RECENT PROGRESSES OF NON-CODING RNAS IN BIOLOGICAL AND MEDICAL RESEARCH,
2nd Edition

EDITED BY: Yun Zheng and Philipp Kapranov

frontiers Research Topics

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: researchtopics@frontiersin.org

RECENT PROGRESSES OF NON-CODING RNAS IN BIOLOGICAL AND MEDICAL RESEARCH,
2nd Edition

Topic Editors:
Yun Zheng, Kunming University of Science and Technology, China
Philipp Kapranov, Huaqiao University, China

Table of Contents

Editorial: Recent Progresses of Non-coding RNAs in Biological and Medical Research

Yun Zheng[1]* and Philipp Kapranov[2]*

[1] *Yunnan Key Lab of Primate Biomedicine Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, China,* [2] *Institute of Genomics, School of Biomedical Sciences, Huaqiao University, Xiamen, China*

Editorial on the Research Topic

Recent Progresses of Non-coding RNAs in Biological and Medical Research

INTRODUCTION

Short (<200 nt) and long (>200 nt) non-coding (nc) RNAs account for majority of mammalian transcriptional output and encompass RNA species critical for various aspects of development and disease (Ambros, 2001; Kapranov et al., 2002, 2007; Bartel, 2004; Carninci et al., 2005). We have witnessed an ever-increasing pace of discovery of these transcripts in the last decade, in a large measure owing to the wide-spread application of high-throughput sequencing technologies for RNA analysis. These ncRNAs include, but not limited to, novel members of known classes such as miRNAs and siRNAs; new classes of small RNAs, for example, those associated with promoters and termini of genes; new classes of long non-coding (lnc) RNAs; plethora of antisense transcripts; circular RNAs derived from exons and introns; and many others (Laurent et al., 2015; Li et al., 2016; Kristensen et al., 2019; Zhang et al., 2019). Non-coding RNAs have been associated with almost every important biological process and human disease (Calin et al., 2004; Esteller, 2011; Wapinski and Chang, 2011; Mendell and Olson, 2012). However, our understanding of most of these transcripts is still at the initial stages.

Deeper insight into these enigmatic RNA species clearly requires efforts from both wet-lab and computational avenues of research (Zheng et al., 2017). Therefore, this Research Topic aimed to provide works from both directions to converge on generation of new insights into the functionalities of ncRNAs. Thirteen papers included in it serve as a collection of recent results and advances across multiple areas of ncRNA research field.

WET-LAB EXPERIMENTAL STUDIES OF NCRNAS

Lin et al. identified miR-30c secreted by bovine embryos as a potential biomarker for hampered preimplantation. Two miRNAs, i.e., miR-30c and miR-10b, were found at much higher levels in conditioned medium of slow cleaving embryos compared to intermediately cleaving ones (Lin et al.). One of them, miR-30c, directly repressed cyclin-dependent kinase 12 (CDK12) through a complementary site in the 3′ UTR (Lin et al.). Several DNA damage response (DDR) genes were significantly downregulated after introducing miR-30c or repressing CDK12, suggesting that miR-30c regulates embryo development through the DDR pathway (Lin et al.).

Mature hair follicles in mammals undergo periodic self-renewal processes called hair follicle cycles. Understanding the molecular regulatory mechanisms of the renewal cycle is important in medicine and developmental biology. Zhao et al. examined deregulated miRNAs, lncRNAs and circRNAs in the hair follicle cycle of Angora Rabbit (*Oryctolagus cuniculus*) and provides comprehensive repository of ncRNAs potentially relevant to this process.

Wang et al. profiled lncRNAs in the CD4$^+$ T cells in the mouse model of acute asthma. They found 36 up- and 98 down-regulated lncRNAs in the disease compared with the control samples (Wang et al.). The potential functions of deregulated lncRNA were analyzed by performing miRNA binding analysis (Wang et al.).

It has been well-established that miRNAs work by guiding RNA-induced silencing complex (RISC) to their target RNA binding sites in cytoplasm (Bartel, 2004). However, a steady stream of evidence shows that some miRNAs localize and potentially function in nucleus (Place et al., 2008; Ritland Politz et al., 2009; Liu et al., 2018). Xun et al. proposed an efficient experimental method to find miRNA binding sequences in genomic DNA *in vivo*, thus potentially identifying miRNA binding sites in the regulatory regions of genes.

COMPUTATIONAL STUDIES OF NCRNAS

Ou-Yang et al. proposed a novel method called two-side sparse self-representation (TSSR) for predicting lncRNA-disease associations. TSSR significantly outperformed other tested methods and identified some candidate lncRNA-disease associations (Ou-Yang et al.).

Zhang et al. proposed a method called CRlncRC2 for predicting associations between lncRNAs and cancers. More than four hundred cancer-related lncRNA candidates were identified, which were evaluated by examining the Lnc2Cancer database, reviewing literature, and performing statistical analysis of multiple relevant data sources containing information on mutations and differential gene expression in cancers (Zhang et al.). These results demonstrated that CRlncRC2 is an effective and accurate method for identification of cancer-related lncRNAs (Zhang et al.).

LncRNAs are assumed to realize their functions by interacting with other molecules, such as proteins, chromatin and other RNA species. Shen et al. proposed a new method for identifying lncRNA-protein interactions by employing Kernel Ridge Regression, based on Fast Kernel Learning (LPI-FKLKRR). LPI-FKLKRR demonstrated a superior performance compared with a series of other methods as judged by area under precision recall curve.

Huang et al. introduced a computational method to predict interactions between lncRNAs and miRNAs leveraging the information of expression profile data for these transcripts and the graph convolution technique. The proposed model is based on the assumption that the interaction between an lncRNA and a miRNA could be deciphered from their co-expression pattern. Compared with the conventional miRNA-target prediction algorithms based on sequence matching, their work presents a new approach to predict lncRNA:miRNA interactions.

Fukunaga et al. introduced a web server, called LncRRIsearch, for predicting lncRNA:lncRNA and lncRNA:mRNA interactions in human and mouse. The tissue-specific expression and cellular localization data of lncRNAs are integrated in this web server to explore tissue-specific or subcellular-localized lncRNA interactions (Fukunaga et al.).

REVIEWS AND PERSPECTIVES

Li and Liu summarizing recent evidences suggesting that coding and non-coding properties are inherent to both coding and non-coding transcripts. In other words, some lncRNAs and circRNAs could be used to produce short peptides, i.e., have coding capabilities. On the other hand, 3′ and 5′ UTRs of coding genes have non-coding functions such as recruiting RNA-binding proteins (Li and Liu).

Smith et al. reviewed the miRNAs and lncRNAs that play key roles in the initiation and progression of pediatric solid tumors. Pediatric tumors, due to lower mutation load compared with adult ones, are assumed to arise from mis-regulation of networks normally functioning during development at transcriptional level (Smith et al.). The authors summarized accumulating evidence of involvement of miRNAs and lncRNAs in the regulatory networks functioning during oncogenesis.

Watson et al. explored small RNAs in neurodegenerative diseases. This comprehensive review discusses roles of various small RNAs in multiple neurodegenerative diseases, including Alzheimer's, Parkinson's, multiple sclerosis, Amyotrophoic lateral sclerosis, and Huntington's disease.

Recent evidences show that ncRNAs, both miRNAs and lncRNAs, could serve as communication factors between cells (Bayraktar et al., 2017; Bär et al., 2019). Ramón y Cajal et al. proposed that the interactions between miRNAs and lncRNAs might contribute to the cell-type specific outcomes and to the determination of cell fate. In one model, miRNAs could be competitively sequestered by tissue-specific lncRNAs. In another context, miRNAs released to extracellular space as ligands could interact with lncRNAs in different organs as receptors to either sequester the miRNAs or induce degradation of the miRNAs or the lncRNAs.

SUMMARY

Non-coding RNAs have been associated with various biological processes and human diseases. These phenomena were further expanded and reviewed by several studies in this Research Topic. A number of wet lab and computational methods as well as database resources reported in the Topic should help to refine the connections between ncRNAs and diseases and identify the mechanisms of actions of the former, thus further contributing to the advancement of the ncRNA field.

AUTHOR CONTRIBUTIONS

YZ and PK conceived of the work and wrote the manuscript.

FUNDING

REFERENCES

Ambros, V. (2001). microRNAs: tiny regulators with great potential. *Cell* 107, 823–826. doi: 10.1016/S0092-8674(01)00616-X

Bär, C., Thum, T., and de Gonzalo-Calvo, D. (2019). Circulating miRNAs as mediators in cell-to-cell communication. *Epigenomics* 11, 111–113. doi: 10.2217/epi-2018-0183

Bartel, D. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/s0092-8674(04)00045-5

Bayraktar, R., Van Roosbroeck, K., and Calin, G. (2017). Cell to cell communication: microRNAs as hormones. *Mol. Oncol.* 11, 1673–1686. doi: 10.1002/1878-0261.12144

Calin, G., Sevignani, C., Dumitru, C., Hyslop, T., Noch, E., Yendamuri, S., et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad Sci. U.S.A.* 101, 2999–3004. doi: 10.1073/pnas.0307323101

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M., Maeda, N., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563. doi: 10.1126/science.1112014

Esteller, M. (2011). Non-coding RNAS in human diseases. *Nat. Rev. Genet.* 12, 861–74. doi: 10.1038/nrg3074

Kapranov, P., Cawley, S. E., Drenkow, J., Bekiranov, S., Strausberg, R. L., Fodor, S. P., et al. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296, 916–919. doi: 10.1126/science.1068597

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. doi: 10.1126/science.1138341

Kristensen, L. S., Andersen, M., Stagsted, L., Ebbesen, K., Hansen, T., 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

Laurent, G. S., Wahlestedt, C., and Kapranov, P. (2015). The Landscape of long noncoding RNA classification. *Trends Genet.* 31, 239–251. doi: 10.1016/j.tig.2015.03.007

Li, Z., Wang, S., Cheng, J., Su, C., Zhong, S., Liu, Q., et al. (2016). Intron lariat RNA inhibits microRNA biogenesis by sequestering the dicing complex in *Arabidopsis. PLoS Genet.* 12:e1006422. doi: 10.1371/journal.pgen.1006422

Liu, H., Lei, C., He, Q., Pan, Z., Xiao, D., and Tao, Y. (2018). Nuclear functions of mammalian MicroRNAs in gene regulation, immunity and cancer. *Mol. Cancer* 17:64. doi: 10.1186/s12943-018-0765-5

Mendell, J., and Olson, E. (2012). MicroRNAs in stress signaling and human disease. *Cell* 148, 1172–1187. doi: 10.1016/j.cell.2012.02.005

Place, R. F., Li, L.-C., Pookot, D., Noonan, E. J., and Dahiya, R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1608–1613. doi: 10.1073/pnas.0707594105

Ritland Politz, J. C., Hogan, E. M., and Pederson, T. (2009). MicroRNAs with a nucleolar location. *RNA* 15, 1705–1715. doi: 10.1261/rna.1470409

Wapinski, O., and Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends Cell Biol.* 21, 354–361. doi: 10.1016/j.tcb.2011.04.001

Zhang, X., Zhang, Y., Wang, T., Li, Z., Cheng, J., Ge, H., et al. (2019). A comprehensive map of intron branchpoints and lariat RNAs in plants. *Plant Cell* 31, 956–973. doi: 10.1105/tpc.18.00711

Zheng, Y., Liu, L., and Shukla, G. C. (2017). A comprehensive review of web-based non-coding RNA resources for cancer research. *Cancer Lett.* 407, 1–8. doi: 10.1016/j.canlet.2017.08.015

Multivariate Information Fusion With Fast Kernel Learning to Kernel Ridge Regression in Predicting LncRNA-Protein Interactions

Cong Shen[1], Yijie Ding[2], Jijun Tang[1,3] and Fei Guo[1]*

[1] School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China,
[2] School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou, China,
[3] Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, United States

Long non-coding RNAs (lncRNAs) constitute a large class of transcribed RNA molecules. They have a characteristic length of more than 200 nucleotides which do not encode proteins. They play an important role in regulating gene expression by interacting with the homologous RNA-binding proteins. Due to the laborious and time-consuming nature of wet experimental methods, more researchers should pay great attention to computational approaches for the prediction of lncRNA-protein interaction (LPI). An in-depth literature review in the state-of-the-art *in silico* investigations, leads to the conclusion that there is still room for improving the accuracy and velocity. This paper propose a novel method for identifying LPI by employing Kernel Ridge Regression, based on Fast Kernel Learning (LPI-FKLKRR). This approach, uses four distinct similarity measures for lncRNA and protein space, respectively. It is remarkable, that we extract Gene Ontology (GO) with proteins, in order to improve the quality of information in protein space. The process of heterogeneous kernels integration, applies Fast Kernel Learning (FastKL) to deal with weight optimization. The extrapolation model is obtained by gaining the ultimate prediction associations, after using Kernel Ridge Regression (KRR). Experimental outcomes show that the ability of modeling with LPI-FKLKRR has extraordinary performance compared with LPI prediction schemes. On benchmark dataset, it has been observed that the best Area Under Precision Recall Curve (AUPR) of 0.6950 is obtained by our proposed model LPI-FKLKRR, which outperforms the integrated LPLNP (AUPR: 0.4584), RWR (AUPR: 0.2827), CF (AUPR: 0.2357), LPIHN (AUPR: 0.2299), and LPBNI (AUPR: 0.3302). Also, combined with the experimental results of a case study on a novel dataset, it is anticipated that LPI-FKLKRR will be a useful tool for LPI prediction.

Keywords: lncRNA-protein interactions, multiple kernel learning, fast kernel learning, kernel ridge regression, gene ontology

1. INTRODUCTION

Long non-coding RNAs (lncRNAs) constitute a large class of transcribed molecules. They have a characteristic length of more than 200 nucleotides which do not encode proteins (St Laurent et al., 2015). Existing research has proven that lncRNAs can control gene expression during the transcriptional, post-transcriptional, and epigenetic procedures through interacting with the homologous RNA-binding proteins (Guttman and Rinn, 2012; Quan et al., 2015; Tee et al., 2015). A most recent research found that, a kind of lncRNA named lnc-Lsm3b can refrain the activity of the receptor RIG-I, by the induction of viruses during the regulation of immune response (Jiang et al., 2018). This is consistent with previous studies which have proven that lncRNAs are playing potential roles in complex human diseases (Li et al., 2013). Due to the laborious and time-consuming nature of wet experimental methods in molecular biology, many state-of-the-art computational researches have been carried out dealing with the conundrum, in an effort to enhance accuracy and time efficiency (Zou et al., 2012; Jalali et al., 2015; Han et al., 2018).

Since it is very difficult to extract any actual details on the 3D structures of lncRNAs and relative proteins, many sequence-based and secondary structure-based approaches for the prediction of lncRNA-protein interaction (LPI) have been published in the literature. Bellucci et al. have established the well-known catRAPID (Bellucci et al., 2011) by leveraging both physicochemical properties and secondary structure information, which could be employed as compound information to handle the problem of predicting LPI. Meanwhile, the hybrid schema RPISeq has been introduced by Muppirala et al. (2011), which employs both Support Vector Machines (SVM) and Random Forest (RF). Wang et al. have proposed a classifier combining Naive Bayes (NB) and Extended NB (ENB) classifier to extrapolate LPI (Wang et al., 2012). Lu et al. have established lncPro, which translates each LPI into numerical form, and applies matrix multiplication (Lu et al., 2013). Suresh et al. developed RPI-Pred based on SVM, by using the structure and sequence information of lncRNAs and proteins (Suresh et al., 2015).

In contrast to the aforementioned works, Li et al. have introduced the LPIHN by employing an heterogeneous network, assembled with a kind of random walk on lncRNA-protein association profile, with a restart mechanism (RWR) (Li et al., 2015). Ge et al. have used resource allocation mode on a dichotomous network, and they have published the algorithm as LPBNI (Ge et al., 2016). Lately, Hu et al. have proposed a kind of semi-supervised link prediction scheme, entitled LPI-ETSLP (Hu et al., 2017), which was soon upgraded to the IRWNRLPI. This method actually integrates RWR and matrix factorization (Zhao et al., 2018).

Zhang et al. have suggested two classes of state-of-the-art computational intelligence approaches (Zhang et al., 2017). The first includes supervised LPI binary classifiers, which do not require prior knowledge of interactions as negative instances (Bellucci et al., 2011; Muppirala et al., 2011; Wang et al., 2012; Lu et al., 2013; Suresh et al., 2015). second category includes semi-supervised approaches which combine known interactions

to suggest unknown LPI. The following are characteristic cases of this class: LPIHN (Li et al., 2015), LPBNI (Ge et al., 2016), LPI-ETSLP (Hu et al., 2017), and IRWNRLPI (Zhao et al., 2018).

Transfer learning (Jonathan et al., 1995), which can recognize and leverage skills or knowledge learned in previous tasks to novel tasks, is viewed as a kind of burgeoning machine learning branch. Whereas, zero-shot learning in pairwise learning with two-step Kernel Ridge Regression (KRR) (Stock et al., 2016), is a special type of transfer learning, constructing predictors from a dataset which contains both labeled and unlabeled samples. Hence, it is a kind of effective mechanism which can reduce the need of labeled data. In order to detect the pairwises of lncRNAs and proteins that can interact with each other, the state-of-the-art statistical methods have been exploited, such as Recursive Least Squares (RLS), Kronecker RLS, Sparse Representation based Classifier (SRC), and Multiple Kernel Learning (MKL). All these techniques have already been applied in predicting Protein-Protein Interactions (PPIs) (Ding et al., 2016; Liu X. et al., 2016), Drug-Target Interactions (DTIs) (Xia Z. et al., 2010; Laarhoven et al., 2011; Twan and Elena, 2013; Nascimento et al., 2016; Shen et al., 2017b), binding sites of biomolecules (Ding et al., 2017; Shen et al., 2017a) identification of disease-resistant genes (Xia J. et al., 2010), and microRNA-disease associations (Zou et al., 2015; Peng et al., 2017) with comparative consequences.

With reference to the above researches, we have enriched the categories of similarity measures adopted during LPI prediction. Integration of the heterogeneous kinds of similarity information is achieved by applying Fast Kernel Learning (FastKL) which deals with kernel weight optimization. This is done through the integration of the prediction architectures for weighting heterogeneous kernels. This research proposes a kind of two-step Kernel Ridge Regression (KRR) applied in the field of LPI prediction. LPI-FKLKRR has proven to be a more reliable and effective approach for LPI prediction, compared with other competitive methods. The core of the algorithm proposed herein has been evaluated on the benchmark dataset of LPIs. What is especially encouraging, is that many of the LPI predictions made by our method have been confirmed, with a high degree of correlation. Also, we have conducted a comparative testing on a novel dataset to illustrate the stable performance of the LPI-FKLKRR.

2. METHODS

In this section, we focus on the elaboration of architecture for our model. Its basic structural components-entities are the following: The known interactions matrix of LPI and the multivariate information that consists of lncRNA expressions, the local network, the sequence information and moreover the Gene Ontology (GO). It is imperative to combine all the similarity information together with the respective combination weights. Finally, we have developed and employed the *LPI with Fast Kernel Learning based on Kernel Ridge Regression Prediction* (LPI-FKLKRR) identification strategy, which utilizes a kind of two-stage Kernel Ridge Regression in LPI prediction.

2.1. Problem Specification

Suppose there are m lncRNAs and n proteins involved in LPI. We formally define two kinds of molecules as $\mathcal{L} = \{l_i \mid i = 1, 2, \cdots, m\}$ and $\mathcal{P} = \{p_j \mid j = 1, 2, \cdots, n\}$, respectively. Hence, the interactions between lncRNAs and proteins can be intuitively and succinctly expressed as an adjacency matrix $\mathbf{F}$ with $m \times n$, which can be formulated as Equation (1)

$$\mathbf{F} = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,j} & \cdots & f_{1,n} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,j} & \cdots & f_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ f_{i,1} & f_{i,2} & \cdots & f_{i,j} & \cdots & f_{i,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ f_{m,1} & f_{m,2} & \cdots & f_{m,j} & \cdots & f_{m,n} \end{bmatrix}_{m \times n} \quad (1)$$

where $f_{i,j}$ in matrix $\mathbf{F}$ corresponds to the prediction value of pairwise $\langle l_i, p_j \rangle$, $1 \leq i \leq m$, $1 \leq j \leq n$, and $m, n \in \mathbb{N}^*$. If lncRNA l_i can interact with protein p_j, the value of $f_{i,j}$ is marked as 1, otherwise it is marked as 0.

Obviously, the identification of new interactions between lncRNAs and proteins can be viewed as a task suitable for a recommender system (Koren et al., 2009) of a bipartite network, which can mine and detect the potential associated individuals. To this end, we use Multiple Kernel Learning (MKL) to design the optimization with respect to the prediction of LPI. In the following chapter, we will support the argument that the similarity matrix is equivalent to a kernel.

2.2. lncRNA Kernels and Protein Kernels

In order to conduct MKL, it is inevitable to construct similarity matrices of molecules in lncRNA and protein kernel spaces, respectively. Specifically, lncRNA expression, protein GO, lncRNA sequence, protein sequence, and known interactions between one lncRNA and all proteins are considered in our framework. In addition, the training adjacency matrix $\mathbf{F}_{train}$ is obtained by masking the known pairwise $\langle l_i, p_j \rangle$, where the partial known elements in the matrix are set to 0 for the validation set, which are represented in **Figure 1**.

2.2.1. Gaussian Interaction Profile Kernel

Interactions can be reflected in the connectivity behavior in the subjacent network (Laarhoven et al., 2011; Twan and Elena, 2013). For the lncRNAs, we extract information of lncRNA interactions corresponding to each row of the training adjacency matrix $\mathbf{F}_{train}$. Then we use a broadly applicable Gaussian Interaction Profile (GIP) kernel to device interaction kernel defined for lncRNA l_i and l_k $(i, k = 1, 2, \cdots, m)$. GIP about protein p_j and p_s $(j, s = 1, 2, \cdots, n)$ can be generated in a similar way. As a summary, each element value in GIP can be represented as follows:

$$\mathbf{K}_{GIP}^{lnc}(l_i, l_k) = exp(-\sigma_{lnc}\|\mathbf{F}_{l_i} - \mathbf{F}_{l_k}\|^2) \quad (2a)$$

$$\mathbf{K}_{GIP}^{pro}(p_j, p_s) = exp(-\sigma_{pro}\|\mathbf{F}_{p_j} - \mathbf{F}_{p_s}\|^2) \quad (2b)$$

where $\mathbf{F}_{l_i}$, $\mathbf{F}_{l_k}$ and $\mathbf{F}_{p_j}$, $\mathbf{F}_{p_s}$ are the matrices of interactions for lncRNA l_i, l_k and protein p_j, p_s, respectively. The Gaussian kernel

bandwidths σ_{lnc} and σ_{pro} are initialized to the value of 1 in the experiments. Practically, when employing 5-fold CV and LOOCV, the GIP kernel similarity should be recalculated each time based on the training samples.

2.2.2. Sequence Similarity Kernel

A sequence S with length d is an ordered list of characters, which can be written as $S = c_1 c_2 \cdots c_h \cdots c_d$ $(1 \leq h \leq d)$. Enlightened by state-of-the-art methods (Yamanishi et al., 2008; Nascimento et al., 2016), we use normalized Smith-Waterman (SW) score (Smith and Waterman, 1981) to measure the sequence similarity. The formulations are represented as follows:

$$\mathbf{K}_{SW}^{lnc}(l_i, l_k) = SW(S_{l_i}, S_{l_k})/\sqrt{SW(S_{l_i}, S_{l_i})SW(S_{l_k}, S_{l_k})} \quad (3a)$$

$$\mathbf{K}_{SW}^{pro}(p_j, p_s) = SW(S_{p_j}, S_{p_s})/\sqrt{SW(S_{p_j}, S_{p_j})SW(S_{p_s}, S_{p_s})} \quad (3b)$$

where $SW(\cdot, \cdot)$ stands for Smith-Waterman score; S_{l_i} and S_{l_k} are the sequences for lncRNA l_i and l_k; S_{p_j} and S_{p_s} denote the sequences for protein p_j and p_s.

2.2.3. Sequence Feature Kernel

We obtain the sequence feature kernel by extracting the feature of the sequences about lncRNAs and proteins. In practice, Conjoint Triad (CT) (Shen et al., 2007) and Pseudo Position-Specific Score Matrix (Pse-PSSM) (Chou and Shen, 2007) are adopted to describe lncRNA and protein sequences, respectively. Both Sequence Feature kernels (SF) $\mathbf{K}_{SF}^{lnc}$ and $\mathbf{K}_{SF}^{pro}$ are constructed based on a Radial Basis Function kernel (RBF) with bandwidth equals to 1.

2.2.4. lncRNA Expression Kernel

It is interesting to identify genes with concordant behaviors because different genes always show different behaviors (Lai et al., 2017). Expression profiles of lncRNAs refers to 24 cell types which come from NONCODE database (Xie et al., 2014). After expressing each lncRNA as a 24-dimensional expression profile vector, the kernel of lncRNAs expression $\mathbf{K}_{EXP}^{lnc}$ can be generated according to the RBF, and kernel bandwidth is also set to 1.

2.2.5. GO Kernel

Inspired by a former research (Zheng et al., 2012), similar Gene Ontology (GO) with proteins are expected to act in similar biological processes, or to reside in similar cell compartments, or to have similar molecular functions. Therefore, GO annotations are employed in this paper to generate a similarity matrix in protein space. The files of Gene Ontology (GO) terms have been downloaded from the GOA database (Wan et al., 2013).

Semantic similarity is always based on the overlap of the terms associated with two proteins (Wu et al., 2013). Jaccard value which we exploited in measuring the semantic similarity of two GO terms t_j and t_s related to proteins p_j and p_s is defined as follows:

$$Jaccard(t_j, t_s) = \frac{|t_j \cap t_s|}{|t_j \cup t_s|} \quad (4)$$

FIGURE 1 | Technical flow chart of our LPI prediction model. **(A)** LncRNAs and proteins belong to two separated and independent spaces, respectively; **(B)** Fast kernel learning is applied to estimate the weight of each kernel in the corresponding space; **(C)** Kronecker Product is adopted in generating the final kernel matrix; **(D)** Kernel Ridge Regression (KRR) is applied in LPI prediction.

where $t_j \cap t_s$ denotes the common terms between p_j and p_s, and $t_j \cup t_s$ refers to total number of terms of p_j and p_s. However, there has not been any formal definition with GO common terms $t_j \cap t_s$ given before.

We denote that, if the two sequences are completely consistent, two sequences S_1 and S_2 have common terms of GO. For example, given three sequences $S_1 = \langle 3, 1, 5 \rangle$, $S_2 = \langle 3, 2, 5 \rangle$, and $S_3 = \langle 3, 2, 5 \rangle$, if we only follow that all the corresponding locations of three sequences have non-zero values, then all three sequences have common terms. Nevertheless, for sequence S_2, it can be said that S_2 has common terms with S_3, but does not have common terms with S_1, because the second characters of S_1 and S_2 are different. Thus, we obtain a more sparse GO similarity matrix $\mathbf{K}_{GO}^{pro}$ which can facilitate the computation.

2.3. Fast Kernel Learning

In MKL, we need to find an optimal mapping vector $\mathbf{w}$, i.e., we require to choose a kind of optimal weighting strategy so that object similarity matrices can be appropriately constructed. Concretely, the vector of parameter weight values for lncRNA kernels and protein kernels are represented as $\mathbf{w}^{lnc}$ and $\mathbf{w}^{pro}$, respectively. We have already described that there are four kernels in lncRNA space including $\mathbf{K}_{GIP}^{lnc}$, $\mathbf{K}_{SW}^{lnc}$, $\mathbf{K}_{SF}^{lnc}$, and $\mathbf{K}_{EXP}^{lnc}$, and four kernels in protein space including $\mathbf{K}_{GIP}^{pro}$, $\mathbf{K}_{SW}^{pro}$, $\mathbf{K}_{SF}^{pro}$, and $\mathbf{K}_{GO}^{pro}$, respectively. The optimal lncRNA and protein kernels are given as follows:

$$\mathbf{K}_{lnc} = \sum_{a=1}^{4} w_a^{lnc} \mathbf{K}_a^{lnc}, \quad \mathbf{K}_a^{lnc} \in \mathfrak{R}^{m \times m} \tag{5a}$$

$$\mathbf{K}_{pro} = \sum_{a=1}^{4} w_a^{pro} \mathbf{K}_a^{pro}, \quad \mathbf{K}_a^{pro} \in \mathfrak{R}^{n \times n} \tag{5b}$$

where w_a^{lnc} and w_a^{pro} denote each element in $\mathbf{w}^{lnc}$ and $\mathbf{w}^{pro}$; $\mathbf{K}_a^{lnc}$ and $\mathbf{K}_a^{pro}$ correspond each kind of normalized similarity matrix among the heterogenous similarity kernels in lncRNA and protein spaces.

According to the description of Fast Kernel Learning (FastKL) (He et al., 2008), $\mathbf{w}$ is used as a substitute for the required optimal solution $\mathbf{w}^{lnc}$ or $\mathbf{w}^{pro}$, and $\mathbf{K}$ denotes kernel matrix $\mathbf{K}_{lnc}$ or $\mathbf{K}_{pro}$. FastKL is not only minimizing the distance between $\mathbf{K}$ and $\mathbf{Y}$, where $\mathbf{Y} = \mathbf{y}\mathbf{y}^{T}$, $\mathbf{y}$ is a matrix corresponds to all training set labels. It considers the regularization term $\|\mathbf{w}\|^2$ that is used to prevent overfitting. To this end, $\mathbf{w}$ can be drawn from the Formula 6 as follows:

$$\min_{\mathbf{w}, \mathbf{K}} \quad \|\mathbf{K} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{w}\|^2$$

$$s.t. \quad \sum_{a}^{J} w_a = 1 \tag{6}$$

where F represents Frobenius norm and λ is the tradeoff parameter. In practice, we set λ 10000 when selecting the optimal parameter value.

As a step forward to deduct Equation (6), since the Frobenius norm of a matrix equals to the trace about the product between the matrix itself and matrix of its transformation, i.e., $\|\mathbf{X}\|_F^2 = tr(\mathbf{X}\mathbf{X}^T)$, the object function with respect to the optimal solution $\mathbf{w}$ can be simplified as follows:

$$\min_{\mathbf{w}} \quad \mathbf{w}^T(\mathbf{A} + \lambda \mathbf{I})\mathbf{w} - 2\mathbf{b}^T\mathbf{w}$$

$$s.t. \quad \sum_{a}^{J} w_a = 1 \tag{7}$$

$$A_{u,v} = tr(\mathbf{K}_u^T \mathbf{K}_v)$$

$$b_v = tr(\mathbf{Y}^T \mathbf{K}_v)$$

where $tr(\cdot)$ is the symbol of the trace operator; $A_{u,v}$ represents each element in matrix $\mathbf{A}$; $\mathbf{K}_u$ and $\mathbf{K}_v$ denote two different kernel matrices.

Recapitulating the above statement, through gaining the final $\mathbf{w}^{lnc}$ and $\mathbf{w}^{pro}$, we have achieved the goal of MKL for fusing all kinds of similarity matrices so that the input matrix of KRR can be generated.
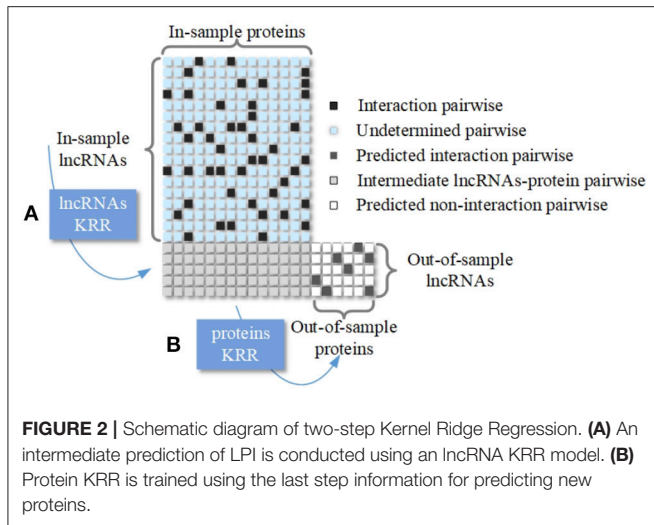
FIGURE 2 | Schematic diagram of two-step Kernel Ridge Regression. **(A)** An intermediate prediction of LPI is conducted using an lncRNA KRR model. **(B)** Protein KRR is trained using the last step information for predicting new proteins.

2.4. Kernel Ridge Regression

Stock et al. developed a scenario of pairwise learning, called Kernel Ridge Regression (KRR) (Stock et al., 2016), which can be applied in binary classification. The basic idea of KRR is to minimize a suitable objective function with an L_2-complexity penalty so that it can fit the labeled dyads as much as possible. Specifically, the KRR prediction for the LPI pairwise $\langle l_i, p_j \rangle$ has two steps which are shown in **Figure 2**.

In the first step, a prediction with respect to the new protein for all intermediate LPI pairwise is obtained as an $1 \times n$ vector $\mathbf{f}_{i,\cdot}$, which can be computed as follows:

$$\mathbf{f}_{i,\cdot} = \mathbf{k}_{lnc}^{\mathrm{T}}(\mathbf{K}_{lnc} + \lambda_l \mathbf{I})^{-1}\mathbf{F} \tag{8}$$

where $\mathbf{k}_{lnc}$ denotes the vector of lncRNA kernel evaluation between lncRNAs in the training set and a protein in the test set, and λ_l is the regularization parameter.

In the second step, we can obtain each element $f_{i,j}^*$ in the prediction matrix $\mathbf{F}^*$ by using another regularization parameter λ_p as following Equation (9):

$$f_{i,j}^* = \mathbf{k}_{pro}^{\mathrm{T}}(\mathbf{K}_{pro} + \lambda_p \mathbf{I})^{-1}\mathbf{f}_{i,\cdot}^{\mathrm{T}} \tag{9}$$

Considering the optimal lncRNAs and proteins kernels $\mathbf{K}_{lnc}$ and $\mathbf{K}_{pro}$, the general objective function of the two-step KRR is defined as follows:

$$\min_{\mathbf{F}^*} \sum_{(i,j,f) \in \mathbf{F}} (f_{i,j} - f_{i,j}^*)^2 + vec(\mathbf{F}^*)^{\mathrm{T}} \Xi^{-1} vec(\mathbf{F}^*) \tag{10}$$

where $vec(\cdot)$ is a vectorization operator that can rearrange the matrix elements in one row; $\mathbf{F}^*$ denotes the prediction of the original matrix $\mathbf{F}$ which can be estimated with the application of the LPI-KRR. Objective function in Equation (10) need to be minimized by iterations, and the iterations usually gets converged in about 5–10 iterations.

The kernel matrix Ξ that is used in Equation (10) is defined as Equation (11):

$$\Xi = \mathbf{K}_{pro} \otimes \mathbf{K}_{lnc}(\lambda_l \lambda_p \mathbf{I} \otimes \mathbf{I} + \lambda_p \mathbf{I} \otimes \mathbf{K}_{lnc} + \lambda_l \mathbf{K}_{pro} \otimes \mathbf{I})^{-1} \tag{11}$$

By using the lncRNAs, the proteins' kernels and the two regularization parameters λ_l and λ_p, each element in matrix $\mathbf{F}^*$ can be represented as Equation (12):

$$\mathbf{F}^* = \mathbf{K}_{lnc}(\mathbf{K}_{lnc} + \lambda_l \mathbf{I})^{-1}\mathbf{F}(\mathbf{K}_{pro} + \lambda_p \mathbf{I})^{-1}\mathbf{K}_{pro} \tag{12}$$

The LPI-FKLKRR calculation framework is illustrated in the following Algorithm 1.

Algorithm 1 Fast Kernel Learning based on Kernel Ridge Regression (LPI-FKLKRR).

Input: $\mathbf{K}_{GIP}^{lnc}$, $\mathbf{K}_{SW}^{lnc}$, $\mathbf{K}_{SF}^{lnc}$, $\mathbf{K}_{EXP}^{lnc} \in \Re^{m \times m}$ and $\mathbf{K}_{GIP}^{pro}$, $\mathbf{K}_{SW}^{pro}$, $\mathbf{K}_{SF}^{pro}$, $\mathbf{K}_{GO}^{pro} \in \Re^{n \times n}$; $\mathbf{F} \in \Re^{m \times n}$.
Output: $\mathbf{F}^*$.
1: Calculate $\mathbf{w}^{lnc}$ and $\mathbf{w}^{pro}$ and adjust the parameter λ by Eq.7;
2: Calculate $\mathbf{K}_{lnc}$ and $\mathbf{K}_{pro}$ by using Eq.5a and Eq.5b;
3: Calculate the prediction value in matrix $\mathbf{F}^*$ by Eq.12;
4: Adjust the parameters λ_l and λ_p by using Eq.10 and Eq.11, and produce the optimal $\mathbf{F}^*$.

---

3. RESULTS

This section provides a quantitative evaluation that employ benchmark dataset to assess our approach. We first show a result of 5-fold cross validation, then conduct an independent analyzing about performance of single kernel. Moreover, LPI-FKLKRR is not only compared with mean weighted model but also be assessed in parallel comparison including other outstanding methods. Furthermore, we utilize the case study to evaluate our method in predicting unknown lncRNA-protein interactions. What's more, there is also a comparison between LPI-FKLKRR and state-of-the-art work on a novel dataset.

3.1. Benchmark Dataset

Although there exists a high volume of web-based resources (Park et al., 2014), available datasets should be carefully selected. We have acquired the benchmark dataset according to the state-of-the-art work by Zhang et al. (2017). They have experimentally determined lncRNA-protein interactions with 1114 lncRNAs and 96 proteins from NPInter V2.0 (Yuan et al., 2014). Non-coding RNAs and sequence information of proteins were gleaned from NONCODE (Xie et al., 2014) and SUPERFAMILY database (Gough et al., 2001), respectively. Zhang et al. also removed lncRNAs and proteins whose expression or sequence information were unavailable in order to reduce the pressure of computation. Those lncRNAs and proteins with only one interaction were removed for the same reason. A dataset with 4158 lncRNA-protein interactions which contains 990 lncRNAs and 27 proteins were finally collected.

3.2. Evaluation Measurements

To gauge the stability of our model, 5-fold Cross Validation (5-fold CV) has been employed. The Area Under ROC curve (AUC) and Area Under the Precision-Recall curve (AUPR) measures have been utilized to evaluate our approach. We would like to emphasize that AUPR is more significant than AUC as a quality measurement because of the sparsity of the true lncRNA-protein interactions.

3.3. Experimental Environment

The proposed LPI-FKLKRR algorithm, has been implemented by using MATLAB as the development and compilation platform. All programs have been validated on a computer with 3.7 GHz 4-core CPU, 20 GB of memory, and 64-bit Windows Operating Systems.

3.4. Parameter Optimization

Grid search schema has been adopted to get the optimized values of the parameters λ_l and λ_p. The range of λ_l is from 20 to 980 while λ_p parameter ranges from 2 to 27. The criteria used to select the optimal values of λ_l and λ_p were the highest AUPR value and the lowest values of λ_l and λ_p, due to the fact that the smaller values of λ_l and λ_p, the less is the running time of the algorithm.

We have found that $\lambda_l = 20.89$ and $\lambda_p = 0.02$ are the best values for the two parameters (AUPR: 0.6950).

3.5. Performance Analysis

After testing different kinds of kernels on the benchmark dataset, we obtain that the AUPRs of GIP kernel, sequence feature kernel, sequence similarity kernel and gene expression & protein GO kernel are 0.6429, 0.4885, 0.5024, and 0.2663, respectively. The detailed results are listed in **Table 1**. It is obvious that GIP kernel has the highest AUPR value (among the single Kernels). Multiple kernels with the FastKL weighted model achieves AUPR equal to 0.6950, which is an outstanding performance. In **Figure 3**, we can see that the FastKL performs better than the other models. It is

TABLE 1 | The AUPR and AUC of different kernels on benchmark dataset.

Kernel type	AUPR	AUC
GIP kernel	0.6429	0.8671
Sequence feature kernel	0.4885	0.8250
Sequence similarity kernel	0.5024	0.8342
Gene expression & protein GO	0.2663	0.6626
Multiple kernels with mean weighted	0.6433	0.8840
Multiple kernels with FastKL weighted	**0.6950**	**0.9063**

Bold values represent the best value in columns.

FIGURE 4 | The kernel weights in the experiment of LPI-FKLKRR on benchmark dataset.
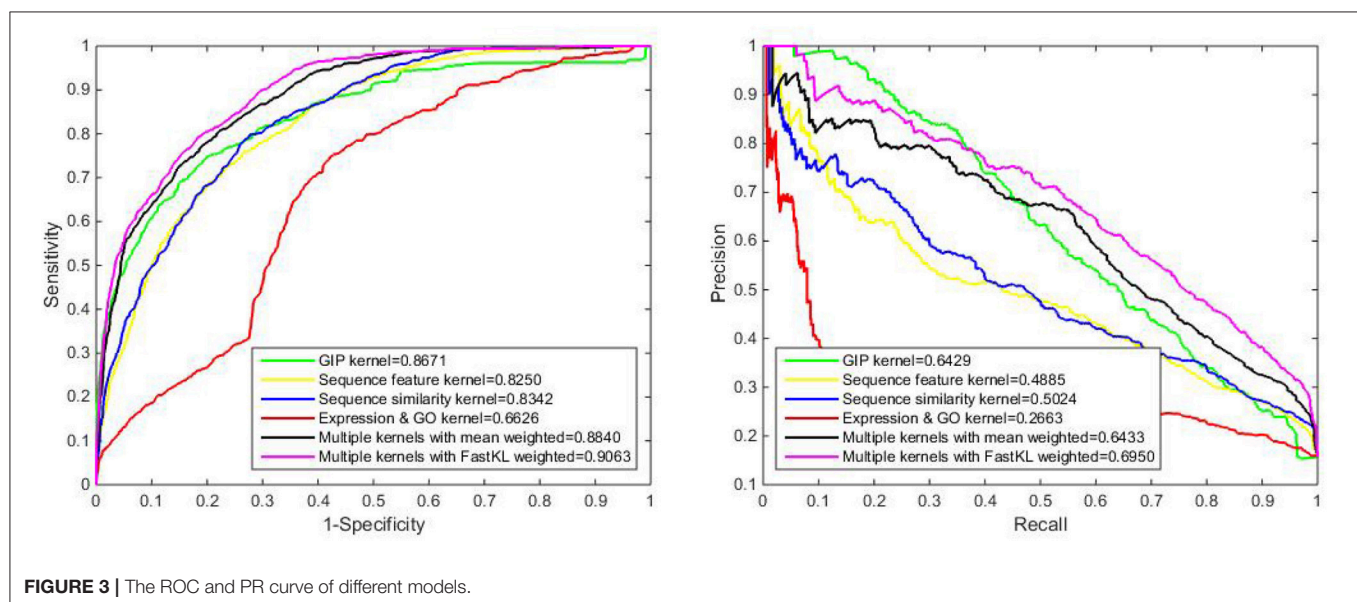
FIGURE 3 | The ROC and PR curve of different models.

clear that the FastKL is effective in improving the performance of LPIs prediction.

In addition, **Figure 4** shows the weight of each kernel, including lncRNA space and protein space in a 5-fold CV experiment. Conspicuously, weights of GIP kernel obtain the largest values on the lncRNA space. However, four kinds of protein similarity matrices equally divide the weights in protein space. This occasion should be explained that four kinds of protein similarity have low degree of overlapping in the representation space, i.e., each kind of protein similarity presents a specific aspect of protein feature.

3.6. Comparing to Existing Predictors

The comparison between our approach and other existing methods are showed in **Table 2**. It should be mentioned that the highest AUPR 0.6950 is achieved by our proposed approach, which is superior to all others. The AUPR values for the other established methods are the following: integrated LPLNP (AUPR: 0.4584) (Zhang et al., 2017), RWR (AUPR: 0.2827) (Gan, 2014), CF (AUPR: 0.2357) (Sarwar et al., 2001), LPIHN (AUPR: 0.2299) (Li et al., 2015), and LPBNI (AUPR: 0.3302) (Ge et al.,

2016). There are two well-founded reasons for the successful improved performance of our method. Firstly, FastKL effectively combines multivariate information by employing multiple kernel learning. Simultaneously, LPI-KRR is an effective prediction algorithm employing two-step KRR to fuse lncRNA and protein feature spaces. Due to the fact that there are extrapolation difficulties for the imbalanced datasets, PRC is more effective than ROC on highly imbalanced datasets. Therefore, we have obtained acquire competitive AUC value, compared to the state-of-the-art algorithms. From all the above we conclude that our approach can be a useful tool in the prediction of LPI.

3.7. Case Study

We have also used Local Leave-One-Out Cross-Validation (LOOCV) to evaluate the predictive performance. Local LOOCV masks the relationship between one protein and all lncRNAs. Our model is trained by the rest of the known information no matter if they are interacting or not and it is tested on a masked relationship. For a protein not appearing in the trial, our approach can predict the strength of interactions between this protein and gross 990 lncRNAs in the experiment. We have ranked these values of interactions in descending order,

TABLE 2 | Comparison to existing methods via 5-fold CV on benchmark dataset.

Method	AUPR	AUC
LPI-FKLKRR	**0.6950**	0.9063
Integrated LPLNP*	0.4584	**0.9104**
RWR*	0.2827	0.8134
CF*	0.2357	0.7686
LPIHN*	0.2299	0.8451
LPBNI*	0.3302	0.8569

Results are derived from Zhang et al. (2017). Bold values represent the best value in columns.

TABLE 3 | The AUPR and AUC of different kernels by local LOOCV on benchmark dataset.

Kernel	AUPR	AUC
GIP kernel	0.1690	0.5189
Sequence feature kernel	0.2814	0.6800
Sequence similarity kernel	0.3546	0.7333
Gene expression & protein GO	0.3101	0.7301
Multiple kernels with mean weighted	0.4956	0.7898
Multiple kernels with FastKL weighted	**0.5506**	**0.7937**

Bold values represent the best value in columns.

FIGURE 5 | The ROC and PR curve by local LOOCV on benchmark dataset.
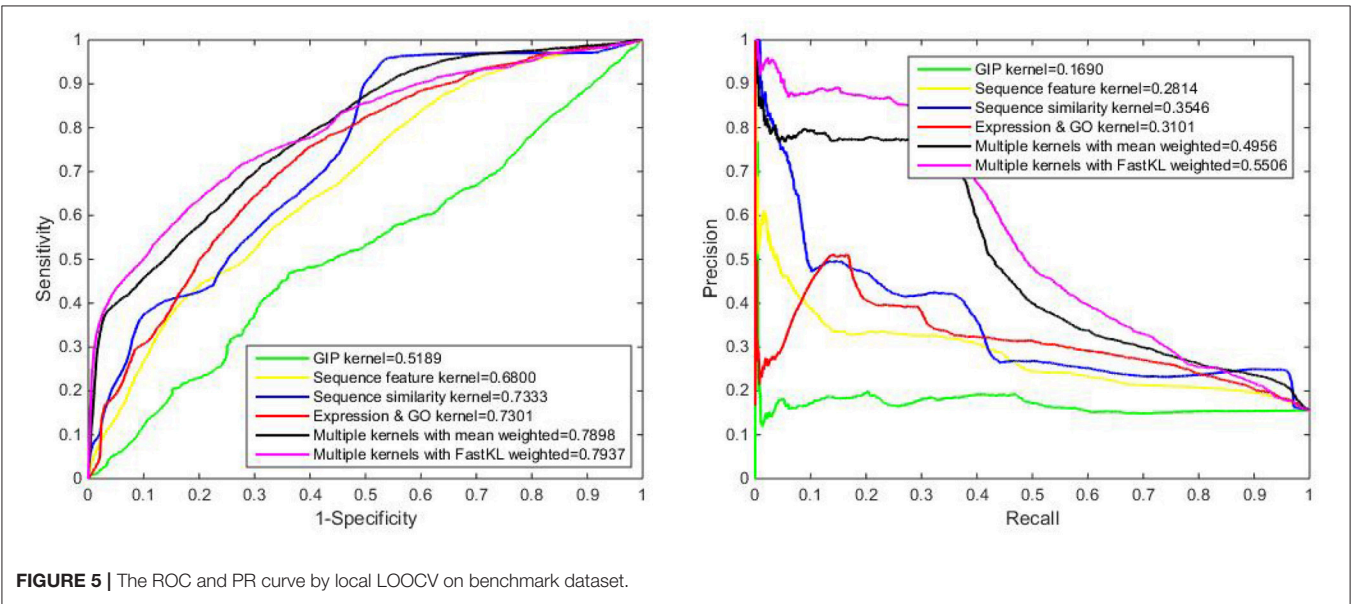
since high ranking is connected to high interaction possibility. In **Figure 5**, we can see that the performance of single kernel, average weighted kernels and weighted kernels with FastKL have failed. The FastKL weighted model using Multiple kernels, gains the best performance with values 0.5506 and 0.7937 for the AUPR and the AUC respectively. The detailed results are listed in **Table 3**.

As shown in **Table 4**, two cases of the top 20 interactions (including proteins ENSP00000309558 and ENSP00000401371), have been extrapolated by LPI-FKLKRR. Also, two cases in **Table 5** including lncRNAs, NONHSAT145960 and NONHSAT031708 of the top 10 interactions have been extrapolated by the LPI-FKLKRR. We check them up in the masked relationship between one protein and all lncRNAs, or one lncRNA and all proteins. Our approach achieves successful identification proportion equal to 11/20 and 12/20

on the proteins ENSP00000309558 and ENSP00000401371, respectively, and it achieves identification proportion equal to 6/10 and 6/10 on lncRNAs NONHSAT145960 and NONHSAT031708.

3.8. Speed Comparison on Benchmark Dataset

Practically, running speed is also play an important role in predicting LPI. The state-of-the-art methods of peer groups, such as LPLNP, can produce high-accuracy performances. Hence, the overall evaluation of the success of each approach, should also consider the Running Time (RT). Thus, a comparison between the RT of LPLNP and LPI-FKLKRR, has been performed. The comparative RT analysis between LPLNP and LPI-FKLKRR after running the available source code of LPLNP from the network, is illustrated in **Table 6**.

TABLE 4 | Top 20 interactions rank on protein ENSP00000309558 and ENSP00000401371.

lncRNA ID	Protein ID	Rank	Confirm?	lncRNA ID	Protein ID	Rank	Confirm?
NONHSAT011652	ENSP00000309558	1	Confirmed	NONHSAT002344	ENSP00000401371	1	Confirmed
NONHSAT027070	ENSP00000309558	2	Confirmed	NONHSAT104639	ENSP00000401371	2	–
NONHSAT104991	ENSP00000309558	3	Confirmed	NONHSAT027070	ENSP00000401371	3	Confirmed
NONHSAT001511	ENSP00000309558	4	Confirmed	NONHSAT104991	ENSP00000401371	4	Confirmed
NONHSAT079374	ENSP00000309558	5	–	NONHSAT101154	ENSP00000401371	5	–
NONHSAT009703	ENSP00000309558	6	Confirmed	NONHSAT041921	ENSP00000401371	6	Confirmed
NONHSAT138142	ENSP00000309558	7	Confirmed	NONHSAT042032	ENSP00000401371	7	–
NONHSAT104639	ENSP00000309558	8	Confirmed	NONHSAT131038	ENSP00000401371	8	Confirmed
NONHSAT135796	ENSP00000309558	9	Confirmed	NONHSAT084827	ENSP00000401371	9	–
NONHSAT077129	ENSP00000309558	10	–	NONHSAT021830	ENSP00000401371	10	Confirmed
NONHSAT023404	ENSP00000309558	11	–	NONHSAT001953	ENSP00000401371	11	Confirmed
NONHSAT063901	ENSP00000309558	12	Confirmed	NONHSAT145923	ENSP00000401371	12	Confirmed
NONHSAT099046	ENSP00000309558	13	–	NONHSAT039675	ENSP00000401371	13	–
NONHSAT031489	ENSP00000309558	14	–	NONHSAT135796	ENSP00000401371	14	Confirmed
NONHSAT041921	ENSP00000309558	15	Confirmed	NONHSAT011652	ENSP00000401371	15	Confirmed
NONHSAT013639	ENSP00000309558	16	–	NONHSAT044002	ENSP00000401371	16	–
NONHSAT027206	ENSP00000309558	17	–	NONHSAT112849	ENSP00000401371	17	–
NONHSAT134595	ENSP00000309558	18	–	NONHSAT114444	ENSP00000401371	18	Confirmed
NONHSAT054716	ENSP00000309558	19	–	NONHSAT007429	ENSP00000401371	19	Confirmed
NONHSAT122291	ENSP00000309558	20	Confirmed	NONHSAT123220	ENSP00000401371	20	–

TABLE 5 | Top 10 interactions rank on lncRNA NONHSAT145960 and NONHSAT031708.

lncRNA ID	Protein ID	Rank	Confirm?	lncRNA ID	Protein ID	Rank	Confirm?
NONHSAT145960	ENSP00000258962	1	–	NONHSAT031708	ENSP00000385269	1	Confirmed
NONHSAT145960	ENSP00000240185	2	Confirmed	NONHSAT031708	ENSP00000258962	2	–
NONHSAT145960	ENSP00000385269	3	–	NONHSAT031708	ENSP00000240185	3	Confirmed
NONHSAT145960	ENSP00000349428	4	Confirmed	NONHSAT031708	ENSP00000349428	4	–
NONHSAT145960	ENSP00000379144	5	Confirmed	NONHSAT031708	ENSP00000258729	5	Confirmed
NONHSAT145960	ENSP00000338371	6	Confirmed	NONHSAT031708	ENSP00000338371	6	–
NONHSAT145960	ENSP00000401371	7	Confirmed	NONHSAT031708	ENSP00000379144	7	–
NONHSAT145960	ENSP00000254108	8	–	NONHSAT031708	ENSP00000254108	8	Confirmed
NONHSAT145960	ENSP00000258729	9	Confirmed	NONHSAT031708	ENSP00000401371	9	Confirmed
NONHSAT145960	ENSP00000413035	10	–	NONHSAT031708	ENSP00000371634	10	Confirmed

Although LPLNP and LPI-FKLKRR have competitive AUC (according to results shown in **Table 2**) it is clear that the LPI-FKLKRR achieves better average running performance using only 11.48 s to accomplish the prediction task of LPI. This is much faster than the 352.93 s of the LPLNP (as shown in **Table 6**. Moreover, the standard deviation also manifest that LPI-FKLKRR is both fast and stable. Furthermore, considering the higher AUPR value of LPI-FKLKRR, we can strongly suggest that LPI-FKLKRR can be both a time-saving and useful tool for LPI prediction.

3.9. Evaluation on Novel Dataset

To support the results of the benchmark experiments, we have employed another dataset which is published by Zheng et al. The size of the novel dataset is larger than the benchmark dataset, which is shown in **Table 7**.

Originated from the same databases as the benchmark dataset, the novel dataset consists of 4467 LPIs, including 1050 unique lncRNAs and 84 unique proteins. We have conducted the comparison of LPI-FKLKRR and PPSNs (Zheng et al., 2017) by applying 5-fold CV on novel dataset, and list the results in **Table 8**. The AUC value for the LPI-FKLKRR algorithm is equal to 0.9669, which is higher than the one of PPSNs. What's more, the AUPR value which is equal to 0.7062 for the novel dataset proves the robustness performance of the LPI-FKLKRR on an imbalanced dataset.

Apart from the baseline methods that we have done test in **Figure 2**, we make a new comparison on the dataset that proposed by Zheng et al. with methods including NRLMF and CF. NRLMF, which is also capable of integrating various data sources, achieved good performance for both MDA prediction (Yan et al., 2017; He et al., 2018) and DTI prediction (Liu Y. et al., 2016). And CF method that has proposed by Sarwar et al., is another state-of-the-art work. From **Table 8**, we notice that no matter from the aspect of AUPR or AUC, the value of LPI-FKLKRR are higher than NRLMF (AUPR:0.4010, AUC:0.8287) and CF (AUPR:0.4267, AUC:0.8103).

Both the 5-fold CV and local LOOCV are also done in the novel dataset experiment. After testing different kinds of kernels on the novel dataset, we obtain that in the 5-fold CV, the AUPRs of GIP kernel, sequence feature kernel, sequence similarity kernel and gene expression & protein GO kernel are 0.6812, 0.4819, 0.4846, and 0.2379, respectively. Multiple kernels with the FastKL weighted model achieves AUPR equal to 0.7076, which is an outstanding performance. In **Figures 6, 7**, we can see that the FastKL performs better than the other models.

This result is consistent with the consequence on benchmark dataset.

CONCLUSIONS AND DISCUSSION

In this paper, we have proposed a novel prediction method for the prediction of lncRNAs-protein interactions by using Kernel Ridge Regression, combined with a multiple kernel learning approach (LPI-FKLKRR). LPI-FKLKRR employs fast kernel learning to fuse lncRNA and protein similarity matrices, respectively. A two-step Kernel Ridge Regression is adopted to forecast the interactions between lncRNAs and proteins. The 5-fold cross validation (5-fold CV) testing of the proposed LPI-FKLKRR algorithm, achieved very reliable and promising results when applied on the benchmark dataset (AUPR: 0.6950). Furthermore, LPI-FKLKRR achieves satisfactory prediction performances compared with the state-of-the-art approaches. A comparison on a novel dataset illustrates the stability performance of our model.

From the view point of the classification method about the prediction, the problem setting of lncRNA-protein interaction prediction can be the same with miRNA-disease interaction prediction and drug-target interaction prediction (Ezzat et al., 2018). For instance, CF method, which has proposed by Sarwar et al, has a recent work named MSCMF, which projects drugs and targets into a common low-rank feature space Zheng et al. (2013). This method can be transfered to the area of LPI prediction. Ezzat et al. have supposed that chemogenomic methods can be categorized into five types, including neighborhood models, bipartite local models, network diffusion models, matrix factorization models, and feature-based classification models. Consequently, in the future we will improve the predicting performance by adding information such as available 3D structure data, by constructing more heterogeneous similarity matrices, by changing weighting strategy or by drawing other effective regression models.

TABLE 7 | The information of two datasets in the experiment.

Dataset	Number of lncRNAs	Number of proteins	LPIs
benchmark dataset*	990	27	4,158
novel dataset*	1,050	84	4,467

The benchmark dataset and the novel dataset come from the paper of Zhang et al. (2017) and Zheng et al. (2017), respectively.

TABLE 8 | The AUPR and AUC of different methods on novel dataset.

Method	AUPR	AUC
LPI-FKLKRR	**0.7062**	**0.9669**
PPSNs	_[a]	0.9098
NRLMF	0.4010	0.8287
CF	0.4267	0.8103

[a]*AUPR is not exploited by Zheng et al. (2017). Bold values represent the best value in columns.*

TABLE 6 | Comparison of running time between LPI-FKLKRR and LPLNP in 10 times.

Method	Average running time(s)	Standard deviation(s)
LPI-FKLKRR	**11.48**	**0.2126**
LPLNP[a]	352.93	2.6656

[a]*The address of LPLNP is given by Zhang et al. (2017). Bold values represent the best value in columns.*

FIGURE 6 | The ROC and PR curve of different models with novel dataset by 5-fold CV.

FIGURE 7 | The ROC and PR curve by local LOOCV on novel dataset.
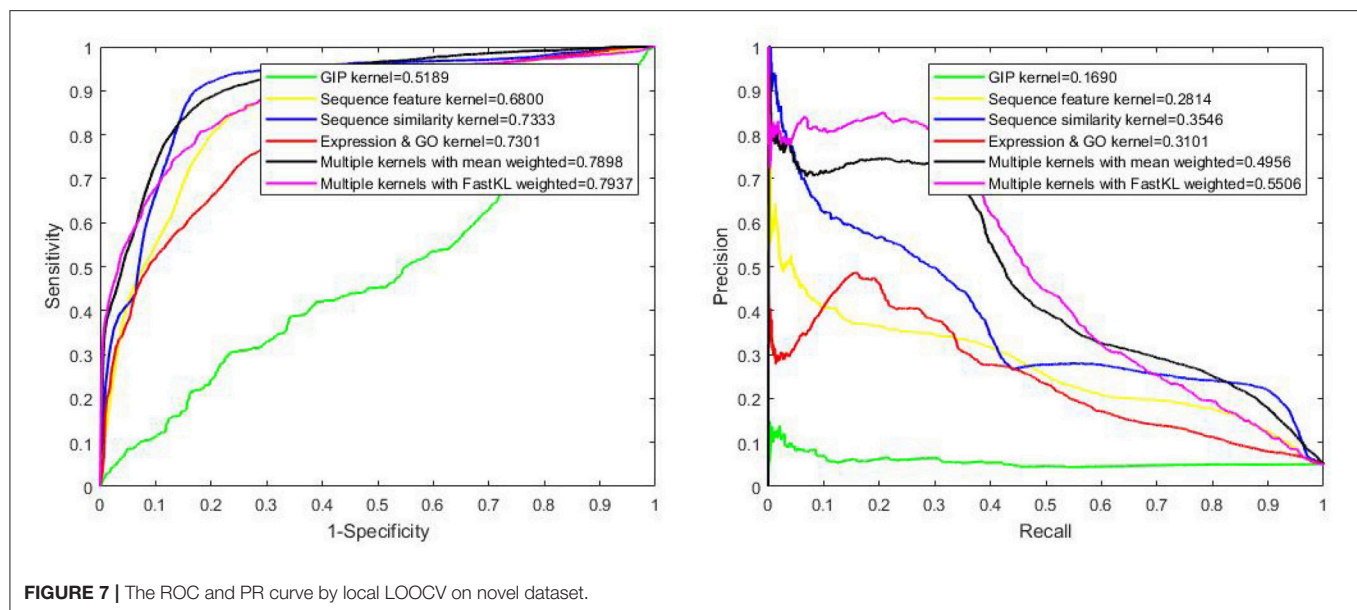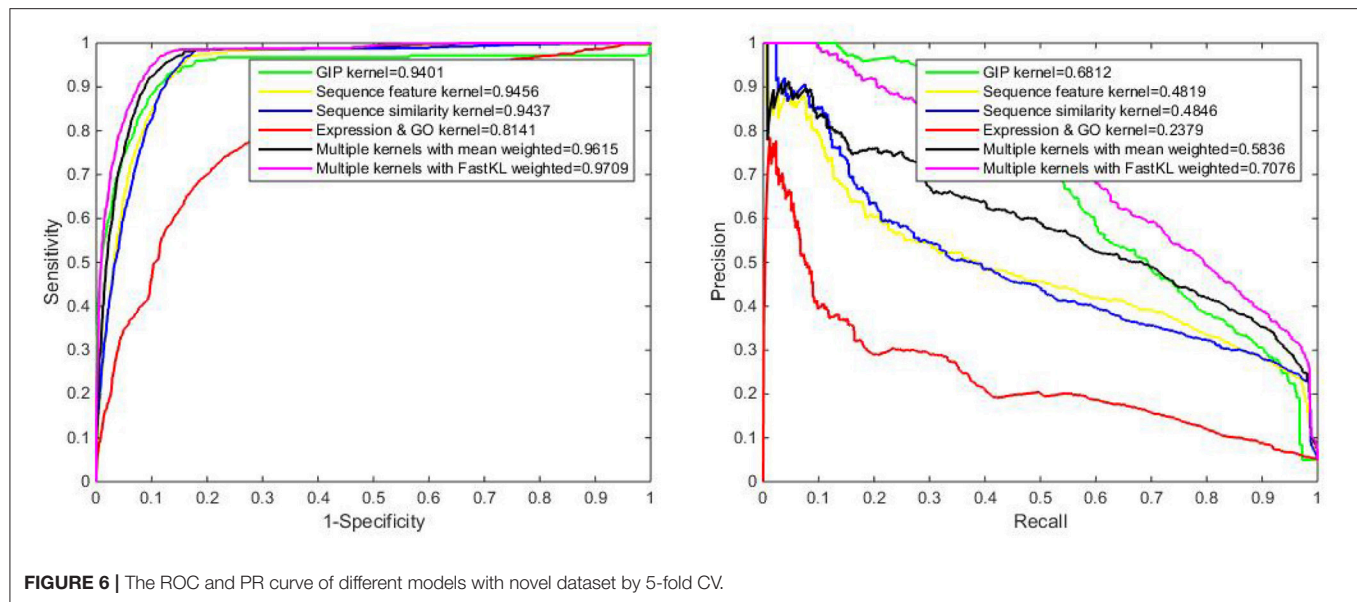
DATA AVAILABILITY STATEMENT

The datasets and codes for this study can be found in website https://github.com/6gbluewind/LPI_FKLKRR.

AUTHOR CONTRIBUTIONS

FG, YD, and CS conceived and designed the experiments. CS and YD performed the experiments and analyzed the data. FG and CS wrote the paper. FG and JT supervised the experiments and reviewed the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G. G. (2011). Predicting protein associations with long noncoding RNAs. *Nat. Methods* 8, 444–445. doi: 10.1038/nmeth.1611

Chou, K. C., and Shen, H. B. (2007). MemType-2L: a web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. *Biochem. Biophys. Res. Commun.* 360, 339–345. doi: 10.1016/j.bbrc.2007.06.027

Ding, Y., Tang, J., and Guo, F. (2016). Identification of protein-protein interactions via a novel matrix-based sequence representation model with amino acid contact information. *Int. J. Mol. Sci.* 17:1623. doi: 10.3390/ijms17101623

Ding, Y., Tang, J., and Guo, F. (2017). Identification of protein-ligand binding sites by sequence information and ensemble classifier. *J. Chem. Inform. Model.* 57, 3149–3161. doi: 10.1021/acs.jcim.7b00307

Ezzat, A., Wu, M., Li, X. L., and Kwoh, C. K. (2018). Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey. *Brief. Bioinform.* doi: 10.1093/bib/bby002. [Epub ahead of print].

Gan, M. (2014). Walking on a user similarity network towards personalized recommendations. *PLoS ONE* 9:e114662. doi: 10.1371/journal.pone.0114662

Ge, M., Ao, L., and Wang, M. (2016). A bipartite network-based method for prediction of long non-coding RNA-protein interactions. *Genomics Proteomics Bioinformatics* 14, 62–71. doi: 10.1016/j.gpb.2016.01.004

Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden markov models that represent all proteins of known structure. *J. Mol. Biol.* 313, 903–919. doi: 10.1006/jmbi.2001.5080

Guttman, M., and Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482, 339–346. doi: 10.1038/nature10887

Han, N., Miao, H., Yu, T., Xu, B., Yang, Y., Wu, Q., et al. (2018). Enhancing thermal tolerance of *Aspergillus niger* PhyA phytase directed by structural comparison and computational simulation. *BMC Biotechnol.* 18:36. doi: 10.1186/s12896-018-0445-y

He, B., Qu, J., and Zhao, Q. (2018). Identifying and exploiting potential miRNA-disease associations with neighborhood regularized logistic matrix factorization. *Front. Genet.* 9:303. doi: 10.3389/fgene.2018.00303

He, J., Chang, S. F., and Xie, L. (2008). "Fast kernel learning for spatial pyramid matching," in *Computer Vision and Pattern Recognition* (Anchorage, AL), 1–7.

Hu, H., Zhu, C., Ai, H., Zhang, L., Zhao, J., Zhao, Q., et al. (2017). LPI-ETSLP: lncRNA-protein interaction prediction using eigenvalue transformation-based semi-supervised link prediction. *Mol. Biosyst.* 13, 1781–1787. doi: 10.1039/C7MB00290D

Jalali, S., Kapoor, S., Sivadas, A., Bhartiya, D., and Scaria, V. (2015). Computational approaches towards understanding human long non-coding RNA biology. *Bioinformatics* 31, 2241–2251. doi: 10.1093/bioinformatics/btv148

Jiang, M., Zhang, S., Yang, Z., Lin, H., Zhu, J., Liu, L., et al. (2018). Self-recognition of an inducible host lncRNA by RIG-I feedback restricts innate immune response. *Cell* 173, 906–919.e13. doi: 10.1016/j.cell.2018.03.064

Jonathan, B., Rich, C., Tom, M., Lorien, Y. P., Daniel, L. S., and Sebastian, T. (1995). "Learning to learn: knowledge consolidation and transfer in inductive systems," in *Post-NIPS*95 Workshop on Transfer in Inductive Systems* (Denver, CO), 1–6.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Comput. J.* 42, 30–37. doi: 10.1109/MC.2009.263

Laarhoven, T. V., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics* 27, 3036–3043. doi: 10.1093/bioinformatics/btr500

Lai, Y., Zhang, F., Nayak, T. K., Modarres, R., Lee, N. H., and Mccaffrey, T. A. (2017). An efficient concordant integrative analysis of multiple large-scale two-sample expression data sets. *Bioinformatics* 33, 3852–3860. doi: 10.1093/bioinformatics/btx061

Li, A., Ge, M., Zhang, Y., Peng, C., and Wang, M. (2015). Predicting long noncoding RNA and protein interactions using heterogeneous network model. *Biomed. Res. Int.* 2015:671950. doi: 10.1155/2015/671950

Li, J., Xuan, Z., and Liu, C. (2013). Long non-coding RNAs and complex human diseases. *Int. J. Mol. Sci.* 14, 18790–18808. doi: 10.3390/ijms140918790

Liu, X., Zou, Q., Wu, Y., Li, D., and Zeng, J. (2016). An empirical study of features fusion techniques for protein-protein interaction prediction. *Curr. Bioinform.* 11, 4–12. doi: 10.2174/1574893611666151119221435

Liu, Y., Wu, M., Miao, C., Zhao, P., and Li, X. L. (2016). Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS Comput. Biol.* 12:e1004760. doi: 10.1371/journal.pcbi.1004760

Lu, Q., Ren, S., Lu, M., Zhang, Y., Zhu, D., Zhang, X., et al. (2013). Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics* 14:651. doi: 10.1186/1471-2164-14-651

Muppirala, U. K., Honavar, V. G., and Dobbs, D. (2011). Predicting RNA-Protein interactions using only sequence information. *BMC Bioinformatics* 12:489. doi: 10.1186/1471-2105-12-489

Nascimento, A. C. A., Prudâncio, R. B. C., and Costa, I. G. (2016). A multiple kernel learning algorithm for drug-target interaction prediction. *BMC Bioinformatics* 17:61. doi: 10.1186/s12859-016-0890-3

Park, C., Yu, N., Choi, I., Kim, W., and Lee, S. (2014). lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics* 30, 2480–2485. doi: 10.1093/bioinformatics/btu325

Peng, W., Lan, W., Zhong, J., Wang, J., and Pan, Y. (2017). A novel method of predicting microRNA-disease associations based on microRNA, disease, gene and environment factor networks. *Methods* 124:69. doi: 10.1016/j.ymeth.2017.05.024

Quan, M., Chen, J., and Zhang, D. (2015). Exploring the secrets of long noncoding RNAs. *Int. J. Mol. Sci.* 16, 5467–96. doi: 10.3390/ijms16035467

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). "Item-based collaborative filtering recommendation algorithms," in *International Conference on World Wide Web* (Hong Kong), 285–295.

Shen, C., Ding, Y., Tang, J., Song, J., and Guo, F. (2017a). Identification of DNA-protein binding sites through multi-scale local average blocks on sequence information. *Molecules* 22, 2079. doi: 10.3390/molecules22122079

Shen, C., Ding, Y., Tang, J., Xu, X., and Guo, F. (2017b). An ameliorated prediction of drug-target interactions based on multi-scale discrete wavelet transform and network features. *Int. J. Mol. Sci.* 18:1781. doi: 10.3390/ijms18081781

Shen, J., Zhang, J., Luo, X., Zhu, W., Yu, K., Chen, K., et al. (2007). Predicting protein-protein interactions based only on sequences information. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4337–4341. doi: 10.1073/pnas.0607879104

Smith, T. F., and Waterman, M. S. (1981). Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195–197.

St Laurent, G., Wahlestedt, C., and Kapranov, P. (2015). The landscape of long noncoding RNA classification. *Trends Genet.* 31, 239–251. doi: 10.1016/j.tig.2015.03.007

Stock, M., Pahikkala, T., Airola, A., Baets, B. D., and Waegeman, W. (2016). Efficient pairwise learning using kernel ridge regression: an exact two-step method. *arXiv:1606.04275*

Suresh, V., Liu, L., Adjeroh, D., and Zhou, X. (2015). RPI-Pred: predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res.* 43, 1370–1379. doi: 10.1093/nar/gkv020

Tee, A. E., Liu, P., Maag, J., Song, R., Li, J., Cheung, B. B., et al. (2015). The long noncoding RNA MALAT1 promotes hypoxia-driven angiogenesis by upregulating pro-angiogenic gene expression in neuroblastoma cells. *Cancer Res.* 75(15 Suppl.):146. doi: 10.1158/1538-7445.AM2015-146

Twan, V. L., and Elena, M. (2013). Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. *PLoS ONE* 8:e66952. doi: 10.1371/journal.pone.0066952

Wan, S., Mak, M. W., and Kung, S. Y. (2013). GOASVM: a subcellular location predictor by incorporating term. *J. Theor. Biol.* 323, 40–48. doi: 10.1016/j.jtbi.2013.01.012

Wang, Y., Chen, X., Liu, Z. P., Huang, Q., Wang, Y., Xu, D., et al. (2012). De novo prediction of RNA-protein interactions from sequence information. *Mol. Biosyst.* 9, 133–142. doi: 10.1039/C2MB25292A

Wu, X., Pang, E., Lin, K., and Pei, Z. M. (2013). Improving the measurement of semantic similarity between gene ontology terms and gene products: insights from an edge- and ic-based hybrid method. *PLoS ONE* 8:e66745. doi: 10.1371/journal.pone.0066745

Xia, J., Hu, X., Shi, F., Niu, X., and Zhang, C. (2010). Support vector machine method on predicting resistance gene against *Xanthomonas oryzae* pv. oryzae in rice. *Expert Syst. Appl.* 37, 5946–5950. doi: 10.1016/j.eswa.2010.02.010

Xia, Z., Wu, L. Y., Zhou, X., and Wong, S. T. (2010). Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst. Biol.* 4(Suppl. 2), 1–16. doi: 10.1186/1752-0509-4-S2-S6

Xie, C., Yuan, J., Li, H., Li, M., Zhao, G., Bu, D., et al. (2014). NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.* 42, 98–103. doi: 10.1093/nar/gkt1222

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, i232–i240. doi: 10.1093/bioinformatics/btn162

Yan, C., Wang, J., Ni, P., Lan, W., Wu, F., and Pan, Y. (2017). "DNRLMF-MDA:predicting microRNA-disease associations based on similarities of microRNAs and diseases," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics.* doi: 10.1109/TCBB.2017.2776101

Yuan, J., Wu, W., Xie, C., Zhao, G., Zhao, Y., and Chen, R. (2014). NPInter v2.0: an updated database of ncRNA interactions. *Nucleic Acids Res.* 42, D104–D108. doi: 10.1093/nar/gkt1057

Zhang, W., Qu, Q., Zhang, Y., Wang, W., Zhang, W., Qu, Q., et al. (2017). The linear neighborhood propagation method for predicting long non-coding RNA-protein interactions. *Neurocomputing* 273, 526–534. doi: 10.1016/j.neucom.2017.07.065

Zhao, Q., Zhang, Y., Hu, H., Ren, G., Zhang, W., and Liu, H. (2018). IRWNRLPI: integrating random walk and neighborhood regularized logistic matrix factorization for lncRNA-protein interaction prediction. *Front. Genet.* 9:239. doi: 10.3389/fgene.2018.00239

Zheng, W., Tao, P., Wei, H., and Zhang, H. (2012). High-throughput sequencing to reveal genes involved in reproduction and development inbactrocera dorsalis (diptera: Tephritidae). *PLoS ONE* 7:e36463. doi: 10.1371/journal.pone.0036463

Zheng, X., Ding, H., Mamitsuka, H., and Zhu, S. (2013). "Collaborative matrix factorization with multiple similarities for predicting drug-target interactions," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Chicago, IL), 1025–1033.

