bioeng.* 15, 161–173. doi:10.1007/s12195-022-00718-y

Sava, V., Fihurka, O., Khvorova, A., and Sanchez-Ramos, J. (2021). Kinetics of HTT lowering in brain of YAC 128 mice following single and repetitive intranasal dosing of siRNA packaged in chitosan-based nanoparticle. *J. Drug Deliv. Sci. Technol.* 63, 102517. doi:10.1016/j.jddst.2021.102517

Sewing, S., Roth, A. B., Winter, M., Dieckmann, A., Bertinetti-Lapatki, C., Tessier, Y., et al. (2017). Assessing single-stranded oligonucleotide drug-induced effects *in vitro* reveals key risk factors for thrombocytopenia. *PLoS One* 12, e0187574. doi:10.1371/journal.pone.0187574

Shen, W., De Hoyos, C. L., Migawa, M. T., Vickers, T. A., Sun, H., Low, A., et al. (2019). Chemical modification of PS-ASO therapeutics reduces cellular protein-binding and improves the therapeutic index. *Nat. Biotechnol.* 37, 640–650. doi:10.1038/s41587-019-0106-2

Shi, L., Zhang, J., Zhao, M., Tang, S., Cheng, X., Zhang, W., et al. (2021). Effects of polyethylene glycol on the surface of nanoparticles for targeted drug delivery. *Nanoscale* 13, 10748–10764. doi:10.1039/d1nr02065j

Song, H., Zhang, J., Liu, B., Xu, J., Cai, B., Yang, H., et al. (2022). Biological roles of RNA m(5)C modification and its implications in Cancer immunotherapy. *Biomark. Res.* 10, 15. doi:10.1186/s40364-022-00362-8

Standifer, K. M., Chien, C. C., Wahlestedt, C., Brown, G. P., and Pasternak, G. W. (1994). Selective loss of delta opioid analgesia and binding by antisense oligodeoxynucleotides to a delta opioid receptor. *Neuron* 12, 805–810. doi:10.1016/0896-6273(94)90333-6

Sun, K., Zheng, X., Jin, H., Yu, F., and Zhao, W. (2022). Exosomes as CNS drug delivery tools and their applications. *Pharmaceutics* 14, 2252. doi:10.3390/pharmaceutics14102252

Takakusa, H., Iwazaki, N., Nishikawa, M., Yoshida, T., Obika, S., and Inoue, T. (2023). Drug metabolism and pharmacokinetics of antisense oligonucleotide therapeutics: typical profiles, evaluation approaches, and points to consider compared with small molecule drugs. *Nucleic Acid. Ther.* 33, 83–94. doi:10.1089/nat.2022.0054

Tang, L., Zhang, R., Wang, Y., Zhang, X., Yang, Y., Zhao, B., et al. (2023). A simple self-assembly nanomicelle based on brain tumor-targeting peptide-mediated siRNA delivery for glioma immunotherapy via intranasal administration. *Acta Biomater.* 155, 521–537. doi:10.1016/j.actbio.2022.11.013

Thakkar, H., Vaghela, D., and Patel, B. P. (2021). Brain targeted intranasal *in-situ* gelling spray of paroxetine: formulation, characterization and *in-vivo* evaluation. *J. Drug Deliv. Sci. Technol.* 62, 102317. doi:10.1016/j.jddst.2020.102317

Thomsen, M. S., Johnsen, K. B., Kucharz, K., Lauritzen, M., and Moos, T. (2022). Blood–brain barrier transport of transferrin receptor-targeted nanoparticles. *Pharmaceutics* 14, 2237. doi:10.3390/pharmaceutics14102237

Traber, G. M., and Yu, A. M. (2023). RNAi-based therapeutics and novel RNA bioengineering technologies. *J. Pharmacol. Exp. Ther.* 384, 133–154. doi:10.1124/jpet.122.001234

Valenzuela, A., Ayuso, M., Buyssens, L., Bars, C., Van Ginneken, C., Tessier, Y., et al. (2023). Platelet activation by antisense oligonucleotides (ASOs) in the gottingen minipig, including an evaluation of glycoprotein VI (GPVI) and platelet factor 4 (PF4) ontogeny. *Pharmaceutics* 15, 1112. doi:10.3390/pharmaceutics15041112

Vasquez, G., Migawa, M. T., Wan, W. B., Low, A., Tanowitz, M., Swayze, E. E., et al. (2022). Evaluation of phosphorus and non-phosphorus neutral oligonucleotide backbones for enhancing therapeutic index of gapmer antisense oligonucleotides. *Nucleic Acid. Ther.* 32, 40–50. doi:10.1089/nat.2021.0064

Veerapandiyan, A., Eichinger, K., Guntrum, D., Kwon, J., Baker, L., Collins, E., et al. (2020). Nusinersen for older patients with spinal muscular atrophy: a real-world clinical setting experience. *Muscle Nerve* 61, 222–226. doi:10.1002/mus.26769

Wahlestedt, C., Golanov, E., Yamamoto, S., Yee, F., Ericson, H., Yoo, H., et al. (1993a). Antisense oligodeoxynucleotides to NMDA-R1 receptor channel protect cortical neurons from excitotoxicity and reduce focal ischaemic infarctions. *Nature* 363, 260–263. doi:10.1038/363260a0

Wahlestedt, C., Pich, E. M., Koob, G. F., Yee, F., and Heilig, M. (1993b). Modulation of anxiety and neuropeptide Y-Y1 receptors by antisense oligodeoxynucleotides. *Science* 259, 528–531. doi:10.1126/science.8380941

Wave Life Sciences Usa, I. (2022). *Wave life Sciences announces positive update from phase 1b/2a SELECT-HD trial with initial results indicating allele-selective target engagement with WVE-003 in Huntington's disease.*

Xie, H., Li, L., Sun, Y., Wang, Y., Gao, S., Tian, Y., et al. (2019). An available strategy for nasal brain transport of nanocomposite based on PAMAM dendrimers via *in situ* gel. *Nanomater. (Basel)* 9, 147. doi:10.3390/nano9020147

Yamamoto, Y., Sanwald Ducray, P., Bjornsson, M., Smart, K., Grimsey, P., Vatakuti, S., et al. (2023). Development of a population pharmacokinetic model to characterize the pharmacokinetics of intrathecally administered tominersen in cerebrospinal fluid and plasma. *CPT Pharmacometrics Syst. Pharmacol.* doi:10.1002/psp4.13001

Yazdani, S., Jaldin-Fincati, J. R., Pereira, R. V. S., and Klip, A. (2019). Endothelial cell barriers: transport of molecules between blood and tissues. *Traffic* 20, 390–403. doi:10.1111/tra.12645

Yoshida, M., Oda, C., Mishima, K., Tsuji, I., Obika, S., and Shimojo, M. (2023). An antisense amido-bridged nucleic acid gapmer oligonucleotide targeting SRRM4 alters REST splicing and exhibits anti-tumor effects in small cell lung cancer and prostate cancer cells. *Cancer Cell. Int.* 23, 8. doi:10.1186/s12935-022-02842-1

Zhang, L., Wu, T., Shan, Y., Li, G., Ni, X., Chen, X., et al. (2021). Therapeutic reversal of Huntington's disease by *in vivo* self-assembled siRNAs. *Brain* 144, 3421–3435. doi:10.1093/brain/awab354

Zhou, Y., Zhu, F., Liu, Y., Zheng, M., Wang, Y., Zhang, D., et al. (2020). Blood-brain barrier-penetrating siRNA nanomedicine for Alzheimer's disease therapy. *Sci. Adv.* 6, eabc7031. doi:10.1126/sciadv.abc7031

Zingariello, C. D., Brandsema, J., Drum, E., Henderson, A. A., Dubow, S., Glanzman, A. M., et al. (2019). A multidisciplinary approach to dosing nusinersen for spinal muscular atrophy. *Neurol. Clin. Pract.* 9, 424–432. doi:10.1212/CPJ.0000000000000718

Zogg, H., Singh, R., and Ro, S. (2022). Current advances in RNA therapeutics for human diseases. *Int. J. Mol. Sci.* 23, 2736. doi:10.3390/ijms23052736

Zubair, A., and De Jesus, O. (2022). *Ommaya reservoir*. Treasure Island (FL): StatPearls Publishing.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 September 2023
DOI 10.3389/fgene.2023.1254226

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Jianhua Jia,
Jingdezhen Ceramic Institute, China
Cuncong Zhong,
University of Kansas, United States

*CORRESPONDENCE
François Major,
francois.major@umontreal.ca

RECEIVED 06 July 2023
ACCEPTED 16 August 2023
PUBLISHED 04 September 2023

CITATION
Chasles S and Major F (2023), Automatic recognition of complementary strands: lessons regarding machine learning abilities in RNA folding.
Front. Genet. 14:1254226.
doi: 10.3389/fgene.2023.1254226

Automatic recognition of complementary strands: lessons regarding machine learning abilities in RNA folding

Simon Chasles[1,2] and François Major[1,2]*

[1]Institute for Research in Immunology and Cancer, Montréal, QC, Canada, [2]Department of Computer Science and Operations Research, Université de Montréal, Montréal, QC, Canada

Introduction: Prediction of RNA secondary structure from single sequences still needs substantial improvements. The application of machine learning (ML) to this problem has become increasingly popular. However, ML algorithms are prone to overfitting, limiting the ability to learn more about the inherent mechanisms governing RNA folding. It is natural to use high-capacity models when solving such a difficult task, but poor generalization is expected when too few examples are available.

Methods: Here, we report the relation between capacity and performance on a fundamental related problem: determining whether two sequences are fully complementary. Our analysis focused on the impact of model architecture and capacity as well as dataset size and nature on classification accuracy.

Results: We observed that low-capacity models are better suited for learning with mislabelled training examples, while large capacities improve the ability to generalize to structurally dissimilar data. It turns out that neural networks struggle to grasp the fundamental concept of base complementarity, especially in lengthwise extrapolation context.

Discussion: Given a more complex task like RNA folding, it comes as no surprise that the scarcity of useable examples hurdles the applicability of machine learning techniques to this field.

KEYWORDS

RNA folding, base complementarity, machine learning, neural networks, binary classification, artificial data

Introduction

Identifying potential structural candidates for a single RNA sequence is a computationally demanding task. The Zuker-style dynamic programming approach to fold an RNA sequence of length L without pseudoknots requires time complexity in $\mathcal{O}(L^3)$ (Zuker and Stiegler 1981; Hofacker et al., 1994). Algorithms that take into account pseudoknots are even more complex and have been reported to require significantly more computational power ranging from $\mathcal{O}(L^4)$ to $\mathcal{O}(L^6)$ (Rivas and Eddy 1999; Condon et al., 2004), or higher (Marchand et al., 2022).

Machine learning (ML) algorithms offer an alternative to traditional methods for identifying RNA structural candidates. In particular, neural networks can compute structures in an end-to-end fashion, allowing for quick inference in a single feedforward

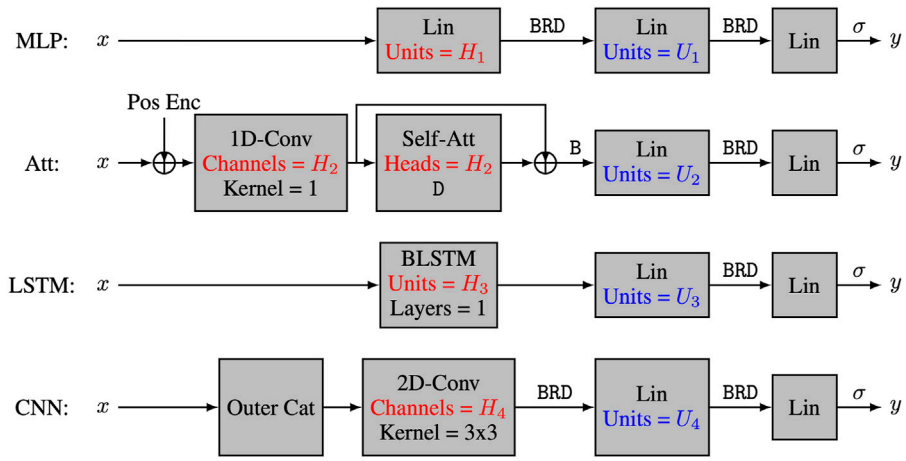


FIGURE 1
The four neural network architectures. The 4 tested neural network architectures take nucleotide encodings as input and output the positivity score. All models have three layers, with the first layer being a characteristic layer, and the last two layers having the form Lin-B-R-D-Lin-σ, where Lin refers to a linear layer, Conv refers to a convolutional layer, Self-Att refers to a multi-head self-attention layer, BLSTM refers to a bidirectional long-short term memory layer, Pos Enc refers to positional encodings, and letters σ, B, R and D refer respectively to sigmoid activation, batch normalization, ReLU activation and dropout regularization. The capacity of the models is controlled by hyperparameters H_i and U_i.

TABLE 1 Hyper-parameters used to control the capacity of each model.

$\log_{10}C$	≈ 2.70	≈ 3.00	≈ 3.50	≈ 4.15	≈ 4.75	≈ 5.50
MLP (H_1, U_1)	(5, 5)	(10, 10)	(30, 30)	(80, 80)	(200, 200)	(500, 500)
Att (H_2, U_2)	(2, 9)	(4, 12)	(6, 24)	(12, 60)	(24, 120)	(80, 180)
LSTM (H_3, U_3)	(2, 5)	(3, 8)	(4, 18)	(10, 35)	(16, 78)	(48, 145)
CNN (H_4, U_4)	(1, 1)	(1, 2)	(2, 4)	(4, 9)	(8, 16)	(20, 36)

H_i refers to the capacity of the characteristic layer and U_i indicates the number of units in the second to last linear layer. The quantity $\log_{10}C$ gives an order of magnitude for the number C of trainable parameters.

pass, especially when running on GPU (Chen et al., 2020; Fu et al., 2022). However, the training phase can take several days, which can complicate software updates (Shen et al., 2022). Regardless of the approach, predicting RNA structure requires significant computational resources due to the inherent complexity of RNA folding. As prediction methods must be at least as complex as the problems they aim to solve, it is important in the case of ML to avoid overfitting to this task.

To match the inherent complexity of RNA structure prediction, ML algorithms require high capacity, meaning they should be capable of learning a wide variety of mathematical functions (Goodfellow et al., 2016). Actually, because statistical learning relies heavily on training data, ML algorithms require high finite-sample expressivity to learn effectively (Zhang et al., 2021). However, high expressivity can lead to overfitting if the neural networks perform well on the training data without being able to generalize to structurally dissimilar testing data (LeCun et al., 1989). Therefore, it is crucial to balance expressivity with generalization to ensure accurate predictions on unseen data.

Several ML algorithms have been developed for RNA secondary structure prediction in recent years (Chen et al., 2020; Fu et al., 2022; Zakov et al., 2011; Singh et al., 2019; Wang et al., 2019). However, many of these algorithms are suspected of overfitting (Rivas et al., 2012; Sato et al., 2021; Zhao et al., 2021) and have limited ability to generalize across RNA families (Szikszai et al., 2022). Generalization is crucial for accurate RNA structure prediction since known RNA structures only represent a small fraction of the entire RNA structure space. Prediction algorithms must be able to accurately predict structures for molecules that are similar and dissimilar to known structures. Therefore, it is important to develop ML algorithms with stronger generalization properties to accurately predict RNA structures across a wider range of sequences and structures.

The aim of this study is to explore the performance and behavior of ML algorithms on a fundamental RNA-related task: determining if two RNA strands are fully complementary. Specifically, we examined the behavior of four families of neural networks with a focus on overfitting. We tackled three major challenges encountered when applying ML to RNA folding: 1) learning with mislabelled training examples (mislabels), 2) generalizing to structurally dissimilar data, and 3) training with limited examples (Rivas et al., 2012; Flamm et al., 2021; Burley et al., 2022; Danaee et al., 2018).

Our results indicate that low-capacity models are more effective for learning with mislabels, as they have the ability to

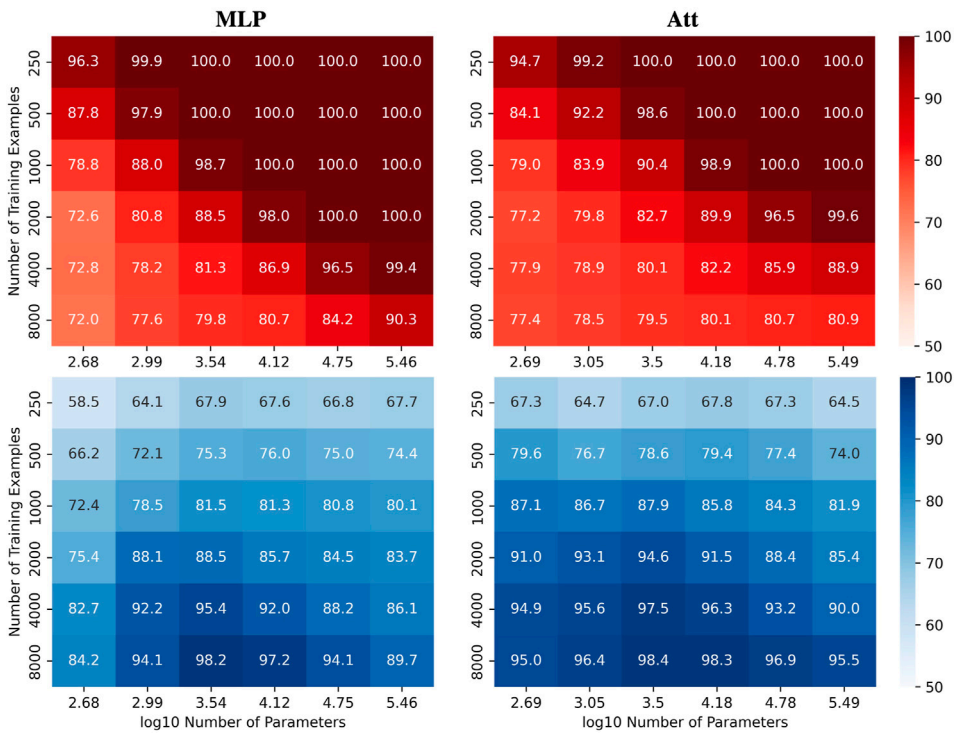


FIGURE 2
Performance of MLP and Att models when learning with mislabels. Train (red) and test (blue) mean accuracies over 50 simulations reported for MLP and Att models. Sequence length and mislabelling probability are fixed to $(L, \mu) = (8, 0.2)$.

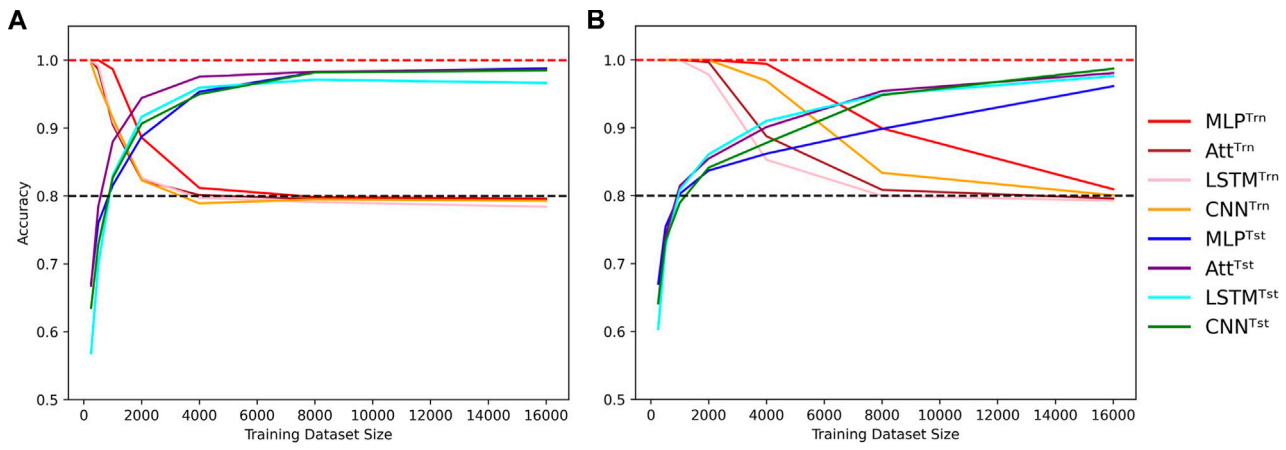


FIGURE 3
Cross-over behavior when learning with mislabels. Influence of the training dataset size over train (Trn) and test (Tst) accuracies for low-capacity models **(A)** and high-capacity models **(B)**. Sequence length is fixed to $L = 8$ and mislabelling probability is fixed to $\mu = 0.2$, with low capacity meaning $\log_{10}C \approx 3.5$ and high capacity meaning $\log_{10}C \approx 5.5$. Dotted lines indicate the 100% and 80% accuracy marks since μ =20%.

ignore them. Conversely, high-capacity models demonstrate better generalization performance in length-wise extrapolation context. On top of that, learning with few examples poses challenges for both low and high-capacity models, highlighting the importance of problem representation and architecture choice.

Materials and methods

Learning task

Given the RNA alphabet $\Sigma = \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{U}\}$, complementary strands are those in which each nucleotide on one strand pairs

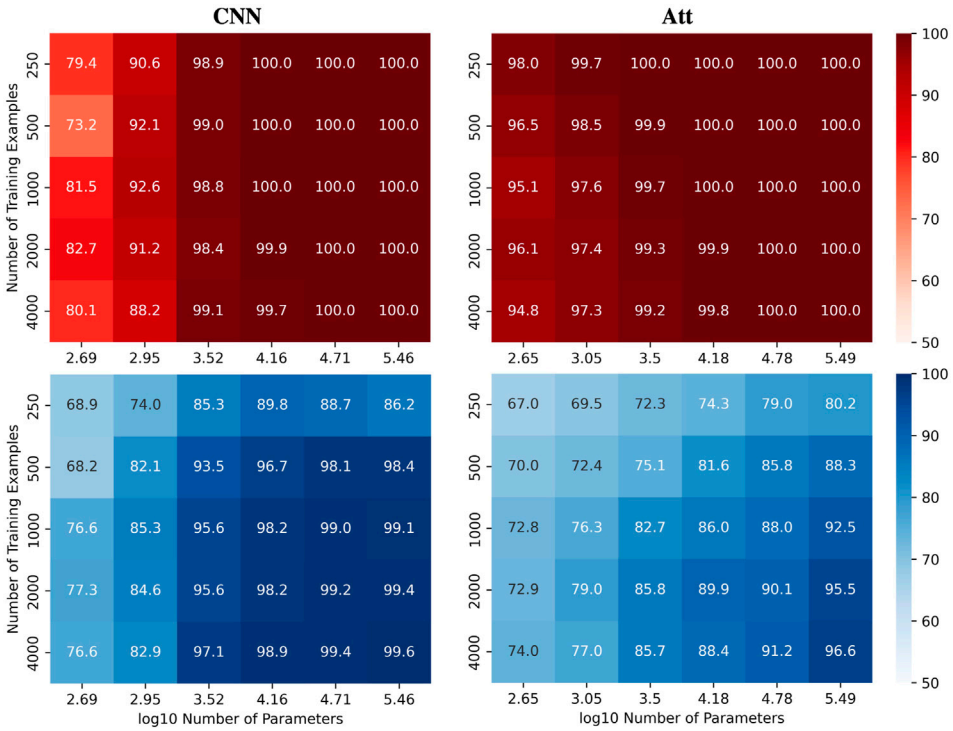


FIGURE 4
Performance of CNN and Att models in length-wise extrapolation context. Train (red) and test (blue) mean accuracies over 50 simulations reported for CNN and Att models. Models were trained on sequences of length 6 before being tested on sequences of length 8, without mislabelling in the training set.

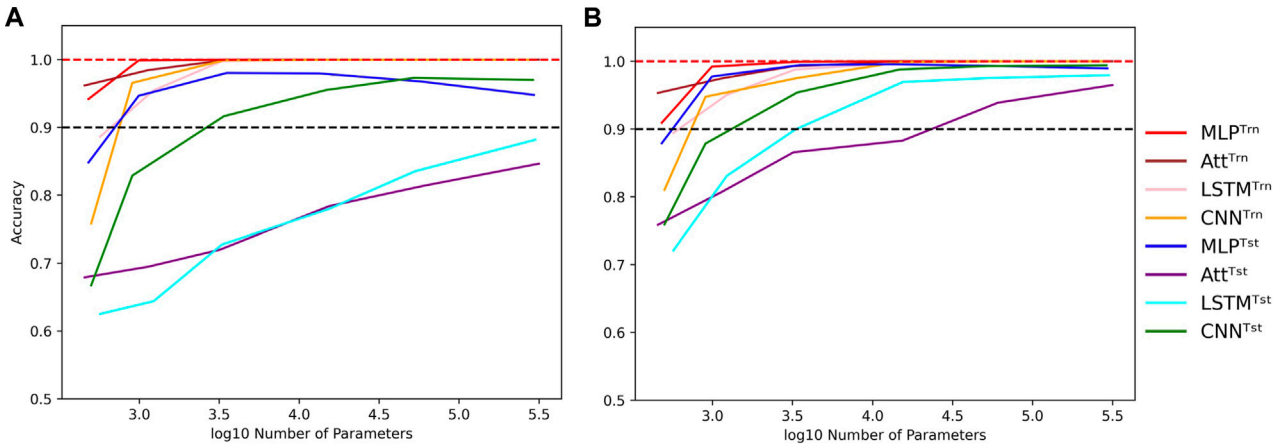


FIGURE 5
Influence of capacity in length-wise extrapolation context. Impact of the number of trainable parameters over train (Trn) and test (Tst) accuracies in length-wise extrapolation context. Results when training with 500 examples of length 5 shown in **(A)** and 2000 examples of length 6 in **(B)**. Sequence length is set to 8 in the testing datasets and mislabelling probability is set to 0 in the training datasets. Dotted lines indicate the 100% and 90% accuracy marks to highlight acceptable test performance.

with its Watson-Crick partner on the other strand (A with U and C with G). For example, the RNA strand 5′-AGUCAG-3′ is complementary to 5′-CUGACU-3′. We defined the task of automatic recognition of complementary strands as a binary classification problem that involves comparing and determining whether pairs of RNA strands of the same length are complementary or not. The target is True if the strands are fully complementary and False otherwise.

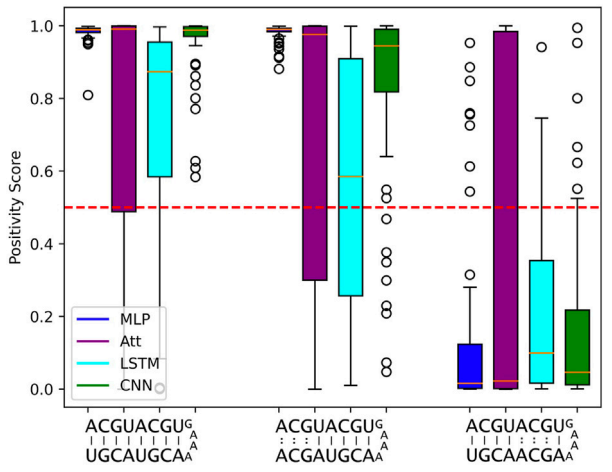


FIGURE 6
Limits to length-wise extrapolation with zero-padding and fixed-size inputs. Distribution of model outputs (positivity score) on specific sequences of length 8 when trained on 500 sequences of length 5. The first sequence is a positive example and the second and third ones are the same positive example where 3 mismatches have been introduced respectively at base pair positions 6 to 8 and 2 to 4. We use the best tested capacity for each model so $\log_{10}C \approx 3.5$ for the MLP and $\log_{10}C \approx 5.5$ for all other models. The classification threshold is represented by the red dotted line.

To have a fixed-size noisy input and simulate the structure of a hairpin loop, we restricted the maximum length of an RNA strand to 10 nucleotides and inserted an apical loop of 4 random nucleotides between each pair of RNA strands. During the experiments, the lengths of the apical loops varied from 3 to 7 nucleotides, but no significative impact was observed based on apical loop length, so we used the results computed with tetraloops as representatives of our key findings. More specifically, given strands $s, \bar{s} \in \Sigma^L$ with $L \leq 10$ and $a \in \Sigma^4$, we represented the input as the sequence $0^{10-L}sas\bar{0}^{10-L}$, where 0 denotes zero-padding.

The target of the classification is $t = 1$ if for all $i \in \{1, \ldots, L\}$, s_i is paired with its Watson-Crick complement on the other strand, that is, if $s_i = c(\bar{s}_{L+1-i})$, where c: $\Sigma \rightarrow \Sigma$ is the Watson-Crick complementarity function defined as $c(\mathtt{A}) = \mathtt{U}$, $c(\mathtt{C}) = \mathtt{G}$, $c(\mathtt{G}) = \mathtt{C}$, and $c(\mathtt{U}) = \mathtt{A}$. Otherwise, the target is $t = 0$.

Each nucleotide was represented by a vector of size 4, and a word embedding layer was used to update these vectors during training. Therefore, the inputs are fixed-size real-valued matrices $x \in \mathbb{R}^{24\times4}$ and the outputs are real values $y \in (0, 1)$. The output can be interpreted as the probability of the RNA sequence being a positive example, which we also refer to as the positivity score.

The loss function used to train the models is the binary cross-entropy with adjustments made to account for varying ratios of positive and negative examples. Specifically, for a training dataset $\mathcal{D} = \{(x^{(i)}, t^{(i)})\}_{i=1}^{N}$ with positive example ratio $\alpha = \frac{1}{N}\sum_{i=1}^{N} t^{(i)}$, the loss value for a model prediction $y^{(i)} = f(x^{(i)})$ was computed by Eq. 1.

$$l(y^{(i)}, t^{(i)}) = -\left[\left(\frac{1-\alpha}{\alpha}\right)t^{(i)}\log(y^{(i)}) + (1-t^{(i)})\log(1-y^{(i)})\right] \quad (1)$$

This correction ensured that the loss for positive examples was scaled up or down relative to the loss for negative examples, subject to the dataset's positive example ratio. For instance, if there were twice as many negative as positive examples in a dataset (i.e., $\alpha = 1/3$), the loss value for each positive example would be multiplied by a factor of 2, effectively balancing the contribution of positive and negative examples to the training loss. The parameter α could be controlled when creating synthetic datasets.

Artificial data

Our data generation process aimed to create diverse training and testing datasets that would enable us to evaluate our models in various scenarios. To achieve this, we introduced structural and

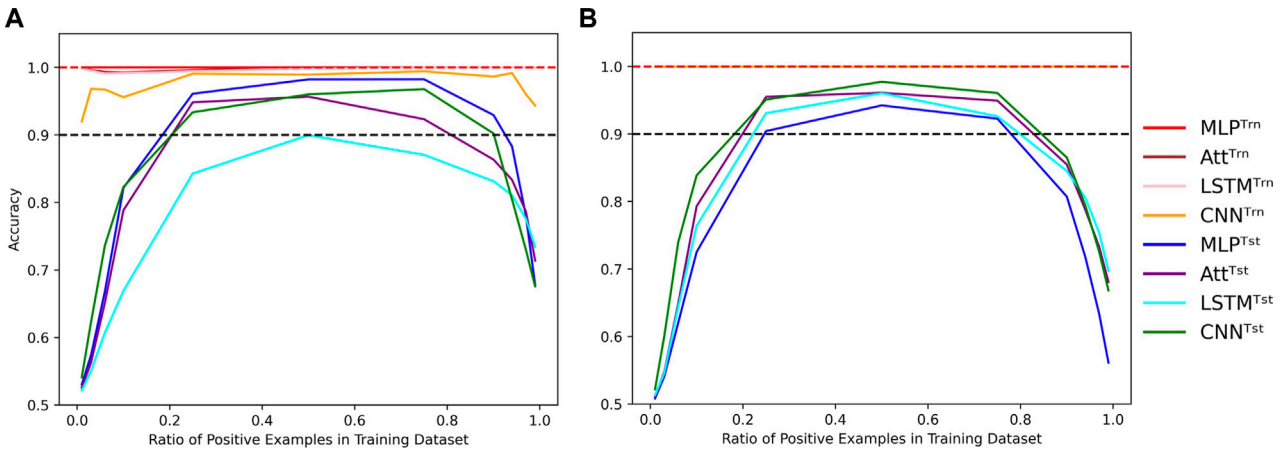


FIGURE 7
Evolution of performance in the extrapolation context of various positivity rates. Influence of the concentration of positive examples in the training set on the train (Trn) and test (Tst) accuracies for low-capacity models **(A)** and high-capacity models **(B)**. Parameters were fixed to $(N, L, \mu) = (500, 6, 0)$. Dotted lines indicate the 100% and 90% accuracy marks to highlight acceptable test performance.

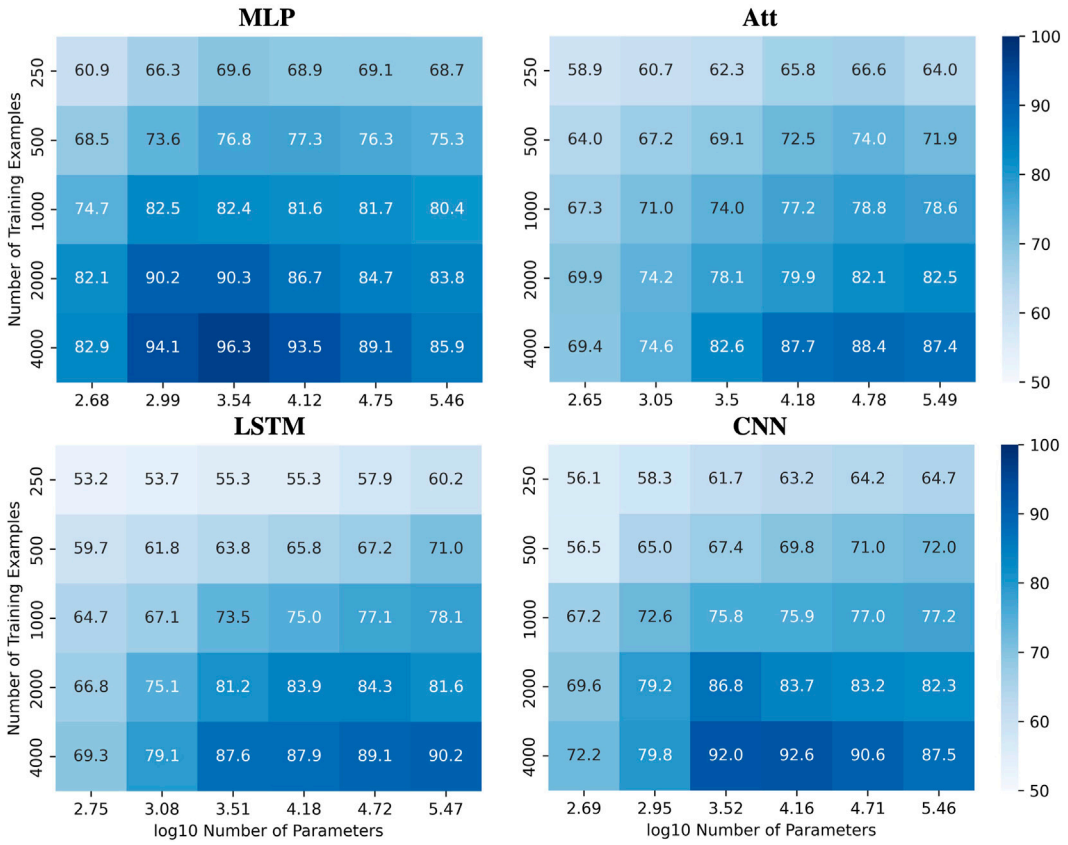


FIGURE 8
Performances for all models when learning with mislabels in extrapolation context. Test accuracies for all models when tested on sequences of length 8 after being trained on sequences of length 6 with 20% mislabelled training examples and 40% positivity rate.

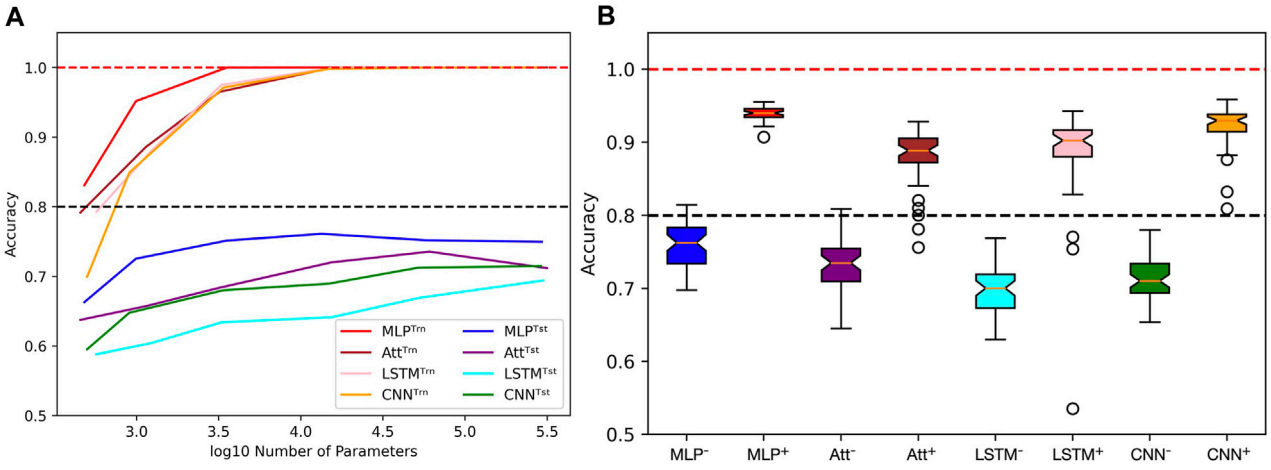


FIGURE 9
Limits when learning with few examples with mislabels in extrapolation context. **(A)** Influence of capacity on train (Trn) and test (tst) accuracies. The training sets are unbalanced and mislabelled as parameters (N, L, μ, α) are fixed to (500, 6, 0.2, 0.4). The testing sets on the opposite are balanced and correctly labelled with parameters (N, L, μ, α) being fixed to (4^8, 8, 0.0, 0.5). **(B)** With the same parameters, distributions of test accuracies over 50 simulations are reported, with models being trained on 500 examples (e.g., MLP$^-$) or 4,000 examples (e.g., MLP$^+$). The capacity used for each situation is the best capacity achieved on the test sets with respect to the heatmaps in Figure 8. Dotted lines indicate the 100% and 80% accuracy marks since μ = 20%.

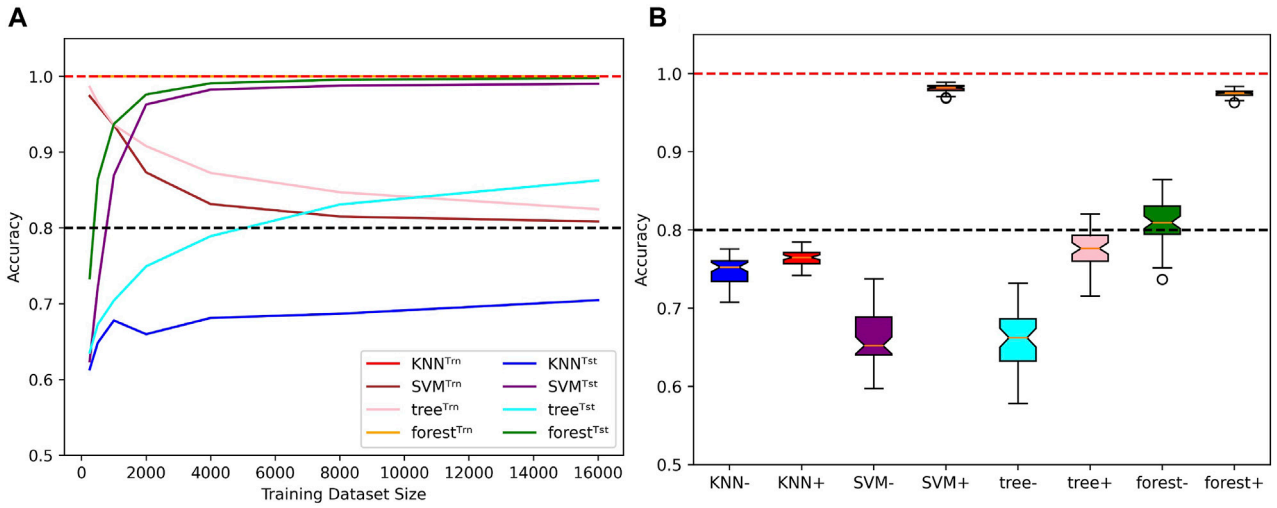


FIGURE 10
Behaviors of classical ML methods when extrapolating with mislabels. **(A)** Influence of the training dataset size over train (Trn) and test (Tst) accuracies for specific classical ML algorithms. Sequence length and mislabelling probability are fixed to $(L, \mu) = (8, 0.2)$. **(B)** With $(\mu, \alpha) = (0.2, 0.4)$, the models are trained on sequences of length 6 before being tested on the whole set of sequences of length 8. Distributions of test accuracies over 50 simulations are reported. The dataset size varies between $N = 500$ (e.g., KNN$^-$) and $N = 4{,}000$ (e.g., KNN$^+$). Dotted lines indicate the 100% and 80% accuracy marks since $\mu = 20\%$.

statistical dissimilarities between the datasets, including differences in quality, sequence length and positivity rate. Mainly, we aimed at simulating the use of all known data (a training set) to infer predictions on a portion of the unseen data of interest (a corresponding testing set).

To produce a pair of training and testing datasets, we ensured that there was no overlap between the training and testing examples by randomly partitioning the set of all 4^L sequences of length L into two sets. For each example, we concatenated a sequence $s \in \Sigma^L$ with a randomly generated apical loop a and a complementary sequence $\bar{s}$. Specifically, for positive examples, we set $\bar{s}$ to be the exact complement of s (denoted by $s^\star$), while for negative examples, we chose $\bar{s}$ randomly in $\Sigma^L \setminus \{s^\star\}$. We carefully controlled the ratio α of positive examples in the training sets and ensured that the testing sets had an equal number of positive and negative examples, thereby avoiding bias when measuring the performance on a test set.

To introduce structural dissimilarities, we varied the sequence length and the positive example ratio between the training and testing datasets. This allowed us to measure the extrapolation abilities of our models. Additionally, to control the quality of a training dataset, we introduced a proportion $\mu \in [0, \frac{1}{2}]$ of mislabelled examples, where a positive example could have label 0 with probability μ, and a negative example could have label 1 with probability μ. However, we ensured that all examples in the testing datasets were correctly labelled. By doing so, we tested the models' ability to ignore errors and avoid overfitting.

Overall, our data generation process produced realistic and diverse datasets, allowing us to evaluate our models under various conditions. Fully aware of the similarity between the task defined here and the prediction of binding sites of microRNAs, we are mostly interested in the abilities and behaviors of ML algorithms (and neural networks in particular) when trained and tested in various conditions, which we better control using artificial data.

Performance measure

To evaluate the performance of our models, we measured their classification accuracy on the testing datasets using a classification threshold $\theta \in (0, 1)$ to distinguish between examples of class 0 and 1. Specifically, given a model $f(\cdot)$ and a dataset $\mathcal{D} = \{(x^{(i)}, t^{(i)})\}_{i=1}^{N}$, the accuracy of $f(\cdot)$ on $\mathcal{D}$ with threshold θ was calculated by Eq. 2.

$$A(f(\cdot), \mathcal{D}, \theta) = \frac{1}{N}\sum_{i=1}^{N}\left(t^{(i)}\mathbb{1}\{f(x^{(i)}) > \theta\} + (1 - t^{(i)})\mathbb{1}\{f(x^{(i)}) \leq \theta\}\right) \tag{2}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function that equals 1 if the condition inside the bracket is true, and 0 otherwise.

While threshold θ could be optimized through methods such as SVM (Smola and Schölkopf 2004) or maximizing accuracy on the training dataset (Singh et al., 2019), we set $\theta = 1/2$ to study the behavior of our neural networks independently of any additional optimization steps. We considered the mean accuracy over 50 simulations, each of which generated datasets as described above, trained a model as described in the next section, and evaluated its performance using the aforementioned accuracy metric.

Architectures and training

The four tested models are relatively small representatives of four types of neural networks that have been recently used to predict RNA structures (Chen et al., 2020; Singh et al., 2019; Sato et al., 2021). All models have only three layers, where the first layer is unique to each architecture and the last two layers have the form linear—batchnorm—ReLU—dropout—linear-sigmoid (Goodfellow

et al., 2016; Ioffe and Szegedy 2015; Nair and Hinton 2010; Hinton et al., 2012). We used a dropout rate of 0.1 and a weight decay of 10^{-3} for all models, and the number of epochs was set to $\frac{8\times10^4}{N}$ to ensure that all models were trained for the same number of iterations, regardless of the number of training examples N. We used the Adam optimizer with a learning rate of 10^{-3} and default parameters in PyTorch (Kingma and Ba 2014) with a batch size of 256.

The four tested models and their capacity control are summarized in Figure 1. The multi-layer perceptron model (**MLP**) (Goodfellow et al., 2016) has a first layer of the form linear—batchnorm—ReLU—dropout, and its capacity is controlled by the number H_1 of hidden units in the linear module. For the multi-head self-attention model (**Att**), the first layer uses a skip connection of the form $h_1 = \texttt{conv}(x + \texttt{positionalencoding})$ and $h_2 = \texttt{batchnorm}(h_1 + \texttt{dropout}(\texttt{multi-head-self-attention}(h_1)))$, where the multi-head attention and positional encoding are described by Vaswani and co-workers (Vaswani et al., 2017). The capacity of this layer is controlled by the number H_2 of heads in the multi-head attention, so the 1D-convolution uses a kernel of size 1 to project the 4-dimensional embedding into a H_2-dimensional input for the attention module. For the long-short term memory model (**LSTM**) (Hochreiter and Schmidhuber 1997; Sak et al., 2014), the first layer is a one-layer bidirectional LSTM without dropout, and its capacity is controlled by the number H_3 of hidden units in the LSTM module. Finally, for the convolutional neural network (**CNN**) (LeCun et al., 1989; Dumoulin and Visin 2016), the first layer has the form outer concatenation—3 × 3conv—batchnorm—ReLU—dropout with a stride and padding of size 1. The outer concatenation operation takes the 24 × 4 input matrix and returns a 24 × 24 × 8 tensor, where the 8-dimensional vector at position i, j is the concatenation of the 4-dimensional encodings at positions i and j in the initial matrix. The capacity of the CNN is controlled by the number H_4 of feature maps produced by the convolution module.

To further manipulate the capacity of the models, we also varied the number U_i of hidden units in the second layer of our four architectures. Our main interest lies in the number C of parameters in each model, and more specifically in $\log_{10}C$. The actual number of heads, hidden units, and feature maps for each value of C is provided in Table 1. As discussed earlier, we trained 50 instances of each model on 50 different training datasets and reported the corresponding mean test accuracy. Our goal was to investigate how the test accuracy is affected by N and C for different values of L, μ and α.

Results and discussion

Learning with mislabels

The concept of mislabelling introduced in this section can be seen as a form of double standard. Our goal was to complicate the learning process deliberately by labelling some positive examples as class 0 and some negative examples as class 1. Thus, the term mislabelling is used because a ground truth classification is defined based on sequence complementarity. However, mislabelling on its own is not necessarily a problem for the learning process. A model $f(\cdot)$ that generalizes well on a completely mislabelled dataset could perform equally well on a correctly labelled dataset sampled from the same distribution by simply outputting $1 - f(\cdot)$. Therefore, the learning process is complicated when a proportion $\mu \in (0, 1)$ of the training dataset is mislabelled, creating a double standard as different pairs of complementary strands are labelled differently.

Moreover, since the labels can be interchanged without loss of generality, we focused on mislabelling probabilities $\mu \in (0, \frac{1}{2})$. The special case where $\mu = \frac{1}{2}$ behaves as if the labels were randomly assigned (Zhang et al., 2021). In this study, we used a mislabelling probability of $\mu = 0.2$, which we find to be a good representative of the key findings presented in this section.

In comparison to the actual RNA structure prediction task, learning with mislabels can be seen as analogous to learning on a dataset that contains structures with non-canonical base pairs as well as structures within which non-canonical base pairs are ignored, even though such pairs exist and can be identified. Similarly, the same concept applies when training on a dataset aimed at predicting pseudoknots and triplets when these kinds of interactions are only reported for some structures but not for others. Without defining the structural elements that need to be predicted, situations of double standard can arise when certain types of base pairs can be labelled as present (positive) as well as absent (negative) in a single dataset.

We were interested in investigating how the number of parameters C and the number of training examples N affect the test accuracy of the models for the automatic recognition of complementary strands. In particular, we focused on fixed sequence length $L = 8$ and mislabelling probability $\mu = 0.2$. The results for the MLP and Att models are presented in Figure 2, where heatmaps depict the performance of the models for different values of C and N. Shades of red depict train accuracies; blue testing. See Supplementary Figure S1 for the equivalent LSTM and CNN results. As expected, the accuracy increases with N, and overfitting behavior is observable for $\log_{10}C \geq 4.15$.

Remarkably, these results demonstrate the models' ability to handle mislabels, as they achieve a test accuracy of over 80% even when 20% of the training set is mislabelled. Moreover, the models can achieve near-perfect test accuracies as long as they are trained on a sufficiently large dataset. This finding suggests that models with relatively low capacity can still learn effectively when trained on low-quality high-quantity datasets.

Indeed, it appears that low-capacity models ($\log_{10}C \approx 3.5$) are more likely to achieve test accuracies above $1 - \mu$ than high-capacity models ($\log_{10}C \approx 5.5$). This is likely due to the fact that mislabels introduce irregularities into the sample space that are difficult for low-capacity models to account for. Low-capacity models tend to compute smoother functions than high-capacity models. They are thus less capable of capturing the intricate patterns that arise from mislabelling, especially in large training datasets. They have however enough capacity to estimate the function of interest, yielding test accuracies near 100% and train accuracies around $1 - \mu$.

On the other hand, high-capacity models can account for the irregularities introduced by the mislabels for larger training datasets, delaying the cross-over point of train and test accuracies to larger N. They thus require a broader view of the sample space to attain high test accuracies, coupled with sufficient regularization to prevent overfitting. The graphs in Figure 3 illustrate these behaviors. As in

Figure 2, shades of red and blue depict train and test accuracies respectively.

However, it is important to note that these reported accuracies may be too optimistic because the models were trained and tested on sequences of the same length with the same ratios of positive and negative examples. Although no training sequence was repeated in the testing set, this setup only evaluates a model's ability to generalize within structurally similar data, but not to extrapolate to structurally dissimilar data. While this section highlights the effectiveness of low-capacity models in learning with mislabels, the following section presents a contrast as high-capacity models appear to be more suitable for generalizing to structurally dissimilar data.

Generalizing to structurally dissimilar data

In the context of the automatic recognition of complementary strands, the ability to generalize to structurally dissimilar data refers to the ability of a model to make accurate predictions over a subspace of the sample space that it has not or poorly seen during training. This means that models could be trained and tested on datasets containing different sequence lengths and positivity rates to evaluate their understanding of sequence complementarity. For example, a model can be trained on sequences of length 5 or 6 and then tested on sequences of length 8, or trained on datasets with few or a lot of positive examples and then tested on balanced datasets. This approach allows us to investigate how well a model can generalize and extrapolate its understanding of the problem. In light of current discussions regarding extrapolation in ML conditions (Berrada et al. 2020; Balestriero et al. 2021), the concept of extrapolation is used here to convey how sequences of length 8 cannot belong to the convex hull of a set of sequences of length 6 with zero-padding.