Zheng, X., Wang, Y., Tian, K., Zhou, J., Guan, J., Luo, L., et al. (2017). Fusing multiple protein-protein similarity networks to effectively predict lncRNA-protein interactions. *BMC Bioinformatics* 18(Suppl. 12):420. doi: 10.1186/s12859-017-1819-1

Zou, Q., Jiang, Y., Qu, Y., Huang, Y., and Wei, L. (2012). Computational analysis of miRNA target identification. *Curr. Bioinform.* 7, 512–525. doi: 10.2174/157489312803900974

Zou, Q., Li, J., Hong, Q., Lin, Z., Wu, Y., Shi, H., et al. (2015). Prediction of microRNA-disease associations based on social network analysis methods. *Biomed. Res. Int.* 2015:810514. doi: 10.1155/2015/810514

Interplay Between ncRNAs and Cellular Communication: A Proposal for Understanding Cell-Specific Signaling Pathways

Santiago Ramón y Cajal[1,2,3]*, Miguel F. Segura[4] and Stefan Hümmer[2,3]

[1] Department of Pathology, Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona, Barcelona, Spain,
[2] Translational Molecular Pathology, Vall d'Hebron Research Institute, Barcelona, Spain, [3] Spanish Biomedical Research Network Centre in Oncology (CIBERONC), Barcelona, Spain, [4] Group of Translational Research in Child and Adolescent Cancer, Vall d'Hebron Research Institute, Barcelona, Spain

Intercellular communication is essential for the development of specialized cells, tissues, and organs and is critical in a variety of diseases including cancer. Current knowledge states that different cell types communicate by ligand–receptor interactions: hormones, growth factors, and cytokines are released into the extracellular space and act on receptors, which are often expressed in a cell-type-specific manner. Non-coding RNAs (ncRNAs) are emerging as newly identified communicating factors in both physiological and pathological states. This class of RNA encompasses microRNAs (miRNAs, well-studied post-transcriptional regulators of gene expression), long non-coding RNAs (lncRNAs) and other ncRNAs. lncRNAs are diverse in length, sequence, and structure (linear or circular), and their functions are described as transcriptional regulation, induction of epigenetic changes and even direct regulation of protein activity. They have also been reported to act as miRNA sponges, interacting with miRNA and modulating its availability to endogenous mRNA targets. Importantly, lncRNAs may have a cell-type-specific expression pattern. In this paper, we propose that lncRNA–miRNA interactions, analogous to receptor–ligand interactions, are responsible for cell-type-specific outcomes. Specific binding of miRNAs to lncRNAs may drive cell-type-specific signaling cascades and modulate biochemical feedback loops that ultimately determine cell identity and response to stress factors.

Keywords: cancer, cellular communication, cell signaling, long non-coding RNA, microRNA, epigenetics

INTRODUCTION

Cancer is a complex disease and a major cause of death worldwide. Development of neoplastic disease is a multistep process involving the accumulation of numerous molecular changes. These changes impact cellular function within the tumor and its microenvironment, ultimately resulting in the hallmarks of cancer (Hanahan and Weinberg, 2011).

To date, most researchers have aimed to define the molecular mechanisms of tumorigenesis and cancer progression based on the classical gene expression theory – transcription of coding genes followed by protein synthesis. Following this approach, numerous genetic (e.g., mutations and genomic aberrations) and epigenetic alterations have been identified to have an association with carcinogenesis (Lengauer et al., 1998; Kagohara et al., 2018). In addition, alterations in post-transcriptional regulation of gene expression (e.g., splicing), mRNA translation (e.g., miRNAs)

and post-translational protein modification (e.g., phosphorylation) have been reported in almost every type of cancer. When different combinations are taken into consideration, the potential alterations are almost infinite. However, studies have mainly been based on around 20,000 protein-coding genes, corresponding to approximately 2% of the whole transcribed genome (Bertone et al., 2004; Carninci et al., 2005); the other transcripts include a large variety of non-coding RNAs (ncRNAs). Continuous generation of RNA sequencing (RNAseq) data shows that ncRNAs are strongly deregulated in pathological processes – particularly in multifactorial diseases like cancer (Cipolla et al., 2018). Hence, current limitations to deciphering the molecular mechanisms of cancer might be due to the fact that the putative implications of a large part of the genome remain undefined.

While ncRNA genes were for years considered as an irrelevant part the genome there is growing evidence that mammalian cells produce them in their thousands (Bertone et al., 2004; Carninci et al., 2005). Yet, in the absence of experimental verification of their function, most (>95%) of these transcripts are still considered transcriptional noise (Ponjavic et al., 2007). Studies dating back to the early 1990s indicated that certain lncRNAs may have similar functions to common mRNAs (Brannan et al., 1990; Brown et al., 1991), and since then, detailed studies on certain well-characterized non-coding transcripts have provided mechanistic insights. However, the variety in their mode of action, ranging from protein activity regulation to epigenetic control and regulation of other ncRNAs, implies that we are just beginning to understand their importance for a multitude of biochemical and cellular functions. Importantly, the role of these transcripts in certain tumor types is beginning to become apparent, and lncRNA expression profile has been proposed as a strong prognostic factor (Jiang et al., 2016; Bolha et al., 2017; Ali et al., 2018). It is therefore conceivable that elucidating the function of lncRNAs in normal cells and their deregulation in cancer cells will be one of the next milestones toward a more detailed understanding of the molecular mechanisms of cancer.

In this review, we propose a new role of ncRNAs in cancer. This model, analogous to the well-established ligand–receptor interactions, proposes intercellular communication via ncRNA interactions as a fundamental concept in cancer. This model could provide a hypothetical basis to explain the different types of biochemical feedback in tissues, which in turn could be linked to the differing response to drugs in tumors harboring similar genetic alterations, the different sites of tumor metastases, and the activation of different microRNA profiles depending on the tumor type and location.

THE NON-CODING TRANSCRIPTOME

RNAs comprise a diverse range of molecules. In addition to the well-characterized RNAs with established functions such as coding genes (mRNAs), protein synthesis (rRNAs and tRNAs), or mRNA splicing (snRNAs), a multitude of additional non-protein-coding RNAs has been described in recent years. According to their size, ncRNAs are categorized as small (<200 bp) or long non-coding RNAs (lncRNAs, >200 bp).

SMALL NON-CODING RNAS

The group of short ncRNAs consists of microRNAs (miRNAs), small interference RNAs (siRNAs), small nucleolar (snoRNAs), and Piwi-interacting RNAs (piRNAs) (Zamore and Haley, 2005). Of these, miRNAs have been best characterized in terms of their function, regulation and role in multiple human diseases such as cancer. Initiated by the discovery of the first miRNA, lin-4 in *Caenorhabditis elegans* (Lee et al., 1993; Wightman et al., 1993), large research efforts have led to a detailed characterization of miRNA biogenesis and regulatory functions in recent years (Ambros, 2004; Bartel, 2004; He and Hannon, 2004; Ebert and Sharp, 2012). Binding of miRNAs to specific sites in their target transcripts, called miRNA recognition elements (MRE), results in either transcript degradation or translational inhibition (Lee et al., 1993; Shivdasani, 2006). Currently there are just under 2,000 human high-confidence annotated miRNAs (Kozomara et al., 2018), and it is believed that they collectively regulate at least one third of the genome (Hammond, 2015). Importantly, gene expression profiling studies have demonstrated altered miRNA expression in a wide range of human diseases, including cancer.

miRNAs AND CANCER

miRNAs have been shown to participate in cancer throughout the various stages: from tumor origin, to immortalization, metastatic steps and interactions with the host tissue (Saliminejad et al., 2018), and they are able to regulate oncogenes and tumor suppressor genes. In the clinical setting, miRNA expression signatures are emerging as important diagnostic and prognostic predictors (Cho, 2007; Kong et al., 2012; Hayes et al., 2014; Drusco and Croce, 2017). Functional studies clearly support a relevant role of certain miRNAs in cancer. However, a remaining challenge is to understand the exact signaling pathways altered by miRNA deregulation. Several factors contribute to this complexity: (1) the 3′ UTR of a particular target gene contains multiple MREs; (2) multiple MREs can act either alone or cooperatively and (3) the same miRNA can regulate different targets (Doench and Sharp, 2004; Grimson et al., 2007).

In addition to cell-type-specific gene expression profiles, deregulation of miRNAs can result in tumor suppressive or oncogenic effects in a context-dependent manner (Zhang et al., 2007). Furthermore, the existence of feedback loops involving certain transcription factors such as c-Myc, which is both a regulator of miRNA expression and a target of miRNAs, adds yet another layer of complexity to the role of miRNAs in cancer (Jackstadt and Hermeking, 2015).

miRNA TRAFFICKING

While the main mechanism of action is to control mRNA stability or translation in the cytoplasm, miRNAs can be found in unexpected cellular compartments such as the nucleus, mitochondria, and endoplasmic reticulum (Leung, 2015).

They are also found in the extracellular space – this was first described in 2008, when it was proposed that circulating miRNAs may serve as biomarkers of certain cancers (Mitchell et al., 2008). Since then, miRNAs have been discovered in various extracellular environments including blood (Weber et al., 2010), urine (Gidlof et al., 2011), saliva (Park et al., 2009; Patel et al., 2011), and ascitic fluid (Husted et al., 2011). These findings opened up two new fields of investigation on miRNA. First, as they are easy to detect in body fluids, clinical research has focused on the use of extracellular miRNAs as biomarkers as an alternative, non-invasive method for diagnosis and disease monitoring (Duttagupta et al., 2011; Ajit, 2012). Second, because miRNAs are found in the extracellular space, it was proposed that they may act not only in the cells in which they are transcribed but also in neighboring cells. This intriguing idea of horizontal transfer of genetic material subsequently gained major attention. To date, five non-exclusive mechanisms of miRNA release from donor cells have been proposed: (i) miRNA bound to RNA-protein complexes (e.g., in complex with Argonaut) (Arroyo et al., 2011; Turchinovich et al., 2011); (ii) transport via lipid or lipoprotein particles (Rayner and Hennessy, 2013); (iii) vesicles shed directly from the plasma membrane (Hunter et al., 2008; Callis et al., 2009; Shefler et al., 2010; Jaiswal et al., 2012); (iv) vesicles of endosomal origin (exosomes) (Valadi et al., 2007; Skog et al., 2008; Rechavi et al., 2009) and (v) vesicles from apoptotic bodies (Zernecke et al., 2009; Hergenreider et al., 2012). Currently, the coexistence of these different forms of miRNA transport is supported in the literature, but improved biochemical methods and molecular tools with higher temporal and spatial resolution are required to strengthen the evidence (Tkach and Thery, 2016).

Many groups have clearly demonstrated that isolated miRNA containing fractions (e.g., exosomes) from donor cells are capable of inducing phenotypic alterations in the recipient cells. However, studies on the underlying mechanisms are rare thus far, and basic questions concerning the amount of miRNA (signaling-like or enzymatic function) or the type of miRNA [individual or pool of miRNA(s)] remain unaddressed.

LONG NON-CODING RNAs

All RNAs longer than 200 nucleotides that are not translated into proteins are collectively categorized as lncRNAs (Carninci et al., 2005; Guttman et al., 2009; Ponting et al., 2009; Cabili et al., 2011; Derrien et al., 2012; Housman and Ulitsky, 2016; Lagarde et al., 2017). Similarly to mRNAs, lncRNAs are transcribed by RNA polymerase II and are often subject to post-transcriptional modification like 5′ capping, 3′ poly-adenylation and splicing (Quinn and Chang, 2016). Although less well-studied than miRNAs, several thousand lncRNAs have already been described, and thanks to technical advances in RNAseq techniques and computational prediction methods, the total number of lncRNAs identified continues to increase (Quinn and Chang, 2016; Hon et al., 2017). Solely defined by their length, lncRNAs constitute the largest class of ncRNAs

in the mammalian genome, and can be further categorized into long intergenic ncRNAs (lincRNAs), antisense RNAs (asRNAs), pseudogenes, and circular RNAs (circRNAs) (Quinn and Chang, 2016; Lagarde et al., 2017). lncRNAs fulfill a variety of functions by interacting with DNA, RNA and proteins, and they may be described according to their mechanism of action toward their interacting molecule as enhancers, decoys, guides or scaffolds (Wang and Chang, 2011; Fok et al., 2017; Kopp and Mendell, 2018).

The first unsupervised clustering analysis of individual transcripts in different tissues revealed that 78% of lncRNAs (in comparison to 19% of mRNAs) were expressed in a tissue-specific manner (Cabili et al., 2011). As sequencing techniques advanced, this specificity of expression has been observed at the individual cell level, and even differential cell-to-cell expression has been observed by Lv et al. (2016); we could confirm this in our own (unpublished) studies in the triple negative breast cancer cell line MB-MDA231. In the field of cancer biology, lncRNA expression profile has recently been proposed as a strong prognostic factor and even as a therapeutic target (Chen et al., 2014; Yarmishyn and Kurochkin, 2015).

Despite the growing catalog of lncRNAs, the majority of detected transcription products remain functionally unannotated. Gene function prediction based on sequence homology for protein-coding genes is challenging (Ulitsky, 2016). Therefore, lncRNA classification requires either specialized computational tools or genome-wide functional studies as currently performed by the use of gene knockout techniques (e.g., CRISPERi) (Gilbert et al., 2013; Goyal et al., 2017).

lncRNAs IN CANCER

Differential expression of lncRNAs has been described in a variety of pathological conditions including cardiovascular, autoimmune, neurodegenerative diseases and particularly in cancer (Walsh et al., 2014; Nguyen and Carninci, 2016; Zhang et al., 2016; Aune et al., 2017; Bhan et al., 2017; Viereck and Thum, 2017; Wan et al., 2017). The first reported lncRNA with an aberrant expression in cancer was prostate cancer associated 3 (PCA3) (Bussemakers et al., 1999), which was identified via differential display analysis of transcripts in normal human prostate cancer. PCA3 was the first FDA-approved lncRNA-based biomarker for use in clinical practice. Since then, it has proven a useful, non-invasive test for prostate cancer (Wei et al., 2014). In subsequent years, many other lncRNAs have been identified as having a highly predictive value in the diagnosis of different cancers. Among those, deregulated expression of the lncRNA HOTAIR, originally described in breast cancer, is associated with cancer progression in 26 human tumor types (Gupta et al., 2010; Bhan and Mandal, 2015; Teschendorff et al., 2015). Recently, the application of next-generation sequencing in a variety of different cancer transcriptomes uncovered thousands of lncRNAs with aberrant expression in different cancer types (Huarte, 2015). These numbers increase further when single nucleotide polymorphisms

(SNPs) that are associated with cancer are taken into account (Freedman et al., 2011).

Regarding metastasis, some lncRNAs have been associated with more aggressive, metastatic tumors and even with cancer cell colonization to specific organ sites (Li et al., 2018). For example, in colorectal carcinoma, expression of the lncRNA CCAT2 correlated with a higher incidence of liver metastasis (Ling et al., 2013). Another example is the association of elevated levels of HOTAIR with a higher incidence of liver metastasis in gastric cancer (Zhang et al., 2015) and with brain metastasis in non-small cell lung carcinoma (Nakagawa et al., 2013).

Despite the enormous number of lncRNAs described as having aberrant expression in different cancer types, to date, only a few have been functionally characterized (reviewed in Huarte, 2015); existing studies have identified tumor-suppressor and oncogenic functions of lncRNAs (Prensner and Chinnaiyan, 2011; Huarte, 2015). In many cases, the functional role of these lncRNAs has been linked to well-known oncogenic pathways like p53 and c-myc or participating in different steps of classical cancer processes such as epithelial mesenchymal transition (EMT). Functional analysis will probably expand with the recent application of the CRISPR/Cas9 system, which will provide the tools required to study the function of lncRNAs in genome-wide studies. These unsupervised studies will be crucial to functionally annotate the role of lncRNAs in cancer and shed further light on the regulation of the underlying molecular events (Han et al., 2014; Ho et al., 2015).

lncRNA AND miRNA INTERPLAY

Different lncRNAs are known to interact with DNA, RNA, and proteins; therefore, the functions of the class of lncRNAs appear to be pleiotropic, ranging from chromatin remodeling to the regulation of transcription, splicing, and translation (Wang and Chang, 2011; Fok et al., 2017; Kopp and Mendell, 2018). A subclass of lncRNA has recently been shown to regulate gene expression in *trans* by acting as miRNA "sponges" (Karreth et al., 2011; Salmena et al., 2011; Tay et al., 2011; Karreth and Pandolfi, 2013; Su et al., 2013; Yoon et al., 2014; Ulitsky, 2018). These lncRNAs belong to a group of RNAs named ceRNA (competitor of endogenous RNA) (Tay et al., 2014; Thomson and Dinger, 2016; Zhong et al., 2018). While there is no unifying definition of ceRNAs, their function as miRNA "sponges" minimally requires their cytoplasmic localization and the presence of MRE in their sequence. ceRNAs contain MREs for one or multiple miRNAs, and binding is thought to sequester miRNAs and thereby enable translation of endogenous miRNA targets (**Figure 1**).

In recent years, this concept has been described for the expression of different genes involved in tumor progression (de Giorgio et al., 2013; Cheng et al., 2015; Wang et al., 2016; Song et al., 2017). An example of an oncogenic lncRNA is UCA1, which controls the availability of miR-18a and thereby determines the expression of the oncogene YAP1 (Zhu et al., 2018; **Figure 2A**). In contrast, the tumor suppressor gene PTEN is regulated by the lncRNA CCAT2 by acting as a competing
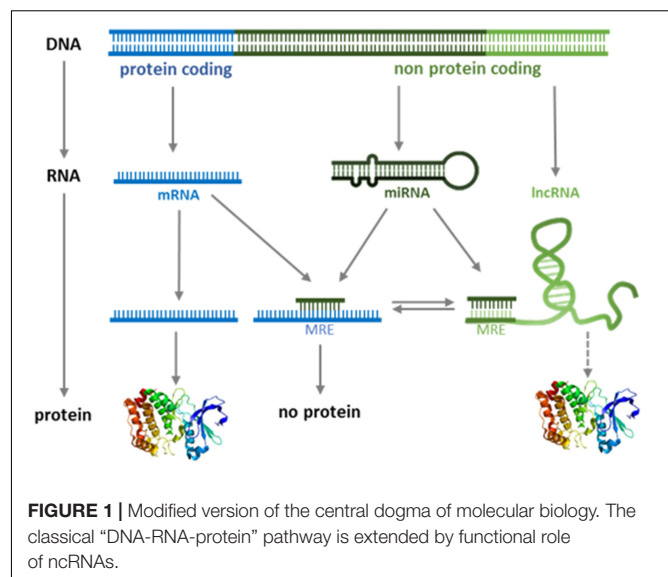
FIGURE 1 | Modified version of the central dogma of molecular biology. The classical "DNA-RNA-protein" pathway is extended by functional role of ncRNAs.
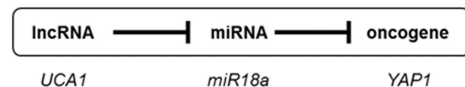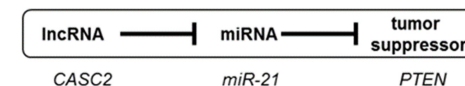
FIGURE 2 | Competing lncRNAs (ceRNAs) in cancer. LncRNAs can act either as oncogenes **(A)** or as tumor suppressor **(B)** as exemplified by the role UCA1 and CASC2. A detailed literature survey of the reported roles ceRNAs in cancer is summarized in **Supplementary Table S1**.

RNA for miR-21 (**Figure 2B**; Xie et al., 2017). In recent years, many more ceRNAs have been described in cancer. We have summarized the known lncRNA/miRNA/mRNA combinations in **Supplementary Table S1**.

In addition to sequestering miRNAs, lncRNAs have also been reported to compete with miRNAs by binding directly to mRNAs (Faghihi et al., 2010). miRNAs have been reported to induce destabilization of lncRNAs, yet some lncRNAs contain miRNAs precursors (Cui et al., 2016). This suggests a complex interplay between lncRNAs and miRNAs, which ultimately determines stability and translation of protein-coding mRNAs. Notably, recent studies have revealed that ceRNAs have significant roles in cancer pathogenesis (Cheng et al., 2015; Wang et al., 2016). For example, alterations in the expression of key factors in oncogenic signaling pathways, like BRAF, have been linked to changes in the level of ceRNAs (Ergun and Oztuzcu, 2015; Karreth et al., 2015). We are just starting to understand these complex molecular interactions, their place in functional regulatory networks controlling cellular processes, and their implications in

cancer (Liz and Esteller, 2016; Yang et al., 2016; Cao et al., 2017; Chan and Tay, 2018).

PROPOSAL: TISSUE-SPECIFIC INTERPLAY BETWEEN miRNA AND lncRNA SUPPORTS CONTEXT-DEPENDENT CELL SIGNALING

Currently, the mechanisms involved in cancer origin and progression have been elaborated according to the alterations found in protein-coding genes, barely 2% of the translated genome. Following this approach, most of the described alterations merge into few biochemical routes such as the PI3K/AKT/mTOR or RAS/MAPK pathways. However, the outcome of these alterations is sometimes tissue-specific and cell-context-dependent. It is, for example, still unclear how certain oncogenic mutations progress to tumor formation only in a particular set of tissues. There are clear examples with germline mutations in genes like adenomatous polyposis coli (APC), cadherin 1 (CDH1), BRCA1, von Hippel-Lindau tumor suppressor (VHL), and ataxia telangiectasia mutated (ATM) which are causative for the development of cancer in specific types of tissue (Schneider et al., 2017). Another paradigmatic example is the initial good response to specific inhibitors in melanomas harboring the BRAF^{V600E} mutation, compared to the protumor effect of the same inhibitors in colon adenocarcinomas carrying similar BRAF mutations (Prahallad et al., 2012; Sclafani et al., 2013). Finally, the development of metastases in different types of cancer is often restricted to certain organs (organotropic metastasis), and even clonal subpopulations within the primary tumor display preferences for certain organs (Obenauf and Massague, 2015).

Therefore, in addition to the classical oncogenic signaling pathways (PI3K/AKT/mTOR or RAS/MAPK), additional layers of regulation must exist, to explain tissue- and organ-specific processes. We hypothesize that these cell-type-specific outcomes may be caused by interplay between lncRNAs and miRNAs. Building on the concept of ceRNAs, the tissue and cell-type-specific expression of lncRNAs might be a key mechanism to support tissue-specific regulation of oncogenic signaling pathways. In this sense, the translational profile (i.e., proteome) of miRNA-regulated mRNAs can be controlled by the cell-type-specific expression of lncRNAs. Consequently, the interplay between lncRNAs and miRNAs might affect signaling cascades by regulating the abundance of proteins within these pathways in a cell-type-specific manner (**Figure 3A**).

This core mechanism, acting on the intracellular level, can also be extended to communication between different cells. This is because miRNAs not only act in the cells in which they are transcribed but can also be transferred into different cells (intercellular level). In this regard, the interplay between miRNAs and lncRNAs suggests many parallelisms to ligand–receptor interactions. Ligands like cytokines, hormones and growth factors are released as soluble factors by the donor cell,

FIGURE 3 | Proposal on the role of tissue specific ceRNAs in cancer **(A)** Tissue specific expression of lncRNAs can induce malignant transformation in cells expressing an equal set of miRNAs and mRNAs. **(B)** Tissue specific response to secreted miRNAs dependent on the expression profile of lncRNAs.
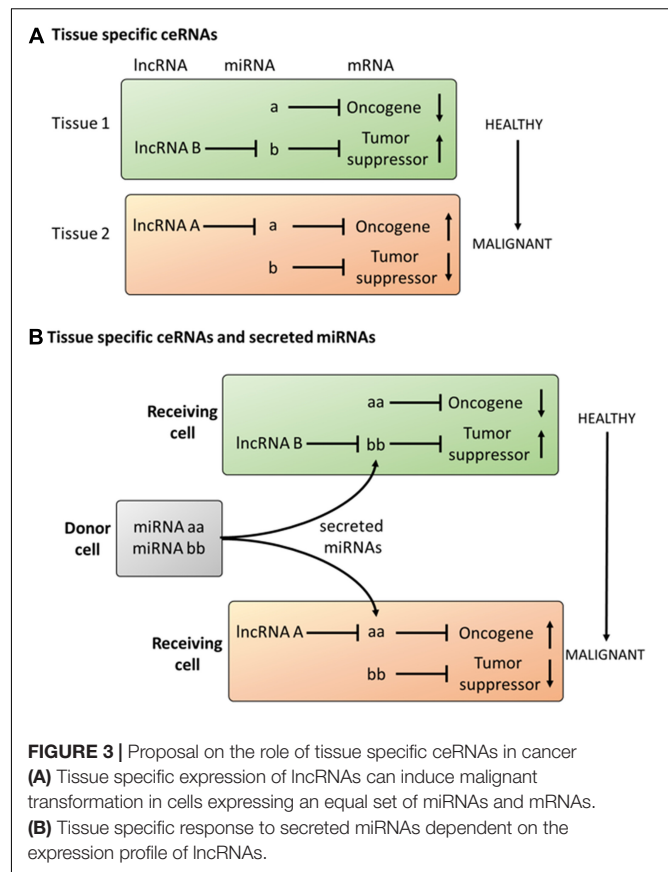
selectively interact with the target cell receptors and activate signaling cascades which ultimately alter the phenotype of the target cell. miRNAs have been shown to be released into the extracellular space, and lncRNAs are expressed in an organ-, tissue- or cell-type-specific manner. As described above, the abundance of lncRNAs may ultimately determine the effect of miRNAs on the expression of protein-coding genes. Therefore, as occurs in the ligand-receptor model, exposure to the same miRNA may result in cell-type-specific alterations dependent on the expression of lncRNAs (**Figure 3B**).

In summary, the interplay between lncRNAs and miRNAs at an intra- and intercellular level may provide a framework for understanding context-specific phenomena in cancer. In the following section, we will exemplify these putative mechanisms in two concrete cases.

INTRACELLULAR INTERPLAY BETWEEN lncRNAs AND miRNAs – BRAF^{V600E} IN COLON ADENOCARCINOMA VS MELANOMA

Transduction of a signal from an activated receptor is dependent on the levels of kinases and phosphatases in these pathways and is regulated by positive and negative feedback loops. Alterations in the stoichiometry of factors involved in signaling cascades

FIGURE 4 | The potential role of lncRNAs in the failure of Vemurafenib in the treatment of colon cancer. Upper panel: BRAF[V600E] mutation induces malignant transformation in melanoma and colon cancer. Lower panel (left): treatment with Vemurafenib inhibits malignant transformation in melanoma. Bypassing the inhibition of BRAF[V600E] by Vemurafenib might be prevented by miRNAs (e.g., miR-34a or miR-218a) inhibiting the expression of ARAF (Lal et al., 2011; Agarwal et al., 2015). Lower panel (right): sequestering of the ARAF controlling miRNA(s) by lncRNAs in colon cancer enables the bypass of Vemurafenib inhibition, resulting in malignant transformation.

can be crucial for the ultimate cellular effect. A paradigmatic example for such alterations in the same oncogenic pathway is described for the BRAF[V600E] mutation. While treatment of

BRAF[V600E]–bearing tumors with BRAF inhibitors has a good response in melanomas, a protumor effect has been described in colon adenocarcinomas. It has been proposed that the local

feedback in the different cell-signaling pathways could partly explain this unexpected contradictory effect (Schneider et al., 2017). In this respect, miRNAs have been described as key players in fine-tuning the expression of proteins involved in the Ras/Raf signaling pathway (reviewed in Masliah-Planchon et al., 2016). However, the expression of the majority of these miRNAs is not specific for single tissues; therefore, additional layers of regulation must exist, to explain the different responses. We propose that the tissue-specific interplay between ncRNAs might partly explain the different sensitivity to BRAF inhibitors in colon adenocarcinoma and melanoma harboring the same oncogenic mutation in BRAF (**Figure 4**).

INTERCELLULAR INTERPLAY BETWEEN lncRNAs AND miRNAs – ORGANOTROPIC METASTASIS

miRNAs are released into the extracellular space and transferred to target cells. lncRNAs (the "receptors") show organ-, tissue- and cell-type-specific expression. Binding of miRNAs to lncRNAs (akin to a ligand-receptor complex) occurs via a sequence-specific MRE within the lncRNA. Finally, this interaction results in miRNA sequestration and/or degradation of miRNA or lncRNA. As the ligand-receptor model would predict, these interactions should result in the activation of signaling cascades enabling significant alterations in biochemical and cellular functions.

Applying this model to cancer, the mechanisms underlying organotropic metastasis might be explained in part by the interplay between miRNAs and lncRNAs. Even though there is no direct experimental evidence for this hypothesis, there are theoretical possibilities for how this interplay could contribute to the development of site-specific metastasis.

High levels of secreted miRNAs from cancer cells have been reported for almost all types of tumor cells. Upon arrival of a disseminated tumor cell to a distant tissue (e.g., lung, bone, or brain), these miRNAs are first taken up by tissue-specific endothelial cells. Transfer of miRNAs to endothelial cells in turn has been shown to alter the expression of proteins required for maintenance of the endothelial barrier (Zhou et al., 2014; Tominaga et al., 2015). Therefore, the expression profile of competing lncRNAs within endothelial cells might determine if the barrier function can be sustained in the presence of exogenous miRNAs, and metastasis will be favored in organs in which ceRNAs are absent (**Figure 5**).

In addition, aberrant expression of lncRNAs has been reported in different tumor types, and differential cell-to-cell expression has even been observed in certain tumor types (Lv et al., 2016). Therefore, the presence of sponging lncRNAs might alter (a) the miRNA secretome of tumor cells, and (b) the responsiveness to secreted miRNAs from other cells. The latter case in particular might account for clonal populations of tumor cells with site-specific patterns of metastasis (Obenauf and Massague, 2015).
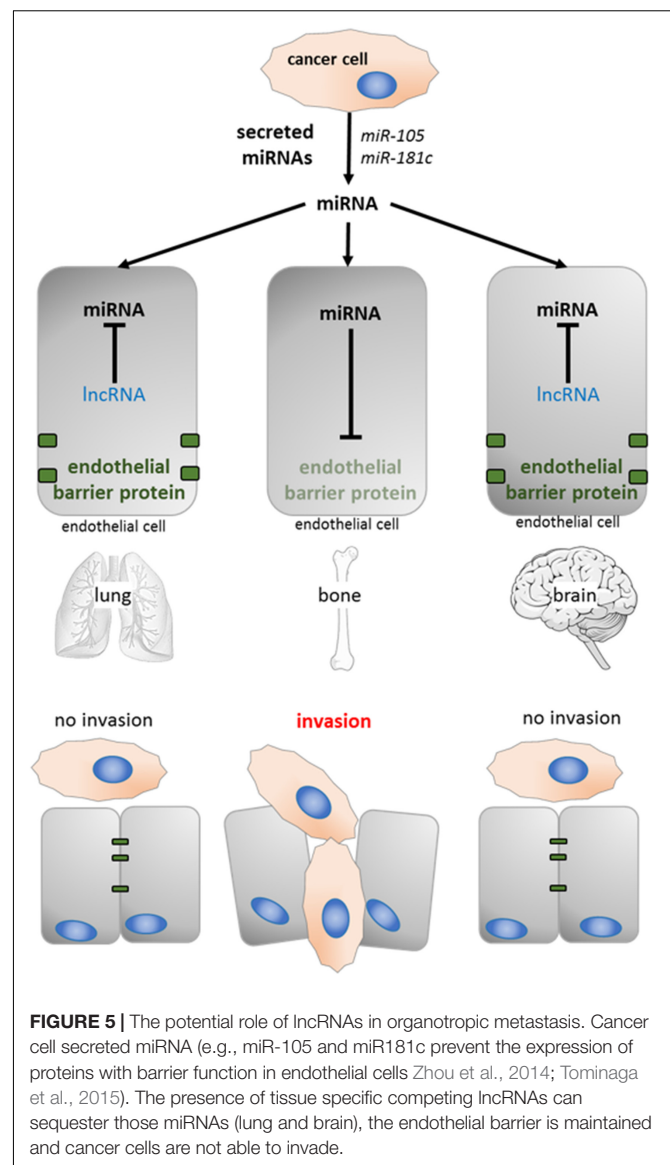
FIGURE 5 | The potential role of lncRNAs in organotropic metastasis. Cancer cell secreted miRNA (e.g., miR-105 and miR181c prevent the expression of proteins with barrier function in endothelial cells Zhou et al., 2014; Tominaga et al., 2015). The presence of tissue specific competing lncRNAs can sequester those miRNAs (lung and brain), the endothelial barrier is maintained and cancer cells are not able to invade.

It is tempting to speculate that the interplay between lncRNAs and miRNAs combined with the tissue- or clone-specific expression of lncRNAs might favor the formation of metastatic niches in an organ- or tissue-specific manner. However, future detailed studies will be required to prove this hypothesis.

FUTURE DIRECTIONS

In our opinion, the continuous accumulation of large amounts of data by RNAseq of whole cancer transcriptomes or extracellular miRNAs might be only partially helpful to move forward. We propose exploring the mechanistic basics in cell-culture-based systems or even model organisms with a reduced complexity. Models derived from these studies could later be used to predict how the extracellular miRNA composition combined

with a distinct set of intracellular lncRNAs might impact on disease progression.

In summary, current limitations in our understanding on the molecular mechanisms of cancer might be due to the fact that, until now, only 2% of the genome has been taken into account. Therefore, future studies should aim at expanding our current view of cancer by including the role of ncRNAs in the interpretation of cancer as a multifactorial disease. The proposed model, combining lncRNA-miRNA interactions with intercellular communication might be particularly helpful in understanding the tissue-specificity of many cancers, hitherto one of the least understood phenomena of cancer.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/pubmed/.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00281/full#supplementary-material

REFERENCES

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4:e05005 doi: 10.7554/eLife.05005

Ajit, S. K. (2012). Circulating microRNAs as biomarkers, therapeutic targets, and signaling molecules. *Sensors* 12, 3359–3369. doi: 10.3390/s120303359

Ali, M. M., Akhade, V. S., Kosalai, S. T., Subhash, S., Statello, L., Meryet-Figuiere, M., et al. (2018). PAN-cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and biomarkers. *Nat. Commun.* 9:883. doi: 10.1038/s41467-018-03265-3261

Ambros, V. (2004). The functions of animal microRNAs. *Nature* 431, 350–355. doi: 10.1038/nature02871

Arroyo, J. D., Chevillet, J. R., Kroh, E. M., Ruf, I. K., Pritchard, C. C., Gibson, D. F., et al. (2011). Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5003–5008. doi: 10.1073/pnas.1019055108

Aune, T. M., Crooke, P. S. 3rd, Patrick, A. E., Tossberg, J. T., Olsen, N. J., Spurlock, C. F. 3rd, et al., (2017). Expression of long non-coding RNAs in autoimmunity and linkage to enhancer function and autoimmune disease risk genetic variants. *J. Autoimmun.* 81, 99–109. doi: 10.1016/j.jaut.2017.03.014

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5

Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science* 306, 2242–2246. doi: 10.1126/science.1103388

Bhan, A., and Mandal, S. S. (2015). LncRNA HOTAIR: a master regulator of chromatin dynamics and cancer. *Biochim. Biophys. Acta* 1856, 151–164. doi: 10.1016/j.bbcan.2015.07.001

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long noncoding RNA and cancer: a new paradigm. *Cancer Res.* 77, 3965–981. doi: 10.1158/0008-5472.CAN-16-2634

Bolha, L., Ravnik-Glavac, M., and Glavac, D. (2017). Long noncoding RNAs as biomarkers in cancer. *Dis. Mark.* 2017:7243968. doi: 10.1155/2017/7243968

Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The product of the H19 gene may function as an RNA. *Mol. Cell Biol.* 10, 28–36. doi: 10.1128/MCB.10.1.28

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., et al. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349, 38–44. doi: 10.1038/349038a0

Bussemakers, M. J., van Bokhoven, A., Verhaegh, G. W., Smit, F. P., Karthaus, H. F., Schalken, J. A., et al. (1999). DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.* 59, 5975–5979.

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611

Callis, T. E., Pandya, K., Seok, H. Y., Tang, R. H., Tatsuguchi, M., Huang, Z. P., et al. (2009). MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J. Clin. Invest.* 119, 2772–2786. doi: 10.1172/JCI36154

Cao, M. X., Jiang, Y. P., Tang, Y. L., and Liang, X. H. (2017). The crosstalk between lncRNA and microRNA in cancer metastasis: orchestrating the epithelial-mesenchymal plasticity. *Oncotarget* 8, 12472–12483. doi: 10.18632/oncotarget.13957

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563. doi: 10.1126/science.1112014

Chan, J. J., and Tay, Y. (2018). Noncoding RNA:RNA regulatory networks in cancer. *Int. J. Mol. Sci.* 19:1310. doi: 10.3390/ijms19051310

Chen, J., Wang, R., Zhang, K., and Chen, L. B. (2014). Long non-coding RNAs in non-small cell lung cancer as biomarkers and therapeutic targets. *J. Cell Mol. Med.* 18, 2425–2436. doi: 10.1111/jcmm.12431

Cheng, D. L., Xiang, Y. Y., Ji, L. J., and Lu, X. J. (2015). Competing endogenous RNA interplay in cancer: mechanism, methodology, and perspectives. *Tumour Biol.* 36, 479–488. doi: 10.1007/s13277-015-3093-z

Cho, W. C. (2007). OncomiRs: the discovery and progress of microRNAs in cancers. *Mol. Cancer* 6:60. doi: 10.1186/1476-4598-6-60

Cipolla, G. A., de Oliveira, J. C., Salviano-Silva, A., Lobo-Alves, S. C., Lemos, D. S., Oliveira, L. C., et al. (2018). Long non-coding RNAs in multifactorial diseases: another layer of complexity. *Noncoding RNA* 4:13. doi: 10.3390/ncrna4020013

Cui, M., You, L., Ren, X., Zhao, W., Liao, Q., and Zhao, Y. (2016). Long non-coding RNA PVT1 and cancer. *Biochem. Biophys. Res. Commun.* 471, 10–14. doi: 10.1016/j.bbrc.2015.12.101

de Giorgio, A., Krell, J., Harding, V., Stebbing, J., and Castellano, L. (2013). Emerging roles of competing endogenous RNAs in cancer: insights from the regulation of PTEN. *Mol. Cell Biol.* 33, 3976–3982. doi: 10.1128/MCB. 00683-613

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

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

Drusco, A., and Croce, C. M. (2017). MicroRNAs and cancer: a long story for short RNAs. *Adv. Cancer Res.* 135, 1–24. doi: 10.1016/bs.acr.2017.06.005

Duttagupta, R., Jiang, R., Gollub, J., Getts, R. C., and Jones, K. W. (2011). Impact of cellular miRNAs on circulating miRNA biomarker signatures. *PLoS One* 6:e20769. doi: 10.1371/journal.pone.0020769

Ebert, M. S., and Sharp, P. A. (2012). Roles for microRNAs in conferring robustness to biological processes. *Cell* 149, 515–524. doi: 10.1016/j.cell.2012.04.005

Ergun, S., and Oztuzcu, S. (2015). Oncocers: ceRNA-mediated cross-talk by sponging miRNAs in oncogenic pathways. *Tumour Biol.* 36, 3129–3136. doi: 10.1007/s13277-015-3346-x

Faghihi, M. A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M. P., Nalls, M. A., et al. (2010). Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol.* 11:R56. doi: 10.1186/gb-2010-11-5-r56

Fok, E. T., Scholefield, J., Fanucchi, S., and Mhlanga, M. M. (2017). The emerging molecular biology toolbox for the study of long noncoding RNA biology. *Epigenomics* 9, 1317–1327. doi: 10.2217/epi-2017-2062

Freedman, M. L., Monteiro, A. N., Gayther, S. A., Coetzee, G. A., Risch, A., Plass, C., et al. (2011). Principles for the post-GWAS functional characterization of cancer risk loci. *Nat. Genet.* 43, 513–518. doi: 10.1038/ng.840

Gidlof, O., Andersson, P., van der Pals, J., Gotberg, M., and Erlinge, D. (2011). Cardiospecific microRNA plasma levels correlate with troponin and cardiac function in patients with ST elevation myocardial infarction, are selectively dependent on renal elimination, and can be detected in urine samples. *Cardiology* 118, 217–226. doi: 10.1159/000328869

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451. doi: 10.1016/j.cell.2013.06.044

Goyal, A., Myacheva, K., Gross, M., Klingenberg, M., Duran Arque, B., and Diederichs, S. (2017). Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. *Nucleic Acids Res.* 45:e12. doi: 10.1093/nar/gkw883

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27, 91–105. doi: 10.1016/j.molcel.2007.06.017

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. doi: 10.1038/nature08975

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227. doi: 10.1038/nature07672

Hammond, S. M. (2015). An overview of microRNAs. *Adv. Drug Deliv. Rev.* 87, 3–14. doi: 10.1016/j.addr.2015.05.001

Han, J., Zhang, J., Chen, L., Shen, B., Zhou, J., Hu, B., et al. (2014). Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biol.* 11, 829–835. doi: 10.4161/rna.29624

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013

Hayes, J., Peruzzi, P.P., and Lawler, S. (2014). MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol. Med.* 20, 460–469. doi: 10.1016/j.molmed.2014.06.005

He, L., and Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 522–531. doi: 10.1038/nrg1379

Hergenreider, E., Heydt, S., Treguer, K., Boettger, T., Horrevoets, A. J., Zeiher, A. M., et al. (2012). Atheroprotective communication between endothelial cells and smooth muscle cells through miRNAs. *Nat. Cell Biol.* 14, 249–256. doi: 10.1038/ncb2441

Ho, T. T., Zhou, N., Huang, J., Koirala, P., Xu, M., Fung, R., et al. (2015). Targeting non-coding RNAs with the CRISPR/Cas9 system in human cell lines. *Nucleic Acids Res.* 43:e17. doi: 10.1093/nar/gku1198

Hon, C. C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J., Gough, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* 543, 199–204. doi: 10.1038/nature21374

Housman, G., and Ulitsky, I. (2016). Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. *Biochim. Biophys. Acta* 1859, 31–40. doi: 10.1016/j.bbagrm.2015.07.017

Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* 21, 1253–1261. doi: 10.1038/nm.3981

Hunter, M. P., Ismail, N., Zhang, X., Aguda, B. D., Lee, E. J., Yu, L., et al. (2008). Detection of microRNA expression in human peripheral blood microvesicles. *PLoS One* 3:e3694. doi: 10.1371/journal.pone.0003694

Husted, S., Sokilde, R., Rask, L., Cirera, S., Busk, P. K., Eriksen, J., et al. (2011). MicroRNA expression profiles associated with development of drug resistance in Ehrlich ascites tumor cells. *Mol. Pharm.* 8, 2055–2062. doi: 10.1021/mp200255d

Jackstadt, R., and Hermeking, H. (2015). MicroRNAs as regulators and mediators of c-MYC function. *Biochim. Biophys. Acta* 1849, 544–553. doi: 10.1016/j.bbagrm.2014.04.003

Jaiswal, R., Gong, J., Sambasivam, S., Combes, V., Mathys, J. M., Davey, R., et al. (2012). Microparticle-associated nucleic acids mediate trait dominance in cancer. *FASEB J.* 26, 420–429. doi: 10.1096/fj.11-186817

Jiang, C., Li, X., Zhao, H., and Liu, H. (2016). Long non-coding RNAs: potential new biomarkers for predicting tumor invasion and metastasis. *Mol. Cancer* 15:62. doi: 10.1186/s12943-016-0545-z

Kagohara, L. T., Stein-O'Brien, G. L., Kelley, D., Flam, E., Wick, H. C., Danilova, L. V., et al. (2018). Epigenetic regulation of gene expression in cancer: techniques, resources and analysis. *Brief. Funct. Genomics* 17, 49–63. doi: 10.1093/bfgp/elx018

Karreth, F. A., and Pandolfi, P. P. (2013). ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov.* 3, 1113–1121. doi: 10.1158/2159-8290.CD-13-0202

Karreth, F. A., Reschke, M., Ruocco, A., Ng, C., Chapuy, B., Leopold, V., et al. (2015). The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell* 161, 319–332. doi: 10.1016/j.cell.2015.02.043

Karreth, F.A., Tay, Y., Perna, D., Ala, U., Tan, S.M., Rust, A.G., et al. (2011). In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* 147, 382–395. doi: 10.1016/j.cell.2011.09.032

Kong, Y. W., Ferland-McCollough, D., Jackson, T. J., and Bushell, M. (2012). microRNAs in cancer management. *Lancet Oncol.* 13, e249–e258. doi: 10.1016/S1470-2045(12)70073-70076

Kopp, F., and Mendell, J. T. (2018). Functional classification and experimental dissection of long noncoding RNAs. *Cell* 172, 393–407. doi: 10.1016/j.cell.2018.01.011

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2018). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47:D155–D162 doi: 10.1093/nar/gky1141

Lagarde, J., Uszczynska-Ratajczak, B., Carbonell, S., Perez-Lluch, S., Abad, A., Davis, C., et al. (2017). High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat. Genet.* 49, 1731–1740. doi: 10.1038/ng.3988

Lal, A., Thomas, M. P., Altschuler, G., Navarro, F., O'Day, E., Li, X. L., et al. (2011). Capture of microRNA-bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. *PLoS Genet.* 7:e1002363. doi: 10.1371/journal.pgen.1002363

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

Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature* 396, 643–649. doi: 10.1038/25292

Leung, A. K. L. (2015). The whereabouts of microRNA Actions: cytoplasm and beyond. *Trends Cell Biol.* 25, 601–610. doi: 10.1016/j.tcb.2015.07.005

Li, Y., Egranov, S. D., Yang, L., and Lin, C. (2018). Molecular mechanisms of long noncoding RNAs-medicated cancer metastasis. *Genes Chromosomes Cancer* 58, 200–207 doi: 10.1002/gcc.22691

Ling, H., Spizzo, R., Atlasi, Y., Nicoloso, M., Shimizu, M., Redis, R. S., et al. (2013). CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res.* 23, 1446–1461. doi: 10.1101/gr.152942.112

Liz, J., and Esteller, M. (2016). lncRNAS and microRNAs with a role in cancer development. *Biochim. Biophys. Acta* 1859, 169–176. doi: 10.1016/j.bbagrm.2015.06.015

Lv, D., Wang, X., Dong, J., Zhuang, Y., Huang, S., Ma, B., et al. (2016). Systematic characterization of lncRNAs' cell-to-cell expression heterogeneity in glioblastoma cells. *Oncotarget* 7, 18403–18414. doi: 10.18632/oncotarget.7580

Masliah-Planchon, J., Garinet, S., and Pasmant, E. (2016). RAS-MAPK pathway epigenetic activation in cancer: miRNAs in action. *Oncotarget* 7, 38892–38907. doi: 10.18632/oncotarget.6476

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10513–10518. doi: 10.1073/pnas.0804549105

Nakagawa, T., Endo, H., Yokoyama, M., Abe, J., Tamai, K., Tanaka, N., et al. (2013). Large noncoding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. *Biochem. Biophys. Res. Commun.* 436, 319–324. doi: 10.1016/j.bbrc.2013.05.101

Nguyen, Q., and Carninci, P. (2016). Expression specificity of disease-associated lncRNAs: toward personalized medicine. *Curr. Top. Microbiol. Immunol.* 394, 237–258. doi: 10.1007/82_2015_464

Obenauf, A. C., and Massague, J. (2015). Surviving at a distance: organ-specific metastasis. *Trends Cancer* 1, 76–91. doi: 10.1016/j.trecan.2015.07.009

Park, N. J., Zhou, H., Elashoff, D., Henson, B. S., Kastratovic, D. A., Abemayor, E., et al. (2009). Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. *Clin. Cancer Res.* 15, 5473–5477. doi: 10.1158/1078-0432.CCR-09-0736

Patel, R. S., Jakymiw, A., Yao, B., Pauley, B. A., Carcamo, W. C., Katz, J., et al. (2011). High resolution of microRNA signatures in human whole saliva. *Arch. Oral Biol.* 56, 1506–1513. doi: 10.1016/j.archoralbio.2011.05.015

Ponjavic, J., Ponting, C. P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* 17, 556–565. doi: 10.1101/gr.6036807

Ponting, C. P., Oliver, P. L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell* 136, 629–641. doi: 10.1016/j.cell.2009.02.006

Prahallad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., et al. (2012). Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* 483, 100–103. doi: 10.1038/nature10868

Prensner, J. R., and Chinnaiyan, A. M. (2011). The emergence of lncRNAs in cancer biology. *Cancer Discov.* 1, 391–407. doi: 10.1158/2159-8290.CD-11-0209

Quinn, J. J., and Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* 17, 47–62. doi: 10.1038/nrg.2015.10

Rayner, K. J., and Hennessy, E. J. (2013). Extracellular communication via microRNA: lipid particles have a new message. *J. Lipid Res.* 54, 1174–1181. doi: 10.1194/jlr.R034991

Rechavi, O., Erlich, Y., Amram, H., Flomenblit, L., Karginov, F. V., Goldstein, I., et al. (2009). Cell contact-dependent acquisition of cellular and viral nonautonomously encoded small RNAs. *Genes Dev.* 23, 1971–1979. doi: 10.1101/gad.1789609

Saliminejad, K., Khorram Khorshid, H. R., Soleymani Fard, S., and Ghaffari, S. H. (2018). An overview of microRNAs: biology, functions, therapeutics, and analysis methods. *J. Cell Physiol.* 234, 5451–5465 doi: 10.1002/jcp.27486

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014

Schneider, G., Schmidt-Supprian, M., Rad, R., and Saur, D. (2017). Tissue-specific tumorigenesis: context matters. *Nat. Rev. Cancer* 17, 239–253. doi: 10.1038/nrc.2017.5

Sclafani, F., Gullo, G., Sheahan, K., and Crown, J. (2013). BRAF mutations in melanoma and colorectal cancer: a single oncogenic mutation with different tumour phenotypes and clinical implications. *Crit. Rev. Oncol. Hematol.* 87, 55–68. doi: 10.1016/j.critrevonc.2012.11.003

Shefler, I., Salamon, P., Reshef, T., Mor, A., and Mekori, Y. A. (2010). T cell-induced mast cell activation: a role for microparticles released from activated T cells. *J. Immunol.* 185, 4206–4212. doi: 10.4049/jimmunol.1000409

Shivdasani, R. A. (2006). MicroRNAs: regulators of gene expression and cell differentiation. *Blood* 108 3646–3653. doi: 10.1182/blood-2006-01-030015

Skog, J., Wurdinger, T., van Rijn, S., Meijer, D. H., Gainche, L., Sena-Esteves, M., et al. (2008). Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol.* 10, 1470–1476. doi: 10.1038/ncb1800

Song, Y. X., Sun, J. X., Zhao, J. H., Yang, Y. C., Shi, J. X., Wu, Z. H., et al. (2017). Non-coding RNAs participate in the regulatory network of CLDN4 via ceRNA mediated miRNA evasion. *Nat. Commun.* 8:289. doi: 10.1038/s41467-017-00304-301

Su, X., Xing, J., Wang, Z., Chen, L., Cui, M., and Jiang, B. (2013). microRNAs and ceRNAs: RNA networks in pathogenesis of cancer. *Chin. J. Cancer Res.* 25, 235–239. doi: 10.3978/j.issn.1000-9604.2013.03.08

Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S. M., Ala, U., et al. (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 147, 344–357. doi: 10.1016/j.cell.2011.09.029

Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505, 344–352. doi: 10.1038/nature12986

Teschendorff, A. E., Lee, S. H., Jones, A., Fiegl, H., Kalwa, M., Wagner, W., et al. (2015). HOTAIR and its surrogate DNA methylation signature indicate carboplatin resistance in ovarian cancer. *Genome Med.* 7:108. doi: 10.1186/s13073-015-0233-234.

Thomson, D. W., and Dinger, M. E. (2016). Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet.* 17, 272–283. doi: 10.1038/nrg.2016.20

Tkach, M., and Thery, C. (2016). Communication by extracellular vesicles: where we are and where we need to go. *Cell* 164, 1226–1232. doi: 10.1016/j.cell.2016.01.043

Tominaga, N., Kosaka, N., Ono, M., Katsuda, T., Yoshioka, Y., Tamura, K., et al. (2015). Brain metastatic cancer cells release microRNA-181c-containing extracellular vesicles capable of destructing blood-brain barrier. *Nat. Commun.* 6:6716. doi: 10.1038/ncomms7716

Turchinovich, A., Weiz, L., Langheinz, A., and Burwinkel, B. (2011). Characterization of extracellular circulating microRNA. *Nucleic Acids Res.* 39, 7223–7233. doi: 10.1093/nar/gkr254

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* 17, 601–614. doi: 10.1038/nrg.2016.85

Ulitsky, I. (2018). Interactions between short and long noncoding RNAs. *FEBS Lett.* 592, 2874–2883. doi: 10.1002/1873-3468.13085

Valadi, H., Ekstrom, K., Bossios, A., Sjostrand, M., Lee, J. J., and Lotvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* 9, 654–659. doi: 10.1038/ncb1596.

Viereck, J., and Thum, T. (2017). Circulating noncoding RNAs as biomarkers of cardiovascular disease and injury. *Circ. Res.* 120, 381–399. doi: 10.1161/CIRCRESAHA.116.308434

Walsh, A. L., Tuzova, A. V., Bolton, E. M., Lynch, T. H., and Perry, A. S. (2014). Long noncoding RNAs and prostate carcinogenesis: the missing 'linc'? *Trends Mol. Med.* 20, 428–436. doi: 10.1016/j.molmed.2014.03.005

Wan, P., Su, W., and Zhuo, Y. (2017). The role of long noncoding RNAs in neurodegenerative diseases. *Mol. Neurobiol.* 54, 2012–2021. doi: 10.1007/s12035-016-9793-9796

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Wang, Y., Hou, J., He, D., Sun, M., Zhang, P., Yu, Y., et al. (2016). The emerging function and mechanism of ceRNAs in cancer. *Trends Genet.* 32, 211–224. doi: 10.1016/j.tig.2016.02.001

Weber, D. G., Casjens, S., Rozynek, P., Lehnert, M., Zilch-Schoneweis, S., Bryk, O., et al. (2010). Assessment of mRNA and microRNA stabilization in peripheral human blood for multicenter studies and biobanks. *Biomark. Insights* 5, 95–102. doi: 10.4137/BMI.S5522

Wei, J. T., Feng, Z., Partin, A. W., Brown, E., Thompson, I., Sokoll, L., et al. (2014). Can urinary PCA3 supplement PSA in the early detection of prostate cancer? *J. Clin. Oncol.* 32, 4066–4072. doi: 10.1200/JCO.2013.52.8505

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

Xie, P., Cao, H., Li, Y., Wang, J., and Cui, Z. (2017). Knockdown of lncRNA CCAT2 inhibits endometrial cancer cells growth and metastasis via sponging miR-216b. *Cancer Biomark.* 21, 123–133. doi: 10.3233/CBM-170388

Yang, C., Wu, D., Gao, L., Liu, X., Jin, Y., Wang, D., et al. (2016). Competing endogenous RNA networks in human cancer: hypothesis, validation, and perspectives. *Oncotarget* 7, 13479–13490. doi: 10.18632/oncotarget.7266

Yarmishyn, A. A., and Kurochkin, I. V. (2015). Long noncoding RNAs: a potential novel class of cancer biomarkers. *Front. Genet.* 6:145. doi: 10.3389/fgene.2015.00145

Yoon, J. H., Abdelmohsen, K., and Gorospe, M. (2014). Functional interactions among microRNAs and long noncoding RNAs. *Semin. Cell Dev. Biol.* 34, 9–14. doi: 10.1016/j.semcdb.2014.05.015

Zamore, P. D., and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. *Science* 309, 1519–1524. doi: 10.1126/science.1111444

Zernecke, A., Bidzhekov, K., Noels, H., Shagdarsuren, E., Gan, L., Denecke, B., et al. (2009). Delivery of microRNA-126 by apoptotic bodies induces CXCL12-dependent vascular protection. *Sci. Signal.* 2:ra81. doi: 10.1126/scisignal.2000610

Zhang, B., Pan, X., Cobb, G. P., and Anderson, T. A. (2007). microRNAs as oncogenes and tumor suppressors. *Dev. Biol.* 302, 1–12. doi: 10.1016/j.ydbio.2006.08.028

Zhang, R., Xia, L. Q., Lu, W. W., Zhang, J., and Zhu, J. S. (2016). LncRNAs and cancer. *Oncol. Lett.* 12, 1233–1239. doi: 10.3892/ol.2016.4770

Zhang, Z. Z., Shen, Z. Y., Shen, Y. Y., Zhao, E. H., Wang, M., Wang, C. J., et al. (2015). HOTAIR long noncoding RNA promotes gastric cancer metastasis through suppression of poly r(C)-binding protein (PCBP) 1. *Mol. Cancer Ther.* 14, 1162–1170. doi: 10.1158/1535-7163.MCT-14-0695

Zhong, Y., Du, Y., Yang, X., Mo, Y., Fan, C., Xiong, F., et al. (2018). Circular RNAs function as ceRNAs to regulate and control human cancer progression. *Mol. Cancer* 17:79. doi: 10.1186/s12943-018-0827-828

Zhou, W., Fong, M. Y., Min, Y., Somlo, G., Liu, L., Palomares, M. R., et al. (2014). Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis. *Cancer Cell* 25, 501–515. doi: 10.1016/j.ccr.2014.03.007

Zhu, H. Y., Bai, W. D., Ye, X. M., Yang, A. G., and Jia, L. T. (2018). Long non-coding RNA UCA1 desensitizes breast cancer cells to trastuzumab by impeding miR-18a repression of Yes-associated protein 1. *Biochem. Biophys. Res. Commun.* 496, 1308–1313. doi: 10.1016/j.bbrc.2018.02.006

Bovine Embryo-Secreted microRNA-30c Is a Potential Non-invasive Biomarker for Hampered Preimplantation Developmental Competence

Xiaoyuan Lin[1], Evy Beckers[1], Séan Mc Cafferty[1], Yannick Gansemans[2], Katarzyna Joanna Szymańska[3], Krishna Chaitanya Pavani[4], João Portela Catani[1], Filip Van Nieuwerburgh[2], Dieter Deforce[2], Petra De Sutter[5], Ann Van Soom[4] and Luc Peelman[1]*

[1] Department of Nutrition, Genetics and Ethology, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium, [2] Department of Pharmaceutics, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium, [3] Physiology Group, Department of Basic Medical Sciences, Ghent University, Ghent, Belgium, [4] Reproduction, Obstetrics and Herd Health, Ghent University, Merelbeke, Belgium, [5] Department of Uro-Gynaecology, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

Recently, secreted microRNAs (miRNAs) have received a lot of attention since they may act as autocrine factors. However, how secreted miRNAs influence embryonic development is still poorly understood. We identified 294 miRNAs, 114 known, and 180 novel, in the conditioned medium of individually cultured bovine embryos. Of these miRNAs, miR-30c and miR-10b were much more abundant in conditioned medium of slow cleaving embryos compared to intermediate cleaving ones. MiR-10b, miR-novel-44, and miR-novel-45 were higher expressed in the conditioned medium of degenerate embryos compared to blastocysts, while the reverse was observed for miR-novel-113 and miR-novel-139. Supplementation of miR-30c mimics into the culture medium confirmed the uptake of miR-30c mimics by embryos and resulted in increased cell apoptosis, as also shown after delivery of miR-30c mimics in Madin-Darby bovine kidney cells (MDBKs). We also demonstrated that miR-30c directly targets Cyclin-dependent kinase 12 (*CDK12*) through its 3′ untranslated region (3′-UTR) and inhibits its expression. Overexpression and downregulation of *CDK12* revealed the opposite results of the delivery of miRNA-30c mimics and inhibitor. The significant down-regulation of several tested DNA damage response (DDR) genes, after increasing miR-30c or reducing *CDK12* expression, suggests a possible role for miR-30c in regulating embryo development through DDR pathways.

Keywords: bovine embryos, secreted miRNAs, miR-30c, *CDK12*, cell cycle, DNA damage response, individual *in vitro* production

INTRODUCTION

Many studies have indicated that the timing of cell division during the early embryonic stages is crucial for normal development and can be used as an indicator of embryo development competence (Fenwick et al., 2002; Zernicka-Goetz, 2002, 2006; Plusa et al., 2005; Hiiragi et al., 2006; Terriou et al., 2007). For example, the delay of cell division might be a consequence of chromosomal aberrations and DNA damage (Milewski et al., 2018) and slow cleaving embryos have a higher caspase activity in comparison to fast cleavers (Vandaele et al., 2007). In general, faster cleaving embryos have a significantly higher probability of reaching advanced developmental stages compared to slower cleaving embryos (Van Soom, 1997; Meirelles et al., 2004; Vandaele et al., 2006), while some studies also demonstrated that cleaving divisions that are too fast or too slow are indicative of poor embryo quality (Market Velker et al., 2012; Gutierrez-Adan et al., 2015). During these early embryonic stages, miRNA levels undergo dynamic changes (Mineno et al., 2006; Tang et al., 2007; Yang et al., 2008; Viswanathan et al., 2009; Goossens et al., 2013), indicating their potential role in embryonic development.

With this study we wanted to investigate if one or more miRNAs have potential as a non-invasive biomarker for preimplantation developmental competence according to cleavage patterns and blastocyst formation. It has been reported that miRNAs are not only localized intracellularly but also secreted via exosomes (Valadi et al., 2007). In addition, miRNAs have been reported to be transferable to other cells, and can be functional in the new location (Valadi et al., 2007; Sohel et al., 2013; Vilella et al., 2015). More specifically, they can be taken up into cells from the extracellular environment, leading to a corresponding endogenous miRNA increase in transfected cells. Recently, secreted miRNA expression was reported to correlate with developmental competence and sexual dimorphism in bovine (Kropp and Khatib, 2015; Gross et al., 2017) and human embryos (Rosenbluth et al., 2014). Although the precise mechanisms of miRNA release in the cellular environment are poorly understood, their selective secretion and high stability (resistant to RNase digestion and other harsh conditions) make miRNAs good candidates for use as biomarkers (Luo et al., 2009; Donker et al., 2012). Potential limitations for their use as biomarkers are their general low abundance, and the high sequence identity among family members.

miR-30 family members are involved in the regulation of p53-induced mitochondrial fission and cell apoptosis (Li et al., 2010). As a member of the miR-30 family, miR-30c has been shown to regulate the cell cycle and proliferation in human and mouse (Li et al., 2012; Quintavalle et al., 2013; Shukla et al., 2015; Liu et al., 2016). One of its potential targets as determined by our study is *CDK12* mRNA. *CDK12* is a protein kinase responsible for mature mRNA synthesis transcriptional elongation (Bartkowiak et al., 2010; Liang et al., 2015). This kinase has been reported to be crucial for the development of the inner cell mass in mouse embryos (Juan et al., 2016) and to maintain genomic stability as Cyclin K/*CDK12* complex (Blazek et al., 2011) through regulating DDR genes. In this study, we demonstrated that miR-30c is secreted and taken up by bovine embryos and functions as a negative regulator of cell growth by targeting *CDK12*, indicating that miR-30c can be considered as a promising biomarker for bovine early embryonic development. These findings may provide new insights into understanding the regulatory role of secreted miRNAs in the process of intercellular communication.

MATERIALS AND METHODS

In vitro Embryo Production and CM Collection

All animal handlings were approved by the Ethical Committee of the Faculty of Veterinary Medicine (EC2013/118) of Ghent University. All methods were performed in accordance with the relevant guidelines and regulations. Bovine blastocysts were produced according to the previously used routine *in vitro* fertilization (IVF) methods in our lab (Wydooghe et al., 2014a). Briefly, ovaries were collected from the local slaughterhouse and processed within 2 h. The collected ovaries were washed three times in warm physiological saline supplemented with 5 mg/ml kanamycin (GIBCO-BRL Life Technologies, Merelbeke, Belgium). Subsequently, cumulus oocytes complexes were aspirated from 4 to 8 mm diameter follicles and cultured in groups of 60 in 500 µl maturation medium-containing TCM199 (Life Technologies, Ghent, Belgium) supplemented with 20% heat-inactivated fetal bovine serum (FBS) (Biochrom AG, Berlin, Germany) for 22 h at 38.5°C in 5% CO_2 in the air. Frozen-thawed bovine spermatozoa from Holstein bulls were separated through a 45% and 90% Percoll gradient (GE healthcare Biosciences, Uppsala, Sweden). The final sperm concentration of 1×10^6 spermatozoa/ml was adjusted in IVF-Tyrode's albumin-lactate-pyruvate (IVF-TALP), consisting of bicarbonate-buffered Tyrode solution supplemented with 6 mg/ml and 25 µg/ml heparin bovine serum albumin (BSA) (Sigma, Schnelldorf, Germany). Matured oocytes were washed in 500 µl IVF-TALP medium and were incubated with spermatozoa. After incubation for 21 h, presumed zygotes were vortexed for 3 min to remove spermatozoa and cumulus cells, washed with IVF-TALP and transferred to 20 µl drops of synthetic oviductal fluid supplemented with ITS (5 µg/ml Insulin + 5 µg/ml Transferrin + 5 ng/ml Selenium) and 4 mg/ml BSA. Culture occurred individually in drops of 20 µl, covered with mineral oil at 38.5°C in 5% CO_2, 5% O_2 and 90% N_2.

Bovine embryos were divided into groups according to the first cleavage patterns, as described previously (Dinnyes, 1999; Amarnath et al., 2007; Sugimura et al., 2017). Time points of first cleavage [24.2–33.8 h post insemination (hpi)] were listed up and were divided in quartiles. The first quartile was considered as "fast," the second and third quartiles were considered as "intermediate" and the last quartile was considered as "slow." More specifically, individual droplets were viewed microscopically at two time points (26.6 and 31.4 hpi), and three groups were produced according to the embryos' cleavage pattern: "fast" (cleavage occurred before 26.6 hpi), "intermediate" (cleavage occurred between 26.6 and 31.4 hpi) and "slow" (cleavage had not occurred yet at 31.4 hpi). Additionally, the

developmental competence of each embryo was microscopically viewed and assessed at 8 days post insemination (dpi), enabling a division into two subgroups (degenerate embryos and blastocysts). Eventually, the embryos were divided into six groups: FB (fast cleaving blastocyst), IB (intermediate cleaving blastocyst), SB (slow cleaving blastocyst), FD (fast cleaving degenerate), ID (intermediate cleaving degenerate), and SD (slow cleaving degenerate). Conditioned medium of single embryos was collected (17.5 μl each droplet) and pooled for each of the six groups.

miRNA Extraction

At 8 dpi, the CM was collected and miRNA was extracted with the miRNeasy Serum/Plasma kit (Qiagen, Germantown, United States). To meet the miRNA-sequencing minimum concentration requirement, RNA was extracted from CM (three replicates of 3 mL each) and was concentrated with the RNeasy MinElute Cleanup kit (Qiagen, Germantown, United States). Finally, the quality and concentration of the RNA samples were examined using an RNA 6000 Pico Chip (Agilent Technologies, Carlsbad, CA, United States) and a Quant-iT RiboGreen RNA Assay kit (Life Technologies, Carlsbad, CA, United States), respectively. The total RNA isolated from CM ranged from 1.982 to 2.448 ng/μl. The FB and FD group were excluded because the required amount of secreted miRNAs from the IVF culture system for sequencing was not obtained.

Small RNA Library Construction and Deep Sequencing

Small RNA library construction was performed with the Tailormix v2 kit (SeqMatic, Fremont, CA, United States). The quality-ensured RNA-seq libraries were pooled and sequencing was performed in triplicate on the Illumina Miseq (NxtGnt sequencing facility, Gent, Belgium).

Small RNA-Sequencing Data Analysis and Differential Expression Analysis

Identification of known miRNAs, prediction of putative novel miRNAs and reading counting were done using the mirPRo pipeline (Shi et al., 2015). MicroRNA data from the miRBase (v21) (Griffiths-Jones et al., 2006) and the annotated cow genome (GCA_000003055.3) were used as reference. Differential expression between sample groups was statistically tested in R (Ihaka and Gentleman, 1996) with both EdgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2014) via the SARTools wrapper (Varet et al., 2016). Two comparisons were made after RNA-Sequencing: IB vs. SB; (I + S) Degenerate vs. (I + S) Blastocysts. The results were considered statistically significant when the Benjamini-Hochberg corrected p-value was <0.05.

Pathway Analysis

The functional analysis of the differentially expressed genes between the groups was performed using DAVID (Huang et al., 2008, 2009) (predicted target genes as input) and miRWalk (Dweep and Gretz, 2015) (miR-30c and miR-10b as input) in terms of enrichment of gene ontologies (GO). In addition,

a pathway analysis was performed using the KEGG database to identify the significant pathways affected by the differentially expressed miRNAs. The Benjamini-Hochberg corrected p-values <0.05 were considered statistically significant.

RT-qPCR

To verify the results of the miRNA sequencing, five mature miRNAs were quantified using RT-qPCR (real-time quantitative PCR). Accordingly, total RNA samples (including miRNAs) isolated from CM (three additional biological replicates of 200 μl each) were reverse transcribed using a miScript II RT kit (Qiagen, Germantown, MD, United States) and subsequently quantified with a miScript SYBR Green Kit containing 10 × miScript Universal Primer (Qiagen, Germantown, MD, United States). U6 (Mondou et al., 2012; Abd El Naby et al., 2013) was quantified to normalize miRNA expression levels.

To check the intracellular expression of the differentially released miRNAs and if miR-30c is taken up by embryos, miRNAs were quantified using RT-qPCR. Total RNA samples (including miRNAs) isolated from embryos (three replicates of approximate 5 embryos each) using the miRNeasy Mini kit (Qiagen, Germantown, United States) and reverse transcribed using a miScript II RT and subsequently quantified with a miScript SYBR Green Kit containing 10 × miScript Universal. U6 was quantified to normalize miRNA expression levels.

Additionally, embryos and MDBKs were used to analyze mRNA abundance of CDK12 and DDR genes. Total RNA samples were isolated from embryos (three replicates of approximate 5 embryos each) and MDBKs using the RNeasy Micro kit (Qiagen, Germantown, MD, United States) and reverse transcribed using the iScript cDNA synthesis kit (BioRad, Brussels, Belgium). The mRNA levels were quantified with a SsoAdvanced Universal SYBR Green Supermix kit (BioRad, Brussels, Belgium). GAPDH (Herrmann et al., 2013; Li et al., 2016), which has been proved to be a stable reference gene in our sample (data not shown), was quantified to normalize mRNA expression levels.

All reactions were performed in triplicate, and the $2^{-\Delta\Delta Ct}$ method was used to analyze the data. The primer sequences used for RT-qPCR are listed in **Supplementary Table S1**.

miR-30c Mimics Supplementation to Embryos Culture Medium

Since individually cultured embryos have less tolerance when compared to group cultured embryos (Goovaerts et al., 2009; Wydooghe et al., 2014a,b) and they easily die after changing the culture environment, group culture was performed for miR-30c functional analysis instead of individual culture. The IVF embryos were produced according to the previously described protocol. This time, however, presumed zygotes were vortexed for 3 min after 21 h incubation, washed with IVF-TALP and transferred to drops of SOF supplemented with ITS, BSA and miR-30c mimics (chemically synthesized, double-stranded RNAs which mimic mature endogenous miRNAs after delivery to cells) or control mimics (chemically synthesized, double-stranded RNAs which have no homology to any known microRNA or mRNA sequences) (Qiagen, Germantown,

United States) with a final concentration of 1 µM according to the instructions. Culture occurred in groups of 25 in drops of 50 µl, covered with mineral oil at 38.5°C in 5% CO_2, 5% O_2, and 90% N_2. On 8 dpi, blastocyst rates were calculated. Blastocysts were collected for RT-qPCR or assessed with apoptosis staining.

TUNEL Staining and Differential Apoptotic Staining

TUNEL staining was performed using a previously described protocol (Ortiz-Escribano et al., 2017) with an *in situ* cell death detection kit (Sigma, St. Louis, MO, United States). Briefly, ∼20 blastocysts for each group were collected and fixed in 4% paraformaldehyde at room temperature (RT) for 1 h, and then permeabilized in 0.1% Triton X-100 at RT for 10 min. Afterward, blastocysts were stained with 20 µl TUNEL mixture for 1 h at 37C and subsequently stained with 10 µg/ml DAPI for 10 min. The embryos were mounted on the slides and were examined using a 20× water immersion objective on a Leica TCS-SP8 X confocal microscope (Leica microsystems, Wetzlar, Germany). The apoptosis ratio was expressed as the total number of TUNEL-positive cells relative to the total number of the cells per blastocyst.

Differential apoptotic staining was performed using previously described protocols (Wydooghe et al., 2011; Lu et al., 2019). The first day, ∼20 blastocysts for each group were fixed in 4% paraformaldehyde for 1 h and put in a 4-well dish in permeabilization solution (0.5% Triton X-100 + 0.05% Tween) in phosphate buffered saline (PBS) at RT for 1 h. After washing the blastocysts 3 times during 2 min in PBS-BSA, they were incubated in 2N HCl at RT for 20 min and then in 100 mM Tris–HCl at RT for 10 min. The blastocysts were washed (3 times during 2 min) and then put into 500 µl of blocking solution at 4C overnight. The second day, the blastocysts were washed again (3 times during 2 min) and incubated in primary CDX-2 antibody (Biogenex, San Ramon, United States) at 4C overnight. On the third day, the blastocysts were washed twice for 15 min and subsequently incubated in blocking solution containing the rabbit active caspase-3 antibody (Cell Signaling Technology, Leiden, Netherlands) overnight at 4C. On day four, the blastocysts were incubated in blocking solution containing the goat anti-mouse Texas Red antibody at RT for 1 h and were subsequently incubated in blocking solution containing the goat anti-rabbit FITC antibody at RT for 1 h. The blastocysts were washed twice for 15 min and incubated at RT for 20 min in a dilution 1: 200 Hoechst in PBS-BSA in the dark. All slides were examined using a 63 × water immersion objective on a Leica TCS-SP8 X confocal microscope. The apoptosis ratio was expressed as the total number of Caspase-3-positive cells relative to the total number of the cells per blastocyst.

Plasmid Construction

The full-length coding sequence of *CDK12* (4473 bp) (NM_001205701.1) was amplified from MDBK cDNA and was inserted into a pEGFP-N1 vector via NheI and XhoI sites for construction of the *CDK12*-overexpressing vector. The empty vector (mock) was used as a negative control. The *CDK12* 3′-UTR (282 bp) containing the predicted miR-30c binding site was amplified from bovine genomic DNA and inserted into a psi-CHECK2 vector (Promega, Madison, United States) via NotI and XhoI sites and confirmed by sequencing. To test whether the predicted miR-30c target site in the *CDK12* 3′-UTR is critical for the miR-30c-mediated repression of *CDK12* expression, the seed sequence of the predicted miR-30c's binding site was changed (Wu et al., 2017; **Figure 4A**). Primers for vector construction are listed in **Supplementary Table S1**.

Dual-Luciferase Reporter Assay

The miR-30c mimics/control mimics and luciferase reporter plasmids were co-transfected into HEK293T cells using Lipofectamine 2000 (Invitrogen, Carlsbad, United States). After 24 h of transfection, the Renilla and Firefly luciferase were assayed using the Dual Luciferase Reporter Kit (Promega, Madison, WI, United States).

Cell Culture and Transfection

The HEK293T cells and MDBK cells were cultured at 37C in 5% CO_2 in DMEM media (Thermo Fisher Scientific, Waltham, MA, United States) supplemented with 10% FBS (VWR, Radnor, United States), 100 U/ml penicillin and 100 mg/ml streptomycin. miR-30c mimics/inhibitor and their negative controls were delivered into MDBK cells using Hiperfect reagent (Qiagen, Germantown, MD, United States) following the manufacturer's instructions. The short-interfering RNA (siRNA) targeting *CDK12* and a non-target control siRNA (si-NTC) were purchased from Qiagen (Germantown, MD, United States). SiRNA or the overexpressing vector was transfected into MDBK cells using Lipofectamine 2000 according to the manufacturer's instructions. Protein or total RNA were extracted for western blotting (WB) or RT-qPCR 48 or 24 h after transfection.

Western Blotting

Cells were collected after 48 h of transfection and lysed using Radioimmunoprecipitation lysis buffer consisting of 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate and protease inhibitors. The samples were denatured at 100C for 10 min before loading onto 10% SDS-polyacrylamide gels. Separated proteins were then transferred onto nitrocellulose membranes and blocked with 5% non-fat milk in PBS with 0.1% Tween-20 for overnight. Membranes were then incubated overnight with 1/1000 rabbit anti-*CDK12* (Novus Biologicals, Abingdon, United Kingdom) and 1/1000 rabbit anti-β-actin (Novus Biologicals, Abingdon, United Kingdom). After three washes, the membranes were incubated with HRP-conjugated goat anti-rabbit IgG (H + L) (Novus Biologicals, Abingdon, United Kingdom) for 2 h at room temperature. Signals were revealed by autograph using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific, Waltham, United States).

Cell Cycle Assays: PI Staining and Flow Cytometry

Madin-Darby bovine kidney cells were cultured in 6-well plates for 48 h after transfection and were stained with propidium iodide (PI) at a final concentration of 50 μg/ml PI and 100 μg/ml RNase A in PBS. Then, the cells were analyzed using Accuri™ C6 flow cytometry (BD, Erembodegem, Belgium) collecting 50000 events. All experiments were replicated three times.

Cell Proliferation Assays: WST-1 Colorimetric Assay

WST-1(4-(3-(4-iodophenyl)-2-(4-nitrophenyl)-2H-5-tetrazolio)-1,3-benzene disulfonate) (Merck, Kenilworth, United States) was used for cell proliferation analysis. The assay was performed using 96-well plates with ∼20000 cells. After 48 h of transfection, 10 μl of WST-1 was added to 90 μl samples. The samples were measured at 450 nm wavelength (570 nm as a reference wavelength) using an EZ read 400 microplate reader (Biochrom, Holliston, United States). Cell viability was then calculated by comparing the absorbance values of sample groups after background subtraction. All experiments were replicated three times.

Statistical Analysis

The data are presented as mean ± S.D and derived from at least three independent experiments. The statistical analyses were performed using ANOVA followed by Tukey's test or Student's t test. For each analysis, $P < 0.05$ was considered significant.

RESULTS

Intermediate Cleaving Embryos Result in a Higher Blastocyst Rate Compared to Slow Cleaving Embryos

According to the timing of the first cell division, 1808 individually cultured embryos for each of three replicate were labeled as either fast, intermediate or slow cleaving and evaluated at 8 dpi for developmental competence. Intermediate embryos produced significantly ($P = 0.027$) more blastocysts in comparison to the slow embryos (41.16 and 18.7%, respectively; **Figure 1**). No statistically distinctive differences ($P = 0.24$) were found between fast and intermediate cleaving embryos (50.65 and 41.16%, respectively; **Figure 1**). The fast group was excluded for sequencing because not enough RNA was obtained due to the low number of embryos belonging to this group.

miRNAs Secreted by Bovine Embryos

In total 294 miRNAs were found in conditioned media (CM) after sequencing (MicroRNAs sequencing data are available in the GEO database under the accession number PRJNA492220): 114 known miRNAs and 180 potential novel miRNAs. The uncorrected p-value was indicative of differential secretion from embryos with different cleavage patterns and different development competences for the following miRNAs: miR-30c
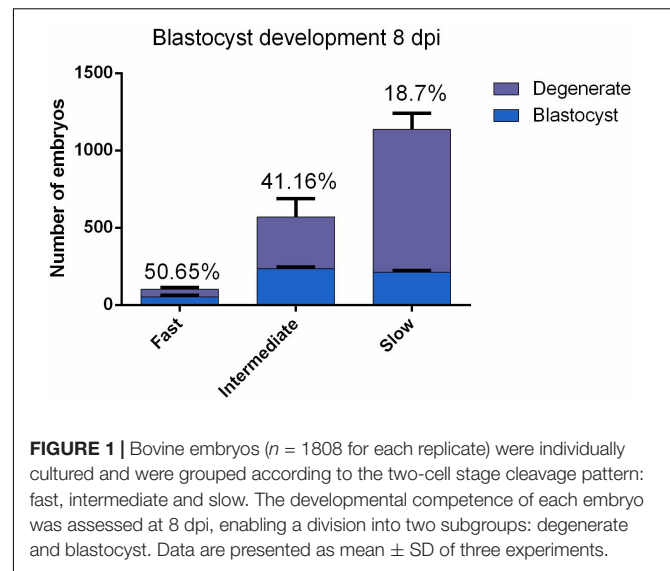
FIGURE 1 | Bovine embryos (n = 1808 for each replicate) were individually cultured and were grouped according to the two-cell stage cleavage pattern: fast, intermediate and slow. The developmental competence of each embryo was assessed at 8 dpi, enabling a division into two subgroups: degenerate and blastocyst. Data are presented as mean ± SD of three experiments.

TABLE 1 | The differentially expressed miRNAs ($p < 0.05$) in CM content from individually cultured bovine embryos (I, intermediate cleaving; S, slow cleaving).

	(I + S) Degenerate embryos vs. (I + S) Blastocysts		
Name	Fold Change	p-value	Benjamini-Hochberg corrected p-value
bta-miR-10b	3.047	0.006246833	0.224885994
bta-novel-miR-113	0.268	0.012946778	0.233041999
bta-novel-miR-45	3.934	0.031555548	0.255968877
bta-novel-miR-44	3.969	0.034171109	0.255968877
bta-novel-miR-139	0.124	0.035551233	0.255968877
SB vs. IB			
bta-miR-30c	17.857	0.041729634	0.707989068
bta-miR-10b	3.831	0.048368042	0.707989068

and miR-10b were secreted more in slow cleaving embryos' CM compared with the CM of intermediate cleaving embryos; miR-10b, miR-novel-44, and miR-novel-45 were more abundant in CM from degenerate embryos than in that of blastocysts, while miR-novel-113 and miR-novel-139 were more abundant in blastocyst's CM than degenerate's CM (**Table 1**). However, with the low sample size, due to the practical difficulty to obtain enough CM, it was unsurprising that none of the differences remained significant after multiple testing with the Benjamini-Hochberg corrected p-value. Consequently, the sequencing results of 5 of the 6 above mentioned miRNAs were confirmed using RT-qPCR (novel-miR-44 has the same mature sequence as novel-miR-45) (**Figure 2**). RT-qPCR showed that miR-30c and miR-10b have an 18 ($P = 0.00072$) and 30 ($P = 0.00017$) fold higher expression in the CM from slow cleaving embryos in comparison to intermediate cleaving embryos (**Figures 2A,B**). The expression levels of both these two miRNAs in the CM of fast cleaving embryos and intermediate cleaving embryos showed no significant difference (**Figures 2A,B**). MiR-10b and novel-miR-45 showed a 55 ($P = 0.00000$) and 8 ($P = 0.0068$) fold
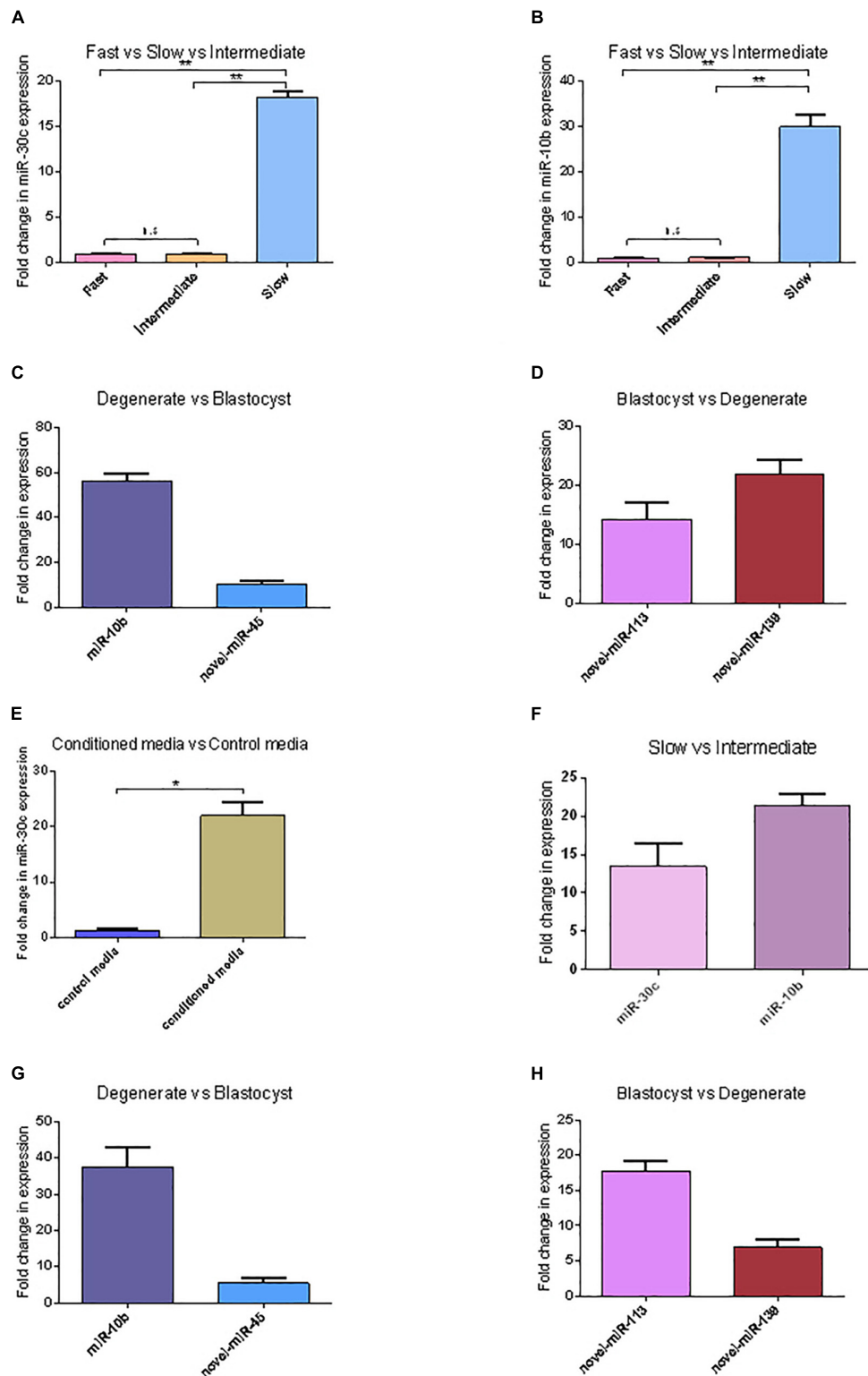
FIGURE 2 | miRNAs' differential expression between embryos with different cleavage patterns and different blastocyst formation. **(A,B)** The relative expression levels of miR-30c and miR-10b between the CM of fast, intermediate and slow developing embryos were detected using RT-qPCR. **(C,D)** Relative expression of miRNAs between CM of degenerate embryos and blastocyst. **(E)** Relative expression of miR-30c between conditioned media and control media. **(F)** Relative expression of miR-30c and miR-10b between the intermediate and slow developing embryos. **(G,H)** Relative expression of miRNAs between CM of degenerate embryos and blastocyst. Data are presented as mean ± SD of three experiments. (*$P < 0.05$, **$P < 0.01$, n.s, no significance).

FIGURE 3 | Top-ranking GO terms and KEGG pathways enriched in genes differentially expressed in IB vs. SB. **(A,B)** Enriched GO and KEGG pathways with Benjamini-Hochberg corrected *p*-value <0.05.
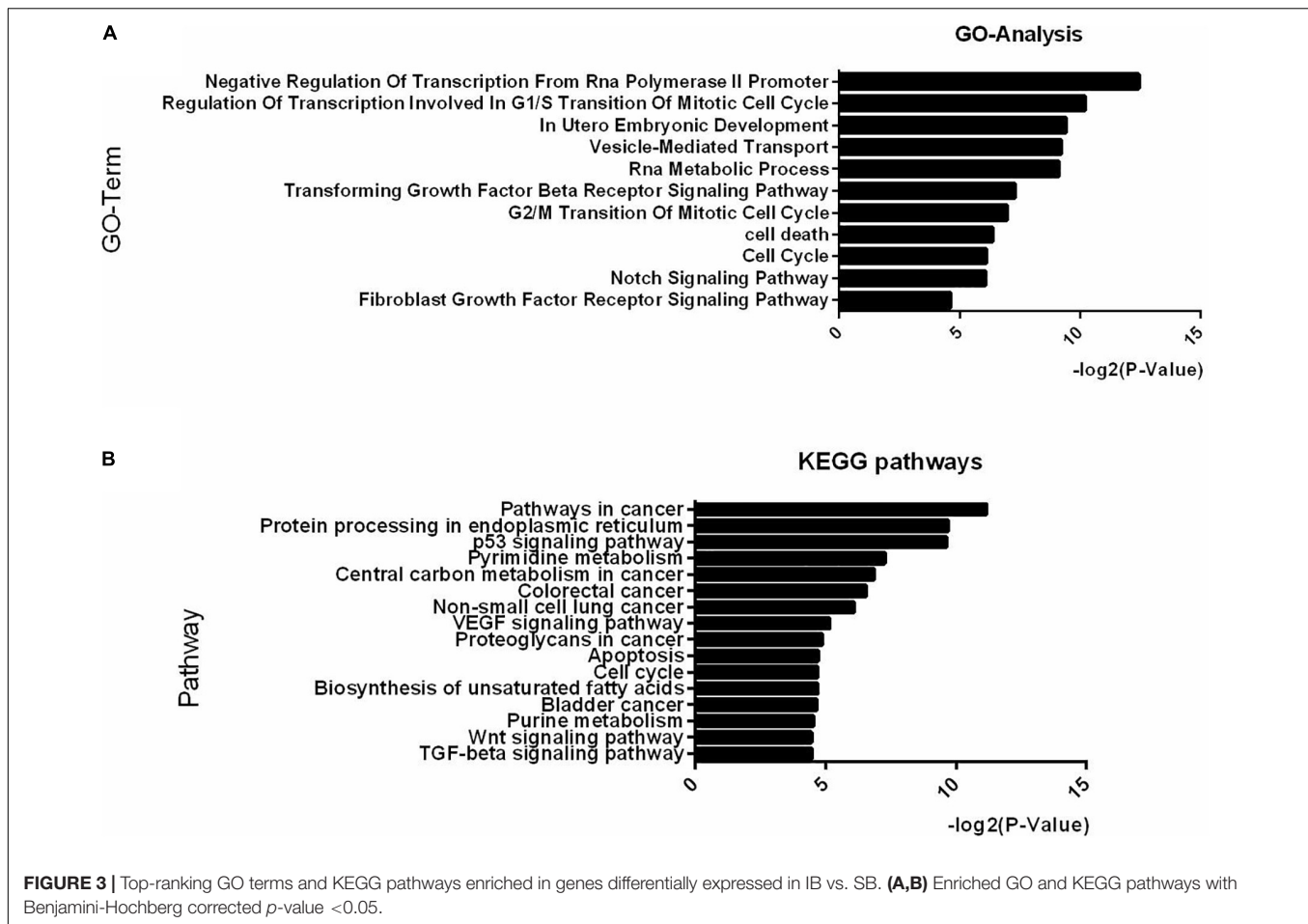
higher expression in the CM from degenerate embryos compared to blastocysts (**Figure 2C**). Novel-miR-113 and novel-miR-139 displayed, respectively 14 (*P* = 0.0027) and 22 (*P* = 0.00033) fold higher expression in the CM of blastocysts than in that of degenerate embryos (**Figure 2D**). In addition, miR-30c was found to be 20 (*P* = 0.00067) times more abundant in CM compared to control media (**Figure 2E**).

The intracellular miRNAs expression was also validated using RT-qPCR and similar results were obtained. miR-30c and miR-10b have a 13 (*P* = 0.0031) and 21 (*P* = 0.00044) times higher expression in slow cleaving embryos in comparison to intermediate cleaving embryos (**Figure 2F**). MiR-10b and novel-miR-45 show a 37 (*P* = 0.0004) and 5 (*P* = 0.0081) times higher expression in degenerate embryos compared to blastocysts (**Figure 2G**). Novel-miR-113 and novel-miR-139 displayed 18 (*P* = 0.00091) and 7 (*P* = 0.0062) times higher expression in blastocysts than degenerate embryos (**Figure 2H**).

Pathway Analysis

Examination of the GO analysis results of the differentially expressed miRNAs between IB and SB revealed that 11 biological processes, among which "in utero embryonic development," "cell cycle," "fibroblast growth factor receptor signaling pathway," and "Notch signaling pathway" were over-represented (**Figure 3A**).

Additionally, 16 KEGG pathways, with as top hits: the p53 signaling pathway, the Wnt signaling pathway, the TGF-beta signaling pathway and apoptosis were over-represented (**Figure 3B**). These GO-terms and pathways enriched with targets provide an intriguing clue to the biological consequences of miRNAs differential secretion from embryos with different cleavage patterns.

miR-30c Mimics Can Be Taken Up by Bovine Embryos and Increase Embryo Apoptosis

miR-30c has been shown to regulate cell cycle and proliferation in human breast cancer cells, glioma cells, hematopoietic cells, osteoblast cells and mice embryonic carcinoma cells (Li et al., 2012; Quintavalle et al., 2013; Shukla et al., 2015; Liu et al., 2016), thus, combining the above sequencing/RT-qPCR results with information from the literature we hypothesized that miR-30c can be taken up by embryos and might influence embryonic development through regulation of the cell cycle. To test this hypothesis, we added the miR-30c mimics into the IVF culture medium at 21 hpi, thus allowing miR-30c mimics to influence embryos for at least 5 to 10 h before they reach the 2-cell stage (26–31 hpi). RT-qPCR results showed
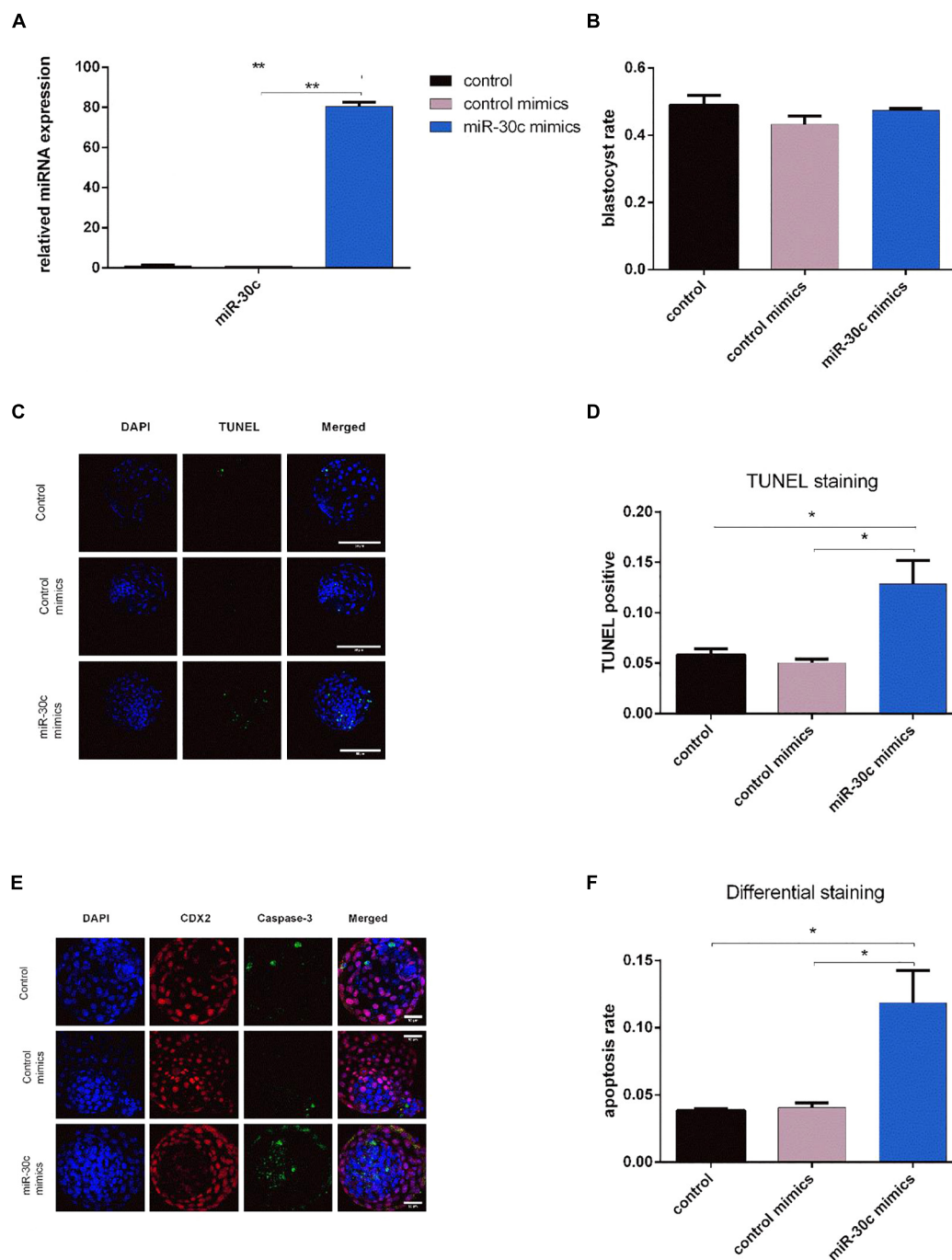
FIGURE 4 | Effects of miR-30c mimics on miR-30c expression in embryos, embryo growth and apoptosis **(A)** Embryos were treated with miR-30c mimics or control mimics and miR-30c expression was evaluated using RT-qPCR. The blastocyst rate was assessed at 8 dpi **(B)** and cell apoptosis was determined by TUNEL staining **(C,D)** and differential apoptotic staining **(E,F).** Data are presented as mean ± SD of three experiments. (*$P < 0.05$, **$P < 0.01$).

that the miR-30c levels were approximate 80 times higher in miR-30c mimics treated embryos compared to the control mimics group (**Figure 4A**), indicating that miR-30c was taken up by the embryos.

No significant difference was found in blastocyst rate between miR-30c mimics group and control mimics group (**Figure 4B**). However, TUNEL staining showed that the miR-30c mimics group had an apoptosis rate of 12.86% whereas that of the control

mimics group was 5.05% (**Figures 4C,D**). Similarly, differential apoptotic staining showed that the miR-30c mimics group had an apoptosis rate of 11.85% whereas that of the control mimics group was 4.05% (**Figures 4E,F**).

miR-30c Directly Targets Cell Progression Regulator *CDK12*

Different miRNA target prediction methods may produce different results, thus we adopted the method from Ozen (Ozen et al., 2007) and Li (Li et al., 2011). If a target was identified by at least three of six used different algorithms (TargetScan, miRDB, PicTar, miRanda, miRWalk and Tarbase), it was considered likely to be a miRNA target. Of the putative target genes identified in this way, *CDK12* (identified by Targetscan, miRDB and miRanda) was chosen for further analysis. This gene was previously shown to be required for the prevention of apoptosis

(Bartkowiak et al., 2015; Juan et al., 2016) and to protect cells from genomic instability and inhibit cell differentiation (Blazek et al., 2011; Dai et al., 2012) through the regulation of DDR genes in human and mouse. The 3′-UTR segment of the bovine *CDK12* gene containing the putative miR-30c target binding site region (**Figure 5A**) was amplified and cloned into luciferase reporter vector psi-CHECK2 and subsequently transfected to HEK293T cells. As shown in **Figure 5B**, the miR-30c mimics dramatically suppressed the activities of wild-type (WT) 3′-UTRs of *CDK12*, while the mutated 3′-UTR binding site (MUT) was unaffected. To further confirm the regulatory relationship between miR-30c and *CDK12*, RT-qPCR, and WB were performed to determine the *CDK12* mRNA and protein levels in MDBKs. The results showed that *CDK12* was suppressed by miR-30c mimics and enhanced by miR-30c inhibitors at the protein level (**Figure 5C**) rather than the mRNA level (**Figure 5D**). The direct target relationship was also analyzed
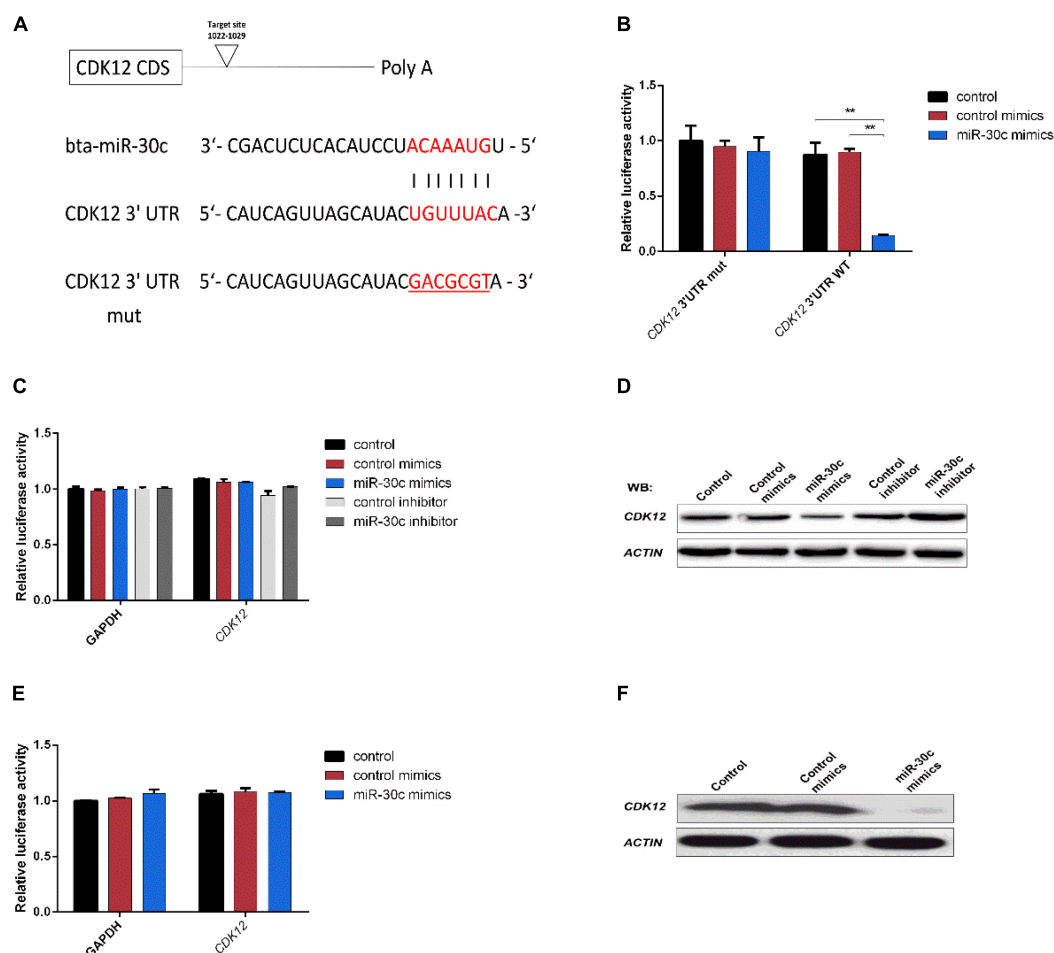
FIGURE 5 | *CDK12* is a direct target of miR-30c. **(A)** 3′-UTR analysis of *CDK12* containing putative regions that match the seed sequence of miR-30c. The mutated nucleotides are underlined. **(B)** Overexpression of miR-30c inhibited the Renilla luciferase activities. HEK-293T cells were cotransfected with 5 nM miR-30c mimics and 500 ng of reporter plasmid containing the WT or MUT-type UTRs. 24 h later, Renilla luciferase values were normalized against firefly luciferase and presented. **(C,D)** miR-30c mimics or inhibitor were transfected into MDBKs. After 48 or 24 h, cells were harvested for western blot or RT-qPCR. **(E,F)** Embryos were treated with miR-30c mimics or control mimics and the relative levels of *CDK12* was detected using RT-qPCR and WB. Data are presented as mean ± SD of three experiments. (**$P < 0.01$).

in embryos: miR-30c mimics were supplemented into embryos culture medium and then *CDK12* expression was evaluated using RT-qPCR and WB. Not surprisingly, embryos showed the similar results as MDBKs (**Figures 5E,F**). Collectively, these results show that miR-30c directly targets *CDK12* and inhibits its translation instead of degrading mRNA.

miR-30c Overexpression and *CDK12* Downregulation Direct Transcription of Key DDR Genes

Given that *CDK12* is involved in DNA repair (Paculová et al., 2017) and has been proven to be a target gene inhibited by miR-30c in our study, we hypothesized that miR-30c may suppress cell cycle progression by inhibiting DDR pathways. A previous study on mouse embryos showed that four DDR genes, namely *Brca1*, *Fancd2*, *Fanci*, and *Atr*, had a reduced expression in the absence of *CDK12* (Juan et al., 2016). To our knowledge, in bovine, the relationship among miR-30c, *CDK12*, and DDR pathway has not been investigated yet. Here we examined the expression of these four genes using RT-qPCR after supplementing miR-30c mimics into embryos culture medium and modulating *CDK12* expression in MDBKs. As shown in **Figure 6A**, the delivery of miR-30c significantly decreased mRNA levels of all four investigated DDR genes *BRCA1*, *FANCD2*, *FANCI*, and *ATR* in embryos. As shown in **Figure 6B**, downregulation of *CDK12* also significantly decreased mRNA levels of the above four genes. We also examined the expression of DDR genes after overexpressing *CDK12* using the previously mentioned vector construct. As shown in **Figure 6C**, overexpression of *CDK12* did not have a significant effect on the mRNA level of these DDR genes.

miR-30c Suppresses the Cell Cycle, While *CDK12* Promotes the Cell Cycle

Although miR-30c has been shown to regulate cell progression in human and mouse (Quintavalle et al., 2013; Liu et al., 2016), this regulatory relationship is still unclear in bovine cells. Considering the fact that the compaction of embryos makes it difficult to use them for flow cytometry analysis, further studies were performed using the bovine cell line MDBKs. PI staining was used to determine the effect of miR-30c mimics or inhibitors on the MDBK cell cycle. As shown in **Figure 7A**, cell cycle phase distribution determined by flow cytometry displayed 8% increase of treated cells in the G1 phase after delivery of miR-30c mimics, indicating the cell growth suppression, while delivery of miR-30c inhibitors resulted in an 8% decrease of cells in G1 phase.

CDK12 expression was assessed after siRNA or vector transfection. RT-qPCR (**Figure 7C**) and WB (**Figure 7D**) showed that *CDK12* expression was indeed upregulated by the vector construct and downregulated by siRNA at transcriptional level, showing their usefulness for the next experiments. Transfection of the *CDK12* overexpressing construct resulted in an 8% decrease of cells in the G1 phase compared with the control group, whereas knockdown of *CDK12* using siRNA resulted in a 20% increase of cells in the G1 phase and a 15% decrease in the S phase compared with si-NTC (**Figure 7E**).
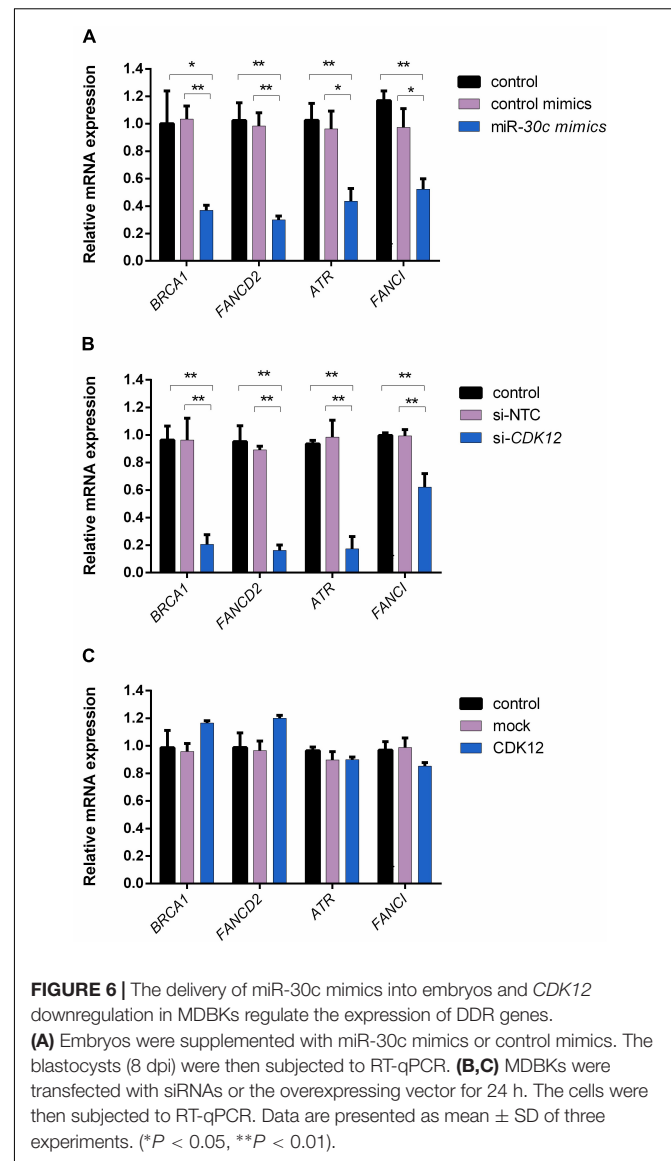
FIGURE 6 | The delivery of miR-30c mimics into embryos and *CDK12* downregulation in MDBKs regulate the expression of DDR genes. **(A)** Embryos were supplemented with miR-30c mimics or control mimics. The blastocysts (8 dpi) were then subjected to RT-qPCR. **(B,C)** MDBKs were transfected with siRNAs or the overexpressing vector for 24 h. The cells were then subjected to RT-qPCR. Data are presented as mean ± SD of three experiments. (*$P < 0.05$, **$P < 0.01$).

miR-30c Decreases Cell Viability, While *CDK12* Increases Cell Viability

The MDBKs cellular metabolic activity, indicative of the cell proliferation, was monitored after addition of miR-30c mimics or inhibitors using the WST-1 assay. As shown in **Figure 7B**, miR-30c mimics led to a significant decrease in cell viability (37%), while miR-30c inhibitors increased cell viability (57%).

As shown in **Figure 7F**, *CDK12* overexpression increased cell viability (84%), while *CDK12* inhibition led to a 49% decrease in cell viability.

DISCUSSION

Timing of cleavage is regarded as an important marker to assess embryo quality (Gutierrez-Adan et al., 2015) and it has been shown that rapid cleaving embryos are of better quality
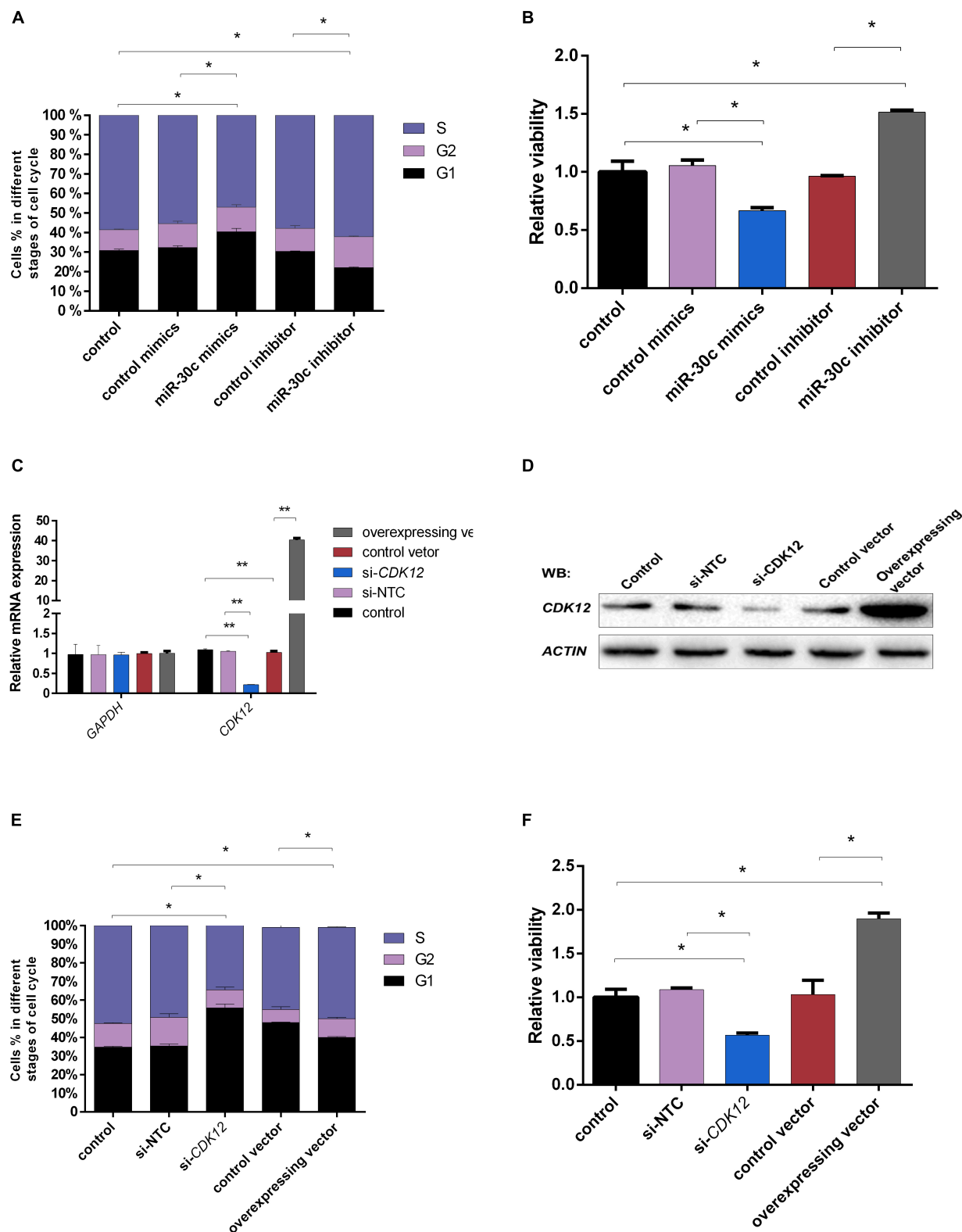
FIGURE 7 | Effects of miR-30c mimics/inhibitor and *CDK12* overexpression/downregulation on cell process. **(A,B)** MDBK cells were reverse transfected with miR-30c mimics or inhibitor for 48 h. The cells were then subjected to cell cycle assay by PI staining and cell proliferation assay by WST-1 assay. **(C–F)** MDBK cells were reverse transfected with siRNAs or overexpressed vector for 48 h. The cells were then subjected to RT-qPCR, WB, cell cycle assay by PI staining and cell proliferation assay by WST-1 assay. Data are presented as mean $\pm$ SD of three experiments. (*$P < 0.05$, **$P < 0.01$).

than slower cleaving embryos (Meirelles et al., 2004; Vandaele et al., 2006). Because it has been demonstrated that an embryo's potential is determined more in the early developmental stages than in later developmental stages (Wong et al., 2010; Milewski and Ajduk, 2017; Milewski et al., 2018), we chose the 2-cell stage to assess the samples regarding evaluation of embryos quality, instead of the 4-cell stage or the morula stage. In our study, fast and intermediate cleaving embryos produced significantly more blastocysts compared to the slow cleaving embryos (50.65 and 41.16% vs. 18.7%), confirming the above theory.

In addition to their intracellular function, secreted miRNAs may play a significant role in intercellular communications (Vickers et al., 2011; Boon and Vickers, 2013; Yang et al., 2018). However, the dynamics of miRNA secretion and their transfer mechanisms are still poorly understood. Secreted miRNAs have been found to be related to cell growth, invasion, migration, dissemination as well as metastasis and impairment of the immune system response (Schwarzenbach et al., 2014). Furthermore, they have potential as biomarkers for cancer and benign diseases, thus raising the questions whether and how secreted miRNAs influence embryo development and if they can be used as non-invasive biomarkers for embryo quality. Given the current methods for miRNA detection, the main limitation is the low abundance of miRNAs in CM. However, miRNAs secreted by a single human embryo have been successfully detected and extracted (Capalbo et al., 2016), indicating the potential application for bovine embryos. In our study, to obtain a sufficient amount of miRNAs for sequencing, we concentrated the CM from 167 embryos for each replicate and thus achieved at least 1 million raw reads. The potential of secreted miRNAs as biomarkers relies mainly on their high stability and their capacity to reflect embryo developmental status and their prognostic abilities in relation to IVF success and pregnancy outcome. Although there are several recent studies focusing on miRNAs in culture media (Rosenbluth et al., 2014; Kropp and Khatib, 2015) and body fluids, such as follicular fluid (Sohel et al., 2013) and endometrium (Vilella et al., 2015), providing an indication of the developmental competence of embryos, improvements in detection techniques and more knowledge of the miRNA signaling is needed in order to use secreted miRNA as biomarkers in embryonic development. In addition, not only technical aspects currently limit the use of secreted miRNAs as biomarkers in culture media and also in other body fluids; to date the source of secreted miRNAs is not clear. Therefore, more extensive studies are necessary to clarify whether secreted miRNAs detected in extracellular environment are the product of dead cells or are secreted in a tissue-specific manner. Furthermore, studies with large samples sizes are needed and some aspects of experimental reliability must be assessed before secreted miRNAs can be used as biomarkers.

Apart from the easy detection, a biomarker should be clearly discriminatory for the state to be defined, in casu the developmental competence of the embryo. Here, we demonstrated that miRNAs are differentially secreted from bovine embryos with different cleavage patterns and different qualities: miR-30c and miR-10b were differentially expressed between slow and intermediate cleaving embryos'

CM; miR-10b, miR-novel-113, miR-novel-44, miR-novel-45, and miR-novel-139 were differentially expressed between blastocyst's and degenerate's CM.

Among the differentially expressed miRNAs, miR-30c was found to be 18 times more abundant in slow cleaving embryos' CM vs. intermediate cleaving embryos' CM. This distinct difference makes it a suitable biomarker candidate for the developmental capacity of bovine early embryos. To gauge the effect of miR-30c uptake by bovine embryos in correlation with the cleavage pattern and the proposed roles of miR-30c in cell proliferation in mouse (Liu et al., 2016), cell apoptosis in human and mouse (Li et al., 2010; Quintavalle et al., 2013; Liu et al., 2016), cell differentiation in human and mouse (Karbiener et al., 2011; Wu et al., 2012) and cell damage in human (Li et al., 2012), apoptosis assays were performed. RT-qPCR results confirmed that miR-30c was indeed taken up by bovine embryos and they showed a higher apoptosis rate, which is in agreement with previous findings that miRNAs could be both released and taken up by embryos (Kropp and Khatib, 2015; Vilella et al., 2015; Gross et al., 2017). The effect was further investigated using the bovine cell line MDBK. The delivery of miR-30c mimics to the MDBKs led to reduced cell proliferation and an arrest at G1 stage, while the delivery of miR-30c inhibitors resulted in the opposite effects, as expected. Previous studies on human embryos suggested that miR-30c can serve as a potential marker of blastocyst implantation potential (Capalbo et al., 2016; Noli et al., 2016). This is not surprising because although miR-30c is highly conserved between different species, it has been shown to act differently among different species. For instance, miR-30c was found to increase cell proliferation in mouse embryonal carcinoma cells (Liu et al., 2016), while it was also found to be a tumor suppressor miRNA in human cancers (Poudel et al., 2013; Shukla et al., 2015).

We also demonstrated for the first time that miR-30c downregulates *CDK12* expression at a post-transcriptional level both on bovine embryos and MDBKs. *CDK12* is a transcription-associated CDK that exerts control over Pol II-mediated transcription (Ekumi et al., 2015) and is essential for splicing and differentiation (Chilà et al., 2016). Intriguingly, recent research has shown that *CDK12* is essential for embryonic development and the maintenance of genomic stability by regulating the expression of DDR genes, and reduced expression of some of these DDR genes will subsequently trigger apoptosis (Juan et al., 2016; Chen et al., 2017). During early embryonic development, DNA replication is prominent and highly efficient DNA repair is crucial for proper embryo development. For instance, *Atr*- and *Brca1*-lacking embryos were reported to display growth retardation in mice (Liu, 1996; Brown and Baltimore, 2000). In our study, both the supplementation of miR-30c mimics into bovine embryos culture medium and the *CDK12* knockdown in MDBKs caused a decreased expression level of key DDR genes *BRCA1*, *FANCD2*, *FANCI*, and *ATR*. These results present evidence that miR-30c overexpression or *CDK12* downregulation reduces the expression of these DDR genes at the transcriptional level, leading to a potential failure of DNA damage repair. Interestingly, while *CDK12* overexpression increased cell cycle progression and cell proliferation, it had no effect on the mRNA

level of those key DDR genes. This indicates that *CDK12* overexpression might influence cell cycle progression at other levels or through other mechanisms. For instance, in breast cancer cells, *CDK12* overexpression led to altered alternative last exon splicing of a subset of genes (Tien et al., 2017) and increased the invasiveness of a breast cancer cell line by decreasing the expression of the long isoform of DNAJB6 (Paculová and Kohoutek, 2017). A potential weakness of our study is that due to technical difficulties, part of the functional analysis of *CDK12* was done on MDBKs. It would be better if we can validate this mechanism in bovine embryos.

In summary, we have found 114 known miRNAs and 180 potential novel miRNAs in CM of bovine embryos. We have also identified miR-30c, which can be secreted and taken up by bovine embryos, as a novel potential biomarker related to bovine embryo apoptosis and reduced development. As miR-30c directly targets *CDK12* and downregulates DDR genes, it may exert its effects on cell cycle progression by inhibiting the DDR pathways.

AUTHOR CONTRIBUTIONS

XL performed the experiments and wrote the manuscript. YG provided the bioinformatics analysis for miRNA sequencing. EB contributed the qPCR experiments. KS was responsible for the embryo staining. KP helped to produce embryos. SMC, JC, PS, FVN, DD, AVS, and LP participated in the study design. All authors reviewed the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00315/full#supplementary-material

REFERENCES

Abd El Naby, W. S., Hagos, T. H., Hossain, M. M., Salilew-Wondim, D., Gad, A. Y., Rings, F., et al. (2013). Expression analysis of regulatory microRNAs in bovine cumulus oocyte complex and preimplantation embryos. *Zygote* 21, 31–51. doi: 10.1017/S0967199411000566

Amarnath, D., Kato, Y. and Tsunoda, Y. (2007) Effect of the timing of first cleavage on in vitro developmental potential of nuclear-transferred bovine oocytes receiving cumulus and fibroblast cells. *J. Reprod. Dev.* 53, 491–497. doi: 10.1262/jrd.18112

Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., et al. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* 24, 2303–2316. doi: 10.1101/gad.1968210

Bartkowiak, B., Yan, C., and Greenleaf, A. L. (2015). Engineering an analog-sensitive CDK12 cell line using CRISPR/Cas. *Biochim. Biophys. Acta* 1849, 1179–1187. doi: 10.1016/j.bbagrm.2015.07.010

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., et al. (2011). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes, *Genes Dev.* 25, 2158–2172. doi: 10.1101/gad.16962311

Boon, R. A., and Vickers, K. C. (2013). Intercellular Transport of MicroRNAs. *Arterioscler. Thromb. Vasc. Biol.* 33, 186–192. doi: 10.1161/ATVBAHA.112.300139

Brown, E. J., and Baltimore, D. (2000). ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.* 14, 397–402.

Capalbo, A., Ubaldi, F. M., Cimadomo, D., Noli, L., Khalaf, Y., Farcomeni, A., et al. (2016). MicroRNAs in spent blastocyst culture medium are derived from trophectoderm cells and can be explored for human embryo reproductive competence assessment. *Fertil. Steril.* 105, 225–235.e3. doi: 10.1016/j.fertnstert.2015.09.014

Chen, H. R., Juan, H. C., Wong, Y. H., Tsai, J. W., and Fann, M. J. (2017). Cdk12 regulates neurogenesis and late-arising neuronal migration in the developing cerebral cortex. *Cereb. Cortex* 27, 2289–2302.

Chilà, R., Guffanti, F., and Damia, G. (2016). Role and therapeutic potential of CDK12 in human cancers. *Cancer Treat. Rev.* 50, 83–88. doi: 10.1016/j.ctrv.2016.09.003

Dai, Q., Lei, T., Zhao, C., Zhong, J., Tang, Y. Z., Chen, B., et al. (2012). Cyclin K-containing kinase complexes maintain self-renewal in murine embryonic stem cells. *J. Biol. Chem.* 287, 25344–25352. doi: 10.1074/jbc.M111.321760

Dinnyes, A. (1999) Timing of the first cleavage post-insemination affects cryosurvival of in vitro–produced bovine blastocysts. *Mol. Reprod. Dev.* 53, 318–324 doi: 10.1002/(SICI)1098-2795(199907)53:3<318::AID-MRD7>3.0.CO;2-O

Donker, R. B., Mouillet, J. F., Chu, T., Hubel, C. A., Stolz, D. B., Morelli, A. E., et al. (2012). The expression profile of C19MC microRNAs in primary human trophoblast cells and exosomes. *Mol. Hum. Reprod.* 18, 417–424. doi: 10.1093/molehr/gas013

Dweep, H., and Gretz, N. (2015). miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat. Methods* 12:697. doi: 10.1038/nmeth.3485

Ekumi, K. M., Paculova, H., Lenasi, T., Pospichalova, V., Bösken, C. A., Rybarikova, J., et al. (2015). Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Res.* 43, 2575–2589. doi: 10.1093/nar/gkv101

Fenwick, J., and Platteau, P., Murdoch, A. P., Herbert, M. (2002). Time from insemination to first cleavage predicts developmental competence of human preimplantation embryos in vitro. *Hum. Reprod.* 17, 407–412. doi: 10.1093/humrep/17.2.407

Goossens, K., Mestdagh, P., Lefever, S., Van Poucke, M., Van Zeveren, A., Van Soom, A., et al. (2013). Regulatory microRNA network identification in bovine blastocyst development. *Stem Cells Dev.* 22, 1907–1920. doi: 10.1089/scd.2012.0708

Goovaerts, I. G., Leroy, J. L., Van Soom, A., De Clercq, J. B., Andries, S., and Bols, P. E. (2009). Effect of cumulus cell coculture and oxygen tension on the in vitro developmental competence of bovine zygotes cultured singly. *Theriogenology* 71, 729–738. doi: 10.1016/j.theriogenology.2008.09.038

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140–D144. doi: 10.1093/nar/gkj112

Gross, N., Kropp, J., and Khatib, H. (2017) Sexual dimorphism of miRNAs secreted by bovine in vitro-produced embryos. *Front. Genet.* 8:39. doi: 10.3389/fgene.2017.00039

Gutierrez-Adan, A., White, C. R., Van Soom, A., and Mann, M. R. (2015). Why we should not select the faster embryo: lessons from mice and cattle. *Reprod. Fertil. Dev.* 27, 765–775. doi: 10.1071/RD14216

Herrmann, D., Dahl, J. A., Lucas-Hahn, A., Collas, P., and Niemann, H. (2013). Histone modifications and mRNA expression in the inner cell mass and trophectoderm of bovine blastocysts. *Epigenetics* 8, 281–289. doi: 10.4161/epi.23899

Hiiragi, T., Louvet-Vallée, S., Solter, D., and Maro, B. (2006). Does prepatterning occur in the mouse egg?. *Nature* 442:E3. doi: 10.1038/nature04907

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4:44. doi: 10.1038/nprot.2008.211

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314.

Juan, H. C., Lin, Y., Chen, H. R., and Fann, M. J. (2016). Cdk12 is essential for embryonic development and the maintenance of genomic stability. *Cell Death Differ.* 23, 1038–1048. doi: 10.1038/cdd.2015.157

Karbiener, M., Neuhold, C., Opriessnig, P., Prokesch, A., Bogner-Strauss, J. G., and Scheideler, M. (2011). MicroRNA-30c promotes human adipocyte differentiation and co-represses PAI-1 and ALK2. *RNA Biol.* 8, 850–860. doi: 10.4161/rna.8.5.16153

Kropp, J., and Khatib, H. (2015). Characterization of microRNA in bovine in vitro culture media associated with embryo quality and development. *J. Dairy Sci.* 98, 6552–6563. doi: 10.3168/jds.2015-9510

Li, J., Donath, S., Li, Y., Qin, D., Prabhakar, B. S., and Li, P. (2010). miR-30 regulates mitochondrial fission through targeting p53 and the dynamin-related protein-1 pathway. *PLoS Genet.* 6:e1000795. doi: 10.1371/journal.pgen.1000795

Li, T., Chen, J. X., Fu, X. P., Yang, S., Zhang, Z., Chen Kh, H., et al. (2011). microRNA expression profiling of nasopharyngeal carcinoma. *Oncol. Rep.* 25, 1353–1363.

Li, W., Goossens, K., Van Poucke, M., Forier, K., Braeckmans, K., Van Soom, A., et al. (2016). High oxygen tension increases global methylation in bovine 4-cell embryos and blastocysts but does not affect general retrotransposon expression. *Reprod. Fertil. Dev.* 28, 948–959. doi: 10.1071/RD14133

Li, X. H., Ha, C. T., Fu, D., and Xiao, M. (2012). Micro-RNA30c negatively regulates REDD1 expression in human hematopoietic and osteoblast cells after gamma-irradiation. *PLoS One* 7:e48700. doi: 10.1371/journal.pone.0048700

Liang, K., Gao, X., Gilmore, J. M., Florens, L., Washburn, M. P., Smith, E., et al. (2015). Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol. Cell. Biol.* 35, 928–938. doi: 10.1128/MCB.01426-14

Liu, P. (1996). Inactivation of the mouse Brca1 gene leads to failure in the morphogenesis of the egg cylinder in early postimplantation development. *Genes Dev.* 10, 1835–1843. doi: 10.1101/gad.10.14.1835

Liu, X., Li, M., Peng, Y., Hu, X., Xu, J., Zhu, S., et al. (2016). miR-30c regulates proliferation, apoptosis and differentiation via the Shh signaling pathway in P19 cells. *Exp. Mol. Med.* 48:e248. doi: 10.1038/emm.2016.57

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Lu, Z., Zhang, C., Han, C., An, Q., Cheng, Y., Chen, Y., et al. (2019). Plasticizer bis(2-ethylhexyl) phthalate causes meiosis defects and decreases fertilization ability of mouse oocytes in vivo. *J. Agric. Food Chem.* [Epub ahead of print] doi: 10.1021/acs.jafc.9b00121

Luo, S. S., Ishibashi, O., Ishikawa, G., Ishikawa, T., Katayama, A., Mishima, T., et al. (2009). Human villous trophoblasts express and secrete placenta-specific micrornas into maternal circulation via exosomes. *Biol. Reprod.* 81, 717–729. doi: 10.1095/biolreprod.108.075481

Market Velker, B. A., Denomme, M. M., and Mann, M. R. (2012). Loss of genomic imprinting in mouse embryos with fast rates of preimplantation development in culture. *Biol. Reprod.* 86, 143, 1–16.

Meirelles, F. V., Schwarz, K. L., Merighe, G. K. F., Carambula, S. F., and Watanabe, Y. F. (2004). 158apoptosis in in vitro produced bovine embryos according

to developmental kinetics. *Reprod. Fertil. Dev.* 16, 201–201. doi: 10.1071/RDv16n1Ab158

Milewski, R., and Ajduk, A. (2017). Time-lapse imaging of cleavage divisions in embryo quality assessment. *Reproduction* 154, R37–R53. doi: 10.1530/REP-17-0004

Milewski, R., Szpila, M., and Ajduk, A. (2018). Dynamics of cytoplasm and cleavage divisions correlates with preimplantation embryo development. *Reproduction* 155, 1–14. doi: 10.1530/REP-17-0230

Mineno, J., Okamoto, S., Ando, T., Sato, M., Chono, H., Izu, H., et al. (2006). The expression profile of microRNAs in mouse embryos. *Nucleic Acids Res.* 34, 1765–1771. doi: 10.1093/nar/gkl096

Mondou, E., Dufort, I., Gohin, M., Fournier, E., and Sirard, M. A. (2012). Analysis of microRNAs and their precursors in bovine early embryonic development. *Mol. Hum. Reprod.* 18, 425–434. doi: 10.1093/molehr/gas015

Noli, L., Capalbo, A., Dajani, Y., Cimadomo, D., Bvumbe, J., Rienzi, L., et al. (2016). Human embryos created by embryo splitting secrete significantly lower levels of miRNA-30c. *Stem Cells Dev.* 25, 1853–1862. doi: 10.1089/scd.2016.0212

Ortiz-Escribano, N., Szymańska, K. J., Bol, M., Vandenberghe, L., Decrock, E., Van Poucke, M., et al. (2017). Blocking connexin channels improves embryo development of vitrified bovine blastocysts[†]. *Biol. Reprod.* 96, 288–301. doi: 10.1095/biolreprod.116.144121

Ozen, M., Creighton, C. J., Ozdemir, M., and Ittmann, M. (2007). Widespread deregulation of microRNA expression in human prostate cancer. *Oncogene* 27:1788. doi: 10.1038/sj.onc.1210809

Paculová, H., and Kohoutek, J. (2017). The emerging roles of CDK12 in tumorigenesis. *Cell Div.* 12:7. doi: 10.1186/s13008-017-0033-x

Paculová, H., Kramara, J., Šimečková, Š., Fedr, R., Souček, K., Hylse, O., et al. (2017). BRCA1 or CDK12 loss sensitizes cells to CHK1 inhibitors. *Tumor Biol.* 39:1010428317727479. doi: 10.1177/1010428317727479

Plusa, B., Hadjantonakis, A. -K., Gray, D., Piotrowska-Nitsche, K., Jedrusik, A., Papaioannou, V. E., et al. (2005). The first cleavage of the mouse zygote predicts the blastocyst axis. *Nature* 434:391. doi: 10.1038/nature03388

Poudel, S., Song, J., Jin, E. J., and Song, K. (2013). Sulfuretin-induced miR-30C selectively downregulates cyclin D1 and D2 and triggers cell death in human cancer cell lines. *Biochem. Biophys. Res. Commun.* 431, 572–578. doi: 10.1016/j.bbrc.2013.01.012

Quintavalle, C., Donnarumma, E., Iaboni, M., Roscigno, G., Garofalo, M., Romano, G., et al. (2013). Effect of miR-21 and miR-30b/c on TRAIL-induced apoptosis in glioma cells. *Oncogene* 32, 4001–4008. doi: 10.1038/onc.2012.410

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

Rosenbluth, E. M., Shelton, D. N., Wells, L. M., Sparks, A. E. T., and Van Voorhis, B. J. (2014). Human embryos secrete microRNAs into culture media—a potential biomarker for implantation. *Fertil. Steril.* 101, 1493–1500. doi: 10.1016/j.fertnstert.2014.01.058

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. doi: 10.1038/nrclinonc.2014.5

Shi, J., Dong, M., Li, L., Liu, L., Luz-Madrigal, A., Tsonis, P. A., et al. (2015). mirPRo–a novel standalone program for differential expression and variation analysis of miRNAs. *Sci. Rep.* 5:14617. doi: 10.1038/srep14617

Shukla, K., Sharma, A. K., Ward, A., Will, R., Hielscher, T., Balwierz, A., et al. (2015). MicroRNA-30c-2-3p negatively regulates NF-kappaB signaling and cell cycle progression through downregulation of TRADD and CCNE1 in breast cancer. *Mol. Oncol.* 9, 1106–1119. doi: 10.1016/j.molonc.2015.01.008

Sohel, M. M., Hoelker, M., Noferesti, S. S., Salilew-Wondim, D., Tholen, E., Looft, C., et al. (2013). Exosomal and non-exosomal transport of extra-cellular microRNAs in follicular fluid: implications for bovine oocyte developmental competence. *PLoS One* 8:e78505. doi: 10.1371/journal.pone.0078505

Sugimura, S., Akai, T., and Imai, K. (2017). Selection of viable in vitro-fertilized bovine embryos using time-lapse monitoring in microwell culture dishes. *J. Reprod. Dev.* 63, 353–357. doi: 10.1262/jrd.2017-041

Tang, F., Kaneda, M., O'Carroll, D., Hajkova, P., Barton, S. C., Sun, Y. A., et al. (2007). Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.* 21, 644–648. doi: 10.1101/gad.418707

Terriou, P., Giorgetti, C., Hans, E., Salzmann, J., Charles, O., Cignetti, L., et al. (2007) Relationship between even early cleavage and day 2 embryo score and

assessment of their predictive value for pregnancy. *Reprod. Biomed. Online* 14, 294–299. doi: 10.1016/S1472-6483(10)60870-X

Tien, J. F., Mazloomian, A., Cheng, S. W. G., Hughes, C. S., Chow, C. C. T., Canapi, L. T., et al. (2017). CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. *Nucleic Acids Res.* 45, 6698–6716. doi: 10.1093/nar/gkx187

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., and Lötvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* 9, 654–659. doi: 10.1038/ncb1596

Van Soom, A. (1997) Relationship between timing of development,morula morphology, and cell allocation to innercell mass and trophectoderm *in vitro-*producedbovine embryos. *Mol. Reprod. Dev.* 47, 47–56. doi: 10.1002/(SICI)1098-2795(199705)47:1<47::AID-MRD7>3.0.CO;2-Q

Vandaele, L., Mateusen, B., Maes, D., de Kruif, A., and Van Soom, A. (2006). Is apoptosis in bovine in vitro produced embryos related to early developmental kinetics and in vivo bull fertility? *Theriogenology* 65, 1691–1703. doi: 10.1016/j.theriogenology.2005.09.014

Vandaele, L., Mateusen, B., Maes, D. G., de Kruif, A., and Van Soom, A. (2007). Temporal detection of caspase-3 and -7 in bovine in vitro produced embryos of different developmental capacity. *Reproduction* 133, 709–718. doi: 10.1530/REP-06-0109

Varet, H., Brillet-Guéguen, L., Coppée, J. -Y., and Dillies, M. -A. (2016). SARTools: a DESeq2- and EdgeR-based R pipeline for comprehensive differential analysis of RNA-seq data. *PLoS One* 11:e0157022. doi: 10.1371/journal.pone.0157022

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, 423–433. doi: 10.1038/ncb2210

Vilella, F., Moreno-Moya, J. M., Balaguer, N., Grasso, A., Herrero, M., Martínez, S., et al. (2015). Hsa-miR-30d, secreted by the human endometrium, is taken up by the pre-implantation embryo and might modify its transcriptome. *Development* 142, 3210–3221. doi: 10.1242/dev.124289

Viswanathan, S. R., Mermel, C. H., Lu, J., Lu, C. W., Golub, T. R., and Daley, G. Q. (2009). microRNA expression during trophectoderm specification. *PLoS One* 4:e6143. doi: 10.1371/journal.pone.0006143.

Wong, C. C., Loewke, K. E., Bossert, N. L., Behr, B., De Jonge, C. J., Baer, T. M., et al. (2010). Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nat. Biotechnol.* 28, 1115–1121. doi: 10.1038/nbt.1686

Wu, H., Zhao, J., Fu, B., Yin, S., Song, C., Zhang, J., et al. (2017). Retinoic acid-induced upregulation of miR-219 promotes the differentiation of embryonic stem cells into neural cells. *Cell Death Dis.* 8:e2953. doi: 10.1038/cddis.2017.336

Wu, T., Zhou, H., Hong, Y., Li, J., Jiang, X., and Huang, H. (2012). miR-30 family members negatively regulate osteoblast differentiation. *J. Biol. Chem.* 287, 7503–7511. doi: 10.1074/jbc.M111.292722

Wydooghe, E., Heras, S., Dewulf, J., Piepers, S., Van den Abbeel, E., De Sutter, P., et al. (2014a). Replacing serum in culture medium with albumin and insulin, transferrin and selenium is the key to successful bovine embryo development in individual culture. *Reprod. Fertil. Dev.* 26, 717–724. doi: 10.1071/RD13043

Wydooghe, E., Vandaele, L., Piepers, S., Dewulf, J., Van den Abbeel, E., De Sutter, P., et al. (2014b). Individual commitment to a group effect: strengths and weaknesses of bovine embryo group culture. *Reproduction* 148, 519–529. doi: 10.1530/REP-14-0213

Wydooghe, E., Vandaele, L., Beek, J., Favoreel, H., Heindryckx, B., De Sutter, P., et al. (2011). Differential apoptotic staining of mammalian blastocysts based on double immunofluorescent CDX2 and active caspase-3 staining. *Anal. Biochem.* 416, 228–230. doi: 10.1016/j.ab.2011.05.033

Yang, J., Li, C., Zhang, L., and Wang, X. (2018). Extracellular vesicles as carriers of non-coding RNAs in liver diseases. *Front. Pharmacol.* 9:415. doi: 10.3389/fphar.2018.00415

Yang, Y., Bai, W., Zhang, L., Yin, G., Wang, X., Wang, J., et al. (2008). Determination of microRNAs in mouse preimplantation embryos by microarray. *Dev. Dyn.* 237, 2315–2327. doi: 10.1002/dvdy.21666

Zernicka-Goetz, M. (2002). Patterning of the embryo: the first spatial decisions in the life of a mouse. *Development* 129:815.

Zernicka-Goetz, M. (2006). The first cell-fate decisions in the mouse embryo: destiny is a matter of both chance and choice. *Curr. Opin. Genet. Dev.* 16, 406–412. doi: 10.1016/j.gde.2006.06.011

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

frontiers
in Genetics

Small Non-coding RNAs: New Class of Biomarkers and Potential Therapeutic Targets in Neurodegenerative Disease

Callum N. Watson[1,2]*, Antonio Belli[1,2] and Valentina Di Pietro[1,2,3]

[1] Neuroscience and Ophthalmology Research Group, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, United Kingdom, [2] National Institute for Health Research Surgical Reconstruction and Microbiology Research Centre, Queen Elizabeth Hospital Birmingham, Birmingham, United Kingdom, [3] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign, Urbana, IL, United States

Neurodegenerative diseases (NDs) are becoming increasingly prevalent in the world, with an aging population. In the last few decades, due to the devastating nature of these diseases, the research of biomarkers has become crucial to enable adequate treatments and to monitor the progress of disease. Currently, gene mutations, CSF and blood protein markers together with the neuroimaging techniques are the most used diagnostic approaches. However, despite the efforts in the research, conflicting data still exist, highlighting the need to explore new classes of biomarkers, particularly at early stages. Small non-coding RNAs (MicroRNA, Small nuclear RNA, Small nucleolar RNA, tRNA derived small RNA and Piwi-interacting RNA) can be considered a "relatively" new class of molecule that have already proved to be differentially regulated in many NDs, hence they represent a new potential class of biomarkers to be explored. In addition, understanding their involvement in disease development could depict the underlying pathogenesis of particular NDs, so novel treatment methods that act earlier in disease progression can be developed. This review aims to describe the involvement of small non-coding RNAs as biomarkers of NDs and their potential role in future clinical applications.

Keywords: small non-coding RNAs, microRNAs, neurodegenerative disease, biomarkers, new therapeutic targets

INTRODUCTION

Neurodegenerative diseases (NDs) are classified as a class of disorders affecting the central nervous system and they are characterized by the progressive loss of neuronal tissues. NDs are age-dependent disorders which are increasing internationally, due to the ever increasing elderly population, which is leaving greater numbers of people subjected to the chronic, debilitating nature of these incurable diseases (Heemels, 2016). Currently, the most represented NDs are: Alzheimer'disease (AD) with 5 million people affected in America only, followed by Parkinson's diseases (PD) with 1 million people; multiple sclerosis (MS) 400,000; Amyotrophoic lateral sclerosis (ALS) 30,000 and Huntington's disease (HD) with 3,000 incidents (Agrawal and Biswas, 2015).

Some treatments for ND have aimed to reduce the syndrome of NDs; these include L-dopa and deep brain stimulation in PD (Groiss et al., 2009; Nagatsua and Sawadab, 2009). However, very

few have aimed to slow or reverse ND development, and those that have been investigated e.g., stem cell therapy (Chung et al., 2002; Rachakonda et al., 2004) highlight the requirement for more research. Late diagnosis leads to strategic treatment being ineffective due to irreversible disease progression (Sheinerman and Umansky, 2013). This has been reported for example, on anti-AD therapies in late-stage clinical trials (including dimebon of Medivation and Pfizer, solanezumab of Eli Lilly and bapineuzumab of Pfizer and Johnson & Johnson). Biomarkers for early diagnosis could prevent or limit disease development through prophylactic or early treatment, which has ignited interest. Currently, the most accurate diagnosis relies on neuropathology, mainly based on autopsy, or in the measurement of cerebrospinal fluid (CSF) proteins, such as tau or Aβ- in AD, which requires invasive procedures. However, blood proteins, such as Aβ1-42 peptide in AD or cytokines for ALS or HD (Agrawal and Biswas, 2015), as well as genetics diagnostics markers such as ApoE isoforms in AD or α-synuclein or Parkin for PD, have also demonstrated potential clinical utility (Agrawal and Biswas, 2015).

Neuroimaging techniques can also help to make the correct diagnosis and monitor the progress of NDs. Magnetic resonance imaging (MRI) is one of the most widely used neuroimaging techniques used for AD (Jack et al., 2011; McKhann et al., 2011) and for dementia with Lewy bodies (DLB) (Ciurleo et al., 2014). Magnetic resonance spectroscopy (MRS) has also showed promise in early diagnosis of PD and traumatic brain injury, measuring metabolic dysfunctions and irreversible neuronal damage (Vagnozzi et al., 2008).

Recently, a new class of circulating RNAs – non-coding RNAs – have been re-evaluated and are being considered as potential biomarkers. After years of the belief that 98% of the genome was "junk" due to its non-coding nature it was realized these genes had biologically functionality. Non-coding genes include introns, pseudogenes, repeat sequences and *cis/trans*-regulatory elements that function as RNA without translation. Estimations have suggested that 99% of total RNA content is made up of non-coding RNA, with numbers of validated non-coding RNAs (ncRNAs) increasing every year (Palazzo and Lee, 2015).

Currently ncRNAs can be defined by length – small 18–200 nts and long >200nts – or functionality with housekeeping ncRNAs such as ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) or regulatory ncRNAs like microRNAs (miRNAs), small nuclear RNAs (snRNAs), piwi-interacting RNA (piRNAs), tRNA derived small RNAs (tsRNAs) and long non-coding RNAs (lncRNAs) (Dozmorov et al., 2013). Nonetheless, difficulty distinguishing categories persists due to the crossover of properties.

Small non-coding RNAs (sncRNAs) have diverse roles, which in conjunction with other molecules involve gene regulation through either RNA interference, RNA modification or spliceosomal involvement (**Table 1**). Consequently, during disease progression their expression can alter. MiRNAs are the most studied sncRNA as biomarkers with involvement in various diseases including cancers, aging and neurodegenerative disease (Calin and Croce, 2006; Grasso et al., 2014; Di Pietro et al., 2017). Other sncRNAs have shown promise as biomarkers,

TABLE 1 | Classification of types of small non-coding RNAs.

Type of small non-coding RNA	Size (nts)	Function
MicroRNA (miRNA)	∼22	Ago – RNAi
Small nuclear RNA (snRNA)	∼150	Spliceosome components
Small nucleolar RNA (snoRNA)	60–140	RNA modification
Piwi-interacting RNA (piRNA)	26–31	PIWI – RNAi
tRNA derived small RNA (tsRNA)	15–50	Diverse

with links to neurodegenerative disease (Munoz-Culla et al., 2016). There is the potential for multiple sncRNA biomarkers for neurodegenerative diseases, which if found, could aid diagnosis in a clinical setting while demonstrating the processes underpinning the disease development. In future, this could produce novel therapies to treat neurodegenerative diseases using original methodologies.

In this review, we consider the evolving role of sncRNAs and discuss their involvement in neurodegenerative disease with particular emphasis on their potential as biomarkers.

MICRORNA

MiRNAs are the most studied sncRNA. Their biogenesis commences with the formation of a pri-miRNA made up of two stem-loop structure. A Drosha and DGCR8 complex cleaves the pri-miRNA to form a single stem-loop pre-miRNA. Dicer cleaves the pre-miRNA to create a double stranded miRNA, which is loaded onto Argonaute family of proteins to form the miRISC complex (**Figure 1Ai**). Accompanied to the miRISC complex, miRNAs regulate gene expression post-transcriptionally through degradation and repression of mRNA sequences by an Argonaute family protein mediated method (**Figure 1Aii**; O'Brien et al., 2018). A single miRNA can have multiple targets, likewise a target mRNA can be bound to by many different miRNAs, to enable more diverse signaling patterns.

MiRNAs show specific signaling in the brain, and were also found differentially expressed in bio-fluids. Although there is no consistent consensus on particular miRNAs or brain area yet, and no specific miRNA overlap between brain tissues and bio-fluids (as reported in **Table 2**) these findings certainly provide insights in the study of NDs pathogenesis.