In contrast, ML algorithms used for predicting actual RNA structure face different challenges. For instance, hardware limitations may restrict the maximum length of training sequences, but the model is still expected to predict the structure of longer sequences. The presence of non-canonical base pairs, pseudoknots, and base triples can increase the number of base pairs per nucleotide, making it difficult for models trained on sequences with a lower base pair density to generalize to more sophisticated structures. Moreover, different RNA families may have varying frequencies of certain structural motifs (Moore 1999), further complicating the generalization of models to unseen families. Despite these challenges, the ultimate goal of predicting RNA structure for all families remains the same, regardless of their structural similarity to previously known RNA structures.

To better visualize the impact of training example count N and the number of model parameters C on test accuracies for the automatic recognition of complementary strands, we report experiments on CNN and Att models. Specifically, the models were trained on correctly labelled sequences of length 6 and then tested on the full set of 4^8 sequences of length 8. We present heatmaps (Figure 4) that illustrate the performance of the models as the number of parameters and training examples are varied. See Supplementary Figure S2 for the equivalent MLP and LSTM results. We observed that higher model capacity tends to yield better performance, regardless of the number of training examples.

Nevertheless, as we increase the capacity, signs of overfitting can still be observed. Notably, in the graphs presented in Figure 5, we observe signs of overfitting from the MLP model when $\log_{10}C \geq 4$, for both $(L, N) = (5, 500)$ on the left and $(L, N) = (6, 2000)$ on the right.

Interestingly, we observe that the MLP and CNN models appear to be better suited for extrapolation than the Att and LSTM models. The former can achieve test accuracies of over 90% with ease, while the latter struggle to do so. Additionally, the Att and LSTM models exhibit greater variability in their performances. For instance, the Att model of capacity $\log_{10}C \approx 4.15$ with $(L, N) = (6, 2000)$ has a mean test accuracy of 84.8% over 50 simulations. However, its best accuracy is over 99% and its worst accuracy is around 50%. In Figure 5, the error bars have been omitted to better highlight the performance of the four models. Furthermore, as it is challenging to perform well on bigger sequences than those on which the models were trained, the reverse can be equally challenging. It is notably the case for the low-capacity LSTM which presents poorer generalization performances as the absolute difference between train and test sequence lengths gets bigger (Supplementary Figure S4).

Despite the impressive statistical performance of the MLP model, it fails to capture the nuances that make two RNA sequences complementary. Specifically, when trained on 500 sequences of length 5, the MLP model tends to mistake a negative example with mismatches in the upper three base pairs for a positive example. The MLP model with $\log_{10}C \approx 3.5$ achieves a statistical accuracy of 98.1%, yet it incorrectly classifies the sequence ACGUACGUGAAAACGUAGCA as a positive example with a mean positivity score of 0.9810 (Figure 6). Interestingly, all other models exhibit a similar trend, as they assign a comparable (albeit slightly lower) mean positivity score to this negative sequence as they do to its corresponding positive sequence.

Furthermore, the negative example with mismatches in the lower five base pair positions is correctly classified most of the time by all models. This suggests that the use of zero-padding to have fixed-size intputs limits the length-wise extrapolation abilities of ML models to the nucleotide positions seen during training. The MLP model seems to be particularly affected by the zero-padding, while the LSTM tends to produce lower positivity scores for the negative sequence with mismatches in the upper three base pair positions, probably due to its sequential calculation.

It is worth noting that all results presented so far were obtained by training and testing on datasets with an equal number of positive and negative examples. However, we can manipulate the proportion α of positive examples in the training dataset while still testing on a balanced dataset. Despite the loss function correction that accounts for an equal number of positive and negative examples, when the concentration of positive examples is too low or too high, all four models perform poorly (Figure 7); even when classification threshold θ is equal to α.

It is worth mentioning that in Figure 7, the test accuracy curves are skewed to the left, indicating that the models perform better with too many positive examples than too few. This behavior can likely be due to the fact that the positive examples require complete Watson-Crick complementarity. To be fair to the models, a distinction could be made between negative examples with 0% Watson-Crick pairs and 90% Watson-Crick pairs. In this line of thought, we also formulated the problem as a regression task where we wanted to predict the concentration of Watson-Crick pairs within each example. Even so, the main conclusions are not affected by such formulation of the task, so

we stick to the binary classification formulation since we are motivated by structured prediction problems that can be modeled using a collection of binary classification tasks.

If we aim to extrapolate accurately without mislabels, high-capacity models are favorable, which means that the training datasets must be sufficiently large to support this level of expressiveness. Moreover, when our training dataset contains mislabelling, the models require a substantial number of training examples before they can disregard the errors. In the next section, we will address the challenge of learning from a small training dataset while also consolidating the insights from the previous two sections on capacity.

Learning with few training examples

Here we focus on learning with small training datasets, but this at the same time provides us with an opportunity to explore how all our initial challenges interact with each other, both for neural networks and some classical ML methods. Thus, it is an important section that consolidates the insights gained from the previous sections.

We can note from last sections that learning with high capacity is advantageous for extrapolating without mislabels, while learning with low capacity is necessary to account for mislabels. However, when both challenges are combined, model behaviors become more complex.

To illustrate this complexity, we present heatmaps showing the behavior of the four tested architectures when trained with various quantities of sequences of length 6 with a mislabelling probability of 20% and a positivity ratio of 40% (Figure 8), the later being an arbitrary value to further challenge the extrapolation abilities of the models. The test accuracies are reported over the whole set of sequences of length 8 with as many positive and negative examples. See Supplementary Figure S3 for the corresponding train accuracies. These heatmaps demonstrate that the ability to ignore errors is conserved even when extrapolating, provided enough training examples are supplied. However, the required capacity to maximize generalization performance varies among the models and can be influenced by the number of training examples. For instance, the MLP and CNN require low capacity to attain their best generalization performance, while the Att and LSTM require high capacity. Additionally, the CNN's best generalization performance requires high capacity with fewer training examples, but low capacity with more training examples.

However, achieving acceptable performances with test accuracy over 80% requires a substantial number of training examples ($N \geq 2000$) when extrapolating with mislabels. When only a few training examples are available ($N \leq 500$), test accuracies over 80% become challenging to achieve. Figure 9A shows the influence of capacity on the accuracies when trained on 500 sequences of length 6 with 20% mislabelling and 40% positivity rate, tested on the whole balanced set of correctly labelled sequences of length 8. It appears that variations in capacity have a low impact on the model's performance, although a general slight improvement can be observed as capacity increases.

In order to gain more insight into the behavior of the four models when dealing with few or many training examples, we compared their distributions of performances (Figure 9B). To produce this figure, we used the capacities that maximize accuracy based on the heatmaps presented in Figure 8. As a result, a wide range of capacities was needed for the models, with the MLP using $C \approx 3{,}500$ and the CNN using $C \approx 15{,}000$ when trained on many examples. On the other hand, the Att and LSTM required much higher capacities, with the Att using $C \approx 60{,}000$ and the LSTM using $C \approx 300{,}000$ regardless of the number of training examples. Analyzing these distributions, we observe that the MLP model appears to be the most suitable for extrapolating with mislabels since its test accuracies have a higher mean and lesser variance than the other three models.

With such a fundamental binary classification task, the behaviors of some specific classical ML algorithms can put the results on neural networks into perspective. Focusing on the *k*-nearest neighbors (KNN), the support vector machine (SVM), the decision tree (tree) and the random forest (forest) algorithms, we measured the performances of such methods when learning on few examples with mislabels in a length-wise extrapolation context. All algorithms use default sklearn implementation except the KNN uses 35 neighbors, the tree uses a maximum depth of 12 decisions and the forest uses 400 classification trees. These parameters were determined by grid search maximizing the generalization performance for most training dataset sizes.

First of all, we measured the ability of these algorithms to account for mislabels in the training dataset (Figure 10A). In comparison to the four tested neural networks, the method that behaves in the most similar way is the SVM, as the test accuracy tends to 100% while the train accuracy tends to $1 - \mu$ as $N \to \infty$. The decision tree bahaves similarly, but its generalization performances are not as good. The KNN also performs poorly, but its train accuracies are stuck at 100%. The most surprising behavior is held by the random forest algorithm: Even tough it fits 100% of the training datasets, which include 20% of mislabelled examples, the test accuracies can simultaneously reach above 99%, which shows an ability to minimize both test and training risks (Belkin et al. 2019; Peters and Schuld 2022). The accuracy distributions presented in Figure 10B allow to further visualize this diversity of behaviors when algorithms are trained with 500 and 4,000 examples of length 6 before being tested on the whole set of sequences of length 8. Note that despite the relatively good performances of the SVM and random forest algorithms, they also suffer from the same misclassification problem presented in Figure 6.

Overall, this section highlights the importance of considering multiple challenges simultaneously when measuring the impact of capacity and training dataset size on the generalization performance of a variety of ML models. It also underscores the need for a better understanding of the trade-offs involved in choosing appropriate models and capacity for specific tasks, especially when dealing with small training datasets.

Conclusion

In conclusion, the use of statistical learning through neural networks holds great promise for gaining insight into the complex mechanisms that govern RNA folding. Even though over-parameterized models may be adequate for learning useful representations, the fact remains that the quality of a ML model highly depends on the data on which it has been trained. Our study highlights three main challenges researchers face when working with current RNA structure data and provides suggestions for overcoming them.

A binary classification task has been defined to measure the abilities of a variety of ML algorithms to learn Watson-Crick complementarity. The definition of the task has allowed us to generate synthetical datasets to properly test our models in light of specific challenges one can encounter when dealing with RNA structure prediction. An emphasis has been put on four types of neural networks that act as representatives of four families of commonly used neural networks in the field.

Specifically, when dealing with mislabels, low-capacity models may be preferable, as long as enough training examples are provided. Moreover, for tasks that require extrapolating to structurally dissimilar data, high-capacity models may provide better performance. With the fixed-size input involving limitations regarding length-wise generalization, we propose using models that can adapt to different RNA sequence lengths like recurrent neural networks or fully convolutional networks. Finally, we recommend exploring the behavior of a variety of neural networks on synthetic data to better understand their specific risks and benefits in predicting RNA structure. Overall, by addressing these challenges, machine learning could provide valuable insights into RNA folding and contribute to the development of new approaches for studying biological systems involving RNA.

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

SC: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. FM: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

Acknowledgments

We would like to express our sincere gratitude to Sébastien Lemieux and all the participants of the Computational Approaches to RNA Structure and Function conference held in Benasque, Spain in 2022 for their valuable suggestions and feedback.

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/fgene.2023.1254226/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Performance of LSTM and CNN models when learning with mislabels. Train (red) and test (blue) mean accuracies over 50 simulations reported for LSTM and CNN models. Sequence length and mislabelling probability are respectively fixed to $L = 8$ and $\mu = 0.2$.

SUPPLEMENTARY FIGURE S2
Performance of MLP and LSTM models in length-wise extrapolation context. Train (red) and test (blue) mean accuracies over 50 simulations reported for MLP and LSTM models. Models were trained on sequences of length 6 before being tested on sequences of length 8, without mislabelling in the training set.

SUPPLEMENTARY FIGURE S3
Training performances for all models when learning with mislabels in extrapolation context. Train accuracies for all models when trained on sequences of length 6 with 20% mislabelled training examples and 40% positivity rate.

SUPPLEMENTARY FIGURE S4
Performance for low-capacity LSTM when trained and tested on a variety of sequence lengths. Test accuracies for the LSTM model with $\log_{10}C \approx 3.5$ when trained and tested on 500 sequences of length 5 to 10 without mislabelling with a 50% positivity rate.

References

Balestriero, P., LeCunBalestriero, R., Pesenti, J., and LeCun, Y. (2021). Learning in high dimension always amounts to extrapolation. Available at: https://arxiv.org/abs/2110.09485.

Belkin, H., MandalBelkin, M., Hsu, D., Ma, S., and Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proc. Natl. Acad. Sci.* 116, 15849–15854. doi:10.1073/pnas.1903070116

Berrada, Z., KumarBerrada, L., Zisserman, A., and Kumar, M. P. (2020). Training neural networks for and by interpolation. Available at: https://arxiv.org/abs/1906.05661.

Burley, B., Bi, B., Chen, C., Bhikadiya, C., Bi, C., and Bittrich, S. (2022). Rcsb protein data bank: celebrating 50 years of the pdb with new tools for understanding and visualizing biological macromolecules in 3d. *Protein Sci.* 31, 187–208. doi:10.1002/pro.4213

Chen, L., Umarov, G., SongChen, X., Li, Y., Umarov, R., Gao, X., et al. (2020). Rna secondary structure prediction by learning unrolled algorithms. Available at: https://arxiv.org/abs/2002.05810.

Condon, D., Rastegari, Z., TarrantCondon, A., Davy, B., Rastegari, B., Zhao, S., et al. (2004). Classifying rna pseudoknotted structures. *Theor. Comput. Sci.* 320, 35–50. doi:10.1016/j.tcs.2004.03.042

Danaee, P., Rouches, M., Wiley, M., Deng, D., Huang, L., and Hendri, D. (2018). bprna: large-scale automated annotation and analysis of rna secondary structure. *Nucleic acids Res.* 46, 5381–5394. doi:10.1093/nar/gky285

Dumoulin, V., and Visin, F. (2016). A guide to convolution arithmetic for deep learning. Available at: https://arxiv.org/abs/1603.07285.

Flamm, C. H., Wielach, J., Wolfinger, M. T., Badelt, S., Lorenz, R., and Hofacker, I. L. (2021). Caveats to deep learning approaches to rna secondary structure prediction. *Front. Bioinform* 2, 835422. doi:10.3389/fbinf.2022.835422

Fu, L., Cao, Y., Wu, J., Peng, Q., Nie, Q., and Xie, X. (2022). Ufold: fast and accurate rna secondary structure prediction with deep learning. *Nucleic acids Res.* 50, e14. doi:10.1093/nar/gkab1074

Goodfellow, B., CourvilleGoodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. Cambridge: MIT press.

Hinton, S., Krizhevsky, S., Hinton, S. G. E., Srivastava, N., Krizhevsky, A., et al. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *Available at:* https://arxiv.org/abs/1207.0580.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Stadler, P. F. (1994). Fast folding and comparison of rna secondary structures. *Monatsh. fur Chem.* 125, 167–188. doi:10.1007/bf00818163

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *Int. Conf. Mach. Learn. (pmlr)* 37, 448–456. doi:10.48550/arXiv.1502.03167

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. Available at: https://arxiv.org/abs/1412.6980.

LeCun, Y., et al. (1989). Generalization and network design strategies. *Connect. perspective* 19, 18.

Marchand, W., Berkemer, B., PontyMarchand, B., Will, S., Berkemer, S., Bulteau, L., et al. (2022). "Automated design of dynamic programming schemes for RNA folding with pseudoknots," in *Wabi 2022 - 22nd Workshop on Algorithms in bioinformatics* (Potsdam, Germany: EasyChair Smart CFP).

Moore, P. B. (1999). Structural motifs in rna. *Annu. Rev. Biochem.* 68, 287–300. doi:10.1146/annurev.biochem.68.1.287

Nair, V., and Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10). Haifa, Israel, June 10, 807–814.

Peters, E., and Schuld, M. (2022). Generalization despite overfitting in quantum machine learning models. Available at: https://arxiv.org/abs/2209.05523.

Rivas, E., and Eddy, S. R. (1999). A dynamic programming algorithm for rna structure prediction including pseudoknots. *J. Mol. Biol.* 285, 2053–2068. doi:10.1006/jmbi.1998.2436

Rivas, L., Eddy. Rivas, E., Lang, R., and Eddy, S. R. (2012). A range of complex probabilistic models for rna secondary structure prediction that includes the nearest-neighbor model and more. *RNA* 18, 193–212. doi:10.1261/rna.030049.111

Sak, H., Senior, A. W., and Beaufays, F. (2014). *Long short-term memory recurrent neural network architectures for large scale acoustic modeling*, 338–342.

Sato, A., Sakakibara Sato, K., Akiyama, M., and Sakakibara, Y. (2021). Rna secondary structure prediction using deep learning with thermodynamic integration. *Nat. Commun.* 12, 941. doi:10.1038/s41467-021-21194-4

Shen, H., Peng, C., Xiong, H., Shen, T., Hu, Z., and Peng, Z. (2022). E2efold-3d: end-to-end deep learning method for accurate de novo rna 3d structure prediction. Available at: https://arxiv.org/abs/2207.01586.

Singh, J., Hanson, J., Paliwal, K., and Zhou, Y. (2019). Rna secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat. Commun.* 10, 5407. doi:10.1038/s41467-019-13395-9

Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics Comput.* 14, 199–222. doi:10.1023/b:stco.0000035301.49549.88

Szikszai, W., Datta, W., MathewsSzikszai, M., Wise, M., Datta, A., Ward, M., et al. (2022). Deep learning models for rna secondary structure prediction (probably) do not generalize across families. *Bioinformatics* 38, 3892–3899. doi:10.1093/bioinformatics/btac415

Vaswani, S., Parmar, U., Jones, G., Shazeer, N., Parmar, N., and Uszkoreit, J. (2017). Attention is all you need. *Adv. neural Inf. Process. Syst.* 30. doi:10.48550/arXiv.1706.03762

Wang, L., Zhong, L., Lu, L., Liu, Y., Zhong, X., and Liu, H. (2019). Dmfold: A novel method to predict rna secondary structure with pseudoknots based on deep learning and improved base pair maximization principle. *Front. Genet.* 10, 143. doi:10.3389/fgene.2019.00143

Zakov, S., Goldberg, Y., Elhadad, M., and Ziv-Ukelson, M. (2011). Rich parameterization improves rna structure prediction. *J. Comput. Biol.* 18, 1525–1542. doi:10.1089/cmb.2011.0184

Zhang, B., Hardt, R., VinyalsZhang, C., Bengio, S., Hardt, M., Recht, B., et al. (2021). Understanding deep learning (still) requires rethinking generalization. *Commun. ACM* 64, 107–115. doi:10.1145/3446776

Zhao, Q., Zhao, Z., Fan, X., Yuan, Z., Mao, Q., and Yao, Y. (2021). Review of machine learning methods for rna secondary structure prediction. *PLoS Comput. Biol.* 17, e1009291. doi:10.1371/journal.pcbi.1009291

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large rna sequences using thermodynamics and auxiliary information. *Nucleic acids Res.* 9, 133–148. doi:10.1093/nar/9.1.133

frontiers | Frontiers in Genetics

TYPE Mini Review
PUBLISHED 15 September 2023
DOI 10.3389/fgene.2023.1231486

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Siti Aishah Sulaiman,
National University of Malaysia, Malaysia

*CORRESPONDENCE
Shahidee Zainal Abidin,
shahidee.zainal@umt.edu.my
Adila A. Hamid,
adilahamid@ppukm.ukm.edu.my

RECEIVED 30 May 2023
ACCEPTED 25 August 2023
PUBLISHED 15 September 2023

CITATION
Abidin SZ, Mat Pauzi NA, Mansor NI, Mohd Isa NI and Hamid AA (2023), A new perspective on Alzheimer's disease: microRNAs and circular RNAs.
Front. Genet. 14:1231486.
doi: 10.3389/fgene.2023.1231486

A new perspective on Alzheimer's disease: microRNAs and circular RNAs

Shahidee Zainal Abidin[1,2]*, Nurul Asykin Mat Pauzi[1], Nur Izzati Mansor[3], Nurul Iffah Mohd Isa[4] and Adila A. Hamid[4]*

[1]Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, Terengganu, Malaysia, [2]Biological Security and Sustainability (BIOSIS) Research Interest Group, Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, Terengganu, Malaysia, [3]Department of Nursing, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia, [4]Department of Physiology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

microRNAs (miRNAs) play a multifaceted role in the pathogenesis of Alzheimer's disease (AD). miRNAs regulate several aspects of the disease, such as Aβ metabolism, tau phosphorylation, neuroinflammation, and synaptic function. The dynamic interaction between miRNAs and their target genes depends upon various factors, including the subcellular localization of miRNAs, the relative abundance of miRNAs and target mRNAs, and the affinity of miRNA-mRNA interactions. The miRNAs are released into extracellular fluids and subsequently conveyed to specific target cells through various modes of transportation, such as exosomes. In comparison, circular RNAs (circRNAs) are non-coding RNA (ncRNA) characterized by their covalently closed continuous loops. In contrast to linear RNA, RNA molecules are circularized by forming covalent bonds between the 3′and 5′ends. CircRNA regulates gene expression through interaction with miRNAs at either the transcriptional or post-transcriptional level, even though their precise functions and mechanisms of gene regulation remain to be elucidated. The current stage of research on miRNA expression profiles for diagnostic purposes in complex disorders such as Alzheimer's disease is still in its early phase, primarily due to the intricate nature of the underlying pathological causes, which encompass a diverse range of pathways and targets. Hence, this review comprehensively addressed the alteration of miRNA expression across diverse sources such as peripheral blood, exosome, cerebrospinal fluid, and brain in AD patients. This review also addresses the nascent involvement of circRNAs in the pathogenesis of AD and their prospective utility as biomarkers and therapeutic targets for these conditions in future research.

KEYWORDS

microRNA, circular RNA, Alzheimer's disease, neurodegenerative disease, biomarkers

1 Introduction

Alzheimer's disease (AD) is a common form of dementia that causes progressive mental decline in the elderly. This is due to impaired neurons and synapses, eventually changing patients' behavior, memory, language, and cognition (Lu et al., 2021). AD is characterized by the accumulation of disposition of amyloid-β (Aβ), neurofibrillary tangle (NFT) formation, and extensive neuronal degeneration (Maciotta et al., 2013). The etiology of AD remains incompletely comprehended. The early indications of AD are frequently characterized by

mild symptoms, including memory lapses and challenges in recollecting recent occurrences. As the disease advances, affected individuals may manifest symptoms such as cognitive impairment, perceptual disorientation, linguistic deficits, affective instability, and behavioral alterations.

The utilization of microRNA (miRNA) panels, which consist of multiple miRNAs, has been studied by researchers to enhance diagnostic accuracy. The panels typically consist of a group of differentially expressed miRNAs that, when combined, provide a more comprehensive disease profile. miRNA is a short non-coding RNA that plays an essential role in biological and pathological mechanisms in humans, animals, and plants (Harper et al., 2019; Toivonen et al., 2020; Beylerli et al., 2022). In recent years, many miRNAs have been studied and found to be involved with the development of neurodegenerative diseases, including AD. This makes them a potential biomarker for neurodegenerative diseases.

On the other hand, circular RNAs (CircRNAs) function as regulators of miRNA activity by sequestering them as miRNA sponges or competitive endogenous RNAs (ceRNAs) due to the presence of miRNA binding sites within circRNAs (Hansen et al., 2013a). When miRNAs form complexes with circRNAs, they undergo a process of sequestration or "sponging" from their respective target messenger RNAs (mRNAs) (Hansen et al., 2013b). Furthermore, circRNA regulates host gene expression through various mechanisms, such as acting as protein sponges, facilitating protein translocation and translation, and promoting protein-protein interaction (Zhao et al., 2022). This interaction can be especially important when the targeted miRNAs regulate biological processes or disease pathways. This review provides an overview of miRNAs and circRNAs' biosynthetic mechanisms and functional roles. Additionally, the most relevant miRNAs and circRNAs, as well as the relationship miRNA-circRNA linked to AD, are discussed. Furthermore, our attention will be directed toward how the levels of multiple miRNAs in peripheral blood, serum, exosomes derived from serum, cerebrospinal fluid, and brain are altered in individuals with AD.

2 Biogenesis of microRNA and circular RNA

Most miRNAs are known to be synthesized via a canonical pathway, which undergoes several processing steps before developing into functional miRNAs. In the canonical pathway, primary miRNAs (pri-miRNAs) are transcribed by RNA polymerase II within the nucleus, which generates distinguish hairpin structures with polyadenylated and capped. The hairpin structures within pri-miRNAs are recognized by DGCR8 (DiGeorge syndrome critical region gene 8) and cleaved DROSHA to form precursor miRNAs (pre-miRNAs). The pre-miRNAs are exported to the cytoplasm by exportin 5 (XPO5) via Ran-GTP-dependent mechanisms (Komatsu et al., 2023).

The pre-miRNA is further cleaved off by dicer and its catalytic partner trans-activator RNA-binding protein (TRBP) into ~22 nt miRNA duplexes. The miRNA duplex is loaded into Argonaute (AGO) protein before unwinding into two strands, guide, and passenger strand. Generally, the RNA-induced silencing complex's component 3 promoter (C3PO) (RISC) will degrade the passenger strand. At the same time, the guide strand will be incorporated with AGO2 to initiate the RNA silencing mechanism. In some cases, both strands can be incorporated into the RISC complex. The RISC complex then recognizes target mRNAs based on sequence complementary between the miRNA and the mRNA 3′untranslated region (UTR). This recognition leads to mRNA degradation or translation inhibition, subsequently reducing protein expression.

CircRNAs exist as a continuous loop structure. circRNAs are ubiquitously expressed in mammalian cells and can also be found in a tissue-specific manner (Han et al., 2018). In contrast to linear mRNA, circRNA is generated through back-splicing, RNA-binding protein (RBP)-mediated circularisation, and exon skipping driven by lariat or intron-pairing mechanisms from pre-messenger RNA (Sekar and Liang, 2019; Guria et al., 2020). This process results in a circular transcript structure that arises from the fusion of the 3′non-co-linearly splice site with either its upstream 5′or another upstream exon during splicing events (Zhou et al., 2020).

The circRNA molecules that contain exonic or exon-intronic sequences are formed through a process known as back-splicing (Patop et al., 2019). CircRNAs can also be produced using lariat-driven circularisation forming a circular exonic RNA with a large lariat of introns (Guria et al., 2020). In addition, the RBP can bind specific motifs in the flanking introns to promote circularisation (Patop et al., 2019). Several splicing factors, including QK1, ADAR (adenine deaminase acting on RNA), ESRP1, and FUS, have been shown to facilitate circRNA production (Conn et al., 2015; Ivanov et al., 2015). Then circRNAs are transported into the cytoplasm in a size-dependent manner through either ATP-dependent RNA helicase DDX39A (also referred to as nuclear RNA helicase URH49) or spliceosome RNA helicase DDX39B (also known as DEAD box protein UAP56) (Huang et al., 2018).

3 microRNA and Alzheimer's disease

Many studies have demonstrated that dysregulation of miRNAs was associated with the pathogenesis of AD. One of the studies identified a set of miRNAs differentially expressed in the brain, cerebrospinal fluid (CSF), and blood of AD patients compared to healthy controls (Takousis et al., 2019). Although miRNAs are predominantly intracellular, a sizeable fraction of them are migratory and can be identified in extracellular fluids (Zen and Zhang, 2012). These miRNAs are known as circulating miRNAs and can be detected in bodily fluids such as blood, urine, saliva, seminal fluid, breast milk, microvesicles, and exosomes (Weber et al., 2010; Vickers et al., 2011; O'Driscoll, 2015). This review presents a comprehensive summary of miRNA expression in three distinct compartments, blood, CSF, and brain tissues, in relation to AD (Table 1).

3.1 Blood sample

miRNAs exhibit considerable stability in blood and can be conveniently obtained through minimally invasive procedures. Studies have revealed distinct miRNAs with altered expression profiles in the bloodstreams of AD patients compared to those in

TABLE 1 The expression of miRNAs in peripheral blood, cerebrospinal fluids (CSF) and brain tissues derived from Alzheimer's patients.

miRNA	Expression pattern	Experimental approach	Patient demographic	Reference
Whole Blood				
miR-112, *miR-161*, *let-7d*-3p, *miR-5010*-3p, *miR-151a*-3p	Upregulated	Small RNA-Sequencing validated by RT-qPCR	AD patient • Age: 72.7 ± 10.4 • MMSE Score: 18.9 ± 3.4	Leidinger et al. (2013)
miR-103a-3p, *miR-107*, *miR-532*–5p, *miR-26b*-5p, *let-7f*-5p	Downregulated		MCI patient • Age: 73.9 ± 6.2 • MMSE Score: 25.3 ± 1.4	
miR-9-5p, *miR-106a*-5p, *miR-106b*-5p, *miR-107*	Downregulated	RT-qPCR	Late-onset AD (>60 years)	Yilmaz et al. (2016)
Blood (Platelet)				
let-7i-5p, *miR-125a*-5p, *miR-1233*-5p	Downregulated	miRNA microarray validated by RT-qPCR	Aβ(−) MCI • Age: 74.2 ± 5.67 • MMSE Score: 24.5 ± 3.03 Aβ(+) MCI • Age: 74.6 ± 2.48 • MMSE Score: 24.0 ± 0.53	Lee et al. (2020)
Serum				
miR-98-5p, *miR-885*-5p, *miR-483*-3p, *miR-342*-3p, *miR-191*-5p, *let-7d*-5p	Downregulated	miRNA sequencing validated by RT-qPCR	Age: 77.83 ± 7.40 MMSE Score: 10.45 ± 2.21	Tan et al. (2014)
miR-210	Downregulated	RT-qPCR	AD patient (Age: 60—84 years) MCI patient (Age: 61—82 years)	Zhu et al. (2015)
miR-519	Upregulated	RT-qPCR	Age: 81.36 ± 13.25	Jia and Liu (2016)
miR-29, miR-125b, miR-223	Downregulated		MMSE Scores: 18.3 ± 5.2	
miR-613	Upregulated	RT-qPCR	MCI patient (Age: 64.8 ± 7.2) DAT patient (Age: 65.5 ± 6.8)	Li et al. (2016)
miR-455-3p, *miR-4668*-5p	Upregulated	Microarray validated by RT-qPCR	AD patient • Age: 75.3 ± 10.14 • MMSE Score: 22.4 ± 4.6 MCI patient • Age: 72.4 ± 7.53 • MMSE Score: 25 ± 2.05	Kumar et al. (2017)
miR-195-5p, *miR-146a*-5p, *miR-106b*-3p, *miR-20b*-5p, *miR-497*-5p	Upregulated	Small RNA sequencing validated RT-qPCR	Age: 72.1 ± 8.5	Wu et al. (2017)
miR-93-5p, *miR-29c*-3p, *miR-125b*-3p, *miR-19b*-3p	Downregulated		MMSE Scores: 28.9 ± 2.9	
miR-28-3p	Upregulated	RT-qPCR	AD patient • Age: 70.12 ± 2.09 • MMSE Score: 15.48 ± 1.68 MCI patient	Zhao et al. (2020)

(Continued on following page)

TABLE 1 (*Continued*) The expression of miRNAs in peripheral blood, cerebrospinal fluids (CSF) and brain tissues derived from Alzheimer's patients.

miRNA	Expression pattern	Experimental approach	Patient demographic	Reference
			• Age: 69.68 ± 2.11	
			• MMSE Score: 22.67 ± 0.73	
***miR-331*-3p**	Downregulated	RT-qPCR	Age: 72.69 ± 6.64	Liu and Lei (2021)
			MMSE Scores: 16.05 ± 2.69	
miR-6761-3p, *miR-6747*-3p, *miR-6875*-3p, *miR-6754*-3p, *miR-6736*-3p, *miR-6762*-3p, *miR-6787*-3p, *miR-208a*-5p, *miR-6740*-3p, *miR-6778*-3p, *miR-6753*-3p, *miR-6716*-3p, *miR-4747*-3p, *miR-3646*, *miR-595*, *miR-4435*	Upregulated	Data were obtained from GEO repository (GSE120584)	Patients are composed of MCI, VaD, and DLB with ages >60 years	Lu et al., 2021, Shigemizu et al., 2019
miR-125a-3p, ***miR-22*-3p**, *miR-24*-3p, *miR-6131*, *miR-125b-1*-3p	Downregulated			
miR-128	Upregulated	RT-qPCR	AD patient	Zhang et al. (2021)
			• Age: 72.57 ± 8.13	
			• MMSE Score: 16.88 ± 2.04	
Plasma				
let-7d-5p, *let-7g*-5p, *miR-15b*-5p, *miR-142*-3p, *miR-191*-5p, *miR-545*-3p, *miR-301a-3p*	Downregulated	Nanostring platform validated by RT-qPCR	Cohort 1 (Composed of AD and MCI patients)	Kumar et al. (2013)
			• Age: 78.65 ± 6	
			• MMSE Score: 20.95 ± 4.4	
			Cohort 2 (Only AD patient)	
			• Age: 69.3 ± 6.18	
			• MMSE Score: 16.5 ± 3.97	
miR-34c	Upregulated	RT-qPCR	Age: 56–90	Bhatnagar et al. (2014)
			Sporadic late-onset AD	
miR-34a, *miR-146a*	Downregulated	RT-qPCR	Age: 80.7 ± 5.8	Kiko et al. (2014)
			MMSE Score: 21.1 ± 3.5	
			CSF-Aβ (1–42) (pg/mL): 378.4 ± 112.2	
			CSF-tau (pg/mL): 480.7 ± 218.9	
			Consist of cerebral atrophy no cerebrovascular lesion	
miR-384	Downregulated	RT-qPCR	MCI patient (Age: 63.2 ± 6.1)	Liu et al. (2014)
			DAT patient (Age: 64.2 ± 5.8)	
miR-483-5p, *miR-486*-5p, *miR-200a*-3p, *miR-502*-3p	Upregulated	miRCURY LNA miRNA PCR panel	AD patient	Nagaraj et al. (2017)
			• Age: 67 ± 8	
			• MMSE Score: 20.4 ± 4	
			• Tau (pg/mL): 774.3 ± 347	
			• pTau (pg/mL): 94.8 ± 38	
			• $A\beta_{1-42}$ (pg/mL): 387.4 ± 163	
miR-30b-5p, *miR-142*-3p	Downregulated		MCI patient	
			• Age: 62 ± 6	
			• MMSE Score: 25.9 ± 2	

(Continued on following page)

TABLE 1 (*Continued*) The expression of miRNAs in peripheral blood, cerebrospinal fluids (CSF) and brain tissues derived from Alzheimer's patients.

miRNA	Expression pattern	Experimental approach	Patient demographic	Reference
			• Tau (pg/mL): 444 ± 243	
			• pTau (pg/mL): 67.3 ± 30	
			• $A\beta_{1-42}$ (pg/mL): 631.9 ± 351	
miR-1908	Upregulated	RT-qPCR	Age: 71.3 ± 2.5	Wang et al. (2018)
			MMSE score: 16.8 ± 3.5	
			CDR score: 1.5 ± 0.3	
miR-103, *miR-107*	Downregulated	RT-qPCR	Age: 72.5 ± 7.7	Wang et al. (2020)
			MMSE Score: 16.8 ± 3.0	
			Mild dementia	
Serum (Exosome)				
miR-135a, ***miR-384***	Upregulated	RT-qPCR	MCI (Age: 61.63 ± 7.32)	Yang et al. (2018a)
miR-193b	Downregulated		DAT (Age: 74.15 ± 7.93)	
***miR-22*-3p**, *miR-378a*-3p	Upregulated	miRNA sequencing validated by RT-qPCR	Age: 75.1 ± 7.5	Dong et al. (2021)
miR-30b-5p	Downregulated			
miR-342-5p	Downregulated	RT-qPCR	Age: 75.1 ± 7.5	Dong et al. (2022)
			MMSE score: 6.55 ± 2.86	
Plasma (Exosome)				
miR-342-3p, *miR-141*-3p, *miR-342*-5p, *miR-23b*-3p, ***miR-125b*-5p**, *miR-24*–3p, *miR-152*–3p	Downregulated	Small RNA deep sequencing	Age: Between 50—75	Lugli et al. (2015)
miR-423-5p, *miR-369*-5p, *miR-23a*-5p	Upregulated	Small RNA sequencing	Age: 67.8 ± 5.72	Nie et al. (2020)
miR-204-5p, *miR-125a*-5p, *miR-1468*-5p, *miR-375*, *let-7e*-5p	Downregulated		MMSE Score: 13.8 ± 7.43	
Cerebrospinal fluids				
let-7b	Upregulated	RT-qPCR	Age: 67.84 ± 6.6	Lehmann et al. (2012)
			MMSE Score: 22.54 ± 3.13 t-Tau (pg/mL): 685.77 ± 282.23	
			$A\beta_{1-42}$ (pg/mL): 408.38 ± 89.34	
miR-27a-3p	Downregulated	RT-qPCR	Cohort 1	Frigerio et al. (2013)
			• Age: 66.6 ± 6.9	
			• MMSE Score: 24.4 ± 2.2	
			• Tau (pg/mL): 677 ± 309.6	
			• pTau (pg/mL): 93.6 ± 36.3	
			• $A\beta_{1-42}$ (pg/mL): 334 ± 131.8	
			Cohort 2	
			• Age: 75.8 ± 8.6	
			• Tau (pg/mL): 864.7 ± 298	
			• pTau (pg/mL): 122.1 ± 26	
			• $A\beta_{1-42}$ (pg/mL): 812.2 ± 213.5	

(Continued on following page)

TABLE 1 (*Continued*) The expression of miRNAs in peripheral blood, cerebrospinal fluids (CSF) and brain tissues derived from Alzheimer's patients.

miRNA	Expression pattern	Experimental approach	Patient demographic	Reference
miR-29a, *miR-29b*	Upregulated	RT-qPCR	Age: 80.7 ± 5.8	Kiko et al. (2014)
			MMSE Score: 21.1 ± 3.5	
			CSF-Aβ (1–42) (pg/mL): 378.4 ± 112.2	
miR-34a, *miR-125b*, ***miR-146a***	Downregulated		CSF-tau (pg/mL): 480.7 ± 218.9	
			Consist of cerebral atrophy	
miR-384	Downregulated	RT-qPCR	MCI patient (Age: 63.2 ± 6.1)	Liu et al. (2014)
			DAT patient (Age: 64.2 ± 5.8)	
miR-146a, *miR-100, miR-505#, miR-4467, miR-766, miR-3622b*-3p, *miR-296*	Upregulated	TaqMan OpenArray Human MicroRNA	Age: 72.1 ± 8.5	Denk et al. (2015)
miR-1274a, *miR-375, miR-708, , miR-219, miR-103*	Downregulated		Total tau (pg/mL): 708.5 ± 282.9 pTau (pg/mL): 92 ± 93.3 Aβ1-42 (pg/mL): 446.7 ± 164.1	
miR-210	Downregulated	RT-qPCR	AD patient (Age: 60—84 years)	Zhu et al. (2015)
			MCI patient (Age: 61—82 years)	
			MCI patient (Age: 64.8 ± 7.2)	
			DAT patient (Age: 65.5 ± 6.8)	
miR-613	Upregulated	RT-qPCR		Li et al. (2016)
miR-29a	Upregulated	RT-qPCR	Age: 70.4 ± 9.1	Müller et al. (2016)
			MMSE Score: 19.7 ± 3.2	
miR-378a-3p*, *miR-1291*	Upregulated	TaqMan microRNA array	Age: 69.62 ± 7.27	Lusardi et al. (2017)
miR-143-3p, *miR-142*-3p, *miR-328*-3p, *miR-193a*-5p, *miR-19b*-3p, *miR-30d*-5p, *miR-340*-5p, *miR-140*-5p, ***miR-125b*-5p**, *miR-223*-3p	Downregulated		MMSE Score: 18.28 ± 6.40	
let-7b, *let-7e*	Upregulated	RT-qPCR	AD patient	Derkow et al. (2018)
			• Age: 71.5 ± 8.5	
			• MMSE Score: 21.9 ± 4.6	
			• Tau (pg/mL): 615.6 ± 322.7	
			• $A\beta_{1\text{-}42}$ (pg/mL): 523.8 ± 94.9	
			FTD patient	
			• Age: 64 ± 11.5	
			• MMSE Score: 22.9 ± 4.9	
			• Tau (pg/mL): 430.6 ± 195.5	
			• $A\beta_{1\text{-}42}$ (pg/mL): 1,227 ± 524.4	
***miR-125*-5p**	Upregulated	miRNA array validated by RT-qPCR	YOAD patient	McKeever et al. (2018)
			• Age: 60.9 ± 4.6	
			• Memory immediate recall: 8.9 ± 4.6	
			• Delayed recall: 2.8 ± 4.4	
			• Delayed recognition: 15.3 ± 3.7	
			• Visuospatial: 15.7 ± 10.7	
			• $A\beta_{1\text{-}42}$ (pg/mL): 356.0 ± 159.1	
miR-451a, *miR-605*-5p	Downregulated		• Total tau (pg/mL): 744.5 ± 375.0	

(Continued on following page)

TABLE 1 (*Continued*) The expression of miRNAs in peripheral blood, cerebrospinal fluids (CSF) and brain tissues derived from Alzheimer's patients.

miRNA	Expression pattern	Experimental approach	Patient demographic	Reference
			• pTau (pg/mL): 101.7 ± 37.9	
			LOAD patient	
			• Age: 75.5 ± 4.6	
			• $A\beta_{1-42}$ (pg/mL): 431.3 ± 139.4	
			• Tau (pg/mL): 721.6 ± 245.1	
			• pTau (pg/mL): 97.1 ± 19.7	
miR-455-3p	Upregulated	RT-qPCR	Age: 79.09 ± 6.17	Kumar and Reddy (2021b)
miR-16-5p, ***miR-331***-3p, *miR-409*-3p, *miR-454-3p*	Upregulated	miRNA array validated by RT-qPCR	AD e−3,3	Sandau et al. (2022)
			• Age: 72.9 ± 10.0	
			• MMSE score: 21.8 ± 3.6	
			AD e−3,4	
			• Age: 73.3 ± 5.1	
			• MMSE score: 21.2 ± 2.2	
Brain (Hippocampus)				
miR-140	Upregulated	RT-qPCR	• Age: 82 ± 7	Akhter et al. (2018)
			• Plague score: Sparse—Frequent	
			Braak stage: IV-VI	
miR-142-5p, *miR-146*-5p, *miR-155*-5p, *miR-455*-5p	Upregulated	RT-qPCR	Hippocampal tissue samples were obtained from the London Neurodegenerative Diseases Brain Bank	Sierksma et al. (2018)
miR-143-3p	Downregulated	RT-qPCR	Age: 79.33 ± 12.80	Wang et al. (2022b)
			CERAD: C	
Brain (Entorhinal Cortex)				
miR-101-3p	Upregulated	RT-qPCR	Braak stage I - II	Kikuchi et al. (2020)
			• Age: 78.5 ± 2.4	
			Braak stage III - IV	
			• Age: 82.6 ± 4.7	
			Braak stage V - VI	
			• Age: 82.9 ± 4.7	
Brain (Frontal Cortex)				
miR-212, miR-23a	Downregulated	miRNA microarray validated by RT-qPCR	AD patient	Weinberg et al. (2015)
			• Age: 88.6 ± 7.0	
			• MMSE Score: 17.5 ± 8.1	
			• Global cognitive score: −1.3 ± 0.8	
			MCI patient	
			• Age: 82.9 ± 4.9	
			• MMSE Score: 28.0 ± 1.3	
			• Global cognitive score: −0.03 ± 0.4	

(Continued on following page)

TABLE 1 (*Continued*) The expression of miRNAs in peripheral blood, cerebrospinal fluids (CSF) and brain tissues derived from Alzheimer's patients.

miRNA	Expression pattern	Experimental approach	Patient demographic	Reference
miR-346	Downregulated	RT-qPCR	Age: 80.8 ± 1.7	Long et al. (2019)
Brain (Broadman's Area 10)				
miR-455-3p	Upregulated	RT-qPCR	Age: 79.85 ± 8.29	Kumar and Reddy (2018)

Data were shown as Mean ± SD; MMSE, Mini-Mental State Examination; MCI, mild cognitive impairment; VaD, vascular dementia; DLB, dementia with lewy bodies; CDR, clinical dementia rating; DAT, dementia of the alzheimer type; FTD, frontotemporal disorder; YOAD, Young On-Set Alzheimer's Disease; LOAD, Late On-Set Alzheimer's Disease; CERAD, The Consortium to Establish a Registry for Alzheimer's Disease. Contradictory results were seen for the miRNA (bold).

good health. The miRNAs that exhibit differential expression have the potential to be used as biomarkers in the diagnosis and monitoring of progression in AD. Previous studies have shown differences in the expression profile of miRNAs between various blood components, including whole blood, serum, plasma, and platelets (Table 1). However, only a few miRNAs demonstrate similar expression patterns. Out of the miRNAs that were assessed for their expression patterns, only miR-107 and miR-125a-5p were found to have similar expression profiles in both whole blood and plasma (Leidinger et al., 2013; Yilmaz et al., 2016; Lee et al., 2020; Nie et al., 2020; Wang et al., 2020; Lu et al., 2021). In contrast, there were seven miRNAs (let-7d-5p, miR-384, miR-191-5p, miR-24-3p, miR-30b-5p, miR-342-3p, and miR-342-5p) that showed similar expression profiles between serum and plasma. (Kumar et al., 2013; Liu et al., 2014; Tan et al., 2014; Lugli et al., 2015; Nagaraj et al., 2017; Yang et al., 2018; Dong et al., 2022; 2021; Lu et al., 2021).

The miR-107 was found to be downregulated in plasma, blood, and platelet. This miRNA was found to be involved in the initiation of blood coagulation. In addition, miR-107, together with other miRNAs such as miR-96, miR-200b, miR-485, miR-107, and miR-223, play a crucial role in platelet, specifically in reactivity, aggregation, secretion, and adhesion (Gatsiou et al., 2012). A previous study has indicated that the levels miR-107 were reduced in the neocortex of individuals with AD compared to the control group (Nelson and Wang, 2010). Furthermore, it has been discovered that miR-107 has targeted 3′UTR of β-site amyloid protein-cleaving enzyme 1 (BACE1) that participates in Aβ production (Wang et al., 2008). In the progression of AD, there was a tendency for the levels of BACE1 mRNA to rise concomitantly with the decline in miR-107 levels.

Besides miR-107, miR-125a-5p has been implicated in regulating von Willebrand factor (VWF), a protein involved in platelet adhesion and aggregation (Bhatlekar et al., 2020). In a study of patients with acute myocardial infarction, decreased levels of miR-125a-5p were associated with increased VWF expression and platelet activation, suggesting that miR-125a-5p may have a role in regulating platelet function and blood clotting. While this study did not specifically investigate the role of miR-125a-5p in AD, the involvement of this miRNA in regulating blood clotting factors suggests they may play a role in AD pathogenesis. Growing evidence indicates that blood clotting factors play a role in the development and progression of AD (Paul et al., 2007; Zamolodchikov et al., 2015). One of the most widely studied clotting factors in relation to AD is fibrinogen, which is involved in blood clotting and inflammation.

In AD, fibrinogen is bound with Aβ in the brain tissue and blood vessel, thus leading to the fibrillization of Aβ and the formation of fibrin clots resistant to breaking down. Reducing fibrinogen levels lowers cerebral amyloid angiopathy (CAA) and blood-brain barrier (BBB) permeability, reduces microglial activation, and improves cognitive performance in AD mouse models (Cortes-Canteli et al., 2012). Also, a prothrombotic state in AD is shown by more clots forming, less fibrinolysis, and higher amounts of coagulation factors and activated platelets. Abnormal fibrin deposition and persistence in AD may be caused by alterations in blood clotting and Aβ-fibrinogen binding. This could lead to Aβ deposition, reduced cerebral blood flow, worsened neuroinflammation, and eventually neurodegeneration (Cortes-Canteli et al., 2012).

Furthermore, the expression of let-7d-5p, miR-384, miR-191-5p, miR-24-3p, miR-30b-5p, miR-342-3p, and miR-342-5p were downregulated in serum and plasma. The expression of miR-191-5p was reduced in microglia and hippocampal tissues of APP/PS1 mice stimulated with Aβ1-42 (Wan et al., 2021). In addition, the overexpression of miR-191-5p was observed to enhance cell viability and suppress the apoptosis rate in microglia treated with Aβ1-42. The inhibition of Aβ1-42-induced microglial cell injury was due to the inactivation of the MAPK signaling pathway, in which Map3k12 was targeted by miR-191-5p. Moreover, miR-191–5p reduced tau phosphorylation and enhanced neurite outgrowth *in vitro* (Wang et al., 2022a). This study also found that miR-191-5p reduced the levels of phosphorylated amyloid precursor protein (APP) and the generation of Aβ. The negative effects of miR-191-5p on tau phosphorylation, Aβ secretion, and neuronal cell death were due to its direct targeting of DAPK1.

Moreover, the miR-22-3p was also found to have different expressions pattern between serum and serum exosomes (Dong et al., 2021; Lu et al., 2021). Exosomes can be considered a stable source of miRNA since they can prevent RNase degradation and recover the miRNAs (Koga et al., 2011; Thind and Wilson, 2016). This condition can be one of the reasons why the expression pattern was different between serum and serum exosome.

3.2 Cerebrospinal fluid (CSF) sample

Besides blood, miRNA profiling also has been performed in AD's CSF. CSF encompasses the central nervous system (CNS) and is a reliable indicator of the biochemical alterations in this region. Numerous studies have examined miRNA expression in the CSF of individuals with AD, and certain miRNAs exhibit differential

expression in AD compared to those deemed healthy controls. Among these miRNAs, only let-7b, miR-125b, miR-146a, and miR-29a were reported in several studies that have deregulation in Alzheimer's patients (Lehmann et al., 2012; Kiko et al., 2014; Denk et al., 2015; Müller et al., 2016; Lusardi et al., 2017; Derkow et al., 2018; McKeever et al., 2018). However, two studies revealed downregulated miR-125b (Kiko et al., 2014; Lusardi et al., 2017), whereas McKeever et al. found this miRNA increased in AD (McKeever et al., 2018). Additionally, Kiko et al. found that miR-146a was downregulated (Kiko et al., 2014), but Denk et al. found that miR-146a was increased in AD (Denk et al., 2015). The variability in outcomes across studies can be attributed to various unidentified factors.

The study revealed a significant rise in let-7b levels correlated with AD progression. The escalation of let-7b was predominantly attributed to $CD4^+$ T cells in the CSF. Nevertheless, the precise role of let-7b in the onset and course of AD has yet to be entirely understood because the involvement of this miRNA in AD is still in the preliminary phases. Based on Pearson correlation coefficients (R) analysis, it was revealed that there was a significant association between the levels of let-7b in the CSF and total tau (*t*-Tau) in the subjective memory complaints (SMC), mild cognitive impairment (MCI), and AD subjects. A significant correlation was also observed between the levels of let-7b and phosphorylated tau (p-Tau) in both MCI and AD subjects. Hence, it can be inferred that let-7b correlates with *t*-Tau and p-Tau, namely, in individuals diagnosed with MCI and AD (Liu et al., 2018).

The overexpression of miR-125b suppressed cellular proliferation, induced apoptotic responses, and enhanced inflammatory and oxidative stress in mouse neuroblastoma Neuro2a APPswe/Δ9 cells. In addition, the upregulation of miR-125b resulted in a significant increase in the expression of APP and β-secretase 1 (BACE1) and the production of Aβ peptide (Jin et al., 2018). Furthermore, overexpression of miR-125b in primary neurons results in the hyperphosphorylation of tau and an increase in the expression of p35, cdk5, and p44/42-MAPK signaling. This miRNA was also found to directly target and downregulate the phosphatases DUSP6 and PPP1CA, subsequently leading to an increase in tau hyperphosphorylation (Banzhaf-Strathmann et al., 2014). On the other hand, another study has reported conflicting results. It was found that the decrease in levels of miR-34a-5p and miR-125b-5p led to an increase in the expression of BACE1 in AD patients and cell cultures (MCN and N2a cells) that were treated with Aβ (Li et al., 2020).

3.3 Brain samples

Although obtaining brain tissue samples for miRNA analysis is challenging due to the invasive nature of the procedure, some studies have investigated miRNA expression in post-mortem brain tissue from AD patients. These studies have revealed altered expression of specific miRNAs in different brain regions affected by AD pathology, including but not limited to the hippocampus and frontal cortex. It is important to note that miRNA expression profiles may vary depending on the stage and severity of AD and the brain regions examined. Additionally, the findings from different studies can sometimes be inconsistent, highlighting the need for further study to validate and establish robust miRNA biomarkers for AD.

In AD patients' brains, five miRNAs (miR-140, miR-142-5p, miR-146-5p, miR-155-5p, and miR-455-5p) were upregulated, and only miR-143-3p was downregulated in the hippocampus (Akhter et al., 2018; Sierksma et al., 2018; Wang et al., 2022b). Other studies demonstrated that miR-101-3p and miR-455-3p were significantly higher in the entorhinal cortex and Broadman's area 10 of Alzheimer's patients (Kumar and Reddy, 2018; Kikuchi et al., 2020). While miR-346 was reduced in the frontal cortex of AD (Long et al., 2019).

In APP/PS1 AD mouse model, the expression of miR-455–3p was upregulated, and its targeted gene, cytoplasmic polyadenylation element-binding 1 (CPED1), was downregulated in the hippocampus of the mouse at the age of 9 months (Xiao et al., 2021). Inhibition of CPED1 by miR-455-5p caused suppression of α-Amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) receptor expressions, subsequently mediated cognitive deficits. Nevertheless, Kumar et al.'s study showed conflicting results with these findings. Kumar and colleagues have generated a transgenic miR-455-5p (miR-455-3p TG) mouse model and demonstrated that the lifespan of this mouse was 5 months longer than the wild-type mice (Kumar et al., 2021a). However, the knockout (KO) mice had a lifespan that was 4 months shorter than the WT mice. Based on behavior studies, miR-455-3p TG mice exhibited enhanced cognitive behavior, spatial learning, and memory compared to age-matched WT mice and miR-455-3p KO mice.

4 Circular RNA and microRNA in Alzheimer's disease

A new class of non-coding RNAs known as circRNAs has been identified. The circRNAs are produced by a non-canonical splicing event called back-splicing (Kristensen et al., 2019). This RNA was initially identified in 1976 as a viroid consisting of a single-stranded circular RNA molecule isolated from an infected tomato plant (Sanger et al., 1976). This RNA molecule exhibits a structure that is covalently closed-like structure, with a high degree of self-complementarity and base-pairing. The circRNAs have been identified as miRNA sponges due to having highly abundant miRNA binding sites (Hansen et al., 2013b). CircRNAs play a crucial role in developing and maintaining brain homeostasis in the brain. Several studies have shown that a healthy mammalian brain has the highest circRNA expression level and varies across different brain regions (Rybak-Wolf et al., 2015; You et al., 2015; Li et al., 2017).

The circRNAs were highly abundant in the cerebellum, followed by the prefrontal cortex and hippocampus (Rybak-Wolf et al., 2015). Interestingly, circRNAs expression was higher in neurons than in astrocytes in the cerebral cortex and cerebellum (Gokool et al., 2020). In neurons, they are highly concentrated in the synapses of the hippocampus, indicating their potential involvement in synaptic plasticity and cognitive processes (You et al., 2015). Many circRNAs have been shown to interact with disease-associated miRNAs, suggesting that circRNAs could play major roles in the development of diseases and as a prospective prognostic biological marker. For instance, a study by Zhao and co-workers

showed that ubiquitin-protein ligase A (UBE2A) expression in AD is affected by the sponging activity of CDR1-as and miR-7 (Zhao et al., 2016). The miR-138-circHDAC9 complex is an additional mechanism that regulates the metabolism of Aβ in AD. The level of circHDAC9 was reduced in the serum of individuals with AD, which may lead to upregulation of miR-138 expression and downregulation of Sirt1 and ADAM10. Consequently, the processing of amyloid APP was redirected from the β-secretase pathway to the β-secretase pathway, resulting in an elevation of amyloid accumulation (Lu et al., 2019).

ciRS-7, the most characterized circRNA, comprises more than 70 conserved binding sites for miR-7 (Hansen et al., 2013b). The ciRS-7 is a robust and stable expression in various tissue, particularly in neural tissues (Ma et al., 2020). The downregulation of ciRS-7 has negatively correlated with miR-7 (Kristensen et al., 2019). The upregulation of miR-7 results in the downregulation of targets associated with AD, namely, UBE2A, which hinders the degradation of APP and β-secretase in the brain (Zhao et al., 2016). Frontotemporal lobar degeneration (FTLD-TAU) patients with 53 MAPT gene mutations are strongly associated with AD (Ghetti et al., 2015). Back-splicing exon 12 to 7 (tau circ 12→7) or 12 to 10 (tau circ 12 → 10), the MAPT gene produces two circRNAs in the brain, contributing to AD pathology. Only tau circ 12 → 7 has a start codon, but 12 → 10 circRNA may require ADAR activity to start translation (Welden et al., 2022). Despite the absence of stop codons in both circRNAs, protein translation can still occur at various locations in circular patterns, producing abnormal proteins. These circRNAs proteins tend to self-aggregate and form neurofibrillary tangles, possibly contributing to FTLD progression (Welden et al., 2022).

Another circRNA involved in tau phosphorylation is circPCCA. The circPCCA might competitively bind to miR-138-5p, inhibiting miR-138-5p′s ability to induce glycogen synthase kinase-3β activation and facilitate tau phosphorylation. These findings indicate that high circPCCA expression can potentially reduce disease severity in AD (Wang et al., 2015; Li et al., 2020).

Based on microarray analysis of CSF derived from AD patients, there was an upregulation of 112 circRNAs and a downregulation of 51 circRNAs when compared to the control subjects (Li et al., 2020). These circRNAs were enriched in pathways associated with AD, such as the neurotrophin signaling pathway, natural killer cell-mediated cytotoxicity, and cholinergic synapse. Further validation using RT-qPCR has demonstrated an increase in the expression of circLPAR1, circAXL, and circGPHN, while circPCCA, circHAUS4, circKIF18B, and circTTC39C were found to be decreased in AD patients.

5 Remarks, challenges, and future direction

Despite the intense research on miRNAs expression profiles for diagnosing AD, they are still in their early stage of development. One of the reasons is the complexity of the disorders to understand their pathological cause, which involves many different pathways and targets. Certain miRNAs displayed a distinct expression pattern specific to particular tissues or developmental stages. These miRNAs play a significant role in preserving tissue identity and function by contributing to various biological processes. Several miRNAs have been identified as brain-specific miRNAs, including miR-143, miR-125a/b, miR-138, miR-708, and miR-9 (Guo et al., 2014). Interestingly, these miRNAs have been identified in various other biological compartments, including blood, serum, plasma, and CSF (Table 1), which suggests that these miRNAs have the potential to be valuable candidates for biomarkers in AD in the future.

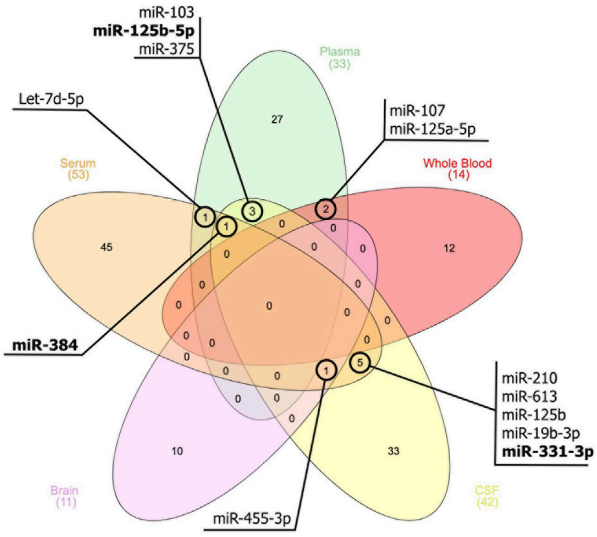


FIGURE 1
Venn diagram illustrates the miRNAs that exhibit similar expression patterns across peripheral blood, cerebrospinal fluids (CSF), and brain tissue. The miRNAs that are bold indicate a different expression pattern between the samples.

Moreover, based on the thorough assessment of the miRNAs expression profile (Figure 1), this present study identified several miRNAs also hold promise as a potential tool for AD diagnosis, monitoring disease progression, and gaining insights into the underlying molecular mechanisms of the disease. These miRNAs (Figure 1) were found to be associated with genes such as MAPK, BACE1, PTGS2, STAT3, SNAP25, and BDNF. This association subsequently contributes to the development of AD. For instance, the upregulation of miR-125 in AD promotes Tau hyperphosphorylation by regulating MAPK kinases (Banzhaf-Strathmann et al., 2014). This activation is most likely achieved by down-regulating the expression of phosphatase genes that miR-125 targets, which are DUSP6, Bcl-W, and PPP1CA (Banzhaf-Strathmann et al., 2014). BDNF, an additional gene implicated in AD, is linked to cognitive decline, particularly in immediate memory. The expression of miR-613 was found to have a negative regulatory effect on the expression of BDNF (Li et al., 2016). This relationship was observed in the hippocampus of APP/PS1 mice, where high levels of miR-613 coincided with low levels of BDNF expression (Li et al., 2016). In addition, the upregulation of miR-210-5p was observed to cause a reduction in the number of synapses in primary hippocampal neurons (Ren et al., 2018). Conversely, inhibition of miR-210-5p led to an increase in synaptic formation. This condition can be attributed to the downregulation of SNAP25, which is negatively inhibited by miR-210-5p (Ren et al., 2018).

There was a positive correlation between the expression of serum miR-331-3p and MMSE scores in patients with AD (Liu and Lei, 2021). The SH-SY5Y cell viability was significantly increased by overexpressing miR-331-3p while inhibiting miR-331-3p reduced cell viability. Meanwhile, the levels of proinflammatory cytokines, such as IL-1β, IL-6, and TNF-α, were decreased when miR-331-3p was overexpressed. Conversely, when miR-331-3p was knocked down, these cytokine levels increased. The treatment of Aβ resulted in a significant decrease in the expression of miR-331-3p. This finding aligns with the observed reduction of miR-331-3p in serum samples from individuals with AD (Liu and Lei, 2021).

STAT3 phosphorylation was markedly elevated in SH-SY5Y cells after transfection with an inhibitor of miR-29c-3p or miR-19b-3p (Wu et al., 2017). Based on computational predictions, miR-29c-3p, and miR-19b-3p target the same sequence in the 3′-UTR of STAT3. The hippocampus of the AD mice model and AD post-mortem brain shows substantial increases in phosphorylation of STAT3. In addition, STAT3 has served as a transcriptional regulator of BACE1, the crucial enzyme involved in producing Aβ (Millot et al., 2020). Other miRNAs that were observed to decrease their expression when the BACE1 level was increased were miR-16-5p and miR-19b-3p (Wu et al., 207). The overexpression of miR-16-5p or miR-19b-3p reduced the adverse effects of Aβ on cell viability and apoptosis in SH-SY5Y cells. Conversely, the knockdown of these miRNAs promoted the injury induced by Aβ.

In addition to miR-16-5p and miR-19b-3p, miR-107 and miR-384 also have complementary binding sites on the 3′UTR of BACE1, which leads to the suppression of its expression (Wang et al., 2008; Liu et al., 2014; Zhang et al., 2020). The relationship between miR-107 and circ_0049472 was confirmed through a dual-luciferase reporter and RNA pull-down assays. Circ_0049472 was found to be overexpressed in CSF and serum samples derived from AD patients. This gene was also overexpressed in SK-N-SH and CHP-212 cells induced with Aβ. The cell viability and apoptosis induced by the downregulation of circ_0049472 in Aβ-induced SK-N-SH and CHP-212 cells were abolished when miR-107 was silenced (Zheng et al., 2022).

Another gene that is related to AD is APP. A previous study discovered that miR-455 regulates APP by binding to its 3′UTR. This regulation was protective against mitochondrial and synaptic abnormalities induced by mutant APP in AD (Kumar et al., 2019). The PTGS2 was predicted to be targeted by miR-103, and it was discovered that PTGS2 is regulated in the opposite direction by miR-103 expression. In the PC12 cellular AD model, transfection with PTGS2 and miR-103 mimic plasmid decreased total neurite growth compared to the miR-103 mimic group. Additionally, there was an increase in cell apoptosis, suggesting that PTGS2 mimic mitigated the impact of miR-103 mimic on AD progression (Yang et al., 2018). In addition, circ_0000950 decreased the expression of miR-103 and increased the expression of PTGS2 in rat pheochromocytoma cell line PC12 cells and rat cerebral cortex neurons AD models. Overexpression of circ_0000950 resulted in increased neuron apoptosis, decreased neurite outgrowth, and elevated levels of IL-1β, IL-6, and TNF-α. On the other hand, the knockdown of circ_0000950 inhibited neuron apoptosis, promoted neurite outgrowth, and reduced levels of IL-1β, IL-6, and TNF-α. These effects were observed through the direct sponging of miR-103 (Yang et al., 2019).

Certain miRNAs, such as miR-455-3p and miR-210, have been found to possess protective mechanisms. According to Table 1, the expression of these miRNAs was upregulated and downregulated in AD patients, respectively. This condition has raised speculation that these miRNAs are protective in reducing AD severity. There are variations in the expression patterns of different miRNAs, such as miR-331-3p, miR-125-5p, and miR-384, across various biological sources. These changes in their environment potentially lead to their ability to respond, or they may also have a protective role in AD severity. Nevertheless, the precise mechanisms governing the expression of these microRNAs remain unidentified based on current knowledge.

circRNAs are known for their high conservation and structural stability. They have been found to play a crucial role in the onset and progression of numerous diseases. Thus, they could be valuable biomarkers or targets for therapy. Nevertheless, there are still barriers that must be addressed. This includes investigating how circRNAs interact with specific miRNAs, mRNAs, and proteins. By understanding these interactions, we can uncover a regulatory network that plays a role in AD development. To confirm the clinical relevance of circRNAs in AD, further investigation is required to understand their mechanisms and establish correlations. This will involve analyzing a large cohort of patient samples. Furthermore, there is an urgent need for technical advancements in accurately quantifying a specific circRNA and effectively silencing it without impacting the expression of the parent linear transcript.

Also, miRNAs can be detected in the bloodstream and potentially be used as biomarkers for the disease. It is essential to mention that miRNA can be detected in CSF. However, miRNA has the unique ability to traverse the blood-brain barrier and maintain its integrity, protected against degradation. This is achieved by its interaction with protein complexes and its containment within membrane-bound vesicles, such as exosomes (Hill, 2019). miRNA levels in circulation have the potential to indicate neuronal function and dysfunction accurately. This suggests they could be used as innovative therapeutic targets for treating dementia. Several recent meta-analyses have been conducted to establish a unified miRNA signature for AD. For example, in their study, Swarbrick et al. (2019) discovered a peripheral blood microRNA signature comprising ten molecules that could be linked to Braak Stage III.

In contrast, Bottero and Potashkin (2019) made predictions regarding miRNAs that were expected to influence the expression of genes known to exhibit differential expression in individuals with MCI and AD. Diagnostic tools and the significant heterogeneity among studies have limited the use of miRNAs as biomarkers and identifying specific miRNAs associated with AD. The heterogeneity observed in this context can be attributed to several factors, including variations in sample handling and the utilization of different profiling techniques such as microarrays, NGS, RT-qPCR, and other analytical approaches. Hence, further research is required to develop consistent protocols, determine dependable biomarkers, and comprehend the functional consequences of miRNA dysregulation in AD.

6 Conclusion

In summary, miRNA and circRNA represent promising avenues for developing non-invasive biomarkers for Alzheimer's disease. The continued exploration of these non-coding RNA molecules in the context of AD has the potential to revolutionize AD diagnosis,

monitoring, and therapeutic interventions, ultimately contributing to better disease management and improved quality of life for affected individuals.

Author contributions

SA and AH contributed to the conception and design of the review. SA, AH, and NIMI organised the structure of the review content. SA wrote the first draft of the manuscript. SA, NAMP, NIM, and NIMI wrote sections of the manuscript. SA and AH contributed equally to this work and share corresponding authorship. All authors contributed to the article and approved the submitted version.

Funding

The research was funded by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme under the grant number FRGS/1/2019/SKK08/UKM/03/2 and by the Faculty of Medicine, Universiti Kebangsaan Malaysia (project code: FF-2020-147).

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

Akhter, R., Shao, Y., Shaw, M., Formica, S., Khrestian, M., Leverenz, J. B., et al. (2018). Regulation of ADAM10 by miR-140-5p and potential relevance for Alzheimer's disease. *Neurobiol. Aging* 63, 110–119. doi:10.1016/j.neurobiolaging.2017.11.007

Banzhaf-Strathmann, J., Benito, E., May, S., Arzberger, T., Tahirovic, S., Kretzschmar, H., et al. (2014). MicroRNA-125b induces tau hyperphosphorylation and cognitive deficits in Alzheimer's disease. *EMBO J.* 33, 1667–1680. doi:10.15252/embj.201387576

Beylerli, O., Ilgiz Gareev, I., Sufianov, A., Ilyasova, T., and Zhang, F. (2022). The role of microRNA in the pathogenesis of glial brain tumors. *Non-coding RNA Res.* 7 (2), 71–76. doi:10.1016/j.ncrna.2022.02.005

Bhatlekar, S., Manne, B. K., Basak, I., Edelstein, L. C., Tugolukova, E., Stoller, M. L., et al. (2020). miR-125a-5p regulates megakaryocyte proplatelet formation via the actin-bundling protein L-plastin. *Blood* 135 (15), 1760–1772. doi:10.1182/blood.2020005230

Bhatnagar, S., Chertkow, H., Schipper, H. M., Yuan, Z., Shetty, V., Jenkins, S., et al. (2014). Increased microRNA-34c abundance in Alzheimer's disease circulating blood plasma. *Front. Mol. Neurosci.* 7 (2), 2. doi:10.3389/FNMOL.2014.00002

Bottero, V., and Potashkin, J. A. (2019). Meta-analysis of gene expression changes in the blood of patients with mild cognitive impairment and Alzheimer's disease dementia. *Int. J. Mol. Sci.* 20 (21), 5403. doi:10.3390/ijms20215403

Conn, S. J., Pillman, K. A., Toubia, J., Conn, V. M., Salmanidis, M., Phillips, C. A., et al. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* 6, 1125–1134. doi:10.1016/j.cell.2015.02.014

Cortes-Canteli, M., Zamolodchikov, D., Ahn, H. J., Strickland, S., and Norris, E. H. (2012). Fibrinogen and altered hemostasis in Alzheimer's disease. *J. Alzheimer's Dis.* 32 (3), 599–608. doi:10.3233/JAD-2012-120820

Denk, J., Boelmans, K., Siegismund, C., Lassner, D., Arlt, S., and Jahn, H. (2015). MicroRNA profiling of CSF reveals potential biomarkers to detect Alzheimer`s disease. *PLOS ONE* 10 (5), e0126423. doi:10.1371/journal.pone.0126423

Derkow, K., Rossling, R., Schipke, C., Kruger, C., Bauer, J., Fahling, M., et al. (2018). Distinct expression of the neurotoxic microRNA family let-7 in the cerebrospinal fluid of patients with Alzheimer's disease. *PLOS ONE* 13 (7), e0200602. doi:10.1371/journal.pone.0200602

Dong, Z. W., Gu, H. J., Guo, Q., Liang, S., Xue, J., Yao, F., et al. (2021). Profiling of serum exosome miRNA reveals the potential of a miRNA panel as diagnostic biomarker for Alzheimer's disease. *Mol. Neurobiol.* 58 (7), 3084–3094. doi:10.1007/s12035-021-02323-y

Dong, Z. W., Gu, H. J., Guo, Q., Liu, X. L., Li, F. F., 2u, H. L., et al. (2022). Circulating small extracellular vesicle-derived miR-342-5p ameliorates beta-amyloid formation via targeting beta-site APP cleaving enzyme 1 in Alzheimer's disease. *Cells* 11 (23), 3830. doi:10.3390/cells11233830

Frigerio, C. S., Lau, P., Salta, E., Tournoy, J., Bossers, K., Vandenberghe, R., et al. (2013). Reduced expression of hsa-miR-27a-3p in CSF of patients with Alzheimer disease. *Neurology* 81 (24), 2103–2106. doi:10.1212/01.WNL.0000437306.37850.22

Gatsiou, A., Boeckel, J-N., Randriamboavonjy, V., and Stellos, K. (2012). MicroRNAs in platelet biogenesis and function: implications in vascular homeostasis and inflammation. *Curr. Vasc. Pharmacol.* 10, 524–531. doi:10.2174/157016112801784611

Ghetti, B., Oblak, A. L., Boeve, B. F., Johnson, K. A., Dickerson, B. C., and Goedert, M. (2015). Invited review: frontotemporal dementia caused by microtubule-associated protein tau gene (MAPT) mutations: A chameleon for neuropathology and neuroimaging. *Neuropathology Appl. Neurobiol.* 41 (1), 24–46. doi:10.1111/nan.12213

Gokool, A., Anwar, F., and Voineagu, I. (2020). The landscape of circular RNA expression in the human brain. *Biol. Psychiatry* 87 (3), 294–304. doi:10.1016/j.biopsych.2019.07.029

Guo, Z., Maki, M., Ding, R., Yang, Y., Zhang, B., and Xiong, L. (2014). Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues. *Sci. Rep.* 4, 5150. doi:10.1038/srep05150

Guria, A., Sharma, P., Natesan, S., and Pandi, G. (2020). Circular RNAs-the road less travelled. *Front. Mol. Biosci.* 10 (6), 146.

Han, B., Chao, J., and Yao, H. (2018). Circular RNA and its mechanisms in disease: from the bench to the clinic. *Pharmacol. Ther.* 187, 31–44. doi:10.1016/j.pharmthera.2018.01.010

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013a). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi:10.1038/nature11993

Hansen, T. B., Kiems, J., and Damgaard, C. K. (2013b). Circular RNA and miR-7 in cancer. *Cancer Res.* 73, 5609–5612. doi:10.1158/0008-5472.CAN-13-1568

Harper, S. J., Cowell, S. J., and Dawson, W. O. (2019). Changes in host microRNA expression during citrus tristeza virus induced disease. *J. Citrus Pathology* 6, 41164. doi:10.5070/C461041164

Hill, A. F. (2019). Extracellular vesicles and neurodegenerative diseases. *J. Neurosci.* 39 (47), 9269–9273. doi:10.1523/JNEUROSCI.0147-18.2019

Huang, C., Liang, D., Tatomer, D. C., and Wilusz, J. E. (2018). A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. *Genes & Dev.* 32, 639–644. doi:10.1101/gad.314856.118

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H. T., Orejuela, M. R., et al. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* 10, 170–177. doi:10.1016/j.celrep.2014.12.019

Jia, L-H., and Liu, Y-H. (2016). Downregulated serum miR-223 servers as biomarker in Alzheimer's disease. *Cell Biochem. Funct.* 34, 233–237. doi:10.1002/cbf.3184

Jin, Y., Tu, Q., and Liu, M. (2018). MicroRNA-125b regulates Alzheimer's disease through SphK1 regulation. *Mol. Med. Rep.* 18 (2), 2373–2380. doi:10.3892/mmr.2018.9156

Kiko, T., Nakagawa, K., Tsuduki, T., Furukawa, K., Arai, H., and Miyazawa, T. (2014). MicroRNAs in plasma and cerebrospinal fluid as potential markers for Alzheimer's disease. *J. Alzheimer's Dis.* 39 (2), 253–259. doi:10.3233/JAD-130932

Kikuchi, M., Sekiya, M., Hara, N., Miyashita, A., Kuwano, R., Ikeuchi, T., et al. (2020). Disruption of a RAC1-centred network is associated with Alzheimer's disease pathology and causes age-dependent neurodegeneration. *Hum. Mol. Genet.* 29 (5), 817–833. doi:10.1093/hmg/ddz320

Koga, Y., Yasunaga, M., Moriya, Y., Akasu, T., Fujita, S., Yamamoto, S., et al. (2011). Exosome can prevent RNase from degrading microRNA in feces. *J. Gastrointest. Oncol.* 2 (4), 215–222. doi:10.3978/j.issn.2078-6891.2011.015

Komatsu, S., Kitai, H., and Suzuki, H. I. (2023). Network regulation of microRNA biogenesis and target interaction. *Cells* 12 (2), 306. doi:10.3390/cells12020306

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, 675–691. doi:10.1038/s41576-019-0158-7

Kumar, P., Dezso, Z., MacKenzie, C., Oestreicher, J., Agoulnik, S., Byrne, M., et al. (2013). Circulating miRNA biomarkers for Alzheimer's disease. *PLOS ONE* 8 (7), e69807. doi:10.1371/journal.pone.0069807

Kumar, S., Morton, H., Sawant, N., Orlov, E., Bunguin, L. E., Pradeepkiran, J. A., et al. (2021a). MicroRNA-455-3p improves synaptic, cognitive functions and extends lifespan: relevance to Alzheimer's disease. *Redox Biol.* 48, 102182. doi:10.1016/j.redox.2021.102182

Kumar, S., Reddy, A. P., Yin, X., and Reddy, H. (2019). Novel MicroRNA-455-3p and its protective effects against abnormal APP processing and amyloid beta toxicity in Alzheimer's disease. *Biochimica Biophysica Acta - Mol. Basis Dis.* 1865 (9), 2428–2440. doi:10.1016/j.bbadis.2019.06.006

Kumar, S., and Reddy, P. H. (2021b). Elevated levels of MicroRNA-455-3p in the cerebrospinal fluid of Alzheimer's patients: A potential biomarker for Alzheimer's disease. *Biochimica Biophysica Acta-Molecular Basis Dis.* 1867 (4), 166052. doi:10.1016/j.bbadis.2020.166052

Kumar, S., and Reddy, P. H. (2018). MicroRNA-455-3p as a potential biomarker for Alzheimer's disease: an update. *Front. Aging Neurosci.* 10, 41. doi:10.3389/fnagi.2018.00041

Kumar, S., Vijayan, M., and Reddy, H. (2017). MicroRNA-455-3p as a potential peripheral biomarker for Alzheimer's disease. *Hum. Mol. Genet.* 26 (19), 3808–3822. doi:10.1093/hmg/ddx267

Lee, B. K., Kim, M. H., Lee, S. Y., Son, S. J., Hong, C. H., and Jung, Y. S. (2020). Downregulated platelet miR-1233-5p in patients with Alzheimer's pathologic change with mild cognitive impairment is associated with a beta-induced platelet activation via P-selectin. *J. Clin. Med.* 9 (6), 1642. doi:10.3390/jcm9061642

Lehmann, S. M., Krüger, C., Park, B., Derkow, K., Rosenberger, K., Baumgart, J., et al. (2012). An unconventional role for miRNA: let-7 activates toll-like receptor 7 and causes neurodegeneration. *Nat. Neurosci.* 15 (6), 827–835. doi:10.1038/NN.3113

Leidinger, P., Backes, C., Deutscher, S., Schmitt, K., Mueller, S. C., Frese, K., et al. (2013). A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biol.* 14 (7), R78–R16. doi:10.1186/gb-2013-14-7-r78

Li, L., Zheng, Y. C., Kayani, M. U. R., Xu, W., Wang, G. Q., Sun, P., et al. (2017). Comprehensive analysis of circRNA expression profiles in humans by RAISE. *Int. J. Oncol.* 51 (6), 1625–1638. doi:10.3892/ijo.2017.4162

Li, P., Xu, Y., Wang, B., Huang, J., and Li, Q. (2020a). miR-34a-5p and miR-125b-5p attenuate Aβ-induced neurotoxicity through targeting BACE1. *J. Neurological Sci.* 413, 116793. doi:10.1016/j.jns.2020.116793

Li, W., Li, X., Xin, X., Kan, P-C., and Yan, Y. (2016). MicroRNA-613 regulates the expression of brain-derived neurotrophic factor in Alzheimer's disease. *Biosci. Trends* 10 (5), 372–377. doi:10.5582/bst.2016.01127

Li, Y., Fan, H., Sun, J., Ni, M., Zhang, L., Chen, C., et al. (2020b). Circular RNA expression profile of Alzheimer's disease and its clinical significance as biomarkers for the disease risk and progression. *Int. J. Biochem. Cell Biol.* 123, 105747. doi:10.1016/j.biocel.2020.105747

Liu, C-G., Wang, J-L., Li, L., and Wang, P-C. (2014). MicroRNA-384 regulates both amyloid precursor protein and β-secretase expression and is a potential biomarker for Alzheimer's disease. *Int. J. Mol. Med.* 34, 160–166. doi:10.3892/ijmm.2014.1780

Liu, Q., and Lei, C. (2021). Neuroprotective effects of miR-331-3p through improved cell viability and inflammatory marker expression: correlation of serum miR-331-3p levels with diagnosis and severity of Alzheimer's disease. *Exp. Gerontol.* 144, 111187. doi:10.1016/j.exger.2020.111187

Liu, Y., He, X., Li, Y., and Wang, T. (2018). Cerebrospinal fluid CD4+ T lymphocyte-derived miRNA-let-7b can enhances the diagnostic performance of Alzheimer's disease biomarkers. *Biochem. Biophysical Res. Commun.* 495 (1), 1144–1150. doi:10.1016/j.bbrc.2017.11.122

Long, J. M., Maloney, B., Rogers, J. T., and Lahiri, D. K. (2019). Novel upregulation of amyloid-beta precursor protein (APP) by microRNA-346 via targeting of APP mRNA 5 '-untranslated region: implications in Alzheimer's disease. *Mol. Psychiatry* 24 (3), 345–363. doi:10.1038/s41380-018-0266-3

Lu, L., Dai, W. Z., Zhu, X. C., and Ma, T. (2021). Analysis of serum miRNAs in Alzheimer's disease. *Am. J. Alzheimer's Dis. Other Dementias* 36, 15333175211021712–15333175211021719. doi:10.1177/15333175211021712

Lu, Y., Tan, L., and Wang, X. (2019). Circular HDAC9/microRNA-138/Sirtuin-1 pathway mediates synaptic and amyloid precursor protein processing deficits in Alzheimer's disease. *Neurosci. Bull.* 35, 877–888. doi:10.1007/s12264-019-00361-0

Lugli, G., Cohen, A. M., Bennett, D. A., Shah, R. C., Fields, C. J., Hernandez, A. G., et al. (2015). Plasma exosomal miRNAs in persons with and without alzheimer disease: altered expression and prospects for biomarkers. *PLOS ONE* 10 (10), e0139233. doi:10.1371/journal.pone.0139233

Lusardi, T. A., Phillips, J. I., Wiedrick, J. T., Harrington, C. A., Lind, B., Lapidus, J. A., et al. (2017). MicroRNAs in human cerebrospinal fluid as biomarkers for Alzheimer's disease. *J. Alzheimer's Dis.* 55 (3), 1223–1233. doi:10.3233/JAD-160835

Ma, Y., Liu, Y., and Jiang, Z. (2020). CircRNAs: A new perspective of biomarkers in the nervous system. *Biomed. Pharmacother.* 128, 110251. doi:10.1016/j.biopha.2020.110251

Maciotta, S., Meregalli, M., and Torrente, Y. (2013). The involvement of microRNAs in neurodegenerative diseases. *Front. Cell. Neurosci.* 7 (DEC), 265. doi:10.3389/fncel.2013.00265

McKeever, P. M., Schneider, R., Taghdiri, F., Weichert, A., Multani, N., Brown, R. A., et al. (2018). MicroRNA expression levels are altered in the cerebrospinal fluid of patients with young-onset Alzheimer's disease. *Mol. Neurobiol.* 55 (12), 8826–8841. doi:10.1007/s12035-018-1032-x

Millot, P., San, C., Bennana, E., Hugon, J., Paquet, C., Hosten, B., et al. (2020). STAT3 inhibition reverses neuroinflammation and Aβ metabolism induced by systemic inflammation. *Basic Sci. Pathogenesis* 16 (2), e041019. doi:10.1002/alz.041019

Müller, M., Jäkel, L., Bruinsma, I. B., Claassen, J. A., Kuiperij, H. B., and Verbeek, M. M. (2016). MicroRNA-29a is a candidate biomarker for Alzheimer's disease in cell-free cerebrospinal fluid. *Mol. Neurobiol.* 53, 2894–2899. doi:10.1007/s12035-015-9156-8

Nagaraj, S., Laskowska-Kaszub, K., Dębski2, K. J., Wojsiat, J., Dąbrowski, M., Gabryelewicz, T., et al. (2017). Profile of 6 microRNA in blood plasma distinguish early-stage Alzheimer's disease patients from non-demented subjects. *Oncotarget* 8 (10), 16122–16143. doi:10.18632/oncotarget.15109

Nelson, P. T., and Wang, W-X. (2010). MiR-107 is reduced in Alzheimer's disease brain neocortex: validation study. *J. Alzheimer's Dis.* 21 (1), 75–79. doi:10.3233/JAD-2010-091603

Nie, C., Sun, Y. Z., Zhen, H. F., Guo, M., Ye, J. Y., Liu, Z. L., et al. (2020). Differential expression of plasma exo-miRNA in neurodegenerative diseases by next-generation sequencing. *Front. Neurosci.* 14, 438. doi:10.3389/fnins.2020.00438

O'Driscoll, L. (2015). Expanding on exosomes and ectosomes in cancer. *N. Engl. J. Med.* 372 (24), 2359–2362. doi:10.1056/NEJMCIBR1503100

Patop, I. L., Wüst, S., and Kadener, S. (2019). Past, present, and future of circRNAs. *EMBO J.* 38 (16), e100836. doi:10.15252/embj.2018100836

Paul, J., Strickland, S., and Melchor, J. P. (2007). Fibrin deposition accelerates neurovascular damage and neuroinflammation in mouse models of Alzheimer's disease. *J. Exp. Med.* 204 (8), 1999–2008. doi:10.1084/jem.20070304

Ren, Z., Yu, J., Wu, Z., Si, W., Li, X., Liu, Y., et al. (2018). MicroRNA-210-5p contributes to cognitive impairment in early vascular dementia rat model through targeting SNAP25. *Front. Mol. Neurosci.* 11, 388. doi:10.3389/fnmol.2018.00388

Rybak-Wolf, A., Stottmeister, C., Glažar, P., Jens, M., Pino, N., Giusti, S., et al. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58 (5), 870–885. doi:10.1016/j.molcel.2015.03.027

Sandau, U. S., McFarland, T. J., Smith, S. J., Galasko, D. R., Quinn, J. F., and Saugstad, J. A. (2022). Differential effects of APOE genotype on microRNA cargo of cerebrospinal fluid extracellular vesicles in females with Alzheimer's disease compared to males. *Front. Cells Dev. Biol.* 10, 864022. doi:10.3389/fcell.2022.864022

Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., and Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U. S. A.* 73 (11), 3852–3856. doi:10.1073/pnas.73.11.3852

Sekar, S., and Liang, W. S. (2019). Circular RNA expression and function in the brain. *Noncoding RNA Res.* 4 (1), 23–29. doi:10.1016/j.ncrna.2019.01.001

Shigemizu, D., Akiyama, S., Asanomi, Y., Boroevich, K. A., Sharma, A., Tsunoda, T., et al. (2019). Risk prediction models for dementia constructed by supervised principal component analysis using miRNA expression data. *Commun. Biol.* 2, 77. doi:10.1038/s42003-019-0324-7

Sierksma, A., Lu, A., Salta, E., Vanden Eynden, E., Callaerts-Vegh, Z., D'Hooge, R., et al. (2018). Deregulation of neuronal miRNAs induced by amyloid-beta or TAU pathology. *Mol. Neurodegener.* 13, 54. doi:10.1186/s13024-018-0285-1

Swarbrick, S., Wragg, N., Ghosh, S., and Stolzing, A. (2019). Systematic review of miRNA as biomarkers in Alzheimer's Disease. *Mol. Neurobiol.* 56 (9), 6156–6167. doi:10.1007/s12035-019-1500-y

Takousis, P., Sadlona, A., Schulzb, J., Wohlersc, I., Dobricicc, V., Middletona, L., et al. (2019). Differential expression of microRNAs in Alzheimer's disease brain, blood, and cerebrospinal fluid. *Alzheimer's Dementia* 15, 1468–1477. doi:10.1016/j.jalz.2019.06.4952

Tan, L., Yu, J. T., Tan, M. S., Liu, Q. Y., Wang, H. F., Zhang, W., et al. (2014). Genome-wide serum microRNA expression profiling identifies serum biomarkers for Alzheimer's disease. *J. Alzheimer's Dis.* 40 (4), 1017–1027. doi:10.3233/JAD-132144

Thind, A., and Wilson, C. (2016). Exosomal miRNAs as cancer biomarkers and therapeutic targets. *J. Extracell. Vesicles* 5, 31292. doi:10.3402/jev.v5.31292

Toivonen, J. M., Sanz-Rubio, D., López-Pérez, Ó., Marín-Moreno, A., Bolea, R., Osta, R., et al. (2020). MicroRNA alterations in a Tg501 mouse model of prion disease. *Biomolecules* 10 (6), 908. doi:10.3390/biom10060908

Vickers, K. C., Palmisano, B. T., Shoucri, B. M., Shamburek, R. D., and Remaley, A. T. (2011). MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell Biol.* 13 (4), 423–433. doi:10.1038/ncb2210

Wan, W., Liub, G., Lic, X., Liud, Y., Wang, Y., Pand, H., et al. (2021). MiR-191-5p alleviates microglial cell injury by targeting Map3k12 (mitogen-activated protein kinase

kinase kinase 12) to inhibit the MAPK (mitogen-activated protein kinase) signaling pathway in Alzheimer's disease. *Bioengineered* 12 (2), 12678–12690. doi:10.1080/21655979.2021.2008638

Wang, J., Chen, C., and Zhang, Y. (2020). An investigation of microRNA-103 and microRNA-107 as potential blood-based biomarkers for disease risk and progression of Alzheimer's disease. *J. Clin. Laboratory Analysis* 34 (1), e23006. doi:10.1002/jcla.23006

Wang, L., Shui, X. D., Mei, Y. X., Xia, Y. F., Lan, G. H., Hu, L., et al. (2022b). miR-143-3p inhibits aberrant tau phosphorylation and amyloidogenic processing of APP by directly targeting DAPK1 in Alzheimer's disease. *Int. J. Mol. Sci.* 23 (14), 7992. doi:10.3390/ijms23147992

Wang, L., Shui, X., Zhang, M., Mei, Y., Xia, Y., Lan, G., et al. (2022a). MiR-191-5p attenuates tau phosphorylation, Aβ generation, and neuronal cell death by regulating death-associated protein kinase 1. *ACS Chem. Neurosci.* 13 (24), 3554–3566. doi:10.1021/acschemneuro.2c00423

Wang, W-X., Rajeev, B. W., Stromberg, A. J., Ren, N., Tang, G., Huang, Q., et al. (2008). The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of β-site amyloid precursor protein cleaving enzyme 1. *J. Neurosci.* 28 (5), 1213–1223. doi:10.1523/JNEUROSCI.5065-07.2008

Wang, X., Tan, L., Lu, Y., Peng, J., Zhu, Y., Zhang, Y., et al. (2015). MicroRNA-138 promotes tau phosphorylation by targeting retinoic acid receptor alpha. *FEBS Lett.* 589 (6), 726–729. doi:10.1016/j.febslet.2015.02.001

Wang, Z., Qin, W., Wei, C. B., Tang, Y., Zhao, L. N., Jin, H. M., et al. (2018). The microRNA-1908 up-regulation in the peripheral blood cells impairs amyloid clearance by targeting ApoE. *Int. J. Geriatric Psychiatry* 33 (7), 980–986. doi:10.1002/gps.4881

Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., Huang, K. H., Lee, M. J., et al. (2010). The microRNA spectrum in 12 body fluids. *Clin. Chem.* 56 (11), 1733–1741. doi:10.1373/CLINCHEM.2010.147405

Weinberg, R. B., Mufson, E. J., and Counts, S. E. (2015). Evidence for a neuroprotective microRNA pathway in amnestic mild cognitive impairment. *Front. Neurosci.* 9, 430. doi:10.3389/fnins.2015.00430

Welden, J. R., Margvelani, G., Arizaca Maquera, K. A., Gudlavalleti, B., Miranda Sardón, S. C., Campos, A. R., et al. (2022). RNA editing of microtubule-associated protein tau circular RNAs promotes their translation and tau tangle formation. *Nucleic Acids Res.* 50 (22), 12979–12996. doi:10.1093/nar/gkac1129

Wu, Y. Q., Xu, J., Xu, J., Cheng, J., Jiao, D. M., Zhou, C., et al. (2017). Lower serum levels of miR-29c-3p and miR-19b-3p as biomarkers for Alzheimer's disease. *J. Exp. Med.* 242 (2), 129–136. doi:10.1620/tjem.242.129

Xiao, G., Chen, Q., and Zhang, X. (2021). MicroRNA-455–5p/CPEB1 pathway mediates Aβ-related learning and memory deficits in a mouse model of Alzheimer's disease. *Brain Res. Bull.* 177, 282–294. doi:10.1016/j.brainresbull.2021.10.008

Yang, H., Wang, H., Shang, H., Chen, X., Yang, S., Qu, Y., et al. (2019). Circular RNA circ_0000950 promotes neuron apoptosis, suppresses neurite outgrowth and elevates inflammatory cytokines levels via directly sponging miR-103 in Alzheimer's disease. *Cell Cycle* 18 (18), 2197–2214. doi:10.1080/15384101.2019.1629773

Yang, H., Wang, H., Shu, Y., and Li, X. (2018b). miR-103 promotes neurite outgrowth and suppresses cells apoptosis by targeting prostaglandin-endoperoxide synthase 2 in cellular models of Alzheimer's disease. *Front. Cell. Neurosci.* 12, 91. doi:10.3389/fncel.2018.00091

Yang, T. T., Liu, C. G., Gao, S. C., Zhang, Y., and Wang, P. C. (2018a). The serum exosome derived MicroRNA-135a, -193b, and -384 were potential Alzheimer's disease biomarkers. *Biomed. Environ. Sci.* 31 (2), 87–96. doi:10.3967/bes2018.011

Yılmaz, S. G., Erdal, M. E., Özge, A. A., and Sungur, M. A. (2016). Can peripheral microRNA expression data serve as epigenomic (upstream) biomarkers of Alzheimer's disease? *OMICS A J. Integr. Biol.* 20, 456–461. doi:10.1089/omi.2016.0099

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., et al. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat. Neurosci.* 18 (4), 603–610. doi:10.1038/nn.3975

Zamolodchikov, D., Chen, Z-L., Conti, B. A., and Strickland, S. (2015). Activation of the factor XII-driven contact system in Alzheimer's disease patient and mouse model plasma. *Biol. Sci.* 112 (13), 4068–4073. doi:10.1073/pnas.1423764112

Zen, K., and Zhang, C. Y. (2012). Circulating microRNAs: A novel class of biomarkers to diagnose and monitor human cancers. *Med. Res. Rev.* 32 (2), 326–348. doi:10.1002/MED.20215

Zhang, M., Han, W., Xu, Y., Li, D., and Xue, Q. (2021). Serum miR-128 serves as a potential diagnostic biomarker for Alzheimer's disease. *Neuropsychiatric Dis. Treat.* 17, 269–275. doi:10.2147/NDT.S290925

Zhang, N., Li, W-W., Lv, C-M., Gao, Y-W., Liu, X-L., and Li, Z. (2020). miR-16-5p and miR-19b-3p prevent amyloid β-induced injury by targeting BACE1 in SH-SY5Y cells. *Neuroreport* 31 (3), 205–212. doi:10.1097/WNR.0000000000001379

Zhao, X., Wang, S., and Sun, W. (2020). Expression of miR 28 3p in patients with Alzheimer's disease before and after treatment and its clinical value. *Exp. Ther. Med.* 20, 2218–2226. doi:10.3892/etm.2020.8920

Zhao, X., Zhong, Y., Wang, X., Shen, J., and An, W. (2022). Advances in circular RNA and its applications. *Int. J. Med. Sci.* 19 (6), 975–985. doi:10.7150/ijms.71840

Zhao, Y., Alexandrov, P. N., Jaber, V., and Lukiw, W. J. (2016). Deficiency in the ubiquitin conjugating enzyme UBE2A in Alzheimer's disease (AD) is linked to deficits in a natural circular miRNA-7 Sponge (circRNA; ciRS-7). *Genes* 7 (12), 116. doi:10.3390/genes7120116

Zheng, C., Xing, H., Chen, M., Li, C., Li, P., Wu, X., et al. (2022). Circ_0049472 regulates the damage of Aβ-induced SK-N-SH and CHP-212 cells by mediating the miR-107/KIF1B axis. *Exp. Brain Res.* 240, 2299–2309. doi:10.1007/s00221-022-06401-y

Zhou, W. Y., Cai, Z. R., Liu, J., Wang, D-S., Ju, H-Q., and Xu, R-H. (2020). Circular RNA: metabolism, functions and interactions with proteins. *Mol. Cancer* 19, 172. doi:10.1186/s12943-020-01286-3

Zhu, Y., Li, C., Sun, A., Wang, Y., and Zhou, S. (2015). Quantification of microRNA-210 in the cerebrospinal fluid and serum: implications for Alzheimer's disease. *Exp. Ther. Med.* 9, 1013–1017. doi:10.3892/etm.2015.2179

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 09 October 2023
DOI 10.3389/fgene.2023.1233842

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Sunny Sharma,
Rutgers, The State University of New Jersey, United States
Constantinos Stathopoulos,
University of Patras, Greece

*CORRESPONDENCE
Mark Bartlam,
bartlam@nankai.edu.cn
Gerlof Sebastiaan Winkler,
sebastiaan.winkler@nottingham.ac.uk

RECEIVED 02 June 2023
ACCEPTED 25 September 2023
PUBLISHED 09 October 2023

CITATION
Zhao Q, Pavanello L, Bartlam M and Winkler GS (2023), Structure and function of molecular machines involved in deadenylation-dependent 5′-3′ mRNA degradation.
Front. Genet. 14:1233842.
doi: 10.3389/fgene.2023.1233842

Structure and function of molecular machines involved in deadenylation-dependent 5′-3′ mRNA degradation

Qi Zhao[1], Lorenzo Pavanello[2], Mark Bartlam[1]* and Gerlof Sebastiaan Winkler[2]*

[1]State Key Laboratory of Medicinal Chemical Biology, College of Life Sciences, Nankai International Advanced Research Institute (Shenzhen Futian), Nankai University, Tianjin, China, [2]School of Pharmacy, University of Nottingham, University Park, Nottingham, United Kingdom

In eukaryotic cells, the synthesis, processing, and degradation of mRNA are important processes required for the accurate execution of gene expression programmes. Fully processed cytoplasmic mRNA is characterised by the presence of a 5′cap structure and 3′poly(A) tail. These elements promote translation and prevent non-specific degradation. Degradation via the deadenylation-dependent 5′-3′ degradation pathway can be induced by trans-acting factors binding the mRNA, such as RNA-binding proteins recognising sequence elements and the miRNA-induced repression complex. These factors recruit the core mRNA degradation machinery that carries out the following steps: i) shortening of the poly(A) tail by the Ccr4-Not and Pan2-Pan3 poly (A)-specific nucleases (deadenylases); ii) removal of the 5′cap structure by the Dcp1-Dcp2 decapping complex that is recruited by the Lsm1-7-Pat1 complex; and iii) degradation of the mRNA body by the 5′-3′ exoribonuclease Xrn1. In this review, the biochemical function of the nucleases and accessory proteins involved in deadenylation-dependent mRNA degradation will be reviewed with a particular focus on structural aspects of the proteins and enzymes involved.

KEYWORDS

RNA, poly(A), gene expression, RNA degradation pathway, RNA decay, nuclease, deadenylase complex

Introduction

In eukaryotic cells, mature mRNAs transported from the nucleus to the cytoplasm contain a 3′poly(A) tail and a 7-methylguanylate (m^7G) cap structure at the 5′end (Wilusz et al., 2001; Vicens et al., 2018; Passmore and Coller, 2022). These features are present on virtually all mRNAs with the notable exception of the histone coding mRNAs, which are characterised by the absence of a poly(A) tail. The poly(A) tail prevents non-specific degradation by 3′-5′ exoribonucleases, while the m^7G cap structure carries out a similar protective role of the 5′end of the mRNA. In addition to their role in preventing premature degradation, both modifications are also required for efficient translation (Wilusz et al., 2001; Vicens et al., 2018; Passmore and Coller, 2022).

Cis-acting sequence elements located in the 3′untranslated region (UTR) of mRNAs can induce degradation (Raisch and Valkov, 2022; Pavanello et al., 2023). The trans-acting factors recruited by these sequence elements typically induce mRNA degradation via deadenylation coupled to the 5′-3′ degradation pathway (Muhlrad et al., 1994; Stoecklin

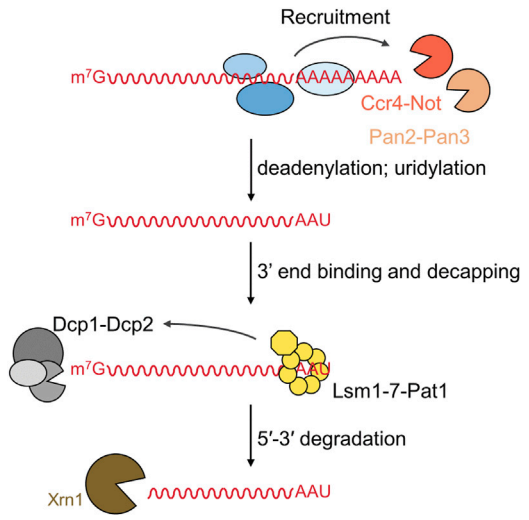


FIGURE 1
Overview of the deadenylation-dependent 5′-3′ mRNA degradation pathway. The 5′-3′ degradation pathway involves the following sequential steps: (i) deadenylation by the Pan2-Pan3 and Ccr4-Not complexes; (ii) uridylation of oligoadenylated RNA; (iii) 3′end binding by the Lsm1-7-Pat1 complex followed by recruitment of and decapping by the Dcp1-Dcp2 complex; and (iv) degradation of the mRNA by the 5′-3′ exoribonuclease Xrn1.

et al., 2006; Bonisch et al., 2007). Well-characterised examples include the A/U-rich element (ARE) (Stoecklin et al., 2006), microRNA binding sites (Behm-Ansmant et al., 2006; Piao et al., 2010), and sequence elements recognised by developmental regulators such as Smaug and Bicaudal-C (Zaessinger et al., 2006; Chicoine et al., 2007; Pekovic et al., 2023). In addition to recruitment by mRNA-specific factors, the 5′-3′ degradation pathway can be induced by general mechanisms involving the interaction between the cytoplasmic poly(A)-binding protein (PABPC) 1 and members of the BTG/Tob family of proteins, which include TOB1 and its paralogue TOB2, and the BTG1 and BTG2 paralogues that are frequently mutated in non-Hodgkin lymphoma (Mauxion et al., 2009; Winkler, 2010; Yuniati et al., 2018; Almasmoum et al., 2021).

The components involved in the deadenylation-dependent 5′-3′ degradation pathway are conserved, and detailed insight into the role of the core components has become available through a large number of studies investigating the structure of the enzymes and proteins involved from various model organisms, including *Saccharomyces cerevisiae*, *Schizosaccharomyes pombe*, *Kluyveromyces lactis*, *Drosophila melanogaster*, and *Homo sapiens*. Here, we will review the structure and function of the core components in deadenylation-dependent 5′-3′ degradation and focus on the wealth of structural insight obtained in the past decade.