MiRNAs are best studied in Alzheimer's disease (AD), which manifests itself as deposition of neurofibrillary tangles (NFT) and extracellular amyloid-β (Aβ), before neuronal degeneration and clinical symptoms materialize in the form of behavioral changes such as memory issues. NFT, Aβ and neuronal degeneration have been associated with dysregulation of miRNA gene expression, which could emanate from altered Aβ or Tau metabolism. MiRNAs effect Aβ metabolism by interacting with amyloid precursor protein (APP) through direct binding of the 3′untranslated region (3′UTR) to the APP mRNA, indirect inhibition through downregulation of Beta-secretase 1 (BACE1) and ATP-binding cassette transporter (ABCA1) or regulating alternative APP splicing. MiRNAs also affect Tau through regulation of microtubule associated protein tau (MAPT)
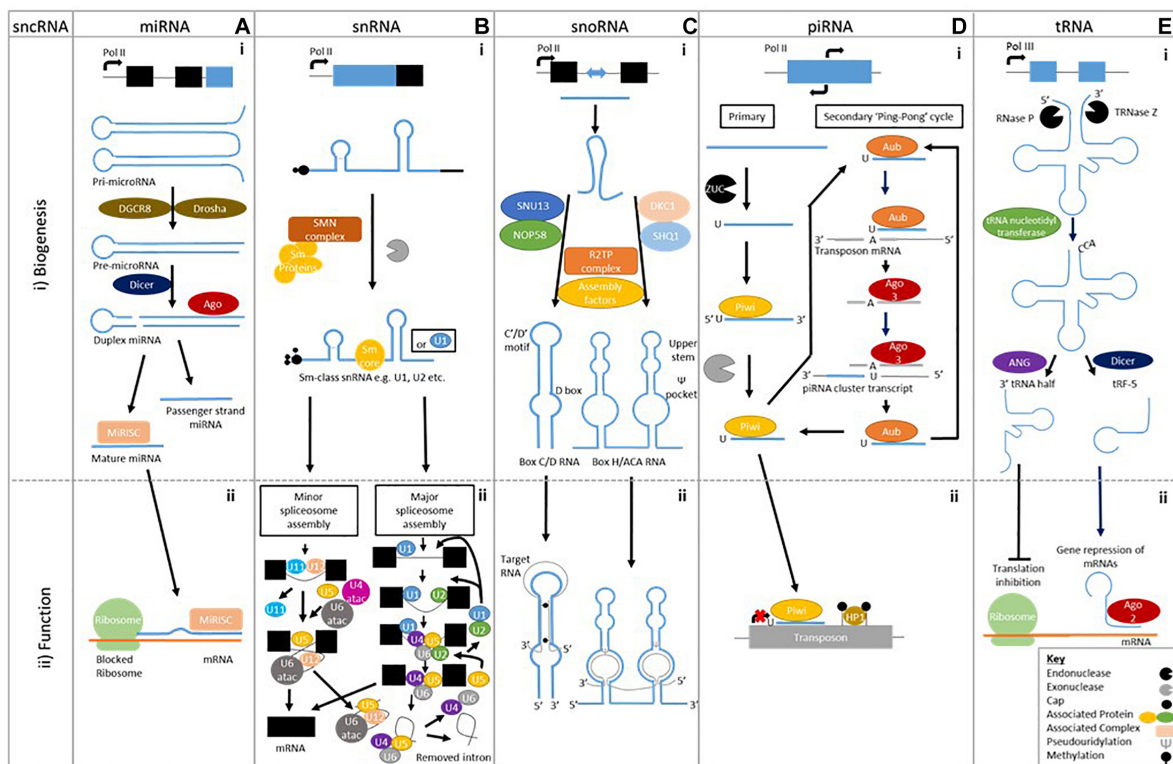
FIGURE 1 | Biogenesis of sncRNAs and an example of their biological function. **A (i)** MicroRNAs are single stranded ∼22 bp sequences formed from double stranded precursors **(ii)** that prevent mRNA translation. **B (i)** Small nuclear RNAs biogenesis is made up of two classes Sm class snRNA and Lsm-class snRNA (Not shown), **(ii)** which form the major and minor spliceosome. **C (i)** Small nucleolar RNAs have two different classes formed using different machinery; Box C/D RNA and Box H/ACA RNA, **(ii)** which cause methylation and pseudouridylation respectively. **D (i)** Piwi interacting RNAs are formed by either primary alone or by both primary and secondary biogenesis **(ii)** that prevent transposon translation through methylation. **E (i)** Transfer RNA cleavage forms transfer RNA derived fragments to be formed, **(ii)** which can prevent translation or cause gene repression.

splicing, affecting tau isoforms 3R and 4R. Direct or indirect binding either modulates phosphorylated Tau-associated protein kinases or influences degradation of phosphorylated tau by binding 3′-UTR BCL2 associated athanogene 2 (BAG2) mRNA (Zhao et al., 2017).

MiRNAs have an established involvement in neurobiological functions and pathogenesis of numerous other neurodegenerative diseases (Serafin et al., 2014; Fransquet and Ryan, 2018; Ricci et al., 2018). Mitochondrial dysfunction caused by miRNA dysregulation leads to oxidative stress, which causes cell death, α-synuclein aggregation and neurodegeneration known to be present in PD (Spano et al., 2015). In ALS, both TAR DNA binding protein (TARDBP) and fused in sarcoma (FUS) are well-established causative genes, which are involved in miRNA processing. TARDBP has specific roles in facilitation of post-transcriptional processing achieved through association directly with miRNA or processing factors such as Dicer (Kawahara and Mieda-Sato, 2012). FUS regulates miRNA-mediated gene silencing through facilitation of the interaction between miRNA, mRNA and RISC components (Zhang et al., 2018). In HD, a miRNA formulation is being trailed as therapeutic agents to alter the aberrant Huntingtin (HTT) protein expression (Aronin and DiFiglia, 2014).

MiRNA involvement in ND development has demonstrated the capability of distinguishing between disease subtypes and shown promise for future stratification. For example in AD, 30 differentially regulated miRNAs found in the brain and blood of AD patients were assigned to different Braak stages, a methodology for classifying AD pathology, with 10 associated with Braak stage III (hsa-mir-107, hsa-mir-26b, hsa-mir-30e, hsa-mir-34a, hsa-mir-485, hsa-mir200c, hsa-mir-210, hsa-mir-146a, hsa-mir-34c, and hsa-mir-125b) (Swarbrick et al., 2019). Likewise in PD, miR-331-5p is differentially expressed in plasma of early onset Parkinson's disease (EOPD) patients, which was not seen in late onset Parkinson's disease (LOPD) patients (Cardo et al., 2013; **Table 2**). Studies comparing between subtypes of NDs are still in the minority and more are required to understand the true capability of miRNA markers in stratification of NDs.

SMALL NUCLEAR RNAs

Small nuclear RNAs (snRNAs), the component parts of the spliceosome – responsible for removal of non-coding introns from precursor mRNA – are highly conserved uridine rich sequences with five snRNAs making up its spine; U1, U2, U4, U5,

TABLE 2 | MiRNAs with an involvement in the neurodegenerative disease development.

Neurodegenerative disease	Source	miRNA
Alzheimer's disease	Whole blood	miR-106b-5p, miR-106a-5p, miR-107, miR-9-5p (Yilmaz et al., 2016a) Let-7d-5p, Let-7f-5p, miR-107, miR-26a-5p, miR-26b-5p (Leidinger et al., 2013) miR-142-5p (Sorensen et al., 2016)
	Brain tissues	miR-9, miR-128 (Lukiw, 2007) miR-26a, miR-27b, miR-30e-5p, miR-34a, miR-92, miR-125, miR-145, miR-200c, miR-381, miR-422a, miR-423 miR-9, miR-132, miR-146b, miR-212 (Cogswell et al., 2008) miR-197, miR-511, miR-320, let-7i, miR-101, miR-106b, miR-15a, miR-181c, miR-19b, miR-22, miR-210, miR-26b, miR-29b-1, miR-363, miR-9, miR-93 (Hebert et al., 2008) miR-29a (Shioya et al., 2010) miR-26b (Absalon et al., 2013) miR-370, miR-328, miR-138, miR-132 and miR-15a (Bekris et al., 2013)
	CSF	let-7f, miR-105, miR-125a, miR-135a, miR-138, miR-141, miR-151, miR-186, miR-191, miR-197, miR-204, miR-205, miR-216, miR-302b, miR-30a5p, miR-30a3p, miR-30b, miR-30c, miR-30d, miR-32, miR-345, miR-362, miR-371, miR-374, miR-375, miR-380-3p, miR-429, miR-448, miR-449, miR-494, miR-501, miR-517, miR-517b, miR-518b, miR-518f, miR-520a*, miR-526a, miR-10a, miR-10b, miR-125, miR-126*, miR-127, miR-142-5p, miR-143, miR-146b, miR-154, miR-15b, miR-181a, miR-181c, miR-194, miR-195, miR-199a*, miR-214, miR-221, miR-328b, miR-422, miR-451, miR-455, miR-497, miR-99a (Bekris et al., 2013) miR-9, miR-125b, miR-146a, miR-155 (Alexandrov et al., 2012) let-7b (Lehmann et al., 2012) miR-146a, miR-155 (Lukiw et al., 2012) miR27a-3p (Sala Frigerio et al., 2013) miR-100, miR-146a, miR-296, miR-3622b-3p, miR-4467, miR-505, miR-766, miR-103, miR1274a, miR-375, miR-708, miR-219 (Denk et al., 2015)
	Plasma	miR-142-3p, miR-483-5p (Nagaraj et al., 2017) miR-125b-5p, miR-29b-3p, miR-3065-5p, miR-342-3p/5p (Lugli et al., 2015) miR-107 (Wang et al., 2008) miR-34a (Bhatnagar et al., 2014) miR-146a, miR-34a (Kiko et al., 2014) Let-7d-5p, Let-7g-5p, miR-142-3p, miR-15b-5p, miR-191-5p (Kumar et al., 2013)
	Serum	miR-143, miR-146a, miR-93 (Dong et al., 2015) miR-125b, miR-26b (Galimberti et al., 2014) Let-7d-5p, Let-7g-5p, miR-191-5p, miR-26b-3p, miR-30e-5p, miR-342-3p, miR-483-3p (Tan et al., 2014b) miR-125b, miR-181c, miR-9 (Tan et al., 2014a) miR-106b-3p, miR-181c-3p, miR-26a-5p (Guo R. et al., 2017) Let-7f-5p, miR-26b-5p, miR-501-3p (Hara et al., 2017) miR-125b (Jia and Liu, 2016) miR-106a-5p, miR-106b-3p, miR-143-3p, miR-15b-3p, miR-3065-5p, miR-30e-5p, miR-342-3p, miR-93-5p (Cheng et al., 2015) miR-181c (Geekiyanage et al., 2012) miR-455-3p (Kumar et al., 2017) miR-222 (Zeng et al., 2017) miR-29c-3p, miR-19b-3p (Wu et al., 2017)
	PBMCs	miR-29b (Villa et al., 2013) Let-7f, miR-34a (Schipper et al., 2007)
Early onset Parkinsons disease (EOPD)	Whole blood	miR-1, miR-22, miR-29a (Margis et al., 2011)
	Brain tissues	miR-34b,c (Minones-Moyano et al., 2011)
	Plasma	miR-331-5p (Cardo et al., 2013)
	Serum	miR-141, miR-146b-5p, miR-193a-3p, miR-214 (Dong et al., 2016)
Late onset Parkinson's disease (LOPD)	Whole Blood	miR-103a, miR29a, miR-30b (Serafin et al., 2015) miR-3143, miR-335-3p, miR-4671-3p, miR-561-3p, miR-579-3p (Yilmaz et al., 2016b)
	Brain tissues	miR-34b,c (Minones-Moyano et al., 2011) miR-181a,b,c,d, miR-22, miR-29a,b,c (Liao et al., 2013) miR-106a, miR-21, miR-224, miR-26b, miR-301b, miR-373 (Alvarez-Erviti et al., 2013) miR-205 (Cho et al., 2013) miR-135b, miR-198, miR-485-5p, miR-548d (Cardo et al., 2014) Let-7i-3p/5p, miR-10b-5p, miR-1224, miR-127-3p, miR-127-5p, miR-16-5p, miR-184, miR-29a-3p (Hoss et al., 2016) miR-144, miR-145, miR-199b, miR-221, miR-488, miR-543, miR-544, miR-7 (Tatura et al., 2016) miR-225, miR-236, miR-46 (Wake et al., 2016)

(Continued)

TABLE 2 | Continued

Neurodegenerative disease	Source	miRNA
	CSF	Let-7g-3p, miR-1224-5p, miR-127-3p, miR-128, miR-132-5p, miR-19a,b, miR-212-3p, miR-370, miR-409-3p, miR-4448, miR-485-5p, miR-873-3p (Burgos et al., 2014)
		Let-7g-3p, miR-1, miR-103a, miR-10a-5p, miR-119a, miR-126, miR-127-3p, miR-132-5p, miR-136-3p, miR-151, miR-153, miR-16-2, miR-19b-3p, miR-22, miR-26a, miR-28, miR-29a,c, miR-301a, miR-30b, miR-331-5p, miR-370, miR-374, miR-409-3p, miR-433, miR-485-5p, miR-873-3p (Gui et al., 2015)
		miR-1249, miR-1274b, miR-150, miR-16, miR-18b, miR-199b, miR-20a, miR-21, miR-320a,b, miR-378c, miR-4293, miR-671, miR-769, miR-92b (Soreq et al., 2013)
	Plasma	miR-222, miR-505, miR-626 (Khoo et al., 2012)
	Serum	miR-19b, miR-29a,c (Botta-Orfila et al., 2014)
		miR-133b (Zhao et al., 2014)
		miR-29a,b,c (Bai et al., 2017)
		miR-146a, miR-214, miR-221, miR-29c (Ma et al., 2016)
		miR-1294, miR-16-2-3p, miR-30a,e, miR-338-3p (Burgos et al., 2014)
		miR-148b, miR-223, miR-24, miR-30c, miR-324-3p (Vallelunga et al., 2014)
		miR-15b, miR-181a, miR-185, miR-195, miR-221 (Ding et al., 2016)
	PBMCs	miR-126, miR-126*, miR-147, miR-151-3p,5p, miR-199a-3p,5p, miR-199b, miR-19b, miR-26a, miR-28-5p, miR-29b,c, miR-301a, miR-30b,c, miR-335, miR-374a,b (Martins et al., 2011)
		miR-155, miRNA-146a (Caggiu et al., 2018)
Amyotrophic lateral sclerosis (ALS)	Whole Blood	let-7a-5p, let-7d-5p, let-7f-5p, let-7g-5p, let-7i-5p, miR-103a-3p, miR-106b-3p, miR-128-3p, miR-130a-3p, miR-130b-3p, miR-144-5p, miR-148a-3p, miR-148b-3p, miR-15a-5p, miR-15b-5p, miR-151a-5p, miR-151b, miR-16-5p, miR-182-5p, miR-183-5p, miR-186-5p, miR-22-3p, miR-221-3p, miR-223-3p, miR-23a-3p, miR-26a-5p, miR-26b-5p, miR-27b-3p, miR-28-3p, miR-30b-5p, miR-30c-5p, miR-342-3p, miR-425-5p, miR-451a, miR-532-5p, miR-550a-3p, miR-584-5p, miR-93-5p (Liguori et al., 2018)
	CSF	miR-150, miR-99b, miR-146a, miR-27b, miR-328, miR-532-3p (Butovsky et al., 2012)
		miR-132-5p, miR-132-3p, miR-143-3p, miR-143-5p, miR-574-5p (Freischmidt et al., 2013)
		miR-338-3p (De Felice et al., 2014)
		miR-181a-5p, miR-21-5p, miR-195-5p, miR-148-3p, miR-15b-5p, miR-let7a-5p, miR-let7b-5p, miR-let7f-5p (Benigni et al., 2016)
		miR-124-3p, miR-127-3p, miR-143-3p, miR-125b-2-3p, miR-9-5p, miR-27b-3p, miR-486-5p, miR-let7f-5p, miR-16-5p, miR-28-3p, miR-146a-3p, miR-150-5p, miR-378a-3p, miR-142-5p, miR-92a-5p (Waller et al., 2017b)
	Plasma	miR-4649-5p, miR-4299 (Takahashi et al., 2015)
		miR-424, miR-206 (de Andrade et al., 2016)
		miR-206, Pairs miR-206/miR-338-3p, miR-9*/miR-129-3p, miR-335-5p/miR-338-3p (Sheinerman et al., 2017)
	Serum	miR-132-3p, miR-132-5p, miR-143-3p, miR-143-5p, let-7b (Freischmidt et al., 2013)
		miR-206, miR-106b (Toivonen et al., 2014)
		miR-4745-5p, miR-3665, miR-4530, miR-1915-3p (Freischmidt et al., 2014)
		miR-1825, miR-1234-3p (Freischmidt et al., 2015)
		miR-206, miR-133a, miR-133b, miR-146a, miR-149*, miR-27a (Tasca et al., 2016)
		miR-1, miR-133a-3p, miR-133b, miR-144-5p, miR-192-3p, miR-195-5p, miR-19a-3p, let-7d-3p, miR-320a, miR-320b, miR-320c, miR-425-5p, miR-139-5p (Raheja et al., 2018)
		miR-206, miR-143-3p, miR-374b-5p (Waller et al., 2017a)
		miR-142-3p, miR-1249-3p (Matamala et al., 2018)
Huntington's disease	Brain tissues	miR-9/miR-9*, miR-124a, miR-132 (Packer et al., 2008)
		miR-10b-5p, miR-196a-5p, miR-615-3p, miR-10b-3p, miR-1298-3p, miR-196b-5p, miR-302a-3p, miR-1247-5p, miR-144-3p, miR-223-3p, miR-3200-3p, miR-302a-5p, miR-1264, miR-6734-5p, miR-144-5p, miR-138-2-5p, miR-431-5p, miR-132-3p, miR-200c-3p, miR-23b-5p, miR-448, miR-486-3p, miR-490-5p, miR-5695, miR-885-5p, miR-1224-5p, miR-1298-5p, miR-142-3p, miR-346, miR-891a-5p, miR-16-2-3p, miR-363-3p, miR-148a-3p, miR-199a-5p, miR-4449, miR-106a-5p, miR-142-5p, miR-549a, miR-214-5p, miR-141-3p, miR-5680, miR-3065-5p, miR-224-5p, miR-4787-3p, miR-452-5p, miR-129-1-3p, miR-4443, miR-101-5p, miR-483-5p, miR-2114-5p, miR-1185-1-3p, miR-670-3p, miR-129-5p, miR-135b-5p, miR-194-5p, miR-208b-3p, miR-4488, miR-888-5p, miR-126-5p, miR-34c-5p, miR-218-1-3p, miR-150-5p, miR-486-5p, miR-433-3p, miR-219b-3p, miR-548n, miR-663b, miR-148a-5p, miR-29a-3p, miR-320b, miR-181a-3p, miR-153-5p, miR-28-5p, miR-7-2-3p, miR-877-5p, miR-3687, miR-4516, miR-3139, miR-663a, miR-34b-3p, miR-1538 (Hoss et al., 2015a)
	CSF	miR-520f-3p, miR-135b-3p, miR-4317, miR-3928-5p, miR-8082, miR-140-5p (Reed et al., 2018)
	Plasma	miR-10b-5p, miR-486-5p (Hoss et al., 2015b)
		miR-34b (Gaughwin et al., 2011)
		miR-877-5p, miR-223-3p/5p, miR-30d-5p, miR-128, miR-22-5p, miR-222-3p, miR-338-3p, miR-130b-3p, miR-425-5p, miR-628-3p, miR-361-5p, miR-942 (Diez-Planelles et al., 2016)

*Passenger miRNA strand.

TABLE 3 | Interactions of small non-coding RNAs with Neurodegenerative diseases.

sncRNA	Disease	Interaction
snRNA major spliceosome	AD	U1 snRNPs present in cytoplasmic aggregates
	SMA	SMN1 gene dysregulation alters U snRNA levels
	Neurodegeneration	U2 snRNA mutation alters pre-mRNA splicing
	ALS	A disease related di-peptide repeat
	FTD	C90RF72 interacts with U2 snRNP
	RP	Mutation found in PRPF4 which encodes U4/U6 di-snRNP protein
snRNA minor spliceosome	ALS	Decreased U12 snRNA in spinal motor neurones
		Decreased TDP-43 disrupts U12 mediated pre-mRNA splicing
		FUS mutants cannot bind U11 so decreased minor intron splicing
snoRNA	AD	Differential expression of two C/D box snoRNAs e307 and e470 in mouse model
	ASD	SNORD115 duplication in mouse causes abnormal brain development
piRNA	AD	9 piRNAs found to be differentially regulated in AD risk variant patients (6 APOE and 3 RNU6-560P)
	PD	70 differentially expressed piRNAs in combined patient tissue and cells
tsRNA	ALS	ANG mutants implicated in pathogenesis
	PD	A subset of ALS-associated ANG mutants
	Intellectual disability	NSun2 mutation causes 5′tiRNA accumulation
	PCH	CLP1 gene mutation disruption of tRNA splicing
	Neurodegenerative patient	KAE1 gene mutation alters tRNA modification

and U6. These snRNAs combine with partner proteins to form the small nuclear ribonucleoprotein (snRNPs) complex, which is essential pre-mRNA splicing to enable production of functional mRNA for protein translation.

Sm-class snRNAs are synthesized by RNA polymerase II and after transcription contain a 7-methylguanosine cap, Sm-protein binding site and 3′ stem-loop. The latter two are recognized by the SMN complex, which recruits a set of Sm proteins to create the Sm-core RNP. Following this, the cap undergoes hypermethylation by trimethylguanosine synthase-1 (TSG1) creating a 2,2,7-trimethylguanosine cap. The 3′ end is then trimmed by an unknown exonuclease before subsequent maturation through modifications (Matera et al., 2007; **Figure 1Bi**).

Two types of spliceosome "major" and "minor" (0.35% of all introns) can be assembled. Major spliceosome assembly commences by U1 interacting with the 5′ splice site while U2 snRNP binds to the branch point sequence. This leads to the recruitment of the premade U4/U6.U5 tri-snRNP complex, in this state the spliceosome is inactive. After destabilization or

release of either U1 or U4, the spliceosome becomes active. The active spliceosome undergoes two phases of catalysis leading to its dissociation – including U2, U5, and U6 that are recycled – when it releases the mRNA, as mRNP (Wahl et al., 2009; **Figure 1Bii**). The minor spliceosome has divergent and highly conserved 5′ splice site and branch point sequences, which interact with U5 as well as alternative factors U11, U12, and U4atac/U6atac that are functional analog of its major counterpart (Verma et al., 2018; **Figure 1Bii**). Both spliceosomes show the capability to contribute to the development of neurodegenerative disease, demonstrating snRNA involvement (Bai et al., 2013; Tsuiji et al., 2013; Ratti and Buratti, 2016; Jutzi et al., 2018).

In sporadic and familial AD, U1 snRNP subunits – including U1-70K and U1A – were present in cytoplasmic aggregates, which occurs by the basic-acidic dipeptide (BAD) domain binding to tau in U1-70K (Bishof et al., 2018). Inordinate levels of unspliced RNA also reside, caused by dysregulation of RNA processing. In conjunction with evidence that inhibition of U1 snRNP increases APP, this implicates U1 snRNP dysregulation in the pathogenesis of AD (Bai et al., 2013; Hales et al., 2014a,b). Recent evidence has shown abnormal expression of U1 snRNA can cause premature cleavage of pre-mRNA via polyadenylation (PCPA) at the 3′ poly-A site. This affects splicing and could demonstrate a novel AD causing pathology (Cheng et al., 2017) (**Table 3**).

U snRNAs are also associated with spinal muscular atrophy (SMA). SMN1 gene dysregulation alters U snRNA levels through its role in U snRNA biosynthesis; nonetheless, the underlying pathology is still unclear (Zhang et al., 2013). Many studies have proposed a reduction in U snRNAs is key to SMA pathology due to their involvement in mRNA processing, with U1 and U11 of particular interest (Gabanella et al., 2007; Zhang et al., 2008). In contrast, U snRNAs can accumulate in the motor neurons of ALS patient spinal cords when compared to control patients, to cause defects showing that U snRNA level can depict disease state, depending of cell type (Tsuiji et al., 2013).

More recently, when considering induced pluripotent stem cell (iPSC) derived motor neurones cultures, a study suggested that an imbalanced ratio of variant U1 to U1 might cause the SMA phenotype rather than an overall reduction in U1 snRNA (Vazquez-Arango et al., 2016). Demonstrating that purely measuring U snRNA level may be an oversimplified measurement and variant U snRNA could indicate the underlying pathophysiology of aberrant spliceosome related neurodegeneration.

Other U snRNAs studied in neurodegenerative disease include U2. A U2 snRNA mutation causes neuron degeneration, through altering pre-mRNA splicing at select splice sites that are associated with alternative pre-mRNA splicing (Jia et al., 2012). In addition, a dipeptide repeat (C90RF72) linked to both ALS and frontotemporal dementia (FTD), interacts and interferes with U2 snRNP. In patient derived cells, this led to mislocalisation but mis-splicing linked to ALS/FTD has yet to be established (Yin et al., 2017).

Mutations found within the gene PRPF4 – which encodes hPrp4 a U4/U6 di-snRNP protein – undertake an important role in the development of retinitis pigmentosa (RP) (Chen et al., 2014). hPrp4 is known to interact with

CypH and hPrp3 to regulate the stability of the tri-snRNP, U4/U6.U5. Thus, aberrant splicing could cause RP through direct or indirect mechanisms that have been hypothesized, but not defined.

The minor spliceosome has ND relevance as in ALS, TDP-43 functionality decreases (Colombrita et al., 2012), which reduces the number Gemini of coiled bodies (GEMs). GEMs contribute to U12 snRNA biogenesis, so in spinal motor neurones of ALS patients there was a decrease of U12 snRNA and U11/U12 snRNP, which may disrupts pre-mRNA splicing (Ishihara et al., 2013). Additionally, an ALS mutant (P525L) cannot promote minor intron splicing due to an aberrant FUS gene that routinely binds to U11 snRNP to direct splicing. This leads to mislocalisation of FUS-trapped U11 and U12 snRNAs, which form aggregates in the cytoplasm so incorrect splicing results (Reber et al., 2016). In addition, a cerebral ataxia mutation RNU12 causes minor intron retention in homozygous mutant patients (Elsaid et al., 2017). When combined this demonstrates a likely role for minor intron splicing in motor neurone maintenance.

SMALL NUCLEOLAR RNAs

Small nucleolar RNAs (SnoRNAs) modify RNA through there conserved motifs, with boxes C/D guiding methylation and H/ACA guiding pseudouridylation, respectively (Ohtani, 2017; **Figure 1Cii**). Each class of snoRNAs displays a unique secondary structure composed of conserved proteins to form the defined C/D and H/ACA snoRNPs. SnoRNAs mainly target rRNA to modify functionally important regions of the ribosome (Decatur and Fournier, 2002) but other purposes include pre-rRNA endonucleolytic processing (Tollervey and Hurt, 1990), guiding snRNAs such as U6 snRNA (Tycowski et al., 1998) and more recently mRNA guiding (Sharma et al., 2016) or regulation of alternative splicing in pre-mRNAs (Falaleeva et al., 2016).

Box C/D snoRNP biogenesis commences when a protein complex of SNU13 and NOP58 is pre-formed and loaded onto the snoRNA with the help of HSP90/R2TP. This recruits assembly factors and the pre-snoRNPs are transferred to the Cajal bodies where final processing occurs. Box H/ACA RNPs biogenesis starts by SHQ1 and DKC1 combining to prevent to non-specific RNAs binding. SHQ1 is released with the help of the R2TP complex allowing DKC1 to bind H/ACA RNAs at the site of transcription. Numerous assembly factors including NHP2, NOP10, and NAF1 are present during this pre-snoRNP form. When NAF1 – which binds the C-terminal domain of RNA polymerase II to keep H/ACA RNP inactive – is replaced by GAR1, mature and functional H/ACA RNPs are produced. Both forms are transported to the nucleolus to elicit their actions (Massenet et al., 2017; **Figure 1Ci**).

A study showed differential regulation of two C/D box snoRNAs (e307 and e470) prior to the development of AD in mouse model. After formation of a β-amyloid plaque, this differential expression is no longer present, demonstrating that they could be useful in early diagnosis. No clear evidence of pathogenesis just hypothesized using bioinformatics methods (Gstir et al., 2014) (**Table 3**).

Despite the fact that autism spectrum disorder (ASD) might not be considered a neurodegenerative disease. Studies have found links in ASD with numerous snoRNA genes found to be differentially expressed using RNA-seq (Wright et al., 2017). Duplication of SNORD115 in mouse chromosome 7 that mirrors human chromosome 15q11-13 – duplication of this is one of the most common chromosomal abnormalities in ASD – has been shown to increase SNORD115 levels and results in abnormal brain development. In addition, SNORD115 (HBII-48 and HBII-52) levels are dysregulated in superior temporal gyrus of human ASD brain samples, which could explain 5-HT changes (Gabriele et al., 2014) and alternative splicing seen in ASD (Voineagu et al., 2011) as HBII-52 may regulate 5-HT2C receptor mRNA levels (Stamova et al., 2015) as well as alternative splicing (Kishore et al., 2010).

Another study demonstrated that maternal alcohol consumption in pregnancy alters the C/D box RNA levels in brain cells during abnormal fetal development. DNA methylation, microRNA and snoRNA levels altered with emphasis on SNORD115 increasing and SNORD116 decreasing (Laufer et al., 2013).

PIWI-INTERACTING RNA

Piwi-Interacting RNAs (PiRNAs) are a diverse range of small RNAs that are highly enriched in the germline tissues. They interact with PIWI-class Argonaute proteins with sequence bias for only the first 5′ nucleotide to be a Uracil. This diverse population can be mapped back to distinct areas of the genome known as piRNA clusters, which contain highly enriched areas of fragmented dysfunctional transposable element (TE) sequences. These are thought to emanate from the memory of previous TE invasions, and can be utilized to protect against TEs (Toth et al., 2016). In addition, PIWI proteins function at the chromatin level by guiding DNA methylation and deposition of repressive histone marks to silence TE transcription (Le Thomas et al., 2013; **Figure 1Dii**).

The biogenesis of piRNAs gives rise to two different forms primary and secondary of 26–30 bps in length, stemming from single-stranded precursors (Yan et al., 2011; Mani and Juliano, 2013), which are best studied in Drosophila. Primary piRNAs biogenesis is poorly defined but precursors of around 200 bp stemming nearly entirely from piRNA clusters are cleaved – Zucchini (ZUC) is thought to do this – to enable loading onto a PIWI protein in association with other factors (**Figure 1Di**). This piRNA-PIWI complex interacts with TEs to prevent insertion through methylation or transcriptional repression, thereby affecting gene expression (Toth et al., 2016).

In Drosophila, secondary piRNAs are formed through a more defined "ping-pong" pathway, which utilizes the primary piRNAs formed from TE fragments present in piRNA clusters loaded onto Aubergine (AUB) to find complementary antisense TE transcripts (**Figure 1D**). Once found the complementary TE mRNA binds, and is cleaved ten nucleotides along from the 5′ end by AUB, which terminates its function. Additionally it creates a new 5′ end and piRNA precursor, which accompanied by

AGO3 is processed into secondary piRNA. The secondary piRNA promotes the development of more cluster-derived piRNAs – it is representative of the sense TE strand – through complementary cluster transcripts to develop a greater repertoire against active TEs (Toth et al., 2016; **Figure 1Di**).

Originally piRNAs were solely thought to be present in germline cells, more recently they have been found in other areas of the body including blood (Yang et al., 2015), blood plasma (Freedman et al., 2016) and the brain (Roy et al., 2017) as well as interacting with diseases in the liver (Rizzo et al., 2016), cardiovascular system (Loche and Ozanne, 2016) and brain (Roy et al., 2017) demonstrating their roles are far-reaching. In neurodegenerative disease there have been recent studies on PD and AD.

Risk variants APOE (rs2075650) and RNU6-560P (rs10792835 + rs3851179) have been linked with AD through genome-wide association studies (GWAS). These risk variants were significantly correlated with nine (6 APOE and 3 RNU6-560P) different piRNAs, showing regulatory capabilities (Guo X. et al., 2017). PiRNA dysregulation may be integral to the development of AD through aberrant downstream signaling. The link to pathogenesis in AD was clarified in three AD dysregulated piRNAs (piR-38240, piR-34393, and piR-40666) after establishing complementary target genes (CYCS, KPNA6, and RAB11A) through inverse expression correlation (Roy et al., 2017). The target genes were known to regulate AD pathways through oxidative stress induced neurodegeneration, apoptosis and vesicular trafficking of Aβ. This demonstrates a regulatory role for piRNAs in preventing AD and so monitoring dysregulation could allow early diagnosis and implicate a treatment method.

There was a difference found in piRNA expression between PD- and control- patient derived cells. Patient tissue samples showed the same trend, with 70 different piRNAs overlapping between both (**Table 3**). Two distinct trends come from these piRNAs, up or down regulation (Schulze et al., 2018). In the down-regulated piRNA fraction, those that were short-interspersed nuclear elements (SINE) and long-interspersed nuclear elements (LINE) derived elements in cell lines and LINE in tissues, showed significant enrichment when compared to genome-wide expression (Schulze et al., 2018). This is indicative of an inability to silence SINE and LINE derived elements in PD-derived neurones, which could show a pathogenesis of PD disease.

TRANSFER RNAs

Transfer RNAs (tRNAs) are the most abundant form of sncRNA, making up 4–10% of all cellular RNAs. Previously thought to be static contributors to gene expression, acting as an adaptor molecule in translation. Recently it has been found that small non-coding tRNAs have unique function that enable wider signaling and dynamic regulation of various functions (Gebetsberger and Polacek, 2013).

Mature tRNA is formed through transcription of precursor tRNA (pre-tRNA) using RNA polymerase III. Endonucleolytic ribonuclease P (RNase P) and ribonuclease Z cleave the transcribed pre-tRNA at the 5′ leader sequence and 3′ polyuracil (poly –U) tail, respectively, before tRNA nucleotidyl transferase adds a 3′CCA tail (**Figure 1Ei**). Many post-transcriptional modifications will occur during maturation and only tRNAs appropriately processed will leave the nucleus via nuclear receptor-mediated export process, with wrongly processed terminating. The mature tRNAs are between 73–90 nts in length and contain a clover-leaf shaped secondary structure, composing of a D-loop, an anticodon loop, a T-loop, a variable loop and an amino acid acceptor stem (Kirchner and Ignatova, 2015). The mature of pre-tRNA can be cleaved – into specific products unlike previously thought – into two main categories of cleaved tRNAs have been categorized; (1) tRNA-halves, (2) tRNA derived fragments.

tRNA halves are produced by cleavage of the anticodon loop giving rise to two halves; 30–35 nt 5′-tRNA halves and 40–50 nt 3′tRNA halves (Li and Hu, 2012; **Figure 1Eii**). A subtype of tRNA halves known as tRNA-derived stress-induced RNAs (tiRNAs) are by-products of stress. They induce cleavage by angiogenin (ANG) – a ribonuclease – of mature cytoplasmic tRNAs (Yamasaki et al., 2009).

tRNA derived fragments (tRFs) are produced from either pre-tRNAs or mature tRNAs (**Figure 1Eii**). Four main types have been established stemming from the fragment location on tRNAs: 5-tRFs, 3-tRFs, 1-tRFs, and 2 tRFs. 5-tRFs – located most abundantly in the nucleus – are generated from cleavage of the D-loop of tRNAs by Dicer, with adenine being present at the 3′ ends. Further subdivision classifies 5-tRFs isoforms into "a" (∼15 nts), "b" (∼22 nts) and "c" (∼30 nts) (Kumar et al., 2015; Lee et al., 2009). 3-tRFs result from cleavage by Dicer, ANG or another member of the Ribonuclease A superfamily of the T-loop, containing a CCA tail sequence (18–22 nts) (Lee et al., 2009; Maraia and Lamichhane, 2011; Kumar et al., 2015). 1-tRFs are formed by the cleavage of the 3′-trailer fragment of pre-tRNAs by either RNaseZ or ELAC2, this usually commences after the 3′-ends of mature tRNA and contains a poly-U 3′-end (Lee et al., 2009; Liao et al., 2010). 2-tRFs, less known about but may be formed from the anticodon loop (Goodarzi et al., 2015).

Numerous neurodegenerative disorders are associated with tRFs. ANG mutants show reduced ribonuclease (RNase) activity and were first implicated in the pathogenesis of amyotrophic lateral sclerosis (ALS) (Greenway et al., 2006). Latterly, a subset of the ALS-associated ANG mutants were observed in Parkinson's disease (PD) patients (van Es et al., 2011). Recombinant ANG can improve life span and motor function in an ALS [SOD1 (G93A)] mouse model, demonstrating that tRFs may have an important role in motor neuron survival (Kieran et al., 2008) (**Table 3**).

The link between ANG-induced tiRNAs, cellular stress and neurodevelopment disorders was strengthened with the finding of NSun2 (Blanco et al., 2014). Mutations in the cytosine-5 RNA methyltransferase NSun2 have been shown to cause intellectual disability and a Dubowitz-like syndrome in humans (Abbasi-Moheb et al., 2012; Martinez et al., 2012). NSun2 methylates two different cytosine residues of tRNA. Without NSun2, cytosine-5 RNAs are not methylated, which increases the stress-induced ANG-mediated endonucleolytic cleavage of tRNAs and

so 5′-tiRNAs accumulate. Accumulation of these factors leads to cell death in hippocampal and striatal neurons because of translational repression leading to cellular stress. Subsequently, NSun2 knockout mice show reduced neuronal size and impaired formation of synapses, which could explain the impairment of NSun2 gene mutation patients (Blanco et al., 2014).

A mutation in CLP1 gene (R140A) – a RNA kinase involved in tRNA splicing – is present in pontocerebellar hypoplasia (PCH) patients, a heterogeneous group of inherited neurodegenerative disorders characterized by the loss of motor neurons, muscle paralysis, impaired development of various parts of the brain and differential tRNA splicing (Karaca et al., 2014; Schaffer et al., 2014). The role of CLP1 in RNA splicing means the mutant gene has reduced kinase activity and affinity to the tRNA endonuclease complex (TSEN), impairing pre-tRNA cleavage and elevating unspliced pre-tRNAs in patient derived neurons (Schaffer et al., 2014). TSEN cuts the transcript at 3′ intron-extron junctions, so the absence of CLP1 means 5′-unphosphorylated tRF cannot interact with the pre-tRNA^{tyr} 3′-exon and subsequent splicing steps are interrupted (Cassandrini et al., 2010).

N^6-threonyl-carbamoyl-adenosine (t^{6}A) is a complex modification of adenosine involved in cytoplasmic tRNA modification. It is located next to the anticodon loop of many tRNAs that decode ANN codons, at position 37 (t^{6}A37). Recently, a biosynthetic defect in the t^{6}A molecule resulting from a mutation in the kinase-associated endopeptidase (KAE1) gene, which is part of the kinase, endopeptidase and other proteins of small size (KEOPS) complex was found in two phenotypically neurodegenerative patients, implicating tRNA modification in neuronal maintenance (Edvardson et al., 2017).

Although, tRNA-derived small non-coding RNAs, have already demonstrated a role in cancer progression (Sun et al., 2018), their role as biomarkers in NDs has not been fully investigated yet.

However, animal studies showed 13 dysregulated tRFs in brain samples of SAMP8 mouse model for AD. In particular, four were upregulated (AS-tDR-011775, AS-tDR-011438, AS-tDR-006835 and AStDR-005058) and 9 down regulated (AS-tDR-013428, AS-tDR-011389, AS-tDR-009392, AS-tDR012690, AS-tDR-010654, AS-tDR-008616, AS-tDR-010789, AS-tDR-011670, and AS-tDR-007919), demonstrating their potential involvement of tRFs in early detection of AD.

CONCLUSION

The key problem with the ND field is the lack of understanding in the events preceding the development of protein-based markers – such as Tau – currently used to diagnose NDs. By this stage, the diseases become more difficult to treat.

SncRNAs play an important regulatory role in the maintenance of the homeostatic brain. Therefore, changes in their concentration levels can be indicative of mechanistic changes that could precede protein-based markers. One single sncRNA biomarker is unlikely to differentiate between diseases. However, a combination of sncRNA biomarkers could be illustrative of the mechanistic development of NDs to enable early diagnosis, enhanced disease monitoring as well as defining subtle differences between NDs. Consequently, novel treatment methods directly related to their mechanistic underpinning of specific NDs, and potentially other brain related pathologies can be envisaged.

Novel, less-well studied sncRNAs could be integral to understanding the overall disease progression. So new methodologies may be necessary to quantify these changes and allow for future biomarker development.

AUTHOR CONTRIBUTIONS

CW drafted the manuscript. AB and VDP critically revised the manuscript.

FUNDING

REFERENCES

Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., et al. (2012). Mutations in NSUN2 cause autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* 90, 847–855. doi: 10.1016/j.ajhg.2012.03.021

Absalon, S., Kochanek, D. M., Raghavan, V., and Krichevsky, A. M. (2013). MiR-26b, upregulated in Alzheimer's disease, activates cell cycle entry, tau-phosphorylation, and apoptosis in postmitotic neurons. *J. Neurosci.* 33, 14645–14659. doi: 10.1523/jneurosci.1327-13.2013

Agrawal, M., and Biswas, A. (2015). Molecular diagnostics of neurodegenerative disorders. *Front. Mol. Biosci.* 2:54. doi: 10.3389/fmolb.2015.00054

Alexandrov, P. N., Dua, P., Hill, J. M., Bhattacharjee, S., Zhao, Y., and Lukiw, W. J. (2012). microRNA (miRNA) speciation in Alzheimer's disease (AD) cerebrospinal fluid (CSF) and extracellular fluid (ECF). *Int. J. Biochem. Mol. Biol.* 3, 365–373.

Alvarez-Erviti, L., Seow, Y., Schapira, A. H., Rodriguez-Oroz, M. C., Obeso, J. A., and Cooper, J. M. (2013). Influence of microRNA deregulation on chaperone-mediated autophagy and alpha-synuclein pathology in Parkinson's disease. *Cell Death Dis.* 4:e545. doi: 10.1038/cddis.2013.73

Aronin, N., and DiFiglia, M. (2014). Huntingtin-lowering strategies in Huntington's disease: antisense oligonucleotides, small RNAs, and gene editing. *Mov. Disord.* 29, 1455–1461. doi: 10.1002/mds.26020

Bai, B., Hales, C. M., Chen, P. C., Gozal, Y., Dammer, E. B., Fritz, J. J., et al. (2013). U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16562–16567. doi: 10.1073/pnas.1310249110

Bai, X., Tang, Y., Yu, M., Wu, L., Liu, F., Ni, J., et al. (2017). Downregulation of blood serum microRNA 29 family in patients with Parkinson's disease. *Sci. Rep.* 7:5411. doi: 10.1038/s41598-017-03887-3

Bekris, L. M., Lutz, F., Montine, T. J., Yu, C. E., Tsuang, D., Peskind, E. R., et al. (2013). MicroRNA in Alzheimer's disease: an exploratory study in brain, cerebrospinal fluid and plasma. *Biomarkers* 18, 455–466. doi: 10.3109/1354750x.2013.814073

Benigni, M., Ricci, C., Jones, A. R., Giannini, F., Al-Chalabi, A., and Battistini, S. (2016). Identification of miRNAs as potential biomarkers in cerebrospinal fluid

from amyotrophic lateral sclerosis patients. *Neuromolecular Med.* 18, 551–560. doi: 10.1007/s12017-016-8396-8

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. doi: 10.3389/fnmol.2014.00002

Bishof, I., Dammer, E. B., Duong, D. M., Kundinger, S. R., Gearing, M., Lah, J. J., et al. (2018). RNA-binding proteins with basic-acidic dipeptide (BAD) domains self-assemble and aggregate in Alzheimer's disease. *J. Biol. Chem.* 293, 11047–11066. doi: 10.1074/jbc.RA118.001747

Blanco, S., Dietmann, S., Flores, J. V., Hussain, S., Kutter, C., Humphreys, P., et al. (2014). Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *EMBO J.* 33, 2020–2039. doi: 10.15252/embj. 201489282

Botta-Orfila, T., Morato, X., Compta, Y., Lozano, J. J., Falgas, N., Valldeoriola, F., et al. (2014). Identification of blood serum micro-RNAs associated with idiopathic and LRRK2 Parkinson's disease. *J. Neurosci. Res.* 92, 1071–1077. doi: 10.1002/jnr.23377

Burgos, K., Malenica, I., Metpally, R., Courtright, A., Rakela, B., Beach, T., et al. (2014). Profiles of extracellular miRNA in cerebrospinal fluid and serum from patients with Alzheimer's and Parkinson's diseases correlate with disease status and features of pathology. *PLoS One* 9:e94839. doi: 10.1371/journal.pone. 0094839

Butovsky, O., Siddiqui, S., Gabriely, G., Lanser, A. J., Dake, B., Murugaiyan, G., et al. (2012). Modulating inflammatory monocytes with a unique microRNA gene signature ameliorates murine ALS. *J. Clin. Invest.* 122, 3063–3087. doi: 10.1172/jci62636

Caggiu, E., Paulus, K., Mameli, G., Arru, G., Sechi, G. P., and Sechi, L. A. (2018). Differential expression of miRNA 155 and miRNA 146a in Parkinson's disease patients. *eNeurologicalSci* 13, 1–4. doi: 10.1016/j.ensci.2018.09.002

Calin, G. A., and Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6, 857–866. doi: 10.1038/nrc1997

Cardo, L. F., Coto, E., de Mena, L., Ribacoba, R., Moris, G., Menendez, M., et al. (2013). Profile of microRNAs in the plasma of Parkinson's disease patients and healthy controls. *J. Neurol.* 260, 1420–1422. doi: 10.1007/s00415-013-6900-8

Cardo, L. F., Coto, E., Ribacoba, R., Menendez, M., Moris, G., Suarez, E., et al. (2014). MiRNA profile in the substantia nigra of Parkinson's disease and healthy subjects. *J. Mol. Neurosci.* 54, 830–836. doi: 10.1007/s12031-014-0428-y

Cassandrini, D., Biancheri, R., Tessa, A., Di Rocco, M., Di Capua, M., Bruno, C., et al. (2010). Pontocerebellar hypoplasia: clinical, pathologic, and genetic studies. *Neurology* 75, 1459–1464. doi: 10.1212/WNL.0b013e3181f88173

Chen, X., Liu, Y., Sheng, X., Tam, P. O., Zhao, K., Chen, X., et al. (2014). PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* 23, 2926–2939. doi: 10.1093/hmg/ddu005

Cheng, L., Doecke, J. D., Sharples, R. A., Villemagne, V. L., Fowler, C. J., Rembach, A., et al. (2015). Prognostic serum miRNA biomarkers associated with Alzheimer's disease shows concordance with neuropsychological and neuroimaging assessment. *Mol. Psychiatry* 20, 1188–1196. doi: 10.1038/mp. 2014.127

Cheng, Z., Shang, Y., Gao, S., and Zhang, T. (2017). Overexpression of U1 snRNA induces decrease of U1 spliceosome function associated with Alzheimer's disease. *J. Neurogenet.* 31, 337–343. doi: 10.1080/01677063.2017.1395425

Cho, H. J., Liu, G., Jin, S. M., Parisiadou, L., Xie, C., Yu, J., et al. (2013). MicroRNA-205 regulates the expression of Parkinson's disease-related leucine-rich repeat kinase 2 protein. *Hum. Mol. Genet.* 22, 608–620. doi: 10.1093/hmg/dds470

Chung, S., Sonntag, K. C., Andersson, T., Bjorklund, L. M., Park, J. J., Kim, D. W., et al. (2002). Genetic engineering of mouse embryonic stem cells by Nurr1 enhances differentiation and maturation into dopaminergic neurons. *Eur. J. Neurosci.* 16, 1829–1838.

Ciurleo, R., Di Lorenzo, G., Bramanti, P., and Marino, S. (2014). Magnetic resonance spectroscopy: an in vivo molecular imaging biomarker for Parkinson's disease? *Biomed Res. Int.* 2014:519816. doi: 10.1155/2014/519816

Cogswell, J. P., Ward, J., Taylor, I. A., Waters, M., Shi, Y., Cannon, B., et al. (2008). Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways. *J. Alzheimers Dis.* 14, 27–41.

Colombrita, C., Onesto, E., Megiorni, F., Pizzuti, A., Baralle, F. E., Buratti, E., et al. (2012). TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional

fate in motoneuron-like cells. *J. Biol. Chem.* 287, 15635–15647. doi: 10.1074/jbc. M111.333450

de Andrade, H. M., de Albuquerque, M., Avansini, S. H., de, S. R. C., Dogini, D. B., Nucci, A., et al. (2016). MicroRNAs-424 and 206 are potential prognostic markers in spinal onset amyotrophic lateral sclerosis. *J. Neurol. Sci.* 368, 19–24. doi: 10.1016/j.jns.2016.06.046

De Felice, B., Annunziata, A., Fiorentino, G., Borra, M., Biffali, E., Coppola, C., et al. (2014). miR-338-3p is over-expressed in blood, CFS, serum and spinal cord from sporadic amyotrophic lateral sclerosis patients. *Neurogenetics* 15, 243–253. doi: 10.1007/s10048-014-0420-2

Decatur, W. A., and Fournier, M. J. (2002). rRNA modifications and ribosome function. *Trends Biochem. Sci.* 27, 344–351.

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:e0126423. doi: 10.1371/journal.pone.0126423

Di Pietro, V., Ragusa, M., Davies, D., Su, Z., Hazeldine, J., Lazzarino, G., et al. (2017). MicroRNAs as novel biomarkers for the diagnosis and prognosis of mild and severe traumatic brain injury. *J. Neurotrauma* 34, 1948–1956. doi: 10.1089/neu.2016.4857

Diez-Planelles, C., Sanchez-Lozano, P., Crespo, M. C., Gil-Zamorano, J., Ribacoba, R., Gonzalez, N., et al. (2016). Circulating microRNAs in Huntington's disease: emerging mediators in metabolic impairment. *Pharmacol. Res.* 108, 102–110. doi: 10.1016/j.phrs.2016.05.005

Ding, H., Huang, Z., Chen, M., Wang, C., Chen, X., Chen, J., et al. (2016). Identification of a panel of five serum miRNAs as a biomarker for Parkinson's disease. *Parkinsonism Relat. Disord.* 22, 68–73. doi: 10.1016/j.parkreldis.2015. 11.014

Dong, H., Li, J., Huang, L., Chen, X., Li, D., Wang, T., et al. (2015). Serum MicroRNA profiles serve as novel biomarkers for the diagnosis of Alzheimer's disease. *Dis. Markers* 2015:625659. doi: 10.1155/2015/625659

Dong, H., Wang, C., Lu, S., Yu, C., Huang, L., Feng, W., et al. (2016). A panel of four decreased serum microRNAs as a novel biomarker for early Parkinson's disease. *Biomarkers* 21, 129–137. doi: 10.3109/1354750x.2015.1118544

Dozmorov, M. G., Giles, C. B., Koelsch, K. A., and Wren, J. D. (2013). Systematic classification of non-coding RNAs by epigenomic similarity. *BMC Bioinformatics* 14(Suppl. 14):S2. doi: 10.1186/1471-2105-14-S14-S2

Edvardson, S., Prunetti, L., Arraf, A., Haas, D., Bacusmo, J. M., Hu, J. F., et al. (2017). tRNA N6-adenosine threonylcarbamoyltransferase defect due to KAE1/TCS3 (OSGEP) mutation manifest by neurodegeneration and renal tubulopathy. *Eur. J. Hum. Genet.* 25, 545–551. doi: 10.1038/ejhg.2017.30

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

Falaleeva, M., Pages, A., Matuszek, Z., Hidmi, S., Agranat-Tamir, L., Korotkov, K., et al. (2016). Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1625–E1634. doi: 10.1073/pnas.1519292113

Fransquet, P. D., and Ryan, J. (2018). Micro RNA as a potential blood-based epigenetic biomarker for Alzheimer's disease. *Clin. Biochem.* 58, 5–14. doi: 10.1016/j.clinbiochem.2018.05.020

Freedman, J. E., Gerstein, M., Mick, E., Rozowsky, J., Levy, D., Kitchen, R., et al. (2016). Diverse human extracellular RNAs are widely detected in human plasma. *Nat. Commun.* 7:11106. doi: 10.1038/ncomms11106

Freischmidt, A., Muller, K., Ludolph, A. C., and Weishaupt, J. H. (2013). Systemic dysregulation of TDP-43 binding microRNAs in amyotrophic lateral sclerosis. *Acta Neuropathol. Commun.* 1:42. doi: 10.1186/2051-5960-1-42

Freischmidt, A., Muller, K., Zondler, L., Weydt, P., Mayer, B., von Arnim, C. A., et al. (2015). Serum microRNAs in sporadic amyotrophic lateral sclerosis. *Neurobiol. Aging* 36, 2660.e15–2660.e20. doi: 10.1016/j.neurobiolaging.2015. 06.003

Freischmidt, A., Muller, K., Zondler, L., Weydt, P., Volk, A. E., Bozic, A. L., et al. (2014). Serum microRNAs in patients with genetic amyotrophic lateral sclerosis and pre-manifest mutation carriers. *Brain* 137, 2938–2950. doi: 10.1093/brain/ awu249

Gabanella, F., Butchbach, M. E., Saieva, L., Carissimi, C., Burghes, A. H., and Pellizzoni, L. (2007). Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. *PLoS One* 2:e921. doi: 10.1371/journal.pone.0000921

Gabriele, S., Sacco, R., and Persico, A. M. (2014). Blood serotonin levels in autism spectrum disorder: a systematic review and meta-analysis. *Eur. Neuropsychopharmacol.* 24, 919–929. doi: 10.1016/j.euroneuro.2014.02.004

Galimberti, D., Villa, C., Fenoglio, C., Serpente, M., Ghezzi, L., Cioffi, S. M., et al. (2014). Circulating miRNAs as potential biomarkers in Alzheimer's disease. *J. Alzheimers Dis.* 42, 1261–1267. doi: 10.3233/jad-140756

Gaughwin, P. M., Ciesla, M., Lahiri, N., Tabrizi, S. J., Brundin, P., and Bjorkqvist, M. (2011). Hsa-miR-34b is a plasma-stable microRNA that is elevated in pre-manifest Huntington's disease. *Hum. Mol. Genet.* 20, 2225–2237. doi: 10.1093/hmg/ddr111

Gebetsberger, J., and Polacek, N. (2013). Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol.* 10, 1798–1806. doi: 10.4161/rna.27177

Geekiyanage, H., Jicha, G. A., Nelson, P. T., and Chan, C. (2012). Blood serum miRNA: non-invasive biomarkers for Alzheimer's disease. *Exp. Neurol.* 235, 491–496. doi: 10.1016/j.expneurol.2011.11.026

Goodarzi, H., Liu, X., Nguyen, H. C., Zhang, S., Fish, L., and Tavazoie, S. F. (2015). Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. *Cell* 161, 790–802. doi: 10.1016/j.cell.2015.02.053

Grasso, M., Piscopo, P., Confaloni, A., and Denti, M. A. (2014). Circulating miRNAs as biomarkers for neurodegenerative disorders. *Molecules* 19, 6891–6910. doi: 10.3390/molecules19056891

Greenway, M. J., Andersen, P. M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., et al. (2006). ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nat. Genet.* 38, 411–413. doi: 10.1038/ng1742

Groiss, S. J., Wojtecki, L., Südmeyer, M., and Schnitzler, A. (2009). Review: deep brain stimulation in Parkinson's disease. *Ther. Adv. Neurol. Disord.* 2, 20–28. doi: 10.1177/1756285609339382

Gstir, R., Schafferer, S., Scheideler, M., Misslinger, M., Griehl, M., Daschil, N., et al. (2014). Generation of a neuro-specific microarray reveals novel differentially expressed noncoding RNAs in mouse models for neurodegenerative diseases. *RNA* 20, 1929–1943. doi: 10.1261/rna.047225.114

Gui, Y., Liu, H., Zhang, L., Lv, W., and Hu, X. (2015). Altered microRNA profiles in cerebrospinal fluid exosome in Parkinson disease and Alzheimer disease. *Oncotarget* 6, 37043–37053. doi: 10.18632/oncotarget.6158

Guo, R., Fan, G., Zhang, J., Wu, C., Du, Y., Ye, H., et al. (2017). A 9-microRNA signature in serum serves as a noninvasive biomarker in early diagnosis of Alzheimer's disease. *J. Alzheimers Dis.* 60, 1365–1377. doi: 10.3233/jad-170343

Guo, X., Qiu, W., Garcia-Milian, R., Lin, X., Zhang, Y., Cao, Y., et al. (2017). Genome-wide significant, replicated and functional risk variants for Alzheimer's disease. *J. Neural Transm.* 124, 1455–1471. doi: 10.1007/s00702-017-1773-0

Hales, C. M., Dammer, E. B., Diner, I., Yi, H., Seyfried, N. T., Gearing, M., et al. (2014a). Aggregates of small nuclear ribonucleic acids (snRNAs) in Alzheimer's disease. *Brain Pathol.* 24, 344–351. doi: 10.1111/bpa.12133

Hales, C. M., Seyfried, N. T., Dammer, E. B., Duong, D., Yi, H., Gearing, M., et al. (2014b). U1 small nuclear ribonucleoproteins (snRNPs) aggregate in Alzheimer's disease due to autosomal dominant genetic mutations and trisomy 21. *Mol. Neurodegener.* 9:15. doi: 10.1186/1750-1326-9-15

Hara, N., Kikuchi, M., Miyashita, A., Hatsuta, H., Saito, Y., Kasuga, K., et al. (2017). Serum microRNA miR-501-3p as a potential biomarker related to the progression of Alzheimer's disease. *Acta Neuropathol. Commun.* 5:10. doi: 10.1186/s40478-017-0414-z

Hebert, S. S., Horre, K., Nicolai, L., Papadopoulou, A. S., Mandemakers, W., Silahtaroglu, A. N., et al. (2008). Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6415–6420. doi: 10.1073/pnas.0710263105

Heemels, M. T. (2016). Neurodegenerative diseases. *Nature* 539:179. doi: 10.1038/539179a

Hoss, A. G., Labadorf, A., Beach, T. G., Latourelle, J. C., and Myers, R. H. (2016). microRNA profiles in Parkinson's disease prefrontal cortex. *Front. Aging Neurosci.* 8:36. doi: 10.3389/fnagi.2016.00036

Hoss, A. G., Labadorf, A., Latourelle, J. C., Kartha, V. K., Hadzi, T. C., Gusella, J. F., et al. (2015a). miR-10b-5p expression in Huntington's disease brain relates to age of onset and the extent of striatal involvement. *BMC Med. Genomics* 8:10. doi: 10.1186/s12920-015-0083-3

Hoss, A. G., Lagomarsino, V. N., Frank, S., Hadzi, T. C., Myers, R. H., and Latourelle, J. C. (2015b). Study of plasma-derived miRNAs mimic differences in Huntington's disease brain. *Mov. Disord.* 30, 1961–1964. doi: 10.1002/mds.26457

Ishihara, T., Ariizumi, Y., Shiga, A., Kato, T., Tan, C. F., Sato, T., et al. (2013). Decreased number of Gemini of coiled bodies and U12 snRNA level in amyotrophic lateral sclerosis. *Hum. Mol. Genet.* 22, 4136–4147. doi: 10.1093/hmg/ddt262

Jack, C. R. Jr., Albert, M. S., Knopman, D. S., McKhann, G. M., Sperling, R. A., Carrillo, M. C., et al. (2011). Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 257–262. doi: 10.1016/j.jalz.2011.03.004

Jia, L. H., and Liu, Y. N. (2016). Downregulated serum miR-223 servers as biomarker in Alzheimer's disease. *Cell Biochem. Funct.* 34, 233–237. doi: 10.1002/cbf.3184

Jia, Y., Mu, J. C., and Ackerman, S. L. (2012). Mutation of a U2 snRNA gene causes global disruption of alternative splicing and neurodegeneration. *Cell* 148, 296–308. doi: 10.1016/j.cell.2011.11.057

Jutzi, D., Akinyi, M. V., Mechtersheimer, J., Frilander, M. J., and Ruepp, M.-D. (2018). The emerging role of minor intron splicing in neurological disorders. *Cell Stress* 2, 40–54. doi: 10.15698/cst2018.03.126

Karaca, E., Weitzer, S., Pehlivan, D., Shiraishi, H., Gogakos, T., Hanada, T., et al. (2014). Human CLP1 mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function. *Cell* 157, 636–650. doi: 10.1016/j.cell.2014.02.058

Kawahara, Y., and Mieda-Sato, A. (2012). TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3347–3352. doi: 10.1073/pnas.1112427109

Khoo, S. K., Petillo, D., Kang, U. J., Resau, J. H., Berryhill, B., Linder, J., et al. (2012). Plasma-based circulating MicroRNA biomarkers for Parkinson's disease. *J. Parkinsons Dis.* 2, 321–331. doi: 10.3233/jpd-012144

Kieran, D., Sebastia, J., Greenway, M. J., King, M. A., Connaughton, D., Concannon, C. G., et al. (2008). Control of motoneuron survival by angiogenin. *J. Neurosci.* 28, 14056–14061. doi: 10.1523/jneurosci.3399-08.2008

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. Alzheimers Dis.* 39, 253–259. doi: 10.3233/jad-130932

Kirchner, S., and Ignatova, Z. (2015). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* 16, 98–112. doi: 10.1038/nrg3861

Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P. J., Stefan, M., et al. (2010). The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum. Mol. Genet.* 19, 1153–1164. doi: 10.1093/hmg/ddp585

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:e69807. doi: 10.1371/journal.pone.0069807

Kumar, P., Mudunuri, S. B., Anaya, J., and Dutta, A. (2015). tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res.* 43, D141–D145. doi: 10.1093/nar/gku1138

Kumar, S., Vijayan, M., and Reddy, P. H. (2017). MicroRNA-455-3p as a potential peripheral biomarker for Alzheimer's disease. *Hum. Mol. Genet.* 26, 3808–3822. doi: 10.1093/hmg/ddx267

Laufer, B. I., Mantha, K., Kleiber, M. L., Diehl, E. J., Addison, S. M., and Singh, S. M. (2013). Long-lasting alterations to DNA methylation and ncRNAs could underlie the effects of fetal alcohol exposure in mice. *Dis. Model. Mech.* 6, 977–992. doi: 10.1242/dmm.010975

Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., et al. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 27, 390–399. doi: 10.1101/gad.209841.112

Lee, Y. S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* 23, 2639–2649. doi: 10.1101/gad.1837609

Lehmann, S. M., Kruger, 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, 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:R78. doi: 10.1186/gb-2013-14-7-r78

Li, S., and Hu, G. F. (2012). Emerging role of angiogenin in stress response and cell survival under adverse conditions. *J. Cell. Physiol.* 227, 2822–2826. doi: 10.1002/jcp.23051

Liao, J. Y., Ma, L. M., Guo, Y. H., Zhang, Y. C., Zhou, H., Shao, P., et al. (2010). Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3′ trailers. *PLoS One* 5:e10563. doi: 10.1371/journal.pone.0010563

Liao, X. Y., Wang, W. W., Yang, Z. H., Wang, J., Lin, H., Wang, Q. S., et al. (2013). Microarray analysis of transcriptome of medulla identifies potential biomarkers for Parkinson's disease. *Int. J. Genomics* 2013:606919. doi: 10.1155/2013/606919

Liguori, M., Nuzziello, N., Introna, A., Consiglio, A., Licciulli, F., D'Errico, E., et al. (2018). Dysregulation of MicroRNAs and target genes networks in peripheral blood of patients with sporadic amyotrophic lateral sclerosis. *Front. Mol. Neurosci.* 11:288. doi: 10.3389/fnmol.2018.00288

Loche, E., and Ozanne, S. E. (2016). Early nutrition, epigenetics, and cardiovascular disease. *Curr. Opin. Lipidol.* 27, 449–458. doi: 10.1097/mol.0000000000000338

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:e0139233. doi: 10.1371/journal.pone.0139233

Lukiw, W. J. (2007). Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus. *Neuroreport* 18, 297–300. doi: 10.1097/WNR.0b013e3280148e8b

Lukiw, W. J., Alexandrov, P. N., Zhao, Y., Hill, J. M., and Bhattacharjee, S. (2012). Spreading of Alzheimer's disease inflammatory signaling through soluble micro-RNA. *Neuroreport* 23, 621–626. doi: 10.1097/WNR.0b013e32835542b0

Ma, W., Li, Y., Wang, C., Xu, F., Wang, M., and Liu, Y. (2016). Serum miR-221 serves as a biomarker for Parkinson's disease. *Cell Biochem. Funct.* 34, 511–515. doi: 10.1002/cbf.3224

Mani, S. R., and Juliano, C. E. (2013). Untangling the web: the diverse functions of the PIWI/piRNA pathway. *Mol. Reprod. Dev.* 80, 632–664. doi: 10.1002/mrd.22195

Maraia, R. J., and Lamichhane, T. N. (2011). 3′ processing of eukaryotic precursor tRNAs. *Wiley Interdiscip. Rev. RNA* 2, 362–375. doi: 10.1002/wrna.64

Margis, R., Margis, R., and Rieder, C. R. (2011). Identification of blood microRNAs associated to Parkinsonis disease. *J. Biotechnol.* 152, 96–101. doi: 10.1016/j.jbiotec.2011.01.023

Martinez, F. J., Lee, J. H., Lee, J. E., Blanco, S., Nickerson, E., Gabriel, S., et al. (2012). Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. *J. Med. Genet.* 49, 380–385. doi: 10.1136/jmedgenet-2011-100686

Martins, M., Rosa, A., Guedes, L. C., Fonseca, B. V., Gotovac, K., Violante, S., et al. (2011). Convergence of miRNA expression profiling, alpha-synuclein interaction and GWAS in Parkinson's disease. *PLoS One* 6:e25443. doi: 10.1371/journal.pone.0025443

Massenet, S., Bertrand, E., and Verheggen, C. (2017). Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol.* 14, 680–692. doi: 10.1080/15476286.2016.1243646

Matamala, J. M., Arias-Carrasco, R., Sanchez, C., Uhrig, M., Bargsted, L., Matus, S., et al. (2018). Genome-wide circulating microRNA expression profiling reveals potential biomarkers for amyotrophic lateral sclerosis. *Neurobiol. Aging* 64, 123–138. doi: 10.1016/j.neurobiolaging.2017.12.020

Matera, A. G., Terns, R. M., and Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.* 8, 209–220. doi: 10.1038/nrm2124

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. Jr., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005

Minones-Moyano, E., Porta, S., Escaramis, G., Rabionet, R., Iraola, S., Kagerbauer, B., et al. (2011). MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function. *Hum. Mol. Genet.* 20, 3067–3078. doi: 10.1093/hmg/ddr210

Munoz-Culla, M., Irizar, H., Saenz-Cuesta, M., Castillo-Trivino, T., Osorio-Querejeta, I., Sepulveda, L., et al. (2016). SncRNA (microRNA &snoRNA) opposite expression pattern found in multiple sclerosis relapse and remission is sex dependent. *Sci. Rep.* 6:20126. doi: 10.1038/srep20126

Nagaraj, S., Laskowska-Kaszub, K., Debski, K. J., Wojsiat, J., Dabrowski, 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, 16122–16143. doi: 10.18632/oncotarget.15109

Nagatsua, T., and Sawadab, M. (2009). L-dopa therapy for Parkinson's disease: past, present, and future. *Parkinsonism Relat. Disord.* 15(Suppl. 1), S3–S8. doi: 10.1016/s1353-8020(09)70004-5

O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol.* 9:402. doi: 10.3389/fendo.2018.00402

Ohtani, M. (2017). Transcriptional regulation of snRNAs and its significance for plant development. *J. Plant Res.* 130, 57–66. doi: 10.1007/s10265-016-0883-3

Packer, A. N., Xing, Y., Harper, S. Q., Jones, L., and Davidson, B. L. (2008). The bifunctional microRNA miR-9/miR-9* regulates REST and CoREST and is downregulated in Huntington's disease. *J. Neurosci.* 28, 14341–14346. doi: 10.1523/jneurosci.2390-08.2008

Palazzo, A. F., and Lee, E. S. (2015). Non-coding RNA: what is functional and what is junk? *Front. Genet.* 6:2. doi: 10.3389/fgene.2015.00002

Rachakonda, V., Pan, T. H., and Le, W. D. (2004). Biomarkers of neurodegenerative disorders: how good are they? *Cell Res.* 14, 347–358. doi: 10.1038/sj.cr.7290235

Raheja, R., Regev, K., Healy, B. C., Mazzola, M. A., Beynon, V., Von Glehn, F., et al. (2018). Correlating serum microRNAs and clinical parameters in amyotrophic lateral sclerosis. *Muscle Nerve* 58, 261–269. doi: 10.1002/mus.26106

Ratti, A., and Buratti, E. (2016). Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. *J. Neurochem.* 138(Suppl. 1), 95–111. doi: 10.1111/jnc.13625

Reber, S., Stettler, J., Filosa, G., Colombo, M., Jutzi, D., Lenzken, S. C., et al. (2016). Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants. *EMBO J.* 35, 1504–1521. doi: 10.15252/embj.201593791

Reed, E. R., Latourelle, J. C., Bockholt, J. H., Bregu, J., Smock, J., Paulsen, J. S., et al. (2018). MicroRNAs in CSF as prodromal biomarkers for Huntington disease in the PREDICT-HD study. *Neurology* 90, e264–e272. doi: 10.1212/wnl.0000000000004844

Ricci, C., Marzocchi, C., and Battistini, S. (2018). MicroRNAs as biomarkers in amyotrophic lateral sclerosis. *Cells* 7:E219. doi: 10.3390/cells7110219

Rizzo, F., Rinaldi, A., Marchese, G., Coviello, E., Sellitto, A., Cordella, A., et al. (2016). Specific patterns of PIWI-interacting small noncoding RNA expression in dysplastic liver nodules and hepatocellular carcinoma. *Oncotarget* 7, 54650–54661. doi: 10.18632/oncotarget.10567

Roy, J., Sarkar, A., Parida, S., Ghosh, Z., and Mallick, B. (2017). Small RNA sequencing revealed dysregulated piRNAs in Alzheimer's disease and their probable role in pathogenesis. *Mol. Biosyst.* 13, 565–576. doi: 10.1039/c6mb00699j

Sala Frigerio, C., 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, 2103–2106. doi: 10.1212/01.wnl.0000437306.37850.22

Schaffer, A. E., Eggens, V. R., Caglayan, A. O., Reuter, M. S., Scott, E., Coufal, N. G., et al. (2014). CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. *Cell* 157, 651–663. doi: 10.1016/j.cell.2014.03.049

Schipper, H. M., Maes, O. C., Chertkow, H. M., and Wang, E. (2007). MicroRNA expression in Alzheimer blood mononuclear cells. *Gene Regul. Syst. Bio.* 1, 263–274.

Schulze, M., Sommer, A., Plotz, S., Farrell, M., Winner, B., Grosch, J., et al. (2018). Sporadic Parkinson's disease derived neuronal cells show disease-specific mRNA and small RNA signatures with abundant deregulation of piRNAs. *Acta Neuropathol. Commun.* 6:58. doi: 10.1186/s40478-018-0561-x

Serafin, A., Foco, L., Blankenburg, H., Picard, A., Zanigni, S., Zanon, A., et al. (2014). Identification of a set of endogenous reference genes for miRNA expression studies in Parkinson's disease blood samples. *BMC Res. Notes* 7:715. doi: 10.1186/1756-0500-7-715

Serafin, A., Foco, L., Zanigni, S., Blankenburg, H., Picard, A., Zanon, A., et al. (2015). Overexpression of blood microRNAs 103a, 30b, and 29a in

L-dopa-treated patients with PD. *Neurology* 84, 645–653. doi: 10.1212/wnl.0000000000001258

Sharma, E., Sterne-Weiler, T. O., Hanlon, D., and Blencowe, B. J. (2016). Global mapping of human RNA-RNA interactions. *Mol. Cell* 62, 618–626. doi: 10.1016/j.molcel.2016.04.030

Sheinerman, K. S., Toledo, J. B., Tsivinsky, V. G., Irwin, D., Grossman, M., Weintraub, D., et al. (2017). Circulating brain-enriched microRNAs as novel biomarkers for detection and differentiation of neurodegenerative diseases. *Alzheimers Res. Ther.* 9:89. doi: 10.1186/s13195-017-0316-0

Sheinerman, K. S., and Umansky, S. R. (2013). Early detection of neurodegenerative diseases: circulating brain-enriched microRNA. *Cell Cycle* 12, 1–2. doi: 10.4161/cc.23067

Shioya, M., Obayashi, S., Tabunoki, H., Arima, K., Saito, Y., Ishida, T., et al. (2010). Aberrant microRNA expression in the brains of neurodegenerative diseases: miR-29a decreased in Alzheimer disease brains targets neurone navigator 3. *Neuropathol. Appl. Neurobiol.* 36, 320–330. doi: 10.1111/j.1365-2990.2010.01076.x

Sorensen, S. S., Nygaard, A. B., and Christensen, T. (2016). miRNA expression profiles in cerebrospinal fluid and blood of patients with Alzheimer's disease and other types of dementia - an exploratory study. *Transl. Neurodegener.* 5:6. doi: 10.1186/s40035-016-0053-5

Soreq, L., Salomonis, N., Bronstein, M., Greenberg, D. S., Israel, Z., Bergman, H., et al. (2013). Small RNA sequencing-microarray analyses in Parkinson leukocytes reveal deep brain stimulation-induced splicing changes that classify brain region transcriptomes. *Front. Mol. Neurosci.* 6:10. doi: 10.3389/fnmol.2013.00010

Spano, M., Signorelli, M., Vitaliani, R., Aguglia, E., and Giometto, B. (2015). The possible involvement of mitochondrial dysfunctions in Lewy body dementia: a systematic review. *Funct. Neurol.* 30, 151–158.