Overview of the deadenylation-dependent 5′-3′ degradation pathway

The 5′-3′ mRNA degradation involves distinct steps that are completed consecutively (Parker and Song, 2004; Jonas and Izaurralde, 2015) (Figure 1). The first phase in the 5′-3′ degradation pathway is the shortening of the poly(A) tail (deadenylation) (Raisch and Valkov, 2022; Pavanello et al., 2023). This step is carried out by two multi-subunit nuclease complexes that selectively recognise poly(A) RNA. While the Pan2-Pan3 deadenylase may primarily be involved in general deadenylation, degradation of a target mRNA appears to be mainly regulated by recruitment of the Ccr4-Not complex by factors binding in the 3′UTR of the messenger (reviewed in Pavanello et al., 2018; Raisch and Valkov, 2022; Wahle and Winkler, 2013). Following deadenylation, the Lsm1-7-Pat1 complex binds at the 3′end of the mRNA and recruits the Dcp1-Dcp2 decapping complex, which removes the m^7G cap from the 5′end (Parker and Song, 2004; Jonas and Izaurralde, 2015). In *S. pombe* and mammalian cells, uridylation of degradation intermediates takes place after deadenylation, but prior to binding of the Lsm1-7-Pat1 complex (Rissland and Norbury, 2009; Scott and Norbury, 2013; Lim et al., 2014). In the final stage of the 5′-3′ degradation pathway, the mRNA body is degraded by the conserved 5′-3′ exonuclease Xrn1 (Parker and Song, 2004; Nagarajan et al., 2013).

Deadenylation: Shortening of the poly (A) tail

The main deadenylases implicated in the 5′-3′ degradation pathway are the Pan2-Pan3 complex and the multi-subunit Ccr4-Not deadenylase (Boeck et al., 1996; Brown et al., 1996; Tucker et al., 2001; Yamashita et al., 2005; Wahle and Winkler, 2013). Pan2-Pan3 has an intrinsic affinity for cytoplasmic poly(A)-binding protein (PABPC) and prefers long poly(A)-PABPC ribonucleoprotein particles (RNPs). In addition, the GW182 (TNRC6) component of the miRNA repression complex has been implicated in mRNA-specific recruitment of Pan2-Pan3 (Braun et al., 2011; Fabian et al., 2011). By contrast, Ccr4-Not has no direct affinity for PABPC. Instead, it can interact with member so the BTG/Tob family of proteins in metazoan organisms (Mauxion et al., 2009; Winkler, 2010). The BTG1/BTG2 and TOB1/TOB2 components of this family have been shown to interact with PABPC and Ccr4-Not, thereby stimulating deadenylation of poly(A)-PABPC RNPs (Ezzeddine et al., 2007; Stupfler et al., 2016). Ccr4-Not appears to be the dominant deadenylase recruited to target mRNAs as a large number of interactions with RNA-binding proteins have been established, and the mode of recruitment has been determined in several cases in molecular detail (Raisch and Valkov, 2022; Pavanello et al., 2023).

Deadenylation: the Pan2-Pan3 complex

The Pan2-Pan3 deadenylase is one of two deadenylase complexes implicated in the initiation of the 5'-3' decay of mRNA by Xrn1. The Pan2-Pan3 complex is highly conserved (Uchida et al., 2004; Wahle and Winkler, 2013) with an atypical architecture in which an asymmetric Pan3 homodimer is bound by a single Pan2 subunit (Pan2:Pan3 = 1:2) (Jonas et al., 2014; Schafer et al., 2014; Wolf et al., 2014).

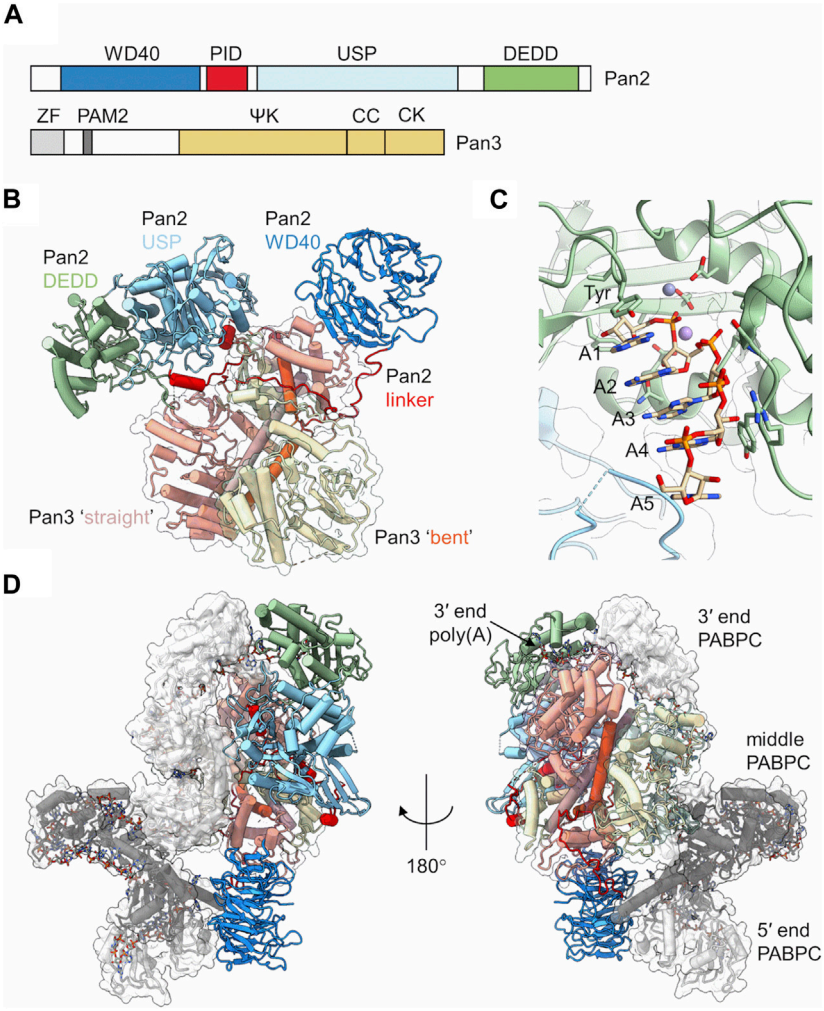


FIGURE 2
The Pan2-Pan3 deadenylase. **(A)** Schematic diagram of the domain organisation of Pan2 and Pan3. PID, Pan3 interacting domain; USP, ubiquitin-specific protease; DEDD, Asp-Glu-Asp-Asp catalytic domain. ZF, zinc finger; ΨK, pseudo-kinase; CC, coiled-coil; CK, C-terminal knob. **(B)** Overview of the *S. cerevisiae* Pan2-$(Pan3)_2$ complex. Indicated are the two Pan3 protomers (tan, yellow) with the long α-helices of the coiled coil highlighted (orange, brown). Also indicated are the domains of Pan2: WD40 (blue), PID linker domain (red), USP (light blue), and catalytic DEDD domain (green). PDB entry: 6R5K (Schafer et al., 2019). **(C)** Poly **(A)** recognition by the Pan2 catalytic domain. Indicated are A5 form A helical RNA stacked onto a conserved tyrosine. PDB entry: 6R9J (Tang et al., 2019). Also indicated are two metal ions obtained by superimposition of the *S. pombe* Pop2 enzyme. PDB entry: 3G0Z (Andersen et al., 2009). **(D)** Recognition of poly**(A)**-PABPC ribonucleoprotein particles by *S. cerevisiae* Pan2-Pan3. Indicated are the domains of Pan2 and Pan3 as shown in panel **(B)**. In addition, three PABPC proteins are indicated (grey; dark grey; white). The RNA strand is shown using a stick model. PDB entry: 6R5K (Schafer et al., 2019).

The catalytic subunit Pan2

Pan2 is the large subunit of the complex (Figure 2A). It contains a C-terminal exoribonuclease that belongs to the DEDD family of exonucleases. It is Mg^{2+} dependent and releases 5′-AMP upon hydrolysis of the poly(A) mRNA tail using a two-metal catalytic mechanism (Uchida et al., 2004; Wahle and Winkler, 2013; Tang et al., 2019). Pan2 displays low affinity for RNA, and has modest catalytic activity in absence of Pan3 (Schafer et al., 2014; Wolf et al., 2014). By contrast, Pan2 shows readily detectable deadenylation activity and specificity for poly(A) upon binding of its complex partner Pan3 (Wolf et al., 2014). In addition to its catalytic domain, Pan2 contains three further conserved regions (Uchida et al., 2004; Wahle and Winkler, 2013; Jonas et al., 2014; Schafer et al., 2014; Wolf et al., 2014). At the N-terminus, a WD40 domain is located, which forms a typical seven-blade β-propeller that mediates interactions with Pan3 (Jonas et al., 2014; Schafer et al., 2014) (Figure 2B). Together with the WD40 domain, a Pan3-Interacting Domain (PID) linker region adjacent to the WD40 domain is also required for complex formation with Pan3 (Jonas et al., 2014; Schafer et al., 2014; Wolf et al., 2014). This sequence contains several conserved residues responsible for binding the coiled coil regions of the Pan3 homodimer (Figure 2B). Upon binding the Pan3 homodimer, the linker sterically prevents a second WD40 domain from binding resulting in a stable complex composed of a single Pan2 subunit, and two Pan3 protomer (Jonas et al., 2014; Schafer et al., 2014). Located between the linker region and the catalytic DEDD domain is a ubiquitin-specific protease (USP) domain. This domain, however, lacks residues in the catalytic triad that are essential

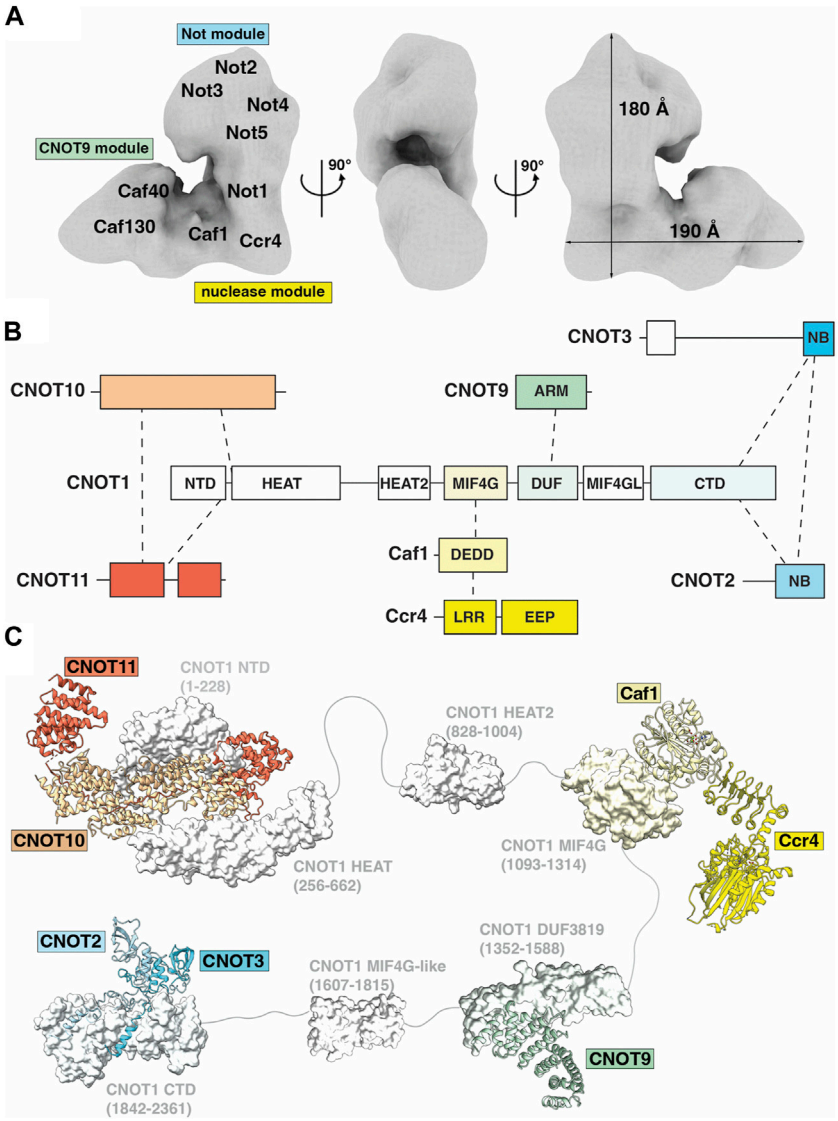


FIGURE 3
The Ccr4-Not deadenylase. **(A)** Surface representation of the three-dimensional 'L-shaped' map from electron microscopic analysis of the *S. cerevisiae* Ccr4-Not complex. EMDB: EMD-1901 (Nasertorabi et al., 2011). Tentative positions of subunits and modules are labelled and approximate dimensions of the complex are given. **(B)** Overview of the Ccr4-Not complex subunit architecture. **(C)** Available structures for the Ccr4-Not complex. Indicated are the N-terminal module composed of the N-terminal region of CNOT1, CNOT10 (light orange) and CNOT11 (dark orange), PDB entry: 8BFI (Mauxion et al., 2023); MIF4G-like domain 1 of CNOT1, PDB entry: 4J8S (Fabian et al., 2013); the nuclease module composed of the CNOT1 MIF4G domain, Caf1/CNOT7 (light yellow) and Ccr4/CNOT6L (dark yellow), PDB entries 3NGQ and 7VOI (Wang et al., 2010; Zhang et al., 2022); the CNOT9 module, PDB entries 4CT6 or 4CRV (Chen et al., 2014; Mathys et al., 2014) composed of the DUF3819 domain of CNOT1 and CNOT9 (green); a second MIF4G-like domain of CNOT1 modelled using AlphaFold (Jumper et al., 2021); and the NOT module composed of the CNOT1 C-terminal domain and the conserved NOT-Box regions located at the C-termini of CNOT2 (light blue) and CNOT3 (blue), PDB entry: 4C0D (Boland et al., 2013). Colours correspond to subunits in panel **(B)**. Numbers in brackets refer to the amino acid residues of CNOT1. Modules are connected by grey linkers to indicate the flexibility of the complex. **(B,C)** are adapted from Figure 2, Pavanello et al. (2023), used under CC BY 4.0.

for activity, and the domain shows no protease activity (Quesada et al., 2004). Instead, the Pan2 USP domain engages in extensive interactions with the Pan2 C-terminal exonuclease domain (Jonas et al., 2014; Schafer et al., 2014; Wolf et al., 2014) (Figure 2B).

The regulatory subunit Pan3

Pan3 forms an asymmetric homodimer when assembled into the yeast Pan2-Pan3 complex (Christie et al., 2013; Jonas et al., 2014; Schafer et al., 2014; Wolf et al., 2014). It displays five conserved regions (Figure 2A). At the N-terminus, a CCCH-type zinc-finger domain is located that has preference for poly(A) binding (Wolf et al., 2014). In addition, a short PAM2 motif is present, which can associate with the C-terminal MLLE domain of cytoplasmic poly(A)-binding protein (Uchida et al., 2004; Siddiqui et al., 2007). The central part of Pan3 contains a pseudokinase (ΨK) domain. The ΨK domain lacks catalytic residues required for kinase activity, but has retained the ability to bind ATP in a Mg^{2+}-dependent manner (Christie et al., 2013).

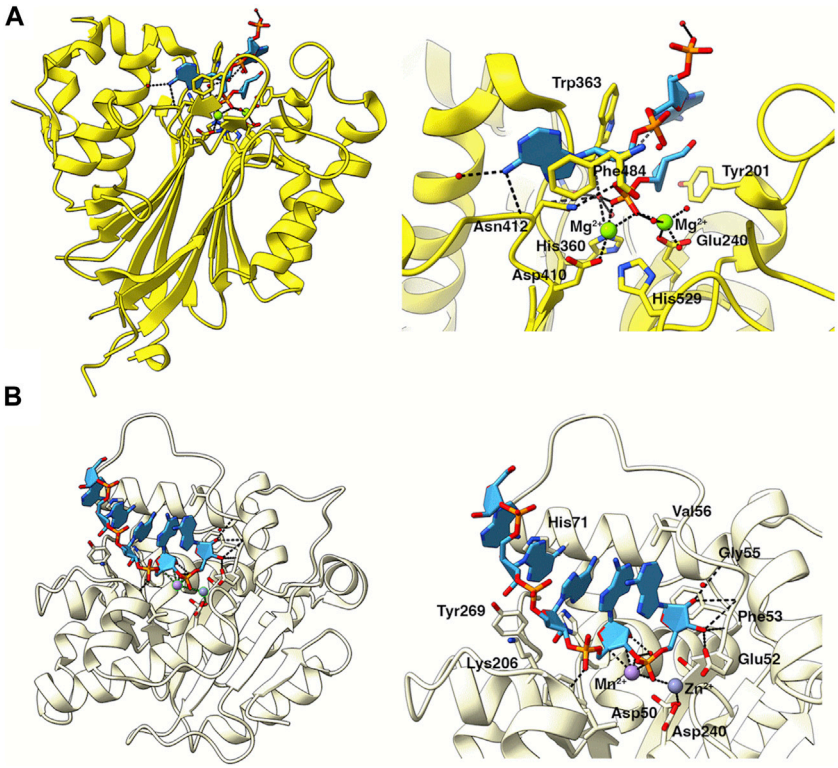


FIGURE 4
Poly **(A)** recognition by the catalytic subunits of Ccr4-Not. **(A)** Structure of the nuclease domain of human Ccr4/CNOT6L (yellow) in complex with poly **(A)** DNA (light blue). PDB entry: 3NGO (Wang et al., 2010). The right panel shows an enlarged view of poly **(A)** in the catalytic site. Key residues involved in catalysis or substrate recognition are shown as sticks as labelled. Mg^{2+} ions are shown as green spheres. Bonds are indicated by black dashed lines. **(B)** Structure of Caf1 (pale yellow), PDB entry 3G0Z (Andersen et al., 2009), with a poly **(A)** RNA substrate (light blue) modelled from the homologous Pan2 structure in complex with poly $(A)_7$, PDB entry: 6R9J (Tang et al., 2019). The right panel shows an enlarged view of poly **(A)** in the catalytic site. Key residues involved in catalysis or substrate recognition are shown as sticks as labelled. Metal ions are shown as spheres. Bonds are indicated by black dashed lines.

Moreover, the ability for nucleotide binding seems required for the ribonuclease activity of the Pan2-Pan3 complex (Christie et al., 2013). A coiled-coil (CC) region connects the ΨK domain and the C-terminal knob (CK) domain in the C-terminus of the protein. The Pan3 homodimer is formed by extensive interactions between the ΨK, CC and CK domains of one Pan3 protomer with the corresponding regions in the second protomer (Christie et al., 2013; Jonas et al., 2014; Schafer et al., 2014; Wolf et al., 2014) (Figure 2B). The resulting homodimer is asymmetric, and a notable difference can be seen in the CC regions of the two Pan3 protomers. In one protomer, the CC region forms a long 'straight' α-helix, whereas the CC region of the second protomer has a pronounced bend (Figure 2B). The Pan3 homodimer binds a single Pan2 subunit via the N-terminal WD40 domain and the PID linker of Pan2 (Jonas et al., 2014; Schafer et al., 2014; Wolf et al., 2014). The WD40 domain of Pan2 binds laterally to the CK domain of the Pan3 protomer containing the CC in the 'bent' orientation. The PID linker of Pan2 wraps around the CC regions and interacts with the CK domain of the protomer containing the 'straight' CC conformation thereby preventing the association of the WD40 domain of a second Pan2 protomer (Jonas et al., 2014; Schafer et al., 2014).

Recognition and deadenylation of poly(A) ribonucleoprotein particles

The catalytic domain of Pan 2 can accommodate up to seven adenosines in the active site (Tang et al., 2019). Intriguingly, the poly(A) substrate is not recognised by selective interactions with the nucleobases. Instead, extensive hydrogen bonding takes place between Pan2 residues and the phosphate-sugar backbone of the poly(A) substrate (Tang et al., 2019) (Figure 2C). Specificity of substrate recognition is based on the intrinsic ability of poly(A) RNA to adopt an A form single stranded helical RNA conformation that depends on multiple base-base stacking interactions within the poly(A) sequence (Figure 2C). The 5′terminal adenosine stacks onto a conserved tyrosine in the active site, which positions the scissile bond towards the metal ions resulting in release of AMP (Figure 2C). The presence of guanosine residues, which disrupt the helical A-form RNA structure, interfere with productive nucleolytic activity, while cytosine and uracil residues, which allow the formation of A-form RNA, are permitted (Tang et al., 2019).

Whereas the N-terminal zinc-finger domain of Pan3 contributes to RNA binding and specificity for poly(A), removal of this domain does not significantly impair *in vitro* deadenylase activity (Wolf

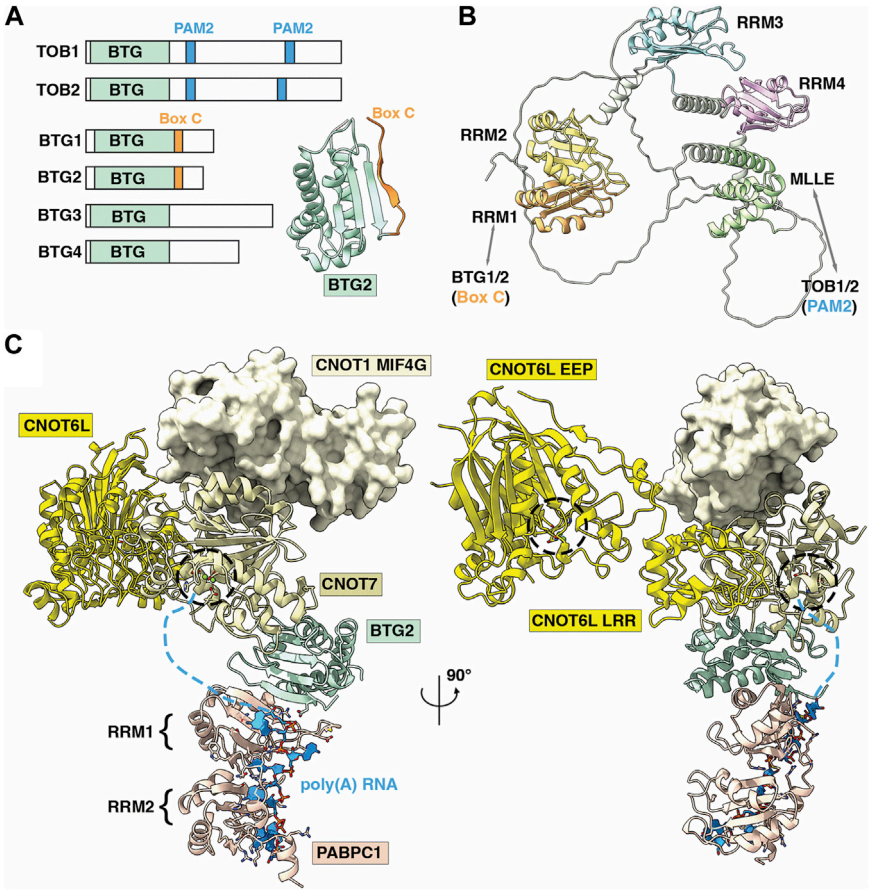


FIGURE 5
Role of the BTG-Tob proteins in deadenylation of poly **(A)**-PABPC ribonucleoprotein particles by Ccr4-Not. **(A)** Overview of BTG/TOB proteins. Shown are the conserved BTG (APRO) domain (light green), PAM2 motifs in TOB1 and TOB2 (light blue), and Box C regions in BTG1 and BTG2 (yellow). Inset is the BTG2 structure showing the APRO domain (light green) and Box C region (orange). PDB entry: 3DJU (Yang et al., 2008). **(B)** AlphaFold model of human PABPC1 (Uniprot P11940) (Jumper et al., 2021). Indicated are RRM1 (orange) that interacts with Box C regions of BTG1/BTG2, RRM2-RRM4 (yellow, light blue, pink), and the MLLE domain (green) that interacts with PAM2 motifs of TOB1/TOB2. **(C)** Model of the human nuclease module, PDB entry 7VOI (Zhang et al., 2022), in complex with BTG2-PABPC1 (RRM1-2)-poly**(A)** (PDB_Dev entry PDBDEV_00000099) (Ameerul et al., 2022). Indicated are PABPC1 (light pink); BTG2 (light green); Caf1/CNOT7 (light yellow); Ccr4/CNOT6L (yellow); CNOT1 (MIF4G domain, pale yellow) and poly**(A)** RNA (light blue). Active sites of Ccr4/CNOT6L and Caf1/CNOT7 are indicated by black dashed circles. Light blue dashed lines indicate the predicted path of poly**(A)** RNA from PABPC1 (3′ end) to the Caf1/CNOT7 active site. **(A,B)** are adapted from Figure 5, Pavanello et al. (2023), used under CC BY 4.0.

et al., 2014; Schafer et al., 2019). However, even though Pan2-Pan3 nuclease activity is stimulated by the presence of PABPC, removal of the PAM2 motif of Pan3 has moderate effect on the activity of the Pan2-Pan3 complex when short oligo(A) substrates are used *in vitro* (Wolf et al., 2014; Schafer et al., 2019). This is in marked contrast to the increased length of poly(A) tails in yeast containing inactivating point mutations in this region of Pan2 (Mangus et al., 2004). However, nuclease activity of Pan2-Pan3 is markedly increased in the presence of long poly(A) substrates containing 70 or 90 nucleotides that can accommodate two or three PABPC subunits (Schafer et al., 2019). The Pan2-Pan3 complex recognises a RNP containing a 90-mer poly(A) bound to three PABPC molecules through two main interactions (Figure 2D). First, Pan2 binds via the wider base of the WD40 domain to the RRM1 domain of the PABPC protein located at the 5′end of the 90-mer substrate. Secondly, RRM1 and RRM2 of PABPC located at the 3′end of the poly(A) tail interact with the USP and catalytic domains of Pan2, and position the 3′residues of the substrate into the active site of Pan2 thereby providing a rationale for enhanced deadenylation of poly(A) substrates containing multiple PABPC proteins (Schafer et al., 2019).

Deadenylation: the Ccr4-Not complex

The Ccr4-Not (carbon catabolite repression–negative on TATA-less) complex is the main deadenylase linked to initiation of mRNA degradation via the 5'-3' pathway. A number of regulators of mRNA stability have been shown to directly interact with the complex thereby initiating shortening of the poly(A) tail of the mRNA target and initiation of the degradation pathway (recently reviewed by Pavanello et al., 2023; Raisch and Valkov, 2022).

Ccr4-Not is a large, multi-subunit protein complex of approximately 675 kDa, with a minimum of five 'canonical' subunits in the human complex; CNOT1, CNOT2, CNOT3/5, Ccr4 and Caf1/Pop2 (Collart and Panasenko, 2012; Wahle and Winkler, 2013). The

first structural view of the Ccr4-Not complex was revealed by electron microscopy analysis of single particles of the 1 MDa *S. cerevisiae* complex, consisting of nine subunits (Nasertorabi et al., 2011). The 33 Å electron microscopy map suggests that Ccr4-Not exists in a flat 'L-shaped' configuration with two arms of similar length connected via a central hinge domain (Figure 3A) (Nasertorabi et al., 2011). At the time of writing, high resolution structural information for the complete Ccr4-Not complex has so far proven elusive (Ukleja et al., 2016; Raisch et al., 2019), most likely due to intrinsic flexibility of the complex.

The largest constituent component of human Ccr4-Not is CNOT1, a 2,376 amino acid protein that contains at least six structured domains identified to date (Figure 3B). CNOT1 serves as a molecular scaffold to provide binding sites for the other Ccr4-Not subunits (Figure 3B). Structural information is currently available for four 'modules' in Ccr4-Not: the 'N-terminal module' encompassing the NTD (N-terminal domain) and HEAT domains of CNOT1 and the CNOT10-CNOT11 heterodimer; the 'Nuclease module' consisting of the CNOT1 MIF4G domain and the catalytic subunits Caf1 and Ccr4; the 'CNOT9 module' formed by the CNOT1 DUF3819 (CN9BD) domain and CNOT9; and the 'NOT module' composed of the CNOT1 CTD (C-terminal domain) and the CNOT2-CNOT3 heterodimer (Figures 3B, C). The roles of these modules will be discussed in greater detail below. An additional MIF4G domain C-terminal to the 'CNOT9 module' has been identified but no function has been assigned. Linkers connecting the structured CNOT1 domains provide it with a degree of flexibility. Experimental depletion of CNOT1 both reduces the amount of other Ccr4-Not subunits and suppresses the formation of P-bodies, cellular aggregations of mRNA decay components (Ito et al., 2011). Liver-specific disruption of CNOT1 leads to increased mRNAs for transcription factors, cell cycle regulators and DNA damage response proteins due to reduced deadenylation, as well as aberrant gene expression associated with lethal hepatitis (Takahashi et al., 2020).

Key differences between yeast and vertebrate Ccr4-Not complexes are evidenced by their subunit compositions. Whereas the N-terminal region of *Drosophila* and human CNOT1 binds the CNOT10-CNOT11 heterodimer (Bawankar et al., 2013; Mauxion et al., 2013; Mauxion et al., 2023), the N-terminal region of Not1 in fungi binds Caf130 which is non-conserved in metazoans (Chen et al., 2001). The *Schizosaccharomyces pombe* Ccr4-Not complex uniquely includes the RNA-interacting subunit Mmi1 as a stable component (Ukleja et al., 2016). The E3 ubiquitin ligase Not4, a partner of the E2 conjugating enzyme UbcH5b (Ubc4/5 in yeast), is stably associated with Not1 in fungi (Collart and Panasenko, 2012; Bhaskar et al., 2015). In contrast, CNOT4 is not stably attached to metazoan Ccr4-Not complexes but interacts with CNOT9 via a short well-conserved C-terminal peptide motif (CBM) and with the NOT module (Keskeny et al., 2019). The N-terminal of CNOT4 is understood to inhibit the interaction of the CBM with Ccr4-Not and some structural reorganization is required to facilitate the interaction.

The N-terminal module

The 'N-terminal module' of metazoan Ccr4-Not includes CNOT10 and CNOT11 assembled around the N-terminal region of CNOT1 (CNOT1N hereafter), as evidenced by purification of CNOT10 and CNOT11 with endogenous Ccr4-Not complexes from human and *Drosophila* cells (Bawankar et al., 2013; Mauxion et al., 2013). However, both CNOT10 and CNOT11 are considered 'non-canonical' subunits. CNOT1N consists of two structured domains: the NTD and HEAT domains. The first structural evidence of CNOT1N revealed HEAT repeats that are implicated in protein-protein interactions (Basquin et al., 2012). In the metazoan complex, the HEAT repeats facilitate the binding of CNOT11, in turn providing a binding surface for CNOT10, and together forming the N-terminal module (Bawankar et al., 2013). Evidence suggests that a domain of unknown function (DUF2363) in CNOT11 is responsible for tethering the protein to CNOT10, as protein fragments containing DUF2363 can bind Not10 with equal efficiency as the full-length protein (Bawankar et al., 2013). A more recent structure of the human CNOT1-CNOT10-CNOT11 complex revealed the detailed architecture of the N-terminal module (Figure 3C) (Mauxion et al., 2023). CNOT10 and CNOT11 form a heterodimer sandwiched between two helical domains of CNOT1. CNOT10 consists of 13 tetratricopeptide repeats (TPR) stacked against each other. CNOT11 comprises of three domains: a globular helical N-terminal domain (CNOT11N), an extended middle domain (CNOT11M), and a C-terminal domain previously known as DUF2363 (CNOT11C). CNOT10 wraps around CNOT11M and packs against CNOT11N, while CNOT11C extends into the solvent and is proposed to function as an 'antenna'. The tumor suppressor/spermatogenic factor GGNBP2 was subsequently identified as an interacting partner of CNOT11C (Mauxion et al., 2023).

The nuclease module

The central 'Nuclease module' of the CNOT1 scaffold includes the MIF4G domain, which provides the binding site for the first of two nucleases within the nuclease module, named Caf1 (Daugeron et al., 2001; Jonstrup et al., 2007; Wahle and Winkler, 2013). Caf1 belongs to the RNase D family of proteins with a DEDD (Aspartate-Glutamate-Aspartate-Aspartate) active site (Figure 3C). Caf1 provides a binding platform for the second nuclease, Ccr4, which belongs to the exonuclease-endonuclease-phosphatase (EEP) family of proteins (Figure 3C) (Wang et al., 2010; Wahle and Winkler, 2013). Ccr4 consists of two domains: an N-terminal leucine-rich repeat domain (LRR) to facilitate interaction with Caf1, and a C-terminal nuclease domain. Two human paralog genes of Ccr4, CNOT6/Ccr4a and CNOT6L/Ccr4b (78% identity and 88% similarity), are mutually exclusive in the Ccr4-Not complex (Lau et al., 2009; Winkler and Balacco, 2013). Moreover, the structure of CNOT7/Caf1 by Horiuchi and colleagues determined in complex with Tob revealed the basis for the interaction with the N-terminal domain of BTG/TOB proteins (Yang et al., 2008; Horiuchi et al., 2009), which link the Ccr4-Not complex to PABPC1 and stimulate deadenylation (Ezzeddine et al., 2007).

Caf1 and Ccr4 have been shown to have non-equivalent roles in cells (Aslam et al., 2009; Mittal et al., 2011; Yi et al., 2018; Mostafa et al., 2020). A differential contribution of Caf1 and Ccr4 has also been shown from biochemical studies of the purified Ccr4-Not

complex or isolated nuclease sub-complexes even though some experiments indicate that both catalytic subunits are required for deadenylation (Maryati et al., 2015; Stowell et al., 2016; Raisch et al., 2019; Chen et al., 2021; Pekovic et al., 2023). Structures of CNOT6L/Ccr4 and CNOT7/Caf1 determined in isolation from the Ccr4-Not complex revealed the molecular basis for their Mg^{2+}dependent activities (Horiuchi et al., 2009; Wang et al., 2010). The human CNOT6L structure revealed five conserved catalytic residues: Asn195, Glu240, Asp410, Asp489 and His529. Two bound magnesium ions were identified in the active site, and deadenylase activity was abolished by an E240A mutant or by loss of Mg^{2+} (Wang et al., 2010). The human structure of CNOT7 in complex with the antiproliferative protein Tob revealed the conserved DEDD residues: Asp40, Glu42, Asp161 and Asp230. CNOT7 was shown to require divalent metal ions for activity, with higher activity in the presence of Mn^{2+} than Mg^{2+} (Horiuchi et al., 2009). Crystallographic studies of the structural homolog *S. pombe* Pop2p (Caf1p) identified two metal sites in the active site, with a preference for Mn^{2+} and Zn^{2+} over Mg^{2+} (Jonstrup et al., 2007; Andersen et al., 2009).

Structural analysis has also revealed the different modes of poly(A) recognition by the two nucleases. Ccr4 selectively recognises poly(A) residues via specific recognition of the adenine bases (Wang et al., 2010), whereas Caf1 forms multiple interactions with the phosphate-sugar backbone with no significant base interactions (Tang et al., 2019). In CNOT6L/Ccr4, a structure with poly(A) DNA identified two complete nucleotides in the deep binding cleft. Specific base interactions with Asn412 and Phe484 explain the strict preference for adenine bases (Figure 4A) (Wang et al., 2010). While no structures of Caf1 in complex with nucleotide substrates have been determined to date, a structure of the homologous Pan2 nuclease with poly(A)$_7$-RNA identified five nucleotides in the shallow binding cleft with a lack of base-specific interactions between Pan2 and adenines, suggesting Pan2 recognizes poly(A) RNA primarily through backbone interactions (Figure 4B) (Tang et al., 2019). Docking of poly(A) RNA into the Caf1 structure suggests that Caf1 recognizes poly(A) via a similar mechanism to Pan2, although poly(A) is more buried in Caf1 than in Pan2 and base-specific contacts cannot be ruled out.

Initial structural studies of the nuclease module revealed the interaction of the Ccr4 LRR domain with Caf1 but were unable to resolve the positions of the two nuclease domains relative to one another. More recent nuclease module structures have established the flexibility of the nuclease domains (Figure 3C). A human Caf1-Ccr4 (CNOT6) dimeric complex structure showed an estimated distance of 46 Å between divalent metal ions in the Caf1 and Ccr4 active sites (Chen et al., 2021), while this distance was increased to approximately 64 Å for a human CNOT1-Caf1-Ccr4 (CNOT6L) complex as evidenced by structural and electron paramagnetic resonance (EPR) spectroscopy analyses (Zhang et al., 2022). Interestingly, the active sites of Ccr4 and Caf1 are both accessible but point away when they are in complex, suggesting a spatial organisation, possibly triggered by co-factors, that would explain an apparent redundancy (Wahle and Winkler, 2013). Alternatively, allosteric regulation might facilitate the action of the two deadenylases in a cooperative fashion (Maryati et al., 2015; Pekovic et al., 2023).

The CNOT9 module

The 'CNOT9 module' is composed of the CNOT1 DUF3819 domain and CNOT9. CNOT9, also known as RQCD1 (Required for Cell Differentiation 1) or Caf40, is a canonical subunit of the Ccr4-Not complex that acts as a transcriptional co-factor in embryo development, is involved in growth control and cell differentiation, and is associated with tumorigenesis (Hiroi et al., 2002; Ajiro et al., 2009; Ajiro et al., 2010; Wong et al., 2015). The DUF3819 domain, also known as CN9BD, is located immediately C-terminal to the MIF4G domain in the mammalian complex (Figure 3C). CNOT9 features a conserved ARM domain which consists almost entirely of armadillo repeats folded into a crescent shape with a positively charged cleft (Garces et al., 2007).

CNOT9 is not catalytically active, but structural evidence has shown that it is a hotspot for protein-protein interactions. The interaction of CNOT9 and CNOT1 DUF3819 reveals W-binding pockets on the convex side that can interact with specific tryptophan residues of tristetraprolin (TTP) and GW182/TNRC6 proteins (Chen et al., 2014; Mathys et al., 2014). The armadillo repeats also provide a peptide-binding pocket on the concave side that can accommodate RNA-binding proteins such as Roquin and Bag-of-marbles (Sgromo et al., 2017; Sgromo et al., 2018), as well as the conserved CBM of the E3 ubiquitin ligase CNOT4 (Keskeny et al., 2019), thus conferring an important regulatory role.

The NOT module

The 'NOT module' of the Ccr4-Not complex located in the C-terminal CTD region of CNOT1 is a trimeric complex with CNOT2 (Not2) and CNOT3 (Not5) (Figure 3C). The CTD of CNOT1, while largely unstructured, contains a conserved CNOT1 superfamily homology (SH) domain (Boland et al., 2013). This SH domain provides a binding surface for CNOT2, itself tethering CNOT3 to the complex and forming the NOT domain (Raisch et al., 2018). CNOT2 and CNOT3 share similar structures at their C-terminus, which is responsible, within a larger region, for the heterodimer assembly (Bawankar et al., 2013; Boland et al., 2013). Both proteins feature a NOT1 anchoring region (NAR), a connector sequence (CS) and a NOT-box domain (Boland et al., 2013). CNOT2 and CNOT3 heterodimerise through the interaction of their NOT-box domains, while the CNOT2-CNOT3 heterodimer is tethered to CNOT1 via the NOT1 anchoring regions (Boland et al., 2013).

Structural information for the N-terminal region of CNOT2 is limited, but the N-terminus of CNOT3 (Not5) is known to form a highly conserved three-helix bundle (Buschauer et al., 2020). Cryo-electron microscopy analysis in *S. cerevisiae* showed that the Ccr4-Not complex is recruited to the ribosome via specific interaction of this Not5 N-terminal domain with the ribosomal E-site, with the requirement that the A-site is empty, tRNA is in the P-site, and the E3 ubiquitin ligase Not4 (CNOT4) is present (Buschauer et al., 2020). Binding of the CNOT3 N-terminal domain into the ribosomal E-site has been shown to be conserved in mammalian cells and requires the presence of CNOT4 (Absmeier et al., 2022).

A model for recruitment of Ccr4-Not to mRNA via PABPC1

The cytoplasmic poly(A)-binding protein 1 (PABPC1) is linked to the Ccr4-Not complex via the BTG/TOB family of proteins. TOB1 and TOB2 interact with the C-terminal MLLE domain of PABPC1 via their PAM2 motifs in their extended C-termini (Ezzeddine et al., 2012) (Figures 5A, B). BTG1 and BTG2, on the other hand, interact with the first RNA recognition motif (RRM1) of PABPC1 via the short Box C motif, and both BTG2 and PABPC1 RRM are sufficient to stimulate Ccr4-Not deadenylase activity (Figures 5A, B) (Stupfler et al., 2016). TOB/BTG proteins have also been shown to interact with the Ccr4-Not subunit Caf1 via BoxA and BoxB motifs (Yang et al., 2008; Horiuchi et al., 2009). A structure of PABPC1 RRM1 and RRM2 motifs in complex with poly(A)$_{11}$ shows that each RRM uses a β-sheet bearing highly conserved RNP1 and RNP2 sequence motifs to recognise poly(A) RNA, with the linkers between RRM domains forming a clamp to hold the RNA (Safaee et al., 2012).

The accumulation of available structures for the Ccr4-Not nuclease module, BTG2, and PABPC1 with poly(A) RNA have facilitated the construction of a model for the recruitment of Ccr4-Not to mRNA via TOB/BTG and PABPC1 (Ameerul et al., 2022). A combination of mutagenesis, NMR chemical shift perturbation and molecular docking facilitated a model for BTG2-PABPC1 in the absence of an experimental structure (Figure 5C). In the model, the 3′ end of the poly(A) RNA bound to PABPC1 is oriented towards the Caf1 active site, which degrades poly(A) RNA in a 3′-5′ manner. Thus, by serving as a bridge between Ccr4-Not (via Caf1) and PABPC1, BTG2 is able to stimulate deadenylation by the Ccr4-Not complex. A BTG2 variant lacking the ability to interact with PABPC1 does not inhibit cell cycle progression, indicating that binding to Ccr4-Not and PABPC1 is key for BTG2 function (Stupfler et al., 2016).

Decapping: role of the Pat1-Lsm1-7 and Dcp1-Dcp2 complexes

Binding of deadenylated RNA by the Lsm1-7-Pat1 complex

Following deadenylation, the oligoadenylated mRNA is bound by the Lsm1-7-Pat1 complex (Bouveret et al., 2000; Tharun et al., 2000; Haas et al., 2010). In *S. cerevisiae*, cells containing loss-of-function alleles of LSm1-7 display impaired mRNA degradation, and oligoadenylated mRNA species accumulate (Bouveret et al., 2000; Tharun et al., 2000). The oligoadenylated degradation intermediates are capped indicating that binding by the LSm1-7-Pat1 complex follows deadenylation mRNA and precedes decapping (Bouveret et al., 2000; Tharun et al., 2000). The LSm1-7-Pat complex contains seven small (MW ~10–20 kDa) proteins that all contain a conserved Sm motif, which consists of five anti-parallel β-strands and a single α-helix, and a single Pat1 subunit (Sharif and Conti, 2013; Montemayor et al., 2020) (Figure 6A). The LSm1-7 complex is cytoplasmic and shares several subunits with the nuclear LSm2-8 complex involved in splicing. The Lsm1-7 proteins form a β-barrel with the α-helices located on one side of the heptameric ring (Sharif and Conti, 2013) (Figure 6B). A unique feature of the Lsm1 subunit is the presence of a C-terminal extension, which forms an extended αhelix that appears to occlude-the opening in the centre of the ring (Sharif and Conti, 2013; Montemayor et al., 2020). The Pat1 protein interacts with the Lsm1-7 ring via the α-helices of the Lsm2 and Lsm3 subunits and the C-terminal domain of Pat1 (Montemayor et al., 2020).

The purified Lsm1-7 complex can bind RNA with high affinity using an interface located in the centre of the Lsm1-7 ring structure (Chowdhury et al., 2014; Montemayor et al., 2020) (Figure 6B). Deletion of the C-terminal extension of Lsm1 increases the affinity of the Lsm1-7 complex for RNA suggesting a possible regulatory role for this part of the Lsm1 protein (Chowdhury et al., 2012). Lsm1-7 binds near the 3′end of RNA and specifically recognises RNA with a short oligo(A) tail (Montemayor et al., 2020). In addition, Lsm1-7-Pat1 has a strong preference of U-rich sequences near the 3′end (Montemayor et al., 2020). Addition of the Pat1 subunit increases the affinity of the complex for oligo(A) RNA via its middle and C-terminal region (Lobel et al., 2019). While the C-terminal domain of Pat1 contains a highly basic surface area with RNA-binding activity, the molecular basis of RNA recognition by Pat1 is not clear (Braun et al., 2010).

In *S. pombe* and mammalian cells, oligoadenylated mRNAs are readily uridylated (Rissland and Norbury, 2009; Lim et al., 2014). In *S. pombe*, the Cid1 enzyme is required for this activity, while TUT4 and TUT7 have been identified as the enzymes in mammalian cells (Rissland and Norbury, 2009; Lim et al., 2014). Uridylation is enhanced on oligo(A)-tailed degradation intermediates in the absence of degradation factors Lsm1, decapping factors, or Xrn1 indicating that uridylation is required for Lsm1-7-Pat1 binding and decapping. Interestingly, degradation of histone mRNA, which do not have a poly(A) tail, requires oligouridylation for decapping and degradation (Mullen and Marzluff, 2008). In addition, C/U residues with the consensus CUCU are added to deadenylated mRNA prior to decapping in the filamentous fungus *Aspergillus nidulans* (Morozov et al., 2010) suggesting that 3′end modification of the oligoadenylated degradation intermediates is a common and conserved event.

Removal of the cap structure by the Dcp1-Dcp2 complex

After binding of the Lsm1-7-Pat1 complex, the oligoadenylated intermediate is prepared for decapping by the Dcp1-Dcp2 heterodimeric complex (Figure 7A). Recruitment of Dcp1-Dcp2 is mediated by the Lsm1-7-Pat1 complex by interactions between Pat1 and Dcp2 (Charenton et al., 2017; Lobel et al., 2019). The Dcp1 protein contains an N-terminal EVH1 domain and a divergent C-terminal region (Arribas-Layton et al., 2013). The EVH1 domain is a protein-protein interaction domain responsible for the interaction with Dcp2, the catalytically active subunit of the complex. Dcp2 contains an N-terminal regulatory domain, a Nudix (nucleotide diphosphate linked to an X moiety) hydrolase domain, that is characterised by a 23 amino acids consensus Nudix pyrophosphatase motif (GX_5EX_7RE (I/L/V)XEEXG (I/L/V)). The disordered C-terminal region contains leucine-rich helical motifs that directly interact with the C-terminal domain of Pat1

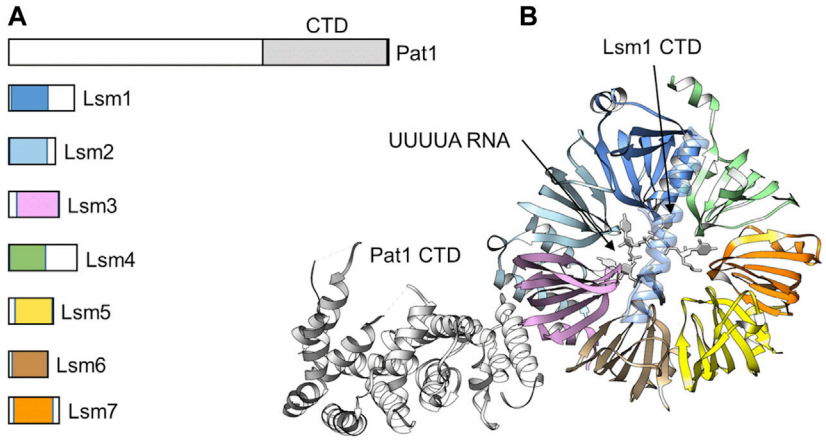


FIGURE 6
The Lsm1-7-Pat1 complex binds 3′oligoadenylated RNA. **(A)** Schematic diagram of the subunits of the Lsm1-7-Pat1 complex. **(B)** Overview of the structure of the Lsm1-7-Pat1 complex. Indicated is the β-propeller formed by Lsm1-7; the C-terminal α-helix of Lsm1; Pat1 (white) binding to Lsm2-Lsm3; and UUUA RNA (white). The model was generated by superposition of the *S. cerevisiae* Lsm1-7-Pat1 complex, PDB entry: 4C8Q (Sharif and Conti, 2013), and the *S. pombe* Lsm1-7 complex bound to UUUUA RNA, PDB entry: 6PPQ (Montemayor et al., 2020).

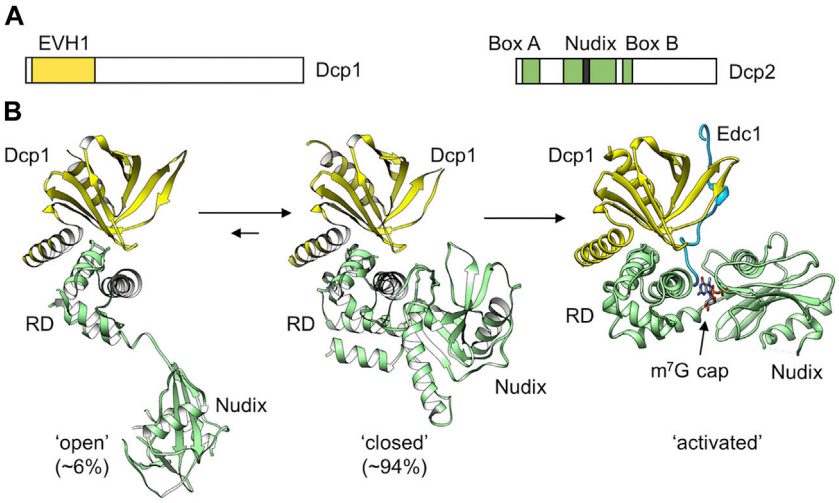


FIGURE 7
The Dcp1-Dcp2 decapping complex. **(A)** Schematic diagram of the subunits of the Dcp1 and Dcp2 proteins. Indicated are the Ena-VASP homology (EVH1) domain of Dcp1; **(B)** Structure of the Dcp1-Dcp2 complex. Indicated are the 'open' (*left*) and 'closed' (*middle*) conformations adopted by the *S. pombe* Dcp1-Dcp2 dimer in solution. PDB entry: 2KQM (She et al., 2008). Also indicated is the active conformation adopted upon binding of enhancers of decapping protein 1 and RNA (Right). PDB entry: 5N2V (Wurm et al., 2017). The percentages 'open' and 'closed' were determined by (Wurm et al., 2017).

(Charenton et al., 2017; Lobel et al., 2019). Box A region, which is part of the N-terminal regulatory domain, interacts with the EVH1 domain of Dcp1, whereas Box B region located in the Nudix domain has an intrinsic ability to bind RNA, and is required for its decapping activity *in vitro* (Piccirillo et al., 2003).

Dcp2 requires divalent metal ions (Mg^{2+} or Mn^{2+}) for its activity to remove the cap structure and release m^7GDP and RNA containing a 5′monophosphate group as the products (Wang et al., 2002; Piccirillo et al., 2003). The enzyme has high specificity for capped RNA, and is unable to bind the isolated cap structure (m^7GpppG) or unmethylated cap structures with high affinity (Piccirillo et al., 2003). The activity of Dcp2 is enhanced in the presence Dcp1 (She et al., 2008). However, full activity requires the binding of enhancers of decapping, such as Edc1-4, which are disordered proteins that bind via proline-rich sequences to the regulatory domain of Dcp1 (Charenton et al., 2016; Valkov et al., 2017; Wurm et al., 2017; Mugridge et al., 2018).