Stamova, B., Ander, B. P., Barger, N., Sharp, F. R., and Schumann, C. M. (2015). Specific regional and age-related small noncoding RNA expression patterns within superior temporal gyrus of typical human brains are less distinct in autism brains. *J. Child Neurol.* 30, 1930–1946. doi: 10.1177/0883073815602067

Sun, C., Yang, F., Zhang, Y., Chu, J., Wang, J., Wang, Y., et al. (2018). tRNA-derived fragments as novel predictive biomarkers for trastuzumab-resistant breast cancer. *Cell. Physiol. Biochem.* 49, 419–431. doi: 10.1159/000492977

Swarbrick, S., Wragg, N., Ghosh, S., and Stolzing, A. (2019). Systematic review of miRNA as biomarkers in Alzheimer's disease. *Mol. Neurobiol.* doi: 10.1007/s12035-019-1500-y

Takahashi, I., Hama, Y., Matsushima, M., Hirotani, M., Kano, T., Hohzen, H., et al. (2015). Identification of plasma microRNAs as a biomarker of sporadic Amyotrophic Lateral Sclerosis. *Mol. Brain* 8:67. doi: 10.1186/s13041-015-0161-7

Tan, L., Yu, J. T., Liu, Q. Y., Tan, M. S., Zhang, W., Hu, N., et al. (2014a). Circulating miR-125b as a biomarker of Alzheimer's disease. *J. Neurol. Sci.* 336, 52–56. doi: 10.1016/j.jns.2013.10.002

Tan, L., Yu, J. T., Tan, M. S., Liu, Q. Y., Wang, H. F., Zhang, W., et al. (2014b). Genome-wide serum microRNA expression profiling identifies serum biomarkers for Alzheimer's disease. *J. Alzheimers Dis.* 40, 1017–1027. doi: 10.3233/jad-132144

Tasca, E., Pegoraro, V., Merico, A., and Angelini, C. (2016). Circulating microRNAs as biomarkers of muscle differentiation and atrophy in ALS. *Clin. Neuropathol.* 35, 22–30. doi: 10.5414/np300889

Tatura, R., Kraus, T., Giese, A., Arzberger, T., Buchholz, M., Hoglinger, G., et al. (2016). Parkinson's disease: SNCA-, PARK2-, and LRRK2- targeting microRNAs elevated in cingulate gyrus. *Parkinsonism Relat. Disord.* 33, 115–121. doi: 10.1016/j.parkreldis.2016.09.028

Toivonen, J. M., Manzano, R., Olivan, S., Zaragoza, P., Garcia-Redondo, A., and Osta, R. (2014). MicroRNA-206: a potential circulating biomarker candidate for amyotrophic lateral sclerosis. *PLoS One* 9:e89065. doi: 10.1371/journal.pone.0089065

Tollervey, D., and Hurt, E. C. (1990). The role of small nucleolar ribonucleoproteins in ribosome synthesis. *Mol. Biol. Rep.* 14, 103–106.

Toth, K. F., Pezic, D., Stuwe, E., and Webster, A. (2016). The piRNA pathway guards the germline genome against transposable elements. *Adv. Exp. Med. Biol.* 886, 51–77. doi: 10.1007/978-94-017-7417-8_4

Tsuiji, H., Iguchi, Y., Furuya, A., Kataoka, A., Hatsuta, H., Atsuta, N., et al. (2013). Spliceosome integrity is defective in the motor neuron diseases

ALS and SMA. *EMBO Mol. Med.* 5, 221–234. doi: 10.1002/emmm.201202303

Tycowski, K. T., You, Z. H., Graham, P. J., and Steitz, J. A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell* 2, 629–638.

Vagnozzi, R., Signoretti, S., Tavazzi, B., Floris, R., Ludovici, A., Marziali, S., et al. (2008). Temporal window of metabolic brain vulnerability to concussion: a pilot 1H-magnetic resonance spectroscopic study in concussed athletes–part III. *Neurosurgery* 62, 1286–1295; discussion 1295–1296. doi: 10.1227/01.neu.0000333300.34189.74

Vallelunga, A., Ragusa, M., Di Mauro, S., Iannitti, T., Pilleri, M., Biundo, R., et al. (2014). Identification of circulating microRNAs for the differential diagnosis of Parkinson's disease and Multiple System Atrophy. *Front. Cell. Neurosci.* 8:156. doi: 10.3389/fncel.2014.00156

van Es, M. A., Schelhaas, H. J., van Vught, P. W., Ticozzi, N., Andersen, P. M., Groen, E. J., et al. (2011). Angiogenin variants in Parkinson disease and amyotrophic lateral sclerosis. *Ann. Neurol.* 70, 964–973. doi: 10.1002/ana.22611

Vazquez-Arango, P., Vowles, J., Browne, C., Hartfield, E., Fernandes, H. J., 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

Verma, B., Akinyi, M. V., Norppa, A. J., and Frilander, M. J. (2018). Minor spliceosome and disease. *Semin. Cell Dev. Biol.* 79, 103–112. doi: 10.1016/j.semcdb.2017.09.036

Villa, C., Ridolfi, E., Fenoglio, C., Ghezzi, L., Vimercati, R., Clerici, F., et al. (2013). Expression of the transcription factor Sp1 and its regulatory hsa-miR-29b in peripheral blood mononuclear cells from patients with Alzheimer's disease. *J. Alzheimers Dis.* 35, 487–494. doi: 10.3233/jad-122263

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., et al. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474, 380–384. doi: 10.1038/nature10110

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

Wake, C., Labadorf, A., Dumitriu, A., Hoss, A. G., Bregu, J., Albrecht, K. H., et al. (2016). Novel microRNA discovery using small RNA sequencing in post-mortem human brain. *BMC Genomics* 17:776. doi: 10.1186/s12864-016-3114-3

Waller, R., Goodall, E. F., Milo, M., Cooper-Knock, J., Da Costa, M., Hobson, E., et al. (2017a). Serum miRNAs miR-206, 143-3p and 374b-5p as potential biomarkers for amyotrophic lateral sclerosis (ALS). *Neurobiol. Aging* 55, 123–131. doi: 10.1016/j.neurobiolaging.2017.03.027

Waller, R., Wyles, M., Heath, P. R., Kazoka, M., Wollff, H., Shaw, P. J., et al. (2017b). Small RNA sequencing of sporadic amyotrophic lateral sclerosis cerebrospinal fluid reveals differentially expressed miRNAs related to neural and glial activity. *Front. Neurosci.* 11:731. doi: 10.3389/fnins.2017.00731

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 beta-site amyloid precursor protein-cleaving enzyme 1. *J. Neurosci.* 28, 1213–1223. doi: 10.1523/jneurosci.5065-07.2008

Wright, C., Shin, J. H., Rajpurohit, A., Deep-Soboslay, A., Collado-Torres, L., Brandon, N. J., et al. (2017). Altered expression of histamine signaling genes in autism spectrum disorder. *Transl. Psychiatry* 7:e1126. doi: 10.1038/tp.2017.87

Wu, Y., Xu, J., Xu, J., Cheng, J., Jiao, D., Zhou, C., et al. (2017). Lower serum levels of miR-29c-3p and miR-19b-3p as biomarkers for Alzheimer's disease. *Tohoku J. Exp. Med.* 242, 129–136. doi: 10.1620/tjem.242.129

Yamasaki, S., Ivanov, P., Hu, G. F., and Anderson, P. (2009). Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J. Cell Biol.* 185, 35–42. doi: 10.1083/jcb.200811106

Yan, Z., Hu, H. Y., Jiang, X., Maierhofer, V., Neb, E., He, L., et al. (2011). Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.* 39, 6596–6607. doi: 10.1093/nar/gkr298

Yang, X., Cheng, Y., Lu, Q., Wei, J., Yang, H., and Gu, M. (2015). Detection of stably expressed piRNAs in human blood. *Int. J. Clin. Exp. Med.* 8, 13353–13358.

Yilmaz, S. G., Erdal, M. E., Ozge, A. A., and Sungur, M. A. (2016a). Can peripheral MicroRNA expression data serve as epigenomic (Upstream) biomarkers of Alzheimer's disease? *OMICS* 20, 456–461. doi: 10.1089/omi.2016.0099

Yilmaz, S. G., Geyik, S., Neyal, A. M., Soko, N. D., Bozkurt, H., and Dandara, C. (2016b). Hypothesis: do miRNAs targeting the leucine-rich repeat kinase 2 gene (LRRK2) influence Parkinson's disease susceptibility? *OMICS* 20, 224–228. doi: 10.1089/omi.2016.0040

Yin, S., Lopez-Gonzalez, R., Kunz, R. C., Gangopadhyay, J., Borufka, C., Gygi, S. P., et al. (2017). Evidence that C9ORF72 dipeptide repeat proteins associate with U2 snRNP to cause Mis-splicing in ALS/FTD Patients. *Cell Rep.* 19, 2244–2256. doi: 10.1016/j.celrep.2017.05.056

Zeng, Q., Zou, L., Qian, L., Zhou, F., Nie, H., Yu, S., et al. (2017). Expression of microRNA222 in serum of patients with Alzheimer's disease. *Mol. Med. Rep.* 16, 5575–5579. doi: 10.3892/mmr.2017.7301

Zhang, T., Wu, Y. C., Mullane, P., Ji, Y. J., Liu, H., He, L., et al. (2018). FUS regulates activity of MicroRNA-mediated gene silencing. *Mol. Cell* 69, 787–801.e8. doi: 10.1016/j.molcel.2018.02.001

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., et al. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* 133, 585–600. doi: 10.1016/j.cell.2008.03.031

Zhang, Z., Pinto, A. M., Wan, L., Wang, W., Berg, M. G., Oliva, I., et al. (2013). Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19348–19353. doi: 10.1073/pnas.1319280110

Zhao, J., Yue, D., Zhou, Y., Jia, L., Wang, H., Guo, M., et al. (2017). The role of MicroRNAs in Abeta deposition and tau phosphorylation in Alzheimer's disease. *Front. Neurol.* 8:342. doi: 10.3389/fneur.2017.00342

Zhao, N., Jin, L., Fei, G., Zheng, Z., and Zhong, C. (2014). Serum microRNA-133b is associated with low ceruloplasmin levels in Parkinson's disease. *Parkinsonism Relat. Disord.* 20, 1177–1180. doi: 10.1016/j.parkreldis.2014.08.016

Systematic Analysis of Non-coding RNAs Involved in the Angora Rabbit (*Oryctolagus cuniculus*) Hair Follicle Cycle by RNA Sequencing

Bohao Zhao[1], Yang Chen[1,2], Shuaishuai Hu[1], Naisu Yang[1], Manman Wang[2], Ming Liu[1], Jiali Li[1], Yeyi Xiao[1] and Xinsheng Wu[1,2]*

[1] College of Animal Science and Technology, Yangzhou University, Yangzhou, China, [2] Joint International Research Laboratory of Agriculture and Agri-Product Safety, Yangzhou University, Yangzhou, China

The hair follicle (HF) cycle is a complicated and dynamic process in mammals, associated with various signaling pathways and gene expression patterns. Non-coding RNAs (ncRNAs) are RNA molecules that are not translated into proteins but are involved in the regulation of various cellular and biological processes. This study explored the relationship between ncRNAs and the HF cycle by developing a synchronization model in Angora rabbits. Transcriptome analysis was performed to investigate ncRNAs and mRNAs associated with the various stages of the HF cycle. One hundred and eleven long non-coding RNAs (lncRNAs), 247 circular RNAs (circRNAs), 97 microRNAs (miRNAs), and 1,168 mRNAs were differentially expressed during the three HF growth stages. Quantitative real-time PCR was used to validate the ncRNA transcriptome analysis results. Gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses provided information on the possible roles of ncRNAs and mRNAs during the HF cycle. In addition, lncRNA–miRNA–mRNA and circRNA–miRNA–mRNA ceRNA networks were constructed to investigate the underlying relationships between ncRNAs and mRNAs. LNC_002919 and novel_circ_0026326 were found to act as ceRNAs and participated in the regulation of the HF cycle as miR-320-3p sponges. This research comprehensively identified candidate regulatory ncRNAs during the HF cycle by transcriptome analysis, highlighting the possible association between ncRNAs and the regulation of hair growth. This study provides a basis for systematic further research and new insights on the regulation of the HF cycle.

Keywords: rabbit, non-coding RNA, sequencing, hair follicle cycle, ceRNA

INTRODUCTION

Hair follicle (HF) development is a complex morphogenetic process that relies on a variety of signaling systems, and on interactions between mesenchymal and epithelial tissues (Hardy, 1992; Oro and Scott, 1998). Under the biological regulation of stem cells, mature HFs undergo a cycling and continuous self-renewal process, with periods of active growth (anagen), followed by regression

(catagen), and rest (telogen) (Cotsarelis et al., 1990; Paus and Cotsarelis, 1999; Fuchs and Segre, 2000; Oshima et al., 2001). In murine HF cycling, key parameters for the recognition of distinct stages have been defined in many studies (Chase et al., 1951; Chase, 1954; Straile et al., 2010). Moreover, the immediate removal of hair shafts could induce homogeneous anagen development in the murine model, which leads to the spontaneous entering of consecutive stages (catagen and telogen). In this way, the methods for the analysis of murine HF growth were provided, and were based on histologic and ultrastructural studies on murine hair cycling (Veen et al., 1999; Müller-Röver et al., 2001). During the anagen phase, the hair root is dividing and adding to the hair shaft. The HFs actively grow, surrounded by dermal fibroblasts that have not reached the subcutis. During the catagen phase, interfollicular dermal fibroblasts fully surrounded the HFs, the blood supply is cut off, and the hair bulb starts to atrophy. Finally, HFs enter the telogen phase, where hair shafts stop growing, and begin to fall due to synthesis and release of hair cycle inhibitor (Stenn and Paus, 2001). The molecular mechanisms underlying the regulation of the hair cycle and of HF development are of interest in medicine and developmental biology (Shirokova et al., 2016; Ahmed et al., 2017; Sardella et al., 2017).

Long non-coding RNAs (lncRNAs), microRNAs (miRNAs), and circular RNAs (circRNAs) are non-coding RNA (ncRNA) that are not translated into proteins but regulate many cell functions and play vital roles in many biological processes (Mattick and Makunin, 2006; Guttman and Rinn, 2012). miRNAs are small ncRNA molecules (∼22 nucleotides length) that repress gene expression by recognizing specific target mRNAs (Ding et al., 2009). An increasing number of studies reported that lncRNAs (non-coding RNAs containing more than 200 base pairs) regulate interactions between genes and proteins, act as decoys that bind to miRNAs or proteins, or bind to enhancer regions or neighboring loci to modulate the transcription of their target gene as enhancers (Winkle et al., 2015; Chen et al., 2016; Li et al., 2016; Song et al., 2017; Lu et al., 2018). CircRNAs consist of continuous loop structures, are more stable than linear mRNAs, and are conserved between different species (Stoffelen et al., 2012; Memczak et al., 2014). As sponges for miRNAs, circRNAs act as competitive inhibitors that interfere with the binding of miRNAs to their target genes (Hansen et al., 2013; Zhong Z. et al., 2016). circRNAs may also regulate the function of RNA-binding proteins and the transcription activity of the host gene (Reut et al., 2014; Li et al., 2015). Although circRNAs have been categorized as ncRNA, they have been reported to have the ability to code proteins as gene regulators (Pamudurti et al., 2017).

Accumulating evidence suggests that lncRNAs are involved in the regulation of the HF cycle (Wang et al., 2017; Song et al., 2018; Zhu Y.B. et al., 2018). Specific lncRNAs, such as HOTAIR, H19, and RP11-766N7.3, have been reported to be differentially expressed in dermal papilla cells after Wnt signaling by using lncRNA microarrays, and integrated analysis by RNA-seq techniques has led to the identification of potential lncRNA, which may play a role during the initiation of secondary HFs (Lin et al., 2015; Yue et al., 2016). Moreover, aberrantly expressed miRNAs may participate in the regulation of the

development of skin and HFs. miRNAs play important roles in several signaling pathways and control gene expression patterns during the HF cycle (Mardaryev et al., 2010; Chao et al., 2013; Ahmed et al., 2014; Zhou et al., 2018). In addition, the expression levels and functions of circRNAs associated with skin color during different skin differentiation stages have been analyzed by RNA-seq (Zhu Z. et al., 2018).

However, only very few studies have systematically investigated ncRNAs during the HF cycle. This study established a HF cycle synchronization model in the rabbit, allowing an integrated analysis of ncRNAs and mRNAs expressed during the different HF cycle phases (anagen, catagen, and telogen). Numerous essential factors related to the HF cycle have been uncovered, contributing to the understanding of HF cycle regulation and suggesting new potential therapies for hair-related diseases.

MATERIALS AND METHODS

Animals

Twelve 6-month-old male Wanxi Angora rabbits were used to establish the HF synchronization model. They were all housed under the same conditions, including temperature, and were fed the same diet (feed pellet and grass). Animals were reared in a controlled environment and had the same length of the hair coat phenotypes. The experimental procedures in this study were approved by the Animal Care and Use Committee of Yangzhou University.

To estimate the wool growth rate and to determine the onset of the anagen phase, the dorsal area of experimental animals was shaved with electronic clippers and entry into anagen was determined by the appearance of light pink skin and by hair regrowth. The length of the hair coat was measured, skin samples were collected after shaving, samples were fixed in 4% formaldehyde, and paraffin sections were stained with hematoxylin–eosin (HE) for histological observations. Longitudinal sections of the HFs showed the skin status and the phase of the HF cycle.

Tissue Collection

Rabbits were anesthetized via ear vein injections of 0.7% pentobarbital sodium (6 mL/kg), dorsal skin samples (1 cm^2) were collected, and placed immediately in liquid nitrogen for RNA extraction. Iodine solution was applied on the wound to prevent bacterial infection. Samples were harvested at different phases of the HF cycle for gene expression profiling: growth (anagen), cessation (catagen), and rest (telogen). Three sample replicates were collected at days 90, 130, and 150 of the HF cycle for ncRNA and mRNA sequencing analysis.

RNA Isolation and RNA Quantification

Total RNA from nine samples was extracted from skin tissue using Trizol reagent (Invitrogen, Carlsbad, CA, United States), according to the manufacturer's instructions. RNA degradation and contamination were monitored by running samples on 1% agarose gels. RNA purity was analyzed via a NanoPhotometer®

spectrophotometer (IMPLEN, CA, United States). RNA concentration was measured using the Qubit® RNA Assay Kit and a Qubit® 2.0 Fluorometer (Life Technologies, CA, United States). RNA integrity was assessed via the RNA Nano 6000 Assay Kit and a Bioanalyzer 2100 system (Agilent Technologies, CA, United States). lncRNAs and miRNAs were quantified following the same procedure used for conventional mRNAs. Quantification of circRNAs was performed adding an exonuclease to degrade non-circRNAs. Briefly, two samples containing the same amount of RNA were collected. In one sample, linear RNA was digested with RNase R (Cat. No. RNR07250, Epicentre Company, United States), leaving only the circRNAs, while the other sample was not treated with RNase R. The two RNA samples were reverse transcribed. The samples subjected to RNase treatment were used to detect circRNAs, whereas the untreated samples were used to detect β-actin.

Library Construction for lncRNA and circRNA Sequencing

A total amount of 3 μg of RNA per sample was used for lncRNA sequencing and of 5 μg for circRNA sequencing. First, ribosomal RNAs were removed with the Epicentre Ribo-zero™ rRNA Removal Kit (Epicentre, United States) and the rRNA-depleted samples were purified by ethanol precipitation. Subsequently, sequencing libraries were generated using the rRNA-depleted RNA and the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, United States), following the manufacturer's recommendations. First strand cDNA was synthesized using random hexamer primers and M-MuLV Reverse Transcriptase (RNaseH). Second strand cDNA synthesis was performed using DNA Polymerase I and RNase H. After adenylation of the 3′ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. To select cDNA fragments with a preferential length of 150∼200 bp, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, MA, United States). Then, 3 μl of USER Enzyme (NEB, United States) was used with size-selected, adaptor-ligated cDNA before the PCR. Finally, the PCR products were purified (AMPure XP system) and the library quality was assessed with the Agilent Bioanalyzer 2100 system.

Library Construction for Small RNA Sequencing

A total amount of 3 μg of RNA per sample was used as input material for the small RNA library. Sequencing libraries were generated using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB, United States), following the manufacturer's recommendations. Briefly, NEB 3′ SR Adaptor was directly and specifically ligated to the 3′ end of miRNAs, siRNAs, and piRNAs. After the 3′ ligation reaction, the SR RT Primer was hybridized to the excess of 3′ SR Adaptor, transforming the single-stranded DNA adaptor into a double-stranded DNA molecule. Then, the 5′ ends adapter was ligated to the 5′ ends of the miRNAs, siRNAs, and piRNAs. First strand cDNA was synthesized using M-MuLV Reverse Transcriptase (RNase H–). DNA fragments of 140–160 bp length (the length of small non-coding RNAs plus the 3′ and 5′ adaptors) were recovered and dissolved in 8 μL of elution buffer. Finally, the library quality was assessed using the Agilent Bioanalyzer 2100 system and DNA High Sensitivity Chips.

Clustering and Sequencing of lncRNAs, circRNAs, and miRNAs

Clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina), according to the manufacturer's instructions. After cluster generation, the lncRNA and circRNA libraries were sequenced on an Illumina Hiseq 4000 platform and 150 bp paired end reads were generated. The miRNA library was sequenced on an Illumina Hiseq 2500 platform and 50 bp single-end reads were generated.

Quality Control

For lncRNA and circRNA sequencing, raw data (raw reads) in fastq format were first processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N, and low-quality reads from the raw data. For miRNA sequencing, raw data (raw reads) in fastq format were first processed through custom perl and python scripts. In this step, clean data (clean reads) were obtained by removing reads containing ploy-N, with 5′ adapter contaminants, without 3′ adapter or the insert tag, containing ploy A or T or G or C, and low-quality reads from raw data. At the same time, the Q20, Q30 scores, and GC-content of the raw data were calculated. A specific length range from the clean reads was selected to conduct all the downstream analyses, based on clean data of high quality.

Genome Mapping, Transcriptome Assembly, and ncRNAs Identification

For lncRNA and circRNA sequences, the reference genome (*Oryctolagus cuniculus* genome obtained from Ensembl OryCun2.0) and annotation files were directly downloaded from the genome website. An index of the reference genome was built using bowtie2 (Langmead and Salzberg, 2012), and paired-end clean reads were aligned to the reference genome using HISAT2 v2.0.4 (Pertea et al., 2016). Also, the small RNA tags were mapped to the reference sequence with bowtie2 (Langmead and Salzberg, 2012) without mismatch to analyze the expression and distribution of miRNA sequences in the reference genome.

The mapped lncRNA and mRNA reads from each sample were assembled by means of StringTie (v1.3.1) (Pertea et al., 2016), following a reference-based approach. The circRNAs were detected and identified using find_circ (Memczak et al., 2014). Alignment of the small RNA tags to miRBase20.0 identified known *Oryctolagus cuniculus* and *Mus musculus* (near-source species) miRNAs. Mirdeep2 software (Friedländer et al., 2011) was used to identify potentially novel miRNAs and to draw the secondary structures and the characteristics of the hairpin structures of miRNA precursors.

Quantification of lncRNA, circRNA, mRNA, and miRNA Expression Levels

Cuffdiff (v2.1.1) was used to calculate fragments per kilo-base millions of exon per million fragments mapped (FPKM) of both lncRNAs and mRNA in each sample (Trapnell et al., 2010). FPKMs of genes were computed by summing the FPKMs of transcripts in each gene group. Also, the raw counts were first normalized using transcripts per million (TPM) (Zhou et al., 2010) and normalized expression levels = (read count*1,000,000)/lib size (lib size is the sum of circRNA read counts). This was used to determine the circRNA expression levels. On the other hand, miRNA expression levels were estimated by TPM based on the following criteria: Normalization formula: Normalized expression = mapped read count/total reads*1,000,000. The differential expression of ncRNAs was determined using the DESeq R package (1.10.1) (Wang et al., 2010).

Target Gene Prediction, GO, and KEGG Enrichment Analysis

In *cis* regulation, lncRNAs can act on neighboring target genes. Coding genes 10 k/100 k upstream or downstream of the lncRNA gene were searched for and their function was analyzed. For *trans* regulation, lncRNAs and their target genes were analyzed based on their expression levels. The correlation between lncRNAs and coding gene expression levels were calculated with custom scripts; then, the genes from different samples were clustered using WGCNA (Langfelder and Horvath, 2008) to search for common expression modules and to analyze the function via functional enrichment analysis. The target genes of miRNAs and miRNA target sites in exons of circRNA loci were identified using miRanda (version 3.3a, main parameter: -sc 140; -en -10; -scale 4; -strict) (Enright et al., 2004). Differentially expressed (DE) ncRNAs were annotated by gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses to investigate their biological functions. Briefly, GO analysis was applied to elucidate genetic regulatory networks of interest by forming hierarchical categories according to the molecular function (MF), cellular component (CC), and biological process (BP) aspects of the differentially expressed genes[1]. KEGG pathway analysis was performed to explore the significantly enriched pathways of DE genes[2].

Quantitative Real-Time PCR

Eight mRNAs, four lncRNAs, and five circRNAs associated with skin and the HF cycle were selected for validation by qRT-PCR analysis. Approximately 1μg of total RNA was used to synthesize cDNA using HiScript II Q Select RT SuperMix for qPCR (Vazyme). qRT-PCR was performed using the AceQ qPCR SYBR® Green Master Mix (Vazyme), according to the manufacturer's instructions, and data were analyzed via QuantStudio® 5 (Applied Biosystems). The specific primer sequences are listed in **Supplementary Table S1**. The expression

levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008), with glyceraldehyde 3-phosphate dehydrogenase (GAPDH) as reference gene.

To confirm the miRNA transcriptome data, three miRNAs were selected for qRT-PCR analysis. Approximately 2μg of total RNA was used to synthesize cDNA after adding a poly (A) tail to the 3′ end of the miRNAs using the miRcute Plus miRNA First-Strand cDNA Synthesis Kit (Tiangen). qRT-PCR was performed using the miRcute miRNA qPCR Detection Kit (SYBR Green), according to the manufacturer's instructions. The specific primers were designed by Beijing Tiangen Co., Ltd. and the product code sets are listed in **Supplementary Table S1**. The U6 small nuclear RNA gene was chosen as internal control. The expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008), and the results of the experiments were normalized to the expression levels of the constitutively expressed U6 gene.

Construction of ncRNAs Regulatory Networks

To investigate the role and interactions between ncRNAs and mRNAs during the HF cycle, ncRNAs regulatory networks were constructed. For the interaction network of lncRNA–miRNA, DE lncRNAs were filtered out according to the homology between lncRNA and miRNA precursor; then, the targeted relationships between lncRNA and miRNA were predicted by miRanda. Then, the regulatory networks of lncRNA–miRNA–mRNA pairs and circRNA–miRNA–mRNA pairs were constructed according to the following steps: (i) the ncRNAs and mRNAs that were upregulated or downregulated were retained; (ii) the interactions of lncRNA–miRNA, miRNA–mRNA, and miRNA–circRNA were predicted by miRanda, which predicts miRNA binding seed sequence sites, as well as overlapping the same miRNA binding site in lncRNAs, circRNAs, and mRNAs; (iii) The lncRNA–miRNA–mRNA pairs network covered two cases: one was the upregulated lncRNA-downregulated miRNA-upregulated mRNA, the other was the downregulated lncRNA-upregulated miRNA-downregulated mRNA. The circRNA–miRNA–mRNA pairs network covered two cases: one was the upregulated circRNA-downregulated miRNA-upregulated mRNA, the other was the downregulated circRNA-upregulated miRNA-downregulated mRNA. Cytoscape software was used to build and visually display the networks.

Luciferase Assay

The dual-luciferase reporter system E1910 (Promega, Madison, WI, United States) was used to perform luciferase activity assays. The miR-320-3p mimic and miR-320-3p negative control mimics were purchased from Shanghai GenePharma Co., Ltd. Wild-type luciferase reporter vectors (pMir-HTATIP2-3'UTR-WT, pMir-LNC_002919-WT, and pMir-novel_circ_0026326-WT) were constructed using the primers shown in **Supplementary Table S2**. Their substitution mutants (pMir-HTATIP2-3'-UTR-MUT, pMir-LNC_002919-MUT, and pMir-novel_circ_0026326-MUT) were synthesized by Beijing Tsingke Co., Ltd. Briefly, the skin fibroblast cells of rabbit

---

[1]http://www.geneontology.org

[2]http://www.genome.jp/kegg/

(RAB-9, ATCC® CRL-1414^{TM}) were cultured in 24-well tissue culture plates. Cells were co-transfected with the pMir-report luciferase reporter, the miRNA (miR-320-3p) mimics and pRL-TK using Lipofectamine^{TM} 2000 (Invitrogen). After 48 h of culture at 37°C, transfected cells were lysed with 100 μl of passive lysis buffer. Next, 20 μl of lysates were mixed with 100 μl of LAR II, and firefly luciferase activity was measured by using a luminometer. As an internal control, 100 μl of Stop & Glo reagent was added to the sample. Firefly luciferase activity was normalized to the corresponding Renilla luciferase activity.

RESULTS

Hair Follicle Cycle Synchronization Model

For the HF cycle synchronization model, Angora rabbits were used. The obtained observations showed that the length of the hair coat increased steadily until day 110. Between days 120 and 150, the growth rate of wool declined rapidly. Then, between days 160 and 180, the wool recovered and once again showed an increased growth rate (**Figure 1A**). Histological analysis showed rapid growth of the hair shaft and increasing depth of the HF between days 0 and 110. Then, the growth of the hair shaft and the depth of the HF decreased between days 120 and 130. Finally, the hair shaft started to fall off and the hair bulbs atrophied between days 140 and 150. After the HF cycle ended, a new HF appeared, the growth of the hair shaft recovered and the HFs moved into a new cycle (**Figure 1B**). In conclusion, the hair cycle of Angora rabbits is characterized by an anagen phase between days 0 and 110, a catagen phase between days 120 and 130, and a telogen phase between days 140 and 150.

Differentially Expressed lncRNAs, mRNAs, miRNAs, and circRNAs

A summary of the lncRNA-seq, miRNA-seq, and circRNA-seq data from the three HF cycle phases is shown in **Supplementary Table S3**, indicating the relatively high quality of the transcriptome data. The lncRNA-seq, miRNA-seq, and circRNA-seq data were deposited in the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the bioproject numbers PRJNA479733, PRJNA495446, and PRJNA495449. DE ncRNAs and mRNAs were analyzed using Cuffdiff software with a criterion of $p < 0.05$. Volcano plots, clustering maps, and Venn diagrams were used to illustrate the distribution of the DE ncRNAs and mRNAs between the three groups (**Figures 2–5**). **Table 1** summarizes the number of DE ncRNAs and mRNAs. Differential expressions of 111 lncRNAs (60 upregulated and 51 downregulated), 247 circRNAs (128 upregulated and 119 downregulated), 97 miRNAs (38 upregulated and 59 downregulated), and 1,168 mRNAs (750 upregulated and 418 downregulated) were found between the three HF cycle stages. Complete information on all DE lncRNAs, circRNAs, miRNAs, and mRNAs is listed in **Supplementary Tables S4–S7**. Several lncRNAs were found to be associated with the HF cycle, such as LNC_002694, LNC_002919, LNC_003354, LNC_003790, LNC_008354, LNC_008931, and LNC_005484, which could regulate gene expression by recognizing their target mRNAs. Based on analysis of their biological function, the candidate lncRNAs associated with the HF cycle are listed in **Supplementary Table S8**. Moreover, analysis of the relationships between circRNAs and genes allowed identification of novel_circ_0004876, novel_circ_0005177, novel_circ_0026326, novel_circ_0034968, and novel_circ_0036671, which may play a role during the HF cycle. In addition, several miRNAs, including miR-128-3p, miR-200a-3p, miR-27a-3p, miR-30e-5p, and miR-320-3p; mRNAs, such as *BMP2*, *CSNK2B*, *KRT17*, *LAMB1*, *FZD4*, *SMAD2*, *HTATIP2*, and *SIAH1* were identified to play pivotal roles during the HF cycle and during skin development.

Validation of Differentially Expressed lncRNAs, circRNAs, miRNAs, and mRNAs by qPCR

To validate the lncRNAs, mRNAs, miRNAs, and circRNAs differential expression results, the relative expression of four DE lncRNAs (LNC_002694, LNC_002919, LNC_003354, and LNC_005484), five DE circRNAs (novel_circ_0004876,

FIGURE 1 | Hair follicle synchronization model in Angora rabbits. **(A)** Measurement of length of the hair coat after shaving the dorsal area of Angora rabbits. **(B)** HE staining of sequential skin samples after shaving the dorsal area of rabbits. The anagen, catagen, telogen, and subsequent anagen stages were determined based on the histomorphology of hair follicles.
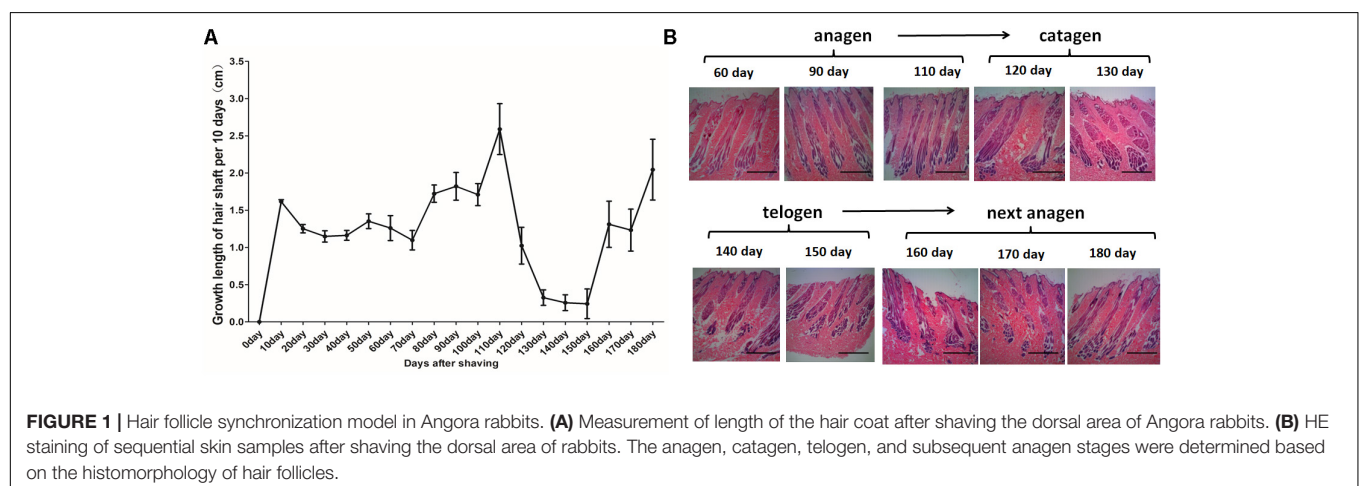
FIGURE 2 | Changes in lncRNA expression during the Angora rabbit hair follicle cycle. **(A–C)** Volcano plots showing up- and down-regulated lncRNAs between days 90, 130, and 150 of the hair follicle cycle. **(D)** Venn diagram showing the number of overlapping differentially expressed lncRNAs between days 90, 130, and 150. **(E)** Heat map of lncRNAs showing hierarchical clustering of DE lncRNAs between days 90, 130, and 150. Up- and down-regulated lncRNAs are shown in red and blue, respectively.
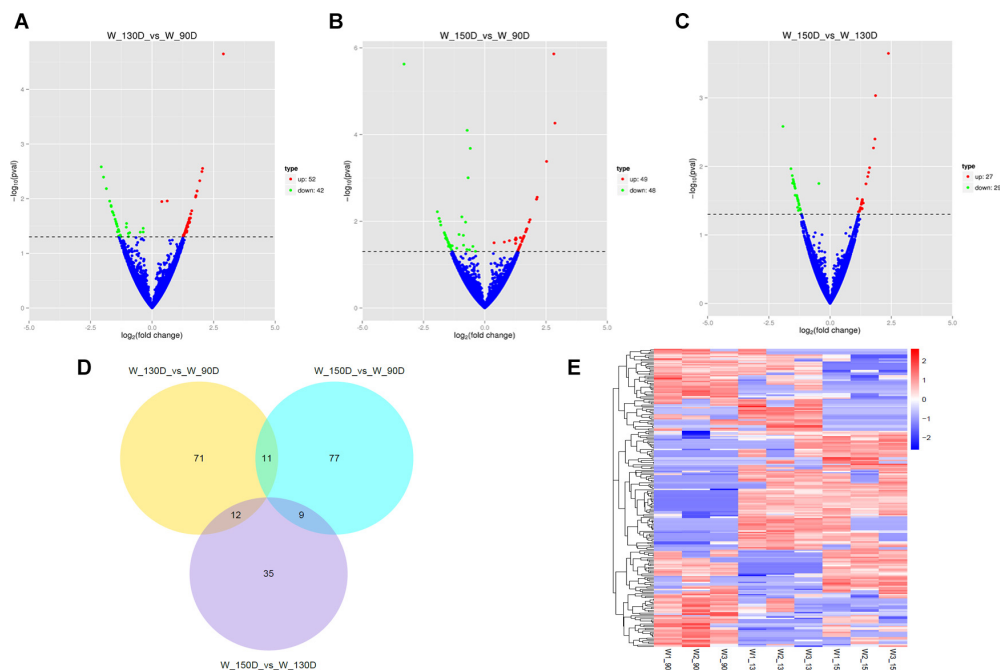
FIGURE 3 | Changes in circRNA expression during the Angora rabbit hair follicle cycle. **(A–C)** Volcano plots showing up- and down-regulated circRNAs between days 90, 130, and 150 of the hair follicle cycle. **(D)** Venn diagram showing the number of overlapping differentially expressed circRNAs between days 90, 130, and 150. **(E)** Heat map of circRNAs showing hierarchical clustering of DE circRNAs between days 90, 130, and 150. Up- and down-regulated circRNAs are shown in red and blue, respectively.
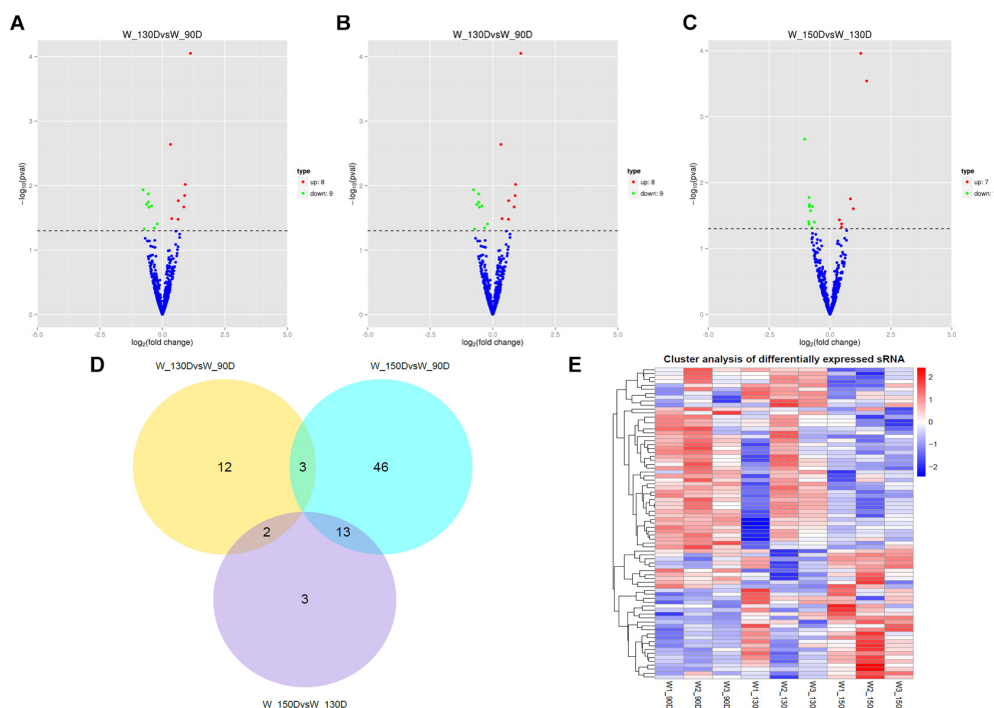
FIGURE 4 | Changes in miRNA expression during the Angora rabbit hair follicle cycle. **(A–C)** Volcano plots showing up- and down-regulated miRNAs between days 90, 130, and 150 of the hair follicle cycle. **(D)** Venn diagram showing the number of overlapping differentially expressed miRNAs between days 90, 130, and 150. **(E)** Heat map of miRNAs showing hierarchical clustering of DE miRNAs between days 90, 130, and 150. Up- and down-regulated miRNAs are shown in red and blue, respectively.

novel_circ_0005177, novel_circ_0026326, novel_circ_0034968, and novel_circ_0036671), four DE miRNA (miR-128-3p, miR-200a-3p, miR-27a-3p, and miR-320-3p), and eight DE mRNAs (*BMP2*, *CSNK2B*, *FAM45A*, *FUOM*, *HTATIP2*, *KRT17*, *ME1*, and *SIAH1*) were measured by qRT-PCR (**Figures 6–9**). The qRT-PCR results were consistent with the transcriptome sequencing data.

GO and KEGG Pathway Analysis

lncRNAs can regulate neighboring protein-coding genes; therefore, a colocalization threshold of 100 kb upstream or downstream of lncRNAs was set for the GO and KEGG analyses. Several GO terms were found that were significantly enriched in the three experimental groups (**Supplementary Table S9**), including skin and HF-related GO terms like HF development (GO: 0001942), hair cycle (GO: 0042633), hair cycle process (GO: 0022405), regulation of HF development (GO: 0051797), and skin morphogenesis (GO: 0043589), among others. The top 20 KEGG pathways associated with DE lncRNAs between days 90, 130, and 150 of the HF cycle based on the function of colocalized mRNAs (**Supplementary Figure S1**) and co-expressed mRNAs (**Supplementary Figure S2**) included the Wnt signaling pathway, TGF-β signaling pathway, MAPK signaling pathway, and JAK/STAT signaling pathway.

In addition, based on the relationship between circRNAs and genes, GO analysis of genes producing DE circRNAs was performed (**Supplementary Table S10**). The GO terms identified

HF development (GO: 0001942), hair cycle process (GO: 0022405), hair cycle (GO: 0042633), and skin development (GO: 0043588), which were all related to skin and HF development. The top 20 KEGG pathways associated with genes producing DE circRNAs between 90, 130, and 150 days (**Supplementary Figure S3**) of the HF cycle were likewise related to skin and HF development, such as the Hedgehog signaling pathway, Wnt signaling pathway, and MAPK signaling pathway.

Furthermore, GO enrichment analysis of genes targeted by DE miRNA (**Supplementary Table S11**) identified GO terms related to HF development, such as HF morphogenesis (GO: 0031069), negative regulation of HF development (GO: 0051799), and regulation of HF development (GO: 0051797), among others. The top 20 KEGG pathways associated with DE miRNAs are shown in **Supplementary Figure S4**. They include pathways related to HF cycle, such as the Hedgehog signaling pathway, NF-κB signaling pathway, and JAK/STAT signaling pathway.

Finally, GO and KEGG analyses of DE mRNAs are shown in **Supplementary Table S12**. The GO terms identified include, for example, skin morphogenesis (GO: 0043589) and positive regulation of HF development (GO: 0051798). The top 20 enriched KEGG pathways for DE genes between the different stages of the HF cycle are shown in **Supplementary Figure S5**. These KEGG pathways include the Wnt signaling pathway, the MAPK signaling pathway, and the TGF-β signaling pathway, which participate in skin development and HF cycle. Differentially expressed genes between days 90, 130, and 150 of
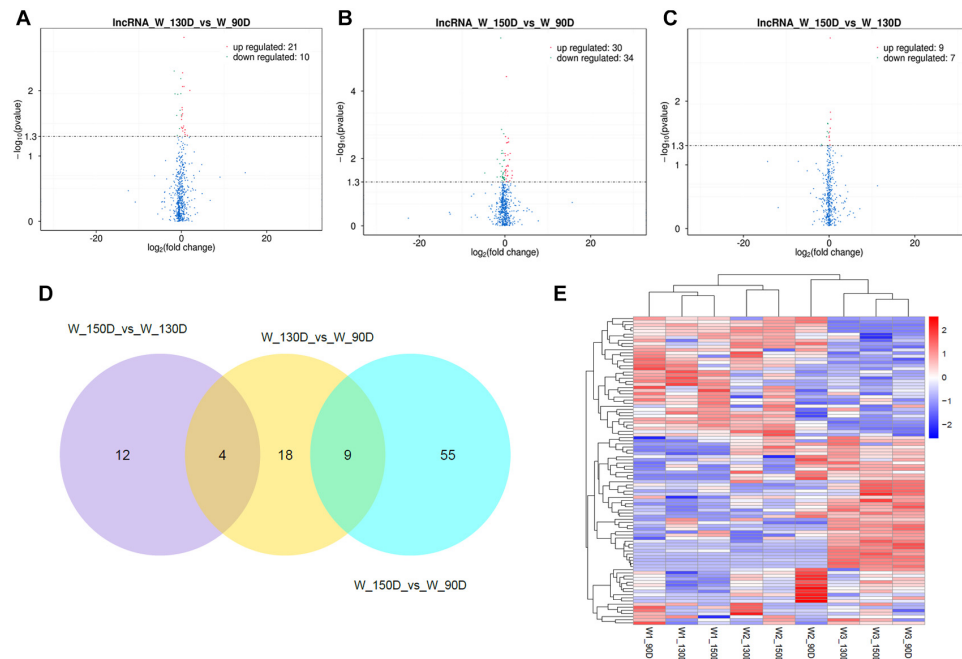
FIGURE 5 | Changes in mRNA expression during the Angora rabbit hair follicle cycle. **(A–C)** Volcano plots showing up- and down-regulated mRNAs between days 90, 130, and 150 of the hair follicle cycle. **(D)** Venn diagram showing the number of overlapping differentially expressed mRNAs between days 90, 130, and 150. **(E)** Heat map of mRNAs showing hierarchical clustering of DE mRNAs between days 90, 130, and 150. Up- and down-regulated mRNAs are shown in red and blue, respectively.

**TABLE 1 |** Summary of the number of differentially expressed ncRNAs and mRNAs.

Groups	Regulation	lncRNA	circRNA	miRNA	mRNA
130 vs. 90 days	Up	21	52	8	152
	Down	10	42	9	64
150 vs. 90 days	Up	30	49	23	387
	Down	34	48	39	136
150 vs. 130 days	Up	9	27	7	211
	Down	7	29	11	218
Total		111	247	97	1168

the HF cycle, as well as their biological functions, are listed in **Supplementary Table S13**.

ceRNA Regulatory Networks

Study of the relationship between ncRNAs and mRNAs may increase our understanding of the molecular mechanisms operating during skin development and HF cycle. According to the competing endogenous RNA (ceRNA) regulatory hypothesis, ncRNAs and mRNAs can compete for the same miRNAs, resulting in additional layers of regulation of gene expression. Based on the analysis of DE lncRNAs, circRNAs, miRNAs, and mRNAs, a network of lncRNAs and miRNAs was first constructed (**Figure 10**). In lncRNA-miRNA-mRNA regulatory networks, miRNA may act as the center, lncRNA as the decoy, and mRNA as the target, which suggests that lncRNAs could act as miRNA

sponges to regulate gene expression (**Figure 11**). In addition, certain circRNAs can competitively bind miRNAs and act as miRNA sponges; therefore, circRNA-miRNA-mRNA triads were constructed with the circRNA as the docoy, miRNA as the center, and mRNA as the target (**Figure 12**).

LNC_002919 and novel_circ_0026326 were identified as ceRNAs for miR-320-3p, which targets *HTATIP2*. A dual-luciferase reporter system was used to verify the binding relationships between the identified lncRNA and miRNA, circRNA and miRNA, and mRNA and miRNA. Luciferase assay showed that miR-320-3p could decrease luciferase activity by binding to sites on LNC_002919, novel_circ_0026326, and the *HTATIP2* 3′UTR (**Figure 13**). The interactions between ncRNAs and mRNA suggest the existence of novel regulatory mechanisms during skin development and HF cycle.

DISCUSSION

The HF cycle is similar in most mammalian species, and many animal models have been used to study the process of hair growth, including mice (Wolbach, 1951; Chase, 1954), rats (Johnson and Ebling, 1964), monkeys (Uno, 1991), cats (Hendriks et al., 1997), and sheep (Hynd et al., 1986). In mice, the hair growth period lasts only 17–19 days, and anterior regions can enter the resting period before the posterior regions regrow (Chase, 1954). By plucking the hairs of rats, the first wave of hair growth was observed between 31 and 22 days,
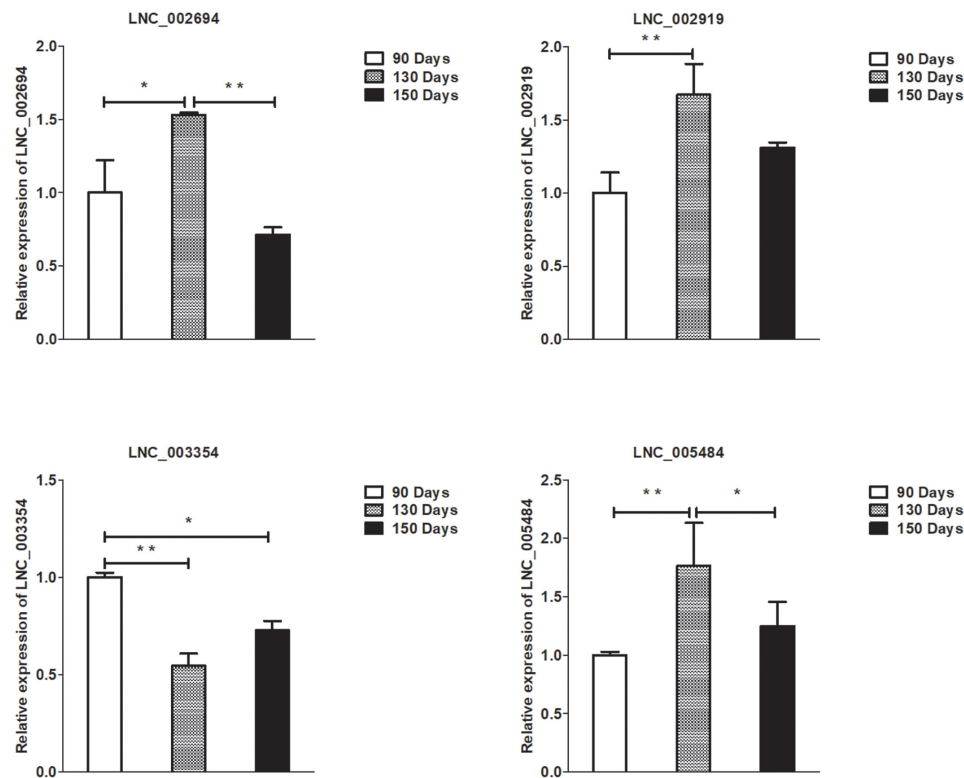
FIGURE 6 | Validation of lncRNA differential expression results at 90, 130, and 150 days. qRT-PCR validation of LNC_002694, LNC_002919, LNC_003354, and LNC_005484 lncRNA expression levels in skin samples between 90, 130, and 150 days. The lncRNA expression levels at 130 and 150 days were normalized to the value at 90 days. Error bars indicate the mean ± SD of triplicate experiments. *$P < 0.05$; **$P < 0.01$.
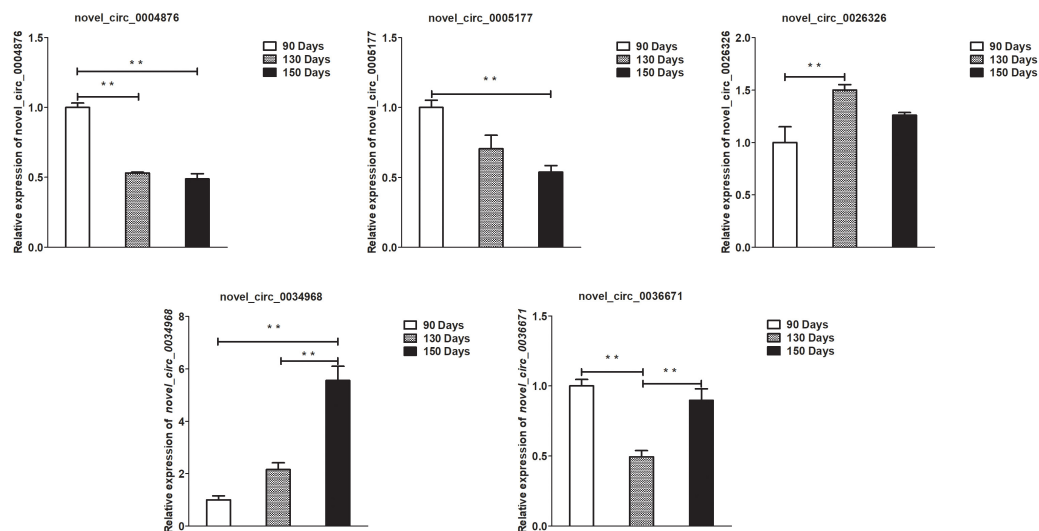
FIGURE 7 | Validation of circRNA differential expression results at 90, 130, and 150 days. qRT-PCR validation of novel_circ_0004876, novel_circ_0005177, novel_circ_0026326, novel_circ_0034968, and novel_circ_0036671 circRNA expression levels in skin samples between 90, 130, and 150 days. The circRNA expression levels at 130 days and 150 days were normalized to the value at 90 days. Error bars indicate the mean ± SD of triplicate experiments. **$P < 0.01$.

and HF from resting clubs were collected at 55 days of age (Johnson and Ebling, 1964). Although animal HFs show a circannual rhythm, the HF cycles producing sheep wool, horse mane, and human scalp hair have special characteristics, including a biological clock that is independent from day and night, season and temperature over a period of 2–6 years
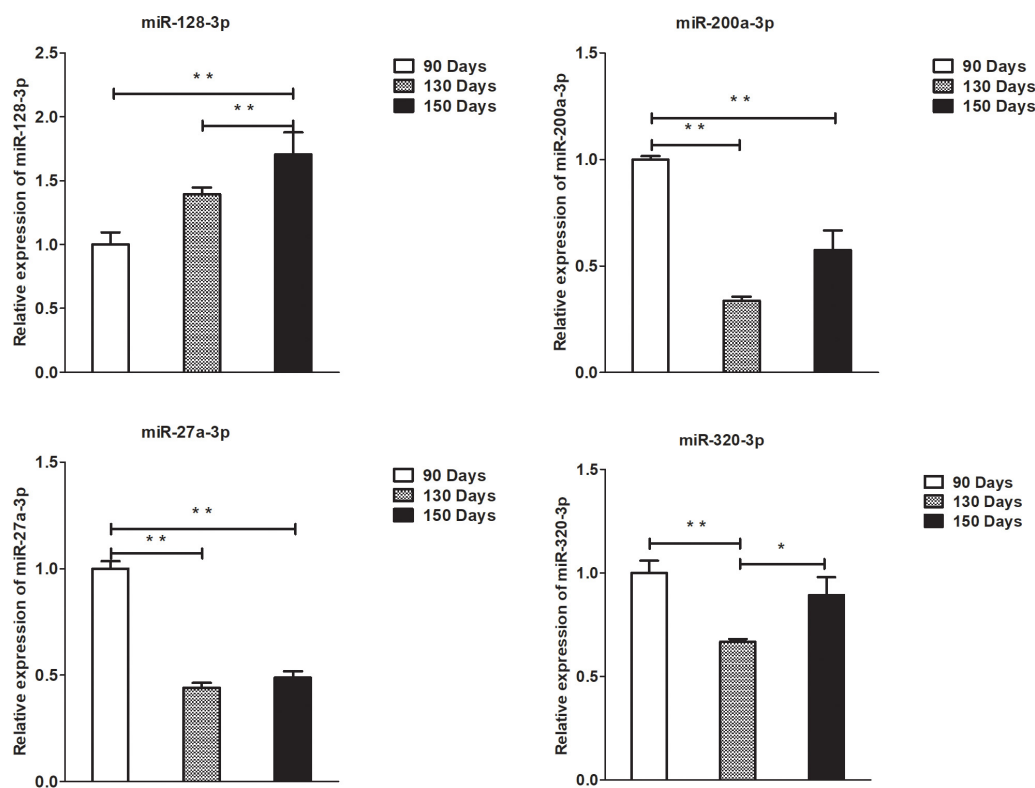
FIGURE 8 | Validation of miRNA differential expression results at 90, 130, and 150 days. qRT-PCR validation of miR-128-3p, miR-200a-3p, 27a-3p, and miR-320-3p miRNA expression levels in skin samples between 90, 130, and 150 days. The miRNA expression levels at 130 days and 150 days were normalized to the value at 90 days. Error bars indicate the mean ± SD of triplicate experiments. $*P < 0.05$; $**P < 0.01$.
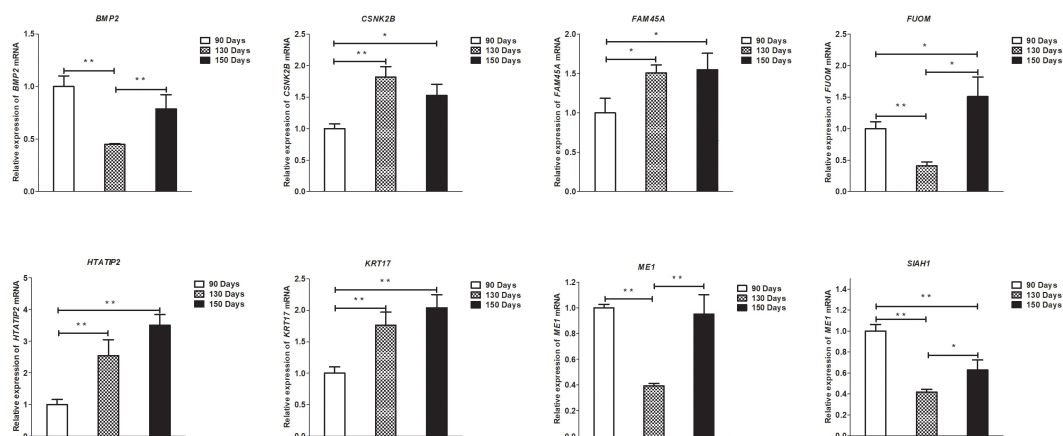
FIGURE 9 | Validation of mRNA differential expression results at 90, 130, and 150 days. qRT-PCR validation of *BMP2*, *CSNK2B*, *FAM45A*, *FUOM*, *HTATIP2*, *KRT17*, *ME1*, and *SIAH1* mRNA expression levels in skin samples between 90, 130, and 150 days. The mRNA expression levels at 130 days and 150 days were normalized to the value at 90 days. Error bars indicate the mean ± SD of triplicate experiments. $*P < 0.05$; $**P < 0.01$.

(Stenn and Paus, 2001). The structure, composition, and growth of hair fibers are similar between Angora rabbits and other rabbit breeds. However, the appearance of a mutation in Angora rabbits leads to a prolongation of the anagen phase, so this phase lasts approximately 5 weeks in New Zealand white rabbits but more than 3 months in Angora rabbits (Moore et al., 1987). The HF

clock in Angora rabbits has its own characteristic chronobiology, with a long growing period, and independence from seasons and temperature. This study established a synchronization model for hair growth in Angora rabbits. The HFs initiated vigorous growth after shaving the dorsal area, and measuring the length of the hair coat and analyzing the histological characteristics

FIGURE 10 | LncRNA–miRNA regulatory networks in Angora rabbit skin. **(A)** Interaction network of lncRNA–miRNA between 130 days and 90 days. **(B)** Interaction network of lncRNA–miRNA between 150 days and 90 days. **(C)** Interaction network of lncRNA–miRNA between 150 days and 130 days.

FIGURE 11 | lncRNA–miRNA–mRNA regulatory networks in Angora rabbit skin. **(A)** Interaction network of lncRNA–miRNA–mRNA between 130 days and 90 days. **(B)** Interaction network of lncRNA–miRNA–mRNA between 150 days and 90 days. **(C)** Interaction network of lncRNA–miRNA–mRNA between 150 days and 130 days.

FIGURE 12 | circRNA–miRNA–mRNA regulatory networks in Angora rabbit skin. **(A)** Interaction network of circRNA–miRNA–mRNA between 130 days and 90 days. **(B)** Interaction network of circRNA–miRNA–mRNA between 150 days and 90 days. **(C)** Interaction network of circRNA–miRNA–mRNA between 150 days and 130 days.

FIGURE 13 | Verification of target binding. **(A)** miR-320-3p putative binding sites in *HTATIP2* 3′-UTR. Blue letters indicate wild type sites and red letters indicate mutated sites in the pMir-report luciferase reporter vector. Luciferase assays were performed in RAB-9 cells co-transfected with pMir-report-HTATIP2-3′-UTR-WT and miRNA-320-3p mimics, or pMir-HTATIP2-3′-UTR-MUT, and miRNA-320-3p mimics. **(B)** Putative binding sites for miR-320-3p in LNC_002919. Blue letters indicate wild type sites and red letters indicate mutated sites in the pMir-report luciferase reporter vector. The luciferase assays were performed in RAB-9 cells co-transfected with pMir-LNC_002919-WT and miRNA-320-3p mimics, or pMir-LNC_002919-MUT, and miRNA-320-3p mimics. **(C)** Putative binding sites for miR-320-3p in novel_circ_0026326. Blue letters indicate wild type sites and red letters indicate mutated sites in the pMir-report luciferase reporter vector. The luciferase assays were performed in RAB-9 cells co-transfected with pMir-novel_circ_0026326-WT and miRNA-320-3p mimics, or pMir-novel_circ_0026326-MUT, and miRNA-320-3p mimics. Error bars indicate the mean ± SD of triplicate experiments. **$P < 0.01$.

showed that the growth phase lasted about 110 days, the regression period started at about 120 days, and the resting period at about 150 days. The HF synchronization model can contribute to the field of research in the chronobiology of HFs. ncRNAs are epigenetic, translational and genetic regulators that may play a role in numerous biological processes in eukaryotes (Mattick and Makunin, 2006). ncRNAs could play complicated and vital roles during the hair cycle; investigation of the regulatory and functional interactions between lncRNAs, circRNAs, miRNAs, and mRNAs may increase understanding of this biological process.

The present study investigated ncRNAs and mRNAs that were significantly up-regulated or down-regulated during the three stages of the HF cycle. Recent studies have shown that DE lncRNAs modulate biological functions in dermal papilla cells, which regulate postnatal hair cycling and HF cycle (Lin et al., 2015). Likewise, RNA-seq technology has been used for the analysis of lncRNAs and mRNAs during the initiation of sheep secondary HFs (Yue et al., 2016). In addition, miRNAs have been the focus of intense research for several years, and have been associated with HF morphogenesis and development (Mardaryev et al., 2010; Ahmed et al., 2014; Hochfeld et al., 2017). However, only few studies analyzed the involvement of circRNA in skin development and HF cycle. circRNAs can act as miRNA sponges, suppressing miRNA activity and resulting in increased RNA expression (Hansen et al., 2013). This study employed high-throughput sequencing for the analysis of DE ncRNAs in the HF during the different hair cycle stages, based on the synchronization model. A total of 111 lncRNAs, 247 circRNAs, 97 mi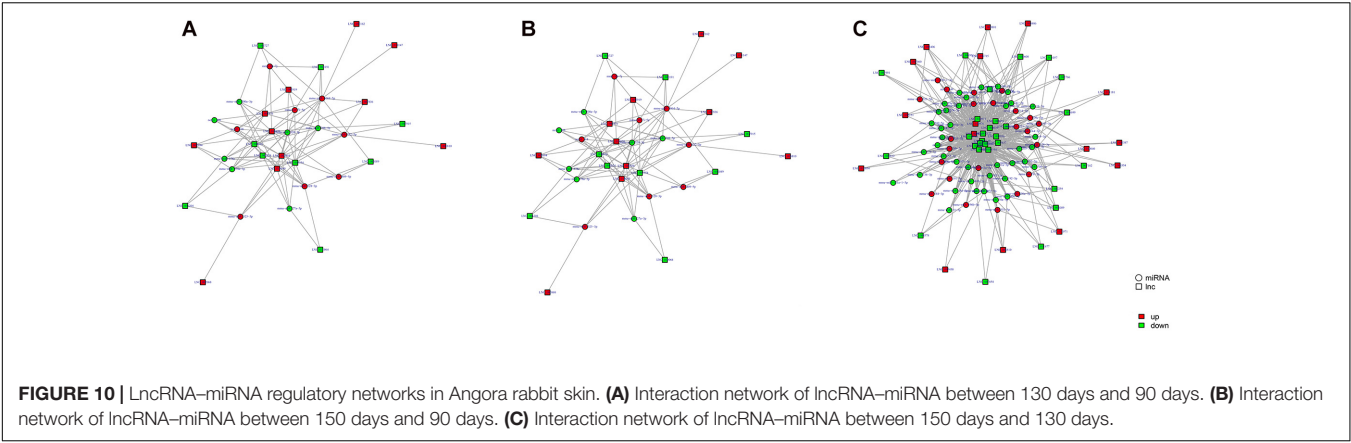RNAs, and 1,168 mRNAs were differentially expressed during the hair cycle stages. Moreover, several differentially expressed mRNAs were identified during hair cycling. As a dermal papilla signature gene, *BMP2* is expressed in the hair matrix and can regulate HF cycling (Nakamura et al., 2003; Rendl et al., 2008). In this study, its

expression in the catagen was significantly decreased. A previous study reported that KRT17 acts as a key factor to regulate the hair cycling, which affects the transition of anagen-catagen (Tong and Coulombe, 2006). The present results showed that *KRT17* is highly expressed in catagen (via the identified candidates mRNA) between days 130 and 90 as well as between days 150 and 90. Moreover, the co-location relationships between LNC_004603 and *KRT17* were obtained via functional analysis of lncRNA, which indicates that LNC_004603 may act as a potential factor for the regulation of hair cycling. Furthermore, miR-200a-3p is highly expressed in the anagen, which has been proved the be preferentially expressed in the epidermis (Yi et al., 2006). In addition, the expression of miR-128-3p significantly increased from days 90 to 150, with high expression in the telogen. In human HF mesenchymal stem cells, miR-128 could regulate the cell differentiation by targeting SMAD2 (Wang et al., 2016).

Gene ontology analysis includes three domains describing the cellular and molecular roles of genes and gene products (MF, CC, and BP) (Harris, 2004). KEGG is a pathway database for the systematic analysis of gene function, linking genomic and functional information (Ogata et al., 2000). GO and KEGG were used to investigate the potential mechanisms of action of the DE ncRNAs in this study. The obtained results suggest that multiple signaling pathways form a complex regulatory network during skin and HF development. These include the Wnt signaling pathway, the Hedgehog signaling pathway, the TGF-β signaling pathway, the MAPK signaling pathway, the BMP signaling pathway, and the JAK/STAT signaling pathway. These signaling pathways have been previously reported to regulate HF morphogenesis and development (Andl et al., 2002; Mill et al., 2003; Jamora et al., 2005; Kulessa et al., 2014; Akilli Öztürk et al., 2015; Harel et al., 2015). Both *SMAD2* and *SIAH1* were enriched in the Wnt signaling pathway, and *SMAD2* was upregulated at day 150 compared to the

differential expression at day 90. In addition, *SIAH1* decreased significantly from days 90 to 130, but increased from days 130 to 150, and was highly expressed when comparing day 150 to day 90. In the cashmere goat, *SIAH1* and *SMAD2* were significantly expressed during the telogen-anagen HF transition. *SIAH1* is highly significantly expressed from telogen to early anagen, and the expression of *SMAD2* increased from telogen to late anagen (Liu et al., 2018). Via functional analysis of lncRNA, the co-expression relationship between LNC_002690 and *SIAH1* was identified, indicating that LNC_002690 might play a central role in hair cycling via regulation of *SIAH1* expression. Hence, these candidates could act as key candidates during HF cycling.

RNA transcripts are regulated by ceRNAs, which compete for the binding of shared miRNAs. miRNA response elements (MREs) are sequences where miRNAs can bind and repress target gene expression. Acting as miRNA sponges, pseudogenes, lncRNAs, circRNAs, and mRNAs can suppress miRNA function through shared MREs (Salmena et al., 2011). Therefore, to try to understand the role of ncRNAs during the HF cycle, lncRNA–miRNA–mRNA and circRNA–miRNA–mRNA regulatory networks were constructed. LNC_002919 and novel_circ_0026326 acted as sponges for miR-320-3p, which targets *HTATIP2*. MiR-320-3p has been reported to either directly or indirectly target genes that regulate the cell cycle and differentiation of the HF (Liu et al., 2013). *HTATIP2* was highly expressed during the catagen and telogen phases, suggesting that *HTATIP2* could inhibit cellular activities during the hair cycle. Decreased or absent HTATIP2 activity modulated through JAK-STAT3 signaling has been shown to play an important role in certain cellular processes. Furthermore, the study shows a link between the JAK-STAT signaling pathway and hair growth (Zhang et al., 2012; Harel et al., 2015). In this analysis of DE lncRNAs, a relationship was found between LNC_002919 and *KRTAP11-1*, suggesting that LNC_002919 could modulate *KRTAP11-1* expression. *KRTAP11-1* influences keratin-bundle assembly and can regulate the physical properties of hair (Fujimoto et al., 2014). Therefore, LNC_002919 could be a potent regulator of the HF cycle. However, the molecular mechanisms underlying the regulation of HTATIP2 by LNC_002919 and novel_circ_0026326, which may act as miR-320-3p sponges, need to be further explored.

CONCLUSION

In summary, this study established a rabbit HF synchronization model and investigated the lncRNA, circRNA, miRNA, and mRNA expression profiles by transcriptome analysis of samples collected at different stages of the HF cycle. GO and KEGG pathway enrichment analyses were carried out to identify candidate ncRNAs and mRNAs involved in the regulation of the HF cycle. In addition, ceRNA networks were constructed, which may be active during the HF cycle. These results provide a basis for an improved understanding of the mechanisms underlying the HF cycle.

DATA AVAILABILITY

The datasets generated for this study can be found in The lncRNA-seq, miRNA-seq, and circRNA-seq data were deposited in the SRA of the NCBI, The lncRNA-seq, miRNA-seq, and circRNA-seq data were deposited in the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the bioproject numbers PRJNA479733, PRJNA495446, and PRJNA495449.

ETHICS STATEMENT

The experimental procedures in this study were approved by the Animal Care and Use Committee of Yangzhou University.

AUTHOR CONTRIBUTIONS

BZ was responsible for the collection and analysis of results and wrote the manuscript. YC, SH, NY, and MW were responsible for construction of hair follicle synchronization model. ML, JL, and YX carried out of the experiments. BZ and XW designed the study and finalized the manuscript. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00407/full#supplementary-material

FIGURE S1 | Top 20 KEGG pathways based on mRNA colocalization with differentially expressed lncRNAs between 90, 130, and 150 days. **(A)** Scatterplot showing KEGG pathway enrichment between 130 and 90 days. **(B)** Scatterplot showing KEGG pathway enrichment between 150 and 90 days. **(C)** Scatterplot showing KEGG pathway enrichment between 150 and 130 days.

FIGURE S2 | Top 20 KEGG pathways based on mRNA co-expression with differentially expressed lncRNAs between 90, 130, and 150 days. **(A)** Scatterplot showing KEGG pathway enrichment between 130 and 90 days. **(B)** Scatterplot showing KEGG pathway enrichment between 150 and 90 days. **(C)** Scatterplot showing KEGG pathway enrichment between 150 and 130 days.

FIGURE S3 | Top 20 KEGG pathways associated with differentially expressed circRNAs between 90, 130, and 150 days. **(A)** Scatterplot showing KEGG pathway enrichment between 130 and 90 days. **(B)** Scatterplot showing KEGG pathway enrichment between 150 and 90 days. **(C)** Scatterplot showing KEGG pathway enrichment between 150 and 130 days.

FIGURE S4 | Top 20 KEGG pathways associated with differentially expressed miRNAs between 90, 130, and 150 days. **(A)** Scatterplot showing KEGG pathway

enrichment between 130 and 90 days. **(B)** Scatterplot showing KEGG pathway enrichment between 150 and 90 days. **(C)** Scatterplot showing KEGG pathway enrichment between 150 and 130 days.

FIGURE S5 | Top 20 KEGG pathways associated with differentially expressed mRNAs between 90, 130, and 150 days. **(A)** Scatterplot showing KEGG pathway enrichment between 130 and 90 days. **(B)** Scatterplot showing KEGG pathway enrichment between 150 and 90 days. **(C)** Scatterplot showing KEGG pathway enrichment between 150 and 130 days.

TABLE S1 | Primers used for the quantitative real-time PCR analysis.

TABLE S2 | Primers used for construction of the luciferase reporter vector.

TABLE S3 | Summary of RNA sequencing for each sample.

TABLE S4 | Analysis of differentially expressed lncRNAs between 90, 130, and 150 days.

TABLE S5 | Analysis of differentially expressed circRNAs between 90, 130, and 150 days.

TABLE S6 | Analysis of differentially expressed miRNAs between 90, 130, and 150 days.

TABLE S7 | Analysis of differentially expressed mRNAs between 90, 130, and 150 days.

TABLE S8 | Differentially expressed lncRNAs associated with the hair follicle cycle by target prediction.

TABLE S9 | Gene ontology classification of differentially expressed lncRNAs between 90, 130, and 150 days.

TABLE S10 | Gene ontology classification of differentially expressed circRNAs between 90, 130, and 150 days.

TABLE S11 | Gene ontology classification of differentially expressed miRNA between 90, 130, and 150 days.

TABLE S12 | Gene ontology classification of differentially expressed mRNAs between 90, 130, and 150 days.

TABLE S13 | Differentially expressed genes between 90, 130, and 150 days of the hair follicle cycle.

REFERENCES

Ahmed, M. I., Alam, M., Emelianov, V. U., Poterlowicz, K., Patel, A., Sharov, A. A., et al. (2014). MicroRNA-214 controls skin and hair follicle development by modulating the activity of the Wnt pathway. *J. Cell Biol.* 207, 549–567. doi: 10.1083/jcb.201404001

Ahmed, N. S., Ghatak, S., El Masry, M. S., Gnyawali, S. C., Roy, S., Amer, M., et al. (2017). Epidermal E-cadherin dependent β-catenin pathway is phytochemical inducible and accelerates anagen hair cycling. *Mol. Ther.* 25, 2502–2512. doi: 10.1016/j.ymthe.2017.07.010

Akilli Öztürk, Ö, Pakula, H., Chmielowiec, J., Qi, J., Stein, S., Lan, L., et al. (2015). Gab1 and mapk signaling are essential in the hair cycle and hair follicle stem cell quiescence. *Cell Rep.* 13, 561–572. doi: 10.1016/j.celrep.2015.09.015

Andl, T., Reddy, S. T., Gaddapara, T., and Millar, S. E. (2002). WNT signals are required for the initiation of hair follicle development. *Dev. Cell* 2, 643–653. doi: 10.1016/s1534-5807(02)00167-3

Chao, Y., Wang, X., Geng, R., He, X., Lei, Q., and Chen, Y. (2013). Discovery of cashmere goat (*Capra hircus*) microRNAs in skin and hair follicles by Solexa sequencing. *BMC Genomics* 14:511. doi: 10.1186/1471-2164-14-511

Chase, H. B. (1954). Growth of the hair. *Physiol. Rev.* 34:113.

Chase, H. B., Rauch, R., and Smith, V. W. (1951). Critical stages of hair development and pigmentation in the mouse. *Physiol. Zool.* 24, 1–8. doi: 10.1086/physzool.24.1.30152098

Chen, Z. Z., Lan, H., Wu, Y. H., Zhai, W. J., Zhu, P. P., and Gao, Y. F. (2016). LncSox4promotes the self-renewal of liver tumour-initiating cells through Stat3-mediated Sox4 expression. *Nat. Commun.* 7:12598. doi: 10.1038/ncomms12598

Cotsarelis, G., Sun, T. T., and Lavker, R. M. (1990). Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* 61, 1329–1337. doi: 10.1016/0092-8674(90)90696-c

Ding, X. C., Weiler, J., and Grosshans, H. (2009). Regulating the regulators: mechanisms controlling the maturation of microRNAs. *Trends Biotechnol.* 27, 27–36. doi: 10.1016/j.tibtech.2008.09.006

Enright, A. J., Bino, J., Ulrike, G., Thomas, T., Chris, S., and Marks, D. S. (2004). MicroRNA targets in *Drosophila*. *Genome Biol.* 5:R1.

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2011). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Fuchs, E., and Segre, J. A. (2000). Stem cells: a new lease on life. *Cell* 100, 143–155.

Fujimoto, S., Takase, T., Kadono, N., Maekubo, K., and Hirai, Y. (2014). Krtap11-1, a hair keratin-associated protein, as a possible crucial element for the physical properties of hair shafts. *J. Dermatol. Sci.* 74, 39–47. doi: 10.1016/j.jdermsci.2013.12.006

Guttman, M., and Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482:339. doi: 10.1038/nature10887

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993

Hardy, M. H. (1992). The secret life of the hair follicle. *Trends Genet. Tig* 8:55. doi: 10.1016/0168-9525(92)90350-d

Harel, S., Higgins, C. A., Cerise, J. E., Dai, Z., Chen, J. C., Clynes, R., et al. (2015). Pharmacologic inhibition of JAK-STAT signaling promotes hair growth. *Sci. Adv.* 1:e1500973. doi: 10.1126/sciadv.1500973

Harris, M. A. (2004). *The Gene Ontology (GO) Database and Informatics Resource*. New York, NY: WCB/McGraw-Hill.

Hendriks, W. H., Tarttelin, M. F., and Moughan, P. J. (1997). Seasonal hair growth in the adult domestic cat (*Felis catus*). *Comp. Biochem. Physiol. Part A Physiol.* 116, 29–35. doi: 10.1016/s0300-9629(96)00113-2

Hochfeld, L. M., Anhalt, T., Reinbold, C. S., Herrerarivero, M., Fricker, N., Nöthen, M. M., et al. (2017). Expression profiling and bioinformatic analyses suggest new target genes and pathways for human hair follicle related microRNAs. *BMC Dermatol.* 17:3. doi: 10.1186/s12895-017-0054-9

Hynd, P. I., Schlink, A. C., Phillips, P. M., and Scobie, D. R. (1986). Mitotic activity in cells of the wool follicle bulb. *Aus. J Biol. Sci.* 39:329. doi: 10.1071/bi9860329

Jamora, C., Lee, P., Kocieniewski, P., Azhar, M., Hosokawa, R., Chai, Y., et al. (2005). A signaling pathway involving tgf-β2 and snail in hair follicle morphogenesis. *PLoS Biol.* 3:e11. doi: 10.1371/journal.pbio.0030011

Johnson, E., and Ebling, F. J. (1964). The effect of plucking hairs during different phases of the follicular cycle. *J. Embryol. Exp. Morphol.* 12, 465–474.

Kulessa, H., Turk, G., and Hogan, B. L. M. (2014). Inhibition of Bmp signaling affects growth and differentiation in the anagen hair follicle. *Embo J.* 19, 6664–6674. doi: 10.1093/emboj/19.24.6664

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, D., Liu, X., Zhou, J., Hu, J., Zhang, D., Liu, J., et al. (2016). LncRNA HULC modulates the phosphorylation of YB-1 through serving as a scaffold of ERK and YB-1 to enhance hepatocarcinogenesis. *Hepatology* 65:1612. doi: 10.1002/hep.29010

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., et al. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* 22:256. doi: 10.1038/nsmb.2959

Lin, C. M., Liu, Y., Huang, K., Chen, X. C., Cai, B. Z., Li, H. H., et al. (2015). Long noncoding RNA expression in dermal papilla cells contributes to hairy gene regulation. *Biochem. Biophys. Res. Commun.* 453, 508–514. doi: 10.1016/j.bbrc.2014.09.119

Liu, G., Liu, R., Li, Q., Tang, X., Yu, M., Li, X., et al. (2013). Identification of microRNAs in wool follicles during anagen, catagen, and telogen phases in Tibetan Sheep. *PLoS One* 8:e77801. doi: 10.1371/journal.pone.0077801

Liu, Z., Yang, F., Zhao, M., Ma, L., Li, H., Xie, Y., et al. (2018). The intragenic mRNA-microRNA regulatory network during telogen-anagen hair follicle transition in the cashmere goat. *Sci. Rep.* 8:14227. doi: 10.1038/s41598-018-31986-2

Lu, Q., Shan, S., Li, Y., Zhu, D., Jin, W., and Ren, T. (2018). Long noncoding RNA SNHG1 promotes non-small cell lung cancer progression by up-regulating MTDH via sponging miR-145-5p. *FASEB J.* 32, 3957–3967. doi: 10.1096/fj.201701237RR

Mardaryev, A. N., Ahmed Mivlahov, N. V., Fessing, M. Y., Gill, J. H., Sharov, A. A., and Botchkareva, N. V. (2010). Micro-RNA-31 controls hair cycle-associated changes in gene expression programs of the skin and hair follicle. *Faseb J.* 24, 3869–3881. doi: 10.1096/fj.10-160663

Mattick, J. S., and Makunin, I. V. (2006). Non-coding RNA. *Hum. Mol. Genet.* 15, R17–R29.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2014). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature11928

Mill, P., Mo, R., Fu, H., Grachtchouk, M., Kim, P. C., Dlugosz, A. A., et al. (2003). Sonic hedgehog-dependent activation of Gli2 is essential for embryonic hair follicle development. *Genes Dev.* 17:282. doi: 10.1101/gad.1038103

Moore, G. P. M., Thébault, R. G., Rougeot, J., Dooren, P. V., and Bonnet, M. (1987). Epidermal growth factor (EGF) facilitates depilation of the Angora rabbit. *Annales De Zootechnie* 36, 433–438. doi: 10.1051/animres%3A19870407

Müller-Röver, S., Foitzik, K., Paus, R., Handjiski, B., Veen, C. V. D., Eichmüller, S., et al. (2001). A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. *J. Invest. Dermatol.* 117, 3–15. doi: 10.1046/j.0022-202x.2001.01377.x

Nakamura, M., Matzuk, M. M., Gerstmayer, B., Bosio, A., Lauster, R., Miyachi, Y., et al. (2003). Control of pelage hair follicle development and cycling by complex interactions between follistatin and activin. *Faseb J.* 17:497. doi: 10.1096/fj.02-0247fje

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34.

Oro, A. E., and Scott, M. P. (1998). Splitting hairs: dissecting roles of signaling systems in epidermal development. *Cell* 95:575.

Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K., and Barrandon, Y. (2001). Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* 104, 233–245. doi: 10.1016/s0092-8674(01)00208-2

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). Translation of CircRNAs. *Mol. Cell.* 66, 9.e7–21.e7. doi: 10.1016/j.molcel.2017.02.021

Paus, R., and Cotsarelis, G. (1999). The biology of hair follicles. *N. Engl. J. Med.* 341, 491–497.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11:1650. doi: 10.1038/nprot.2016.095

Rendl, M., Polak, L., and Fuchs, E. (2008). BMP signaling in dermal papilla cells is required for their hair follicle-inductive properties. *Genes Dev.* 22, 543–557. doi: 10.1101/gad.1614408

Reut, A. F., Markus, M., Nagarjuna Reddy, P., Andranik, I., Osnat, B., Mor, H., et al. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell.* 56, 55–66. doi: 10.1016/j.molcel.2014.08.019

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014

Sardella, C., Winkler, C., Quignodon, L., Hardman, J. A., Toffoli, B., Gmp, G. A., et al. (2017). Delayed hair follicle morphogenesis and hair follicle dystrophy in a lipoatrophy mouse model of pparg total deletion. *J. Invest. Dermatol.* 138, 500–510. doi: 10.1016/j.jid.2017.09.024

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

Shirokova, V., Biggs, L. C., Jussila, M., Ohyama, T., Groves, A. K., and Mikkola, M. L. (2016). Foxi3 deficiency compromises hair follicle stem cell specification and activation. *Stem Cells* 34, 1896–1908. doi: 10.1002/stem.2363

Song, S., Yang, M., Li, Y., Rouzi, M., Zhao, Q., Pu, Y., et al. (2018). Genome-wide discovery of lincRNAs with spatiotemporal expression patterns in the skin of goat during the cashmere growth cycle. *BMC Genomics* 19:495. doi: 10.1186/s12864-018-4864-x

Song, Y., Sun, J., Zhao, J., Yang, Y., Shi, J., Wu, Z., et al. (2017). Non-coding RNAs participate in the regulatory network of CLDN4 via ceRNA mediated miRNA evasion. *Nat. Commun.* 8:289. doi: 10.1038/s41467-017-00304-1

Stenn, K. S., and Paus, R. (2001). Controls of hair follicle cycling. *Physiol. Rev.* 81:449. doi: 10.1152/physrev.2001.81.1.449

Stoffelen, R., Jimenez, M. I., and Dierckxsens, C. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 7:e30733. doi: 10.1371/journal.pone.0030733

Straile, W. E., Chase, H. B., and Arsenault, C. (2010). Growth and differentiation of hair follicles between periods of activity and quiescence. *J. Exp. Zool.* 148, 205–221. doi: 10.1002/jez.1401480304

Tong, X., and Coulombe, P. A. (2006). Keratin 17 modulates hair follicle cycling in a TNFα-dependent fashion. *Genes Dev.* 20:1353. doi: 10.1101/gad.1387406

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Uno, H. (1991). Quantitative models for the study of hair growth in vivo. *Ann. N.Y. Acad. Sci.* 642, 107–124. doi: 10.1111/j.1749-6632.1991.tb24384.x

Veen, C. V. D., Handjiski, B., Paus, R., Müller-Röver, S., Maurer, M., Eichmüller, S., et al. (1999). A comprehensive guide for the recognition and classification of distinct stages of hair follicle morphogenesis. *J. Invest. Dermatol.* 113:523. doi: 10.1046/j.1523-1747.1999.00740.x

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612

Wang, S., Ge, W., Luo, Z., Guo, Y., Jiao, B., Qu, L., et al. (2017). Integrated analysis of coding genes and non-coding RNAs during hair follicle cycle of cashmere goat (*Capra hircus*). *BMC Genomics* 18:767. doi: 10.1186/s12864-017-4145-0

Wang, Z., Pang, L., Zhao, H., Song, L., Wang, Y., Sun, Q., et al. (2016). miR-128 regulates differentiation of hair follicle mesenchymal stem cells into smooth muscle cells by targeting SMAD2. *Acta Histochem.* 118, 393–400. doi: 10.1016/j.acthis.2016.04.001

Winkle, M., Van, D. B. A., Tayari, M., Sietzema, J., Terpstra, M., Kortman, G., et al. (2015). Long noncoding RNAs as a novel component of the Myc transcriptional network. *Faseb J.* 29:2338. doi: 10.1096/fj.14-263889

Wolbach, S. B. (1951). The hair cycle of the mouse and its importance in the study of sequences of experimental carcinogenesis. *Ann. N.Y. Acad. Sci.* 53, 517–536. doi: 10.1111/j.1749-6632.1951.tb31954.x

Yi, R., O'carroll, D., Pasolli, H. A., Zhang, Z., Dietrich, F. S., Tarakhovsky, A., et al. (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat. Genet.* 38:356. doi: 10.1038/ng1744

Yue, Y., Guo, T., Yuan, C., Liu, J., Guo, J., Feng, R., et al. (2016). Integrated analysis of the roles of long noncoding rna and coding RNA expression in sheep (*Ovis aries*) skin during initiation of secondary hair follicle. *PLoS One* 11:e0156890. doi: 10.1371/journal.pone.0156890

Zhang, W., Sun, H. C., Wang, W. Q., Zhang, Q. B., Zhuang, P. Y., Xiong, Y. Q., et al. (2012). Sorafenib down-regulates expression of htatip2 to promote invasiveness and metastasis of orthotopic hepatocellular carcinoma tumors in mice. *Gastroenterology* 143, 1641.e5–1649.e5. doi: 10.1053/j.gastro.2012.08.032

Zhong, Z., Lv, M., and Chen, J. (2016). Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma. *Sci. Rep.* 6:30919. doi: 10.1038/srep30919

Zhou, G., Kang, D., Ma, S., Wang, X., Gao, Y., Yang, Y., et al. (2018). Integrative analysis reveals ncRNA-mediated molecular regulatory network driving secondary hair follicle regression in cashmere goats. *BMC Genomics* 19:222. doi: 10.1186/s12864-018-4603-3

Zhou, L., Chen, J., Li, Z., Li, X., Hu, X., Huang, Y., et al. (2010). Integrated Profiling of MicroRNAs and mRNAs: MicroRNAs Located on Xq27.3 Associate with Clear Cell Renal Cell Carcinoma. *PLoS One* 5:e15224. doi: 10.1371/journal.pone.0015224

Zhu, Y. B., Wang, Z. Y., Yin, R. H., Jiao, Q., Zhao, S. J., Cong, Y. Y., et al. (2018). A lncRNA-H19 transcript from secondary hair follicle of Liaoning cashmere goat: Identification, regulatory network and expression regulated potentially by its promoter methylation. *Gene* 641, 78–85. doi: 10.1016/j.gene.2017.10.028

Zhu, Z., Li, Y., Liu, W., He, J., Zhang, L., Li, H., et al. (2018). Comprehensive circRNA expression profile and construction of circRNA-associated ceRNA network in fur skin. *Exp. Dermatol.* 27, 251–257. doi: 10.1111/exd.13502

Coding or Noncoding, the Converging Concepts of RNAs

Jing Li and Changning Liu*

CAS Key Laboratory of Tropical Plant Resource and Sustainable Use, Xishuangbanna Tropical Botanical Garden,
The Innovative Academy of Seed Design, Chinese Academy of Sciences, Kunming, China

Technological advances over the past decade have unraveled the remarkable complexity of RNA. The identification of small peptides encoded by long non-coding RNAs (lncRNAs) as well as regulatory functions mediated by non-coding regions of mRNAs have further complicated our understanding of the multifaceted functions of RNA. In this review, we summarize current evidence pointing to dual roles of RNA molecules defined by their coding and non-coding potentials. We also discuss how the emerging roles of RNA transform our understanding of gene expression and evolution.

Keywords: messenger RNA, long noncoding RNA, coding potential, ribosome profiling, micropeptide

INTRODUCTION

Benefiting from the advances in science and technology, our understanding of the complexity of organisms is constantly increasing. The "central dogma" of molecular biology states that genetic information is typically processed from DNA to RNA to protein, and this decides cellular and organismal phenotype (Crick, 1970). In the past, RNAs, except for infrastructural RNAs (such as rRNAs and tRNAs), were commonly considered as an intermediate between DNA and proteins. However, over recent decades, the rapid development of high-throughput sequencing technologies has revealed the pervasive transcription of eukaryotic genomes (Okazaki et al., 2002; Carninci et al., 2005; Kapranov et al., 2007; Lander, 2011), thus revealing RNA-mediated gene regulation. The fact that most regulatory RNAs function without involvement in protein translation led us to re-examine the roles of RNAs in the development and evolution of higher organisms.

In higher organisms, only a small fraction of genetic transcripts (less than 3%) have the capability to encode proteins, despite pervasive transcription across genomes. This raises the question of whether the remaining non-protein-coding transcripts are transcriptional "noise" or contain more genetic information. Large-scale projects for the systematic annotation and functional characterization of genes (such as ENCODE and FANTOM) have reported that at least 80% of mammalian genomic DNA is actively transcribed and elaborately regulated, with the vast majority of this considered to be noncoding RNA (ncRNA) genes (Consortium, 2012; Hon et al., 2017). The numbers of ncRNA genes vary between species, and interestingly, the complexity of an organism is

---

Abbreviations: 4E-BP, Eukaryotic translation initiation factor 4E (eIF4E)-binding protein; aa, amino acid; CCR2, C-C chemokine receptor type 2; ceRNAs, competing endogenous RNAs; circRNAs, circular RNAs; FBXW7, F-Box and WD repeat domain containing 7; Foxo, Forkhead box-o; HIST1H1C, histone cluster 1 H1 family member C; hnRNPC, heterogeneous nuclear ribonucleoprotein C; hTR, human telomerase RNA; IRES, internal ribosome entry site; lncRNA, long non-coding RNA; Mbl, Mannose-binding lectin; Mdm2, mouse double minute 2; MLN, myoregulin; MOTS-C, mitochondrial open reading frame of the 12S rRNA-c; MS, mass spectrometry; nt, nucleotide; ribo-seq, ribosome profiling sequencing; Scl, sarcolamb; SERCA, sarco/endoplasmic reticulum Ca2+-ATPase; sORF, small open reading frame; SR, sarcoplasmic reticulum; SRA, steroid receptor RNA activator; SRAP, steroid receptor RNA activator protein; Ube3a1, ubiquitin-protein ligase E3A; UTRs, untranslated regions; VEGF, vascular endothelial growth factor; ZNF609, zinc Finger Protein 609.

highly associated with the abundance of ncRNA genes but not protein-coding genes, implying the potential importance of ncRNAs (Rubin et al., 2000; Stover et al., 2000; Mattick, 2001; Venter et al., 2001; Kapusta and Feschotte, 2014). Among these, lncRNAs that are defined as transcripts longer than 200 nucleotides with low/no protein-coding potential, represent a considerable proportion.

Long non-coding RNAs can regulate gene expression in various ways, including epigenetic, transcriptional, post-transcriptional, translational and protein location effects. Corresponding to functional diversity, the modes of action of lncRNAs are also quite varied. lncRNAs can recruit epigenetic factors to modify chromatin state (Rinn and Chang, 2012), assemble transcriptional machinery to trigger the initiation of transcription (Bonasio and Shiekhattar, 2014), or act as a structural organizer to participate in the formation of subcellular organelles (Naganuma and Hirose, 2013). Additionally, lncRNAs can complementarily bind with other forms of RNA molecules to modulate gene expression at transcriptional, post-transcriptional and translational levels, for example as a moderator of mRNA activity or a decoy/sponge for miRNA (Poliseno et al., 2010; Gong and Maquat, 2011; Bonasio and Shiekhattar, 2014; Tay et al., 2014; Yoon et al., 2014). Moreover, lncRNAs couple with proteins through particular structures to act as a location transferor, or to modulate enzyme activities (Wang and Chang, 2011).

Based on the "noncoding" definition, the modes of action of lncRNAs mentioned above are exerted primarily through ncRNAs. Intriguingly, recent bioinformatics analyses of large-scale data from ribosome-protected RNA fragments (ribosome profiling or ribo-profiling) have revealed that a considerably large part of these transcripts tends to contain sORFs and binds with ribosomes (Aspden et al., 2014; Ruiz-Orera et al., 2014; Anderson et al., 2015; Mackowiak et al., 2015; Olexiouk et al., 2016), suggesting that the coding potential of lncRNAs has been vastly underestimated. Several functional experiments have demonstrated that some lncRNAs can encode small peptides (named "micropeptides" with a length less than 100 aa) that are involved in various biological processes, although this is rare (Hubé and Francastel, 2018). In addition, certain coding transcripts, such as TP53 mRNA, could also function as RNA, without translation to proteins, to regulate significant biological processes (Candeias, 2011; Kloc et al., 2011). Therefore, it seems reasonable to presume that the demarcation of RNA depending on its coding or noncoding status is somewhat blurred, and partially intertwined. That is, RNA roles are likely not tightly constrained (such as RNA functioning only as mRNA or ncRNA), but rather converge and overlap: lncRNAs can function by encoding small peptides, while mRNAs can use their special structural features, such as the 3′ UTR or 5′ UTR, to function (**Figure 1**).

In the present article, we will review current studies of the bilateral functionality of lncRNAs and mRNAs in terms of their coding potential, as well as the advancement of high-throughput techniques that would facilitate a deeper recognition of functional diversity of RNAs. This review will highlight the cases that illuminate the contrapositive roles between lncRNAs and mRNAs, and briefly discuss the biological significance of these discoveries for gene expression and evolution.

LONG NONCODING RNAs ENCODE SMALL PEPTIDES/PROTEINS WITH REGULATORY FUNCTIONS

Peptides/Proteins Encoded by Regular Long Noncoding RNAs

The original definition of lncRNAs concerns their low/non-coding potential. However, with accumulating evidence from bioinformatics and ribosome transcriptome profiling, lncRNAs have been shown to display strong ribosomal associations in many species, varying from plant to animal, indicating a potential coding capacity in lncRNA sORFs (Kageyama et al., 2011; Nam et al., 2016; Yeasmin et al., 2018). In recent years, several micropeptides derived from lncRNAs have been shown to be functional. We have summarized these micropeptides in **Table 1**.

Steroid receptor RNA activator is a prototypic example of lncRNAs with both coding and noncoding products (Lanz et al., 1999; Mattick, 2003; Hubé et al., 2006, 2011). *SRA* was initially identified as a noncoding gene with multiple RNA isoforms, which is critical in many biological processes, such as acting as a co-activator of nuclear receptors and a regulator of steroid receptor-dependent gene expression (Hubé et al., 2006, 2011; Cooper et al., 2011). Interestingly, *SRA* can also encode for a conserved SRAP, which, in turn, represses the transcriptional regulatory activity of the *SRA1* gene by interacting with a specific SRA stem-loop (Emberley et al., 2003; Chooniedass-Kothari et al., 2006; Hubé et al., 2011). The transmissible functionalities between the coding and noncoding *SRA* gene are caused by alternative splicing (AS) of introns/extrons (Colley and Leedman, 2011), suggesting the significance of AS events in the generation of bifunctional RNA.

Of note, among the small number of already-known functional micropeptides, a few are muscle-specific, and have been implicated in the regulation of the activities of SERCA (Anderson et al., 2016; Nelson et al., 2016; Matsumoto et al., 2017). For example, MLN, a 46-aa micropeptide specifically expressed in skeletal-muscle, is encoded by a lncRNA (*LINC00948* in human and *2310015B20Rik* in mouse); it can directly interact with SERCA to decrease the affinity of this ATPase for Ca^{2+} and inhibit Ca^{2+} entry into the SR (Anderson et al., 2015). The Scl micropeptide is encoded by the noncoding *pncr003:2L* gene, and can affect Ca^{2+} traffic in cardiac muscle in the fly; the mutation of this gene triggers an arrhythmic phenotype (Magny et al., 2013). The MOTS-C micropeptide can regulate insulin sensitivity and metabolic homeostasis in the mitochondria of muscle cells, and derives from mitochondrial 12S rRNA (Slavoff et al., 2014; Lee et al., 2015). In the above example, the Scl peptides and their respective regulatory functions in the heart are quite conserved between species, including the fly and humans (Magny et al., 2013). These results indicate that several sORFs embedded in the noncoding region of the genome seem to undergo a relatively stricter natural
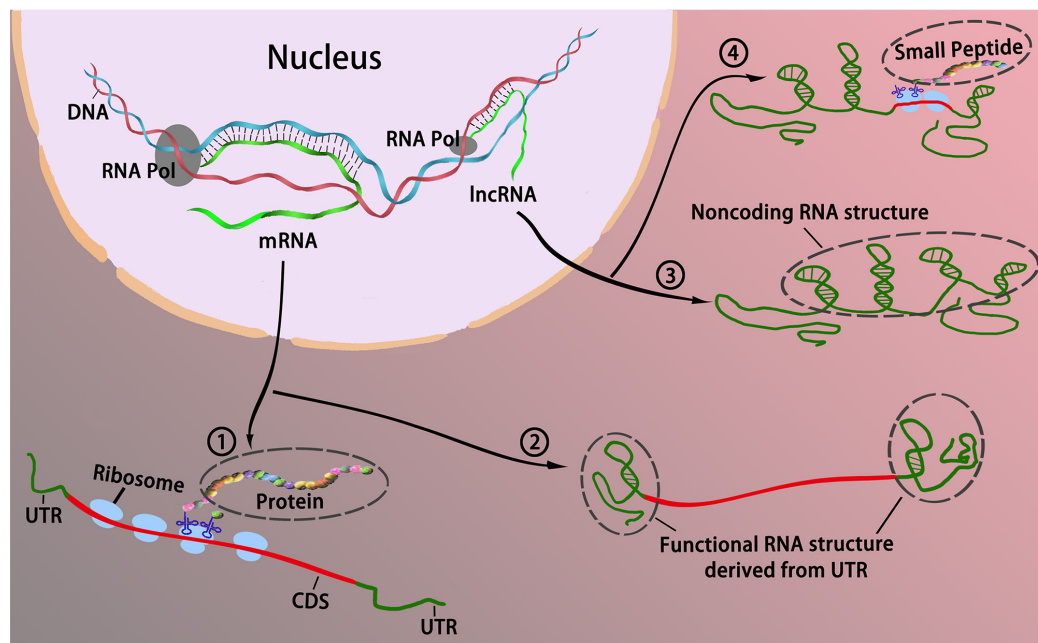
FIGURE 1 | The interchangeable roles between coding and long noncoding RNAs. Traditionally, RNAs could be divided into two categories in accordance with their coding potential, that is, coding RNAs and noncoding RNAs. Coding RNAs generally refers to mRNA that encodes protein ① to act as various components including enzymes, cell structures, and signal transductors. Noncoding RNAs act as cellular regulators without encoding proteins ③. However, it appears that the boundaries blur between coding RNA and noncoding RNA as some coding mRNAs can function without translating to protein via the formation of RNA secondary structure primarily derived from the UTR ②; some lncRNAs can bind with ribosomes, and encode peptides to modulate cellular activities ④.

selection than adjacent sequences, raising the question of whether these sORFs have a capability to sprout into a new gene *in situ* or to be integrated as a component into new genes elsewhere during evolution.

The *tal* gene in *Drosophila* is of vital importance in tarsal morphogenesis in the fly leg, and stage-and position-specific expression have been reported in embryonic development. Although *tal* is regarded as a noncoding gene as none of its ORFs are over 100 aa, deeper analysis has found that the functionality of *tal* is predominantly dependent on the ORF regions (Manak et al., 2006). There are five ORFs in the *tal* gene, four of which contain a similar and conserved 7 aa motif that determines the functionality of the gene, with the shortest peptide of only 11 aa. Phylogenetic analysis revealed that these *tal*-like peptides are conserved in metazoans and represent a new class of eukaryotic genes. The discovery of these mini-peptides further expands the possible scope and function of lncRNA-encoded peptides that are hidden in currently sequenced genomes and the transcriptome (Galindo et al., 2007).

Peptides/Proteins Encoded by Circular RNAs

Circular RNAs pertain to a sub-category of specialized lncRNAs, which are primarily produced by backsplicing the 3′ end to the 5′ end of exons in the same transcript (often a coding gene) via the spliceosome, thereby forming lncRNAs in a circular shape (Ashwal-Fluss et al., 2014; Zhang X.O. et al., 2014; Starke et al., 2015). Through bioinformatics analysis and high-throughput

sequencing, many circRNAs have been identified in multiple species (Sanger et al., 1976; Capel et al., 1993; Danan et al., 2012; Memczak et al., 2013; Jeck and Sharpless, 2014; Wang et al., 2014). However, understanding of circRNA function is still very limited. The reported biological activities of circRNAs include acting as a sponge for microRNAs (Hansen et al., 2013; Memczak et al., 2013), as a competitor during pre-mRNA splicing (Ashwal-Fluss et al., 2014), and as a transcriptional regulator in the nucleus (Li et al., 2015). The majority of circRNAs are chimeric lncRNAs derived from mRNA transcripts and likely in part encompass the exons of protein-coding genes. This poses the question of whether circRNAs have protein coding capabilities. In fact, many studies have demonstrated that circRNAs have coding capabilities both *in vitro* and *in vivo* in terms of cap-independent translation (Chen and Sarnow, 1995; Li and Lytton, 1999; Guo et al., 2014; Jeck and Sharpless, 2014; Abe et al., 2015; Wang and Wang, 2015; Pamudurti et al., 2017). Moreover, some functional protein products are encoded by circRNAs (such as *circ-FBXW7; circ-Mbl*, *circ-ZNF609* and *circ-SHPRH*) (Rybak-Wolf et al., 2015; Pamudurti et al., 2017; Yang et al., 2018; Zhang et al., 2018a).

circ-ZNF609 was initially screened out in a functional genetic screen, and is differentially expressed during myogenesis (Legnini et al., 2017). This circRNA contains an ORF covering almost all ORF regions of the host gene, but has a small variation at the splice junction. Its protein product lacks the zinc-finger domain compared with its linear counterpart, with an obvious impact on myoblast proliferation. Interestingly, heat shock could

TABLE 1 | Peptides encoded by lncRNAs in plants and animals.

Origin	Micropeptides	Gene	Function	Size (AAs)	References
Plants	Early nodulin 40 (Enod 40)	*Enod 40*	Nodule organogenesis	12; 24	Rohrig et al., 2002
Plants	POLARIS (PLS)	*POLARIS*	Leaf morphogenesis	36	Chilley et al., 2006
Plants	ROTUNDIFOLIA (ROT4)	*ROT4*	Leaf morphogenesis	53	Narita et al., 2004
Plants	ROT18/ DLV1	*DEVIL1 (DVL1)*	Plant organogenesis	51	Wen et al., 2004
Plants	Kiss of death (KOD)	*Kiss of death (KOD)*	Programmed cell death regulation	25	Blanvillain et al., 2011
Plants and animals	Brick1 (Brk)	*Brick1 (Brk)*	Leaf morphogenesis	84	Frank and Smith, 2002
Poaceae	Zm401p10; Zm908p11	*Zm401*	Pollen development	89; 97	Ma et al., 2008
Mammals	DWORF	*LOC100507537*	Enhance muscle performance	34	Nelson et al., 2016
Vertebrates	Toddler	*LOC100506013*	Promotes cell migration; activator of APJ/Apelin receptor signaling	54	Pauli et al., 2014
Vertebrates	Myomixer	*LOC101929726*	Functionally involve in controlling muscle performance	84	Bi et al., 2017
Fruit fly	Pri	*polished rice (pri)*	Epidermal morphogenesis in embryogenesis	11; 32	Kondo et al., 2007
Fruit fly	MOTS-c	*12s rRNA*	Insulin sensitivity and metabolic Homeostasis.	16	Lee et al., 2015
Fruit fly	Pgc polypeptide	*pgc*	Positivephosphorylation of transcription elongation factor b (P-TEFb)	71	Hanyu-Nakamura et al., 2008
Fruit fly	sarcolamban (Scl)	*pncr003:2L*	Calcium transport and muscle contraction,	28; 29	Magny et al., 2013
Fruit Fly	Tarsal-less/tal	*LP10384*	Morphogenesis, including tissue morphogenesis and pattern formation	11	Galindo et al., 2007
Mammals	PINT87aa	*LINC-PINT*	Inhibit the transcriptional elongation of multiple oncogenes	87	Zhang et al., 2018b
Mammals	Mitoregulin (Mtln)	*LINC00116*	Mitochondrial respiration, ROS, and Ca2+ retention capacity	56	Stein et al., 2018
Mammals	MRI-2	*C7orf49*	Non-homologous end-joining DNA repair	69	Slavoff et al., 2014
From humans to zebrafish	AGD3	*AGU1*	Human stem Cell differentiation	63	Kikuchi et al., 2009
Mammals	NoBody	*LINC01420/ LOC550643*	mRNA turnover and nonsense-mediated decay (NMD)	67	D'Lima et al., 2017
Mammals	Minion	*LOC101929726*	Muscle development	84	Zhang et al., 2017
Human and mouse	SPAR	*LINC00961*	Regulate muscle regeneration	90	Matsumoto et al., 2017
Primates	HOXB-AS3 peptide	*HOXB-AS3*	Suppresses colon cancer growth, PKM splicing and subsequent metabolic reprogramming	53	Huang et al., 2017
Different species	Humanin	*16s rRNA*	Program cell death	24	Hashimoto et al., 2001
Human	SRAP	*SRA*	Diverse roles in both normal biological processes and pathological changes	224; 236	Emberley et al., 2003; Chooniedass-Kothari et al., 2006
Mouse and human	Myoregulin (MLN)	*LINC00948* in human; *2310015B20Rik* in mouse	Inhibit the SERCA activity and regulate Ca^{2+} flow in muscle	46	Anderson et al., 2015

significantly activate the translation of *circ-ZNF609*; suggesting a possible regulatory role of circRNA translation under specific stimuli (Legnini et al., 2017).

circ-Mbl was first detected in the lodge of the second exon of the splicing factor muscleblind (MBL/MBNL1) in flies and humans, with a function of competing with pre-mRNA splicing (Ashwal-Fluss et al., 2014). Recently, through a bioinformatics analysis of ribosome foot-printing datasets, Pamudurti and co-workers revealed that *circ-Mbl* could encode a peptide in the fly head, as detected through MS. Both *circ-Mbl1* RNA and its protein-related product reside in the synaptosome and can be regulated by the 4E-BP and the transcription factor in forkhead

family – FOXO, suggesting that this circRNA translation might be distinctively important in the brain (Pamudurti et al., 2017).

The observation that circRNAs generate proteins can be traced back to much earlier studies in *Archaea*, where circularized introns produce a site-specific endonuclease (Dalgaard et al., 1993). However, to date, direct experimental evidence for circRNA translation to peptides is still scarce; as a result, it is even tougher to understand the function of their translated products. Considering that most circRNAs stem from coding transcripts and contain complete exons, it is possibly assumed that the circRNAs and their coding-products might provide uncharacterized modes of regulation of gene and

protein expression (Pamudurti et al., 2017). Therefore, it is important to further investigate the possible functions associated with circRNA coding.

Large-Scale Approaches for the Identification of Potential sORFs

To date, hundreds of thousands of lncRNAs have been discovered in various species, and there is a desire to study their relevant functional mechanisms (Okazaki et al., 2002; Liu et al., 2005; Kapranov et al., 2007; Ponting et al., 2009; Ulitsky and Bartel, 2013; Volders et al., 2013). However, it is unpractical to identify lncRNAs and predict their functions using only traditional technical approaches, irrespective of the requirement for intensive validation of the exact mechanisms underlying lncRNA activities. The same is true for the identification of lncRNA coding capacity. Therefore, new large-scale technologic approaches based on computational analysis of transcriptome data and proteomics data have been developed, all of which are mutually reinforcing and cross-validated.

A ribo-seq technique has been recently developed and is widely used to measure the full coding potential of RNA transcripts on a genome scale through deep sequencing of ribosome-protected RNA fragments (Ingolia et al., 2009). By identifying the precise ribosomal positions of RNAs, ribo-seq can plot the potential on-going events of translation in the cytosol, which is useful in identifying potentially functional micropeptides (Ingolia et al., 2011; Ingolia, 2016). With the advent of Ribo-Seq, thousands of translated sORFs were discovered in lncRNAs (Ingolia et al., 2011; Bazzini et al., 2014; Ruiz-Orera et al., 2014; Ji et al., 2015), with a few functional peptides, such as MLN (Anderson et al., 2015) and HOXB-AS3 (Huang et al., 2017). However, the proportion of coding lncRNAs estimated by various ribosome-profiling studies differ widely (Guttman et al., 2013; Ingolia et al., 2014), resulting from false positive and distinct prediction thresholds. Therefore, MS has emerged as a complementary method.

Mass spectrometry demonstrates excellent performance in detecting and characterizing the products of proteins/peptides in a complex biological sample. The detection of lncRNA-encoded peptides is the most direct evidence for lncRNA coding potential. However, to date, the proportion of coding lncRNAs detected by MS-based proteomes is small compared with that in ribo-seq results (Verheggen et al., 2017). The main weakness attributed to this approach is that MS-based proteomics is obviously impacted by the length and concentration of the detected samples. Therefore, specialized methods have been developed to circumvent these detection limitations. Short translation products at low abundance can surmount the threshold of MS detection through the use of peptidomics approaches (Schulz-Knappe et al., 2005) and enrichment protocols (Mustafa et al., 2015).

Both of the above techniques have their respective advantages and shortcomings; therefore, "proteogenomics" has been developed (Nesvizhskii, 2014; Menschaert and Fenyö, 2017; Ruggles et al., 2017). In proteogenomics, proteomics data are systematically integrated and analyzed with genomics and transcriptomic data generated from DNA-sequencing, RNA-sequencing and ribosome-profiling. The predicted sequences of proteins/peptides are tracked back to the genome and transcripts to identify the gene expression patterns and actual translational events. The significance of proteogenomics studies lies in improving genome annotation, and reasonably applying multi-omics data to explore complex and profound mechanisms in biological activities and complex diseases (Zhang B. et al., 2014; Zhang et al., 2016; Mertins et al., 2016).

NONCODING RNA REGULATORY FUNCTIONS EMBEDDED IN mRNAs

3′ UTR Regulatory Roles of mRNAs

Based on current research results, the noncoding regulatory functions discovered in mRNAs are mainly present in the 3′ UTRs, which were previously supposed to be the vital regulative elements for mRNA self-stability and location. Compared with highly conserved coding regions that have to undergo strictly selected pressure, the 3′ UTR displays more flexibility and plasticity between species. Its size varies from a few to hundreds of nucleotides, and likely has a close relationship with biological complexity (Chen et al., 2012; Mayr, 2016). Moreover, for an RNA molecule, other than the impact on base pairing, the changes in sequence are most likely to induce corresponding changes in structure, resulting in information transmitted from RNA to protein (Berkovits and Mayr, 2015).

By comprehensively estimating up-to-date cases where mRNAs regulate biologic activities without translating to protein, we found that the 3′ UTR of mRNA plays a large role as an effectors. Increasing evidence has demonstrated that the 3′ UTRs of mRNAs are actively involved in repressing the occurrence and progression of cancer cells, such as the 3′ UTRs of α-tropomyosin mRNA, prohibitin mRNA and ribonucleotide reductase mRNA (Rastinejad et al., 1993; Fan et al., 1996; Manjeshwar et al., 2003). These studies demonstrate that the 3′ UTRs of some mRNAs can antagonize tumor development, likely through RNA interactions with regulatory factors involved in cellular growth in a post-transcriptional pattern. Indeed, the 3′ UTR can recruit RNA-binding proteins, as in the case of CD47 mRNA. CD47 mRNA has two isoforms of the 3′ UTR, long (CD47-LU) and short (CD47-SU), and only the CD47-LU, which is AU-rich, can interact with the RNA-binding protein TIS11B to form a membraneless organelle with a specific biochemical and biophysical environment which is separate from the cytosol (Ma and Mayr, 2018). However, the most prevalent mode of action of the 3′ UTR is as ceRNAs, such as in the cases of CCR2 mRNA and Ube3a1 RNA, which confer to the function of lncRNAs (Valluy et al., 2015; Hu et al., 2017).

Noncoding Regulatory Roles of mRNA Not Involving the 3′ UTR

Other than the 3′ UTR, the 5′ UTR and ORF can also be involved in RNA-mediated regulatory function, although recent reports of this phenomenon are scare. Two mRNAs, TP53 mRNA

and *HIST1H1C* mRNA, are recognized as being involved in ORF-mediated regulation. TP53 protein is a tumor suppressor implicated in many processes during tumor occurrence and development. However, a triple synonymous mutant (TriMp53) in codons led to a misshapen structure, resulting in loss of the IRES activity of p47 (one isoform of p53) and an abrogated affinity of hnRNPC, but with better binding to Mdm2, which is an E3 ubiquitin-protein ligase in mediating p53/TP53 ubiquitination, and an augmented ability of p53 to activate apoptosis. These facts indicate that TP53 has intricate regulatory roles at both the RNA and protein levels, suggesting that the functions of the RNA and protein molecules are closely intertwined (Candeias, 2011). *HIST1H1C* mRNA participates in regulating telomere length homeostasis. Aside from the protein-related product, a 15-nt long region in the ORF region (nt334–nt348) is attributed to *HIST1H1C*-mRNA-mediated biological activity, through complementation with the terminal stem-loop sequence of the P6b region of hTR, in a base-pairing pattern. These results extend the functional potency of mRNA ORF regions in a non-traditional and noncoding direction (Ivanyi-Nagy et al., 2018).

In terms of the 5′ UTR, there are only two examples. VEGF is a key regulator of angiogenesis during embryonal and cancerous development, and this regulatory function is closely correlated with the 5′ UTR. *vegf* mRNA has an unusually long 5′ UTR of 1,038 nucleotides, and contains two IRES, resulting in an intricate regulation of VEGF expression. In addition, the presence of the 5′ UTR of *vegf* mRNA alone in tumor cells could promote the expression of anti-apoptotic genes but repress pro-apoptotic genes, suggesting an anti-apoptotic role of the *vegf* 5′ UTR, and demonstrating its potential as a target for cancer treatment. To the best of our knowledge, the 5′ UTR of *vegf* mRNA represents the only example of an mRNA UTR which can promote tumor progression (Akiri et al., 1998; Huez et al., 1998; Masuda et al., 2008).

The *c-myc P0* transcript is an isoform transcript from the promoter 0 (P0) of the c-myc gene, which has an extra ˜639-nucleotide extension of the 5′ UTR when compared with two major isoforms (P1 and P2) of *c-myc* mRNA. Ectopic expression of the 5′ UTR of the *c-myc P0* transcript alone in HeLa cells results in significantly increased expression of the c-Myc1 (p67) and c-Myc2 (p64) proteins as well as incremental apoptosis sensitivity, but decreased tumorigenicity, all of which are likely attributable to competitive regulation of gene expression in the c-myc locus. These results demonstrate that the 5′ UTR potentially functions *in trans* to perform gene regulation (Blume et al., 2003).

PERSPECTIVE

In recent years, researchers have begun to pay close attention to the development of bifunctional RNAs, and have discussed the evolved roles of RNAs with multiple functions (Dinger et al., 2011; Ulveling et al., 2011; Kageyama et al., 2011; Hubé and Francastel, 2018). In the early stages of such research, researchers discovered individual gene on a case-by-case basis.

However, in the last ten years, rapid advances in large-scale detection and identification techniques (such as ribo-seq and MS-based proteomics) have facilitated multi-faceted investigations of genomes and vital processes, thus shedding light on the complex activities of various RNA molecules. Bifunctional RNAs raise questions about the concept of a gene, in terms of whether RNA, both coding and noncoding, is an independent gene type or a convergence of coding and noncoding genes which occurred during evolution. In this review, we intend to not only investigate the current status of bifunctional RNAs as reported in recent years, but also discuss the potential pervasiveness of bifunctional RNAs from a global perspective in terms of large-scale data.

With recent estimates of ribosome profiling, small peptides encoded by lncRNAs have significantly expanded the extent and diversity of the proteome, and predictions suggested that a large fraction of the annotated lncRNAs in various eukaryotic organisms would be translated with sORFs (Aspden et al., 2014; Ruiz-Orera et al., 2014; Mackowiak et al., 2015; Olexiouk et al., 2016). Proteogenomic evidence has confirmed that many small peptides which stem from regions of lncRNA genes are expressed differentially in different cell types and during different developmental/disease stages, although their functions are somewhat enigmatic (Nesvizhskii, 2014; Zhu et al., 2018). However, other studies have revealed that mRNAs could also be involved in cellular regulatory processes in a coding-independent manner (Nam et al., 2016). The results from a large-scale RNA structure analysis revealed that the secondary structures of mRNA have an essential regulatory effect on its maturation and stability, even for the evolutionarily conserved RNA silencing pathways of eukaryotes, suggesting that mRNAs partially retain the functionality of structure that exists in many RNA molecules (Katz and Burge, 2003; Li et al., 2012; Taggart et al., 2012). All these facts indicate that the coding potential and biological roles of mRNAs and lncRNAs could be switched in some cases, implying a conceptual blurring between coding and noncoding genes.

Many lncRNAs share similar features with classical mRNAs, such as transcription by polymerase II with a 5′-cap and 3′-polyadenylated tail, and frequent accumulation in the cytoplasm (van Heesch et al., 2014). Therefore, when associated with ribosomes, sORFs embedded in lncRNAs have a significant chance to be translated to peptides. The peptides derived from lncRNAs have a relatively shorter chain length and weaker conservation across different species, and this is consistent with the original lncRNAs which often have few introns, a low expression level and weak phylogenetic conservation (Cabili et al., 2011; Derrien et al., 2012; Kutter et al., 2012; Necsulea et al., 2014). From the perspective of proteins driving evolution, these peptides are likely considered to be an important source for new protein (Ruiz-Orera et al., 2014). Previously reported experimental evidence indicates that noncoding RNAs expressed at low levels could contribute to the birth of novel protein coding genes (Levine et al., 2006; Cai et al., 2008; Reinhardt et al., 2013). Given that several lncRNA-derived peptides have been demonstrated to play essential roles in many biological activities, it

is worth investigating the putative significance of the generation of these lncRNA-derived peptides in gene evolution, expression and regulation.

However, in view of the huge quantity, diversified mechanisms of action, and intricate functions of lncRNAs, it is inappropriate to regard lncRNAs just as a pool for evolved peptides. In terms of RNA alone, its roles are diverse, including the potential to be retro-transcribed into DNA, or to act as an enzyme to participate in complex biochemical processes (Cech, 1986). Moreover, random RNA sequences can inoculate structurally complex and highly active RNA ligases, suggesting that randomness can produce functionality (Ekland et al., 1995). Therefore, it is very likely that RNA molecules alone comprise abundant genetic information, such as particular structural features and ultra-conservative sequence elements, which could regulate the timing and place of gene expression during cellular differentiation and development.

In recent decades, because of the addition of the huge family of noncoding genes, RNAs have provoked great interest for their mysterious roles in organisms. lncRNA-encoded peptides expand the horizon of functional mechanisms for these bio-macromolecules. To date, thousands of peptide products have been identified in human cells, with limited understanding of their function. The current review has summarized the recently discovered micropeptides implicated in various biological processes. We also discussed the potential noncoding roles of mRNAs as a regulator. The continued discovery and functional characterization of bifunctional RNAs will provide new insights into important cellular processes and organismal evolution.

AUTHOR CONTRIBUTIONS

Both authors have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Abe, N., Matsumoto, K., Nishihara, M., Nakano, Y., Shibata, A., Maruyama, H., et al. (2015). Rolling circle translation of circular RNA in living human cells. *Sci. Rep.* 5:16435. doi: 10.1038/srep16435

Akiri, G., Nahari, D., Finkelstein, Y., Le, S. Y., Elroy-Stein, O., and Levi, B. Z. (1998). Regulation of vascular endothelial growth factor (VEGF) expression is mediated by internal initiation of translation and alternative initiation of transcription. *Oncogene* 17, 227–236. doi: 10.1038/sj.onc.1202019

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 160, 595–606. doi: 10.1016/j.cell.2015.01.009

Anderson, D. M., Makarewich, C. A., Anderson, K. M., Shelton, J. M., Bezprozvannaya, S., Bassel-Duby, R., et al. (2016). Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. *Sci. Signal.* 9:ra119. doi: 10.1126/scisignal.aaj1460

Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., et al. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56, 55–66. doi: 10.1016/j.molcel.2014.08.019

Aspden, J. L., Eyre-Walker, Y. C., Phillips, R. J., Amin, U., Mumtaz, M. A., Brocard, M., et al. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *eLife* 3:e03528. doi: 10.7554/eLife.03528

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 33, 981–993. doi: 10.1002/embj.201488411

Berkovits, B. D., and Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature* 522, 363–367. doi: 10.1002/embj.201488411

Bi, P., Ramirez-Martinez, A., Li, H., Cannavino, J., McAnally, J. R., Shelton, J. M., et al. (2017). Control of muscle formation by the fusogenic micropeptide myomixer. *Science* 356, 323–327. doi: 10.1126/science.aam9361

Blanvillain, R., Young, B., Cai, Y. M., Hecht, V., Varoquaux, F., Delorme, V., et al. (2011). The *Arabidopsis* peptide kiss of death is an inducer of programmed cell death. *EMBO J.* 30, 1173–1183. doi: 10.1038/emboj.2011.14

Blume, S. W., Miller, D. M., Guarcello, V., Shrestha, K., Meng, Z., Snyder, R. C., et al. (2003). Inhibition of tumorigenicity by the 5′-untranslated RNA of the human c-myc P0 transcript. *Exp. Cell Res.* 288, 131–142. doi: 10.1016/S0014-4827(03)00182-4

Bonasio, R., and Shiekhattar, R. (2014). Regulation of transcription by long noncoding RNAs. *Annu. Rev. Genet.* 48, 433–455. doi: 10.1146/annurev-genet-120213-092323

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611

Cai, J., Zhao, R., Jiang, H., and Wang, W. (2008). De novo origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* 179, 487–496. doi: 10.1534/genetics.107.084491

Candeias, M. M. (2011). The can and can't dos of p53 RNA. *Biochimie* 93, 1962–1965. doi: 10.1016/j.biochi.2011.06.010

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., et al. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* 73, 1019–1030. doi: 10.1016/0092-8674(93)90279-Y

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563. doi: 10.1126/science.1112014

Cech, T. R. (1986). A model for the RNA-catalyzed replication of RNA. *Proc. Natl. Acad. Sci. U.S.A.* 83, 4360–4363. doi: 10.1073/pnas.83.12.4360

Chen, C. Y., Chen, S. T., Juan, H. F., and Huang, H. C. (2012). Lengthening of 3′UTR increases with morphological complexity in animal evolution. *Bioinformatics* 28, 3178–3181. doi: 10.1093/bioinformatics/bts623

Chen, C. Y., and Sarnow, P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* 268, 415–417. doi: 10.1126/science.7536344

Chilley, P. M., Casson, S. A., Tarkowski, P., Hawkins, N., Wang, K. L., Hussey, P. J., et al. (2006). The POLARIS peptide of *Arabidopsis* regulates auxin transport and root growth via effects on ethylene signaling. *Plant Cell* 18, 3058–3072. doi: 10.1105/tpc.106.040790

Chooniedass-Kothari, S., Hamedani, M. K., Troup, S., Hube, F., and Leygue, E. (2006). The steroid receptor RNA activator protein is expressed in breast tumor tissues. *Int. J. Cancer* 118, 1054–1059. doi: 10.1002/ijc.21425

Colley, S. M., and Leedman, P. J. (2011). Steroid Receptor RNA Activator - A nuclear receptor coregulator with multiple partners: insights and challenges. *Biochimie* 93, 1966–1972. doi: 10.1016/j.biochi.2011.07.004

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

Cooper, C., Vincett, D., Yan, Y., Hamedani, M. K., Myal, Y., and Leygue, E. (2011). Steroid Receptor RNA Activator bi-faceted genetic system: heads or Tails? *Biochimie* 93, 1973–1980. doi: 10.1016/j.biochi.2011.07.002

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi: 10.1038/227561a0

Dalgaard, J. Z., Garrett, R. A., and Belfort, M. (1993). A site-specific endonuclease encoded by a typical archaeal intron. *Proc. Natl. Acad. Sci. U.S.A.* 90, 5414–5417. doi: 10.1073/pnas.90.12.5414

Danan, M., Schwartz, S., Edelheit, S., and Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* 40, 3131–3142. doi: 10.1093/nar/gkr1009

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

Dinger, M. E., Gascoigne, D. K., and Mattick, J. S. (2011). The evolution of RNAs with multiple functions. *Biochimie* 93, 2013–2018. doi: 10.1016/j.biochi.2011.07.018

D'Lima, N. G., Ma, J., Winkler, L., Chu, Q., Loh, K. H., Corpuz, E. O., et al. (2017). A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* 13, 174–180. doi: 10.1038/nchembio.2249

Ekland, E. H., Szostak, J. W., and Bartel, D. P. (1995). Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* 269, 364–370. doi: 10.1126/science.7618102

Emberley, E., Huang, G. J., Hamedani, M. K., Czosnek, A., Ali, D., Grolla, A., et al. (2003). Identification of new human coding steroid receptor RNA activator isoforms. *Biochem. Biophys. Res. Commun.* 301, 509–515. doi: 10.1016/s0006-291x(02)03070-x

Fan, H., Villegas, C., Huang, A., and Wright, J. A. (1996). Suppression of malignancy by the 3′ untranslated regions of ribonucleotide reductase R1 and R2 messenger RNAs. *Cancer Res.* 56, 4366–4369.

Frank, M. J., and Smith, L. G. (2002). A small, novel protein highly conserved in plants and animals promotes the polarized growth and division of maize leaf epidermal cells. *Curr. Biol.* 12, 849–853. doi: 10.1016/s0960-9822(02)00819-9

Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A., and Couso, J. P. (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* 5:e106. doi: 10.1371/journal.pbio.0050106

Gong, C., and Maquat, L. E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* 470, 284–288. doi: 10.1038/nature09701

Guo, J. U., Agarwal, V., Guo, H., and Bartel, D. P. (2014). Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* 15:409. doi: 10.1186/s13059-014-0409-z

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., and Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240–251. doi: 10.1016/j.cell.2013.06.009

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., and Nakamura, A. (2008). *Drosophila* Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature* 451, 730–733. doi: 10.1038/nature06498

Hashimoto, Y., Niikura, T., Tajima, H., Yasukawa, T., Sudo, H., Ito, Y., et al. (2001). A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Abeta. *Proc. Natl. Acad. Sci. U.S.A.* 98, 6336–6341. doi: 10.1073/pnas.101133498

Hon, C. C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J., Gough, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature* 543, 199–204. doi: 10.1038/nature21374

Hu, J., Li, X., Guo, X., Guo, Q., Xiang, C., Zhang, Z., et al. (2017). The CCR2 3′UTR functions as a competing endogenous RNA to inhibit breast cancer metastasis. *J. Cell Sci.* 130, 3399–3413. doi: 10.1242/jcs.202127

Huang, J. Z., Chen, M., Chen, X., Gao, C., Zhu, S., Huang, H., et al. (2017). A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol. Cell* 68, 171–184.e6. doi: 10.1016/j.molcel.2017.09.015

Hubé, F., and Francastel, C. (2018). Coding and Non-coding RNAs, the frontier has never been so blurred. *Front. Genet.* 9:140. doi: 10.3389/fgene.2018.00140

Hubé, F., Guo, J., Chooniedass-Kothari, S., Cooper, C., Hamedani, M. K., Dibrov, A. A., et al. (2006). Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines. *DNA Cell Biol.* 25, 418–428. doi: 10.1089/dna.2006.25.418

Hubé, F., Velasco, G., Rollin, J., Furling, D., and Francastel, C. (2011). Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. *Nucleic Acids Res.* 39, 513–525. doi: 10.1093/nar/gkq833

Huez, I., Creancier, L., Audigier, S., Gensac, M. C., Prats, A. C., and Prats, H. (1998). Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Mol. Cell Biol.* 18, 6178–6190. doi: 10.1128/mcb.18.11.6178

Ingolia, N. T. (2016). Ribosome footprint profiling of translation throughout the genome. *Cell* 165, 22–33. doi: 10.1016/j.cell.2016.02.066

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J., Jackson, S. E., et al. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 8, 1365–1379. doi: 10.1016/j.celrep.2014.07.045

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., and Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223. doi: 10.1126/science.1168978

Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802. doi: 10.1016/j.cell.2011.10.002

Ivanyi-Nagy, R., Ahmed, S. M., Peter, S., Ramani, P. D., Ong, P. F., Dreesen, O., et al. (2018). The RNA interactome of human telomerase RNA reveals a coding-independent role for a histone mRNA in telomere homeostasis. *eLife* 7:e40037. doi: 10.7554/eLife.40037

Jeck, W. R., and Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 32, 453–461. doi: 10.1038/nbt.2890

Ji, Z., Song, R., Regev, A., and Struhl, K. (2015). Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* 4:e08890. doi: 10.7554/eLife.08890

Kageyama, Y., Kondo, T., and Hashimoto, Y. (2011). Coding vs non-coding: translatability of short ORFs found in putative non-coding transcripts. *Biochimie* 93, 1981–1986. doi: 10.1016/j.biochi.2011.06.024

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. doi: 10.1126/science.1138341

Kapusta, A., and Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends Genet.* 30, 439–452. doi: 10.1016/j.tig.2014.08.004

Katz, L., and Burge, C. B. (2003). Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.* 13, 2042–2051. doi: 10.1101/gr.1257503

Kikuchi, K., Fukuda, M., Ito, T., Inoue, M., Yokoi, T., Chiku, S., et al. (2009). Transcripts of unknown function in multiple-signaling pathways involved in human stem cell differentiation. *Nucleic Acids Res.* 37, 4987–5000. doi: 10.1093/nar/gkp426

Kloc, M., Foreman, V., and Reddy, S. A. (2011). Binary function of mRNA. *Biochimie* 93, 1955–1961. doi: 10.1016/j.biochi.2011.07.008

Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y. (2007). Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.* 9, 660–665. doi: 10.1038/ncb1595

Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., et al. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* 8:e1002841. doi: 10.1371/journal.pgen.1002841

Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature* 470, 187–197. doi: 10.1038/nature09792

Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., et al. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97, 17–27. doi: 10.1016/S0092-8674(00)80711-4

Lee, C., Zeng, J., Drew, B. G., Sallam, T., Martin-Montalvo, A., Wan, J., et al. (2015). The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab.* 21, 443–454. doi: 10. 1016/j.cmet.2015.02.009

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell* 66, 22–37.e9. doi: 10.1016/j.molcel.2017.02.017

Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A., and Begun, D. J. (2006). Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9935–9939. doi: 10.1073/pnas.0509809103

Li, F., Zheng, Q., Vandivier, L. E., Willmann, M. R., Chen, Y., and Gregory, B. D. (2012). Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. *Plant Cell* 24, 4346–4359. doi: 10.1105/tpc.112.104232

Li, X. F., and Lytton, J. (1999). A circularized sodium-calcium exchanger exon 2 transcript. *J. Biol. Chem.* 274, 8153–8160. doi: 10.1074/jbc.274.12.8153

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., et al. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* 22, 256–264. doi: 10.1038/nsmb.2959

Liu, C., Bai, B., Skogerbo, G., Cai, L., Deng, W., Zhang, Y., et al. (2005). NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.* 33, D112–D115. doi: 10.1093/nar/gki041

Ma, J., Yan, B., Qu, Y., Qin, F., Yang, Y., Hao, X., et al. (2008). Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize. *J. Cell. Biochem.* 105, 136–146. doi: 10.1002/jcb.21807

Ma, W., and Mayr, C. (2018). A membraneless organelle associated with the endoplasmic reticulum enables 3′UTR-mediated protein-protein interactions. *Cell* 175, 1492–1506.e19. doi: 10.1016/j.cell.2018.10.007

Mackowiak, S. D., Zauber, H., Bielow, C., Thiel, D., Kutz, K., Calviello, L., et al. (2015). Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol.* 16:179. doi: 10.1186/s13059-015-0742-x

Magny, E. G., Pueyo, J. I., Pearl, F. M., Cespedes, M. A., Niven, J. E., Bishop, S. A., et al. (2013). Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* 341, 1116–1120. doi: 10.1126/science.1238802

Manak, J. R., Dike, S., Sementchenko, V., Kapranov, P., Biemar, F., Long, J., et al. (2006). Biological function of unannotated transcription during the early development of *Drosophila melanogaster*. *Nat. Genet.* 38, 1151–1158. doi: 10. 1038/ng1875

Manjeshwar, S., Branam, D. E., Lerner, M. R., Brackett, D. J., and Jupe, E. R. (2003). Tumor suppression by the prohibitin gene 3′untranslated region RNA in human breast cancer. *Cancer Res.* 63, 5251–5256.

Masuda, K., Teshima-Kondo, S., Mukaijo, M., Yamagishi, N., Nishikawa, Y., Nishida, K., et al. (2008). A novel tumor-promoting function residing in the 5′ non-coding region of vascular endothelial growth factor mRNA. *PLoS Med.* 5:e94. doi: 10.1371/journal.pmed.0050094

Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., et al. (2017). mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* 541, 228–232. doi: 10.1038/nature21034

Mattick, J. S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* 2, 986–991. doi: 10.1093/embo-reports/kve230

Mattick, J. S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* 25, 930–939. doi: 10.1002/bies.10332

Mayr, C. (2016). Evolution and biological roles of alternative 3′UTRs. *Trends Cell Biol.* 26, 227–237. doi: 10.1016/j.tcb.2015.10.012

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature11928

Menschaert, G., and Fenyö, D. (2017). Proteogenomics from a bioinformatics angle: a growing field. *Mass Spectrom. Rev.* 36, 584–599. doi: 10.1002/mas.21483

Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 534, 55–62. doi: 10.1038/nature18003

Mustafa, G. M., Larry, D., Petersen, J. R., and Elferink, C. J. (2015). Targeted proteomics for biomarker discovery and validation of hepatocellular carcinoma in hepatitis C infected patients. *World J. Hepatol.* 7, 1312–1324. doi: 10.4254/wjh.v7.i10.1312

Naganuma, T., and Hirose, T. (2013). Paraspeckle formation during the biogenesis of long noncoding RNAs. *RNA Biol.* 10, 456–461. doi: 10.4161/rna.23547

Nam, J. W., Choi, S. W., and You, B. H. (2016). Incredible RNA: dual functions of coding and noncoding. *Mol. Cells* 39, 367–374. doi: 10.14348/molcells.2016.0039

Narita, N. N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., et al. (2004). Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in *Arabidopsis thaliana*. *Plant J.* 38, 699–713. doi: 10.1111/j.1365-313X.2004.02078.x

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640. doi: 10.1038/nature12943

Nelson, B. R., Makarewich, C. A., Anderson, D. M., Winders, B. R., Troupes, C. D., Wu, F., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 351, 271–275. doi: 10.1126/science.aad4076

Nesvizhskii, A. I. (2014). Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* 11, 1114–1125. doi: 10.1038/nmeth.3144

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., et al. (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563–573. doi: 10.1038/nature01266

Olexiouk, V., Crappe, J., Verbruggen, S., Verhegen, K., Martens, L., and Menschaert, G. (2016). sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* 44, D324–D329. doi: 10.1093/nar/gkv1175

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). Translation of CircRNAs. *Mol. Cell* 66, 9–21.e7. doi: 10.1016/j.molcel.2017.02.021

Pauli, A., Norris, M. L., Valen, E., Chew, G. L., Gagnon, J. A., Zimmerman, S., et al. (2014). Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science* 343:1248636. doi: 10.1126/science.1248636

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038. doi: 10.1038/nature09144

Ponting, C. P., Oliver, P. L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell* 136, 629–641. doi: 10.1016/j.cell.2009.02.006

Rastinejad, F., Conboy, M. J., Rando, T. A., and Blau, H. M. (1993). Tumor suppression by RNA from the 3′ untranslated region of alpha-tropomyosin. *Cell* 75, 1107–1117. doi: 10.1016/0092-8674(93)90320-P

Reinhardt, J. A., Wanjiru, B. M., Brant, A. T., Saelao, P., Begun, D. J., and Jones, C. D. (2013). *De novo* ORFs in Drosophila are important to organismal fitness and evolved rapidly from previously noncoding sequences. *PLoS Genet.* 9:e1003860. doi: 10.1371/journal.pgen.1003860

Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166. doi: 10.1146/annurev-biochem-051410-092902

Rohrig, H., Schmidt, J., Miklashevichs, E., Schell, J., and John, M. (2002). Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1915–1920. doi: 10.1073/pnas.022664799

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., et al. (2000). Comparative genomics of the eukaryotes. *Science* 287, 2204–2215. doi: 10.1126/science.287.5461.2204

Ruggles, K. V., Krug, K., Wang, X., Clauser, K. R., Wang, J., Payne, S. H., et al. (2017). Methods, tools and current perspectives in proteogenomics. *Mol. Cell. Proteomics* 16, 959–981. doi: 10.1074/mcp.MR117.000024

Ruiz-Orera, J., Messeguer, X., Subirana, J. A., and Alba, M. M. (2014). Long non-coding RNAs as a source of new peptides. *eLife* 3:e03523. doi: 10.7554/eLife.03523

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, 870–885. doi: 10.1016/j.molcel.2015.03.027

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

Schulz-Knappe, P., Schrader, M., and Zucht, H. D. (2005). The peptidomics concept. *Comb. Chem. High Throughput Screen.* 8, 697–704. doi: 10.2174/1386207054764418

Slavoff, S. A., Heo, J., Budnik, B. A., Hanakahi, L. A., and Saghatelian, A. (2014). A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. *J. Biol. Chem.* 289, 10950–10957. doi: 10.1074/jbc.C113.533968

Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L. H., et al. (2015). Exon circularization requires canonical splice signals. *Cell Rep.* 10, 103–111. doi: 10.1016/j.celrep.2014.12.002

Stein, C. S., Jadiya, P., Zhang, X., McLendon, J. M., Abouassaly, G. M., Witmer, N. H., et al. (2018). Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency. *Cell Rep.* 23, 3710–3720.e8. doi: 10.1016/j.celrep.2018.06.002

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1: an opportunistic pathogen. *Nature* 406, 959–964. doi: 10.1038/35023079

Taggart, A. J., DeSimone, A. M., Shih, J. S., Filloux, M. E., and Fairbrother, W. G. (2012). Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. *Nat. Struct. Mol. Biol.* 19, 719–721. doi: 10.1038/nsmb.2327

Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505, 344–352. doi: 10.1038/nature12986

Ulitsky, I., and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46. doi: 10.1016/j.cell.2013.06.020

Ulveling, D., Francastel, C., and Hubé, F. (2011). Identification of potentially new bifunctional RNA based on genome-wide data-mining of alternative splicing events. *Biochimie* 93, 2024–2027. doi: 10.1016/j.biochi.2011.06.019

Valluy, J., Bicker, S., Aksoy-Aksel, A., Lackinger, M., Sumer, S., Fiore, R., et al. (2015). A coding-independent function of an alternative Ube3a transcript during neuronal development. *Nat. Neurosci.* 18, 666–673. doi: 10.1038/nn.3996

van Heesch, S., van Iterson, M., Jacobi, J., Boymans, S., Essers, P. B., de Bruijn, E., et al. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* 15:R6. doi: 10.1186/gb-2014-15-1-r6

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351. doi: 10.1126/science.1058040

Verheggen, K., Volders, P. J., Mestdagh, P., Menschaert, G., Van Damme, P., Gevaert, K., et al. (2017). Noncoding after all: biases in proteomics data do not explain observed absence of lncRNA translation products. *J. Proteome Res.* 16, 2508–2515. doi: 10.1021/acs.jproteome.7b00085

Volders, P. J., Helsens, K., Wang, X., Menten, B., Martens, L., Gevaert, K., et al. (2013). LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* 41, D246–D251. doi: 10.1093/nar/gks915

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Wang, P. L., Bao, Y., Yee, M. C., Barrett, S. P., Hogan, G. J., Olsen, M. N., et al. (2014). Circular RNA is expressed across the eukaryotic tree of life. *PLoS One* 9:e90859. doi: 10.1371/journal.pone.0090859

Wang, Y., and Wang, Z. (2015). Efficient backsplicing produces translatable circular mRNAs. *RNA* 21, 172–179. doi: 10.1261/rna.048272.114

Wen, J., Lease, K. A., and Walker, J. C. (2004). DVL, a novel class of small polypeptides: overexpression alters *Arabidopsis* development. *Plant J.* 37, 668–677. doi: 10.1111/j.1365-313X.2003.01994.x

Yang, Y., Gao, X., Zhang, M., Yan, S., Sun, C., Xiao, F., et al. (2018). Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis. *J. Natl. Cancer Inst.* 110, 304–315. doi: 10.1093/jnci/djx166

Yeasmin, F., Yada, T., and Akimitsu, N. (2018). Micropeptides encoded in transcripts previously identified as long noncoding RNAs: a new chapter in transcriptomics and proteomics. *Front. Genet.* 9:144. doi: 10.3389/fgene.2018.00144

Yoon, J. H., Abdelmohsen, K., and Gorospe, M. (2014). Functional interactions among microRNAs and long noncoding RNAs. *Semin. Cell Dev. Biol.* 34, 9–14. doi: 10.1016/j.semcdb.2014.05.015

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., et al. (2014). Proteogenomic characterization of human colon and rectal cancer. *Nature* 513, 382–387. doi: 10.1038/nature13438

Zhang, H., Liu, T., Zhang, Z., Payne, S. H., Zhang, B., McDermott, J. E., et al. (2016). Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* 166, 755–765. doi: 10.1016/j.cell.2016.05.069

Zhang, M., Huang, N., Yang, X., Luo, J., Yan, S., Xiao, F., et al. (2018a). A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene* 37, 1805–1814. doi: 10.1038/s41388-017-0019-9

Zhang, M., Zhao, K., Xu, X., Yang, Y., Yan, S., Wei, P., et al. (2018b). A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat. Commun.* 9:4475. doi: 10.1038/s41467-018-06862-2

Zhang, Q., Vashisht, A. A., O'Rourke, J., Corbel, S. Y., Moran, R., Romero, A., et al. (2017). The microprotein Minion controls cell fusion and muscle formation. *Nat. Commun.* 8:15664. doi: 10.1038/ncomms15664

Zhang, X. O., Wang, H. B., Zhang, Y., Lu, X., Chen, L. L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. *Cell* 159, 134–147. doi: 10.1016/j.cell.2014.09.001

Zhu, Y., Orre, L. M., Johansson, H. J., Huss, M., Boekel, J., Vesterlund, M., et al. (2018). Discovery of coding regions in the human genome by integrated proteogenomics analysis workflow. *Nat. Commun.* 9:903. doi: 10.1038/s41467-018-03311-y

LncRNA-Disease Association Prediction Using Two-Side Sparse Self-Representation

Le Ou-Yang [1,2†], Jiang Huang [3†], Xiao-Fei Zhang [4], Yan-Ran Li [3], Yiwen Sun [5], Shan He [6] and Zexuan Zhu [3*]

[1] Guangdong Key Laboratory of Intelligent Information Processing and Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen, China, [2] FJKLMAA (Fujian Key Laborotary of Mathematical Analysis and Applications), Fujian Normal University, Fuzhou, China, [3] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, [4] School of Mathematics and Statistics and Hubei Key Laboratory of Mathematical Sciences, Central China Normal University, Wuhan, China, [5] School of Medicine, Shenzhen University, Shenzhen, China, [6] School of Computer Science, University of Birmingham, Birmingham, United Kingdom

Evidences increasingly indicate the involvement of long non-coding RNAs (lncRNAs) in various biological processes. As the mutations and abnormalities of lncRNAs are closely related to the progression of complex diseases, the identification of lncRNA-disease associations has become an important step toward the understanding and treatment of diseases. Since only a limited number of lncRNA-disease associations have been validated, an increasing number of computational approaches have been developed for predicting potential lncRNA-disease associations. However, how to predict potential associations precisely through computational approaches remains challenging. In this study, we propose a novel two-side sparse self-representation (TSSR) algorithm for lncRNA-disease association prediction. By learning the self-representations of lncRNAs and diseases from known lncRNA-disease associations adaptively, and leveraging the information provided by known lncRNA-disease associations and the intra-associations among lncRNAs and diseases derived from other existing databases, our model could effectively utilize the estimated representations of lncRNAs and diseases to predict potential lncRNA-disease associations. The experiment results on three real data sets demonstrate that our TSSR outperforms other competing methods significantly. Moreover, to further evaluate the effectiveness of TSSR in predicting potential lncRNAs-disease associations, case studies of Melanoma, Glioblastoma, and Glioma are carried out in this paper. The results demonstrate that TSSR can effectively identify some candidate lncRNAs associated with these three diseases.

Keywords: lncRNAs-disease associations prediction, computational approaches, sparse representation, lncRNA similarity, disease similarity

1. INTRODUCTION

Long non-coding RNAs (lncRNAs), which are a class of non-coding transcripts with the lengths longer than 200 nucleotides (Derrien et al., 2012; Harrow et al., 2012; Guttman et al., 2013; Chen et al., 2016b), have been proven to be involved in various biological processes (Chen et al., 2012, 2016b, 2018) and closely correlated with the development of complex diseases, such as cancers

and rheumatic diseases (Bussemakers et al., 1999; Managadze et al., 2011; Bhartiya et al., 2012; Schonrock et al., 2012; Li et al., 2013; Lu et al., 2013; Zhao et al., 2014; Chen et al., 2016b). For example, studies have revealed the roles of lncRNAs in regulating gene expression (Taft et al., 2010; Wapinski and Chang, 2011). As the development of complex diseases are closely related to the mutations and abnormalities of lncRNAs, to understand the pathogenesis of human diseases systematically, and identify the biomarkers of disease progression and prognosis, it is important to predict the potential associations between diseases and lncRNAs (Chen et al., 2016b; Yu et al., 2018). However, only a small number of lncRNA-disease associations have been validated. Therefore, efficient methods for predicting the associations between lncRNAs and diseases are emergent needed (Lu et al., 2018).

In recent years, identifying the associations between diseases and lncRNAs has attracted a lot of attentions (Chen and Yan, 2013; Lu et al., 2018). Prediction methods based on biological experiments or computational approaches are proposed to undertake this task. Due to the limitations of biological experiments such as time-consuming and expensive in cost, computational approaches provide an alternative for biological experiments and have been widely used to identify the associations between lncRNAs and diseases (Chen et al., 2016b). Existing computational approaches for association prediction can be roughly classified into three categories. The first category is based on machine learning approaches. These models predict the associations between diseases and lncRNAs based on known lncRNA-disease associations. For example, Chen et al. proposed a semi-supervised learning-based method named Laplacian Regularized Least Squares for LncRNA-disease Association (LRLSLDA) (Chen and Yan, 2013) to predict the associations between diseases and lncRNAs. Zheng et al. formulated the problem of association prediction as a matrix factorization problem and introduced a collaborative matrix factorization model (CMF) (Zheng et al., 2013) to predict the associations. However, the performance of machine learning-based methods depend on the choice of hyperparameters such as the dimensionality of the latent space in matrix factorization-based methods, and the suitable values for these hyperparameters are usually previously unknown and hard to determine.

The second category is based on random walk. These models identify potential lncRNA-disease associations by integrating known associations between diseases and lncRNAs and similarities among diseases and lncRNAs. For example, Zhou et al. predicted the associations between diseases and lncRNAs by implementing random walk with restart on the constructed similarity networks among lncRNAs and diseases (Zhou M. et al., 2015). The third category is based on data integration. These models focus on integrating multiple heterogeneous data sources. For example, Lu et al. (2018) developed a model named SIMCLDA for identifying the associations between diseases and lncRNAs based on disease-gene and gene-gene ontology associations. However, the above methods rely heavily on the similarity networks or external information (e.g., similarity networks among diseases and lncRNAs, and gene-gene associations) that are inferred based on predefined metrics.

Moreover, the information extracted from other databases or data platforms may include some irrelevant or noise information that may mislead the prediction of associations.

To address the above problems, in this paper, we introduce a novel two-side sparse self-representation (TSSR) model for lncRNA-disease association prediction. Based on known lncRNA-disease associations, our model can adaptively learn two non-negative sparse self-representation matrices which capture the intra-similarities among lncRNAs and diseases respectively. Moreover, our model could also drawn support from the intra-associations among disease and lncRNAs that derived from external information of lncRNAs and diseases to generate more accurate estimation of the representation matrices. Experiment results on three real datasets demonstrate that compared with six state-of-the-art association prediction algorithms, our TSSR model could achieve more accurate prediction results. Furthermore, case studies on three cancers (i.e., Glioblastoma, Glioma, and Melanoma) also demonstrate the effectiveness of TSSR in predicting the associations between lncRNAs and diseases. The source code of TSSR is available at https://github.com/Oyl-CityU/TSSR.