The Dcp2 catalytic subunit can adopt different conformations (She et al., 2008) (Figure 7B). A flexible hinge between the N-terminal regulatory domain and the catalytic Nudix domain allows major changes in the orientation of the two domains and provides a regulatory mechanism for mRNA decapping. In a more

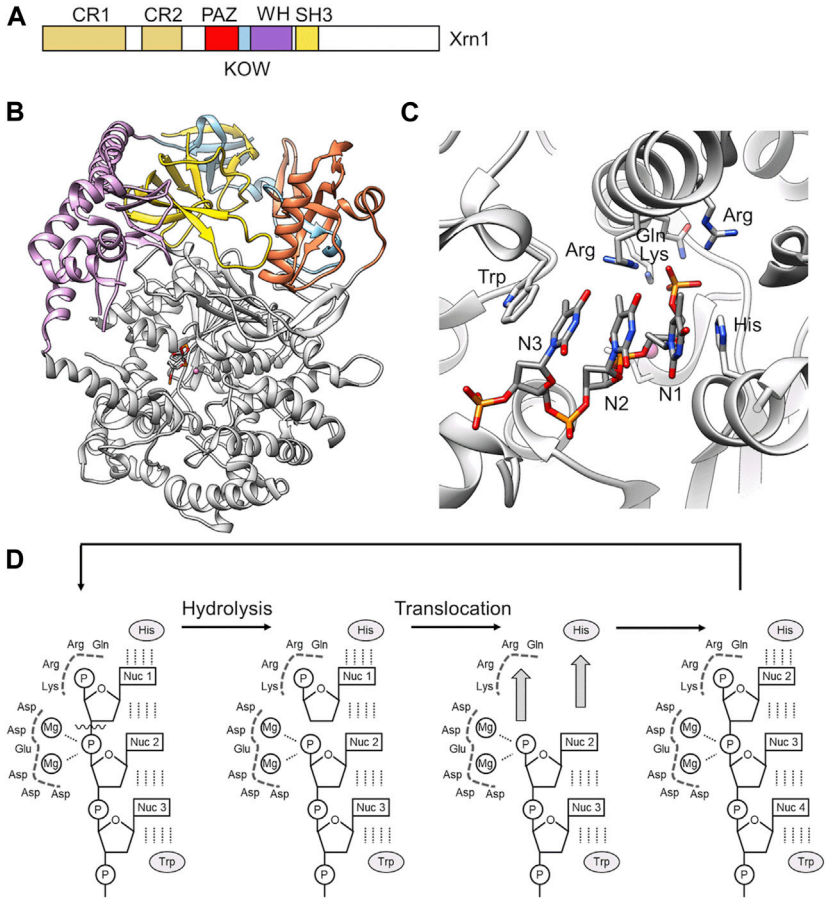


FIGURE 8
RNA degradation by the 5′-3′ exoribonuclease Xrn1. **(A)** Schematic diagram and domain organisation of the Xrn1 nuclease. Indicated are the following regions: CR1, conserved region 1; CR2 conserved region 2; PAZ, Piwi Argonaut and Zwille; KOW, Kyrpides, Ouzounis and Woese; WH, winged helix; SH3, SH3-like domain. The C-terminal region (white) is disordered. **(B)** Overview of the structure of the structured N-terminal region of *D. melanogaster* Xrn1. PDB entry: 2Y35 (Jinek et al., 2011). **(C)** The active site of Xrn1 containing a three-nucleotide DNA substrate. Indicated are residues of the basic pocket (Arg, Lys, Gln, Arg) and the His and Trp residues stacking the nucleotides in the active site. **(D)** Proposed processive mechanism of catalysis by Xrn1 (Jinek et al., 2011).

compact, closed conformation, the N-terminal regulatory domain packs close to the catalytic domain (Figure 7B). This orientation is catalytically inactive, because the cap binding site is separated from the Nudix helix that is required for catalysis (Charenton et al., 2016; Wurm et al., 2017; Mugridge et al., 2018). Moreover, the RNA binding path (Box B) on the catalytic domain is not accessible in the closed conformation. Dcp2 can also adopt an extended conformation where the regulatory domain is distant to the catalytic domain (Figure 7B). However, in the presence of Dcp1, the closed, catalytically inactive conformation is predominantly induced (Wurm et al., 2017). While binding of the enhancer of decapping Edc1 alone does not influence the conformation of the Dcp1-Dcp2 complex, a major conformational change occurs upon binding of capped RNA in the presence of Edc1. In the active conformation, Edc1 interacts with Dcp1 and both the regulatory and catalytic domains of Dcp2 (Figure 7B). The catalytic site is composed of residues from both the N-terminal regulatory and Nudix domain. For instance, the terminal methylated guanosine residue stacks with a conserved tryptophan of the regulatory domain, while 3 Mg^{2+} ions required for catalysis are coordinated by residues of the Nudix domain. In addition, an a positively charged putative RNA binding channel that includes Box B residues of Dcp2 extends from the catalytic site (Charenton et al., 2016; Wurm et al., 2017; Mugridge et al., 2018).

Degradation: the Xrn1 nuclease

In the final step, the decapped mRNA is degraded in the 5′-3′ direction by the Xrn1 nuclease generating 5′monophosphate nucleotides (Stevens, 1980; Nagarajan et al., 2013). In the absence of Xrn1, partially decapped mRNA species lacking a poly(A) tail accumulate (Hsu and Stevens, 1993). Exonucleolytic degradation is linked to decapping by direct interactions between Xrn1 and the Dcp1-Dcp2-Edc4 complex (Braun et al., 2012). Xrn1 is a highly conserved, high molecular weight, single polypeptide exoribonuclease with processive activity (Stevens, 2001) (Figure 8A). Xrn1 can also degrade DNA although at a slower rate compared to its preferred RNA substrate (Stevens, 2001). Enzyme activity requires the presence of divalent metal ions

(Stevens, 1980), which are coordinated by seven, highly conserved acidic residues (Jinek et al., 2011).

In addition to the conserved nuclease domain, which is located at the N-terminus of Xrn1, and conserved PAZ, KOW, Winged helix, and SH3-like regions in the middle part of the protein, Xrn1 contains an extensive disordered C-terminal region (Figure 8A). The crystal structures of the conserved regions of *D. melanogaster* and *K. lactis* Xrn1 provide insight into the overall organisation of the conserved domains as well as substrate recognition (Chang et al., 2011; Jinek et al., 2011). The conserved regions encompassing the catalytic domain of Xrn1 form a largely α-helical, globular conformation with the catalytic site located in its centre (Figure 8B). The conserved PAZ, KOW, Winged helix and SH3-like regions are stacked on top of the globular, catalytic domain. These regions likely contribute to the stability of the globular assembly. In budding yeast, the SH3-like domain provides an essential function of the protein, as the severe growth defect observed in Xrn1Δ cells cannot be rescued by expression of an Xrn1 variant lacking this region (Page et al., 1998). The Winged helix domain extends towards the catalytic centre and may have a role in regulating the activity of Xrn1 (Jinek et al., 2011).

A crystal structure of Dm Xrn1 lacking the disordered C-terminal region is available (Figure 8A). This model also contains a 5′phosphorylated 11-mer oligo (dT) DNA oligonucleotide and provides insight into substrate recognition. In this structure, a highly conserved Asp residue involved in the coordination of a Mg^{2+} ion was substituted with an Ala residue to prevent degradation of the substrate. In the model, Xrn1 only recognises the 5′terminal three nucleotides (Figure 8C), which is consistent with RNA protection analysis that indicate only a short chain of nucleotides are bound by the Xrn1 enzyme (Jinek et al., 2011). The backbone of the three nucleotides form electrostatic interactions with a positively charged surface area, while the nucleobases are stacked between an invariant His and Trp residue (Figure 8C). This binding mode indicates the absence of specific nucleobase interactions, and is consistent with the absence of a sequence preference by Xrn1. The 5′monophosphate group is specifically recognised in a basic pocket containing strictly conserved residues (Figure 8C). The pocket cannot accommodate larger 5′modifications thereby explaining the specific recognition of uncapped RNA. Moreover, the 5′phosphate group makes a critical contribution to substrate binding, because RNA lacking a phosphate at the 5′end are poor substrates.

Substrate binding positions the phosphate ester in close proximity of 2 Mg^{2+} ions. Based on similarities to the FEN-1 and T4 RNase H nucleases (Hwang et al., 1998; Mueser et al., 1996), the latter Mg^{2+} ion may activate a water molecule for nucleophilic attack at the scissile phosphate bond linking the first and second nucleotide (Jinek et al., 2011). Xrn1 is a highly processive enzyme with no partially degraded intermediates observed (Stevens, 2001). In addition to substrate recognition, the invariant His that forms π-π interactions with the 5′nucleotide and the basic pocket binding the 5′phosphate are also required for processivity as demonstrated by the analysis of alanine substitutions. Thus, Jinek et al. (Jinek et al., 2011) proposed a mechanism in which two key interactions drive translocation (Figure 8D). First, π-π stacking between the 5′nucleobase and the invariant His residue, and secondly, interactions between the 5′phosphate and the basic binding pocket (Figure 8D).

Coordination of events in the deadenylation-dependent 5′-3′ degradation pathway

Following initiation of deadenylation, a series of consecutive steps results in the degradation of the target mRNA by Xrn1. As discussed above, specific steps ensure the sequential recruitment of protein complexes. Deadenylation by Pan2-Pan3 and Ccr4-Not result in the specific recognition of oligoadenylated mRNA by the Lsm1-7-Pat1 complex. It is likely that oligoadenylated mRNA is uridylated. This can assist recognition by the Lsm1-7-Pat1 complex, which prefers the presence of U residues (Montemayor et al., 2020). However, the molecular basis for the coordination of uridylation and recognition of oligoadenylated intermediates by specific terminal uridylyl transferase enzymes is unclear. In addition to specific interactions that aid the consecutive completion of the decapping and degradation events (i.e., binding of Pat1 to the Dcp1-Dcp2 complex; interaction between Dcp1-Dcp2 and the Xrn1 nuclease), a number of additional interactions have been identified between the degradation factors involved in 5′-3′ degradation. These include, for example, interactions between Lsm1-7-Pat1 and Xrn1, where the same surface of Pat1 that binds helical leucine-rich motifs in Dcp2 recognise similar motifs present in the C-terminus of Xrn1 (Bouveret et al., 2000; Chowdhury et al., 2007; Charenton et al., 2017). Pat1 can also bind to the Ccr4-Not complex (Haas et al., 2010). Moreover, the C-terminal unstructured region of Xrn1 can also interact with components of the Ccr4-Not complex (Chang et al., 2019).

In addition to multiple interactions between the core components of the 5′-3′ degradation pathway, they can also interact with other factors that module mRNA degradation. For example, the MIF4G domain of CNOT1 binds DDX6 (Dhh1/RCK/p54). This protein is an RNA helicase involved in miRNA regulation, which also activates the decapping pathway and represses translation (Chen et al., 2014; Mathys et al., 2014). Thus, multiple transient, low-affinity interactions between components of the 5′-3′ degradation pathway may result in self-organisation of factors involved in RNA degradation. The resulting local enrichment of degradation factors in the cytoplasm of eukaryotic cells may result in the formation of cytoplasmic foci known as processing bodies (P-bodies) (Van Dijk et al., 2002; Sheth and Parker, 2003; Parker and Sheth, 2007).

Concluding remarks

In recent years, a large body of work has established critical steps in the 5′-3′ degradation pathway, including the molecular basis of the catalytic steps required for deadenylation by Pan2-Pan3, decapping, and degradation of the RNA body by Xrn1. In addition, many interactions between the molecular machines involved are understood at the molecular level. Despite immense progress, however, there are still areas that are poorly understood. For example, the requirement for the Caf1 and Ccr4 catalytic subunits during deadenylation by Ccr4-Not, and their collaborative or unique roles are not clear. In addition, the role of uridylation is not understood in detail, and the molecular basis for selective uridylation of oligoadenylated degradation intermediates is

not known. A third area for future investigations is to decipher in molecular detail how events in the 5′-3′ mRNA degradation pathway are coordinated with other RNA degradation pathways and the regulation of translational efficiency.

Author contributions

QZ and LP wrote sections of the manuscript. MB and GW wrote sections of the manuscript, and edited the final version. All authors contributed to the article and approved the submitted version.

Funding

This work was in part supported by a Vice-Chancellor's Scholarship for Research Excellence of the University of Nottingham (LP). MB was supported by the National Natural Science Foundation of China (NSFC) (grant 31870053) and the Guangdong Province Basic and Applied Research Foundation (grant 2023A1515011818).

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

Absmeier, E., Chandrasekaran, V., O'reilly, F. J., Stowell, J. A., Rappsilber, J., and Passmore, L. A. (2022). Specific recognition and ubiquitination of slow-moving ribosomes by human CCR4-NOT. *Nat. Struct. Mol. Biol.* 30, 1314–1322. doi:10.1038/s41594-023-01075-8

Ajiro, M., Katagiri, T., Ueda, K., Nakagawa, H., Fukukawa, C., Lin, M. L., et al. (2009). Involvement of RQCD1 overexpression, a novel cancer-testis antigen, in the Akt pathway in breast cancer cells. *Int. J. Oncol.* 35, 673–681. doi:10.3892/ijo_00000379

Ajiro, M., Nishidate, T., Katagiri, T., and Nakamura, Y. (2010). Critical involvement of RQCD1 in the EGFR-Akt pathway in mammary carcinogenesis. *Int. J. Oncol.* 37, 1085–1093. doi:10.3892/ijo_00000760

Almasmoum, H. A., Airhihen, B., Seedhouse, C., and Winkler, G. S. (2021). Frequent loss of BTG1 activity and impaired interactions with the Caf1 subunit of the Ccr4-Not deadenylase in non-Hodgkin lymphoma. *Leuk. Lymphoma* 62, 281–290. doi:10.1080/10428194.2020.1827243

Ameerul, A., Almasmoum, H., Pavanello, L., Dominguez, C., and Winkler, G. S. (2022). Structural model of the human BTG2-PABPC1 complex by combining mutagenesis, NMR chemical shift perturbation data and molecular docking. *J. Mol. Biol.* 434, 167662. doi:10.1016/j.jmb.2022.167662

Andersen, K. R., Jonstrup, A. T., Van, L. B., and Brodersen, D. E. (2009). The activity and selectivity of fission yeast Pop2p are affected by a high affinity for Zn^{2+} and Mn^{2+} in the active site. *RNA* 15, 850–861. doi:10.1261/rna.1489409

Arribas-Layton, M., Wu, D., Lykke-Andersen, J., and Song, H. (2013). Structural and functional control of the eukaryotic mRNA decapping machinery. *Biochim. Biophys. Acta* 1829, 580–589. doi:10.1016/j.bbagrm.2012.12.006

Aslam, A., Mittal, S., Koch, F., Andrau, J. C., and Winkler, G. S. (2009). The ccr4-not deadenylase subunits CNOT7 and CNOT8 have overlapping roles and modulate cell proliferation. *Mol. Biol. Cell.* 20, 3840–3850. doi:10.1091/mbc.e09-02-0146

Basquin, J., Roudko, V. V., Rode, M., Basquin, C., Seraphin, B., and Conti, E. (2012). Architecture of the nuclease module of the yeast ccr4-not complex: the not1-caf1-ccr4 interaction. *Mol. Cell.* 48, 207–218. doi:10.1016/j.molcel.2012.08.014

Bawankar, P., Loh, B., Wohlbold, L., Schmidt, S., and Izaurralde, E. (2013). NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. *RNA Biol.* 10, 228–244. doi:10.4161/rna.23018

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes. Dev.* 20, 1885–1898. doi:10.1101/gad.1424106

Bhaskar, V., Basquin, J., and Conti, E. (2015). Architecture of the ubiquitylation module of the yeast Ccr4-Not complex. *Structure* 23, 921–928. doi:10.1016/j.str.2015.03.011

Boeck, R., Tarun, S., Jr., Rieger, M., Deardorff, J. A., Muller-Auer, S., and Sachs, A. B. (1996). The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *J. Biol. Chem.* 271, 432–438. doi:10.1074/jbc.271.1.432

Boland, A., Chen, Y., Raisch, T., Jonas, S., Kuzuoglu-Ozturk, D., Wohlbold, L., et al. (2013). Structure and assembly of the NOT module of the human CCR4-NOT complex. *Nat. Struct. Mol. Biol.* 20, 1289–1297. doi:10.1038/nsmb.2681

Bonisch, C., Temme, C., Moritz, B., and Wahle, E. (2007). Degradation of hsp70 and other mRNAs in Drosophila via the 5' 3' pathway and its regulation by heat shock. *J. Biol. Chem.* 282, 21818–21828. doi:10.1074/jbc.M702998200

Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Séraphin, B. (2000). A Sm-like protein complex that participates in mRNA degradation. *Embo J.* 19, 1661–1671. doi:10.1093/emboj/19.7.1661

Braun, J. E., Tritschler, F., Haas, G., Igreja, C., Truffault, V., Weichenrieder, O., et al. (2010). The C-terminal alpha-alpha superhelix of Pat is required for mRNA decapping in metazoa. *Embo J.* 29, 2368–2380. doi:10.1038/emboj.2010.124

Braun, J. E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol. Cell.* 44, 120–133. doi:10.1016/j.molcel.2011.09.007

Braun, J. E., Truffault, V., Boland, A., Huntzinger, E., Chang, C. T., Haas, G., et al. (2012). A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. *Nat. Struct. Mol. Biol.* 19, 1324–1331. doi:10.1038/nsmb.2413

Brown, C. E., Tarun, S. Z., Jr., Boeck, R., and Sachs, A. B. (1996). PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 16, 5744–5753. doi:10.1128/mcb.16.10.5744

Buschauer, R., Matsuo, Y., Sugiyama, T., Chen, Y. H., Alhusaini, N., Sweet, T., et al. (2020). The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science* 368, eaay6912. doi:10.1126/science.aay6912

Chang, J. H., Xiang, S., Xiang, K., Manley, J. L., and Tong, L. (2011). Structural and biochemical studies of the 5'→3' exoribonuclease Xrn1. *Nat. Struct. Mol. Biol.* 18, 270–276. doi:10.1038/nsmb.1984

Chang, C. T., Muthukumar, S., Weber, R., Levdansky, Y., Chen, Y., Bhandari, D., et al. (2019). A low-complexity region in human XRN1 directly recruits deadenylation and decapping factors in 5'-3' messenger RNA decay. *Nucleic Acids Res.* 47, 9282–9295. doi:10.1093/nar/gkz633

Charenton, C., Taverniti, V., Gaudon-Plesse, C., Back, R., Séraphin, B., and Graille, M. (2016). Structure of the active form of Dcp1-Dcp2 decapping enzyme bound to m(7)GDP and its Edc3 activator. *Nat. Struct. Mol. Biol.* 23, 982–986. doi:10.1038/nsmb.3300

Charenton, C., Gaudon-Plesse, C., Fourati, Z., Taverniti, V., Back, R., Kolesnikova, O., et al. (2017). A unique surface on Pat1 C-terminal domain directly interacts with Dcp2 decapping enzyme and Xrn1 5'-3' mRNA exonuclease in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 114, E9493–e9501. doi:10.1073/pnas.1711680114

Chen, J., Rappsilber, J., Chiang, Y. C., Russell, P., Mann, M., and Denis, C. L. (2001). Purification and characterization of the 1.0 MDa CCR4-NOT complex identifies two novel components of the complex. *J. Mol. Biol.* 314, 683–694. doi:10.1006/jmbi.2001.5162

Chen, Y., Boland, A., Kuzuoglu-Ozturk, D., Bawankar, P., Loh, B., Chang, C. T., et al. (2014). A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. *Mol. Cell.* 54, 737–750. doi:10.1016/j.molcel.2014.03.034

Chen, Y., Khazina, E., Izaurralde, E., and Weichenrieder, O. (2021). Crystal structure and functional properties of the human CCR4-CAF1 deadenylase complex. *Nucleic Acids Res.* 49, 6489–6510. doi:10.1093/nar/gkab414

Chicoine, J., Benoit, P., Gamberi, C., Paliouras, M., Simonelig, M., and Lasko, P. (2007). Bicaudal-C recruits CCR4-NOT deadenylase to target mRNAs and regulates oogenesis, cytoskeletal organization, and its own expression. *Dev. Cell.* 13, 691–704. doi:10.1016/j.devcel.2007.10.002

Chowdhury, A., Mukhopadhyay, J., and Tharun, S. (2007). The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* 13, 998–1016. doi:10.1261/rna.502507

Chowdhury, A., Raju, K. K., Kalurupalle, S., and Tharun, S. (2012). Both Sm-domain and C-terminal extension of Lsm1 are important for the RNA-binding activity of the Lsm1-7-Pat1 complex. *Rna* 18, 936–944. doi:10.1261/rna.029876.111

Chowdhury, A., Kalurupalle, S., and Tharun, S. (2014). Pat1 contributes to the RNA binding activity of the Lsm1-7-Pat1 complex. *RNA* 20, 1465–1475. doi:10.1261/rna.045252.114

Christie, M., Boland, A., Huntzinger, E., Weichenrieder, O., and Izaurralde, E. (2013). Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. *Mol. Cell.* 51, 360–373. doi:10.1016/j.molcel.2013.07.011

Collart, M. A., and Panasenko, O. O. (2012). The Ccr4-not complex. *Gene* 492, 42–53. doi:10.1016/j.gene.2011.09.033

Daugeron, M.-C., Mauxion, F., and Seraphin, B. (2001). The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. *Nucleic Acids Res.* 29, 2448–2455. doi:10.1093/nar/29.12.2448

Ezzeddine, N., Chang, T. C., Zhu, W., Yamashita, A., Chen, C. Y., Zhong, Z., et al. (2007). Human TOB, an antiproliferative transcription factor, is a poly(A)-binding protein-dependent positive regulator of cytoplasmic mRNA deadenylation. *Mol. Cell. Biol.* 27, 7791–7801. doi:10.1128/MCB.01254-07

Ezzeddine, N., Chen, C. Y., and Shyu, A. B. (2012). Evidence providing new insights into TOB-promoted deadenylation and supporting a link between TOB's deadenylation-enhancing and antiproliferative activities. *Mol. Cell. Biol.* 32, 1089–1098. doi:10.1128/MCB.06370-11

Fabian, M. R., Cieplak, M. K., Frank, F., Morita, M., Green, J., Srikumar, T., et al. (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat. Struct. Mol. Biol.* 18, 1211–1217. doi:10.1038/nsmb.2149

Fabian, M. R., Frank, F., Rouya, C., Siddiqui, N., Lai, W. S., Karetnikov, A., et al. (2013). Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. *Nat. Struct. Mol. Biol.* 20, 735–739. doi:10.1038/nsmb.2572

Garces, R. G., Gillon, W., and Pai, E. F. (2007). Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with *in vitro* nucleic acid binding properties. *Protein science. a Publ. Protein Soc.* 16, 176–188. doi:10.1110/ps.062600507

Haas, G., Braun, J. E., Igreja, C., Tritschler, F., Nishihara, T., and Izaurralde, E. (2010). HPat provides a link between deadenylation and decapping in metazoa. *J. Cell. Biol.* 189, 289–302. doi:10.1083/jcb.200910141

Hiroi, N., Ito, T., Yamamoto, H., Ochiya, T., Jinno, S., and Okayama, H. (2002). Mammalian Rcd1 is a novel transcriptional cofactor that mediates retinoic acid-induced cell differentiation. *EMBO J.* 21, 5235–5244. doi:10.1093/emboj/cdf521

Horiuchi, M., Takeuchi, K., Noda, N., Muroya, N., Suzuki, T., Nakamura, T., et al. (2009). Structural basis for the antiproliferative activity of the Tob-hCaf1 complex. *J. Biol. Chem.* 284, 13244–13255. doi:10.1074/jbc.M809250200

Hsu, C. L., and Stevens, A. (1993). Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol. Cell. Biol.* 13, 4826–4835. doi:10.1128/mcb.13.8.4826

Ito, K., Takahashi, A., Morita, M., Suzuki, T., and Yamamoto, T. (2011). The role of the CNOT1 subunit of the CCR4-NOT complex in mRNA deadenylation and cell viability. *Protein Cell.* 2, 755–763. doi:10.1007/s13238-011-1092-4

Jinek, M., Coyle, S. M., and Doudna, J. A. (2011). Coupled 5' nucleotide recognition and processivity in Xrn1-mediated mRNA decay. *Mol. Cell.* 41, 600–608. doi:10.1016/j.molcel.2011.02.004

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* 16, 421–433. doi:10.1038/nrg3965

Jonas, S., Christie, M., Peter, D., Bhandari, D., Loh, B., Huntzinger, E., et al. (2014). An asymmetric PAN3 dimer recruits a single PAN2 exonuclease to mediate mRNA deadenylation and decay. *Nat. Struct. Mol. Biol.* 21, 599–608. doi:10.1038/nsmb.2837

Jonstrup, A. T., Andersen, K. R., Van, L. B., and Brodersen, D. E. (2007). The 1.4-A crystal structure of the *S. pombe* Pop2p deadenylase subunit unveils the configuration of an active enzyme. *Nucleic Acids Res.* 35, 3153–3164. doi:10.1093/nar/gkm178

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

Keskeny, C., Raisch, T., Sgromo, A., Igreja, C., Bhandari, D., Weichenrieder, O., et al. (2019). A conserved CAF40-binding motif in metazoan NOT4 mediates association with the CCR4-NOT complex. *Genes. Dev.* 33, 236–252. doi:10.1101/gad.320952.118

Lau, N. C., Kolkman, A., Schaik, V., Mulder, K. W., Pijnappel, W. W., Heck, A. J., et al. (2009). Human Ccr4-Not complexes contain variable deadenylase subunits. *Biochem. J.* 422, 443–453. doi:10.1042/BJ20090500

Lim, J., Ha, M., Chang, H., Kwon, S. C., Simanshu, D. K., Patel, D. J., et al. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell.* 159, 1365–1376. doi:10.1016/j.cell.2014.10.055

Lobel, J. H., Tibble, R. W., and Gross, J. D. (2019). Pat1 activates late steps in mRNA decay by multiple mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* 116, 23512–23517. doi:10.1073/pnas.1905455116

Mangus, D. A., Evans, M. C., Agrin, N. S., Smith, M., Gongidi, P., and Jacobson, A. (2004). Positive and negative regulation of poly(A) nuclease. *Mol. Cell. Biol.* 24, 5521–5533. doi:10.1128/MCB.24.12.5521-5533.2004

Maryati, M., Airhihen, B., and Winkler, G. S. (2015). The enzyme activities of Caf1 and Ccr4 are both required for deadenylation by the human Ccr4-Not nuclease module. *Biochem. J.* 469, 169–176. doi:10.1042/BJ20150304

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., et al. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol. Cell.* 54, 751–765. doi:10.1016/j.molcel.2014.03.036

Mauxion, F., Chen, C. Y., Seraphin, B., and Shyu, A. B. (2009). BTG/TOB factors impact deadenylases. *Trends Biochem. Sci.* 34, 640–647. doi:10.1016/j.tibs.2009.07.008

Mauxion, F., Preve, B., and Seraphin, B. (2013). C2ORF29/CNOT11 and CNOT10 form a new module of the CCR4-NOT complex. *RNA Biol.* 10, 267–276. doi:10.4161/rna.23065

Mauxion, F., Basquin, J., Ozgur, S., Rame, M., Albrecht, J., Schafer, I., et al. (2023). The human CNOT1-CNOT10-CNOT11 complex forms a structural platform for protein-protein interactions. *Cell. Rep.* 42, 111902. doi:10.1016/j.celrep.2022.111902

Mittal, S., Aslam, A., Doidge, R., Medica, R., and Winkler, G. S. (2011). The Ccr4a (CNOT6) and Ccr4b (CNOT6L) deadenylase subunits of the human Ccr4-Not complex contribute to the prevention of cell death and senescence. *Mol. Biol. Cell.* 22, 748–758. doi:10.1091/mbc.E10-11-0898

Montemayor, E. J., Virta, J. M., Hayes, S. M., Nomura, Y., Brow, D. A., and Butcher, S. E. (2020). Molecular basis for the distinct cellular functions of the Lsm1-7 and Lsm2-8 complexes. *Rna* 26, 1400–1413. doi:10.1261/rna.075879.120

Morozov, I. Y., Jones, M. G., Spiller, D. G., Rigden, D. J., Dattenbock, C., Novotny, R., et al. (2010). Distinct roles for Caf1, Ccr4, Edc3 and CutA in the co-ordination of transcript deadenylation, decapping and P-body formation in Aspergillus nidulans. *Mol. Microbiol.* 76, 503–516. doi:10.1111/j.1365-2958.2010.07118.x

Mostafa, D., Takahashi, A., Yanagiya, A., Yamaguchi, T., Abe, T., Kureha, T., et al. (2020). Essential functions of the CNOT7/8 catalytic subunits of the CCR4-NOT complex in mRNA regulation and cell viability. *RNA Biol.* 17, 403–416. doi:10.1080/15476286.2019.1709747

Mugridge, J. S., Tibble, R. W., Ziemniak, M., Jemielity, J., and Gross, J. D. (2018). Structure of the activated Edc1-Dcp1-Dcp2-Edc3 mRNA decapping complex with substrate analog poised for catalysis. *Nat. Commun.* 9, 1152. doi:10.1038/s41467-018-03536-x

Muhlrad, D., Decker, C. J., and Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. *Genes. Dev.* 8, 855–866. doi:10.1101/gad.8.7.855

Mullen, T. E., and Marzluff, W. F. (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes. Dev.* 22, 50–65. doi:10.1101/gad.1622708

Nagarajan, V. K., Jones, C. I., Newbury, S. F., and Green, P. J. (2013). XRN 5'-->3' exoribonucleases: structure, mechanisms and functions. *Biochim. Biophys. Acta* 1829, 590–603. doi:10.1016/j.bbagrm.2013.03.005

Nasertorabi, F., Batisse, C., Diepholz, M., Suck, D., and Bottcher, B. (2011). Insights into the structure of the CCR4-NOT complex by electron microscopy. *FEBS Lett.* 585, 2182–2186. doi:10.1016/j.febslet.2011.05.071

Page, A. M., Davis, K., Molineux, C., Kolodner, R. D., and Johnson, A. W. (1998). Mutational analysis of exoribonuclease I from *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 26, 3707–3716. doi:10.1093/nar/26.16.3707

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. *Mol. Cell.* 25, 635–646. doi:10.1016/j.molcel.2007.02.011

Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.* 11, 121–127. doi:10.1038/nsmb724

Passmore, L. A., and Coller, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat. Rev. Mol. Cell. Biol.* 23, 93–106. doi:10.1038/s41580-021-00417-y

Pavanello, L., Hall, B., Airhihen, B., and Winkler, G. S. (2018). The central region of CNOT1 and CNOT9 stimulates deadenylation by the Ccr4-Not nuclease module. *Biochem. J.* 475, 3437–3450. doi:10.1042/BCJ20180456

Pavanello, L., Hall, M., and Winkler, G. S. (2023). Regulation of eukaryotic mRNA deadenylation and degradation by the Ccr4-Not complex. *Front. Cell. Dev. Biol.* 11, 1153624. doi:10.3389/fcell.2023.1153624

Pekovic, F., Rammelt, C., Kubikova, J., Metz, J., Jeske, M., and Wahle, E. (2023). RNA binding proteins Smaug and Cup induce CCR4-NOT-dependent deadenylation of the nanos mRNA in a reconstituted system. *Nucleic Acids Res.* 51, 3950–3970. doi:10.1093/nar/gkad159

Piao, X., Zhang, X., Wu, L., and Belasco, J. G. (2010). CCR4-NOT deadenylates mRNA associated with RNA-induced silencing complexes in human cells. *Mol. Cell. Biol.* 30, 1486–1494. doi:10.1128/MCB.01481-09

Piccirillo, C., Khanna, R., and Kiledjian, M. (2003). Functional characterization of the mammalian mRNA decapping enzyme hDcp2. *Rna* 9, 1138–1147. doi:10.1261/rna.5690503

Quesada, V., Diaz-Perales, A., Gutierrez-Fernandez, A., Garabaya, C., Cal, S., and Lopez-Otin, C. (2004). Cloning and enzymatic analysis of 22 novel human ubiquitin-specific proteases. *Biochem. Biophys. Res. Commun.* 314, 54–62. doi:10.1016/j.bbrc.2003.12.050

Raisch, T., and Valkov, E. (2022). Regulation of the multisubunit CCR4-NOT deadenylase in the initiation of mRNA degradation. *Curr. Opin. Struct. Biol.* 77, 102460. doi:10.1016/j.sbi.2022.102460

Raisch, T., Sandmeir, F., Weichenrieder, O., Valkov, E., and Izaurralde, E. (2018). Structural and biochemical analysis of a NOT1 MIF4G-like domain of the CCR4-NOT complex. *J. Struct. Biol.* 204, 388–395. doi:10.1016/j.jsb.2018.10.009

Raisch, T., Chang, C. T., Levdansky, Y., Muthukumar, S., Raunser, S., and Valkov, E. (2019). Reconstitution of recombinant human CCR4-NOT reveals molecular insights into regulated deadenylation. *Nat. Commun.* 10, 3173. doi:10.1038/s41467-019-11094-z

Rissland, O. S., and Norbury, C. J. (2009). Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. *Nat. Struct. Mol. Biol.* 16, 616–623. doi:10.1038/nsmb.1601

Safaee, N., Kozlov, G., Noronha, A. M., Xie, J., Wilds, C. J., and Gehring, K. (2012). Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol. Cell.* 48, 375–386. doi:10.1016/j.molcel.2012.09.001

Schafer, I. B., Rode, M., Bonneau, F., Schussler, S., and Conti, E. (2014). The structure of the Pan2-Pan3 core complex reveals cross-talk between deadenylase and pseudokinase. *Nat. Struct. Mol. Biol.* 21, 591–598. doi:10.1038/nsmb.2834

Schafer, I. B., Yamashita, M., Schuller, J. M., Schussler, S., Reichelt, P., Strauss, M., et al. (2019). Molecular basis for poly(A) RNP architecture and recognition by the Pan2-Pan3 deadenylase. *Cell.* 177, 1619–1631. e21. doi:10.1016/j.cell.2019.04.013

Scott, D. D., and Norbury, C. J. (2013). RNA decay via 3' uridylation. *Biochim. Biophys. Acta* 1829, 654–665. doi:10.1016/j.bbagrm.2013.01.009

Sgromo, A., Raisch, T., Bawankar, P., Bhandari, D., Chen, Y., Kuzuoglu-Ozturk, D., et al. (2017). A CAF40-binding motif facilitates recruitment of the CCR4-NOT complex to mRNAs targeted by Drosophila Roquin. *Nat. Commun.* 8, 14307. doi:10.1038/ncomms14307

Sgromo, A., Raisch, T., Backhaus, C., Keskeny, C., Alva, V., Weichenrieder, O., et al. (2018). Drosophila Bag-of-marbles directly interacts with the CAF40 subunit of the CCR4-NOT complex to elicit repression of mRNA targets. *RNA* 24, 381–395. doi:10.1261/rna.064584.117

Sharif, H., and Conti, E. (2013). Architecture of the lsm1-7-pat1 complex: A conserved assembly in eukaryotic mRNA turnover. *Cell. Rep.* 5, 283–291. doi:10.1016/j.celrep.2013.10.004

She, M., Decker, C. J., Svergun, D. I., Round, A., Chen, N., Muhlrad, D., et al. (2008). Structural basis of dcp2 recognition and activation by dcp1. *Mol. Cell.* 29, 337–349. doi:10.1016/j.molcel.2008.01.002

Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* 300, 805–808. doi:10.1126/science.1082320

Siddiqui, N., Mangus, D. A., Chang, T. C., Palermino, J. M., Shyu, A. B., and Gehring, K. (2007). Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. *J. Biol. Chem.* 282, 25067–25075. doi:10.1074/jbc.M701256200

Stevens, A. (1980). Purification and characterization of a *Saccharomyces cerevisiae* exoribonuclease which yields 5'-mononucleotides by a 5' leads to 3' mode of hydrolysis. *J. Biol. Chem.* 255, 3080–3085. doi:10.1016/s0021-9258(19)85855-6

Stevens, A. (2001). 5'-exoribonuclease 1: xrn1. *Methods Enzymol.* 342, 251–259. doi:10.1016/s0076-6879(01)42549-3

Stoecklin, G., Mayo, T., and Anderson, P. (2006). ARE-mRNA degradation requires the 5'-3' decay pathway. *EMBO Rep.* 7, 72–77. doi:10.1038/sj.embor.7400572

Stowell, J. A. W., Webster, M. W., Kogel, A., Wolf, J., Shelley, K. L., and Passmore, L. A. (2016). Reconstitution of targeted deadenylation by the ccr4-not complex and the YTH domain protein Mmi1. *Cell. Rep.* 17, 1978–1989. doi:10.1016/j.celrep.2016.10.066

Stupfler, B., Birck, C., Seraphin, B., and Mauxion, F. (2016). BTG2 bridges PABPC1 RNA-binding domains and CAF1 deadenylase to control cell proliferation. *Nat. Commun.* 7, 10811. doi:10.1038/ncomms10811

Takahashi, A., Suzuki, T., Soeda, S., Takaoka, S., Kobori, S., Yamaguchi, T., et al. (2020). The CCR4-NOT complex maintains liver homeostasis through mRNA deadenylation. *Life Sci. Alliance* 3, e201900494. doi:10.26508/lsa.201900494

Tang, T. T. L., Stowell, J., Hill, C. H., and Passmore, L. A. (2019). The intrinsic structure of poly(A) RNA determines the specificity of Pan2 and Caf1 deadenylases. *Nat. Struct. Mol. Biol.* 26, 433–442. doi:10.1038/s41594-019-0227-9

Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D., and Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* 404, 515–518. doi:10.1038/35006676

Tucker, M., Valencia-Sanchez, M. A., Staples, R. R., Chen, J., Denis, C. L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell.* 104, 377–386. doi:10.1016/s0092-8674(01)00225-2

Uchida, N., Hoshino, S., and Katada, T. (2004). Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. *J. Biol. Chem.* 279, 1383–1391. doi:10.1074/jbc.M309125200

Ukleja, M., Cuellar, J., Siwaszek, A., Kasprzak, J. M., Czarnocki-Cieciura, M., Bujnicki, J. M., et al. (2016). The architecture of the *Schizosaccharomyces pombe* CCR4-NOT complex. *Nat. Commun.* 7, 10433. doi:10.1038/ncomms10433

Valkov, E., Jonas, S., and Weichenrieder, O. (2017). Mille viae in eukaryotic mRNA decapping. *Curr. Opin. Struct. Biol.* 47, 40–51. doi:10.1016/j.sbi.2017.05.009

Van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E., and Séraphin, B. (2002). Human Dcp2: A catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *Embo J.* 21, 6915–6924. doi:10.1093/emboj/cdf678

Vicens, Q., Kieft, J. S., and Rissland, O. S. (2018). Revisiting the closed-loop model and the nature of mRNA 5'-3' communication. *Mol. Cell.* 72, 805–812. doi:10.1016/j.molcel.2018.10.047

Wahle, E., and Winkler, G. S. (2013). RNA decay machines: deadenylation by the ccr4-not and Pan2-Pan3 complexes. *Biochim. Biophys. Acta* 1829, 561–570. doi:10.1016/j.bbagrm.2013.01.003

Wang, Z., Jiao, X., Carr-Schmid, A., and Kiledjian, M. (2002). The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12663–12668. doi:10.1073/pnas.192445599

Wang, H., Morita, M., Yang, X., Suzuki, T., Yang, W., Wang, J., et al. (2010). Crystal structure of the human CNOT6L nuclease domain reveals strict poly(A) substrate specificity. *EMBO J.* 29, 2566–2576. doi:10.1038/emboj.2010.152

Wilusz, C. J., Wormington, M., and Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell. Biol.* 2, 237–246. doi:10.1038/35067025

Winkler, G. S., and Balacco, D. L. (2013). Heterogeneity and complexity within the nuclease module of the Ccr4-Not complex. *Front. Genet.* 4, 296. doi:10.3389/fgene.2013.00296

Winkler, G. S. (2010). The mammalian anti-proliferative BTG/Tob protein family. *J. Cell. Physiol.* 222, 66–72. doi:10.1002/jcp.21919

Wolf, J., Valkov, E., Allen, M. D., Meineke, B., Gordiyenko, Y., Mclaughlin, S. H., et al. (2014). Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. *EMBO J.* 33, 1514–1526. doi:10.15252/embj.201488373

Wong, S. Q., Behren, A., Mar, V. J., Woods, K., Li, J., Martin, C., et al. (2015). Whole exome sequencing identifies a recurrent RQCD1 P131L mutation in cutaneous melanoma. *Oncotarget* 6, 1115–1127. doi:10.18632/oncotarget.2747

Wurm, J. P., Holdermann, I., Overbeck, J. H., Mayer, P. H. O., and Sprangers, R. (2017). Changes in conformational equilibria regulate the activity of the Dcp2 decapping enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 114, 6034–6039. doi:10.1073/pnas.1704496114

Yamashita, A., Chang, T. C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C. Y., et al. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat. Struct. Mol. Biol.* 12, 1054–1063. doi:10.1038/nsmb1016

Yang, X., Morita, M., Wang, H., Suzuki, T., Yang, W., Luo, Y., et al. (2008). Crystal structures of human BTG2 and mouse TIS21 involved in suppression of CAF1 deadenylase activity. *Nucleic Acids Res.* 36, 6872–6881. doi:10.1093/nar/gkn825

Yi, H., Park, J., Ha, M., Lim, J., Chang, H., and Kim, V. N. (2018). PABP cooperates with the CCR4-NOT complex to promote mRNA deadenylation and block precocious decay. *Mol. Cell.* 70, 1081–1088. e5. doi:10.1016/j.molcel.2018.05.009

Yuniati, L., Scheijen, B., Van Der Meer, L. T., and Van Leeuwen, F. N. (2018). Tumor suppressors BTG1 and BTG2: beyond growth control. *J. Cell. Physiol.* 234, 5379–5389. doi:10.1002/jcp.27407

Zaessinger, S., Busseau, I., and Simonelig, M. (2006). Oskar allows nanos mRNA translation in Drosophila embryos by preventing its deadenylation by Smaug/CCR4. *Development* 133, 4573–4583. doi:10.1242/dev.02649

Zhang, Q., Pavanello, L., Potapov, A., Bartlam, M., and Winkler, G. S. (2022). Structure of the human Ccr4-Not nuclease module using X-ray crystallography and electron paramagnetic resonance spectroscopy distance measurements. *Protein Sci.* 31, 758–764. doi:10.1002/pro.4262

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 23 November 2023
DOI 10.3389/fgene.2023.1281538

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Kaushlendra Tripathi,
University of Alabama at Birmingham, United States
Gaetano Caramori,
University of Messina, Italy

*CORRESPONDENCE
Marc de Perrot,
marc.deperrot@uhn.ca

†These authors share first authorship

RECEIVED 22 August 2023
ACCEPTED 06 November 2023
PUBLISHED 23 November 2023

CITATION
Man HSJ, Moosa VA, Singh A, Wu L, Granton JT, Juvet SC, Hoang CD and de Perrot M (2023), Unlocking the potential of RNA-based therapeutics in the lung: current status and future directions.
Front. Genet. 14:1281538.
doi: 10.3389/fgene.2023.1281538

Unlocking the potential of RNA-based therapeutics in the lung: current status and future directions

H. S. Jeffrey Man[1,2,3,4†], Vaneeza A. Moosa[3,5†], Anand Singh[6], Licun Wu[3,5], John T. Granton[4], Stephen C. Juvet[1,2,3,4], Chuong D. Hoang[6] and Marc de Perrot[2,3,5*]

[1]Temerty Faculty of Medicine, Institute of Medical Science, Toronto, ON, Canada, [2]Department of Immunology, University of Toronto, Toronto, ON, Canada, [3]Latner Thoracic Research Laboratories, Toronto General Hospital Research Institute, Toronto, ON, Canada, [4]Division of Respirology and Critical Care Medicine, Department of Medicine, University Health Network, Toronto, ON, Canada, [5]Division of Thoracic Surgery, Toronto General Hospital, Toronto, ON, Canada, [6]Thoracic Surgery Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States

Awareness of RNA-based therapies has increased after the widespread adoption of mRNA vaccines against SARS-CoV-2 during the COVID-19 pandemic. These mRNA vaccines had a significant impact on reducing lung disease and mortality. They highlighted the potential for rapid development of RNA-based therapies and advances in nanoparticle delivery systems. Along with the rapid advancement in RNA biology, including the description of noncoding RNAs as major products of the genome, this success presents an opportunity to highlight the potential of RNA as a therapeutic modality. Here, we review the expanding compendium of RNA-based therapies, their mechanisms of action and examples of application in the lung. The airways provide a convenient conduit for drug delivery to the lungs with decreased systemic exposure. This review will also describe other delivery methods, including local delivery to the pleura and delivery vehicles that can target the lung after systemic administration, each providing access options that are advantageous for a specific application. We present clinical trials of RNA-based therapy in lung disease and potential areas for future directions. This review aims to provide an overview that will bring together researchers and clinicians to advance this burgeoning field.

KEYWORDS

RNA-based therapy, mRNA vaccine, nanoparticle, noncoding RNA, antisense oligonucleotide, lung disease, RNA, COVID-19

1 Introduction

The COVID-19 pandemic has publicized the value of RNA-based therapies because of the rapid development and widespread use of mRNA vaccines (Kojima et al., 2021). However, the potential of RNA-based therapy reaches far beyond mRNA vaccines, though the adoption of mRNA vaccines itself could be considered revolutionary. A major advantage of RNA-based therapies is that they drastically expand the numbers and types of targets that can be addressed therapeutically. While some proteins are considered difficult to target by other approaches, virtually all proteins and even noncoding RNAs are susceptible to RNA-based therapy (Bennett et al., 2017). RNA-based therapies can address genetic disease, with the ability to be highly

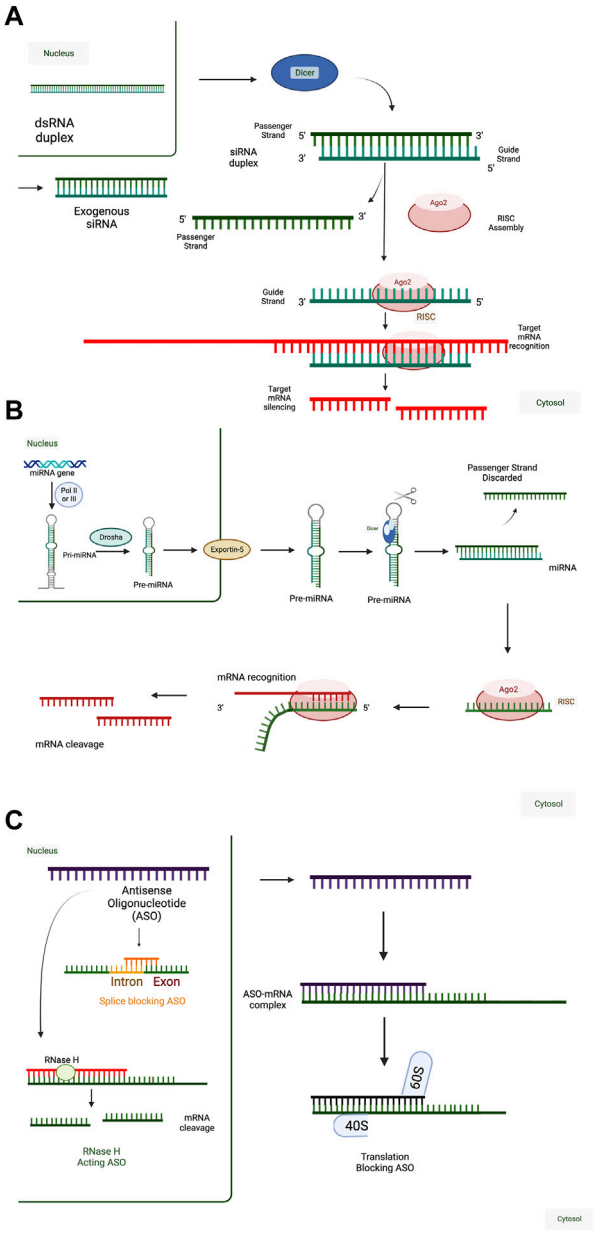


FIGURE 1
Mechanisms of RNA-Mediated Gene Regulation. **(A)** (Short-interfering RNA) siRNA mechanism of action. Double-stranded RNA (dsRNA) transcribed in the nucleus (e.g., from endogenous retroviral elements) are cleaved by Dicer in the cytoplasm into shorter RNA duplexes, which can interact with the RNA-induced silencing complex (RISC). Alternatively, exogenous siRNA (e.g., therapeutic siRNA) can interact with the RISC upon cellular entry. In both situations, the passenger strand of the siRNA duplex is discarded, and only the guide strand is incorporated in RISC. 100% complementarity of the siRNA sequence with the target mRNA leads to cleavage and degradation of the target mRNA. Therapeutic siRNAs are designed to act against one target mRNA and decrease target mRNA expression and function. **(B)** MicroRNA (miRNA) mechanism of action. In the canonical miRNA processing pathway, endogenous miRNAs are transcribed as primary miRNAs (pri-miRNA) and are processed by Drosha into shorter pre-miRNA sequences. Pre-miRNA sequences are then exported into the cytoplasm by Exportin-5 and processed into shorter, mature miRNA sequences by Dicer. Other miRNA processing pathways exist but are less common than the canonical pathway. Mature miRNA sequences, or exogenously administered miRNA duplexes, can interact with the RISC, discarding the passenger strand. miRNAs act through

(*Continued*)

FIGURE 1 (Continued)
incomplete complementarity of the miRNA guide strand with target mRNAs and rely on a 7-8 nucleotide complementary "seed sequence," typically in positions 2-8, to interact with a network of target mRNAs. Interaction of the miRNA-RISC complex with target mRNAs leads to translation inhibition or mRNA target degradation. Therefore, therapeutic miRNAs are designed to act against a network of target mRNAs and decrease target mRNA expression and function. **(C)** Antisense oligonucleotide (ASO) mechanism of action. ASOs are a diverse class of single-stranded, RNA-based therapeutics. ASO sequences can be formed from combinations of RNA, DNA, and modified nucleic acids in any order because their short sequences can synthesized in a base-by-base fashion (solid phase synthesis), as with miRNAs and siRNAs. Their mechanisms of action depend on both the chemistry, including base composition and sequence. ASOs can function in the nucleus or the cytoplasm and either decrease or increase target mRNA expression and function. ASOs with ~10 nucleotide central DNA sequences flanked by ~5 nucleotide RNA sequences with complete complementarity to the target mRNA can be used to decrease target mRNA expression by RNAse H. They can also be designed to bind intron-exon boundaries to block splicing to influence the expression of alternatively spliced RNA variants. In the cytoplasm, ASOs can be designed to block translation initiation sites of an mRNA. When these are designed against the primary translation initiation site, they will decrease the translation of the target mRNA into protein and when these are designed against upstream translation initiation sites that interfere with the primary translation initiation site, they will increase the translation of the target mRNA into protein. can impede translation by specifically targeting mRNA sequences, as translation blocking ASO. Created with BioRender.com.

specific, even targeting sequences with single-base pair mutations (Monga et al., 2017). Furthermore, RNA can be a vehicle for gene delivery into cells, whether as replacement therapy for protein-coding or noncoding genes or foreign sequences, such as with viral vaccines (DeWeerdt, 2019). The lungs are an attractive target for RNA-based therapy and can be accessed both by direct local delivery and systemic delivery. The goal of this review is to create a dialogue between lung-focused researchers, clinicians and developers of RNA-based therapy.