The rest of this paper is organized as follows. In section 2, we formulate our two-side sparse self-representation model and introduce a relaxed Majorization-Minimization algorithm to solve the optimization problem. The experiment results and case studies are given in section 3. In section 4, we conclude our works.

2. METHODS

2.1. Notations and Problem Statement

In this paper, we use $D = \{d_i\}_{i=1}^m$ to represent the set of lncRNAs and $T = \{t_j\}_{j=1}^n$ to represent the set of diseases, where m and n denote the number of lncRNAs and the number of diseases, respectively. A binary matrix $Y = [Y_{ij}] \in \{0, 1\}^{m \times n}$ is introduced to represent the associations between lncRNAs and diseases, where $Y_{ij} = 1$ if there is an association between lncRNA d_i and disease t_j, and $Y_{ij} = 0$ otherwise. Note that there are two reasons that may lead to $Y_{ij} = 0$. The first reason is that it has been experimentally verified that there is no association between d_i and t_j. The second reason is that whether there is an association between d_i and t_j is still unknown. Therefore, we usually refer to the zero elements in Y as unknown pairs. The lncRNA-disease association prediction problem can be formulated as the problem of predicting the scores of unknown pairs in Y, which can be used for ranking the pairs. In this study, we first rank the unknown pairs in Y based on the predicted scores in descending order, and then select the top-ranked pairs as potential association pairs.

In particular, unlike matrix factorization methods that project lncRNAs and diseases into a shared latent space and predict lncRNA-disease associations based on the inner product of their latent vectors, we try to learn the intra-similarities among lncRNAs and diseases from the observed associations in Y, and utilize the learned similarity matrices to reconstruct Y and thus predict the scores of unknown pairs in Y. Here, instead of using predefined metrics to construct the similarity matrices of lncRNAs and diseases (which makes the predicted results sensitive to the selected metrics and input data), we

introduce a novel two-side sparse self-representation (TSSR) model to adaptively learn the intra-similarities among lncRNAs and diseases from the observed associations in Y, and effectively utilize external information of lncRNAs and diseases to enhance the prediction performance.

2.2. Two-Side Sparse Self-Representation Model

Sparse representation techniques which focus on finding a sparse representation of a sample in the form of a linear combination of basic elements (also called atoms) in a dictionary, have been widely used to numerous applications such as computer vision and machine learning (Zhang et al., 2015). In traditional sparse representation models, the objective is to solve the following problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad s.t. \quad \mathbf{y} = D\mathbf{x}. \tag{1}$$

where $\| \cdot \|_0$ denotes L_0 norm, $\mathbf{y} \in R^{m \times 1}$ is a sample vector, D is a $m \times l$ matrix which denotes the dictionary and $\mathbf{x} \in R^{l \times 1}$ is the sparse representation coefficient of $\mathbf{y}$. In practice, L_0 norm is usually replaced with L_1 norm to make the above problem (1) solvable in polynomial time. Since the above problem (1) needs to take extra time to construct the dictionary D and has not data-adaptiveness. Many approaches are proposed to employ the dataset itself as the dictionary, which results in the following sparse self-representation model

$$\min_{X} \|Y - YX\|_F^2 + \beta \|X\|_1. \tag{2}$$

where $\|.\|_F$ is the Frobenius norm, Y denotes the feature set of all samples (each row denotes a feature and each column represents the feature vector of a sample), X is the sparse self-representation coefficient matrix of the columns of Y (each column $X_{.j}$ of X denotes the representation coefficient of j-th sample $Y_{.j}$, with all samples in Y as dictionary) and β is a tuning parameter to control the trade off between the minimization error and the sparsity. By solving the above model (2), X can capture the most similar relationships among the columns of Y, based on the information provided in Y. In this study, $Y \in \{0, 1\}^{m \times n}$ describes the observed associations between lncRNAs and diseases and we would like to predict potential associations between lncRNAs and diseases based on their intra-similarities learned from Y. Thus, instead of just finding the representations of the columns of Y, we prefer to explore the representations of the rows and columns of Y simultaneously, which capture the intra-similarities within lncRNAs and diseases respectively. Based on the idea of sparse self-representation, we introduce a novel two-side sparse self-representation (TSSR) model to handle the task of lncRNA-disease association prediction. In particular, we formulate the framework of TSSR into the following optimization problem

$$\min_{U,V} \|Y - UYV\|_F^2 + \beta(\|U\|_1 + \|V\|_1),$$
$$s.t. \quad U \geq 0, V \geq 0, \sum_{z=1}^{m} U_{iz} = 1, \sum_{k=1}^{n} V_{kj} = 1. \tag{3}$$

where $U = [U_{ii'}] \in \mathbb{R}_+^{m \times m}$ and $V = [V_{jj'}] \in \mathbb{R}_+^{n \times n}$ are two non-negative sparse matrices which represent the row and column representation coefficient matrices of Y, respectively, and β is a tuning parameter which controls the sparsity of U and V. Based on this definition, U denotes the coefficient matrix based on the dictionary YV, which captures the similarities among lncRNAs. For example, $U_{ii'}$ denotes the similarities between the i-th and i'-th lncRNAs, which correspond to the i-th and i'-th rows of Y. On the other hand, V denotes the coefficient matrix based on the dictionary UY, which captures the similarities among diseases. For example, $V_{jj'}$ denotes the similarities between the j-th and j'-th diseases, which correspond to the j-th and j'-th columns of Y. With the sparse regularization term, we can control the sparsity of the learned representation matrices U and V, and find the most similar relationships within lncRNAs and diseases. The constraints $\sum_{z=1}^{m} U_{iz} = 1$ and $\sum_{k=1}^{n} V_{kj} = 1$ are used to guarantee the probability properties of $U_{i.}$ and $V_{.j}$, respectively.

In the above objective function (3), the representation matrices are learned from the original data matrix Y, which means that they will be sensitive to the input data Y. If the input data only includes a small number of known associations, it may be hard to learn a comprehensive representation matrix. With the development of high-throughput experimental techniques and the accumulation of clinical information, we could also collect some functional annotations and phenotype information for lncRNAs and diseases respectively. Based on these prior information, we can infer the intra-associations among diseases and lncRNAs. To utilize these pairwise associations inferred from other databases to promote the estimation of two representation coefficient matrices U and V, two regularization terms are added to Equation (3). Moreover, we introduce a weight matrix W in a similar way to Zheng et al. (2013) to prevent unknown instances (for which association information is not available) from contributing to the determination of the row and column representations of Y (i.e., U and V). The final objective function of our TSSR model is as follows.

$$\min_{U,V} \|W \odot (Y - UYV)\|_F^2 + \beta(\|U\|_1 + \|V\|_1)$$
$$+ \lambda_d \|S_d - U\|_F^2 + \lambda_t \|S_t - V\|_F^2,$$
$$s.t. \quad U \geq 0, V \geq 0, \sum_{z=1}^{m} U_{iz} = 1, \sum_{k=1}^{n} V_{kj} = 1. \tag{4}$$

where λ_d and λ_t are two tuning parameters controlling the influences of prior intra-associations among lncRNAs and diseases, $S_d \in \mathbb{R}^{m \times m}$ and $S_t \in \mathbb{R}^{n \times n}$ denote the affinity matrices of lncRNA and disease respectively, where $(S_d)_{ii'}$ describes the association between lncRNAs d_i and $d_{i'}$, and $(S_t)_{jj'}$ describes the associations between diseases t_j and $t_{j'}$. $\odot$ denotes the element-wise product or Hadamard product of two matrices and $W \in \mathbb{R}^{m \times n}$ is a weight matrix where $W_{ij} = 0$ for unknown entries in Y and $W_{ij} = 1$ for known entries in Y. Consequently, unknown entries in Y do not contribute to the minimization of the first term of Equation (4).

2.3. Optimization Algorithm

Here, to handle the constraints in (4), we employ a relaxed Majorization-Minimization algorithm (Yang and Oja, 2011, 2012) to obtain the solution of objective function (4). For more details about this optimization method, please refer to Yang and Oja (2012). In particular, we denote $\triangledown_U$ as the gradient of our objective function with respect to U.

$$\triangledown_U = -2[W \odot (Y - UYV)]V^T Y^T - 2\lambda_d(S_d - U) + \beta. \quad (5)$$

Let $\triangledown_U^+ = 2W \odot (UYV)V^T Y^T + 2\lambda_d U + \beta$ and $\triangledown_U^- = 2(W \odot Y)V^T Y^T + 2\lambda_d S_d$ denote the positive and negative parts of $\triangledown_U$, respectively. Thus, we have $\triangledown_U = \triangledown_U^+ - \triangledown_U^-$.

Due to the constraint $\sum_{z=1}^{m} U_{iz} = 1$ and $U_{iz} \geq 0$, we obtain the following updating rule for U_{iz}:

$$U_{iz}^{\text{new}} = U_{iz} \cdot \frac{a_i^U(\triangledown_U^-)_{iz} + 1}{a_i^U(\triangledown_U^+)_{iz} + b_i^U}. \quad (6)$$

where a_i^U and b_i^U can be obtained by Equations (7) and (8), respectively.

$$a_i^U = \sum_z \frac{U_{iz}}{(\triangledown_U^+)_{iz}}, \quad (7)$$

$$b_i^U = \sum_z U_{iz} \frac{(\triangledown_U^-)_{iz}}{(\triangledown_U^+)_{iz}}. \quad (8)$$

Similarly, we denote $\triangledown_V$ as the gradient of our objective function with respect to V.

$$\triangledown_V = -2(Y^T U^T)[W \odot (Y - UYV)] - 2\lambda_t(S_t - V) + \beta. \quad (9)$$

Let $\triangledown_V^+ = 2Y^T U^T[W \odot (UYV)] + 2\lambda_t V + \beta$ and $\triangledown_V^- = 2Y^T U^T(W \odot Y) + 2\lambda_t S_t$ denote the positive and negative parts of $\triangledown_V$, respectively, we have $\triangledown_V = \triangledown_V^+ - \triangledown_V^-$.

Similarly, the updating rule for V_{kj} is as follows:

$$V_{kj}^{\text{new}} = V_{kj} \cdot \frac{a_j^V(\triangledown_V^-)_{kj} + 1}{a_j^V(\triangledown_V^+)_{kj} + b_j^V}. \quad (10)$$

where $a_j^V = \sum_k \frac{V_{kj}}{(\triangledown_V^+)_{kj}}$ and $b_j^V = \sum_k V_{kj} \frac{(\triangledown_V^-)_{kj}}{(\triangledown_V^+)_{kj}}$.

The details of the optimization algorithm to the proposed TSSR model are described in Algorithm 1. U and V can be updated by Equations (6) and (10), respectively. In this study, we stop the iteration when the changes of U and V are less than 1e-6, measured by L_1 norm. Finally, the predicted label matrix $\hat{Y}$ can be returned by $\hat{Y} = UYV$ when algorithm arrives at the convergence conditions.

Algorithm 1: Algorithm for the TSSR model

- **Inputs:** Partial label matrix Y, lncRNA affinity matrix S_d, disease affinity matrix S_t, tuning parameter $\lambda_d, \lambda_t, \beta$, weight matrix W.
- **Output:** Predicted label matrix $\hat{Y}$.
- **Main algorithm:**

  1. Initialize U and V;
  2. **While** not converged **do**
  3. Update U according to Equation (6)
     $$U_{iz}^{\text{new}} = U_{iz} \cdot \frac{a_i^U(\triangledown_U^-)_{iz} + 1}{a_i^U(\triangledown_U^+)_{iz} + b_i^U};$$
  4. Update V according to Equation (10)
     $$V_{kj}^{\text{new}} = V_{kj} \cdot \frac{a_j^V(\triangledown_V^-)_{kj} + 1}{a_j^V(\triangledown_V^+)_{kj} + b_j^V};$$
  5. Check the convergence conditions.
  6. **End while**
  7. **Return** $\hat{Y} = UYV$.

---

3. RESULTS

In this section, we demonstrate the performance of various algorithms on three real datasets. Furthermore, case studies of three cancer diseases (i.e., Melanoma, Glioblastoma, and Glioma) are performed to validate the effectiveness of our TSSR model. The materials, experimental settings, and parameter settings are described as follows.

3.1. Materials
3.1.1. LncRNA-Disease Associations
We collect three datasets to evaluate the performance of various prediction algorithms. The first dataset is downloaded from the supplementary data of a article (Lu et al., 2018), which contains 621 experimentally confirmed lncRNA-disease associations between 226 diseases and 285 lncRNAs from the LncRNADisease database[1] established in 2015. The second dataset involving 260 high-quality associations between 95 lncRNAs and 81 human disease is obtained from the supplementary files of the published article (Chen et al., 2015), which retrieved data from MNDR database[2] (Wang et al., 2013) in March 2015 . The third dataset is downloaded from the Lnc2Cancer database [3] in 2015. By getting rid of the duplicate lncRNA-disease associations for the same lncRNA-disease pair, we obtain 677 distinct associations, including 54 human cancers and 436 lncRNAs. The statistics of the three datasets are illustrated in **Table 1**.

3.1.2. Disease Similarities
As previous studies have discovered that diseases with similar phenotypes are usually related with similar dysfunctions of lncRNAs (Chen et al., 2015), incorporating the similarities among diseases estimated from other database may help to infer the

---

[1]http://www.cuilab.cn/lncrnadisease
[2]http://www.rna-society.org/mndr/
[3]http://www.bio-bigdata.com/lnc2cancer/

Datasets	No.of lncRNA	No.of disease	No.of associations	Density
LncRNA Disease	285	226	621	0.01
MNDR	95	81	260	0.03
Lnc2Cancer	436	54	677	0.03

potential associations between diseases and lncRNAs based on known lncRNA-disease associations. Similar to previous studies (Wang et al., 2010; Chen et al., 2015), we construct the similarity matrix S_t of diseases by integrating the disease semantic similarity matrix inferred from the structure of directed acyclic graph that describes the relationships among diseases (Wang et al., 2010; Chen et al., 2015) and disease Gaussian interaction profile kernel similarity matrix inferred from known associations between diseases and lncRNAs (Chen and Yan, 2013; Chen et al., 2015). In particular, we obtain the similarity matrix S_t by averaging the disease similarity matrix and disease Gaussian interaction profile kernel similarity matrix (van Laarhoven et al., 2011; Chen and Yan, 2013; Chen et al., 2015, 2016a).

3.1.3. LncRNA Similarities
Since lncRNAs with similar functions tend to exhibit similar associations with diseases, calculating the similarities among lncRNAs will promotes the identification of potential associations between diseases and lncRNAs. In this study, we calculate the similarity matrix S_d of lncRNAs by integrating the functional similarity matrix calculated by the model of LNCSIM (Chen et al., 2015) and the lncRNA Gaussian interaction profile kernel similarity matrix estimated from known associations between lncRNAs and diseases (Chen and Yan, 2013). Similar to the disease similarity matrix S_t, we obtain the lncRNA similarity matrix S_d by averaging the lncRNA functional similarity matrix and Gaussian interaction profile kernel similarity matrix (van Laarhoven et al., 2011; Chen and Yan, 2013; Chen et al., 2015; Chen et al., 2016a).

3.2. Experimental Settings
To illustrate the effectiveness of our proposed TSSR model, we compare our method with other six state-of-the-art association prediction methods, namely NetlapRLS (Xia et al., 2010), BLM-NII (Mei et al., 2012), CMF (Zheng et al., 2013), PBMDA (You et al., 2017a), PRMDA (You et al., 2017b), and SIMCLDA (Lu et al., 2018). All these methods are designed for predicting the inter-associations between different types of biological entities and all of them can make use of the prior intra-associations among biological entities to improve their performance. Thus, all these algorithms are well suited for undertaking the task of lncRNA-disease association prediction. Moreover, our experiment results show that they are effective in inferring the associations between diseases and lncRNAs. Specifically, 15 repetitions of 10-fold cross validation (CV) are conducted for each model, with receiver operating characteristic (ROC) curve as the main metric to evaluate the performance. By stacking the columns of matrix Y, we obtained the vector, a $mn \times 1$ vector,

denoted as vec(Y). In each repetition of 10-fold CV, we divide vec(Y) into ten disjoint folds randomly. Nine folds are treated as the training set while the remaining one fold is left out as the testing set. The AUC (Area Under Curve) score is calculated for each 10-fold CV repetition, and the final AUC score for each model are obtained by averaging over 15 such repetitions.

3.3. Parameter Settings
As each model has some hyperparameters that need to be predefined, we perform cross validation on the training set to determine the values of these hyperparameters. In particular, the parameter settings for various models are described as follows. For NetLapRLS (Xia et al., 2010), the hyperparameters satisfy $\frac{\gamma_{d2}}{\gamma_{d1}} = \frac{\gamma_{p2}}{\gamma_{p1}}$, $\beta_d = \beta_p$ with their values chosen from $\{10^{-6}, 10^{-5}, \ldots, 10^2\}$. For BLM-NII (Mei et al., 2012), the value of the linear combination weight α is chosen from $\{0, 0.1, 0.2, \ldots, 1.0\}$. The max function is utilized to combine the interaction scores inferred from the disease and lncRNA sides. For the matrix factorization based methods, the dimensionality of the latent space K is selected from $\{50, 100\}$ (Zheng et al., 2013). For CMF (Zheng et al., 2013), the regularization coefficient λ_1 is chosen from $\{2^{-2}, \ldots, 2^1\}$ (Zheng et al., 2013), while the values of λ_d and λ_t are chosen from $\{2^{-3}, 2^{-2}, \ldots, 2^5\}$. For PBMDA (You et al., 2017a), the maximum path length L is set to 3 and the weight threshold T is selected from $\{0.2, 0.3, \ldots, 0.8\}$ with the step size set to 0.1, while the decay factor α is set to 2.26. For SIMCLDA (Lu et al., 2018), we set the values of α_l and α_d from 0.1 to 1 with stepsize 0.1 and select the regularization parameter from $\{10^{-3}, 10^{-2}, \ldots, 10^3\}$. For TSSR, we choose the three parameters β and $\lambda_d = \lambda_t$ from $\{2^{-10}, 2^{-9}, \ldots, 2^9, 2^{10}\}$. Note that the most suitable hyper-parameters of a machine learning model on different datasets are usually different. Therefore, in this work, we adopt grid search (Bergstra and Bengio, 2012) to select the optimal hyperparameters for each model on each dataset.

3.4. Comparison With State-of-the-Art Methods
We conduct the experiments with 10-fold CV to shed light on the performance of TSSR in predicting potential lncRNA-disease associations, compared with other six state-of-the-art methods. Here, the AUC score is used to evaluate the predictive performance of various methods. The experiment results measured by AUC are shown in **Figures 1–3**. As shown in **Figure 1**, on LncRNADisease dataset, TSSR obtains an AUC score of 0.8736, which is higher than other methods (BLM-NII 0.8641, NetLapRLS 0.7837, CMF 0.7273, PBMDA 0.6885, PRMDA 0.7231, SIMCLDA 0.6067), indicating the superiority of our TSSR in predicting lncRNA-disease associations. We can find from **Figure 2** that on MNDR dataset, TSSR achieves the best AUC score (TSSR 0.8369, BLM-NII 0.7929, NetLapRLS 0.8210, CMF 0.8078, PBMDA 0.7722, PRMDA 0.6596, SIMCLDA 0.6187). On Lnc2Cancer dataset (the results are shown in **Figure 3**), TSSR still has competitive performance with other six methods with respect to AUC score (TSSR 0.9814, BLM-NII 0.9859, NetLapRLS 0.9392, CMF 0.9864, PBMDA 0.9680,

PRMDA 0.8179, SIMCLDA 0.6190). Note that on Lnc2Cancer, our TSSR achieves similar performance with BLM-NII and CMF. This may due to the parameter setting of TSSR. In this study, the values of the hyperparameters λ_d and λ_t (which control the influences of prior intra-similarities among lncRNAs and diseases) in our TSSR are set to same for simplicity, which is reasonable when the two data sets are balanced. However, the number of lncRNAs and diseases in Lnc2Cancer dataset are imbalanced. Thus, forcing λ_d and λ_t to be equal may limit the performance of TSSR. If the values of λ_d and λ_t are tuned separately, TSSR could achieve better performance. Moreover, to evaluate the effect of external information on the performance of TSSR, we remove the regularization terms related to the external information (i.e., setting $\lambda_d = \lambda_t = 0$) and show the results in **Figure 4**. As shown in this figure, the performance of TSSR and TSSR without external information (denoted by TSSR_original) is comparable (on LncRNADisease, TSSR 0.8736, TSSR_original 0.8735; on MNDR, TSSR 0.8369, TSSR_original 0.8367; on Lnc2Cancer, TSSR 0.9814, TSSR_original 0.9614), which means the improved performance of TSSR is mainly due to the self-representation learning. Thus, our TSSR does not depend heavily on the external information. All these results demonstrate the effectiveness of the proposed TSSR in predicting potential lncRNA-disease associations.

3.5. Effects of Parameters

The proposed TSSR involves three parameters, λ_d, λ_t, and β, where λ_d and λ_t control the influences of prior intra-associations among lncRNAs and diseases and β controls the sparsity of U and V. We will study how these parameters affect the performance of TSSR.

Figure 5 shows the prediction performance of TSSR on LncRNADisease dataset, MNDR dataset and Lnc2Cancer dataset, measured by AUC with respect to different values of λ_d and λ_t. As shown in **Figure 5**, the optimal value of $\lambda_d = \lambda_t$ for these three datasets is 2^{-10}, 2^0, and 2^2, respectively, while β is set to 2^1, 2^8, and 2^8, respectively. We find that TSSR usually performs well when the values of λ_d and λ_t are relatively small, which means additional use of external information is not always helpful for performance improvement. On the contrary, if the external information contains noise, the performance of TSSR may decrease if we overemphasizing the effect of external information. These results demonstrate that our TSSR can effectively learn the representation matrices from known lncRNA-disease associations, and flexibly utilize external information to promote the prediction of potential lncRNA-disease associations.

In addition, we also study the impact of sparsity control parameter β. **Figure 6** illustrates the AUC scores obtained by TSSR in terms of different values of β. As shown in **Figure 6**, on these three datasets, TSSR achieves the best AUC score when the value of β is 2^1, 2^8, and 2^8, respectively, while $\lambda_d = \lambda_t$ is set to 2^{-10}, 2^0, and 2^2, respectively. We can also find from this figure that larger values of β can generally achieve better performance, which indicates the

importance of controlling the sparsity of the representation matrices U and V.

3.6. Case Studies

To further validate the performance of our algorithm, based on the LncRNADisease dataset, we apply our TSSR model to identify the most possible lncRNAs that associated with three cancers (i.e., Melanoma, Glioma, and Glioblastoma). Here, all the known associations in the LncRNADisease dataset are used to train the model. Then we select the top 20 associated lncRNAs which get the highest predicted ranks for each cancer and verify these predictions based on MNDR and Lnc2Cancer databases. Moreover, the relevant literatures that support the prediction results are listed to indicate whether the predicted lncRNA-disease associations have been experimentally validated. Specially, MNDR database contains both experimental and prediction evidence (Ning et al., 2016; Ping et al., 2018). The results for the three cancers are shown in **Tables 2–4**, respectively. Note that we only show the predictions that are not included in the training set.

Melanoma is a deadly malignancy which develops from the pigment-containing cells with increasing incidence than that of any other types of cancer (Aladowicz et al., 2013). People with low level of skin pigment exposure in excess ultraviolet light (UV) have a high risk to be infected a melanoma (Kanavy and Gerstenblith, 2011). It has been estimated that by 2030, melanoma could overtake colorectal cancer as the fifth most common cancer (Rahib et al., 2014). Therefore, we apply our TSSR model to predict the potential melanoma-associated lncRNAs. According to the results shown in **Table 2** (the complete list of the top 20 identified lncRNAs is shown in **Supplementary Material**), 10 out of the top 20 identified lncRNAs have been verified. For example, Luan et al. (2016) discovered that MALAT1 could promote the cell proliferation, invasion and migration of melanoma. Li et al. observed that MEG3 was obviously decreased in melanoma cells (Li et al., 2018). They also found melanoma cell apoptosis was induced by up-regulation of MEG3, and consequently come to a conclusion that overexpression of MEG3 has a significant repression impact in melanoma cell migration and invasion ability.

Glioma is one of the most common primary malignant tumors originating in the brain, which comprises approximately 30% of all brain tumors (Goodenberger and Jenkins, 2012; Boele et al., 2015). Glioma can be graded from I to IV by World Health Organization (WHO) grading system according to their grade (Louis et al., 2016a,b). The exact causes of glioma are still unclear at the present (Kwiatkowska and Symons, 2013; Li et al., 2015). Studies have revealed the roles of lncRNAs in the development of human disease, including glioma (Zhou et al., 2018). Here, we utilize the TSSR to identify the potential lncRNAs that are more likely to related to glioma. Based on the experiment results, 9 out of the top 20 identified lncRNAs have been validated in the MNDR and Lnc2Cancer databases, and other relevant literatures. The results are shown in **Table 3** (the complete list of the top 20 identified lncRNAs is shown in **Supplementary Material**). For example, Ma et al. discovered that compared with paired
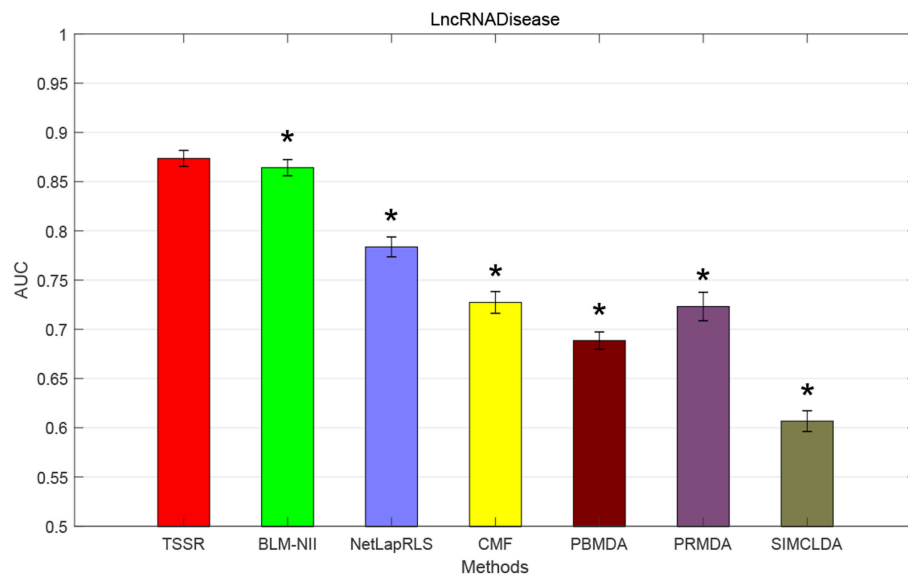
FIGURE 1 | AUC scores of various algorithms in LncRNADisease dataset (* indicates TSSR significantly outperforms the competitor with $p < 0.05$ using t-test, error bars denote 95% confidence intervals).
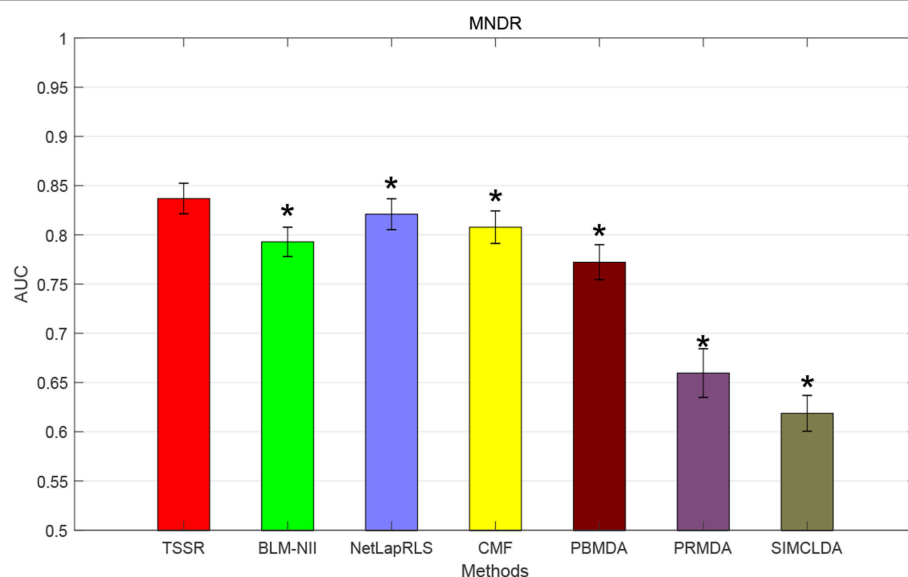
FIGURE 2 | AUC scores of various algorithms in MNDR dataset (* indicates TSSR significantly outperforms the competitor with $p < 0.05$ using t-test, error bars denote 95% confidence intervals).

normal tissues, the expression level of lncRNA MALAT1 was increased in glioma tissues, which means MALAT1 can be treated as a convictive marker for the prognosis of glioma patients (Ma et al., 2015). Zou et al. revealed that glioma patients with high PVT1 expression had low survival rate (Zou et al., 2017). Moreover, patients who received chemotherapy and radiotherapy could improve their survival by down-regulating PVT1. They also indicated that PVT1 could be served as potential target for the treatment of diffuse gliomas.

Glioblastoma, also known as glioblastoma multiform (GBM) (grade IV of Glioma), is the most common and aggressive form of primary brain tumors and kills nearly every patient in a median time of 15 months (Bleeker et al., 2012; Jovčevska et al., 2013). More importantly, there is still no clear way to prevent the disease (Gallego, 2015). Therefore, it is urgent to predict the potential glioblastoma-associated lncRNAs. In this study, we use our TSSR to undertake this task. As shown in **Table 4**, 8 out of the 20 lncRNAs have been verified in
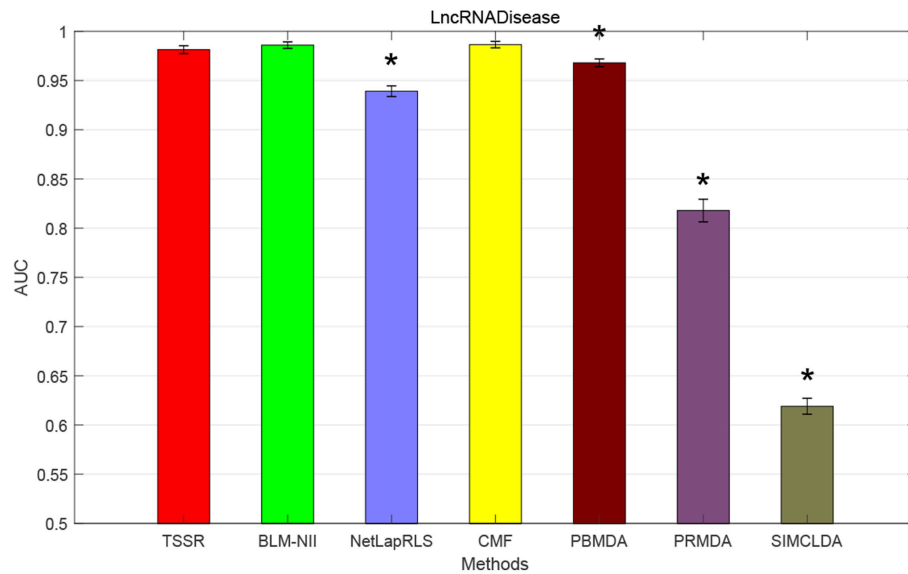
FIGURE 3 | AUC scores of various algorithms in Lnc2Cancer dataset (* indicates TSSR significantly outperforms the competitor with $p < 0.05$ using t-test, error bars denote 95% confidence intervals).
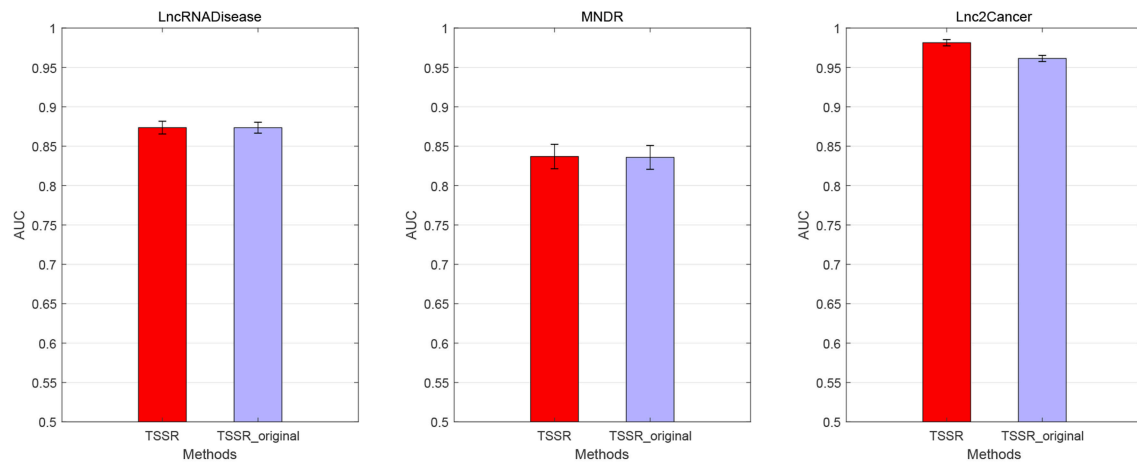
FIGURE 4 | Performance of TSSR with and without external information (denoted by TSSR and TSSR_original, respectively) on LncRNADisease, MNDR, and Lnc2Cancer datasets, measured by AUC (error bars denote 95% confidence intervals).
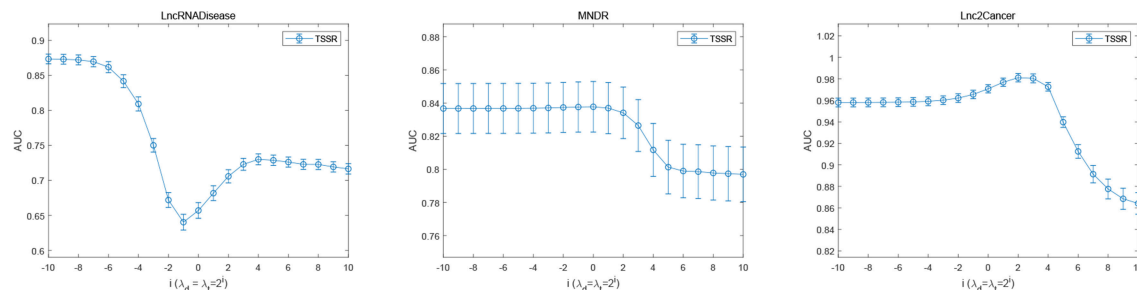
FIGURE 5 | Performance of TSSR on LncRNADisease, MNDR, and Lnc2Cancer datasets, measured by AUC with different values of λ_d and λ_t (error bars denote 95% confidence intervals).
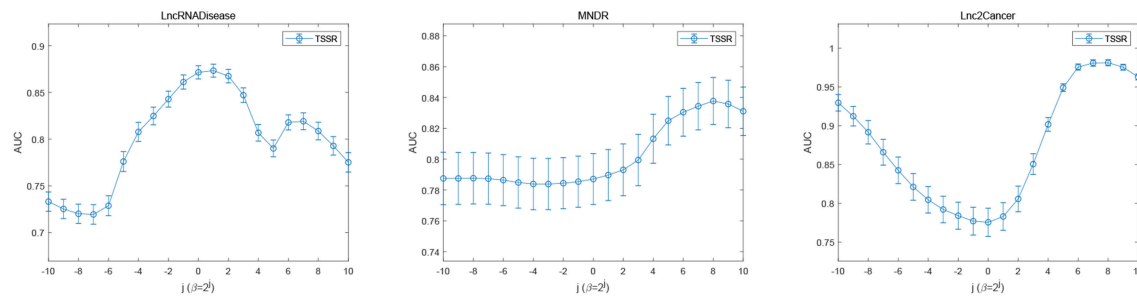
FIGURE 6 | Performance of TSSR on LncRNADisease, MNDR, and Lnc2Cancer datasets, measured by AUC with different values of β (error bars denote 95% confidence intervals).

TABLE 2 | The identified novel lncRNAs that have been verified to be associated with Melanoma.

Rank	lncRNA	Evidence(Database)	Evidence(PMID)	Expression pattern
1	CCAT2	MNDR	Prediction evidence	
2	TUSC7	MNDR	Prediction evidence	
9	GHET1	MNDR	Prediction evidence	
12	MEG3	MNDR/lnc2Cancer	29781534,29808164	Up-regulated, differential expression
13	HOTAIR	MNDR/lnc2Cancer	28067428,23862139	up-regulated
14	SOX2-OT	MNDR	Prediction evidence	
15	MALAT1	MNDR/lnc2Cancer	27725873, 27564100,27966454,24892958,19625619	Up-regulated,differential expression
17	SNHG5	MNDR/lnc2Cancer	26440365	Up-regulated
18	BCAR4	MNDR	Prediction evidence	
19	CCAT1	lnc2Cancer	28409554	Up-regulated

Prediction evidence denotes the prediction associations in MNDR database.

TABLE 3 | The identified novel lncRNAs that have been verified to be associated with Glioma.

Rank	lncRNA	Evidence(Database)	Evidence(PMID)	Expression pattern
2	HOTAIR	MNDR/lnc2Cancer	29323737,28083786 ,29218099, 27277755,24203894	Up-regulated, down-regulated
3	MALAT1	MNDR/lnc2Cancer	28551849,27134488,26649728,25613066,26619802	Up-regulated, down-regulated
4	GAS5	MNDR/lnc2Cancer	26370254,28666797	Up-regulated, down-regulated
7	PVT1	lnc2Cancer	28351322,29108264,29620147,29501773,29046366	Up-regulated, differential expression
11	SPRY4-IT1	MNDR/lnc2Cancer	29467908,27460732,26464658	Up-regulated
12	GHET1	MNDR	Prediction evidence	
15	IGF2-AS	MNDR	Prediction evidence	
18	LincRNA-p21	lnc2Cancer	28689810	Down-regulated
19	SNHG4	MNDR	Prediction evidence	

Prediction evidence denotes the prediction associations in MNDR database.

the MNDR and Lnc2Cancer databases, and other relevant literatures (the complete list of the top 20 identified lncRNAs is shown in **Supplementary Material**). For example, Zhou et al. described that HOTAIR has a significant increased expression in multiple human cancers including GBM and they found HOTAIR is necessary for GBM formation *in vivo* (Zhou X. et al., 2015). Thus, HOTAIR could be a potential therapeutic target in glioblastoma. Liu et al. found that NBAT1 has lower expressions in glioblastoma tissues compared with those in normal brain tissues and they also observed that up-regulated NBAT1 inhibits proliferation of T98 and U87 cells via regulating

Akt, suggesting that NBAT1 may be related to prognosis of glioblastoma (Liu et al., 2018).

Based on the above case studies, we find that our TSSR is effective in identifying novel associations between lncRNAs and diseases based on known lncRNA-disease associations and intra-associations among lncRNAs and diseases.

4. CONCLUSION

Increasing evidences indicate the role of lncRNAs in biological processes, which motivates the development of computational

TABLE 4 | The identified novel lncRNAs that have been verified to be associated with Glioblastoma.

Rank	lncRNA	Evidence(Database)	Evidence(PMID)	Expression pattern
1	MEG3	MNDR/lnc2Cancer	27306825,28187000,22234798,25378224,26111795	Up-regulated
2	HOTAIR	MNDR/lnc2Cancer	27306825,25428914,25823657,26111795,26943771	Up-regulated
6	BCYRN1	MNDR	25561975	Differentially expressed
8	GAS5	MNDR/lnc2Cancer	27784795,23726844	Up-regulated, differentially expressed
10	NEAT1	lnc2Cancer	23046790	Up-regulated
11	HIF1A-AS2	MNDR/lnc2Cancer	27264189	Up-regulated
15	NBAT1	lnc2Cancer	29771423	Up-regulated
17	NDM29	MNDR	25561975	Differentially expressed

Prediction evidence denotes the prediction associations in MNDR database.

models to identify the potential associations between lncRNAs and diseases. Predicting the potential associations between lncRNAs and diseases based on known lncRNA-disease associations is equivalent to a recommendation problem with implicit feedback, where the task is to predict whether the unknown pairs in Y are potential associations or not. In this paper, we present a novel model, named two-side sparse self-representation (TSSR), to predict the scores of unknown pairs in Y. Based on these predicted scores, we could identify potential associations between lncRNAs and diseases. Unlike previous matrix factorization techniques that project lncRNAs and diseases into a shared latent space and predict lncRNA-disease associations based on the inner product of their latent vectors (where the dimension of latent space is previously unknown and hard to determine), our model directly learn the intra-similarities among lncRNAs and diseases from the observed associations in Y, and utilize the learned representation matrices to reconstruct Y by regarding original Y as a dictionary. As shown in Equation (4), our TSSR does not need to make many assumptions of the model in advance. Moreover, by forcing the representation matrices to be sparse, our TSSR could learn the most similar relationships among lncRNAs and diseases based on the observed associations in Y. Thus, our TSSR has data-adaptiveness and avoids the determination of some sensitive parameters such as the dimension of latent space and number of nearest neighbors. Unlike random walk-based or data integration-based methods that rely heavily on the similarity networks inferred from external information with predefined metrics, our model could adaptively learn the self-representations of lncRNAs and diseases according to their performance in reconstructing observed associations in Y. Moreover, in case the input data Y only includes a small number of known associations, our model could draw support from the intra-associations among lncRNAs and diseases derived from external information to enhance the learning of representation matrices. Therefore, our model could effectively predict potential lncRNA-disease associations by leveraging the information provided by known lncRNA-disease associations and external information of lncRNAs and diseases. Experiment results on three real data sets show that our TSSR could achieve better performance than other six state-of-the-art methods. The effectiveness of TSSR in predicting potential lncRNA-disease associations is also evaluated based on three case studies. As a link prediction algorithm, our TSSR model is flexible

and could be used to handle other link prediction tasks in bipartite networks.

Furthermore, since external information of lncRNAs and diseases are utilized to enhance the performance of various methods, we also perform sensitivity analysis to assess the influences of noise information on the performances of various methods. In particular, we generate the similarity matrices S_d and S_t randomly (i.e., the elements in S_d and S_t are generated randomly) and test the performances of various methods. The detailed experiment results are shown in **Tables S4–S6**. As shown in these tables, although the performance of TSSR is affected by the noise information, it could still achieve the best performance, which means our TSSR could be used to undertake the lncRNA-disease prediction task even when the collected external information of lncRNAs and diseases contains a lot of noise.

With the development of high-throughput experimental techniques, an increasing number of data for lncRNAs and diseases are becoming available. We can calculate the similarities among lncRNAs (or diseases) based on different views of data and different metrics. How to efficiently seek the optimal combination of these similarities is an interesting future work. We will try to extend our model to handle this problem.

AUTHOR CONTRIBUTIONS

LO-Y and JH conceived and designed the study, performed the statistical analysis, and drafted the manuscript. ZZ conceived of the study, and participated in its design and coordination and helped to draft the manuscript. X-FZ and Y-RL participated in the design of the study, performed the statistical analysis, and helped to revise the manuscript. YS and SH participated in the design of the study and helped to revise the manuscript. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

REFERENCES

Aladowicz, E., Ferro, L., Vitali, G. C., Venditti, E., Fornasari, L., and Lanfrancone, L. (2013). Molecular networks in melanoma invasion and metastasis. *Future Oncol.* 9, 713–726. doi: 10.2217/fon.1

Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* 13, 281–305. doi: 10.1016/j.chemolab.2011.12.002

Bhartiya, D., Kapoor, S., Jalali, S., Sati, S., Kaushik, K., Sachidanandan, C., et al. (2012). Conceptual approaches for lncRNA drug discovery and future strategies. *Expert Opin. Drug Discov.* 7, 503–513. doi: 10.1517/17460441.2012.682055

Bleeker, F. E., Molenaar, R. J., and Leenstra, S. (2012). Recent advances in the molecular understanding of glioblastoma. *J. Neuro Oncol.* 108, 11–27. doi: 10.1007/s11060-011-0793-0

Boele, F. W., Rooney, A. G., Grant, R., and Klein, M. (2015). Psychiatric symptoms in glioma patients: from diagnosis to management. *Neuropsychiatr. Dis. Treat.* 11:1413. doi: 10.2147/NDT.S65874

Bussemakers, M. J., Van Bokhoven, A., Verhaegh, G. W., Smit, F. P., Karthaus, H. F., Schalken, J. A., et al. (1999). Dd3:: A new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.* 59, 5975–5979.

Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., et al. (2012). LncRNAdisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* 41, D983–D986. doi: 10.1093/nar/gks1099

Chen, L., Ma, D., Li, Y., Li, X., Zhao, L., Zhang, J., et al. (2018). Effect of long non-coding RNA PVT1 on cell proliferation and migration in melanoma. *Int. J. Mol. Med.* 41, 1275–1282. doi: 10.3892/ijmm.2017.3335

Chen, X., Huang, Y.-A., Wang, X.-S., You, Z.-H., and Chan, K. C. (2016a). FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget* 7, 45948–45958. doi: 10.18632/oncotarget.10008

Chen, X., Yan, C. C., Luo, C., Ji, W., Zhang, Y., and Dai, Q. (2015). Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. *Sci. Rep.* 5:11338. doi: 10.1038/srep11338

Chen, X., Yan, C. C., Zhang, X., and You, Z.-H. (2016b). Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief. Bioinformatics* 18, 558–576. doi: 10.1093/bib/bbw060

Chen, X., and Yan, G.-Y. (2013). Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics* 29, 2617–2624. doi: 10.1093/bioinformatics/btt426

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

Gallego, O. (2015). Nonsurgical treatment of recurrent glioblastoma. *Curr. Oncol.* 22:e273. doi: 10.3747/co.22.2436

Goodenberger, M. L., and Jenkins, R. B. (2012). Genetics of adult glioma. *Cancer Genet.* 205, 613–621. doi: 10.1016/j.cancergen.2012.10.009

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., and Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding rnas do not encode proteins. *Cell* 154, 240–251. doi: 10.1016/j.cell.2013.06.009

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). Gencode: the reference human genome annotation for the encode project. *Genome Res.* 22, 1760–1774. doi: 10.1101/gr.135350.111

Jovčevska, I., Kočevar, N., and Komel, R. (2013). Glioma and glioblastoma-how much do we (not) know? *Mol. Clin. Oncol.* 1, 935–941. doi: 10.3892/mco.2013.172

Kanavy, H. E., and Gerstenblith, M. R. (2011). Ultraviolet radiation and melanoma. *Semin. Cutan. Med. Surg.* 30, 222–228. doi: 10.1016/j.sder.2011.08.003

Kwiatkowska A., and Symons M. (2013). "Signaling determinants of glioma cell invasion," in *Glioma Signaling. Advances in Experimental Medicine and Biology,* Vol. 986, ed J. Barańska (Dordrecht: Springer), 121–141.

Li, J., Xuan, Z., and Liu, C. (2013). Long non-coding rnas and complex human diseases. *Int. J. Mol. Sci.* 14, 18790–18808. doi: 10.3390/ijms140918790

Li, J., Yuan, J., Yuan, X., Zhang, C., Li, H., Zhao, J., et al. (2015). Induction effect of microRNA-449a on glioma cell proliferation and inhibition on glioma cell apoptosis by promoting pkcα. *Eur. Rev. Med. Pharmacol. Sci.* 19, 3587–3592. Available online at: https://www.europeanreview.org/article/9593

Li, P., Gao, Y., Li, J., Zhou, Y., Yuan, J., Guan, H., et al. (2018). LncRNA meg3 repressed malignant melanoma progression via inactivating Wnt signaling pathway. *J. Cell. Biochem.* 119, 7498–7505. doi: 10.1002/jcb.27061

Liu, J., Wang, W., Zhang, X., Du, Q., Li, H., and Zhang, Y. (2018). Effect of downregulated lncrna nbat1 on the biological behavior of glioblastoma cells. *Eur. Rev. Med. Pharmacol. Sci.* 22, 2715–2722. doi: 10.26355/eurrev_201805_14968

Louis, D. N., Ohgaki, H., Wiestler, O. D., and Cavenee, W. K. (2016a). *WHO Classification of Tumours of the Central Nervous System.* International Agency for Research on Cancer.

Louis, D. N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016b). The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131, 803–820. doi: 10.1007/s00401-016-1545-1

Lu, C., Yang, M., Luo, F., Wu, F.-X., Li, M., Pan, Y., et al. (2018). Prediction of lncRNA-disease associations based on inductive matrix completion. *Bioinformatics* 34, 3357–3364. doi: 10.1093/bioinformatics/bty327

Lu, Q., Ren, S., Lu, M., Zhang, Y., Zhu, D., Zhang, X., et al. (2013). Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics* 14:651. doi: 10.1186/1471-2164-14-651

Luan, W., Li, L., Shi, Y., Bu, X., Xia, Y., Wang, J., et al. (2016). Long non-coding RNA MALAT1 acts as a competing endogenous RNA to promote malignant melanoma growth and metastasis by sponging miR-22. *Oncotarget* 7, 63901–63912. doi: 10.18632/oncotarget.11564

Ma, K., Wang, H., Li, X, Li, T., Su, G., Yang, P., et al. (2015). Long noncoding RNA MALAT1 associates with the malignant status and poor prognosis in glioma. *Tumor Biol.* 36, 3355–3359. doi: 10.1007/s13277-014-2969-7

Managadze, D., Rogozin, I. B., Chernikova, D., Shabalina, S. A., and Koonin, E. V. (2011). Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. *Genome Biol. Evol.* 3, 1390–1404. doi: 10.1093/gbe/evr116

Mei, J.-P., Kwoh, C.-K., Yang, P., Li, X.-L., and Zheng, J. (2012). Drug–target interaction prediction by learning from local information and neighbors. *Bioinformatics* 29, 238–245. doi: 10.1093/bioinformatics/bts670

Ning, S., Zhang, J., Peng, W., Hui, Z., Wang, J., Yue, L., et al. (2016). Lnc2cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res.* 44(Database issue), D980–D985. doi: 10.1093/nar/gkv1094

Ping, P., Wang, L., Kuang, L., Ye, S., Iqbal, M. F. B., and Pei, T. (2018). A novel method for lncRNA-disease association prediction based on an lncRNA-disease association network. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 16, 688–693. doi: 10.1109/TCBB.2018.2827373

Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., and Matrisian, L. M. (2014). Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* 74, 2913–2921. doi: 10.1158/0008-5472.CAN-14-0155

Schonrock, N., Harvey, R. P., and Mattick, J. S. (2012). Long noncoding RNAs in cardiac development and pathophysiology. *Circul. Res.* 111, 1349–1362. doi: 10.1161/CIRCRESAHA.112.268953

Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M., and Mattick, J. S. (2010). Non-coding RNAs: regulators of disease. *J. Pathol.* 220, 126–139. doi: 10.1002/path.2638

van Laarhoven, T., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics* 27, 3036–3043. doi: 10.1093/bioinformatics/btr500

Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 26, 1644–1650. doi: 10.1093/bioinformatics/btq241

Wang, Y., Chen, L., Chen, B., Li, X., Kang, J., Fan, K., et al. (2013). Mammalian ncRNA-disease repository: a global view of ncRNA-mediated disease network. *Cell Death Dis.* 4:e765. doi: 10.1038/cddis.2013.292

Wapinski, O., and Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends Cell Biol.* 21, 354–361. doi: 10.1016/j.tcb.2011.04.001

Xia, Z., Wu, L.-Y., Zhou, X., and Wong, S. T. (2010). Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst. Biol.* 4:S6. doi: 10.1186/1752-0509-4-S2-S6

Yang, Z., and Oja, E. (2011). Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization. *IEEE Trans. Neural Netw.* 22, 1878–1891. doi: 10.1109/TNN.2011.2170094

Yang, Z., and Oja, E. (2012). "Clustering by low-rank doubly stochastic matrix decomposition," in *Proceedings of the 29th International Coference on International Conference on Machine Learning, ICML'12* (Edinburgh: Omnipress), 707–714.

You, Z.-H., Huang, Z.-A., Zhu, Z., Yan, G.-Y., Li, Z.-W., Wen, Z., et al. (2017a). PBMDA: A novel and effective path-based computational model for MiRNA-disease association prediction. *PLoS Comput. Biol.* 13:e1005455. doi: 10.1371/journal.pcbi.1005455

You, Z.-H., Wang, L.-P., Chen, X., Zhang, S., Li, X.-F., Yan, G.-Y., et al. (2017b). PRMDA: personalized recommendation-based MiRNA-disease association prediction. *Oncotarget* 8, 85568–85583. doi: 10.18632/oncotarget.20996

Yu, J., Ping, P., Wang, L., Kuang, L., Li, X., and Wu, Z. (2018). A novel probability model for lncRNA–disease association prediction based on the naïve bayesian classifier. *Genes* 9:345. doi: 10.3390/genes9070345

Zhang, Z., Xu, Y., Yang, J., Li, X., and Zhang, D. (2015). A survey of sparse representation: algorithms and applications. *IEEE Access* 3, 490–530. doi: 10.1109/ACCESS.2015.2430359

Zhao, W., Luo, J., and Jiao, S. (2014). Comprehensive characterization of cancer subtype associated long non-coding RNAs and their clinical implications. *Sci. Rep.* 4:6591. doi: 10.1038/srep06591

Zheng, X., Ding, H., Mamitsuka, H., and Zhu, S. (2013). "Collaborative matrix factorization with multiple similarities for predicting drug-target interactions," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Chicago, IL: ACM), 1025–1033. doi: 10.1145/2487575.2487670

Zhou, M., Wang, X., Li, J., Hao, D., Wang, Z., Shi, H., et al. (2015). Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncrna and disease network. *Mol. Biosyst.* 11, 760–769.

Zhou, Q., Liu, J., Quan, J., Liu, W., Tan, H., and Li, W. (2018). lncRNAs as potential molecular biomarkers for the clinicopathology and prognosis of glioma: a systematic review and meta-analysis. *Gene* 668, 77–86. doi: 10.1016/j.gene.2018.05.054

Zhou, X., Ren, Y., Zhang, J., Zhang, C., Zhang, K., Han, L., et al. (2015). HOTAIR is a therapeutic target in glioblastoma. *Oncotarget* 6, 8353–8365.

Zou, H., Wu, L.-X., Yang, Y., Li, S., Mei, Y., Liu, Y.-B., et al. (2017). lncRNAs PVT1 and HAR1A are prognosis biomarkers and indicate therapy outcome for diffuse glioma patients. *Oncotarget* 8, 78767–78780.

LncRRIsearch: A Web Server for lncRNA-RNA Interaction Prediction Integrated With Tissue-Specific Expression and Subcellular Localization Data

Tsukasa Fukunaga [1,2†], Junichi Iwakiri [3,4†], Yukiteru Ono [5] and Michiaki Hamada [1,4,6,7,8,9]*

[1] *Department of Electrical Engineering and Bioscience, Faculty of Science and Engineering, Waseda University, Tokyo, Japan,* [2] *Department of Computer Science, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan,* [3] *Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan,* [4] *Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan,* [5] *IMSBIO Co., Ltd., Tokyo, Japan,* [6] *Computational Bio Big-Data Open Innovation Laboratory, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan,* [7] *Institute for Medical-oriented Structural Biology, Waseda University, Tokyo, Japan,* [8] *Graduate School of Medicine, Nippon Medical School, Tokyo, Japan,* [9] *Center for Data Science, Waseda University, Tokyo, Japan*

Long non-coding RNAs (lncRNAs) play critical roles in various biological processes, but the function of the majority of lncRNAs is still unclear. One approach for estimating a function of a lncRNA is the identification of its interaction target because functions of lncRNAs are expressed through interaction with other biomolecules in quite a few cases. In this paper, we developed "LncRRIsearch," which is a web server for comprehensive prediction of human and mouse lncRNA-lncRNA and lncRNA-mRNA interaction. The prediction was conducted using RIblast, which is a fast and accurate RNA-RNA interaction prediction tool. Users can investigate interaction target RNAs of a particular lncRNA through a web interface. In addition, we integrated tissue-specific expression and subcellular localization data for the lncRNAs with the web server. These data enable users to examine tissue-specific or subcellular localized lncRNA interactions. LncRRIsearch is publicly accessible at http://rtools.cbrc.jp/LncRRIsearch/.

Keywords: lncRNA, RNA-RNA interaction, web server, tissue-specific expression, subcellular localization

1. INTRODUCTION

Long non-coding RNAs (lncRNAs) were initially considered to be transcriptional noise or experimental artifacts, but recent research has revealed that lncRNAs play important roles in various biological processes, such as cell differentiation (Fatica and Bozzoni, 2014) and functioning of the immune system (Carpenter et al., 2013). While large-scale RNA sequencing studies have discovered several tens of thousands of lncRNAs in the human transcriptome (Iyer et al., 2015; Hon et al., 2017), the function is known in detail for only a small number of lncRNAs (Quek et al., 2014; de Hoon et al., 2015). To understand the molecular mechanisms of complex biological systems, elucidating the functions of more lncRNAs is an important research topic.

Recent discoveries of lncRNA-mRNA interactions regulating biological processes (Gong and Maquat, 2011; Kretz et al., 2013; Abdelmohsen et al., 2014) suggest that comprehensive lncRNA-mRNA interaction predictions are helpful for the estimation of lncRNA function. Several databases or web services have been developed for the function prediction based on lncRNA-mRNA interactions, but there are no web services for comprehensive prediction of human and mouse lncRNA interaction. RAID contains some lncRNA-mRNA interaction data taken from the literature, but the number of interactions is limited and comprehensiveness is low (Yi et al., 2017). RISE includes experimentally validated lncRNA-RNA interactions based on high-throughput sequencing methods (Lu et al., 2016; Nguyen et al., 2016), but the number of lncRNA interactions is also limited (Gong et al., 2017). The database compiled by Terai et al. (2016) contains predicted lncRNA-mRNA and lncRNA-lncRNA interaction data at transcriptome scale, but the database does not store more than one local base-pairing interaction for each lncRNA-RNA interaction. In addition, the database includes only human lncRNA-RNA interactions.

To address these shortcomings, we have constructed the LncRRIsearch, which is a web server for comprehensive prediction of human and mouse lncRNA-mRNA and lncRNA-lncRNA interactions. We applied RIblast to human and mouse transcriptome to predict RNA-RNA interactions (Fukunaga and Hamada, 2017). LncRRIsearch provides multiple local base-pairing interactions predicted by RIblast for each lncRNA-RNA interaction. In addition, unlike previous databases or web services, we integrated tissue-specific RNA expression and subcellular localization data of lncRNAs with our web service. These data help us to verify the correctness of the predicted interactions. Actually, we showed the tissue-specificity information improves the prediction accuracy for lncRNA-RNA interactions in previous research (Iwakiri et al., 2017). LncRRIsearch is freely accessible at http://rtools.cbrc.jp/LncRRIsearch/.

2. MATERIALS AND METHODS

2.1. Dataset of lncRNA and mRNA Sequences

We downloaded human and mouse RNA sequences from GENCODE version 25 and M14, respectively (Harrow et al., 2012). While we used all lncRNA transcript sequences in our analysis, we used the longest mRNA transcript for each gene to reduce the size of the dataset. In addition, we excluded transcripts in the pseudoautosomal region on the Y-chromosome from the analysis. As a result, we obtained 27,674 lncRNA and 20,360 mRNA transcripts as human RNA dataset, and 16,113 lncRNA and 22,468 mRNA transcripts as mouse RNA dataset. Note that LncRRIsearch contains an additional 175 mRNA and 3,776 lncRNA transcripts in comparison with the database previously compiled by Terai et al. (2016) as human RNA dataset. This difference is derived from the version update of GENCODE.

2.2. Prediction of lncRNA-RNA Interactions

RNA-RNA interaction prediction for long RNAs is time-consuming calculation, and even the fastest programs at present cannot be predict the interactions in real-time. Therefore, we predicted comprehensive human and mouse lncRNA-mRNA and lncRNA–lncRNA interactome in advance, and stored the interaction results in MySQL database. By selecting a query RNA or a target RNA, users can obtain pre-calculated prediction results of the selected RNA.

We used the RIblast program, which has been recently developed by our group, for comprehensive RNA-RNA interaction prediction (Fukunaga and Hamada, 2017). RIblast predicts local base-pairing interactions based on interaction energy that is computed by using both accessibility energy and hybridization energy. Briefly, RIblast considers both effects on stabilization energy derived from hybridization between two RNA sequences and the energy for preventing the formation of intramolecular double-stranded structure. (If an RNA region forms double-stranded structure in the secondary structure, the region does not tend to interact with the other RNA molecules via base-pairing.) RIblast output multiple candidates for local base-pairing interactions for each RNA-RNA pair. The threshold interaction energy was set to −12 or −16 kcal/mol. We regarded the query and target RNA pairs (A, B) and (B, A) as being different because RIblast predicts slightly different interactions for these pairs. Users can sort target transcripts for each query transcript by two criteria: MINENERGY and SUMENERGY. MINENERGY denotes the minimum interaction energy of local base-pairing interaction among all interactions between the query RNA and the target RNA. SUMENERGY means the sum of all interaction energies of local base-parings for the RNA-RNA pair.

We investigated whether the experimentally validated lncRNA-mRNA interactions were predicted by RIblast. We verified that RIblast predicted human 1/2-sbs RNA (ENST00000548810) and SERPINE1 (ENST00000223095) interaction, and human 1/2-sbs RNA and ANKRD57 (ENST00000356454) interaction (Gong and Maquat, 2011). In addition, human 7SL RNA (ENST00000635274) and TP53 mRNA (ENST00000617185) interaction (Abdelmohsen et al., 2014) was also predicted by RIblast. As we did not predict mRNA-mRNA interactions, LncRRIsearch does not provide human TINCR-mRNA interactions (Kretz et al., 2013) (TINCR ENST00000448587) was annotated as mRNA in GENCODE ver.25). In summary, our prediction results include experimentally validated lncRNA-mRNA interactions for lncRNAs.

2.3. Expression Analysis for Tissue-Specific lncRNA-RNA Interaction

Expression levels of human lncRNA and mRNA genes were estimated from RNA-seq data derived from five international consortia. The first RNA-seq dataset was derived from 32 tissues collected from 122 human individuals, which was produced by the Human Protein Atlas Project (Expression Atlas ID: E-MTAB-2836) (Uhlén et al., 2015). The second

RNA-seq dataset was derived from 30 representative tissues, released by the GTEx Consortium (Expression Atlas ID: E-MTAB-2919) (GTEx Consortium, 2015). The third RNA-seq dataset was produced by the Human Body Map Project from 16 tissues (Expression Atlas ID: E-MTAB-513) (Cabili et al., 2011). The fourth RNA-seq dataset, derived from 19 tissues isolated from fetuses with congenital defects, was released by the Epigenome Roadmap Project (Expression Atlas ID: E-MTAB-3871) (Kundaje et al., 2015). The last RNA-seq dataset, the largest collection of primary cells, was derived from 56 tissues produced by FANTOM5 project (Expression Atlas ID: E-MTAB-3358) (Forrest et al., 2014). Note that the second RNA-seq dataset originally contained 53 tissues derived from several cell lines and subregions of a single tissue. To reduce the number of redundant cell types, 30 representative tissues were arbitrarily selected.

In addition, expression levels of mouse lncRNA and mRNA genes were also estimated from RNA-seq data. The first RNA-seq dataset was derived from nine tissues harvested from an adult male C57BL/6 mouse (Expression Atlas ID: E-GEOD-74747) (Huntley et al., 2016). The second RNA-seq dataset was derived from three mouse strains (C57BL/6, DBA/2J, and CD1) (Expression Atlas ID: E-MTAB-2801) (Merkin et al., 2012). In this dataset, gene expression data across eight (C57BL/6 strain) or nine mouse tissues (DBA/2J and CD1 strains) is available.

Tissue-specificities of lncRNA and mRNA genes were investigated based on an outlier analysis of the RNA-seq data using ROKU (Kadota et al., 2006). For each lncRNA and mRNA gene, the tissues in which the gene was specifically expressed were detected based on its extremely high or low expression levels in one or a few tissues. These tissue-specificity data

allow the user to investigate the tissue-specific lncRNAs which regulate the expression levels of their target mRNAs through the base-pairing interactions. The tissue-specific lncRNA-RNA interactions derived from the aforementioned five human RNA-seq datasets and four mouse RNA-seq dataset are provided in LncRRIsearch (**Tables S1–S9**).

2.4. Integration With Subcellular Localization Data to LncRRIsearch

Subcellular localization dataset was downloaded from the LncAtlas database (Mas-Ponte et al., 2017). This dataset includes 15 human cell-line subcellular localization data, and the localization was quantified by "relative concentration index" (RCI), which was defined as log_2-transformed ratio of FPKM between two expression data. For example, high cytoplasmic/nuclear RCI means that the transcript tends to localize in cytoplasm rather than nucleus. For 14 cell-lines, two types of RCIs (cytoplasmic/nuclear and nuclear/cytoplasmic RCIs) are included in the dataset. On the other hand, for the K562 cell-line, five types of RCI data (Chromatin/Nucleus, Nucleolus/Nucleus, Nucleoplasm/Nucleus, Cell membrane/Cytoplasm, and Insoluble fraction/Cytoplasm RCIs) are additionally included in the dataset. These subcellular localized RNA-RNA interactions are also provided in LncRRIsearch (**Tables S10–S12**). The detail of the dataset was described in the original publication (Mas-Ponte et al., 2017). Note that mouse subcellular localization data are not included in LncRRIsearch.

2.5. Database Organization
In LncRRIsearch, tissue-specific expression data and subcellular localization data were stored in a series of MySQL databases. For

FIGURE 1 | Three workflows for investigating lncRNA–RNA interaction using LncRRIsearch.

FIGURE 2 | Investigation of lncRNA–RNA interactions using LncRRIsearch. In this example, detailed information about the interaction between NORAD lncRNA (used as query, Gene ID:ENSG00000260032, Transcript ID: ENST00000565493) and TUBB4A mRNA (used as target, Gene ID:ENSG00000104833, Transcript ID:ENST00000264071) is shown.

RNA-RNA interaction data, all pre-calculated SUMENERGY and MINENERGR scores were also stored in the databases, but the local base-pair data were not stored in the databases because the data size is too large. In the web service, the base-pairs are re-predicted by RIblast in real time when both the query and target RNAs are selected based on SUMENERGY or MINENERGY scores. However, because RIblast cannot predict interactions of long RNAs in real-time, base-pair prediction results for RNA sequences longer than 5,000 nt were stored in the databases, and the data is referenced in the web service.

3. RESULTS

LncRRIsearch provides three types of interaction prediction method (**Figure 1**): a name/ID based method, an expression pattern-based method, and a localization-based method.

3.1. Investigation of an RNA-RNA Interaction Based on Name or ID

Users firstly select target species (human or mouse) and the energy threshold (−12 or −16 kcal/mol), and then inputs name or ID of genes or transcripts (**Figure 1**). LncRRIsearch supports GENCODE gene/transcript names or IDs as input type, and either query lncRNA or target lncRNA/mRNA is required as input RNA. After specifying a gene of interest, several transcript isoforms derived from the gene are listed for selection of a single lncRNA transcript if multiple isoforms are encoded in the gene. For the selected lncRNA transcript (query transcript), all interacting RNAs (target transcripts) predicted by RIblast are provided. After selecting a single target transcript, the details of the RNA-RNA interaction between query and target transcripts are described (**Figure 2**). In this step, all local base-pairing interactions are listed, and users can download the

prediction results as a text file. In addition, the global base-pairing interaction is described as an image (The center left of **Figure 2**). In this figure, the query RNA and the target RNA are represented as a blue line and a red line, respectively, and the predicted interactions are displayed as gray or black lines between two RNAs. The color consistency means strength of interactions. For each local base-pairing interaction, text (output of RIblast) and a graphical view based on VARNA (Darty et al., 2009) are also provided (The lower left and the lower right of **Figure 2**).

3.2. Investigation of Tissue-Specific RNA-RNA Interactions

LncRRIsearch helps users to investigate lncRNA-RNA interactions exhibiting tissue-specific expression patterns (**Figure 1**). Users can select an RNA-seq dataset from four different RNA-seq studies and select a tissue of interest. For the selected tissue, one of three possible tissue-specific expression patterns for the query and target RNA transcripts should be selected: Query and target RNAs are specifically up-regulated in the same tissue; query RNAs are specifically up-regulated and target RNAs are down-regulated in the same tissue; or query RNAs are specifically down-regulated and target RNAs are up-regulated in the same tissue.

After selecting the tissue-specific expression pattern, the corresponding query and target RNAs predicted by RIblast are listed. In this step, once a query RNA is selected, the list of possible target RNAs is automatically updated for the selected query. By selecting the tissue-specific query and target RNAs, detailed information about interactions between the query and target RNAs is provided (**Figure 2**). In addition, the expression values of query and target RNAs are provided as a graphical view in the results page (The upper right of **Figure 2**).

3.3. Investigation of Subcellular Localized RNA-RNA Interactions

Users can investigate subcellular-localized human lncRNA-RNA interactions (**Figure 1**). Users firstly select a energy threshold and select a cell line of interest. For the selected cell line, a type of RCI and the threshold of RCI should be selected. Except for K562 cell line, users can choose which one of the nucleus/cytosol or cytosol/nucleus RCI. For K562 cell line, users have five choices of sub-compartments RCIs in addition to the above-mentioned two RCIs. The subsequent steps are the same as the investigation of tissue-specific RNA-RNA interactions. The RCI values of query and target RNAs are displayed as a graphical view in the results page (The center right of **Figure 2**).

4. DISCUSSION

We developed LncRRIsearch, which is a web server for comprehensive prediction of human and mouse lncRNA-mRNA and lncRNA-lncRNA interactions including tissue-specific expression and subcellular localization data. There are two advantages of LncRRIsearch over other lncRNA-RNA interaction databases or web services; the comprehensiveness of interaction prediction and the ability to investigate tissue-specific or subcellular localized interaction patterns.

We envision three future improvements of LncRRIsearch. The first is the development of real-time RNA-RNA interaction prediction software. Although LncRRIsearch provides comprehensive human and mouse lncRNA-RNA interaction based on GENCODE version 25 and M14, novel lncRNAs will be discovered in the future. Real-time prediction would be useful for the discoverers of new lncRNAs to investigate their interactions. The acceleration of RNA-RNA interaction prediction is still an important research topic. One possible direction is the simplification of the energy model. RIblast uses a complete nearest-neighbor energy model in the search step, but some researchers have reported that the use of an approximated energy model produces a marked increase in the calculation speed in exchange for only a slight decrease in the prediction accuracy (Tafer et al., 2011; Wenzel et al., 2012; Alkan et al., 2017).

The second improvement is the integration of the results of RNA-RNA interaction detection experiments. Recently, several high-throughput sequencing methods for the exhaustive identification of RNA-RNA interaction sites have been developed, including PARIS (Lu et al., 2016) and MARIO (Nguyen et al., 2016). Although only a few lncRNA-related interactions have been detected in these experiments, simultaneously displaying predicted and experimentally verified interactions (where available) should be useful for users. In addition, such data will encourage researchers to develop machine-learning-based RNA-RNA interaction prediction programs.

The third improvement is an increase in the number of target species. This improvement would enable us to not only investigate the lncRNA interactions of newly added species but also compare lncRNA interactomes between species. Nguyen et al. recently showed that the conservation of experimentally confirmed lncRNA-RNA interaction regions is high, although lncRNA generally lacks sequence conservation (Nguyen et al., 2016). This means that conservation information should be useful for the verification of predicted lncRNA-RNA interactions.

DATA AVAILABILITY

All datasets analyzed for this study are included in the manuscript and the **Supplementary Files**. LncRRIsearch is publicly available from http://rtools.cbrc.jp/LncRRIsearch/.

AUTHOR CONTRIBUTIONS

TF, JI, and MH conceived the study and wrote the manuscript. TF, JI, and YO processed the data. YO constructed the database. TF and JI equally contributed to this work. MH supervised this study. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00462/full#supplementary-material

REFERENCES

Abdelmohsen, K., Panda, A. C., Kang, M.-J., Guo, R., Kim, J., Grammatikakis, I., et al. (2014). 7SL RNA represses p53 translation by competing with HuR. *Nucleic Acids Res.* 42, 10099–10111. doi: 10.1093/nar/gku686

Alkan, F., Wenzel, A., Palasca, O., Kerpedjiev, P., Rudebeck, A. F., Stadler, P. F., et al. (2017). RIsearch2: suffix array-based large-scale prediction of RNA–RNA interactions and siRNA off-targets. *Nucleic Acids Res.* 45:e6. doi: 10.1093/nar/gkw1325

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611

Carpenter, S., Aiello, D., Atianand, M. K., Ricci, E. P., Gandhi, P., Hall, L. L., et al. (2013). A long noncoding RNA mediates both activation and repression of immune response genes. *Science* 341, 789–792. doi: 10.1126/science.1240925

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25, 1974–1975. doi: 10.1093/bioinformatics/btp250

de Hoon, M., Shin, J. W., and Carninci, P. (2015). Paradigm shifts in genomics through the FANTOM projects. *Mamm. Genome* 26, 391–402. doi: 10.1007/s00335-015-9593-8

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* 15, 7–21. doi: 10.1038/nrg3606

Forrest, A. R., Kawaji, H., Rehli, M., Baillie, J. K., De Hoon, M. J., Haberle, V., et al. (2014). A promoter-level mammalian expression atlas. *Nature* 507:462. doi: 10.1038/nature13182

Fukunaga, T., and Hamada, M. (2017). RIblast: an ultrafast RNA–RNA interaction prediction system based on a seed-and-extension approach. *Bioinformatics* 33, 2666–2674. doi: 10.1093/bioinformatics/btx287

Gong, C., and Maquat, L. E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* 470, 284–288. doi: 10.1038/nature09701

Gong, J., Shao, D., Xu, K., Lu, Z., Lu, Z. J., Yang, Y. T., et al. (2017). Rise: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res.* 46, D194–D201. doi: 10.1093/nar/gkx864

GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660. doi: 10.1126/science.1262110

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760–1774. doi: 10.1101/gr.135350.111

Hon, C.-C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J., Gough, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature* 543, 199–204. doi: 10.1038/nature21374

Huntley, M. A., Lou, M., Goldstein, L. D., Lawrence, M., Dijkgraaf, G. J., Kaminker, J. S., et al. (2016). Complex regulation of ADAR-mediated RNA-editing across tissues. *BMC Genomics* 17:61. doi: 10.1186/s12864-015-2291-9

Iwakiri, J., Terai, G., and Hamada, M. (2017). Computational prediction of lncRNA-mRNA interactions by integrating tissue specificity in human transcriptome. *Biol. Direct* 12:15. doi: 10.1186/s13062-017-0183-4

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208. doi: 10.1038/ng.3192

Kadota, K., Ye, J., Nakai, Y., Terada, T., and Shimizu, K. (2006). ROKU: a novel method for identification of tissue-specific genes. *BMC Bioinformatics* 7:294. doi: 10.1186/1471-2105-7-294

Kretz, M., Siprashvili, Z., Chu, C., Webster, D. E., Zehnder, A., Qu, K., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493, 231–235. doi: 10.1038/nature11661

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. doi: 10.1038/nature14248

Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., et al. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165, 1267–1279. doi: 10.1016/j.cell.2016.04.028

Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., Pulido, T. H., Guigo, R., and Johnson, R. (2017). Lncatlas database for subcellular localization of long noncoding RNAs. *RNA* 23, 1080–1087. doi: 10.1261/rna.060814.117

Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338, 1593–1599. doi: 10.1126/science.1228186

Nguyen, T. C., Cao, X., Yu, P., Xiao, S., Lu, J., Biase, F. H., et al. (2016). Mapping RNA-RNA interactome and RNA structure *in vivo* by MARIO. *Nat. Commun.* 7:12023. doi: 10.1038/ncomms12023

Quek, X. C., Thomson, D. W., Maag, J. L., Bartonicek, N., Signal, B., Clark, M. B., et al. (2014). lncRNAdb v2. 0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* 43(Database issue), D168–D173. doi: 10.1093/nar/gku988

Tafer, H., Amman, F., Eggenhofer, F., Stadler, P. F., and Hofacker, I. L. (2011). Fast accessibility-based prediction of RNA–RNA interactions. *Bioinformatics* 27, 1934–1940. doi: 10.1093/bioinformatics/btr281

Terai, G., Iwakiri, J., Kameda, T., Hamada, M., and Asai, K. (2016). Comprehensive prediction of lncRNA–RNA interactions in human transcriptome. *BMC Genomics* 17:12. doi: 10.1186/s12864-015-2307-5

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Tissue-based map of the human proteome. *Science* 347:1260419. doi: 10.1126/science.1260419

Wenzel, A., Akbaşli, E., and Gorodkin, J. (2012). RIsearch: fast RNA–RNA interaction search using a simplified nearest-neighbor energy model. *Bioinformatics* 28, 2738–2746. doi: 10.1093/bioinformatics/bts519

Yi, Y., Zhao, Y., Li, C., Zhang, L., Huang, H., Li, Y., et al. (2017). RAID v2.0: an updated resource of RNA-associated interactions across organisms. *Nucleic Acids Res.* 45, D115–D118. doi: 10.1093/nar/gkw1052

Next Generation Sequencing for Long Non-coding RNAs Profile for CD4+ T Cells in the Mouse Model of Acute Asthma

Zhengxia Wang[1†], Ningfei Ji[1†], Zhongqi Chen[1†], Chaojie Wu[1], Zhixiao Sun[1], Wenqin Yu[1,2], Fan Hu[3], Mao Huang[1*] and Mingshun Zhang[4,5*]

[1] Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [2] Department of Infectious Disease, Taizhou People's Hospital, Taizhou, China, [3] State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China, [4] NHC Key Laboratory of Antibody Technique, Nanjing Medical University, Nanjing, China, [5] Department of Immunology, Nanjing Medical University, Nanjing, China

Background and Aims: Although long non-coding RNAs (lncRNAs) have been linked to many diseases including asthma, little is known about lncRNA transcriptomes of CD4+ T cells in asthma. The present study aimed to explore the lncRNAs profile in the CD4+T cells from the mouse model of acute asthma.