A major attraction of RNA-based therapies is that the primary nucleotide sequence forms the basis of the therapeutic effect. This principle has been best demonstrated by the development of COVID-19 mRNA vaccines created by Moderna and Acuitas/BioNTech/Pfizer, which took months rather than the years typically required for drug development (Kojima et al., 2021). Not only were the initial drugs developed quickly, but adaptation to new variants was fast based on the ability to generate a new "drug" once a different sequence was identified. This principle applies not only to mRNA vaccines but also to other mRNA therapeutics and to antisense therapies. Similarly, once delivery vehicles are developed that can target specific tissues and cell types, sequences for various targets can be rapidly developed to treat distinct disorders.

1.1 Timeline of development

Since early descriptions of RNA in the late 1950s (Rich and Davies, 1956), the field of RNA has grown to include many RNA classes, with functions ranging from encoding protein sequences to catalyzing enzymatic reactions and orchestrating gene regulation programs. Messenger RNA (mRNA) encoding for protein was described in 1961, a discovery which has anchored the central dogma of biology: that cellular information flows from DNA-to-

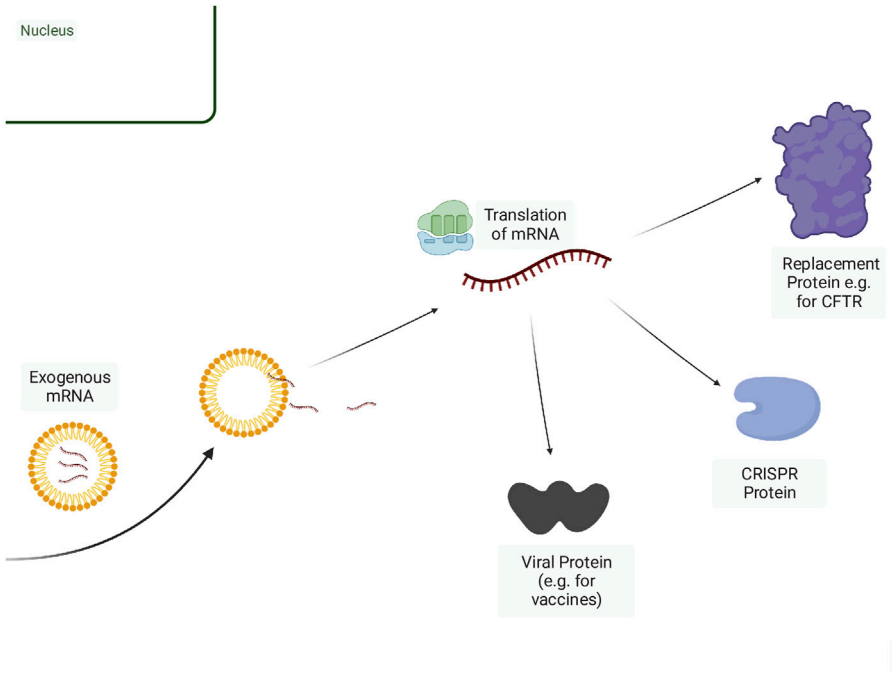


FIGURE 2
Mechanisms of mRNA-based therapy. Exogenous mRNA can be delivered naked but is more efficiently delivered within a delivery vehicle. Once released from the endosome, mRNA can then be translated into protein. Fragments of viral proteins or tumour antigens can be encoded in the mRNA for use as a vaccine. Deficient wild-type proteins can be replaced or augmented via mRNA delivery. mRNAs encoding CRISPR proteins along with a guide RNA can be delivered for genome editing or other adaptations of CRISPR function, including RNA degradation. Created with BioRender.com.

RNA-to-protein (Crick, 1970). This perception of RNA biology dominated until the discovery of RNA interference (RNAi) and microRNAs (miRNAs) highlighted the critical role of small RNAs in gene regulation across species (Lee et al., 1993; Wightman et al., 1993; Fire et al., 1998).

The field of RNA therapeutics dates back as far as 1978 when Stephenson and Zamecnik designed an antisense oligonucleotide (ASO) that utilized RNA base-pairing to inhibit viral replication of the Rous sarcoma virus (Zamecnik and Stephenson, 1978). By 1990, the potential of mRNA transcripts as a vaccine was reported, after observing persistent gene expression in mice following *in vivo* injections of mRNA (Wolff et al., 1990). Further experimentation with mRNA led to the creation of a vaccine for the influenza virus (Martinon et al., 1993). In 1998, the FDA approved the first antisense RNA drug to treat cytomegalovirus retinitis (Roehr, 1998). Subsequently, in 2018, the first siRNA drug was approved for patients with hereditary transthyretin-mediated amyloidosis (Crooke et al., 2018). Despite these early successes, the potential of RNA-based therapy did not hit the public eye until 2020, when the first mRNA vaccines for COVID-19 were granted emergency use authorization, with full approval following in 2021 (Thompson et al., 2021). These therapies transformed the public health response to this rapidly emerging pandemic (Hogan and Pardi, 2022).

2 Spectrum of RNA-based payloads

In theory, RNA payloads falling into any category of RNA (mRNA, microRNA, *etc.*) can be delivered therapeutically. In practice, current RNA-therapeutics fall into three broad categories: 1) RNA designed to inhibit target gene expression, often described as antisense RNA in reference to their mechanism of action via base-pairing (Figure 1); 2) RNA designed to express a protein, often referred to as mRNA-based therapy (Figure 2), and 3) those that target protein (RNA aptamers) (DeWeerdt, 2019). In this review, we focus on the first two categories. RNA aptamers are discussed elsewhere (Tuerk and Gold, 1990; Kang and Lee, 2013; Sundaram et al., 2013; Lei et al., 2023).

2.1 Antisense therapy: RNA-based therapy designed to inhibit gene expression

Most antisense therapies consist of short RNA molecules, often <30 nucleotides, termed oligonucleotides, and include short interfering RNAs (siRNAs), microRNAs (miRNAs), and antisense oligonucleotides (ASOs) (Crooke et al., 2018; Winkle et al., 2021). Antisense therapies target other nucleotides, typically other RNA species, through sequence complementarity following Watson-Crick base pairing rules (A:T/U, C:G) (Crooke et al., 2018; Winkle et al., 2021). Theoretically, this characteristic allows for targeting any known unique nucleotide sequence and positions antisense therapies well for developing treatments that target the ~84% of proteins not currently druggable by other methods (Zhu et al., 2022). We will briefly review RNA therapies that act by: 1) degrading the target RNA through endogenous enzymes, such as the RNA interference (RNAi) pathway or RNAse H, or 2) mechanisms

that do not involve degradation of the target RNA, such as modulation of RNA processing/splicing/polyadenylation and blocking translation into protein. While the double-stranded miRNAs and siRNAs are also "oligonucleotides" designed to act through Watson-Crick hybridization, the term ASO generally refers to single-stranded antisense oligos (Crooke et al., 2021).

Current antisense therapies typically consist of short nucleic acid molecules less than 30 nucleotides in length and can be produced through solid phase synthesis, which is performed in a base-by-base fashion (Juliano, 2016). Thus, the chemistry of nucleotides at specific positions can be adapted to confer desired properties, providing greater versatility and precision in the design of these treatments.

2.1.1 RNAi

RNAi leverages the endogenous RNAi pathway by which mRNA cleavage and degradation is initiated by the RNA-induced silencing complex (RISC) (Lingel and Izaurralde, 2004). Endogenous RNA species that utilize the RNAi pathway include microRNAs (miRNAs), short-interfering RNAs (siRNAs), or Piwi-interacting RNA (piRNA), all of which are short non-coding RNAs (ncRNAs) that are involved in post-transcriptional gene silencing rather than encoding proteins (Fire et al., 1998; Bartel, 2004; Girard et al., 2006). Mature miRNAs and siRNAs are double-stranded ncRNAs, measuring 20–25 base pairs in length. They are loaded onto the RNA-induced silencing complex (RISC) and lead to mRNA degradation and/or inhibit translation, thereby inhibiting the expression of specific genes (Watts et al., 2008; Ferguson et al., 2020; Tian et al., 2021) (Figure 1). This approach of targeting and degrading specific mRNA sequences can be used to control gene expression and provides a promising method for treating a wide range of diseases.

2.1.1.1 miRNA

miRNAs are endogenous small, non-coding RNA molecules that play a crucial role in regulating gene expression. There are currently thought to be 2,656 human miRNAs (miRDB - Statistics, n.d.) (Chen and Wang, 2020) predicted to target 29,161 unique genes. miRNAs are usually transcribed as longer, primary miRNA sequences before being processed into mature double-stranded miRNA that can interact with RISC (Figure 1) (Shukla et al., 2011). One strand is often favoured for incorporation into the RISC complex and silence targets with complementarity to that strand. The RISC contains one strand of the miRNA and various proteins such as Dicer, TRBP, PACT, and Argonaute (Bartel, 2004; Carthew and Sontheimer, 2009). The complex binds to specific sequences on target mRNAs based on a 7-8 nucleotide seed sequence (Doench and Sharp, 2004) on the miRNA and regulates the expression of target mRNAs by translational repression or mRNA degradation [or deadenylation] (Figure 1)]. Because of this short seed sequence, miRNAs typically target multiple genes, and therapeutic applications involving miRNAs can be an efficient strategy to modulate the expression of a set of genes within a miRNA regulatory network (Bartel, 2009; Bartel, 2018).

Although the seed sequence is central to miRNA function, other factors also influence which mRNA targets are regulated. Many miRNAs are expressed in a highly tissue-specific fashion and contribute to tissue-specific expression patterns (Sood et al., 2006). There is also cell- and context-specific action of miRNAs, whereby binding of miRNA to targets is variable across cell types and timelines (Ho et al., 2013; Nowakowski et al., 2018). In this context, the targeting of mRNAs can depend on the level of expression of the mRNA vs. miRNA and RNA-binding proteins that can occupy miRNA binding sites (Ho et al., 2013; Ho et al., 2021). miRNA families consist of multiple miRNAs which share a seed sequence but can have overlapping and distinct target genes and subcellular localization (Horita et al., 2021). These factors are important considerations for the design of miRNA-based therapies.

Several types of RNA-based therapy can leverage the biology of miRNAs and can either potentiate or inhibit miRNA function. miRNA mimics are synthetic miRNAs that can be used to repress a set of miRNA-regulated genes, whereas antisense oligonucleotides such as antagomiRs or blockmiRs are synthetic oligonucleotides that bind to miRNAs or compete for their target sites on mRNAs respectively (Krützfeldt et al., 2005; Davis et al., 2007). miRNA sponges contain multiple miRNA binding sites and can be used to sequester miRNAs from target genes (Ebert et al., 2007)).

miRNA-based therapy has been assessed in non-malignant and malignant lung disease in preclinical studies. miRNAs in the miR-29 family are downregulated in pulmonary fibrosis (Cushing et al., 2015). MRG-229 is a miR-29 mimic that decreases the expression of pro-fibrotic genes in animal models of pulmonary fibrosis (Chioccioli et al., 2022). The multiplexing potential of RNA-based therapies and delivery systems has been leveraged in models of non-small cell lung cancer (NSCLC). To simultaneously target mutation in the Kirsten ras sarcoma viral oncogene homolog (KRAS) and loss of p53 in a single-nanoparticle platform, miR-34a, siKras and cisplatin were formulated layer-by-layer into a multifunctional nanoparticle. This multifunctional nanoparticle therapy was preferentially distributed to the lung and prolonged the survival of mice in an orthotopic lung cancer model (Gu et al., 2017). These early studies demonstrate the versatility and specificity of miRNA-based therapeutics in lung disease.

2.1.1.2 siRNA

siRNAs are a class of double-stranded RNA molecules that can silence specific genes by binding to complementary sequences in the mRNA and triggering degradation or inhibition of translation. Endogenous siRNAs exist and differ from miRNAs in that they are not derived from primary miRNA transcripts, and they do not undergo canonical miRNA processing. Rather, they can be derived from long double-stranded RNA (dsRNA) which have a variety of sources and are cleaved by Dicer to form siRNAs (Carthew and Sontheimer, 2009). In this context, endogenous RNAi acts as a natural defence against viral infections and transposable elements (Boudreau et al., 2011).

siRNAs are also derived exogenously and are used for gene knockdown in research and RNA-based therapeutics. In contrast to miRNAs, siRNAs typically have 100% sequence complementarity with their target genes (Lam et al., 2015). As a result, appropriately designed siRNAs theoretically have a single target, whereas miRNAs have multiple targets. Aside from selectivity for a single target, this feature can be applied to allele-specific gene silencing, where a pathogenic gain-of-function mutation on one allele can be targeted while sparing

the wild-type allele, even with a single nucleotide difference in sequence (Monga et al., 2017).

Despite this theoretical single-sequence specificity, siRNA can still have off-target effects (Woolf and Costigan, 1999). One proposed mechanism is that siRNA sequences can function with incomplete complementarity, with 6-7 nucleotide sequences within the siRNA effectively acting as seed sequences, analogous to miRNAs (Birmingham et al., 2006). Such short seed sequences can target multiple genes. These effects can be minimized by choosing sequences predicted to have fewer off-target effects and by appropriate control experiments, including the assessment of mulitple siRNA sequences for a single target. Chemical modifications can also reduce off-target effects from seed sequence matching (Jackson et al., 2006). Other off-target effects from immune activation and saturation of the RNAi machinery are discussed in more detail later.

siRNAs are a powerful tool in the knockdown of target RNAs and have been a frequently used tool in the research setting to assess gene function in a variety of biological systems (Mello and Conte, 2004). Like miRNAs, siRNAs act through the RISC complex (Carthew and Sontheimer, 2009). One aspect of siRNA design is that the guide strand of the siRNA cannot be chemically modified as it interferes with its incorporation into the RISC. On the other hand, the passenger strand of the siRNA can be chemically modified to ensure its stability (Kuijper et al., 2021).

2.1.2 Antisense oligonucleotide (ASO)

ASOs are a more chemically and mechanistically diverse group of molecules compared to miRNAs and siRNAs, both of which act through the RNAi pathway. ASOs are single-stranded molecules that can alter mRNA expression through a variety of mechanisms, including ribonuclease H (RNAse H) mediated decay, or steric hindrance of splice-sites or translation initiation (Bennett and Swayze, 2010; Crooke et al., 2017). Depending on the design and thus the mechanism of action, ASOs can either downregulate target gene expression through RNAse H or translation inhibition or upregulate target gene expression by increasing translation efficiency or by modulation of splicing (Rinaldi and Wood, 2018; Kim N. et al., 2020). Like siRNAs, ASOs can also be made through solid-phase synthesis.

The mechanism of action can be modulated by ASO biochemistry is a determinant of the mechanism of action. RNAs H is an endogenous enzyme that cleaves the RNA strand in an RNA-DNA duplex (Raal et al., 2010; Roshmi and Yokota, 2019). Thus, RNAse H-dependent ASOs are formed of or contain DNA sequences to target an mRNA. Off-target RNA cleavage can occur due to partial complementarity of only 6-7 nucleotides. Thus, a 20-nucleotide ASO formed completely of deoxynucleic acid (DNA) nucleotides can lead to cleavage of off-target mRNAs (Di Fusco et al., 2019). Thus, to reduce off-target effects, chimeric structures known as gapmers have been developed, with a central 10-nucleotide region of DNA flanked by regions of five RNA-like nucleotides that will not activate RNAse H (Bennett et al., 2017).

ASOs that act as steric blockers in an RNAse H-independent fashion (Figure 1) consist of RNA or modified RNA-like bases without a region of DNA (Bennett et al., 2017; Bandyopadhyay et al., 2018; Chandra Ghosh et al., 2018). Often, they utilize modified chains of synthetic nucleic acids intended to achieve greater stability and longer half-lives (Østergaard et al., 2015), and some have the potential to be delivered without a delivery vehicle (Bennett and Swayze, 2010). These non-enzymatic mechanisms include splicing modulation or inhibition of translation and require that ASO sequences target splice sites and start codons of translation, respectively.

ASOs that bind to intron-exon junctions in pre-mRNA destabilize splicing sites or displace/recruit splicing factors, and result in the exclusion or inclusion of certain exons (Bandyopadhyay et al., 2018; Chandra Ghosh et al., 2018). Open reading frames (ORFs) in the 5′-untranslated region (5′-UTR) of mRNAs, known as uORFs, can affect the translation efficiency of the primary open reading frame (ORF) into protein (Wang et al., 1999), and exist in over half of all human mRNAs (Calvo et al., 2009; Bottorff et al., 2022). ASOs can increase the expression of proteins encoded by target mRNAs by blocking the activity of uORFs, thereby increasing translation efficiency (Liang et al., 2016).

The FDA has approved several ASOs outside of lung disease, such as Nusinersen and Mipomersen (Kim, 2022). In non-malignant lung disease, ASOs have been used to target transforming growth factor-beta (TGF-β) (Kim J. et al., 2020) and NLRP3 (Bai et al., 2019), with the latter having the potential for prevention but not reversal of pulmonary fibrosis. ASOs have also been used to target Jagged 1, a protein involved in the development of goblet cell metaplasia in the lungs, showing decreased mucus production and reduced goblet cells in mice (Carrer et al., 2020). In SARS-CoV-2, ASOs that target a conserved region among immune evasive variants can inhibit viral replication and increase the survival of transgenic mice with human ACE2 (Vora et al., 2022). Other studies have also identified conserved sites in SARS-CoV-2 variants amenable to targeting by ASOs (de Jesus et al., 2021).

2.2 mRNA

The most widespread applications of mRNA therapeutics are the mRNA-based vaccines against SARS-CoV2 (Paunovska et al., 2022). mRNAs are typically longer RNA molecules, with eukaryotic mRNAs being ~2,900 nucleotides on average (Guttman et al., 2010). mRNA therapeutics can encode parts of genes, as with vaccines or whole genes, and are used to introduce genetic information into cells (Rohner et al., 2022) (Figure 2). Because they are not produced on a base-by-base basis via solid-phase synthesis, there are limitations to generating mRNAs with site-specific chemical modifications (Paunovska et al., 2022). However, the introduction of chemically modified nucleotides can improve stability and reduce immunogenicity. For example, twice weekly injection of surfactant protein B (SP-B) mRNA with only 25% replacement of uridine and cytidine with 2-thiouridine and 5-methyl-cytidine can restore surfactant expression to 71% of wild-type in SP-B deficient mice (Kormann et al., 2011). Self-amplifying RNA (sa-RNA) vaccines are a form of mRNA-based vaccines (Frederickson and Herzog, 2021) that encode an RNA-dependent RNA polymerase; this allows for self-replication of the transcript, leading to higher levels and longer duration of protein expression.

When introducing genetic information, such as in vaccines or CRISPR-Cas9-based approaches (Gillmore et al., 2021), an advantage of mRNA therapies compared to DNA-based

therapeutics is the short duration of protein production due to the shorter half-life of mRNAs compared to DNA vectors (Pardi et al., 2015), and lower risk of incorporation into the genome (Schlake et al., 2012). In applications that involve DNA editing (e.g., with CRISPR-Cas9), long-term expression from DNA payloads may lead to more off-target editing events, and there is the potential for DNA to integrate into the genome (Hanlon et al., 2019; Chen et al., 2020). In the lung, non-vaccine mRNA therapies are under development. One example is the development of inhaled mRNA to enhance cystic fibrosis transmembrane receptor (CFTR) expression in cystic fibrosis (Damase et al., 2021; mRNA Technology: Vaccines and Beyond - Sanofi, n.d.).

Beyond vaccines for infectious diseases, mRNA vaccines can be designed as cancer vaccines. For example, dendritic cells can be loaded *ex vivo* with mRNA coding for tumour-associated antigens and administered back to patients to elicit an immune response (Schumacher and Schreiber, 2015; Sahin et al., 2017).

2.3 Other RNA payloads

Other RNA classes and RNA-based molecules have been assessed for therapeutic potential. For example, circular RNAs (circRNAs) are a group of noncoding RNAs that have covalently joined ends and are present in all studied eukaryotic organisms (Sanger et al., 1976; Wilusz, 2018; Pfafenrot et al., 2021). Another examples stems from the discovery that RNA moieties formed the catalytic subunit of ribonuclease P provided evidence that nucleic acids can have inherent enzymatic activity (Kruger et al., 1982; Guerrier-Takada et al., 1983; Guerrier-Takada and Altman, 1984). Since then, at least 21 ribozyme families have been identified, including Hovlinc (human protein vlincRNA localization), a recently evolved class of ribozymes found in human ultra-long intergenic non-coding RNA (Chen et al., 2021; Deng et al., 2023). These discoveries have sparked interest in designing RNA-cleaving nucleic acid therapeutics as gene-silencing agents with modified nucleic acid bases conferring improved biological activity. However, challenges still exist in this field (Wang, 2021).

As with vaccines, genetic information traditionally delivered via DNA-based molecules can be delivered via RNA-based therapies. CRISPR-Cas systems were first identified as prokaryotic adaptive immunity that protects against phages and has been widely adopted in research for genome editing (Doudna and Charpentier, 2014). CRISPR-Cas systems adapted for application in eukaryotic systems are composed of a targeting RNA (guide RNA, sgRNA) and a Cas enzyme. The sgRNA, together with the Cas enzyme, can be delivered as DNA via a viral vector or as RNA via a lipid nanoparticle. Some Cas enzymes target RNA, including long noncoding RNAs, rather than DNA and result in RNA knockdown without relying on endogenous RNAi machinery (Konermann et al., 2018; Li G et al., 2021).

2.4 Challenges and solutions

2.4.1 Activation of immune response

Double-stranded RNA (dsRNA) is a potent activator of innate immune responses that form a natural defence against viral RNA and transposable elements (Sadeq et al., 2021). Thus, the introduction of foreign dsRNA species can/will induce an immune response. Indeed, siRNA-based RNA therapeutics have commonly been seen to induce an innate immune response (Marques and Williams, 2005; Sioud, 2007; Robbins et al., 2008; Morral and Witting, 2012; Meng and Lu, 2017); this unintended effect can confound experimental results and contribute to side effects in a pharmacologic context (Olejniczak et al., 2011). siRNA structure, sequence, and the method of delivery can all contribute to an immune response. In particular, GU-rich sequences similar to viral RNAs are described to activate RNAi in cells (Meng et al., 2013).

Several strategies have been employed to mitigate the issues that can arise and lead to unintended immune stimulation. One is to avoid using sequences with known immunostimulatory motifs, for example, 5′-UGU-3′, 5′-UGUGU-3′ (Judge et al., 2005), and 5′-GUCCUUCAA-3′ (Hornung et al., 2005). Another method seen is to design siRNAs with modified nucleotides that reduce unwanted immune activation; incorporating certain modified nucleotides into the siRNA, such as 2′-O-methyl purines, 2′-fluoropyrimidines, and terminal inverted-dT bases can prevent immune activation (Morrissey et al., 2005).

2.4.2 Saturation of the endogenous RNAi machinery

The endogenous RNAi machinery is critical for normal gene regulation by the miRNAs present in a cell at any given time. When exogenous siRNAs or miRNAs are introduced for research or therapeutic applications, there is the potential to saturate a significant proportion of the endogenous RNAi machinery (i.e., RISC) and competitively inhibit the function of endogenous miRNAs. In this way, exogenous administration of siRNA or miRNA can lead to the upregulation of endogenous miRNA target mRNAs (Khan et al., 2009). Saturation of this critical cellular machinery can have significant consequences *in vivo* (Grimm et al., 2006). Therefore, methods to assess the effects of different doses of RNAs on direct and indirect gene regulation to avoid RISC saturation can be employed (Grimm et al., 2006).

2.4.3 RNA stability

RNA is susceptible to degradation by nucleases (Chioccioli et al., 2022). Specific chemical modifications can be made to the backbone to increase RNA stability. For instance, by substituting sulphur atoms for one of the non-bridging oxygen atoms in the internucleotide phosphate groups or through 2′-O-methyl or 2′-Fluoro-ribose sugar modifications, the resistance to nuclease degradation is enhanced and the circulation time is prolonged (Gaus et al., 2019). Bioactive molecules, such as cholesterol and lipids, can also be covalently attached to oligonucleotides in order to increase directed delivery (Hammond et al., 2021). For example, cholesterol conjugates increase delivery to the liver while reducing delivery to the kidney (Bennett et al., 2017).

3 Vehicles for RNA delivery

The potential versatility of RNA therapies presents great opportunities for disease treatment but requires that the RNA

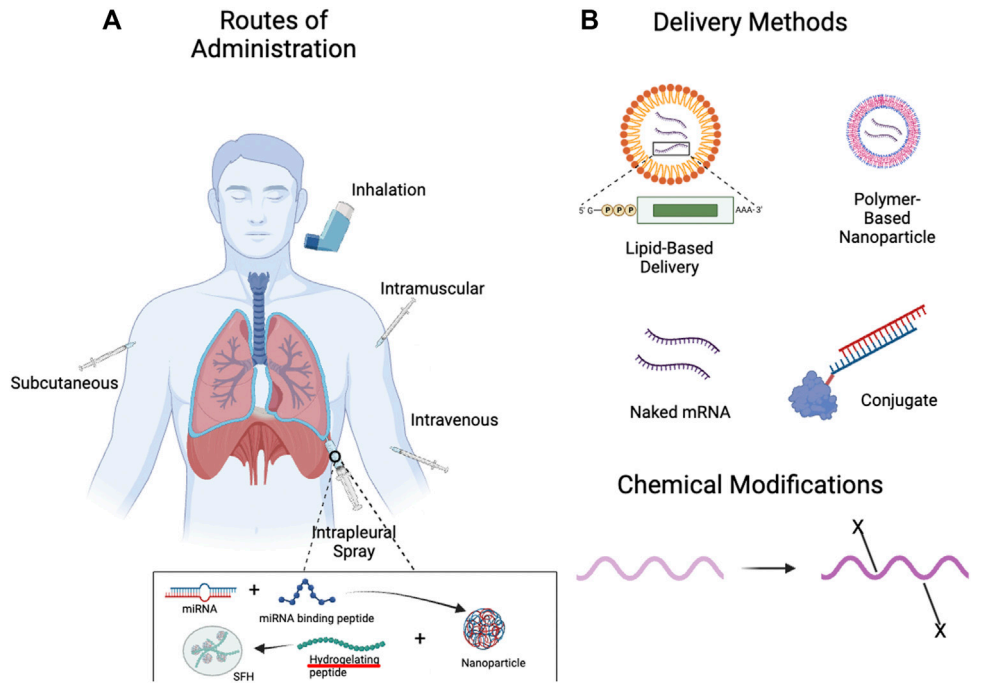


FIGURE 3
Methods of Administration and Delivery Methods for RNA-Based Therapies. **(A)** With advances in delivery vehicles, the lung can be targeted by RNA-based therapies through both local (inhalation, pleural spray/injection) (Majumder et al., 2021) and systemic delivery methods (intravenous, subcutaneous, *etc.*). **(B)** RNA-based therapies can be delivered naked, within delivery vehicles such as lipid- or polymer-based nanoparticles, and can also be conjugated with other moieties for the purposes of targeting. Chemical modifications can confer favourable properties, such as increased stability and reduced immunogenicity. Created with BioRender.com.

payload is delivered to the appropriate cells. Because of their negative charge, RNA molecules will encounter barriers in crossing cell membranes, and native RNA is susceptible to the ubiquitous ribonucleases (RNases) present in the body (Chioccioli et al., 2022). Chemical modifications can improve RNA stability but are generally applicable only to short RNAs that can be synthesized by solid phase synthesis (Gaus et al., 2019). Therefore, small oligonucleotides may be chemically modified to increase stability in the absence of an external vehicle whereas mRNAs must be delivered within a vehicle (Paunovska et al., 2022). Compared to DNA-based carriers of genetic information, which require transcription to RNA in the nucleus, many RNA-based approaches have the advantage that they can be delivered to the cytoplasm to exert their function without the need for nuclear delivery (Schott et al., 2016). As such, viral vectors such as adeno-associated virus (AAV) can be efficient delivery vehicles for RNA but have the disadvantage of triggering adaptive immunity (Tomar et al., 2003).

The development of nanotechnology-based delivery strategies is a critical component of effective RNA-based treatments (Figure 3). Nanotechnology is defined by structures roughly 1–100 nm in at least one dimension, though the term often refers to structures up to several hundred nanometers in size (Farokhzad and Langer, 2009). Delivery vehicles are designed to protect RNA from degradation, deliver the RNA payload to the appropriate cell type, and allow uptake of the RNA payload into the cell via endocytosis while minimizing toxicity (Cheng et al., 2015). Nanotechnology products can also co-deliver multiple payloads and be visualized at delivery sites with imaging (Farokhzad and Langer, 2009). Common nanotechnology vehicles include lipid-based nanoparticles (LNPs), polymer-based nanoparticles and peptide-based nanoparticles. A comprehensive discussion of delivery vehicles is out of the scope of this article and reviewed elsewhere (Paunovska et al., 2022). Here, we will present a selection of delivery vehicles and, where applicable, highlight target specificity relevant to the lung.

3.1 Lipid-based nanoparticles

The most established form of nanodelivery agents are lipid-based (Langer and Folkman, 1976; Adams et al., 2017). Lipid-based nanoparticles (LNPs) have been approved by the FDA for delivery of siRNA to the liver and for mRNA vaccine delivery (Paunovska et al., 2022). LNPs are based on structures that can be formed by phospholipids in aqueous solutions, such as micelles and liposomes (Paunovska et al., 2022). Traditional LNPs contain multiple lipids, including cationic lipids, amphipathic phospholipids, and poly (ethylene glycol) (PEG), as well as cholesterol (Semple et al., 2010; Cheng et al., 2015; Paunovska et al., 2022). Variations of these components can be used to direct LNP uptake to specific vascular beds (Paunovska et al., 2018). LNPs can be used to deliver a range of RNA payloads from siRNA to mRNA (Musunuru et al., 2021; Rothgangl et al., 2021).

Because systemic delivery of RNA often leads to accumulation in the liver (Li et al., 2020), various strategies have been developed to target other organs, including the lung. One strategy has been to covalently conjugate an antibody that binds plasmalemma vesicle-associated protein (PV1) to LNPs. This strategy can increase the protein expression of mRNA delivered via these antibody-conjugated LNPs by 40-fold in the lung (Li et al., 2020). Nanoparticle size is an important factor in lung localization, with 160 nm conjugated LNPs leading to greater protein expression in the lungs compared to smaller 70 nm particles (Anselmo et al., 2015; Li et al., 2020). Similarly, the mRNA:lipid ratio is also an important factor in protein expression from delivered mRNA, with a 3% mRNA:lipid ratio showing a ten-fold increase in mRNA expression over a 10% mRNA:lipid ratio (Kranz et al., 2016; Li et al., 2020).

Specific cell types within the lung including epithelial cells, endothelial cells, and immune cells along with cells outside the lung, such as hematopoietic stem cells can be targeted by specific LNP formulations (Paunovska et al., 2022). A strategy of selective organ targeting has been used to systematically generate multiple classes of lipid nanoparticles that can target the lung, liver or spleen and specific cell types such as epithelial cells, endothelial cells, B cells, T cells and hepatocytes (Cheng et al., 2020; Wang et al., 2023). These methodologies allow the engineering of modified lipid nanoparticles to efficiently deliver ribonucleoprotein complexes (RNPs), including CRISPR/Cas9 RNPs, for multiplex editing of genes in mouse lungs (Wei et al., 2020).

In addition to endothelial uptake in the lung (Dahlman et al., 2014; Khan et al., 2018; Cheng et al., 2020), LNPs have been developed for more effective delivery to the lung via nebulization. Delivery of mRNA encoding a neutralizing antibody to haemagglutinin via optimized LNPs could protect mice from a lethal challenge of the H1N1 influenza A virus (Lokugamage et al., 2021). Systematic evaluation of multi-component LNPs has led to the development of LNPs that enable tissue-selective RNA delivery to the lung, even with intravenous administration (Liu et al., 2021). Similar approaches have developed LNPs to deliver mRNA with traditional (Cre recombinase) and CRISPR-Cas9 gene editing tools *in vivo* to lung epithelium via inhalation (Li et al., 2023). Compared to DNA-based viral gene therapy, this strategy avoids long-term expression of genome editors, which can lead to the accumulation of off-target genome mutations. Compared to viral vectors, which elicit an adaptive immune response, this strategy also allows for repeated delivery.

3.2 Polymer and polymer-based nanoparticles

A polymer is a macromolecule that consists of multiples of a simpler chemical unit called a monomer and can be natural or synthetic. Polymers are good delivery vehicles because they are biocompatible and have simple formulation parameters that can adopt a variety of possible structures and characteristics (Mitchell et al., 2021). Basic traits such as charge, degradability and molecular weight can be varied to generate desired characteristics (Paunovska et al., 2022). Polymers such as poly (ethylenimine) (PEI), poly (L-lysine) (PLL) and poly (beta-amino-ester) (PBAE) can complex to the anionic phosphodiester backbone of RNA via cationic amine groups. Polymers that do not contain cationic groups can be designed to contain separate cationic groups to serve the same purpose.

The respiratory system presents an accessible but challenging drug delivery target. While there is direct access to the environment through the airways, biological barriers designed for protection against foreign materials, such as mucus, must be overcome for effective delivery. PBAE nanoparticles containing mRNA administered intranasally in mice show expression of the protein product and no evidence of toxicity, such as granuloma or chronic inflammation, for up to 3 months (Su et al., 2011). PBAEs have been used to design stable biodegradable DNA nanoparticles that can penetrate highly adhesive human mucus gel layers and could achieve stable and repeatable transgene delivery in a mouse model (Mastorakos et al., 2015). These PBAE based polyplexes delivered via aerosol can have localized delivery without evidence of reporter mRNA expression in other tissues (Patel et al., 2019).

Other polymer-based nanoparticles can also target the lung. Nanoparticles with multi-modular peptides with three different functional modules and poloxamine polymers are able to deliver co-packaged payloads of mRNA and pDNA for CFTR gene knock-in *in vivo* (Guan et al., 2019). The combination payload consisted of a sleeping beauty transposon containing CFTR sequences and a transposase-encoding mRNA. An advantage of this RNA-based system for gene knock-in is that the transposase is expressed only temporarily and reduces the risk of ongoing insertional mutations (Kebriaei et al., 2017). Importantly, there were fewer integration sites within RefSeq mouse genes compared to lentiviral or adeno-associated viral vectors.

To achieve site-specific editing of CFTR, poly (lactic-co-glycolic) acid (PLGA) nanoparticles have been used to deliver peptide nucleic acids (PNAs) to correct F508del CFTR mutation *in vitro* in human bronchial epithelial cells and *in vivo* in a CF murine model (McNeer et al., 2015). Airway mucus can pose a challenge to drug delivery in cystic fibrosis and other lung diseases. Hybrid lipid-polymer nanoparticles comprising a PLGA core and a dipalmitoylphosphatidylcholine (DPPC) shell engineered for inhalation can penetrate the mucous barrier and release PNA cargo within the cytoplasm (Comegna et al., 2021). Furthermore, polymer-based nanoparticles have been designed that can deliver mRNA to immune cells within the lung. These nanoparticles can be administered systemically and may be helpful for pulmonary-specific immunomodulation (Ke et al., 2020).

3.3 Cell-penetrating peptides

Peptide nanoparticles leverage amino acid sequences to impart properties such as structure, charge, solubility, and polarity, which can facilitate interactions with RNA and cellular uptake (Chow et al., 2020). KL4 peptides, which mimic surfactant protein B, can be formulated as a dry powder and deliver both siRNA and mRNA to the lung (Qiu et al., 2017; Qiu et al., 2019). Peptide-based delivery vehicles can also be formulated to target specific cell types. For example, complexes of oligoarginine micelles with high mobility

TABLE 1 Overview of clinical trials investigating RNA therapies in lung disease.

Type of disease	Name	Treatment	Genetic/protein target	Delivery vehicle	Administration method	Disease	Number of patients	ClinicalTrials.gov identifier	Phase
Infectious Disease									
	BNT162b2	mRNA	SARS-CoV2 spike glycoprotein	Lipid nanoparticles	Vaccine	COVID-19	43,448	NCT04368728	Approved
	mRNA-1273	mRNA	SARS-CoV2 spike glycoprotein	Lipid nanoparticles	Vaccine	COVID-19	30,000	NCT04470427	Approved
	mRNA-1273.214	mRNA	SARS-CoV2 spike glycoprotein w/Omicron proteins	Lipid nanoparticles	Vaccine	COVID-19	5,158	NCT04927065	Approved
	ALN-RSV01	siRNA	RSV nucleocapsid (N) protein's mRNA	Lipid nanoparticles	Nasally	RSV	88	NCT00496821	II
	ALN-RSV01	siRNA	RSV nucleocapsid (N) protein's mRNA	Lipid nanoparticles	Aerosolized	Bronchiolitis obliterans syndrome (BOS) in recipients of lung transplant	87	NCT01065935	IIa
Cancer									
	Aprinocarsen	ASO	Inhibits PKC-alpha	-	Orally	NSCLC	14		I
	Aprinocarsen	ASO	Inhibits PKC-alpha	-	Orally	NSCLC	55	NCT00034268	II
	Aprinocarsen	ASO	Inhibits PKC-alpha	-	Orally	NSCLC	880	NCT00034268	III - Terminated
	CV9201	mRNA	Five tumor-associated antigens in non-small cell lung cancer patients (New York esophageal squamous cell carcinoma-1, melanoma antigen family C1/C2, survivin, and trophoblast glycoprotein)	Protamine complexation	Intratumoral Injections	NSCLC	46	NCT00923312	II/IIa
	G3139	ASO	Bcl-2	-	Intravenously	SCLC	16	NCT00005032	I/II
	Atu027	siRNA	PKN3	siRNA lipolex	Intravenously	Advanced solid tumors	27	NCT00938574	I
	TargomiRs	miRNA	EGFR	Minicells	Intravenously	Malignant Pleural Mesothelioma	26	NCT02369198	I
Airways Disease									
	QR-010	ASO	CFTR	-	Nasally	Cystic Fibrosis	18	NCT02564354	I
	QR-011	ASO	CFTR	-	Nasally	Cystic Fibrosis	70	NCT02532764	Ib
	MRT5005	mRNA	CFTR	Lipid nanoparticles	Nasally	Cystic Fibrosis	40	NCT03375047	I/II
	TPI ASM8	ASO	CCR3 & βc	-	Nasally	Asthma	16	NCT01158898	II
	AIR645	ASO	IL-4/IL-13 Receptor (α chain)	-	Nebulization	Asthma	80	NCT00658749	I

10.3389/fgene.2023.1281538

group (HMG) peptide ligands can target activated alveolar macrophages (Choi et al., 2020).

3.4 Hybrid delivery systems

Hybrid systems that contain two or more delivery vectors can benefit from the properties of multiple classes of delivery vehicles. These include lipid-polymer hybrids, lipid-peptide hybrids and polymer-peptide hybrids, amongst other possibilities (Chow et al., 2020). Lipid-polymer nanoparticles can facilitate enhanced lung retention for siRNA *in vivo* (Thanki et al., 2019).

Layer-by-layer assembly of hybrid delivery systems allows the development of purpose-built solutions for special applications such as pleural disease. Surface-fill hydrogel nanocomposite is a materials platform developed to deliver microRNA to complex anatomic surfaces locally. SFH requires two polymers prepared in a two-stage process (Majumder et al., 2021). The first stage involves the assembly of a novel peptide-based miRNA nanoparticle whereby an intrinsically disordered cationic peptide is complexed with double-stranded miRNA mimics with a chemically modified anionic backbone. In the second stage, these particles are encapsulated into a shear-thinning, self-assembling hydrogel. This surface-fill hydrogel (SFH) can be sprayed or injected onto large, anatomically complex surfaces such as the pleura to deliver RNA (Choi et al., 2022).

A key feature of SFH is that it can fill small gaps and crevices during application and change its shape after it has been applied. After application, the positively charged peptide-miRNA nanoparticles (~150 nm diameter) are released from the net-positively charged hydrogel matrix to the adjacent tissues. The design of the peptide-miRNA nanoparticle, including its surface charge state, size, miRNA encapsulation efficiency, and conformation, are important factors in clathrin-mediated cell entry and endosomal trafficking. This platform has been applied in models of malignant pleural mesothelioma (MPM), an aggressive cancer that arises from the mesothelial surface of the pleural cavities (Cho et al., 2021). There is a high rate of recurrence of local disease because MPM covers the pleural surface in a sheet-like manner that technically precludes tumour-free margins, and systemic therapies cannot access sites of residual disease (Baldini et al., 2015). SFH nanocomposites containing either miR-215 or miR-206 mimics reduce rates of local recurrence in pre-clinical models of MPM without detectable miRNA in circulating plasma or observable delivery of miRNA to off-target organs (Singh et al., 2019; 2021; Majumder et al., 2021; Choi et al., 2022).

4 Clinical trials in lung disease

With progress and promise shown in pre-clinical studies, RNA-based therapies are now being explored in clinical trials across several domains of lung disease: infectious disease, cancer, and airway disease (Table 1). RNA payloads, including mRNA, miRNA, siRNA and ASO, have been delivered in these trials. Beyond mRNA vaccines for SARS-CoV-2, RNA-based therapies have yet to be approved for lung disease, and there are few ongoing clinical trials in this growing field (Table 2). These trials suggest acceptable tolerability and toxicity profiles across classes of RNA-based therapy, highlight the ability to target specific genes and reflect the biological challenge of treating complex polygenic disease.

4.1 Infectious disease

The most successful application of RNA-based therapy in lung disease has been the adoption of mRNA vaccines for SARS-CoV-2 and the COVID-19 pandemic (Hall et al., 2021; Zhang et al., 2022). mRNA vaccines were one of three major vaccine platforms for SARS-CoV-2, alongside vaccines based on viral vector and recombinant protein plus adjuvant (Pollard and Bijker, 2021). Two mRNA vaccines, BNT162b2 (Pfizer) and mRNA-1273 (Moderna) were rapidly developed (Walsh et al., 2020) and demonstrated remarkable effectiveness, with early vaccine efficacy of 95% for BNT162b2 (Thomas et al., 2021) and 94% for mRNA-1273 (Baden et al., 2021). However, these numbers decline to 67% and 75%, respectively, at 5–7 months (Rosenberg et al., 2022). The pivotal trial that contributed to emergency FDA approval of BNT162b2 enrolled 43,448 participants aged 16 and older, along with 2,264 participants aged 12–15 (Polack et al., 2020). Both vaccines are encapsulated in lipid nanoparticles, code for the spike glycoprotein (Baden et al., 2021), and elicit $CD4^+$ and $CD8^+$ T-cell responses (Jackson et al., 2020; Zhang et al., 2022). The common side effects observed were injection site pain, headaches, fatigue, fever, chills, muscle pain, nausea, and vomiting (Baden et al., 2021).

Other anti-infective strategies aimed at treating infections have also been explored. ALN-RSV01 is an siRNA directed against the nucleocapsid (N) mRNA of the respiratory syncytial virus (RSV). A proof-of-concept randomized, double-blind, placebo-controlled trial compared saline nasal spray with a nasal spray of ALN-RSV01 2 days before and 3 days after experimental RSV infection in 88 healthy subjects and showed a 38% decrease in culture-defined RSV infections. ALN-RSV01 had a similar safety profile to placebo, and the antiviral effect was independent of pre-existing antibodies and intranasal cytokines (DeVincenzo et al., 2010). RSV infection is associated with an increased incidence of bronchiolitis obliterans syndrome (BOS) in lung transplant recipients. ALN-RSV01 has also been evaluated in a phase IIb randomized, double-blind, placebo-controlled trial of 87 lung transplant recipients. There was a trend toward a decrease in new or progressive BOS in the modified intention-to-treat analysis, which was significant in the per-protocol cohort (Gottlieb J et al., 2016). These results show potential for siRNA-therapeutics in the treatment of lung infection.

4.2 Cancer

In thoracic cancers, a range of RNA-based approaches, from antisense therapies to cancer vaccines, have been assessed in clinical trials. Early reports of RNA-based therapy in lung malignancy focused on ASOs (Coudert et al., 2001). In an early study of the anti-c-raf ASO, ISIS-5132, as monotherapy, toxicity was considered acceptable, but there was no clinical benefit seen in 18 patients with non-small cell lung cancer (NSCLC) (Coudert et al., 2001). A follow-

TABLE 2 Overview of ongoing clinical trials investigating RNA therapies in lung disease.

Type of study	Name	Treatment	Genetic/ protein target	Delivery vehicle	Administration method	Disease	Number of patients	ClinicalTrials.gov identifier	Phase
Active & Recruiting									
	VX-522	mRNA	CFTR	Lipid nanoparticle	Oral inhalation with nebulizer	Cystic Fibrosis	Recruiting	NCT05668741	I
	ARCT-032	mRNA	CFTR	Lipid nanoparticle	Oral inhalation with nebulizer	Cystic Fibrosis	Recruiting	NCT05712538	I
	mRNA personalized tumor vaccine	mRNA	Neoantigens	-	Vaccine	NSCLC	Recruiting	NCT03908671	-
Active									
	RSVpreF + BNTb162b2	mRNA	SARS-CoV2 spike glycoprotein and RSV prefusion-F protein	Lipid nanoparticle	Vaccine	COVID-19 and RSV	1,079	NCT05886777	II
	mRNA-1345	mRNA	Stabilized prefusion F glycoprotein	Lipid nanoparticle	Vaccine	RSV	36,557	NCT05127434	II/III
	CL-0059 or CL-0137	mRNA	Stabilized prefusion F glycoprotein	Lipid nanoparticle	Vaccine	RSV	790	NCT05639894	I/II
	mRNA-1345	mRNA	Stabilized prefusion F glycoprotein	Lipid nanoparticle	Vaccine	RSV	651	NCT04528719	I
	mRNA-1045 or mRNA-1230	mRNA	Influenza virus proteins, SARS-CoV2 spike glycoprotein and RSV prefusion-F protein	Lipid nanoparticle	Vaccine	Influenza, RSV, SARS-CoV-2	392	NCT05585632	I

up phase I trial showed that a combination of paclitaxel, carboplatin, and ISIS-5132 was well tolerated but showed no objective responses in 13 patients (Fidias et al., 2009). Subsequently, custirsen, an ASO targeting the gene clusterin (CLU), showed promise in Stage IIIB/IV non-small cell lung cancer (NSCLC) in a Phase I clinical trial for use in combination with cisplatin and gemcitabine (Laskin et al., 2012). Several trials have evaluated aprinocarsen, an ASO directed against protein kinase C-alpha, in combination with chemotherapy in advanced NSCLC (Vansteenkiste et al., 2005; Paz-Ares et al., 2006; Ritch et al., 2006). In a large Phase III trial of 670 patients that was stopped early due to results of another trial, aprinocarsen did not enhance survival compared to gemcitabine and cisplatin with advanced NSCLC (Paz-Ares et al., 2006). Despite early challenges, other ASOs have been evaluated over a decade later. Apatorsen (OGX-427) is an ASO directed against heat shock protein 27 (Hsp27) mRNA and was evaluated in a randomized, double-blinded, Phase II trial of 155 patients in combination with carboplatin and pemetrexed. Apartorsen was well tolerated but did not improve outcomes in patients with metastatic nonsquamous NSCLC as first-line treatment (Spigel et al., 2019).

Other groups have assessed ASOs in small cell lung cancer (SCLC). G3139 (oblimersen) is an ASO directed against Bcl-2 and underwent a pilot clinical trial (Rudin et al., 2002), and a subsequent Phase I trial, both of which showed tolerability and potential for treatment response (Rudin et al., 2004). Despite this early success, a Phase II trial of oblimersen in combination with carboplatin and etoposide showed worse outcomes in the group with oblimersen, despite evidence supporting a critical role for Bcl-2 in treatment resistance SCLC (Rudin et al., 2008). Inadequate suppression of Bcl-2 *in vivo* was proposed as a reason for the lack of efficacy.

The first clinical experience for miRNA-based therapy in thoracic malignancy was the assessment of mIR-16-based miRNA mimics in malignant pleural mesothelioma (MPM). A Phase I dose escalation open-label trial assessed TargomiRs, a term describing proprietary miRNA mimics based on the miR-16 seed sequence that are encapsulated in a bacterial-derived mini-cell vector targeted to the epidermal growth factor receptor (EGFR) by panitumumab-based antibodies. Of 22 patients who were assessed for response, 1 achieved an objective response to therapy by CT (van Zandwijk N et al., 2017).

Cancer immunotherapy is designed to elicit an immune response to tumour antigens. In stage IIIB/IV NSCLC, a phase I/IIa trial assessed the safety of CV9201, an mRNA-based immunotherapy encoding five NSCLC tumour antigens (New

York esophageal squamous cell carcinoma-1, melanoma antigen family C1/C2, survivin, and trophoblast glycoprotein) (Sebastian M et al., 2019). This application highlights the potential for multiplexing drug delivery for RNA-based therapies. Promise for this therapeutic avenue was shown as no serious adverse events occurred, and immune responses were detected after treatment.

4.3 Airways disease

Most approved RNA-based therapies target monogenic disorders (Egli and Manoharan, 2023). In the lung, cystic fibrosis (CF) represents an attractive target for the design of RNA-based therapeutics, with mutations of the cystic fibrosis transmembrane receptor (CFTR) being the causative factor in the development of the disease (Riordan et al., 1989; Ong and Ramsey, 2023). Eluforsen (QR-010) is an ASO designed to bind CFTR mRNA around the ΔF508 mutation and lead to the translation of wild-type CFTR protein (Sermet-Gaudelus et al., 2019). An open-label study of intranasal eluforsen three times weekly for 4 weeks showed improved chloride transport in a ΔF508 homozygous cohort (Sermet-Gaudelus et al., 2019) and was followed by a randomized, double-blind, placebo-controlled dose escalation Phase Ib trial of adult CF found that eluforsen was well tolerated, with low systemic exposure and improvement in the Cystic Fibrosis Questionnaire-Revised Respiratory Symptom Score after 4 weeks (Drevinek et al., 2020).

In theory, mRNA coding for CFTR could be helpful with any CFTR mutation. MRT5005, a codon-optimized CFTR mRNA delivered by inhaled lipid nanoparticles, is being evaluated in a randomized, double-blind, placebo-controlled Phase I/II study. Interim analysis of 42 subjects found 14 febrile reactions over 28 days, most of which resolved within 1–2 days and allowed ongoing treatment. FEV1 remained stable, but no beneficial effects on FEV1 were observed (Rowe et al., 2023).

ASOs have also been used to target specific factors contributing to airways inflammation in asthma (Gauvreau et al., 2011). Early development of an A1 adenosine receptor ASO, EPI-2010, showed modest efficacy in mild asthma, but product development was discontinued without additional benefit in a Phase II trial (Ball et al., 2003; Liao et al., 2017). TPI-ASM8 contains two ASOs that, together, target the beta subunit of IL-3, IL-5 and GM-CSF receptors and the chemokine receptor CCR3. In a single-center, open-label, ascending dose study, inhaled TPI-ASM9 was well tolerated in stable asthma, and decreased airway responsiveness to methacholine challenge and decreased sputum eosinophils in response to inhalational challenge (Gauvreau et al., 2008; Gauvreau et al., 2011).

5 Future directions

The evolution of chemistry for RNA-based therapeutic payloads and for nanotechnology-based delivery vehicles have advanced the applicability of RNA-based treatments to human disease. By leveraging the ability to selectively target different compartments within the lungs (e.g., airways, alveoli, pleural cavity) and specific cell types (e.g., epithelium, endothelium, immune cell), RNA therapy offers enhanced precision in targeting lung disease (Paunovska et al., 2022). Universal delivery platforms could accelerate the translation of biological findings (Chow et al., 2020), whereas made-for-application platforms could provide the most effective means of drug delivery and targeting. The ability to rapidly develop new therapeutics based on nucleotide sequence reduces the duration of drug development, such that deconvoluting disease biology poses a greater challenge than druggability.

5.1 Airways disease

In obstructive airway diseases, such as asthma and chronic obstructive pulmonary disease (COPD), treatment delivered by inhalation can enhance local delivery while decreasing systemic exposure (Chow et al., 2020). Naked RNA can transfect lung epithelium, though delivery vehicles that enhance transfection efficiency are still a preferred route (Chow et al., 2020). Disease-specific formulations may be important considerations as fluid and pulmonary surfactants can influence drug delivery (Hidalgo et al., 2015), and assessment of fluids may become a feature of individualized therapy. Through efficient delivery methods, it becomes possible to modulate gene expression, reduce inflammation, and improve lung function in patients with airway diseases with absent or limited systemic administration (Patel et al., 2019). Activated T-helper cells subtype 2 (TH2), which contribute to asthma, have been targeted by conjugated polymer nanoparticles in a dry powder formulation. Anti-GATA3 siRNA can be delivered via this system to reduce the production of multiple cytokines en masse rather than blocking or downregulating single cytokines (Keil et al., 2020). As single-cell technologies improve the resolution of disease pathogenesis (Adams et al., 2023; Lim et al., 2023), these targeted approaches will provide therapeutic options that more closely address pathologic processes and cell types while reducing side effects.

5.2 Lung transplantation

Lung transplant recipients face the challenge of maintaining graft function and preventing rejection. Specific intragraft immune cell populations are of prognostic and likely mechanistic importance in the development of chronic lung allograft dysfunction (Bos et al., 2022; Beber et al., 2023). Here, as with asthma, RNA therapy using specific delivery vehicles can play a crucial role by specifically targeting immune cells in the lungs (Ke et al., 2020). Through the development of precise delivery methods that home in on these immune cells, it becomes possible to modulate local immune responses and enhance long-term transplant success rates.

5.3 Pulmonary vascular disease

Local delivery of RNA-based therapy to the lung via inhalation or to the pleural space are attractive options, yet RNA delivered via these routes may not effectively reach the pulmonary endothelium to benefit pulmonary vascular disorders. However, delivery vehicles have been engineered to target lung endothelium, even with systemic

delivery (Dahlman et al., 2014; Khan et al., 2018; Cheng et al., 2020). Loss-of-function mutations in pulmonary arterial hypertension (PAH), such as BMPR2 (Cuthbertson et al., 2023), could be candidates for mRNA-based therapies.

5.4 Critical care

Additionally, critical care conditions like acute respiratory distress syndrome (ARDS) require comprehensive treatment approaches that address both local lung inflammation and systemic consequences. RNA therapy delivered systemically holds promise in treating ARDS (Allegra et al., 2023) by targeting inflammatory pathways locally within the lungs and systemically. In single-cell RNA sequencing of severe COVID-19, loss of a long noncoding RNA (lncRNA) PIRAT in blood monocytes leads to increased inflammatory response by alarmins at the expense of anti-viral response via the JAK-STAT pathway (Aznaourova et al., 2022). These high-resolution insights, facilitated by single-cell technologies, suggest settings which require precise therapeutic targeting that is possible with RNA-based approaches.

5.5 Interstitial lung disease

Recently, the treatment of interstitial lung disease (ILD) has been driven by the recognition of progressive fibrosis as a feature of several ILDs. However, multiple clinical, radiographic, and histologic patterns exist, reflecting distinct pathologic processes (Renzoni et al., 2021). Multiple contributing cell types and gene targets, including miRNAs, are being identified and could lead to a molecular classification of ILD entities (Renzoni et al., 2021). These insights suggest potential benefits for an array of RNA payloads and delivery vehicles to target relevant pathogenic processes and alter functional trajectory in ILD.

5.6 Long noncoding RNAs as targets for RNA-based therapy in the lung

Noncoding RNAs have emerged as major products of the human genome and are involved in all cellular functions. Many classes of noncoding RNA exist, many of which have specific functions and mechanisms, such as miRNAs. Long noncoding RNAs (lncRNAs) have traditionally been defined as noncoding transcripts of more than 200 nucleotides, though a new consensus statement proposes that they be defined as >500 nt to better distinguish from other noncoding RNAs (Mattick et al., 2023). Notably, lncRNAs are likely the major output of genomes in complex organisms and annotated human lncRNAs in the GENCODE database number 19,928 - a number greater than the number of annotated protein-coding genes (GENCODE, 2022). LncRNAs are attractive therapeutic targets as they generally have more specific expression patterns than protein-coding genes (Derrien et al., 2012).

Emerging evidence is defining roles for lncRNAs in lung development, homeostasis and disease. For example, the lncRNA LL18/NANCI contributes to saccule formation/alveolarization in mouse lungs (Herriges and Morrisey, 2014). In hypoxia, the lncRNA GATA2-AS1 was shown to modulate HIF-1α and HIF-2α balance and hypoxia signalling in vascular endothelium (Man et al., 2023). The lncRNA TYKRIL is overexpressed in hypoxia and idiopathic pulmonary arterial hypertension (iPAH) and contributes to a pro-proliferative and anti-apoptotic phenotype in pulmonary artery smooth muscle cells and pericytes (Zehendner et al., 2020). Similarly, the lncRNA SMILR also contributes to smooth muscle proliferation in hypoxic conditions and can be detected in the serum of PAH patients (Lei et al., 2020). These examples illustrate the potential for lncRNAs as specific and effective targets of therapy in lung disease through RNA-based therapies.

6 Conclusion

In conclusion, the future of RNA therapy in lung diseases is promising and offers new avenues for effective treatment. The ability to target specific compartments within the lungs, in combination with various RNA payloads, provides versatile options for precise and personalized therapy. We have briefly overviewed various RNA payloads, delivery vehicles and the clinical state of RNA-based therapy in the lung. Few clinical trials have been completed, and vaccines directed at SARS-CoV-2 are the only FDA-approved treatments. Thus, continued research and development in this field are crucial to fully realize the transformative potential of RNA therapy in the management of lung diseases.

Author contributions

HM: Conceptualization, Project administration, Writing–original draft, Writing–review and editing. VM: Writing–original draft, Writing–review and editing. AS: Writing–review and editing. LW: Writing–review and editing. JG: Writing–review and editing. SJ: Writing–review and editing. CH: Funding acquisition, Supervision, Writing-review and editing. MP: Conceptualization, Funding acquisition, Writing–review and editing, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. HM is supported by a Precision Medicine (PRiME) Clinical Catalyst Fellowship. This work was supported by the NIH Intramural Research Program with funding (ZIA BC 011657) provided to CH. This work was supported by the Mesothelioma Applied Research Foundation, the Princess Margaret Cancer Foundation, and the University Health Network (UHN) Foundation to MP. MP is the recipient of the Canadian Mesothelioma Foundation Professorship in Mesothelioma Research.

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

Adams, B. D., Parsons, C., Walker, L., Zhang, W. C., and Slack, F. J. (2017). Targeting noncoding RNAs in disease. *J. Clin. Invest.* 127, 761–771. doi:10.1172/JCI84424

Adams, T. S., Marlier, A., and Kaminski, N. (2023). Lung cell atlases in health and disease. *Annu. Rev. Physiol.* 85, 47–69. doi:10.1146/annurev-physiol-032922-082826

Allegra, A., Cicero, N., Mirabile, G., Giorgianni, C. M., and Gangemi, S. (2023). Novel biomarkers for diagnosis and monitoring of immune thrombocytopenia. *Int. J. Mol. Sci.* 24, 4438. doi:10.3390/ijms24054438

Anselmo, A. C., Zhang, M., Kumar, S., Vogus, D. R., Menegatti, S., Helgeson, M. E., et al. (2015). Elasticity of nanoparticles influences their blood circulation, phagocytosis, endocytosis, and targeting. *ACS Nano* 9, 3169–3177. doi:10.1021/acsnano.5b00147

Aznaourova, M., Schmerer, N., Janga, H., Zhang, Z., Pauck, K., Bushe, J., et al. (2022). Single-cell RNA sequencing uncovers the nuclear decoy lincRNA PIRAT as a regulator of systemic monocyte immunity during COVID-19. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2120680119. doi:10.1073/pnas.2120680119

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., et al. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* 384, 403–416. doi:10.1056/NEJMoa2035389

Bai, D., Jeff Crosby, C. Z., Zhang, M., and Aaron, D. (2019). Pulmonary delivery of NLRP3 antisense oligonucleotides are effective at preventing but not reversing established bleomycin-induced pulmonary fibrosis in mice. *Am. J. Respir. Crit. Care Med.* 199, A4605. doi:10.1164/ajrccm-conference.2019.199.1_MeetingAbstracts.A4605

Baldini, E. H., Richards, W. G., Gill, R. R., Goodman, B. M., Winfrey, O. K., Eisen, H. M., et al. (2015). Updated patterns of failure after multimodality therapy for malignant pleural mesothelioma. *J. Thorac. Cardiovasc. Surg.* 149, 1374–1381. doi:10.1016/j.jtcvs.2014.10.128

Ball, H. A., Sandrasagra, A., Tang, L., Van Scott, M., Wild, J., and Nyce, J. W. (2003). Clinical potential of respirable antisense oligonucleotides (RASONs) in asthma. *Am. J. Pharmacogenomics* 3, 97–106. doi:10.2165/00129785-200303020-00003

Bandyopadhyay, D., Qureshi, A., Ghosh, S., Ashish, K., Heise, L. R., Hajra, A., et al. (2018). Safety and efficacy of extremely low LDL-cholesterol levels and its prospects in hyperlipidemia management. *J. Lipids* 2018, 8598054. doi:10.1155/2018/8598054

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi:10.1016/s0092-8674(04)00045-5

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi:10.1016/j.cell.2009.01.002

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell* 173, 20–51. doi:10.1016/j.cell.2018.03.006

Beber, S., Moshkelgosha, S., Cheung, M., Hedley, D., Levy, L., Samuels, J., et al. (2023). (1265) exploration of intragraft T cell phenotypes in minimal acute cellular rejection (ACR) using imaging mass cytometry (IMC). *J. Heart Lung Transpl.* 42, S540. doi:10.1016/j.healun.2023.02.1475

Bennett, C. F., Baker, B. F., Pham, N., Swayze, E., and Geary, R. S. (2017). Pharmacology of antisense drugs. *Annu. Rev. Pharmacol. Toxicol.* 57, 81–105. doi:10.1146/annurev-pharmtox-010716-104846

Bennett, C. F., and Swayze, E. E. (2010). RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* 50, 259–293. doi:10.1146/annurev.pharmtox.010909.105654

Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., et al. (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* 3, 199–204. doi:10.1038/nmeth854

Bos, S., Filby, A. J., Vos, R., and Fisher, A. J. (2022). Effector immune cells in chronic lung allograft dysfunction: a systematic review. *Immunology* 166, 17–37. doi:10.1111/imm.13458

Bottorff, T. A., Park, H., Geballe, A. P., and Subramaniam, A. R. (2022). Translational buffering by ribosome stalling in upstream open reading frames. *PLOS Genet.* 18, e1010460. doi:10.1371/journal.pgen.1010460

Boudreau, R., Rodríguez-Lebrón, E., and Davidson, B. (2011). RNAi medicine for the brain: progresses and challenges. *Hum. Mol. Genet.* 20, R21–R27. doi:10.1093/hmg/ddr137

Calvo, S. E., Pagliarini, D. J., and Mootha, V. K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7507–7512. doi:10.1073/pnas.0810916106

Carrer, M., Crosby, J. R., Sun, G., Zhao, C., Damle, S. S., Kuntz, S. G., et al. (2020). Antisense oligonucleotides targeting jagged 1 reduce house dust mite-induced goblet cell metaplasia in the adult murine lung. *Am. J. Respir. Cell Mol. Biol.* 63, 46–56. doi:10.1165/rcmb.2019-0257OC

Carthew, R. W., and Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136, 642–655. doi:10.1016/j.cell.2009.01.035

Chandra Ghosh, G., Bandyopadhyay, D., Ghosh, R. K., Mondal, S., and Herzog, E. (2018). Effectiveness and safety of inclisiran, A novel long-acting RNA therapeutic inhibitor of proprotein convertase subtilisin/kexin 9. *Am. J. Cardiol.* 122, 1272–1277. doi:10.1016/j.amjcard.2018.06.023

Chen, F., Alphonse, M., and Liu, Q. (2020). Strategies for nonviral nanoparticle-based delivery of CRISPR/Cas9 therapeutics. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* 12, e1609. doi:10.1002/wnan.1609

Chen, Y., Qi, F., Gao, F., Cao, H., Xu, D., Salehi-Ashtiani, K., et al. (2021). Hovlinc is a recently evolved class of ribozyme found in human lncRNA. *Nat. Chem. Biol.* 17, 601–607. doi:10.1038/s41589-021-00763-0

Chen, Y., and Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 48 (D1), D127–D131.

Cheng, C. J., Tietjen, G. T., Saucier-Sawyer, J. K., and Saltzman, W. M. (2015). A holistic approach to targeting disease with polymeric nanoparticles. *Nat. Rev. Drug Discov.* 14, 239–247. doi:10.1038/nrd4503

Cheng, Q., Wei, T., Farbiak, L., Johnson, L. T., Dilliard, S. A., and Siegwart, D. J. (2020). Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing. *Nat. Nanotechnol.* 15, 313–320. doi:10.1038/s41565-020-0669-6

Chioccioli, M., Roy, S., Newell, R., Pestano, L., Dickinson, B., Rigby, K., et al. (2022). A lung targeted miR-29 mimic as a therapy for pulmonary fibrosis. *eBioMedicine* 85, 104304. doi:10.1016/j.ebiom.2022.104304

Cho, B. C. J., Donahoe, L., Bradbury, P. A., Leighl, N., Keshavjee, S., Hope, A., et al. (2021). Surgery for malignant pleural mesothelioma after radiotherapy (SMART): final results from a single-centre, phase 2 trial. *Lancet Oncol.* 22, 190–197. doi:10.1016/S1470-2045(20)30606-9

Choi, A. Y., Singh, A., Wang, D., Pittala, K., and Hoang, C. D. (2022). Current state of pleural-directed adjuncts against malignant pleural mesothelioma. *Front. Oncol.* 12, 886430. doi:10.3389/fonc.2022.886430

Choi, M., Jeong, H., Kim, S., Kim, M., Lee, M., and Rhim, T. (2020). Targeted delivery of Chil3/Chil4 siRNA to alveolar macrophages using ternary complexes composed of HMG and oligoarginine micelles. *Nanoscale* 12, 933–943. doi:10.1039/c9nr06382j

Chow, M. Y. T., Qiu, Y., and Lam, J. K. W. (2020). Inhaled RNA therapy: from promise to reality. *Trends Pharmacol. Sci.* 41, 715–729. doi:10.1016/j.tips.2020.08.002

Comegna, M., Conte, G., Falanga, A. P., Marzano, M., Cernera, G., Di Lullo, A. M., et al. (2021). Assisting PNA transport through cystic fibrosis human airway epithelia with biodegradable hybrid lipid-polymer nanoparticles. *Sci. Rep.* 11, 6393. doi:10.1038/s41598-021-85549-z

Coudert, B., Anthoney, A., Fiedler, W., Droz, J. P., Dieras, V., Borner, M., et al. (2001). Phase II trial with ISIS 5132 in patients with small-cell (SCLC) and non-small cell (NSCLC) lung cancer. A European organization for research and treatment of cancer (eortc) early clinical studies group report. *Eur. J. Cancer Oxf. Engl.* 37, 2194–2198. doi:10.1016/s0959-8049(01)00286-6

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi:10.1038/227561a0

Crooke, S. T., Liang, X.-H., Baker, B. F., and Crooke, R. M. (2021). Antisense technology: a review. *J. Biol. Chem.* 296, 100416. doi:10.1016/j.jbc.2021.100416

Crooke, S. T., Wang, S., Vickers, T. A., Shen, W., and Liang, X.-H. (2017). Cellular uptake and trafficking of antisense oligonucleotides. *Nat. Biotechnol.* 35, 230–237. doi:10.1038/nbt.3779

Crooke, S. T., Witztum, J. L., Bennett, C. F., and Baker, B. F. (2018). RNA-targeted therapeutics. *Cell Metab.* 27, 714–739. doi:10.1016/j.cmet.2018.03.004

Cushing, L., Jiang, Z., Kuang, P., and Lü, J. (2015). The roles of microRNAs and protein components of the microRNA pathway in lung development and diseases. *Am. J. Respir. Cell Mol. Biol.* 52, 397–408. doi:10.1165/rcmb.2014-0232RT

Cuthbertson, I., Morrell, N. W., and Caruso, P. (2023). BMPR2 mutation and metabolic reprogramming in pulmonary arterial hypertension. *Circ. Res.* 132, 109–126. doi:10.1161/CIRCRESAHA.122.321554

Dahlman, J. E., Barnes, C., Khan, O. F., Thiriot, A., Jhunjunwala, S., Shaw, T. E., et al. (2014). *In vivo* endothelial siRNA delivery using polymeric nanoparticles with low molecular weight. *Nat. Nanotechnol.* 9, 648–655. doi:10.1038/nnano.2014.84

Damase, T. R., Sukhovershin, R., Boada, C., Taraballi, F., Pettigrew, R. I., and Cooke, J. P. (2021). The limitless future of RNA therapeutics. *Front. Bioeng. Biotechnol.* 9, 628137. doi:10.3389/fbioe.2021.628137

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–W383. doi:10.1093/nar/gkm216

de Jesus, S. F., Santos, L. I., Rodrigues Neto, J. F., Vieira, T. M., Mendes, J. B., D'angelo, M. F. S. V., et al. (2021). Therapeutic perceptions in antisense RNA-mediated gene regulation for COVID-19. *Gene* 800, 145839. doi:10.1016/j.gene.2021.145839

Deng, J., Shi, Y., Peng, X., He, Y., Chen, X., Li, M., et al. (2023). Ribocentre: a database of ribozymes. *Nucleic Acids Res.* 51, D262–D268. doi:10.1093/nar/gkac840

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi:10.1101/gr.132159.111

DeVincenzo, J., Lambkin-Williams, R., Wilkinson, T., Cehelsky, J., Nochur, S., Walsh, E., et al. (2010). A randomized, double-blind, placebo-controlled study of an RNAi-based therapy directed against respiratory syncytial virus. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8800–8805. doi:10.1073/pnas.0912186107

DeWeerdt, S. (2019). RNA therapies explained. *Nature* 574, S2–S3. doi:10.1038/d41586-019-03068-4

Di Fusco, D., Dinallo, V., Marafini, I., Figliuzzi, M. M., Romano, B., and Monteleone, G. (2019). Antisense oligonucleotide: basic concepts and therapeutic application in inflammatory bowel disease. *Front. Pharmacol.* 10, 305. doi:10.3389/fphar.2019.00305

Doench, J. G., and Sharp, P. A. (2004). Specificity of microRNA target selection in translational repression. *Genes Dev.* 18, 504–511. doi:10.1101/gad.1184404

Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096. doi:10.1126/science.1258096

Drevinek, P., Pressler, T., Cipolli, M., Boeck, K. D., Schwarz, C., Bouisset, F., et al. (2020). Antisense oligonucleotide eluforsen is safe and improves respiratory symptoms in F508DEL cystic fibrosis. *J. Cyst. Fibros.* 19, 99–107. doi:10.1016/j.jcf.2019.05.014

Ebert, M. S., Neilson, J. R., and Sharp, P. A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods* 4, 721–726. doi:10.1038/nmeth1079

Egli, M., and Manoharan, M. (2023). Chemistry, structure and function of approved oligonucleotide therapeutics. *Nucleic Acids Res.* 51, 2529–2573. doi:10.1093/nar/gkad067

Farokhzad, O. C., and Langer, R. (2009). Impact of nanotechnology on drug delivery. *ACS Nano* 3, 16–20. doi:10.1021/nn900002m

Ferguson, C. M., Echeverria, D., Hassler, M., Ly, S., and Khvorova, A. (2020). Cell type impacts accessibility of mRNA to silencing by RNA interference. *Mol. Ther. Nucleic Acids* 21, 384–393. doi:10.1016/j.omtn.2020.06.006

Fidias, P., Pennell, N. A., Boral, A. L., Shapiro, G. I., Skarin, A. T., Eder, J. P., et al. (2009). Phase I study of the c-raf-1 antisense oligonucleotide ISIS 5132 in combination with carboplatin and paclitaxel in patients with previously untreated, advanced non-small cell lung cancer. *J. Thorac. Oncol.* 4, 1156–1162. doi:10.1097/JTO.0b013e3181b2793f

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811. doi:10.1038/35888

Frederickson, R., and Herzog, R. W. (2021). RNA-based vaccines and innate immune activation: not too hot and not too cold. *Mol. Ther.* 29, 1365–1366. doi:10.1016/j.ymthe.2021.03.005

Gaus, H. J., Gupta, R., Chappell, A. E., Østergaard, M. E., Swayze, E. E., and Seth, P. P. (2019). Characterization of the interactions of chemically-modified therapeutic nucleic acids with plasma proteins using a fluorescence polarization assay. *Nucleic Acids Res.* 47, 1110–1122. doi:10.1093/nar/gky1260

Gauvreau, G. M., Boulet, L. P., Cockcroft, D. W., Baatjes, A., Cote, J., Deschesnes, F., et al. (2008). Antisense therapy against CCR3 and the common beta chain attenuates allergen-induced eosinophilic responses. *Am. J. Respir. Crit. Care Med.* 177, 952–958. doi:10.1164/rccm.200708-1251OC

Gauvreau, G. M., Pageau, R., Séguin, R., Carballo, D., Gauthier, J., D'Anjou, H., et al. (2011). Dose–response effects of TPI ASM8 in asthmatics after allergen. *Allergy* 66, 1242–1248. doi:10.1111/j.1398-9995.2011.02638.x

GENCODE (2022). *Human release 43*. Available at: https://www.gencodegenes.org/human/ (Accessed June 6, 2023).

Gillmore, J. D., Gane, E., Taubel, J., Kao, J., Fontana, M., Maitland, M. L., et al. (2021). CRISPR-Cas9 *in vivo* gene editing for transthyretin amyloidosis. *N. Engl. J. Med.* 385, 493–502. doi:10.1056/NEJMoa2107454

Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199–202. doi:10.1038/nature04917

Gottlieb, J., Zamora, M. R., Hodges, T., Musk, A. W., Sommerwerk, U., Dilling, D., et al. (2016). ALN-RSV01 for prevention of bronchiolitis obliterans syndrome after respiratory syncytial virus infection in lung transplant recipients. *J. Heart Lung Transpl.* 35, 213–221. doi:10.1016/j.healun.2015.08.012

Grimm, D., Streetz, K. L., Jopling, C. L., Storm, T. A., Pandey, K., Davis, C. R., et al. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* 441, 537–541. doi:10.1038/nature04791

Gu, L., Deng, Z. J., Roy, S., and Hammond, P. T. (2017). A combination RNAi-chemotherapy layer-by-layer nanoparticle for systemic targeting of KRAS/P53 with cisplatin to treat non–small cell lung cancer. *Clin. Cancer Res.* 23, 7312–7323. doi:10.1158/1078-0432.CCR-16-2186

Guan, S., Munder, A., Hedtfeld, S., Braubach, P., Glage, S., Zhang, L., et al. (2019). Self-assembled peptide–poloxamine nanoparticles enable *in vitro* and *in vivo* genome restoration for cystic fibrosis. *Nat. Nanotechnol.* 14, 287–297. doi:10.1038/s41565-018-0358-x

Guerrier-Takada, C., and Altman, S. (1984). Catalytic activity of an RNA molecule prepared by transcription *in vitro*. *Science* 223, 285–286. doi:10.1126/science.6199841

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849–857. doi:10.1016/0092-8674(83)90117-4

Guttman, M., Garber, M., Levin, J. Z., Donaghey, J., Robinson, J., Adiconis, X., et al. (2010). *Ab initio* reconstruction of cell type–specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 28, 503–510. doi:10.1038/nbt.1633

Hall, V. J., Foulkes, S., Charlett, A., Atti, A., Monk, E. J. M., Simmons, R., et al. (2021). SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). *Lancet* 397, 1459–1469. doi:10.1016/S0140-6736(21)00675-9

Hammond, S. M., Aartsma-Rus, A., Alves, S., Borgos, S. E., Buijsen, R. A. M., Collin, R. W. J., et al. (2021). Delivery of oligonucleotide-based therapeutics: challenges and opportunities. *EMBO Mol. Med.* 13, e13243. doi:10.15252/emmm.202013243

Hanlon, K. S., Kleinstiver, B. P., Garcia, S. P., Zaborowski, M. P., Volak, A., Spirig, S. E., et al. (2019). High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat. Commun.* 10, 4439–4511. doi:10.1038/s41467-019-12449-2

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

Hidalgo, A., Cruz, A., and Pérez-Gil, J. (2015). Barrier or carrier? Pulmonary surfactant and drug delivery. *Eur. J. Pharm. Biopharm.* 95, 117–127. doi:10.1016/j.ejpb.2015.02.014

Ho, J. J. D., Man, J. H. S., Schatz, J. H., and Marsden, P. A. (2021). Translational remodeling by RNA-binding proteins and noncoding RNAs. *WIREs RNA* 12, e1647. doi:10.1002/wrna.1647

Ho, J. J. D., Robb, G. B., Tai, S. C., Turgeon, P. J., Mawji, I. A., Man, H. S. J., et al. (2013). Active stabilization of human endothelial nitric oxide synthase mRNA by hnRNP E1 protects against antisense RNA and MicroRNAs. *Mol. Cell. Biol.* 33, 2029–2046. doi:10.1128/MCB.01257-12

Hogan, M. J., and Pardi, N. (2022). mRNA vaccines in the COVID-19 pandemic and beyond. *Annu. Rev. Med.* 73, 17–39. doi:10.1146/annurev-med-042420-112725

Horita, M., Farquharson, C., and Stephen, L. A. (2021). The role of miR-29 family in disease. *J. Cell. Biochem.* 122, 696–715. doi:10.1002/jcb.29896

Hornung, V., Guenthner-Biller, M., Bourquin, C., Ablasser, A., Schlee, M., Uematsu, S., et al. (2005). Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat. Med.* 11, 263–270. doi:10.1038/nm1191

Jackson, A. L., Burchard, J., Schelter, J., Chau, B. N., Cleary, M., Lim, L., et al. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* 12, 1179–1187. doi:10.1261/rna.25706

Jackson, L. A., Anderson, E. J., Rouphael, N. G., Roberts, P. C., Makhene, M., Coler, R. N., et al. (2020). An mRNA vaccine against SARS-CoV-2 — preliminary report. *N. Engl. J. Med.* 383, 1920–1931. doi:10.1056/NEJMoa2022483

Judge, A. D., Sood, V., Shaw, J. R., Fang, D., McClintock, K., and MacLachlan, I. (2005). Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat. Biotechnol.* 23, 457–462. doi:10.1038/nbt1081

Juliano, R. L. (2016). The delivery of therapeutic oligonucleotides. *Nucleic Acids Res.* 44, 6518–6548. doi:10.1093/nar/gkw236

Kang, K.-N., and Lee, Y.-S. (2013). "RNA aptamers: a review of recent trends and applications," in *Future trends in biotechnology advances in biochemical engineering/biotechnology*. Editor J.-J. Zhong (Berlin, Heidelberg: Springer), 153–169. doi:10.1007/10_2012_136

Ke, X., Shelton, L., Hu, Y., Zhu, Y., Chow, E., Tang, H., et al. (2020). Surface-functionalized PEGylated nanoparticles deliver messenger RNA to pulmonary immune cells. *ACS Appl. Mat. Interfaces* 12, 35835–35844. doi:10.1021/acsami.0c08268

Kebriaei, P., Izsvák, Z., Narayanavari, S. A., Singh, H., and Ivics, Z. (2017). Gene therapy with the sleeping beauty transposon system. *Trends Genet. TIG* 33, 852–870. doi:10.1016/j.tig.2017.08.008

Keil, T. W. M., Baldassi, D., and Merkel, O. M. (2020). T-cell targeted pulmonary siRNA delivery for the treatment of asthma. *WIREs Nanomedicine Nanobiotechnology* 12, e1634. doi:10.1002/wnan.1634

Khan, A. A., Betel, D., Miller, M. L., Sander, C., Leslie, C. S., and Marks, D. S. (2009). Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat. Biotechnol.* 27, 549–555. doi:10.1038/nbt.1543

Khan, O. F., Kowalski, P. S., Doloff, J. C., Tsosie, J. K., Bakthavatchalu, V., Winn, C. B., et al. (2018). Endothelial siRNA delivery in nonhuman primates using ionizable low–molecular weight polymeric nanoparticles. *Sci. Adv.* 4, eaar8409. doi:10.1126/sciadv.aar8409

Kim, J., Jeon, S., Kang, S. J., Kim, K.-R., Thai, H. B. D., Lee, S., et al. (2020a). Lung-targeted delivery of TGF-β antisense oligonucleotides to treat pulmonary fibrosis. *J. Control. Release Off. J. Control. Release Soc.* 322, 108–121. doi:10.1016/j.jconrel.2020.03.016

Kim, N., Kim, H. K., Lee, K., Hong, Y., Cho, J. H., Choi, J. W., et al. (2020b). Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* 11, 2285. doi:10.1038/s41467-020-16164-1

Kim, Y.-K. (2022). RNA therapy: rich history, various applications and unlimited future prospects. *Exp. Mol. Med.* 54, 455–465. doi:10.1038/s12276-022-00757-5

Kojima, N., Turner, F., Slepnev, V., Bacelar, A., Deming, L., Kodeboyina, S., et al. (2021). Self-collected oral fluid and nasal swabs demonstrate comparable sensitivity to clinician collected nasopharyngeal swabs for coronavirus disease 2019 detection. *Clin. Infect. Dis.* 73, e3106–e3109. doi:10.1093/cid/ciaa1589

Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., and Hsu, P. D. (2018). Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* 173, 665–676. doi:10.1016/j.cell.2018.02.033

Kormann, M. S. D., Hasenpusch, G., Aneja, M. K., Nica, G., Flemmer, A. W., Herber-Jonat, S., et al. (2011). Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. *Nat. Biotechnol.* 29, 154–157. doi:10.1038/nbt.1733

Kranz, L. M., Diken, M., Haas, H., Kreiter, S., Loquai, C., Reuter, K. C., et al. (2016). Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* 534, 396–401. doi:10.1038/nature18300

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* 31, 147–157. doi:10.1016/0092-8674(82)90414-7

Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K. G., Tuschl, T., Manoharan, M., et al. (2005). Silencing of microRNAs *in vivo* with 'antagomirs. ' *Nat.* 438, 685–689. doi:10.1038/nature04303

Kuijper, E. C., Bergsma, A. J., Pijnappel, W. W. M. P., and Aartsma-Rus, A. (2021). Opportunities and challenges for antisense oligonucleotide therapies. *J. Inherit. Metab. Dis.* 44, 72–87. doi:10.1002/jimd.12251

Lam, J. K. W., Chow, M. Y. T., Zhang, Y., and Leung, S. W. S. (2015). siRNA versus miRNA as therapeutics for gene silencing. *Mol. Ther. Nucleic Acids* 4, e252. doi:10.1038/mtna.2015.23

Langer, R., and Folkman, J. (1976). Polymers for the sustained release of proteins and other macromolecules. *Nature* 263, 797–800. doi:10.1038/263797a0

Laskin, J. J., Nicholas, G., Lee, C., Gitlitz, B., Vincent, M., Cormier, Y., et al. (2012). Phase I/II trial of custirsen (OGX-011), an inhibitor of clusterin, in combination with a gemcitabine and platinum regimen in patients with previously untreated advanced non-small cell lung cancer. *J. Thorac. Oncol.* 7, 579–586. doi:10.1097/JTO.0b013e31823f459c

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854. doi:10.1016/0092-8674(93)90529-y

Lei, B., Wang, C., Snow, K., Graton, M. E., Tighe, R. M., Fager, A. M., et al. (2023). Inhalation of an RNA aptamer that selectively binds extracellular histones protects from acute lung injury. *Mol. Ther. - Nucleic Acids* 31, 662–673. doi:10.1016/j.omtn.2023.02.021

Lei, S., Peng, F., Li, M.-L., Duan, W.-B., Peng, C.-Q., and Wu, S.-J. (2020). LncRNA-SMILR modulates RhoA/ROCK signaling by targeting miR-141 to regulate vascular remodeling in pulmonary arterial hypertension. *Am. J. Physiol.-Heart Circ. Physiol.* 319, H377-H391–H391. doi:10.1152/ajpheart.00717.2019

Li, B., Manan, R. S., Liang, S.-Q., Gordon, A., Jiang, A., Varley, A., et al. (2023). Combinatorial design of nanoparticles for pulmonary mRNA delivery and genome editing. *Nat. Biotechnol.* 41, 1410–1415. doi:10.1038/s41587-023-01679-x

Li, G., Yuan, M., Li, H., Deng, C., Wang, Q., Tang, Y., et al. (2021). Safety and efficacy of artemisinin-piperaquine for treatment of COVID-19: an open-label, non-randomised and controlled trial. *Int. J. Antimicrob. Agents* 57, 106216. doi:10.1016/j.ijantimicag.2020.106216

Li, Q., Chan, C., Peterson, N., Hanna, R. N., Alfaro, A., Allen, K. L., et al. (2020). Engineering caveolae-targeted lipid nanoparticles to deliver mRNA to the lungs. *ACS Chem. Biol.* 15, 830–836. doi:10.1021/acschembio.0c00003

Liang, X., Shen, W., Sun, H., Migawa, M. T., Vickers, T. A., and Crooke, S. T. (2016). Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nat. Biotechnol.* 34, 875–880. doi:10.1038/nbt.3589

Liao, W., Dong, J., Peh, H. Y., Tan, L. H., Lim, K. S., Li, L., et al. (2017). Oligonucleotide therapy for obstructive and restrictive respiratory diseases. *Mol. Basel Switz.* 22, 139. doi:10.3390/molecules22010139

Lim, J., Chin, V., Fairfax, K., Moutinho, C., Suan, D., Ji, H., et al. (2023). Transitioning single-cell genomics into the clinic. *Nat. Rev. Genet.* 24, 573–584. doi:10.1038/s41576-023-00613-w

Lingel, A., and Izaurralde, E. (2004). RNAi: finding the elusive endonuclease. *RNA* 10, 1675–1679. doi:10.1261/rna.7175704

Liu, S., Cheng, Q., Wei, T., Yu, X., Johnson, L. T., Farbiak, L., et al. (2021). Membrane-destabilizing ionizable phospholipids for organ-selective mRNA delivery and CRISPR–Cas gene editing. *Nat. Mat.* 20, 701–710. doi:10.1038/s41563-020-00886-0

Lokugamage, M. P., Vanover, D., Beyersdorf, J., Hatit, M. Z. C., Rotolo, L., Echeverri, E. S., et al. (2021). Optimization of lipid nanoparticles for the delivery of nebulized therapeutic mRNA to the lungs. *Nat. Biomed. Eng.* 5, 1059–1068. doi:10.1038/s41551-021-00786-x

Majumder, P., Singh, A., Wang, Z., Dutta, K., Pahwa, R., Liang, C., et al. (2021). Surface-fill hydrogel attenuates the oncogenic signature of complex anatomical surface cancer in a single application. *Nat. Nanotechnol.* 16, 1251–1259. doi:10.1038/s41565-021-00961-w

Man, H. S. J., Subramaniam, N., Downs, T., Sukumar, A. N., Saha, A. D., Nair, R., et al. (2023). Long noncoding RNA GATA2-AS1 augments endothelial hypoxia inducible factor 1-α induction and regulates hypoxic signaling. *J. Biol. Chem.* 299, 103029. doi:10.1016/j.jbc.2023.103029

Marques, J. T., and Williams, B. R. G. (2005). Activation of the mammalian immune system by siRNAs. *Nat. Biotechnol.* 23, 1399–1405. doi:10.1038/nbt1161

Martinon, F., Krishnan, S., Lenzen, G., Magné, R., Gomard, E., Guillet, J. G., et al. (1993). Induction of virus-specific cytotoxic T lymphocytes *in vivo* by liposome-entrapped mRNA. *Eur. J. Immunol.* 23, 1719–1722. doi:10.1002/eji.1830230749

Mastorakos, P., da Silva, A. L., Chisholm, J., Song, E., Choi, W. K., Boyle, M. P., et al. (2015). Highly compacted biodegradable DNA nanoparticles capable of overcoming the mucus barrier for inhaled lung gene therapy. *Proc. Natl. Acad. Sci.* 112, 8720–8725. doi:10.1073/pnas.1502281112

Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L.-L., et al. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* 24, 430–447. doi:10.1038/s41580-022-00566-8

McNeer, N. A., Anandalingam, K., Fields, R. J., Caputo, C., Kopic, S., Gupta, A., et al. (2015). Nanoparticles that deliver triplex-forming peptide nucleic acid molecules correct F508del CFTR in airway epithelium. *Nat. Commun.* 6, 6952. doi:10.1038/ncomms7952

Mello, C. C., and Conte, D. (2004). Revealing the world of RNA interference. *Nature* 431, 338–342. doi:10.1038/nature02872

Meng, Z., and Lu, M. (2017). RNA interference-induced innate immunity, off-target effect, or immune adjuvant? *Front. Immunol.* 8, 331. doi:10.3389/fimmu.2017.00331

Meng, Z., Zhang, X., Wu, J., Pei, R., Xu, Y., Yang, D., et al. (2013). RNAi induces innate immunity through multiple cellular signaling pathways. *PLOS ONE* 8, e64708. doi:10.1371/journal.pone.0064708

Mitchell, M. J., Billingsley, M. M., Haley, R. M., Wechsler, M. E., Peppas, N. A., and Langer, R. (2021). Engineering precision nanoparticles for drug delivery. *Nat. Rev. Drug Discov.* 20, 101–124. doi:10.1038/s41573-020-0090-8

Monga, I., Qureshi, A., Thakur, N., Gupta, A. K., and Kumar, M. (2017). ASPsiRNA: a resource of ASP-siRNAs having therapeutic potential for human genetic disorders and algorithm for prediction of their inhibitory efficacy. *G3 GenesGenomesGenetics* 7, 2931–2943. doi:10.1534/g3.117.044024

Morral, N., and Witting, S. R. (2012). "shRNA-induced interferon-stimulated gene analysis," in *Cytokine protocols methods in molecular biology*. Editor M. De Ley (Totowa, NJ: Humana Press), 163–177. doi:10.1007/978-1-61779-439-1_10

Morrissey, D. V., Lockridge, J. A., Shaw, L., Blanchard, K., Jensen, K., Breen, W., et al. (2005). Potent and persistent *in vivo* anti-HBV activity of chemically modified siRNAs. *Nat. Biotechnol.* 23, 1002–1007. doi:10.1038/nbt1122

Musunuru, K., Chadwick, A. C., Mizoguchi, T., Garcia, S. P., DeNizio, J. E., Reiss, C. W., et al. (2021). *In vivo* CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* 593, 429–434. doi:10.1038/s41586-021-03534-y

Nowakowski, T. J., Rani, N., Golkaram, M., Zhou, H. R., Alvarado, B., Huch, K., et al. (2018). Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nat. Neurosci.* 21, 1784–1792. doi:10.1038/s41593-018-0265-3

Olejniczak, M., Polak, K., Galka-Marciniak, P., and Krzyzosiak, W. J. (2011). Recent advances in understanding of the immunological off-target effects of siRNA. *Curr. Gene Ther.* 11, 532–543. doi:10.2174/156652311798192770

Ong, T., and Ramsey, B. W. (2023). Cystic fibrosis: a review. *JAMA* 329, 1859–1871. doi:10.1001/jama.2023.8120

Østergaard, M. E., Thomas, G., Koller, E., Southwell, A. L., Hayden, M. R., and Seth, P. P. (2015). Biophysical and biological characterization of hairpin and molecular beacon RNase H active antisense oligonucleotides. *ACS Chem. Biol.* 10, 1227–1233. doi:10.1021/cb500880f

Pardi, N., Tuyishime, S., Muramatsu, H., Kariko, K., Mui, B. L., Tam, Y. K., et al. (2015). Expression kinetics of nucleoside-modified mRNA delivered in lipid nanoparticles to mice by various routes. *J. Control. Release Off. J. Control. Release Soc.* 217, 345–351. doi:10.1016/j.jconrel.2015.08.007

Patel, A. K., Kaczmarek, J. C., Bose, S., Kauffman, K. J., Mir, F., Heartlein, M. W., et al. (2019). Inhaled nanoformulated mRNA polyplexes for protein production in lung epithelium. *Adv. Mat.* 31, 1805116. doi:10.1002/adma.201805116

Paunovska, K., David Loughrey, and Dahlman, J. E. (2022). Drug delivery systems for RNA therapeutics. *Nat. Rev. Genet.* 23, 265–280. doi:10.1038/s41576-021-00439-4

Paunovska, K., Gil, C. J., Lokugamage, M. P., Sago, C. D., Sato, M., Lando, G. N., et al. (2018). Analyzing 2000 *in vivo* drug delivery data points reveals cholesterol structure impacts nanoparticle delivery. *ACS Nano* 12, 8341–8349. doi:10.1021/acsnano.8b03640

Paz-Ares, L., Sanchez, J. M., García-Velasco, A., Massuti, B., López-Vivanco, G., Provencio, M., et al. (2006). A prospective phase II trial of erlotinib in advanced non-small cell lung cancer (NSCLC) patients (p) with mutations in the tyrosine kinase (TK) domain of the epidermal growth factor receptor (EGFR). *J. Clin. Oncol.* 24, 7020. doi:10.1200/jco.2006.24.18_suppl.7020

Pfafenrot, C., Schneider, T., Müller, C., Hung, L.-H., Schreiner, S., Ziebuhr, J., et al. (2021). Inhibition of SARS-CoV-2 coronavirus proliferation by designer antisense-circRNAs. *Nucleic Acids Res.* 49, 12502–12516. doi:10.1093/nar/gkab1096

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. (2020). Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. *N. Engl. J. Med.* 383, 2603–2615. doi:10.1056/NEJMoa2034577

Pollard, A. J., and Bijker, E. M. (2021). A guide to vaccinology: from basic principles to new developments. *Nat. Rev. Immunol.* 21, 83–100. doi:10.1038/s41577-020-00479-7

Qiu, Y., Chow, M. Y. T., Liang, W., Chung, W. W. Y., Mak, J. C. W., and Lam, J. K. W. (2017). From pulmonary surfactant, synthetic KL4 peptide as effective siRNA delivery vector for pulmonary delivery. *Mol. Pharm.* 14, 4606–4617. doi:10.1021/acs.molpharmaceut.7b00725

Qiu, Y., Man, R. C. H., Liao, Q., Kung, K. L. K., Chow, M. Y. T., and Lam, J. K. W. (2019). Effective mRNA pulmonary delivery by dry powder formulation of PEGylated synthetic KL4 peptide. *J. Control. Release* 314, 102–115. doi:10.1016/j.jconrel.2019.10.026

Raal, F. J., Santos, R. D., Blom, D. J., Marais, A. D., Charng, M.-J., Cromwell, W. C., et al. (2010). Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial. *Lancet lond. Engl.* 375, 998–1006. doi:10.1016/S0140-6736(10)60284-X

Renzoni, E. A., Poletti, V., and Mackintosh, J. A. (2021). Disease pathology in fibrotic interstitial lung disease: is it all about usual interstitial pneumonia? *Lancet lond. Engl.* 398, 1437–1449. doi:10.1016/S0140-6736(21)01961-9

Rich, A., and Davies, D. R. (1956). A new two stranded helical structure: polyadenylic acid and polyuridylic acid. *J. Am. Chem. Soc.* 78, 3548–3549. doi:10.1021/ja01595a086

Rinaldi, C., and Wood, M. J. A. (2018). Antisense oligonucleotides: the next frontier for treatment of neurological disorders. *Nat. Rev. Neurol.* 14, 9–21. doi:10.1038/nrneurol.2017.148

Riordan, J. R., Rommens, J. M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245, 1066–1073. doi:10.1126/science.2475911

Ritch, P., Rudin, C. M., Bitran, J. D., Edelman, M. J., Makalinao, A., Irwin, D., et al. (2006). Phase II study of PKC-alpha antisense oligonucleotide aprinocarsen in combination with gemcitabine and carboplatin in patients with advanced non-small cell lung cancer. *Lung Cancer* 52, 173–180. doi:10.1016/j.lungcan.2005.12.012

Robbins, M., Judge, A., Ambegia, E., Choi, C., Yaworski, E., Palmer, L., et al. (2008). Misinterpreting the therapeutic effects of small interfering RNA caused by immune stimulation. *Hum. Gene Ther.* 19, 991–999. doi:10.1089/hum.2008.131

Roehr, B. (1998). Fomivirsen approved for CMV retinitis. *J. Int. Assoc. Physicians AIDS Care* 4, 14–16.

Rohner, E., Yang, R., Foo, K. S., Goedel, A., and Chien, K. R. (2022). Unlocking the promise of mRNA therapeutics. *Nat. Biotechnol.* 40, 1586–1600. doi:10.1038/s41587-022-01491-z

Rosenberg, E. S., Dorabawila, V., Easton, D., Bauer, U. E., Kumar, J., Hoen, R., et al. (2022). Covid-19 vaccine effectiveness in New York state. *N. Engl. J. Med.* 386, 116–127. doi:10.1056/NEJMoa2116063

Roshmi, R. R., and Yokota, T. (2019). Viltolarsen for the treatment of Duchenne muscular dystrophy. *Drugs Today Barc. Spain* 55, 627–639. doi:10.1358/dot.2019.55.10.3045038

Rothgangl, T., Dennis, M. K., Lin, P. J. C., Oka, R., Witzigmann, D., Villiger, L., et al. (2021). *In vivo* adenine base editing of PCSK9 in macaques reduces LDL cholesterol levels. *Nat. Biotechnol.* 39, 949–957. doi:10.1038/s41587-021-00933-4

Rowe, S. M., Zuckerman, J. B., Dorgan, D., Lascano, J., McCoy, K., Jain, M., et al. (2023). Inhaled mRNA therapy for treatment of cystic fibrosis: interim results of a randomized, double-blind, placebo-controlled phase 1/2 clinical study. *J. Cyst. Fibros. Off. J. Eur. Cyst. Fibros. Soc.* S1569-1993 (23), 656–664. –1. doi:10.1016/j.jcf.2023.04.008

Rudin, C. M., Kozloff, M., Hoffman, P. C., Edelman, M. J., Karnauskas, R., Tomek, R., et al. (2004). Phase I study of G3139, a bcl-2 antisense oligonucleotide, combined with carboplatin and etoposide in patients with small-cell lung cancer. *J. Clin. Oncol.* 22, 1110–1117. doi:10.1200/JCO.2004.10.148

Rudin, C. M., Otterson, G. A., Mauer, A. M., Villalona-Calero, M. A., Tomek, R., Prange, B., et al. (2002). A pilot trial of G3139, a bcl-2 antisense oligonucleotide, and paclitaxel in patients with chemorefractory small-cell lung cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 13, 539–545. doi:10.1093/annonc/mdf124

Rudin, C. M., Salgia, R., Wang, X., Hodgson, L. D., Masters, G. A., Green, M., et al. (2008). Randomized phase II study of carboplatin and etoposide with or without the bcl-2 antisense oligonucleotide oblimersen for extensive-stage small-cell lung cancer: CALGB 30103. *J. Clin. Oncol.* 26, 870–876. doi:10.1200/JCO.2007.14.3461

Sadeq, S., Al-Hashimi, S., Cusack, C. M., and Werner, A. (2021). Endogenous double-stranded RNA. *Non-Coding RNA* 7, 15. doi:10.3390/ncrna7010015

Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.-P., Simon, P., Löwer, M., et al. (2017). Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 547, 222–226. doi:10.1038/nature23003

Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., and Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U. S. A.* 73, 3852–3856. doi:10.1073/pnas.73.11.3852

Schlake, T., Thess, A., Fotin-Mleczek, M., and Kallen, K.-J. (2012). Developing mRNA-vaccine technologies. *RNA Biol.* 9, 1319–1330. doi:10.4161/rna.22269

Schott, J. W., Morgan, M., Galla, M., and Schambach, A. (2016). Viral and synthetic RNA vector technologies and applications. *Mol. Ther.* 24, 1513–1527. doi:10.1038/mt.2016.143

Schumacher, T. N., and Schreiber, R. D. (2015). Neoantigens in cancer immunotherapy. *Science* 348, 69–74. doi:10.1126/science.aaa4971

Sebastian, M., Schröder, A., Scheel, B., Hong, H. S., Muth, A., von Boehmer, L., et al. (2019). A phase I/IIa study of the mRNA-based cancer immunotherapy CV9201 in patients with stage IIIB/IV non-small cell lung cancer. *Cancer Immunol. Immunother.* 68, 799–812. doi:10.1007/s00262-019-02315-x

Semple, S. C., Akinc, A., Chen, J., Sandhu, A. P., Mui, B. L., Cho, C. K., et al. (2010). Rational design of cationic lipids for siRNA delivery. *Nat. Biotechnol.* 28, 172–176. doi:10.1038/nbt.1602

Sermet-Gaudelus, I., Clancy, J. P., Nichols, D. P., Nick, J. A., De Boeck, K., Solomon, G. M., et al. (2019). Antisense oligonucleotide eluforsen improves CFTR function in F508del cystic fibrosis. *J. Cyst. Fibros.* 18, 536–542. doi:10.1016/j.jcf.2018.10.015

Shukla, G. C., Singh, J., and Barik, S. (2011). MicroRNAs: processing, maturation, target recognition and regulatory functions. *Mol. Cell. Pharmacol.* 3, 83–92.

Singh, A., Bhattacharyya, N., Srivastava, A., Pruett, N., Ripley, R. T., Schrump, D. S., et al. (2019). MicroRNA-215-5p treatment suppresses mesothelioma progression via the MDM2-p53-signaling Axis. *Mol. Ther.* 27, 1665–1680. doi:10.1016/j.ymthe.2019.05.020

Singh, A., Pruett, N., Pahwa, R., Mahajan, A. P., Schrump, D. S., and Hoang, C. D. (2021). MicroRNA-206 suppresses mesothelioma progression via the Ras signaling axis. *Mol. Ther. - Nucleic Acids* 24, 669–681. doi:10.1016/j.omtn.2021.04.001

Sioud, M. (2007). RNA interference and innate immunity. *Adv. Drug Deliv. Rev.* 59, 153–163. doi:10.1016/j.addr.2007.03.006

Sood, R., Zehnder, J. L., Druzin, M. L., and Brown, P. O. (2006). Gene expression patterns in human placenta. *Proc. Natl. Acad. Sci.* 103, 5478–5483. doi:10.1073/pnas.0508035103

Spigel, D. R., Shipley, D. L., Waterhouse, D. M., Jones, S. F., Ward, P. J., Shih, K. C., et al. (2019). A randomized, double-blinded, phase II trial of carboplatin and pemetrexed with or without apatorsen (OGX-427) in patients with previously untreated stage IV non-squamous-non-small-cell lung cancer: the SPRUCE trial. *Oncol.* 24, e1409–e1416. doi:10.1634/theoncologist.2018-0518

Su, X., Fricke, J., Kavanagh, D. G., and Irvine, D. J. (2011). *In vitro* and *in vivo* mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles. *Mol. Pharm.* 8, 774–787. doi:10.1021/mp100390w

Sundaram, P., Kurniawan, H., Byrne, M. E., and Wower, J. (2013). Therapeutic RNA aptamers in clinical trials. *Eur. J. Pharm. Sci. Off. J. Eur. Fed. Pharm. Sci.* 48, 259–271. doi:10.1016/j.ejps.2012.10.014

Thanki, K., van Eetvelde, D., Geyer, A., Fraire, J., Hendrix, R., Van Eygen, H., et al. (2019). Mechanistic profiling of the release kinetics of siRNA from lipidoid-polymer

hybrid nanoparticles *in vitro* and *in vivo* after pulmonary administration. *J. Control. Release* 310, 82–93. doi:10.1016/j.jconrel.2019.08.004

Thomas, S. J., Moreira, E. D., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. (2021). Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine through 6 months. *N. Engl. J. Med.* 385, 1761–1773. doi:10.1056/NEJMoa2110345

Thompson, M. G., Stenehjem, E., Grannis, S., Ball, S. W., Naleway, A. L., Ong, T. C., et al. (2021). Effectiveness of covid-19 vaccines in ambulatory and inpatient care settings. *N. Engl. J. Med.* 385, 1355–1371. doi:10.1056/NEJMoa2110362

Tian, Z., Liang, G., Cui, K., Liang, Y., Wang, Q., Lv, S., et al. (2021). Insight into the prospects for RNAi therapy of cancer. *Front. Pharmacol.* 12, 644718. doi:10.3389/fphar.2021.644718

Tomar, R. S., Matta, H., and Chaudhary, P. M. (2003). Use of adeno-associated viral vector for delivery of small interfering RNA. *Oncogene* 22, 5712–5715. doi:10.1038/sj.onc.1206733

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510. doi:10.1126/science.2200121

Vansteenkiste, J., Canon, J.-L., Riska, H., Pirker, R., Peterson, P., John, W., et al. (2005). Randomized phase II evaluation of aprinocarsen in combination with gemcitabine and cisplatin for patients with advanced/metastatic non-small cell lung cancer. *Invest. New Drugs* 23, 263–269. doi:10.1007/s10637-005-6736-x

van Zandwijk, N., Pavlakis, N., Kao, S. C., Linton, A., Boyer, M. J., Clarke, S., et al. (2017). Safety and activity of microRNA-loaded minicells in patients with recurrent malignant pleural mesothelioma: a first-in-man, phase 1, open-label, dose-escalation study. *Lancet Oncol.* 18, 1386–1396. doi:10.1016/S1470-2045(17)30621-6

Vora, S. M., Fontana, P., Mao, T., Leger, V., Zhang, Y., Fu, T.-M., et al. (2022). Targeting stem-loop 1 of the SARS-CoV-2 5′ UTR to suppress viral translation and Nsp1 evasion. *Proc. Natl. Acad. Sci.* 119, e2117198119. doi:10.1073/pnas.2117198119

Walsh, E. E., Frenck, R. W., Falsey, A. R., Kitchin, N., Absalon, J., Gurtman, A., et al. (2020). Safety and immunogenicity of two RNA-based covid-19 vaccine candidates. *N. Engl. J. Med.* 383, 2439–2450. doi:10.1056/NEJMoa2027906

Wang, S. (2021). pH-Responsive amphiphilic carboxylate polymers: design and potential for endosomal escape. *Front. Chem.* 9, 645297. doi:10.3389/fchem.2021.645297

Wang, X., Liu, S., Sun, Y., Yu, X., Lee, S. M., Cheng, Q., et al. (2023). Preparation of selective organ-targeting (SORT) lipid nanoparticles (LNPs) using multiple technical methods for tissue-specific mRNA delivery. *Nat. Protoc.* 18, 265–291. doi:10.1038/s41596-022-00755-x

Wang, Y., Spatz, M. K., Kannan, K., Hayk, H., Avivi, A., Gorivodsky, M., et al. (1999). A mouse model for achondroplasia produced by targeting fibroblast growth factor receptor 3. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4455–4460. doi:10.1073/pnas.96.8.4455

Watts, J. K., Deleavey, G. F., and Damha, M. J. (2008). Chemically modified siRNA: tools and applications. *Drug Discov. Today* 13, 842–855. doi:10.1016/j.drudis.2008.05.007

Wei, T., Cheng, Q., Min, Y.-L., Olson, E. N., and Siegwart, D. J. (2020). Systemic nanoparticle delivery of CRISPR-Cas9 ribonucleoproteins for effective tissue specific genome editing. *Nat. Commun.* 11, 3232. doi:10.1038/s41467-020-17029-3

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862. doi:10.1016/0092-8674(93)90530-4

Wilusz, J. E. (2018). A 360° view of circular RNAs: from biogenesis to functions. *Wiley Interdiscip. Rev. RNA* 9, e1478. doi:10.1002/wrna.1478

Winkle, M., El-Daly, S. M., Fabbri, M., and Calin, G. A. (2021). Noncoding RNA therapeutics — challenges and potential solutions. *Nat. Rev. Drug Discov.* 20, 629–651. doi:10.1038/s41573-021-00219-z

Wolff, J. A., Malone, R. W., Williams, P., Chong, W., Acsadi, G., Jani, A., et al. (1990). Direct gene transfer into mouse muscle *in vivo*. *Science* 247, 1465–1468. doi:10.1126/science.1690918

Woolf, C. J., and Costigan, M. (1999). Transcriptional and posttranslational plasticity and the generation of inflammatory pain. *Proc. Natl. Acad. Sci.* 96, 7723–7730. doi:10.1073/pnas.96.14.7723

Zamecnik, P. C., and Stephenson, M. L. (1978). Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proc. Natl. Acad. Sci. U. S. A.* 75, 280–284. doi:10.1073/pnas.75.1.280

Zehendner, C. M., Valasarajan, C., Werner, A., Boeckel, J.-N., Bischoff, F. C., John, D., et al. (2020). Long noncoding RNA TYKRIL plays a role in pulmonary hypertension via the p53-mediated regulation of PDGFRβ. *Am. J. Respir. Crit. Care Med.* 202, 1445–1457. doi:10.1164/rccm.201910-2041OC

Zhang, Z., Mateus, J., Coelho, C. H., Dan, J. M., Moderbacher, C. R., Gálvez, R. I., et al. (2022). Humoral and cellular immune memory to four COVID-19 vaccines. *Cell* 185, 2434–2451.e17. doi:10.1016/j.cell.2022.05.022

Zhu, Y., Zhu, L., Wang, X., and Jin, H. (2022). RNA-based therapeutics: an overview and prospectus. *Cell Death Dis.* 13, 644–715. doi:10.1038/s41419-022-05075-2

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 January 2024
DOI 10.3389/fgene.2024.1271404

OPEN ACCESS

EDITED BY
Yadong Zheng,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Erik Josef Behringer,
Loma Linda University, United States
Isabel Duarte,
University of Algarve, Portugal
Natasha Andressa Jorge,
Leipzig University, Germany

*CORRESPONDENCE
Saadia Zahid,
saadiazahid@hotmail.com,
saadia.zahid@asab.nust.edu.pk

RECEIVED 02 August 2023
ACCEPTED 08 January 2024
PUBLISHED 16 January 2024

CITATION
Amber S and Zahid S (2024), An *in silico* approach to identify potential downstream targets of miR-153 involved in Alzheimer's disease.
Front. Genet. 15:1271404.
doi: 10.3389/fgene.2024.1271404

An *in silico* approach to identify potential downstream targets of miR-153 involved in Alzheimer's disease

Sanila Amber and Saadia Zahid*

Department of Healthcare Biotechnology, Neurobiology Research Laboratory, Atta-Ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, Islamabad, Pakistan

Background: In recent years, microRNAs (miRNAs) have emerged as key players in the pathophysiology of multiple diseases including Alzheimer's disease (AD). Messenger RNA (mRNA) targeting for regulation of gene expression by miRNAs has been implicated in the annotation of disease pathophysiology as well as in the explication of their starring role in contemporary therapeutic interventions. One such miRNA is miR-153 which mediates the survival of cortical neurons and inhibits plaque formation. However, the core mRNA targets of miR-153 have not been fully illustrated.

Objective: The present study aimed to elucidate the potential involvement of miR-153 in AD pathogenesis and to reveal its downstream targets.

Methods: miRanda was used to identify AD-associated targets of miR-153. TargetScan, PicTar, miRmap, and miRDB were further used to validate these targets. STRING 12 was employed to assess the protein-protein interaction network while Gene ontology (GO) analysis was carried out to identify the molecular functions exhibited by these gene targets.

Results: *In silico* analysis using miRanda predicted five important AD-related targets of miR-153, including APP, SORL1, PICALM, USF1, and PSEN1. All five target genes are negatively regulated by miR-153 and are substantially involved in AD pathogenesis. A protein interaction network using STRING 12 uncovered 30 potential interacting partners for SORL1, PICALM, and USF1. GO analysis revealed that miR-153 target genes play a critical role in neuronal survival, differentiation, exon guidance, amyloid precursor protein processing, and synapse formation.

Conclusion: These findings unravel the potential role of miR-153 in the pathogenesis of AD and provide the basis for forthcoming experimental studies.

KEYWORDS

amyloid precursor protein, neurodegeneration, MicroRNAs, Alzheimer's disease, miR-153

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by the formation of neurofibrillary tangles (NFTs) and Amyloid-beta (Aβ) plaques in sub-cortical brain regions that eventually lead to cognitive impairment (Amber et al., 2020). Various genetic, epigenetic, and environmental factors contribute to the development of AD

therefore the identification of informative biomarkers remained a significant challenge. Since the last decade, epigenetic mechanisms gained widespread prominence as the regulators of various important biological processes, and central to these processes are microribonucleic acids (miRNAs) (Filipowicz et al., 2008). miRNAs belong to the class of small non-coding RNAs that modulate gene expression post-transcriptionally either by target mRNA degradation or translational inhibition (Pu et al., 2019). miRNA: mRNA duplex formation necessitates the complementarity between eight nucleotide seed regions within both sequences. The duplex is either directed toward polyribosomes to regulate the mRNA translational process or targeted to the P-bodies for storage/degradation (Filipowicz et al., 2008). miRNAs are known to control the expression of almost 60% of protein-coding genes, therefore, these are considered important biomarkers for early diagnosis of various disorders. Their potential as potent biomarkers can be derived from unique secretory properties as they regulate the expression of multiple genes in various cell types without cell-to-cell contact (Schwarzenbach et al., 2014). Apart from their presence in tissues, miRNAs are also secreted in extracellular fluids, blood plasma, and saliva and therefore can serve as potential non-invasive markers for disease diagnosis (François et al., 2019). The preliminary evidence about the involvement of miRNAs in human diseases originated from cancer studies. Various expression profiling studies revealed the abnormal expression of different miRNAs in cancer samples as compared to the control (Calin et al., 2002).

The miRNAs that were consistently found to be deregulated in AD include; miR-9, miR-29, miR-34, miR-107, miR-181, miR-186, miR-146a, miR-155 and miR-153 (Femminella et al., 2015). The miR-153 is implicated in various diseases such as hypertension, osteosarcoma, glioblastoma, and various other cancers. miR-153 contributes toward the hypertensive state via the downregulation of KCNQ4 (Carr et al., 2016). An increase in miR-153 expression elevated neurogenesis and improved cognition (Qiao et al., 2020). Moreover, a significant reduction in the expression levels of miR-153 is also observed in early, moderate, and severe AD cases as compared to age-matched control specimens. Additionally, an inverse correlation was observed between miR-153 and Aβ plaque burden making it a potential disease biomarker and novel drug target (Long et al., 2012). Ectopic expression of miR-153-3p induced inflammation by increasing the release of IL-1β, TNF-α, and IL-6 and decreased neural stem cell differentiation via regulating GPR55 expression (Dong et al., 2023). Increased expression of miR-153 disrupted synapsin 1 in the hippocampus and impaired glutamatergic vesicle transport thus causing chronic cerebral hypoperfusion in rats (Zhang et al., 2020).

Due to the substantial role of miR-153 in neuronal disorders including AD, it is vital to identify the molecular targets associated with this very same miRNA to elucidate the underlying mechanisms leading to the disease phenotype. The data regarding the regulatory and therapeutic role of miRNAs is scarce due to the limitations of current experimental procedures (Jaberi et al., 2024). Owing to the significance of miRNAs in disease-related processes the pace of miRNA target prediction needs to be improved. Various *in silico* algorithms are available to reveal the molecular targets of a large proportion of miRNAs with relative sensitivity and specificity (Hamzeiy et al., 2014). Therefore, this study aimed to investigate the important AD-related mRNA targets of miR-153 to improve the current understanding of disease at the molecular level. AD-associated mRNA targets of miR-153 are identified via the miRanda algorithm and results are cross-validated by four other publicly available algorithms, TargetScan, PicTar, miRmap, and miRDB.

2 Methods

2.1 Targets prediction of miR-153

Web-based bioinformatic algorithm miRanda (Oliveira et al., 2017) was assessed to predict the mRNA targets of miR-153 and the mirSVR scores were assigned to each predicted target site. The sequence of miR-153 is available in the NCBI database (>LM608503.1 TPA: *Homo sapiens* microRNA hsa-mir-153precursor CTCACAGCTGCCAGTGTCATTTTTGTGATC TGCAGCTAGTATTCTCACTCCAGTTGCATAGTCACAAAAG TGATCATTGGCAGGTGTGGC).

The miRanda algorithm is developed for the prediction of mRNA targets and expression profiles of miRNAs available at MicroRNA.org (http://www.microrna.org); while mirSVR score is a regression model that reveals contextual features and sequence of the predicted miRNA:mRNA duplex and is directly correlated to the downregulation of miRNA and target sites of interest. *Homo sapiens* was selected as a species of choice and all the search was performed using default parameters (MFE threshold: −20 kcal/mol, scaling parameter: 4·00, score threshold: 140.00, gap open and extend penalty: −4·000 and −9.000 respectively).

2.2 Validation of results by different algorithms

The mRNA targets obtained from miRanda were further validated by four other publicly available algorithms, i.e., TargetScan, PicTar, miRDB, and MiRmap. In the TargetScan database, (Release 8, http://www.targetscan.org/), humans were selected as the species of choice. Furthermore, there were two options to find the target, i.e., by entering the gene name or miRNA name. The miRNA-153 was entered as a query and it gave two options such as miR-153–3p and miR-153–5p. Both options were explored for the target genes (Huang et al., 2020).

In the PicTar database, "PicTar target prediction in vertebrates" was selected. Following that, vertebrates was chosen as a species and then, miR-153 was selected from the dropdown menu. (http://pictar.mdc-berlin.de/) (Xue et al., 2020). In miRmap, human was selected as a species and then miR-153 was selected from the dropdown menu (https://mirmap.ezlab.org/) (Vejnar and Zdobnov, 2012).

In miRDB, humans were selected as the species of choice. Furthermore, there were two options to find the target, i.e., by entering the gene name or miRNA name. The miRNA-153 was entered as a query and it gave two selections such as miR-153–3p and miR-153–5p. Both options were explored for the target genes (http://mirdb.org/miRDB/) (Wong and Wang, 2015).

TABLE 1 miR-153 targets and their miSVR scores predicted by miRanda and validated by different software.

Sr No.	miR-153 targets	miSVR score	Algorithms
1	APP	−1.2559	miRanda, Target Scan, PicTar, miRmap
2	SORL1	−0.6425	miRanda, Target Scan, MIRDB, miRmap
3	PICALM	−0.1180	miRanda, Target Scan, miRmap
4	USF1	−0.2466	miRanda, PITA, miRmap
5	PSEN1	−0.1895	miRanda, miRmap

2.3 Protein association, functional enrichment, and post-translational modification analysis

Targets predicted by miRanda were submitted to STRING v.12 (Szklarczyk et al., 2017) (http://string-db.org/) database to explore the functional association networks of target proteins using UniProt accession numbers. *Homo sapiens* was selected from the given list of species. Biological processes, cellular localization, molecular functions, and miRNA targets of the specific miR-153 affected proteins were investigated by GO analysis and microRNA target analysis using the WEB-based Gene SeTAnaLysis Toolkit (WEBGESTALT) (Wang et al., 2017). Swiss-Prot accession numbers of miR-153 target proteins were employed for enrichment analysis.

The phosphorylation modification sites were predicted for the identified target proteins, using NetPhos 3.1 server (www.cbs.dtu.dk/services/NetPhos-3.1) (Arshad et al., 2018) while S-nitrosylation, and N and O glycosylation sites were predicted using GPS-SNO http://sno.biocuckoo.org (Mazina et al., 2023), NetNGlyc 1.0 www.cbs.dtu.dk/services/NetNGlyc/(Azevedo et al., 2018), and NetOGlyc 4.0 www.cbs.dtu.dk/services/NetOGlyc/(Kwan et al., 2021), respectively. Default settings were used for the analysis of posttranslational modification (PTM) sites and the predictions having output scores above 0.5 were only selected to avoid the possibility of false positive results. The FASTA sequence of the targeted proteins was acquired from the NCBI protein database (https://www.ncbi.nlm.nih.gov/pubmed/).

3 Results

3.1 Targets prediction of miR-153

miRanda algorithm returned 5,810 targets for miR-153 which were further screened to identify the targets involved in AD pathophysiology using a literature search. Five of the 5,810 targets found to be most relevant with AD include; sortilin-related receptor 1 (SORL1), amyloid precursor protein (APP), phosphatidylinositol binding clathrin assembly protein (PICALM), upstream stimulatory factor 1 (USF1) and presenilin-1 (PSEN1). miSVR scores indicated that miR-153 downregulates all the target genes. The results were cross-validated by four different freely accessible software TargetScan, PicTar, miRmap, and miRDB. It is observed that all five targets were not predicted by all the software (Table 1). APP and PSEN1 are already reported to be affected by miR-153 so we used SORL1, PICALM, and USF1 for further analysis.

3.2 Protein association network and functional analysis

STRING 12 analysis exhibited a strong association (score >0.7) of miR-153 target proteins with various other proteins, i.e., SORL1 exhibited strong interaction with GGA1, GGA2, APOE, ABCA7, CLU, APP, VPS35, VPS26A, LRPAP1, and NTS; PICALM is strongly associated with CLINT1, AP2A1, EPS15, RPS27A, CLTC, EPN2, EPN3, UBA52, UBB and UBC. Similarly, USF1 also showed significant interaction with ten different proteins such as SP1, ESR1, SMARCD3, EP300, FOSL1, USF2, MED1, RFX5, TAF7, and GTF2I (Figure 1). The functions and complete names of all the interacting partners are listed in Table 2.

3.3 Functional enrichment and plausible port translational modifications analysis

The identified miR-153 target proteins were functionally annotated using WEBGESTALT and Uniprot (www.uniprot.org/). Target proteins were classified based on molecular function, biological process, and cellular localization (Table 3). All three proteins are actively involved in different biological processes. SORL1 is a neuronal apolipoprotein E receptor and its gene is predominantly expressed in the central nervous system (CNS) and is involved in beta-amyloid binding, vesicle-mediated transport, cholesterol metabolic process, negative regulation of neurogenesis, and various other important cellular pathways. PICALM plays an important role in clathrin-mediated endocytosis, vesicle-mediated transport, axonogenesis, neuron projection development, neuronal differentiation, dendrite development, and many other different processes. USF1 acts as a transcription factor and belongs to the basic helix-loop-helix leucine zipper family which is known to regulate the macromolecule metabolic process, cellular metabolic process, coagulation, hypoxia, glucose homeostasis, fibrinolysis, and nutrient levels.

A total of 49 serine, tyrosine, and threonine sites were predicted as plausible phosphorylation sites for USF1, 29 for PICALM, and 368 sites for SORL1. The *S*-nitrosylation prediction analysis revealed 2 cysteine residue sites at positions 229 and 248 for USF1, while 3 and 4 sites were predicted for PICALM and SORL1, respectively. The cysteine modifications for PICALM were predicted at positions

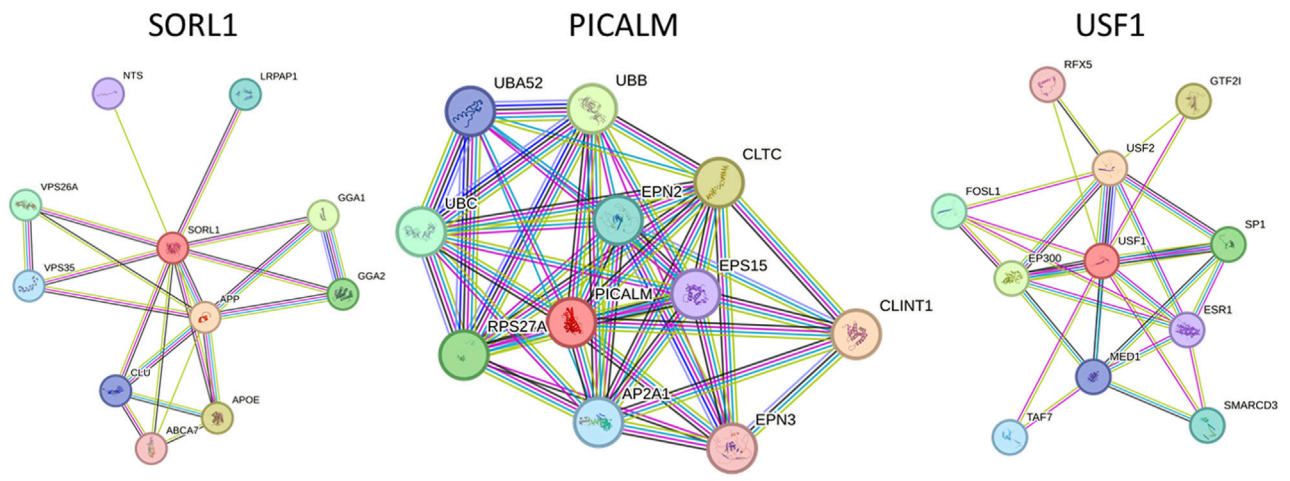


FIGURE 1
Functional association network of miR-153 target proteins. High-confidence protein-protein interaction network of identified proteins derived from the STRING database. Each protein is represented as a node with edged interactions.

27, 48, and 230 whereas positions 942, 1,042, 1,502, and 1,593 of SORL1 are susceptible to cysteine modifications. The N and O glycosylation analysis for the target proteins also showed significant susceptibility for these PTMs. A total of 6 sites were predicted as plausible sites for N-glycosylation in the PICALM sequence at positions 69, 105, 384, 445, 505, and 513. The O-glycosylation was also predicted for 58 sites of PICALM For USF1, 43 sites were predicted for O-glycosylation while no plausible sites were identified for N-glycosylation. The SORL1 has 28 predicted sites for N-glycosylation while 314 sites were found to be susceptible to O-glycosylation (Supplementary Data).

4 Discussion

By regulating the expression of target genes, miRNAs mediate various biological processes. Different miRNAs are reported to associate with AD, however, miR-153 plays a crucial role in regulating the expression of amyloid precursor protein (APP). Its expression is significantly downregulated in early and late-stage AD as observed in the APPswe/PSΔE9 murine model (Liang et al., 2012). SNHG1-mediated suppression of miR-153 increases neurotoxicity in SH-SY5Y cells (Zhao et al., 2020). Inversely, increased expression of miR-153 protects the neurons from cellular death via the upregulation of PRX5 (Xu et al., 2019). Similarly, miR-153–3p reduces LPS-induced neuroinflammation and subsequently cell death by inhibiting the NF-κB signaling pathway (Choi et al., 2022).

miR-153 obstructs APP production in neurons therefore its deregulation may drive over-expression of APP and subsequently leads to AD progression. Apart from APP miR-153 also reduced the expression of APLP2, an (APP homolog), in human fetal brain cultures therefore, it was hypothesized that it may target some of the other critical genes linked to neurodegeneration and AD development (Long et al., 2012). In this study, five main culprits of AD pathogenesis were found to be negatively regulated by miR-153 that include; APP, SORL1, PICALM, USF1, and PSEN1. The relationship between miR-153 and APP expression is well established while PSEN1 is predicted by just one algorithm hence we primarily focused on SORL1, PICALM, and USF1.

Apart from the direct role of these genes in AD, the complex interaction with various important disease-promoting/alleviating entities is revealed by the STRING database. The interaction network exhibits that complex multi-dimensional regulation takes place between key AD players, such as APP, SORL1, PICALM, USF1, PSEN1, and other disease-causing agents. The predicted genes/proteins are significant to neuroprotection, synapse formation, memory and learning, intellectual abilities, and neurodegeneration (Chandrasekaran and Bonchev, 2016). Neuronal sortilin receptor-related gene (SORL1) mediates the intracellular trafficking of APP and dysregulation of the particular process leads to Aβ accumulation and subsequently neuronal apoptosis. The exact underlying mechanisms determining the influence of SORL1 on APP trafficking and export are not explicitly studied therefore opening new avenues to investigate AD from a different perspective (Lee et al., 2008). SORL1 exhibited strong interaction with various proteins, such as GGA1, GGA2, APOE, ABCA7, CLU, APP, VPS35, VPS26A, LRPAP1, and NTS. Apolipoprotein E (APOE) modulates lipid metabolism and is implicated in AD pathogenesis. Lower levels of APOE are linked with a decline in cognitive abilities. Genetic variations in the APOE region alter the plasma expression levels of this gene and increase the risk for AD (Aslam et al., 2023). APOE ε4 allele leads to poor cognitive abilities and increased amyloid beta burden in the brain. Moreover, it alters the microglial immune response by downregulating innate immunity (lysosomal and complement pathways) and inducing stress-like responses (Liu et al., 2023). Apolipoproteins mediate cholesterol metabolism mainly via ABCA1 (ATP-binding cassette transporter A1) (Chen et al., 2013). ABCA1 is widely present in neurons and astrocytes and maintains cholesterol homeostasis in the brain. A recent study reported that amyloid beta-mediated dysfunctional ABCA1 in astrocytes altered the transport of cholesterol from astrocytes to the neurons. It subsequently led to impairment of cholesterol

TABLE 2 Functional association of SORL1, PICALM, and USF1 along with interacting partner derived from the STRING database.

Protein	Interacting partner	Function	Score
SORL1	APOE (Apolipoprotein E)	A protein associated with lipid particles, that mainly functions in lipoprotein-mediated lipid transport between organs via the plasma and interstitial fluids	0.997
	ABCA7 (Phospholipid-transporting ATPase ABCA7)	Catalyzes the translocation of specific phospholipids from the cytoplasmic to the extracellular/lumenal leaflet of membrane coupled to the hydrolysis of ATP	0.841
	APP (Amyloid precursor protein)	Functions as a cell surface receptor and performs physiological functions on the surface of neurons relevant to neurite growth, neuronal adhesion, and axonogenesis	0.999
	GGA1 (Golgi-associated, gamma adaptin ear containing, ARF binding protein 1)	Plays a role in protein sorting and trafficking between the trans-Golgi network (TGN) and endosomes	0.986
	GGA2 (ADP-ribosylation factor-binding protein GGA2)	Mediates the ARF-dependent recruitment of clathrin to the TGN and binds ubiquitinated proteins and membrane cargo molecules with a cytosolic acidic cluster-dileucine (DXXLL) motif	0.951
	VPS26A (Vacuolar protein sorting 26 homolog A)	Acts as a component of the retromer cargo-selective complex (CSC)	0.946
	VPS35 (Vacuolar protein sorting 35 homologs)	Acts as a component of the retromer cargo-selective complex (CSC). The CSC prevents the mis-sorting of selected transmembrane cargo proteins into the lysosomal degradation pathway	0.877
	CLU (Clusterin alpha chain; [Isoform 1])	Functions as an extracellular chaperone that prevents aggregation of non-native proteins	0.862
	LRPAP1 (low-density lipoprotein receptor-related protein associated protein 1)	Molecular chaperone for LDL receptor-related proteins that may regulate their ligand binding activity along the secretory pathway	0.911
	NTS (Neurotensin/neuromedin N)	Neurotensin may play an endocrine or paracrine role in the regulation of fat metabolism	0.852
PICALM	CLINT1 (Clathrin interactor 1)	May have a role in transport via clathrin-coated vesicles from the trans-Golgi network to endosomes	0.982
	RPS27A (biquitin-40S ribosomal protein S27a)	It exists in independent form or is attached to other proteins to modify their functions	0.949
	EPN2 (Epsin-2)	Plays a role in the formation of clathrin-coated invaginations and endocytosis	0.947
	CTLC (clathrin, heavy chain 1)	Clathrin is the major protein of the polyhedral coat of coated pits and vesicles	0.971
	EPN3 (Epsin-3)	Mediates apoptosis	0.944
	AP2A1 (Adaptor-related protein complex 2, alpha 1 subunit)	Adaptor protein complexes function in protein transport via transport vesicles in different membrane traffic pathways	0.947
	UBA52 (biquitin-60S ribosomal protein L40)	It is a component of 60S ribosomal subunit	0.947
	UBB (Polyubiquitin-B)	It exists in independent form or is attached to other proteins to modify their functions	0.950
	UBC (Ubiquitin C)	It exists in independent form or is attached to other proteins to modify their functions	0.949
	EPS15 (Epidermal growth factor receptor pathway substrate 15)	Involved in cell growth regulation	0.946
USF1	SP1 (Transcription factor Sp1)	It can activate or repress transcription in response to physiological and pathological signals	0.901
	ESR1 (estrogen receptor 1)	Involved in the regulation of eukaryotic gene expression and affect cellular proliferation and differentiation in target tissues	0.796
	SMARCD3 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 3)	Stimulates nuclear receptor mediated transcription	0.819

(Continued on following page)

TABLE 2 (*Continued*) Functional association of SORL1, PICALM, and USF1 along with interacting partner derived from the STRING database.

Protein	Interacting partner	Function	Score
	EP300 (Histone acetyltransferase p300)	Functions as histone acetyltransferase and regulates transcription via chromatin remodeling	0.912
	FOSL1 (Fos-related antigen 1)	Modulates cellular transformation, multiplication, and differentiation	0.846
	USF2 (Upstream stimulatory factor 2)	Transcription factor that binds to a symmetrical DNA sequence (E-boxes)	0.999
	MED1(Mediator of RNA polymerase II transcription subunit 1)	A coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes	0.801
	RFX5 (NA-binding protein RFX5)	Activates transcription from class II MHC promoters	0.752
	TAF7 (Transcription initiation factor TFIID subunit 7)	Functions as a component of the DNA-binding general transcription factor complex TFIID.	0.814
	GTF2I (General transcription factor II-I)	Acts as a coregulator for USF1 by binding independently two promoter elements, a pyrimidine-rich initiator (Inr) and an upstream E-box	0.958

TABLE 3 Functional distribution of SORL1, PICALM and USF1 on the basis of biological process, molecular function and cellular compartment.

	SORL1	PICALM	USF1
Biological processes	Vesicle mediated transport	Vesicle mediated transport Endocytosis Receptor mediated endocytosis Plasma membrane part Positive regulation of macromolecule metabolic process Positive regulation of cellular metabolic process Receptor metabolic process Cell part morphogenesis Cell projection morphogenesis Neuron projection development Axonogenesis Cell morphogenesis involved in neuronal differentiation Dendrite development Synapse	Positive regulation of macromolecule metabolic process
	Sterol metabolic process		Positive regulation of cellular metabolic process
	Cholesterol metabolic process		Cellular response to nutrient levels
	Negative regulation of neurogenesis		Coagulation
	Negative regulation of beta-amyloid formation		Response to hypoxia
	Positive regulation of protein catabolic process		Glucose homeostasis Negative regulation to fibrinolysis
Molecular function	Beta-Amyloid binding	Phosphatidylinositol-4,5-bisphosphate 5 phosphatase activity	Transcription factor binding
	Protein transporter activity		MAP Kinase activity
	Beta-aspartyl-transferase activity		
Cellular Components	Early endosome Endoplasmic reticulum Extracellular exosome	Golgi apparatus Clathrin coated vesicles	Nucleoplasm
	Membrane	Neurofibrillary tangle	Nucleus
	Nuclear envelop Lumen	Neuronal cell body Pre and post synaptic membrane	Transcription factor complex
MicroRNA Targets	MIR-17–5p, MIR-20A, MIR-106A, MIR106B, MIR-20B and MIR-519D	MIR-520F	

metabolism, a prominent feature of AD pathogenesis (Azizidoost et al., 2022). Clusterin (CLU) plays a protective role in the brain however, mutations in CLU increase the risk of developing AD. The rs11136000C mutation in CLU causes dysregulation in GABAergic signaling thus promoting AD pathogenesis (Chen et al., 2023).

Phosphatidylinositol-binding clathrin assembly protein (PICALM) is associated with clathrin-mediated endocytosis (Kyriazis et al., 2008). It is predominantly situated in neurons, oligodendrocytes, astrocytes, and endothelial cells where it recruits the adaptor protein 2 (AP-2) and clathrin to the plasma membrane to encapsulate the target proteins (Yao et al., 2003). The clathrin-coated vesicles are further processed in endosomes or lysosomes to be removed from the cell. PICALM is also associated with the removal of Aβ from the cells, therefore, minimizing the plaque burden and preventing AD pathology. Altered PICALM expression levels are reported in AD brain tissues however, it is yet to be determined whether it affects the Aβ transport or is influenced by Aβ levels (Baig et al., 2010). PICALM is strongly associated with various other proteins and alterations in its expression may influence the biological activities of target proteins correspondingly. The interacting partners of PICALM include; CLINT1, AP2A1, EPS15, RPS27A, CLTC,

EPN2, EPN3, UBA52, UBB and UBC. RPS27A is a fusion protein consisting of ubiquitin and S27a (ribosomal protein) (Sayers et al., 2018). An *in silico* analysis revealed the potential role of RPS27A in neurodegenerative disorders by modulating the expression of Il-18 and Cx3cl1 (Khayer et al., 2020). The role of other target proteins is still unclear in AD and needs further research.

Upstream transcription factor 1 (USF1), a ubiquitously expressed gene encodes a transcription factor that stimulates the transcription of various lipid and glucose-metabolizing genes (Lee et al., 2006) including APOE (Salero et al., 2003). USF1 plays a significant role in abnormal lipid aggregation (Guo et al., 2018), neuronal differentiation, and synaptic plasticity, moreover activates the APP promoter thereby affecting Aβ production and processing (Isotalo et al., 2012). USF1 strongly interacts with various other proteins such as SP1, ESR1, SMARCD3, EP300, FOSL1, USF2, MED1, RFX5, TAF7, and GTF2I. ESR1 (Estrogen receptor 1) is implicated in AD progression and it is described that ESR1 mutant (rs9340803) may lead to AD by perturbing cholesterol metabolism and accumulating amyloid beta in the brain. Nevertheless, further studies on larger cohorts are required to confirm the role of the ESR1 variant in AD (Li et al., 2018).

The post-translational modification data for the target proteins revealed a significant number of predicted sites with susceptibility towards phosphorylation, S-nitrosylation, and N and O-glycosylation. There is ample evidence that PTMs play a crucial role in AD pathology (Marcelli et al., 2018). Phosphorylation of tau and amyloid beta is detected in AD mouse models and these modifications affect the functions of microtubules and synapses, respectively (Wang et al., 2023).

Identification and validation of these predicted PTM sites and their pathological correlation with miR-153 targets will also provide substantial data that will be helpful in further elucidation of molecular mechanisms involved in AD pathology.

In this study, bioinformatics analysis predicted some of the important AD-related targets of miR-153. The gene ontology (GO) analysis of putative miR-153 targets revealed their important functions relevant to AD such as regulation of Aβ formation, negative regulation of neurogenesis, neuronal projection development, synapse formation, and NFTs formation. miRNAs perform their regulatory functions by affecting the target genes therefore it is crucial to study the potential targets and their underlying effects. This approach will facilitate the identification of novel regulatory networks of various miRNAs in different disease-related processes.

5 Conclusion

Our findings may aid the understanding of different molecular mechanisms and identification of effective therapeutic targets for AD. Further experimental studies may provide additional insights into the regulatory role of miR-153 and its targets in the development of AD and other neurodegenerative disorders.

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

SA: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing–original draft. SZ: Project administration, Resources, Supervision, Validation, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors are highly obliged to the research facilities of Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST).

Acknowledgments

Authors are grateful to Sadia Nazir, PhD scholar, ASAB-NUST for her help in database search.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1325588/full#supplementary-material

References

Amber, S., Mirza, F. J., Asif, M., Hassan, D., Ahmed, T., Zahid, S., et al. (2020). Amyloid-beta induced neurotoxicity impairs cognition and adult hippocampal neurogenesis in a mouse model for Alzheimer's disease. *Curr. Alzheimer Res.* 17 (11), 1033–1042. doi:10.2174/1567205017666201224162730

Arshad, M., Bhatti, A., and John, P. (2018). Identification and *in silico* analysis of functional SNPs of human TAGAP protein: a comprehensive study. *PloS one* 13 (1), e0188143. doi:10.1371/journal.pone.0188143

Aslam, M. M., Fan, K. H., Lawrence, E., Bedison, M. A., Snitz, B. E., DeKosky, S. T., et al. (2023). Genome-wide analysis identifies novel loci influencing plasma apolipoprotein E concentration and Alzheimer's disease risk. *Mol. Psychiatry* 5, 1–2. doi:10.1038/s41380-023-02170-4

Azevedo, R., Silva, A. M. N., Reis, C. A., Santos, L. L., and Ferreira, J. A. (2018). *In silico* approaches for unveiling novel glycobiomarkers in cancer. *J. Proteomics* 171, 95–106. doi:10.1016/j.jprot.2017.08.004

Azizidoost, S., Babaahmadi-Rezaei, H., Nazeri, Z., Cheraghzadeh, M., and Kheirollah, A. (2022). Amyloid beta increases ABCA1 and HMGCR protein expression, and cholesterol synthesis and accumulation in mice neurons and astrocytes. *Biochimica Biophysica Acta (BBA)-Molecular Cell Biol. Lipids* 1867 (1), 159069. doi:10.1016/j.bbalip.2021.159069

Baig, S., Joseph, S., Tayler, H., Abraham, R., Owen, M., Williams, J., et al. (2010). Distribution and expression of picalm in Alzheimer disease. *J. Neuropathol. Exp. Neurol.* 69, 1071–1077. doi:10.1097/NEN.0b013e3181f52e01

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., et al. (2002). Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci.* 99, 15524–15529. doi:10.1073/pnas.242606799

Carr, G., Barrese, V., Stott, J. B., Povstyan, O. V., Jepps, T. A., Figueiredo, H. B., et al. (2016). MicroRNA-153 targeting of KCNQ4 contributes to vascular dysfunction in hypertension. *Cardiovasc Res.* 112, 581–589. doi:10.1093/cvr/cvw177

Chandrasekaran, S., and Bonchev, D. (2016). Network topology analysis of post-mortem brain microarrays identifies more alzheimer's related genes and micrornas and points to novel routes for fighting with the disease. *PloS One* 11, e0144052. doi:10.1371/journal.pone.0144052

Chen, C., Tang, X., Lan, Z., Chen, W., Su, H., Li, W., et al. (2023). GABAergic signaling abnormalities in a novel CLU mutation Alzheimer's disease mouse model. *Transl. Res.* 260, 32–45. doi:10.1016/j.trsl.2023.05.003

Chen, J., Zhang, X., Kusumo, H., Costa, L. G., and Guizzetti, M. (2013). Cholesterol efflux is differentially regulated in neurons and astrocytes: implications for brain cholesterol homeostasis. *Biochimica Biophysica Acta (BBA)-Molecular Cell Biol. Lipids* 1831 (2), 263–275. doi:10.1016/j.bbalip.2012.09.007

Choi, H. R., Ha, J. S., Kim, E. A., Cho, S. W., and Yang, S. J. (2022). MiR-30a-5p and miR-153-3p regulate LPS-induced neuroinflammatory response and neuronal apoptosis by targeting NeuroD1. *BMB Rep.* 55 (9), 447–452. doi:10.5483/BMBRep.2022.55.9.061

Dong, X., Wang, H., Zhan, L., Li, Q., Li, Y., Wu, G., et al. (2023). miR-153-3p suppresses the differentiation and proliferation of neural stem cells via targeting GPR55. *Aging (Albany NY)* 15 (16), 8518–8527. doi:10.18632/aging.204002

Femminella, G. D., Ferrara, N., and Rengo, G. (2015). The emerging role of micrornas in alzheimer's disease. *Front. Physiol.* 6, 40. doi:10.3389/fphys.2015.00040

Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by micrornas: are the answers in sight? *Nat. Rev. Genet.* 9, 102–114. doi:10.1038/nrg2290

François, M., Bull, C. F., Fenech, M. F., and Leifert, W. R. (2019). Current state of saliva biomarkers for aging and alzheimer's disease. *Curr. Alzheimer Res.* 16, 56–66. doi:10.2174/1567205015666181022094924

Guo, J., Fang, W., Chen, X., Lin, Y., Hu, G., Wei, J., et al. (2018). Upstream stimulating factor 1 suppresses autophagy and hepatic lipid droplet catabolism by activating mTOR. *FEBS Lett.* 592 (16), 2725–2738. doi:10.1002/1873-3468.13203

Hamzeiy, H., Allmer, J., and Yousef, M. (2014). "Computational methods for MicroRNA target prediction," in *miRNomics: MicroRNA biology and computational analysis. Methods in molecular biology (methods and protocols)*. Editors M. Yousef and J. Allmer (Totowa, NJ: Humana Press), 207–221.

Huang, J., Weng, Q., Shi, Y., Mao, W., Zhao, Z., Wu, R., et al. (2020). MicroRNA-155-5p suppresses PD-L1 expression in lung adenocarcinoma. *FEBS Open Bio* 10 (6), 1065–1071. doi:10.1002/2211-5463.12853

Isotalo, K., Kok, E. H., Luoto, T. M., Haikonen, S., Haapasalo, H., Lehtimäki, T., et al. (2012). Upstream transcription factor 1 (USF1) polymorphisms associate with Alzheimer's disease-related neuropathological lesions: tampere Autopsy Study. *Brain Pathol.* 22, 765–775. doi:10.1111/j.1750-3639.2012.00586.x

Jaberi, K. R., Alamdari-Palangi, V., Jaberi, A. R., Esmaeli, Z., Shakeri, A., Hayat, S. M., et al. (2024). The regulation, functions, and signaling of miR-153 in neurological disorders, and its potential as a biomarker and therapeutic target. *Curr. Mol. Med.* 23 (9), 863–875. doi:10.2174/1566524023666220817145638

Khayer, N., Mirzaie, M., Marashi, S. A., and Jalessi, M. (2020). Rps27a might act as a controller of microglia activation in triggering neurodegenerative diseases. *Plos one* 15 (9), e0239219. doi:10.1371/journal.pone.0239219

Kwan, S. H., Wan-Ibrahim, W. I., Juvarajah, T., Fung, S. Y., and Abdul-Rahman, P. S. (2021). Isolation and identification of O-and N-linked glycoproteins in milk from different mammalian species and their roles in biological pathways which support infant growth. *Electrophoresis* 42 (3), 233–244. doi:10.1002/elps.202000142

Kyriazis, G. A., Wei, Z., Vandermey, M., Jo, D. G., Xin, O., Mattson, M. P., et al. (2008). Numb endocytic adapter proteins regulate the transport and processing of the amyloid precursor protein in an isoform-dependent manner implications for alzheimer disease pathogenesis. *J. Biol. Chem.* 283, 25492–25502. doi:10.1074/jbc.M802072200

Lee, J. C., Lusis, A. J., and Pajukanta, P. (2006). Familial combined hyperlipidemia: upstream transcription factor 1 and beyond. *Curr. Opin. Lipidol.* 17, 101–109. doi:10.1097/01.mol.0000217890.54875.13

Lee, J. H., Barral, S., and Reitz, C. (2008). The neuronal sortilin-related receptor gene sorl1 and late-onset alzheimer's disease. *Curr. Neurol. Neurosci. Rep.* 8, 384–391. doi:10.1007/s11910-008-0060-8

Li, X., Zhu, X., Zhang, W., Yang, F., Hui, J., Tan, J., et al. (2018). The etiological effect of a new low-frequency ESR1 variant on Mild Cognitive Impairment and Alzheimer's Disease: a population-based study. *Aging (Albany NY)* 10 (9), 2316–2337. doi:10.18632/aging.101548

Liang, C., Zhu, H., Xu, Y., Huang, L., Ma, C., Deng, W., et al. (2012). MicroRNA-153 negatively regulates the expression of amyloid precursor protein and amyloid precursor-like protein 2. *Brain Res.* 1455, 103–113. doi:10.1016/j.brainres.2011.10.051

Liu, C. C., Wang, N., Chen, Y., Inoue, Y., Shue, F., Ren, Y., et al. (2023). Cell-autonomous effects of APOE4 in restricting microglial response in brain homeostasis and Alzheimer's disease. *Nat. Immunol.* 19, 1854–1866. doi:10.1038/s41590-023-01640-9

Long, J. M., Ray, B., and Lahiri, D. K. (2012). MicroRNA-153 physiologically inhibits expression of amyloid-β precursor protein in cultured human fetal brain cells and is dysregulated in a subset of Alzheimer disease patients. *J. Biol. Chem.* 287, 31298–31310. doi:10.1074/jbc.M112.366336

Marcelli, S., Corbo, M., Iannuzzi, F., Negri, L., Blandini, F., Nistico, R., et al. (2018). The involvement of post-translational modifications in Alzheimer's disease. *Curr. Alzheimer Res.* 15, 313–335. doi:10.2174/1567205014666170505095109

Mazina, A., Shumilina, J., Gazizova, N., Repkin, E., Frolov, A., and Minibayeva, F. (2023). S-nitrosylated proteins involved in autophagy in *Triticum aestivum* roots: a bottom-up proteomics approach and *in silico* predictive algorithms. *Life* 13 (10), 2024. doi:10.3390/life13102024

Oliveira, A. C., Bovolenta, L. A., Nachtigall, P. G., Herkenhoff, M. E., Lemke, N., and Pinhal, D. (2017). Combining results from distinct microRNA target prediction tools enhances the performance of analyses. *Front. Genet.* 16 (8), 59. doi:10.3389/fgene.2017.00059

Pu, M., Chen, J., Tao, Z., Miao, L., Qi, X., Wang, Y., et al. (2019). Regulatory network of miRNA on its target: coordination between transcriptional and post-transcriptional regulation of gene expression. *Cell. Mol. Life Sci.* 76, 441–451. doi:10.1007/s00018-018-2940-7

Qiao, J., Zhao, J., Chang, S., Sun, Q., Liu, N., Dong, J., et al. (2020). MicroRNA-153 improves the neurogenesis of neural stem cells and enhances the cognitive ability of aged mice through the notch signaling pathway. *Cell Death Differ.* 27 (2), 808–825. doi:10.1038/s41418-019-0388-4

Salero, E., Giménez, C., and Zafra, F. (2003). Identification of a non-canonical E-box motif as a regulatory element in the proximal promoter region of the apolipoprotein E gene. *Biochem. J.* 370, 979–986. doi:10.1042/BJ20021142

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Connor, R., et al. (2018). Database resources of the national center for biotechnology information. *Nucleic acids Res.* 46, D8–D13. doi:10.1093/nar/gkx1095

Schwarzenbach, H., Nishida, N., Calin, G. A., and Pantel, K. (2014). Clinical relevance of circulating cell-free microRNAs in cancer. *Nat. Rev. Clin. Oncol.* 11, 145–156. doi:10.1038/nrclinonc.2014.5

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi:10.1093/nar/gkw937

Vejnar, C. E., and Zdobnov, E. M. (2012). MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic acids Res.* 40 (22), 11673–11683. doi:10.1093/nar/gks901

Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 45, W130–W137. doi:10.1093/nar/gkx356

Wang, Q., Xia, C., Zhu, A., Bao, Y., Lu, J., Chen, Y., et al. (2023). Discrepancy of synaptic and microtubular protein phosphorylation in the hippocampus of APP/PS1 and MAPT× P301S transgenic mice at the early stage of Alzheimer's disease. *Metab. Brain Dis.* 9, 1983–1997. doi:10.1007/s11011-023-01209-3

Wong, N., and Wang, X. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* 43, D146–D152. doi:10.1093/nar/gku1104

Xu, C., Wang, C., Meng, Q., Gu, Y., Wang, Q., Xu, W., et al. (2019). miR-153 promotes neural differentiation in the mouse hippocampal HT-22 cell line and increases the expression of neuron-specific enolase. *Mol. Med. Rep.* 20 (2), 1725–1735. doi:10.3892/mmr.2019.10421

Xue, W. X., Zhang, M. Y., Li, R., Liu, X., Yin, Y. H., and Qu, Y. Q. (2020). Serum miR-1228-3p and miR-181a-5p as noninvasive biomarkers for non-small cell lung cancer diagnosis and prognosis. *BioMed Res. Int.* 2020, 1–13. doi:10.1155/2020/9601876

Yao, P., Zhang, P., Mattson, M., and Furukawa, K. (2003). Heterogeneity of endocytic proteins: distribution of clathrin adaptor proteins in neurons and glia. *Neuroscience* 121, 25–37. doi:10.1016/s0306-4522(03)00431-7

Zhang, S., Yan, M. L., Yang, L., An, X. B., Zhao, H. M., Xia, S. N., et al. (2020). MicroRNA-153 impairs hippocampal synaptic vesicle trafficking via downregulation of synapsin I in rats following chronic cerebral hypoperfusion. *Exp. Neurol.* 332, 113389. doi:10.1016/j.expneurol.2020.113389

Zhao, J., Geng, L., Chen, Y., and Wu, C. (2020). SNHG1 promotes MPP+ induced cytotoxicity by regulating PTEN/AKT/mTOR signaling pathway in SH-SY5Y cells via sponging miR-153-3p. *Biol. Res.* 53 (1), 1. doi:10.1186/s40659-019-0267-y

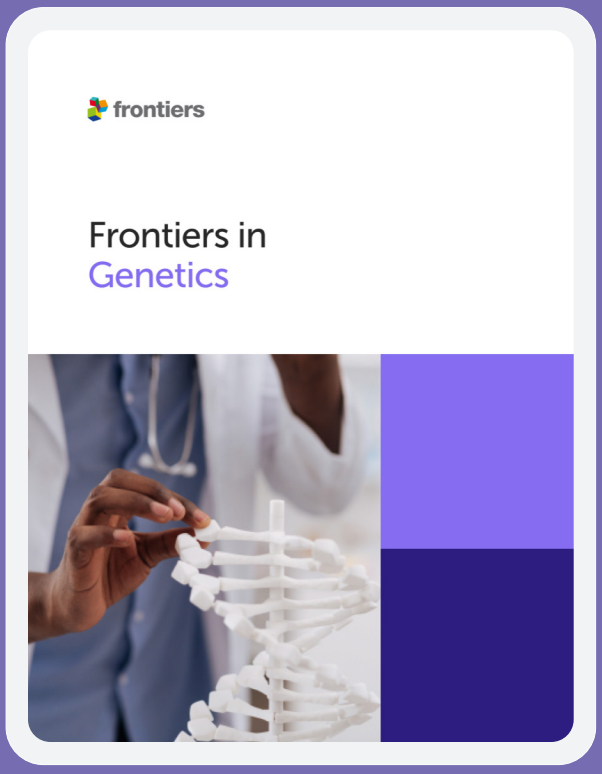
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

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