Methods: Next generation sequencing for lncRNAs and mRNAs was performed on CD4+ T cells from asthma and control mice. Gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) pathway analyses were performed to predict the functions and signal pathways for the aberrant lncRNAs. The selected lncRNAs were further measured using quantitative real-time PCR (polymerase chain reaction) and observed in the fluorescence *in situ* hybridization (FISH). The lncRNA–mRNA co-expression network was constructed via Pearson's correlation coefficient and Cytoscape 3.6.

Results: Next generation sequencing revealed 36 up-regulated lncRNAs and 98 down-regulated lncRNAs in acute asthma compared with controls. KEGG pathway analysis showed that cytokine-cytokine receptor interaction had the highest enrichment scores. A co-expression network was constructed in which 23 lncRNAs and 301 mRNAs altered formed a total of 12424 lncRNA and mRNA pairs. To validate the RNA sequencing results, we measured the 4 different lncRNAs using qPCR. The lncRNA fantom3_9230106C11 was significantly reduced in CD4+ T cells of asthma. Bioinformatics analysis showed that lncRNA fantom3_9230106C11 had the potential to interact with many miRNAs and transcription factors related to Th2 differentiation.

Conclusion: This study provided the first evidence for different expression of lncRNAs of CD4+T cells in asthma and may serve as a template for further, larger functional in-depth analyses regarding asthma molecular lncRNAs.

Keywords: asthma, CD4+ T lymphocyte, long non-coding RNA, mRNA, next-generation sequencing

INTRODUCTION

Allergic asthma is a type of asthma provoked by allergens, including pollen, animal dander, fungal spores, or house dust mites (HDM) (Takyar et al., 2013; Hondowicz et al., 2016). Once processed and presented by dendritic cells, allergens promote the differentiation and expansion of Th2 cells, releasing type 2 cytokines (IL-4, IL-5, IL-13) and expressing the master transcription factor GATA-3. Th2 cells orchestrate eosinophil maturation and survival, airway hyperresponsiveness, and B cell isotype switching to IgE (Holgate, 2012). Consequently, Th2 cells play crucial roles in the cascade reactions of allergic asthma.

Non-coding RNAs emerged as essential players in Th2 cell differentiation and allergic asthma. Recently, we reported the aberrant microRNA (miRNA) profile in CD4[+] T cells from a murine model of acute asthma (Liu et al., 2018). In addition to miRNAs, long non-coding RNAs (lncRNAs), which comprise 200 nucleotides lacking putative open reading frames, may regulate CD4[+] T cell differentiation in asthma (Zhang F. et al., 2017). In allergic asthma, the lncRNA profile has been documented in primary airway smooth muscle cells, CD8[+] T cells or blood samples (Tsitsiou et al., 2012; Austin et al., 2017; Zhu et al., 2018). To our knowledge, lncRNA expression in the CD4[+] T cells in asthma remains elusive.

In this study, we performed next-generation sequencing and data mining to investigate the whole spectrum of transcriptional signatures (mRNAs and lncRNAs) of CD4[+] T cells in the murine model of acute asthma. Some lncRNAs were measured in CD4[+] cells *ex vivo* or in Th2 cells *in vitro* using real-time quantitative reverse transcription PCR (qRT-PCR) and fluorescence *in situ* hybridization (FISH). Moreover, cross-talk between mRNAs and lncRNAs associated with CD4[+] T cell differentiation was revealed.

MATERIALS AND METHODS

Establishment of the Model of Acute Asthma

Specific pathogen-free female C57BL/6J mice (18 to 22 g) aged 6 to 8 weeks were obtained from the College of Veterinary Medicine Yangzhou University (Yangzhou, China). All experiments that involved animal and tissue samples were performed in accordance with the guidelines and procedures approved by the Institutional Animal Care and Use Committee of Nanjing Medical University (IACUC-1709011).

The model of acute asthma was established as described previously (Liu et al., 2018). Briefly, the asthmatic mice ($n = 3$) were sensitized on days 0 and 14 by the intraperitoneal injection of 20 μg ovalbumin (OVA) (Grade V, Sigma-Aldrich) emulsified in 2 mg aluminum hydroxide gel (InvivoGen, San Diego, CA, United States) in a total volume of 200 μl. These sensitized mice were exposed to aerosolized 1% OVA in sterile saline for 30 min from day 20 to day 22, consecutively. The control subjects ($n = 3$) were sensitized and challenged using the same protocol as used for saline alone. All mice were monitored daily and were alive before sacrifice.

Twenty-four hours after the final challenge, lung function was evaluated by the direct measurement of lung resistance and dynamic compliance in restrained, tracheostomized, mechanically ventilated mice via the FinePointe RC System (Buxco Research Systems, Wilmington, NC, United States) under general anesthesia as described previously (Kerzerho et al., 2013). The sera were collected to measure total IgE using an ELISA kit according to the manufacturer's instructions (eBioscience, Thermo Fisher Scientific, United States). To determine lung tissue inflammation, the right upper lung lobe was removed, fixed, dehydrated and embedded in paraffin. The fixed embedded tissues were cut into 5 μm sections on a Leica model 2165 rotary microtome (Leica, Nussloch, Germany), and the tissue slides were stained with hematoxylin and eosin (H&E).

CD4[+] T Cell Purification From the Spleen

Mice were anesthetized by i.p. injection of a mixture of 10 mg/kg xylazine (MTC Pharmaceuticals, Cambridge, ON, Canada) and 200 mg/kg ketamine hydrochloride (Rogar/STB, London, ON, Canada). The anaesthetized mice were sacrificed with cervical dissociation. The spleen was removed, ground and prepared into single cell suspensions. CD4[+] T cells in the spleen were sorted using CD4 (L3T4) micro beads (130-049-201, Miltenyi Biotec, United States). Briefly, the single cell suspensions were incubated with CD4 (L3T4) micro beads. The magnetically labeled cells were flushed and collected. Finally, the purity of CD4[+] T cells was quantified using anti-CD4-APC antibodies (17-0041-81, eBioscience).

RNA Isolation, Library Preparation, and Sequencing

Total RNA was isolated using the miRNeasy Mini Kit (Qiagen, Germany) and stored at −80°C until use. RNA purity was assessed using the ND-1000 Nanodrop. Each RNA sample had an A260:A280 ratio above 1.8 and A260:A230 ratio above

TABLE 1 | Primers used in qRT-PCR.

Name	Sequence
β-actin-F	GAGAAGCTGTGCTATGTTGCT
β-actin-R	CTCCAGGGAGGAAGAGGATG
fantom3_9230106C11-F	CTCTGTCCTGGAAAGTGGTGTC
fantom3_9230106C11-R	TGCTGCCAAGCTAGATGGTC
fantom3_4933428M03-F	TGTCATGCTGTGGTAACAGTGA
fantom3_4933428M03-R	ACACCAGGTAGATATGCAAGGAA
fantom3_F630107E09-F	TTCCTGTTGCCCCAACTGTAG
fantom3_F630107E09-R	GGCAGGGAGTCTTCTACTTCC
T-bet-F	AATCGACAACAACCCCTTTG
T-bet-R	AACTGTGTTCCCGAGGTGTC
IFN-γ-F	ACAATGAACGCTACACACTGC
IFN-γ-R	CTTCCACATCTATGCCACTTGAG
GATA3-F	GAACCGGCCCCTTATCAAG
GATA3-R	CAGGATGTCCCTGCTCTCCTT
IL-4-F	AACTCCATGCTTGAAGAAGAACTC
IL-4-R	CCAGGAAGTCTTTCAGTGATGTG
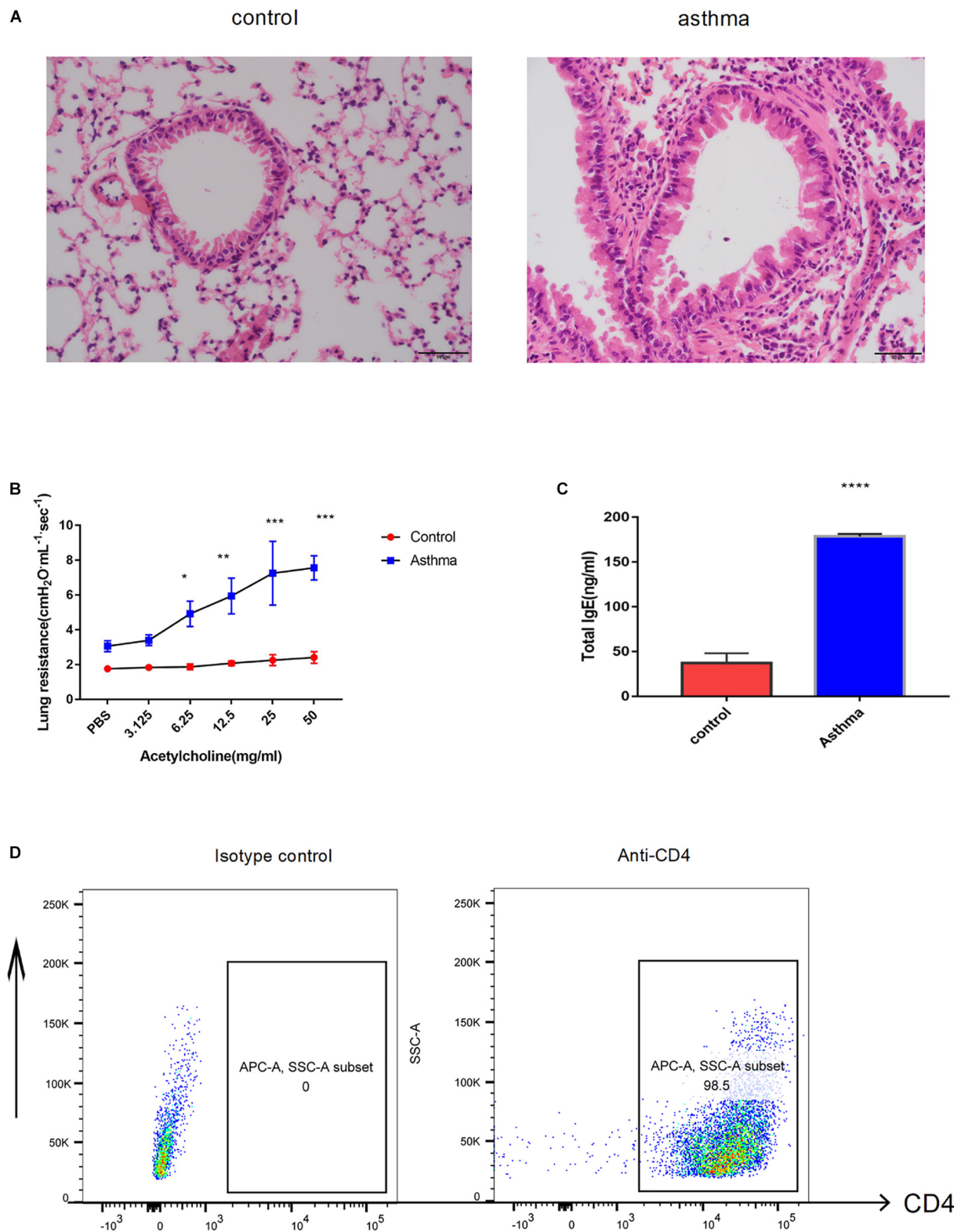
FIGURE 1 | Ovalbumin (OVA)-induced acute model of asthma. **(A)** Pulmonary inflammatory response was increased in the mice challenged with OVA, bar = 20 μm. **(B)** The airway hyperresponse was elevated in the mice from the asthma group, $*p < 0.05$ **(C)**. Sera IgE was significantly increased in the mice from the asthma group, $*p < 0.05$. **(D)** CD4$^+$ T cells magnificently sorted from the spleen were validated by flow cytometry. Representative results of two independent experiments with five mice in each group, $**p < 0.01$, $***p < 0.01$, $****p < 0.0001$.

2.0. RNA integrity was evaluated using the Agilent 2200 TapeStation (Agilent Technologies, United States) and each sample had the RIN above 7.0. Briefly, rRNAs were removed from Total RNA using Epicenter Ribo-Zero rRNA Removal Kit (illumina, United States) and fragmented to approximately 200bp. Subsequently, the purified RNAs were subjected to first strand and second strand cDNA synthesis following by adaptor ligation and enrichment with a low-cycle according to instructions of NEBNext® Ultra^{TM} RNA Library Prep Kit for Illumina (NEB, United States).

The purified library products were evaluated using the Agilent 2200 TapeStation and Qubit® 2.0 (Life Technologies, United States) and then diluted to 10 pM for cluster generation *in situ* on the pair-end flow cell followed by sequencing (2 × 150 bp) on HiSeq 3000. The clean reads were obtained after removal of reads containing adapter, ploy-N and at low quality from raw data. Sequence data were mapped to mouse reference genome mm10 with TopHat v2.0.13. gfold v1.1.2 was subsequently employed to count the number of reads mapped

to each gene. Differential expression was assessed by DEGseq using RPKM as input. Differentially expressed genes were chosen according to the criteria of fold change > 2 and adjusted P-value < 0.05. All the differentially expressed genes were used for heat map analysis and KEGG enrichment analyses. For KEGG enrichment analysis, a P-value < 0.05 was used as the threshold to determine significant enrichment of the gene sets.

Bioinformatics Data Analysis and Data Mining

The transcriptome was assembled using Cufflinks and Scripture based on the reads mapped to the reference genome. The assembled transcripts were annotated using the Cuff compare program from the Cufflinks package. The unknown transcripts were used to screen for putative lncRNAs. Three computational approaches, including CPC/CNCI/Pfam, were combined to sort non-coding RNA candidates from putative protein-coding RNAs in the unknown transcripts. Putative protein-coding RNAs were

FIGURE 2 | Volcano plots and heat maps of lncRNA expression between the asthma and control groups. **(A)** Volcano plot assessment of lncRNA expression in CD4$^+$ T cells between the asthma and control groups. Red dots represent different lncRNAs ($p < 0.05$) showing fold changes >=2 or <= −2. **(B)** Heat map analysis of differentially expressed lncRNAs between the asthma and control groups. Green indicates low expression, and red indicates high expression. A4, A5, and A7 were the control group; B4, B6, and B7 were CD4$^+$ T cells from the asthma group.

filtered out using a minimum length and exon number threshold. Transcripts with lengths greater than 200 nt and with more than two exons were selected as lncRNA candidates and further screened using CPC/CNCI/Pfam, which distinguished protein-coding genes from non-coding genes.

The UCSC genome browser was used to locate lncRNA fantom3_9230106C11. The predicted potential target genes whose loci were within a 10-kb window upstream or downstream of the given aberrantly expressed lncRNA were considered *cis*-regulated genes. Other genes in the co-expression network were identified as *trans*-regulated according to complementary base pairing by LncTar (Li et al., 2015). In addition, Bibiserv (Sczyrba et al., 2003) and RNA22 (Miranda et al., 2006) were used to predict the interactive miRNAs. Different mRNAs in the our NGS were compared with known genes about asthma in the human disease database Malacards associated with asthma[1] and asthma-associated genes[2] using Wayne chart[3].

CD4+ T Cell Differentiation *in vitro*

Naive CD4+ T cells in the spleen were sorted using a Naive CD4+ T Cell isolation kit (130-104-453, Miltenyi Biotec, United States). Briefly, the single cell suspensions were incubated with biotin-antibody cocktails, which depleted the non-T cells and memory CD4+ T cells. The untouched naive CD4+ T lymphocytes were seeded onto anti-CD3 (85-16-0031-85, eBioscience, 1 mg/ml) precoated 96-well plates. For Th1 polarization, naive CD4+ T cells were further stimulated with anti-CD28 (85-16-0281-85, eBioscience, 2 μg/ml), IL-2 (212-12, PeproTech, 20 ng/ml), IL-12 (210-12, PeproTech, 50 ng/ml) and anti-IL-4 (85-16-7041-85, eBioscience, 10 μg/ml). For Th2 polarization, naive CD4+ T cells were further stimulated with anti-CD28 (85-16-0281-85, eBioscience, 2 μg/ml), IL-2 (212-12, PeproTech, 20 ng/ml), IL-4 (214-14, PeproTech, 200 ng/ml), anti-IFN-γ (85-16-7311-85, eBioscience, 10 μg/ml), and anti-IL-12 (85-16-7123-85, eBioscience, 10 μg/ml). Half of the complete RPMI-1640 medium (16000-044, Gibco, United States) with 10% FBS (Gibco, United States) was replaced with fresh medium to maintain the cytokine environments. The CD4+ T cells were harvested and analyzed on day 5.

Flow Cytometry Analysis *ex vivo* and *in vitro*

For nuclear protein assays, spleens from animal experiments were prepared as single-cell suspensions. After RBC (red blood cell) lysis, the cells were stained with CD16/CD32 FcR (Fc Receptor) blocking antibody, Fixable Viability Dye eFluor™ 506 and anti-CD4-APC (eBioscience, 17-0041-83). For intracellular staining, the cells were fixed and permeabilized (eBioscience, 00-5523-00) according to the manufacturer's instructions, and then intracellular products were stained. Flow cytometry was performed with anti-Gata3-PE-Cy7 (BD Biosciences, 560405), anti-T-bet-PE-Cy7 (eBioscience, 25-5825-80) and isotype controls.

[1] https://www.malacards.org/
[2] https://www.ncbi.nlm.nih.gov/gene/?term=asthma
[3] http://bioinformatics.psb.ugent.be/webtools/Venn/

For cytokine assays, T cells were restimulated on day 5 of culture for 5 h with 20 nM PMA (70-CS1001, Multiscience, China) and blocked with BFA (70-CS1002, Multiscience, China). After 5 h, the cells were harvested, gently washed with PBS containing 1% bovine serum albumin (BSA) and fixed in IC fixation buffer (00-822-49, eBioscience, United States). The fixed cells were permeabilized with permeabilization buffer (00-8333-56, eBioscience, United States) and stained with PE-anti-IFN-γ (85-12-7311-84, eBioscience, United States), PE-anti-IL4 (85-12-7041-83, eBioscience, United States) or the isotype control. The stained CD4+ T lymphocytes were analyzed on the BD FACSCalibur, and the data were analyzed by FlowJo software (TreeStar).

Real-Time Quantitative Polymerase Chain Reaction

RNAs from cells or tissues were isolated using a miRNeasy mini kit. The cDNA was synthesized using the Prime Script RT Reagent Kit (TaKaRa, Kyoto, Japan) following the manufacturer's instructions. The quantitative PCR analysis was performed using the CFX96 system (Bio-Rad laboratories) in conjunction with ready-to-use fast-start SYBR Premix Ex Taq II (TaKaRa, Kyoto, Japan). The cycling conditions were 95°C for 30 s, followed by 95°C for 5 s and 60°C for 30 s for up to 40 cycles and dissociation at 95°C for 15 s, 60°C for 30 s and a final extension at 95°C for 15 s. The relative abundance of gene targets was determined by the comparative CT (cycle threshold) number normalized against tested β-actin comparative CT. The primers used are shown in **Table 1**, which were designed by RiboBio Institute (Guangzhou, China) or from primerbank by using primer5.

TABLE 2 | Top 20 lncRNAs in CD4+ T cells between asthma and control groups.

LncRNA	control	asthma	*P*-value
fantom3_E230007F07	8.78	0.61	0.0006
fantom3_4933428M03	14.37	1.08	< 0.0001
fantom3_B230105C15	10.41	0.99	0.0005
fantom3_E030028L20	12.64	1.76	0.0003
fantom3_A530047B01	21.82	4.05	< 0.0001
fantom3_9230106C11	24.67	5.02	0.0002
fantom3_C130024N02	121.92	26.59	< 0.0001
fantom3_9230022E05	100.13	22.58	0.0001
fantom3_D130005M09	37.46	9.67	< 0.0001
fantom3_6430516O08	25.14	7.87	0.0006
fantom3_F630107E09	3.96	20.46	0.0001
fantom3_9930009M05	3.87	20.74	0.0006
fantom3_9530065C24	4.89	27.20	0.0001
fantom3_D130059M16	3.25	21.04	< 0.0001
fantom3_3021401C12	2.21	23.19	< 0.0001
fantom3_I830062N05	5.59	62.54	< 0.0001
fantom3_F730001C01	4.58	52.60	< 0.0001
fantom3_9830118H07	4.92	58.38	< 0.0001
fantom3_E130019F23	1.31	16.65	< 0.0001
fantom3_4930418C01	0.33	9.44	< 0.0001

The P-value indicates the difference between the control and asthma groups.
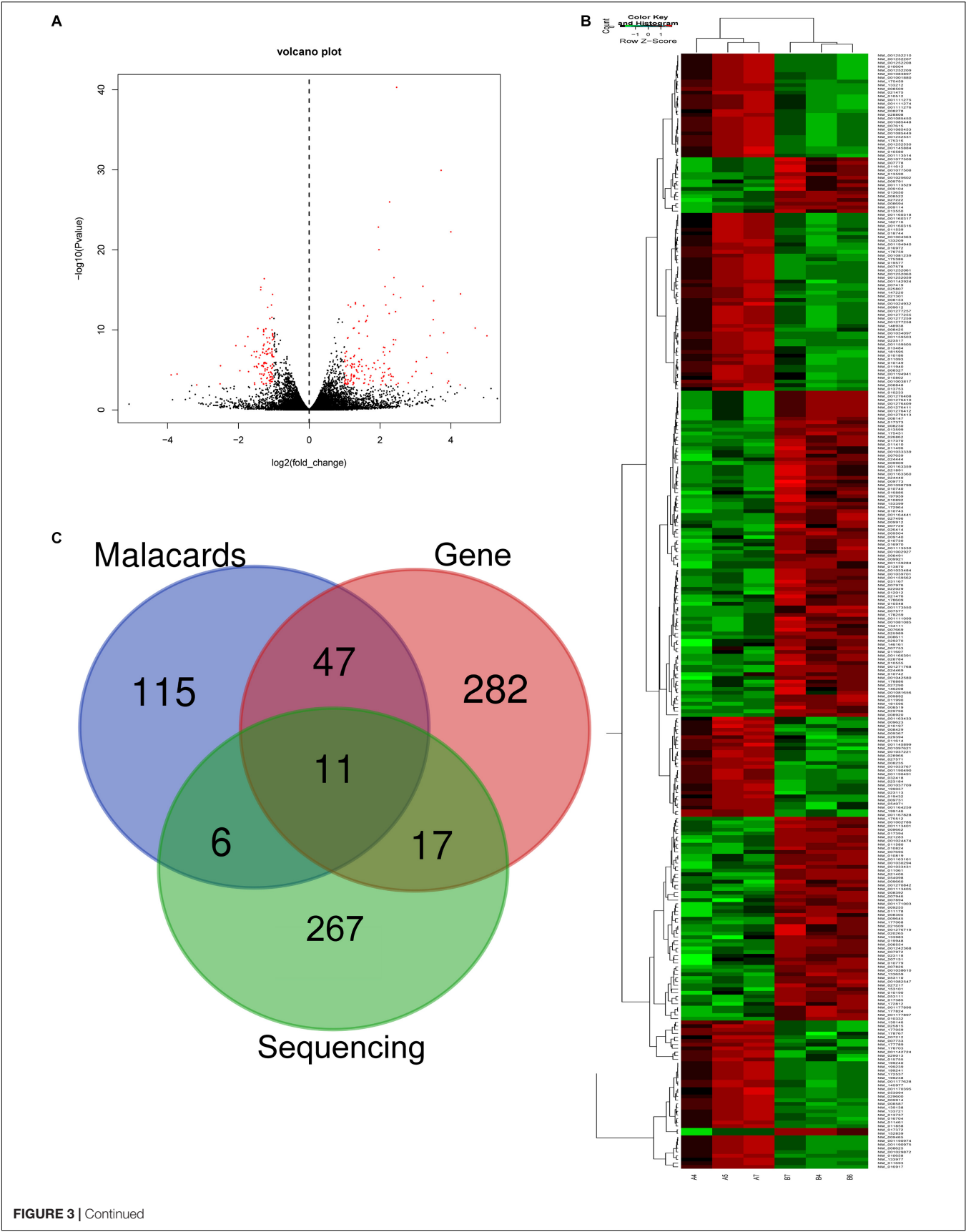
A

B

C

FIGURE 3 | Continued

FIGURE 3 | Volcano plots and heat map of mRNA expression between asthma and control groups. **(A)** Volcano plot assessment of gene expression in CD4$^+$ T cells between asthma and control groups. Red dots represent different mRNAs ($p < 0.05$) showing fold changes >=2 or <= −2. **(B)** Heat map analysis of differentially expressed mRNAs between the asthma and control groups. Green indicates low expression, and red indicates high expression. A4, A5, and A7 were the control groups; B4, B6, and B7 were CD4$^+$ T cells from the asthma group. **(C)** The Venn diagram of genes about asthma from sequencing compared with gene database and malacards database.

Correlation and Co-expression Analysis

The co-expression analysis was based on Pearson's correlation coefficient. Considering the influence of random factors, it is found that using Pearson correlation coefficient only is not strict. Therefore, when the number of samples is less than eight, due to the small number of samples, we use the mixed washing method to calculate the P-value to further screen.

The co-expression relationship was screened according to Pearson correlation coefficient and P-value. Differentially expressed lncRNAs and mRNAs with fold changes ≥2 and $p < 0.05$ were analyzed. For each lncRNA-mRNA pair, the Pearson correlation (COR) was calculated to identify significantly correlated pairs. The Pearson correlation value cutoff was 0.95 and $p < 0.05$. To create a visual representation, a lncRNA-mRNA regulatory network was constructed using Cytoscape 3.6.

Fluorescence *in situ* Hybridization (FISH)

Fluorescence *in situ* hybridization assays were performed using the Ribo^{TM} Fluorescent *in situ* Hybridization Kit and the Ribo^{TM} lncRNA FISH Probe Mix (Ribo, Guangzhou, China) according to the manufacturer's protocols. Briefly, the cells were fixed in 4% formaldehyde for 10 min and then washed with PBS. The cells were incubated with 20 μmol/L lncRNA FISH probe mix at 37°C overnight. After washing, the FISH preparations were counterstained with DAPI observed in confocal microscopy for appropriate fluorescence filter sets (Zeiss, Oberkochen, Germany). The lncRNA probe labeled with Cy3 was designed and synthesized by RiboBio Co., Ltd., Ribo^{TM} U6 and Ribo^{TM} 18S were used as reference controls for the subcellular localization of lncRNA.

Statistical Analysis

The statistical analysis was performed using GraphPad Prism version 5.0 (GraphPad, San Diego, CA, United States). The results for variables that were normally distributed are displayed as the means ± SEM. An ANOVA was performed to establish equal variance, and a 2-tailed Student's t-test with Bonferroni correction was applied to determine statistical significance, which was defined as $P < 0.05$.

RESULTS

Establishment of an Acute Model of Asthma

As the well-established murine model of asthma, OVA challenge provoked evident pulmonary inflammation, airway hypersensitivity and higher IgE in serum (**Figures 1A–C**). CD4$^+$ T cells were magnetically sorted from the spleen, and the purity

was validated by flow cytometry. As shown in **Figure 1D**, 98.5% of the cells were CD4 positive (**Figure 1D**).

LncRNA Profile of CD4$^+$ T Cells in Asthma

To obtain a global overview of CD4$^+$ T cells in the asthma transcriptome, we constructed and sequenced 6 RNA-Seq libraries, including controls ($n = 3$) and asthma mice ($n = 3$). All of the data have been submitted to Sequence Read Archive (SRA^4) with the accession number of PRJNA540404. In total, 134 lncRNAs were significantly altered in expression, including 98 downregulated lncRNAs and 36 upregulated lncRNAs (**Figure 2**). The top 10 decreased lncRNAs and top 10 increased were listed in **Table 2**.

The mRNA Profile of CD4$^+$ T Cells in Asthma

In RNA sequencing, we explored not only the lncRNAs but also the mRNAs from the same samples. With respect to mRNAs, 141 mRNAs were significantly downregulated and 160 mRNAs were significantly upregulated in the CD4$^+$ T cells from

---

^{4}www.ncbi.nlm.nih.gov/sra

TABLE 3 | Top 20 mRNAs in CD4$^+$ T cells between asthma and control groups.

mRNAs	control	asthma	P-value
Kbtbd12	11.39	0.76	< 0.0001
Kbtbd12	14.37	1.08	< 0.0001
Hunk	13.10	1.45	0.0008
Slc6a1	24.04	4.20	0.0006
Vsig4	25.70	4.53	< 0.0001
Fstl4	57.58	13.73	< 0.0001
Col19a1	32.03	8.25	< 0.0001
Trim30d	121.32	34.58	< 0.0001
Trim30d	94.45	28.47	< 0.0001
Agmo	44.13	13.48	< 0.0001
Prg2	22.40	296.72	< 0.0001
F10	2.23	31.00	< 0.0001
Ascl2	2.65	39.68	< 0.0001
9830107B12Rik	0.97	14.54	0.0004
9830107B12Rik	0.97	14.86	0.0003
Irg1	5.78	92.58	< 0.0001
9830107B12Rik	0.67	12.84	0.0008
Epx	4.59	88.82	< 0.0001
Ear6	1.02	26.46	< 0.0001
Ear7	0.69	22.45	< 0.0001

The P-value indicates the difference between the control and asthma groups.

asthma (**Figure 3**). The top 10 decreased and top 10 increased mRNAs are listed in **Table 3**, including reduced Kbtbd12, Hunk, and Slc6a1 and elevated Ear7, Ear6, and Epx. Among of the 301 differently expressed mRNAs, 17 mRNAs were in the human disease database Malacards associated with asthma (see text footnote 1), 28 mRNAs were in the asthma-associated genes[5]. More importantly, 11 mRNAs were overlapped in our sequencing data, MalaCards database and asthma-associated genes, including IL4, IL-10, MMP9, VCAM-1, Il1rl1, Alox5, Il1rn, Ccr3, Cysltr1, Epx, and Ccl24.

---

[5]https://www.ncbi.nlm.nih.gov/gene

The lncRNA-mRNA Co-expression Network

To further examine the function of these differentially expressed lncRNAs in CD4$^+$ T cells in asthma, we constructed an lncRNA-mRNA co-expression network between 301 differentially expressed mRNAs and 23 differentially expressed lncRNAs. The results showed that the co-expression network comprised 12424 connections between lncRNAs and mRNAs. Notably, fantom3_4933428M03 and fantom3_9230106C11 exhibited a high degree of connectivity, suggesting that these two lncRNAs may play key roles (**Figure 4** and **Table 4**). Moreover, our results highlighted the potential internal adjustment correlations
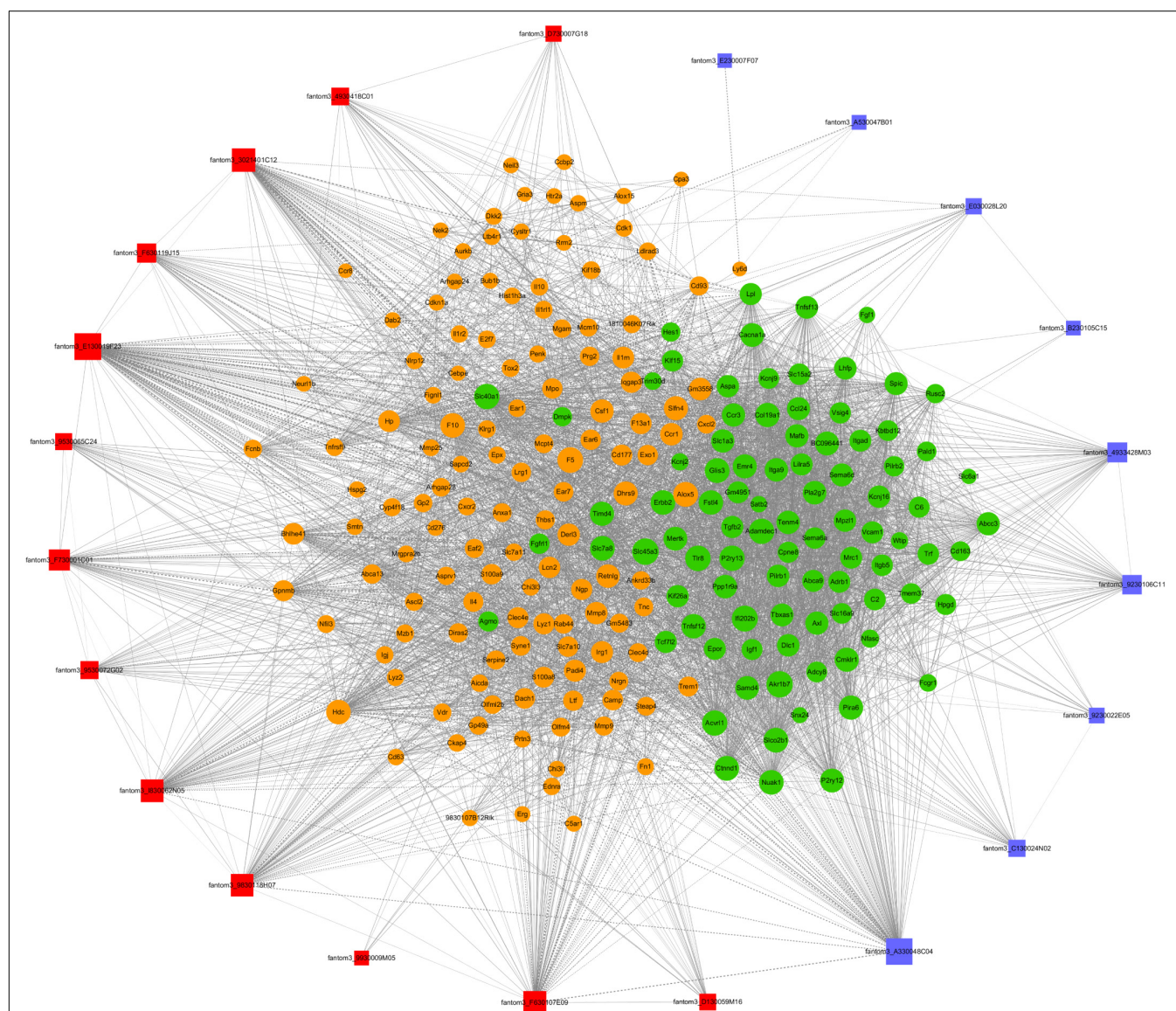
FIGURE 4 | LncRNA–mRNA co-expression networks. Twenty-three lncRNAs and 301 mRNAs were included in the co-expression network | COR| >0.95; $p < 0.05$). The red boxes indicate upregulated lncRNAs, and the blue boxes indicate downregulated lncRNAs. The green nodes indicate downregulated mRNAs, and the yellow nodes indicate upregulated mRNA. The value of "degree" indicates the size of the boxes or nodes. The width of the lines denotes the Pearson correlation coefficient, where the full line indicates a positive correlation and the imaginary line indicates a negative correlation.

TABLE 4 | The "degree" of lncRNAs in the co-expression network.

lncRNA	style	Degree
fantom3_3021401C12	Up	126
fantom3_4930418C01	Up	43
fantom3_9230022E05	Down	102
fantom3_9230106C11	Down	86
fantom3_9530065C24	Up	30
fantom3_9530072G02	Up	39
fantom3_9830118H07	Up	99
fantom3_A330048C04	Down	160
fantom3_C130024N02	Down	39
fantom3_D130059M16	Up	33
fantom3_E130019F23	Up	158
fantom3_F630107E09	Up	108
fantom3_F630119J15	Up	52
fantom3_F730001C01	Up	78
fantom3_A530047B01	Down	120
fantom3_I830062N05	Up	94
fantom3_4933428M03	Down	185

* <30° was not shown.

between the differentially expressed lncRNAs and mRNAs in CD4$^+$ T cells between the asthma and control groups.

Validation of RNA-Seq Data With Real-Time PCR

To confirm the differentially expressed gene data, we further analyzed some dysregulated lncRNAs using qRT-PCR *ex vivo* and *in vitro*. In OVA-induced asthma, Th2 cells orchestrate the pathological cascades reactions. As shown in **Figure 5A**, CD4$^+$ T cells from the asthma group showed higher expression of the Th2 master transcription factor Gata-3 and decreased expression of the Th1 master transcription factor T-bet. Different from the sequencing data, the expression of LncRNA fantom3_4933428M03 or fantom3_F630107E09 was unexpectedly similar in the CD4$^+$ T cells from either the asthma or the control group. However, lncRNA fantom3_9230106C11 was in accordance with sequencing data, which was significantly decreased in the asthma CD4$^+$ T cells (**Figure 5B**). To further explore the data reliability, we induced Th1 or Th2 cells *in vitro* (**Figures 6A,B**). As expected, the expression of lncRNA fantom3_9230106C11 was significantly decreased in Th2 cells (**Figure 6C**). FISH assays indicated that lncRNA fantom3_9230106C11 was localized in the cytoplasm of CD4$^+$ T cells (**Figure 7**).

Prediction of lncRNA fantom3_9230106C11 Targets

The above results indicated that lncRNA fantom3_9230106C11 may be involved in Th2 cell differentiation in asthma. LncRNA fantom3_9230106C11 was located at chr6:34412743–34415062 (GenBank: AK033773.1), overlapping with intron 3, exon 4, and intron 4 of Akr1b7 (chr6: 34412362–34423137). Sequence blast analysis shows that lncRNA fantom3_9230106C11 is
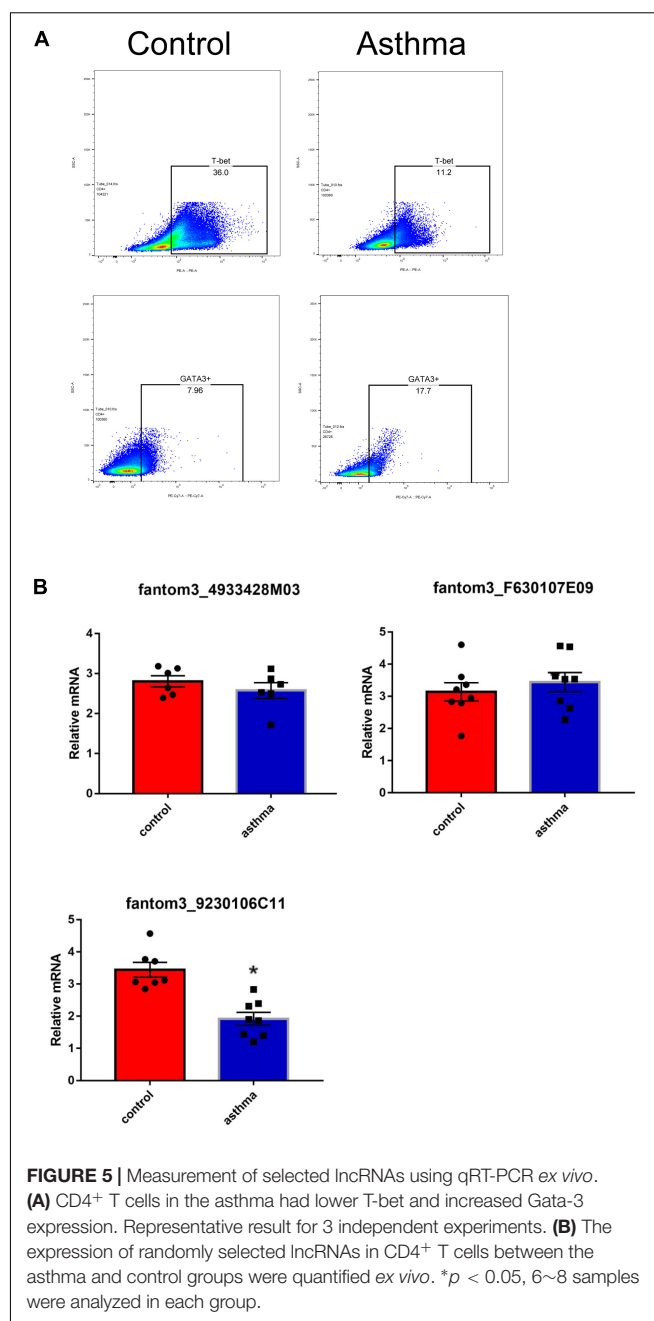
FIGURE 5 | Measurement of selected lncRNAs using qRT-PCR *ex vivo*. **(A)** CD4$^+$ T cells in the asthma had lower T-bet and increased Gata-3 expression. Representative result for 3 independent experiments. **(B)** The expression of randomly selected lncRNAs in CD4$^+$ T cells between the asthma and control groups were quantified *ex vivo*. *$p < 0.05$, 6~8 samples were analyzed in each group.

99.8% similar with Mus musculus lncRNA URS00009B6C6F[6]. In the NONCODE (current version v5.0) http://www.noncode. org/, Mus musculus lncRNA URS00009B6C6F is renamed with NONMMUT056297.2, which is highly expressed in the hippocampus (Data Source:ERP000591).

We further explored the potential candidates that may interact with the lncRNA. Diverse transcription factors, such as T-bet, Gata-3, c-maf, stat4, stat5, stat6, jun-b, Dec-2, IRF4, Notch, Gfi-1, and YY1, may be involved in Th2 differentiation (Hwang et al., 2013; Lee, 2014). We used LncTar (Li et al., 2015) and

---

[6]https://rnacentral.org/rna/URS00009B6C6F

FIGURE 6 | Quantification of LncRNA in Th1 and Th2 cells *in vitro*. **(A)** Typical cytokines and transcription factors for Th1 or Th2 were measured using qRT-PCR. Th1 cells exhibited elevated T-bet and IFN-γ levels. In contrast, Th2 cells expressed higher levels of Gata-3 and IL-4. In each group, 4 samples were analyzed. **(B)** Intracellular cytokines in Th1 and Th2 cells were measured by flow cytometry. Representative result for 3 independent experiments. **(C)** LncRNA fantom3_9230106C11 was measured in Th1 and Th2 cells *in vitro* using qRT-PCR. In each group, 5 samples were analyzed. *$p < 0.05$.

found that lncRNA fantom3_9230106C11 had no effects on the mRNAs of these transcription factors. The Consite Predication Database (Mahmood et al., 2018) showed that transcription factors Gata1, Gklf, SOX17, S8, and Ahr-ARNT may be bound to the promoter of LncRNA fantom3_9230106C11. In addition to transcription factors, a wide range of miRNAs contribute to Th2 differentiation, i.e., miR-17~92, miR-29, miR-126, miR-132-3p, miR-148a, mir-24, mir-27, mir-19, and mir-155 were confirmed to regulate Th2 differentiation (Istomine et al., 2016; Pua et al., 2016). In the LncTar, Bibiserv, and RNA 22 analysis, lncRNA fantom3_9230106C11 was predicted to bind with miR-19 and other miRNAs. Collectively, the potential candidates of

LncRNA fantom3_9230106C11 may include Akr1b7, Gata1 and miRNAs (**Figure 8**).

DISCUSSION

Allergic bronchial asthma is a common chronic inflammation with well-defined pathological features, including intermittent airway hyperresponsiveness, pulmonary eosinophil infiltration, and excessive mucus secretion (Chang et al., 2015). Numerous studies have demonstrated that allergic asthma is driven predominantly by a Th2 type of immune response in both human

FIGURE 7 | Intracellular location of fantom3_9230106C11 expression in CD4$^+$ T cells by FISH. The nucleus was stained with DAPI, and lncRNA was stained with probe-cyt3. Ribo^{TM} U6 and Ribo^{TM} 18S were used as reference controls for subcellular localization.
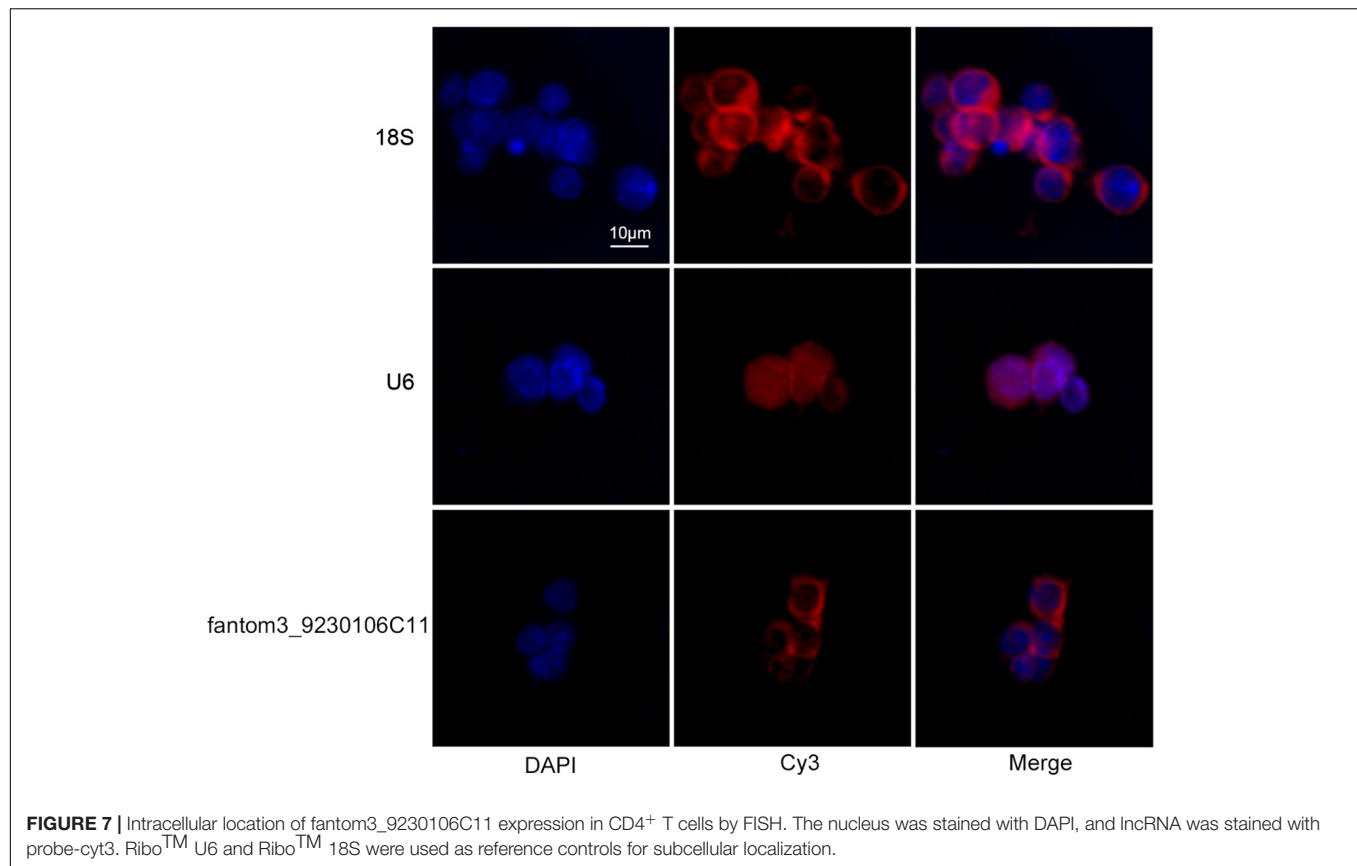
and mouse models of asthma (Kaiko and Foster, 2011). LncRNAs have been linked with airway smooth muscle cells in asthma (Zhang et al., 2016; Austin et al., 2017; Yu et al., 2017; Zhang X.-Y. et al., 2017). However, to our knowledge, there is no report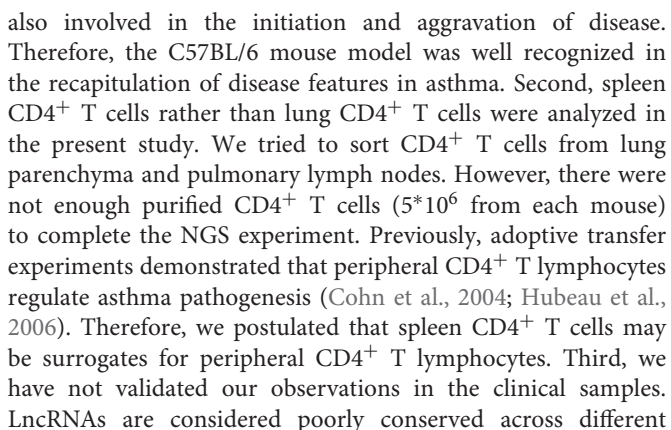 on lncRNA expression in CD4$^+$ T cells in asthma. Therefore, we constructed a mouse model of acute asthma and screened RNA transcripts (mRNAs, lncRNAs) from CD4$^+$ T cells via next-generation sequencing.

In our study, we identified 134 lncRNAs and 301 mRNAs abnormally expressed in CD4$^+$ T cells of asthma compared with controls. In the mRNA expression pattern analysis, elevated Ear7, Ear6 and Epx were closely associated with eosinophils and type 2 inflammations (Ochkur et al., 2017). IL-4, IL-13, and IL-21 (Lajoie et al., 2014), which promoted Th2 differentiation and allergic asthma, were also significantly increased in the CD4$^+$ T cells from the asthma mouse model. However, the Th1 master transcription factor T-bet and the Th2 master transcription factor Gata3 were comparable in CD4$^+$ T cells from both control and asthma mice, suggesting that Th2 differentiation in asthma may be independent of Gata3 (O'Shea and Paul, 2010). In the lncRNA profile analysis, the roles of a wide range of lncRNAs with varied expression in asthma CD4$^+$ T cells were largely unknown. In the lncRNA and mRNA co-expression network, lncRNAs may regulate diverse mRNAs, including Ear6 and Epx, which are required for Th2 differentiation and asthma etiology.

Using real-time qRT-PCR, we demonstrated that lncRNA fantom3_9230106C11 was decreased in CD4$^+$ T cells

from asthma mice *ex vivo* and in Th2 cells *in vitro*. By filtering the aberrantly expressed genes located near the lncRNA fantom3_9230106C11s, we found that lncRNAs might regulate the transcription of Akr1b7 in *cis*. lncRNA fantom3_9230106C11s covered intron 3, exon 4, and intron 4 of Akr1b7. *Akr1b7* encodes aldose reductase-related protein 1, which may catalyze xenobiotic aromatic aldehydes (Liu et al., 2009). The roles of *Akr1b7* in Th2 differentiation or asthma, however, remain uncertain. The bioinformation prediction indicated that the transcription factors Gata1, Gklf, and Ahr-ARNT may be potential candidates for lncRNA fantom3_9230106C11. Gata1 served as a surrogate for Gata3 in its canonic role of programming Th2 gene expression (Sundrud et al., 2005). Gklf (gut-enriched Krüppel-like factor; or Kruppel-like factor 4, KLF4) was required in Th2 cell responses *in vivo* (Tussiwand et al., 2015). Ahr-ARNT, which may regulate IL-33, IL-25, and TSLP, was closely associated with allergic severe asthma (Weng et al., 2018). Moreover, considering that lncRNA fantom3_9230106C11 resides in the cytoplasm, it may interact with miR-19 and other miRNAs, which are closely associated with Th2 differentiation.

Our study was not without limitations. First, BALB/c and C57BL/6 mouse strains are the two most commonly used in asthma models. Compared with the C57BL/6 mice used in this study, BALB/c mice were more prone to a Th2 immune response (Gueders et al., 2009). Although allergic asthma was Th2-dominated, Th1 and other CD4$^+$ T subpopulations were

FIGURE 8 | Bioinformatics analysis of lncRNA fantom3_9230106C11. **(A)** Putative transcription factor binding sites found along the lncRNA fantom3_9230106C11 using Consite and TF scores >95%. **(B)** The Venn diagram of potential miRNA candidates using LncTar, Bibiserv, and RNA22.

also involved in the initiation and aggravation of disease. Therefore, the C57BL/6 mouse model was well recognized in the recapitulation of disease features in asthma. Second, spleen CD4⁺ T cells rather than lung CD4⁺ T cells were analyzed in the present study. We tried to sort CD4⁺ T cells from lung parenchyma and pulmonary lymph nodes. However, there were not enough purified CD4⁺ T cells (5*10⁶ from each mouse) to complete the NGS experiment. Previously, adoptive transfer experiments demonstrated that peripheral CD4⁺ T lymphocytes regulate asthma pathogenesis (Cohn et al., 2004; Hubeau et al., 2006). Therefore, we postulated that spleen CD4⁺ T cells may be surrogates for peripheral CD4⁺ T lymphocytes. Third, we have not validated our observations in the clinical samples. LncRNAs are considered poorly conserved across different species (Ma et al., 2013). However, the conservation may be multidimensional (Diederichs, 2014). The expression of aberrant lncRNAs (lncRNA fantom3_9230106C11) in the clinical samples should be evaluated.

CONCLUSION

In conclusion, we conducted a comprehensive analysis of lncRNA profiles in CD4⁺ T cells from an asthma model using next-generation sequencing. The co-expression network of lncRNAs and mRNAs was constructed. The present study provided a platform for elucidating the roles of lncRNAs in Th2 differentiation and asthma pathogenesis.

ETHICS STATEMENT

AUTHOR CONTRIBUTIONS

MH and MZ conceived the idea and designed the research. ZW, NJ, ZC, and CW conducted the animal and *in vitro* experiments. ZS, WY, and MZ analyzed the results. FH performed the flow cytometry. ZW and MZ wrote the manuscript. All the authors reviewed and approved of the manuscript.

FUNDING

REFERENCES

Austin, P. J., Tsitsiou, E., Boardman, C., Jones, S. W., Lindsay, M. A., Adcock, I. M., et al. (2017). Transcriptional profiling identifies the long noncoding RNA plasmacytoma variant translocation (PVT1) as a novel regulator of the asthmatic phenotype in human airway smooth muscle. *J. Allergy Clin. Immunol.* 139, 780–789. doi: 10.1016/j.jaci.2016.06.014

Chang, P. J., Michaeloudes, C., Zhu, J., Shaikh, N., Baker, J., Chung, K. F., et al. (2015). Impaired nuclear translocation of the glucocorticoid receptor in corticosteroid-insensitive airway smooth muscle in severe asthma. *Am. J. Respir. Criti. Care Med.* 191, 54–62. doi: 10.1164/rccm.201402-0314OC

Cohn, L., Elias, J. A., and Chupp, G. L. (2004). Asthma: mechanisms of disease persistence and progression. *Annu. Rev. Immunol.* 22, 789–815. doi: 10.1146/annurev.immunol.22.012703.104716

Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. *Trends Genet.* 30, 121–123. doi: 10.1016/j.tig.2014.01.004

Gueders, M. M., Paulissen, G., Crahay, C., Quesada-Calvo, F., Hacha, J., Van Hove, C., et al. (2009). Mouse models of asthma: a comparison between C57BL/6 and BALB/c strains regarding bronchial responsiveness, inflammation, and cytokine production. *Inflamm. Res.* 58, 845–854. doi: 10.1007/s00011-009-0054-2

Holgate, S. T. (2012). Innate and adaptive immune responses in asthma. *Nat. Med.* 18, 673–683. doi: 10.1038/nm.2731

Hondowicz, B. D., An, D., Schenkel, J. M., Kim, K. S., Steach, H. R., Krishnamurty, A. T., et al. (2016). Interleukin-2-dependent allergen-specific tissue-resident memory cells drive asthma. *Immunity* 44, 155–166. doi: 10.1016/j.immuni.2015.11.004

Hubeau, C., Apostolou, I., and Kobzik, L. (2006). Adoptively transferred allergen-specific T cells cause maternal transmission of asthma risk. *Am. J. Pathol.* 168, 1931–1939. doi: 10.2353/ajpath.2006.051231

Hwang, S. S., Kim, Y. U., Lee, S., Jang, S. W., Kim, M. K., Koh, B. H., et al. (2013). Transcription factor YY1 is essential for regulation of the Th2 cytokine locus and for Th2 cell differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 276–281. doi: 10.1073/pnas.1214682110

Istomine, R., Pavey, N., and Piccirillo, C. A. (2016). Posttranscriptional and translational control of gene regulation in CD4+ T cell subsets. *J. Immunol.* 196, 533–540. doi: 10.4049/jimmunol.1501337

Kaiko, G. E., and Foster, P. S. (2011). New insights into the generation of Th2 immunity and potential therapeutic targets for the treatment of asthma. *Curr. Opin. Allergy Clin. Immunol.* 11, 39–45. doi: 10.1097/ACI.0b013e328342322f

Kerzerho, J., Maazi, H., Speak, A. O., Szely, N., Lombardi, V., Khoo, B., et al. (2013). Programmed cell death ligand 2 regulates TH9 differentiation and induction of chronic airway hyperreactivity. *J. Allergy Clin. Immunol.* 131, 1048–1057, 1057.e1-2. doi: 10.1016/j.jaci.2012.09.027

Lajoie, S., Lewkowich, I., Herman, N. S., Sproles, A., Pesce, J. T., Wynn, T. A., et al. (2014). IL-21 receptor signalling partially mediates Th2-mediated allergic airway responses. *Clin. Exp. Allergy* 44, 976–985. doi: 10.1111/cea.12341

Lee, G. R. (2014). Transcriptional regulation of T helper type 2 differentiation. *Immunology* 141, 498–505. doi: 10.1111/imm.12216

Li, J., Ma, W., Zeng, P., Wang, J., Geng, B., Yang, J., et al. (2015). LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief. Bioinform.* 16, 806–812. doi: 10.1093/bib/bbu048

Liu, M. J., Takahashi, Y., Wada, T., He, J., Gao, J., Tian, Y., et al. (2009). The aldo-keto reductase Akr1b7 gene is a common transcriptional. *Mol. Pharmacol.* 76, 604–611. doi: 10.1124/mol.109.057455

Liu, Y., Chen, Z., Xu, K., Wang, Z., Wu, C., Sun, Z., et al. (2018). Next generation sequencing for miRNA profile of spleen CD4(+) T cells in the murine model of acute asthma. *Epigenomics* 10, 1071–1083. doi: 10.2217/epi-2018-0043

Ma, L., Bajic, V. B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biol.* 10, 925–933. doi: 10.4161/rna.24604

Mahmood, T., Tahir, T., Munir, F., and Shinwari, Z. K. (2018). Characterization of regulatory elements in OsRGLP2 gene promoter from different rice accessions through sequencing and in silico evaluation. *Comput. Biol. Chem.* 73, 206–212. doi: 10.1016/j.compbiolchem.2018.02.015

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y. S., Tam, W. L., Thomson, A. M., et al. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126, 1203–1217. doi: 10.1016/j.cell.2006.07.031

Ochkur, S. I., Doyle, A. D., Jacobsen, E. A., LeSuer, W. E., Li, W., Protheroe, C. A., et al. (2017). Frontline science: eosinophil-deficient MBP-1 and EPX double-knockout mice link pulmonary remodeling and airway dysfunction with type 2 inflammation. *J. Leukoc. Biol.* 102, 589–599. doi: 10.1189/jlb.3HI1116-488RR

O'Shea, J. J., and Paul, W. E. (2010). Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. *Science* 327, 1098–1102. doi: 10.1126/science.1178334

Pua, H. H., Steiner, D. F., Patel, S., Gonzalez, J. R., Ortiz-Carpena, J. F., Kageyama, R., et al. (2016). MicroRNAs 24 and 27 suppress allergic inflammation and target a network of regulators of T helper 2 cell-associated cytokine production. *Immunity* 44, 821–832. doi: 10.1016/j.immuni.2016.01.003

Sczyrba, A., Kruger, J., Mersch, H., Kurtz, S., and Giegerich, R. (2003). RNA-related tools on the bielefeld bioinformatics server. *Nucleic Acids Res.* 31, 3767–3770. doi: 10.1093/nar/gkg576

Sundrud, M. S., Vancompernolle, S. E., Eger, K. A., Bruno, T. C., Subramaniam, A., Mummidi, S., et al. (2005). Transcription factor GATA-1 potently represses the expression of the HIV-1 coreceptor CCR5 in human T cells and dendritic cells. *Blood* 106, 3440–3448. doi: 10.1182/blood-2005-03-0857

Takyar, S., Vasavada, H., Zhang, J. G., Ahangari, F., Niu, N., Liu, Q., et al. (2013). VEGF controls lung Th2 inflammation via the miR-1-Mpl (myeloproliferative leukemia virus oncogene)-P-selectin axis. *J. Exp. Med.* 210, 1993–2010. doi: 10.1084/jem.20121200

Tsitsiou, E., Williams, A. E., Moschos, S. A., Patel, K., Rossios, C., Jiang, X., et al. (2012). Transcriptome analysis shows activation of circulating CD8+ T cells in patients with severe asthma. *J. Allergy Clin. Immunol.* 129, 95–103. doi: 10.1016/j.jaci.2011.08.011

Tussiwand, R., Everts, B., Grajales-Reyes, G. E., Kretzer, N. M., Iwata, A., Bagaitkar, J., et al. (2015). Klf4 expression in conventional dendritic cells is required for T helper 2 cell responses. *Immunity* 42, 916–928. doi: 10.1016/j.immuni.2015.04.017

Weng, C. M., Wang, C. H., Lee, M. J., He, J. R., Huang, H. Y., Chao, M. W., et al. (2018). Aryl hydrocarbon receptor activation by diesel exhaust particles mediates epithelium-derived cytokines expression in severe allergic asthma. *Allergy* 73, 2192–2204. doi: 10.1111/all.13462

Yu, X., Zhe, Z., Tang, B., Li, S., Tang, L., Wu, Y., et al. (2017). α-Asarone suppresses the proliferation and migration of ASMCs through targeting the lncRNA-PVT1/miR-203a/E2F3 signal pathway in RSV-infected rats. *Acta Biochim. Biophys. Sin.* 49, 598–608. doi: 10.1093/abbs/gmx048

Zhang, F., Liu, G., Wei, C., Gao, C., and Hao, J. (2017). Linc-MAF-4 regulates Th1/Th2 differentiation and is associated with the pathogenesis of multiple sclerosis by targeting MAF. *FASEB J.* 31, 519–525. doi: 10.1096/fj.201600838R

Zhang, X.-Y., Tang, X.-Y., Ma, L.-J., Guo, Y.-L., Li, X.-S., Zhao, L.-M., et al. (2017). Schisandrin B down-regulated lncRNA BCYRN1 expression of airway smooth muscle cells by improving miR-150 expression to inhibit the proliferation and migration of ASMC in asthmatic rats. *Cell Proliferation* 50:e12382. doi: 10.1111/cpr.12382

Zhang, X. Y., Zhang, L. X., Tian, C. J., Tang, X. Y., Zhao, L. M., Guo, Y. L., et al. (2016). LncRNAs BCYRN1 promoted the proliferation and

migration of rat airway smooth muscle cells in asthma via upregulating the expression of transient receptor potential 1. *Am. J. Transl. Res.* 8, 3409–3418.

Zhu, Y. J., Mao, D., Gao, W., and Hu, H. (2018). Peripheral whole blood lncRNA expression analysis in patients with eosinophilic asthma. *Medicine* 97:e9817. doi: 10.1097/MD.0000000000009817

Conflict of Interest Statement: 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.

Identification of Cancer-Related Long Non-Coding RNAs Using XGBoost With High Accuracy

Xuan Zhang[1,2†], Tianjun Li[3†], Jun Wang[4†], Jing Li[1], Long Chen[3] and Changning Liu[1*]*

[1] *CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, China,* [2] *University of Chinese Academy of Sciences, Beijing, China,* [3] *Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China,* [4] *Institute of Medical Sciences, Xiangya Hospital, Central South University, Changsha, China*

In the past decade, hundreds of long noncoding RNAs (lncRNAs) have been identified as significant players in diverse types of cancer; however, the functions and mechanisms of most lncRNAs in cancer remain unclear. Several computational methods have been developed to detect associations between cancer and lncRNAs, yet those approaches have limitations in both sensitivity and specificity. With the goal of improving the prediction accuracy for associations of lncRNA with cancer, we upgraded our previously developed cancer-related lncRNA classifier, CRlncRC, to generate CRlncRC2. CRlncRC2 is an eXtreme Gradient Boosting (XGBoost) machine learning framework, including Synthetic Minority Over-sampling Technique (SMOTE)-based over-sampling, along with Laplacian Score-based feature selection. Ten-fold cross-validation showed that the AUC value of CRlncRC2 for identification of cancer-related lncRNAs is much higher than previously reported by CRlncRC and others. Compared with CRlncRC, the number of features used by CRlncRC2 dropped from 85 to 51. Finally, we identified 439 cancer-related lncRNA candidates using CRlncRC2. To evaluate the accuracy of the predictions, we first consulted the cancer-related long non-coding RNA database Lnc2Cancer v2.0 and relevant literature for supporting information, then conducted statistical analysis of somatic mutations, distance from cancer genes, and differential expression in tumor tissues, using various data sets. The results showed that our approach was highly reliable for identifying cancer-related lncRNA candidates. Notably, the highest ranked candidate, lncRNA AC074117.1, has not been reported previously; however, integrated multi-omics analyses demonstrate that it is the target of multiple cancer-related miRNAs and interacts with adjacent protein-coding genes, suggesting that it may act as a cancer-related competing endogenous RNA, which warrants further investigation. In conclusion, CRlncRC2 is an effective and accurate method for identification of cancer-related lncRNAs, and has potential to contribute to the functional annotation of lncRNAs and guide cancer therapy.

Keywords: cancer, long noncoding RNA, machine learning, Synthetic Minority Over-sampling Technique, XGBoost

INTRODUCTION

Cancer is a leading cause of death worldwide (Siegel et al., 2018) and it is established that cancers are caused by genetic and epigenetic changes (Kanwal and Gupta, 2010; You and Jones, 2012). Hence, high throughput technologies to characterize genes associated with cancer have applications with crucial implications for human health. Long non-coding RNAs (lncRNAs) account for the vast majority of non-coding RNAs longer than 200 nucleotides, and were previously considered "junk" RNA, due to their low coding potential; however, over recent decades, lncRNAs have been recognized as significant regulators of multiple major biological processes impacting development, differentiation, and metabolism (Bhan and Mandal, 2015). In cancer, lncRNAs act via multiple mechanisms, including regulation of chromatin topology in both cis and trans (chromatin remodeling, chromatin interactions), scaffolding of proteins and other RNAs, acting as protein and RNA decoys (competing endogenous RNA, ceRNA), regulating neighboring genes as natural antisense transcripts (NATs), and producing micropeptides (Aab et al., 2016; Ransohoff et al., 2018).

The aberrant expression of lncRNAs has been linked to typical cancer hallmarks, such as continuous proliferation, bypassing apoptosis, genomic instability, drug resistance, invasion, and metastasis (Renganathan and Felley-Bosco, 2017; Bhan et al., 2017; Balas and Johnson, 2018; Wang et al., 2019). For example, the lncRNA growth arrest-specific transcript 5 (*GAS5*), which is down-regulated in almost all tumor tissues, can suppress the tumorigenesis of cervical cancer by downregulating miR-196a and miR-205 (Yang et al., 2017), while *LncRNA-PVT1*, which is up-regulated in non-small cell lung cancer (NSCLC), can improve tumor invasion and metastasis (Yang et al., 2014). Further, Hox transcript antisense intergenic RNA (*HOTAIR*), which contributes to epigenetic regulation of genes, plays an important role in various cellular pathways by interacting with Polycomb Repressive Complex 2 (PRC2) (Mercer and Mattick, 2013). In addition, due to dynamic changes in their expression levels as cancer develops, some lncRNAs are regarded as potential biomarkers and therapeutic targets (Hanahan and Weinberg, 2011; Bhan et al., 2017). The most prominent example of such a biomarker is prostate cancer antigen 3 (*PCA3*), a lncRNA expressed at high levels in prostate cancer (De Kok et al., 2002; Yarmishyn and Kurochkin, 2015). The detection of *PCA3* in urine is a more specific marker for prostate cancer diagnosis than the commonly used factor, prostate specific antigen (PSA), and has been widely applied in the clinic (Hessels et al., 2003; Tinzl et al., 2004). Another example is lncRNA *TUC339*, which is highly enriched in extracellular vesicles secreted by hepatocellular carcinoma cells, where it regulates the growth and adhesion of tumor cells (Kogure et al., 2013). These features of lncRNA prompted us to search for efficient methods to predict functional lncRNAs in cancer, to facilitate deeper understanding

of malignancies and the potential application of lncRNAs as targets for cancer therapies and diagnostics.

Systematic understanding of the contributions of lncRNAs to cancer is challenging, partly due to the unpredictability of lncRNA functional elements, as well as their relatively low conservation, low expression levels, and diverse functional mechanisms. The functions of a single lncRNA, or several lncRNAs, can be determined using experimental methods; however, this approach is time consuming and costly. The successful implementation of machine learning systems for the study of genomics, proteomics, systems biology, and evolution, has been a great inspiration to the field of life sciences more generally (Larranaga et al., 2006). Using machine learning algorithms, we can determine the high dimensional characteristics of functional lncRNAs from an informatics perspective. To successfully apply machine learning to the identification of functional lncRNAs in cancer genomics, it is fundamental to first identify positive and negative sets. For this purpose, there are a number of repositories from which cancer-related lncRNAs can be conveniently obtained, including Lnc2Cancer v2.0, a manually curated database that provides comprehensive experimentally supported associations between lncRNAs and human cancer (Gao et al., 2019), and CRlncRNA, another manually curated database that uses stricter criteria to retain only data related to cancer hallmarks that have been experimentally confirmed (Wang et al., 2018). These databases can be exploited to develop machine learning models to predict and rank cancer-related lncRNAs. There has been relatively little research that has attempted to use machine learning methods to predict functional lncRNAs in cancer. For example, Zhao et al. (2015) presented the first naïve Bayes based machine learning method, and identified 707 cancer-related lncRNA candidates. In our previous work, we used a Random Forest based algorithm, CRlncRC, to classify cancer-related lncRNAs and other lncRNAs, through integration of 85 features (Zhang et al., 2018); however, compared with the computational prediction work reported for cancer-related protein-coding genes, the identification of cancer-related lncRNAs remains preliminary. The sensitivity and specificity of methods to predict cancer-related lncRNAs require further improvement.

In this study, we developed a new cancer-related lncRNA classifier, CRlncRC2. Compared with CRlncRC, CRlncRC2 uses the Laplacian score feature selection method to reduce training time and prevent over-fitting. In addition, unlike the naïve under-sampling method adopted by CRlncRC, we address the data imbalance problem, which is caused by the relatively small size of available positive sets of cancer-related lncRNAs, using the Synthetic Minority Over-sampling Technique (SMOTE) method, to balance imbalanced data, while aiming to retain all important information. Moreover, CRlncRC2 uses a more powerful machine learning model, extreme gradient boosting machine (XGBoost), to improve its predictive performance. Ten-fold cross-validation showed that the area under the receiver operating characteristic curve (AUC or area under ROC curve) score of CRlncRC2 is much higher than those of CRlncRC (0.86 vs. 0.82) and the method developed by Zhao et al. (0.90 vs. 0.79). Finally, 439 possible cancer-related lncRNAs were identified using CRlncRC2, of which 5 in the

---

top 20 were confirmed using the Lnc2Cancer v2.0 database. Further, statistical analyses show that the identified lncRNAs are closer to cancer protein genes, carry more mutations, and are more likely to be differentially expressed in tumor tissues than negative lncRNAs. In addition, survival analysis revealed a significant difference in overall survival between the low and high expression groups of the top 10 predictions. In particular, one lncRNA, AC074117.1 (ENSG00000234072), which was the top ranked of our predictions and has not been reported in the literature, is suggested as being highly likely to be associated with cancer in the lncRNA-related ceRNA network. In conclusion, CRlncRC2 exhibited good performance in both cross-validation and prediction evaluation. We believe our framework will be a useful tool for study of lncRNA–cancer associations.

MATERIALS AND METHODS

Our experiment followed the pipeline illustrated in **Figure 1A**, which consisted of four main steps: Data preparation, Feature engineering, Model training, and Prediction and validation. The detailed processes of feature selection and cross-validation are presented in **Figures 1B**, **C**.

Data Preparation

Cancer-related lncRNAs (positive set) and cancer unrelated lncRNAs (negative set) were downloaded from CRlncRC (https://github.com/xuanblo/CRlncRC). The criteria for cancer-related lncRNA collection include either differentially expressed in cancer (as verified by Real-Time qRT-PCR), co-occurred with a significant relevant clinicopathological parameter (e.g., tumor differentiation, clinical stage, and survival time), or proven by functional experiments (e.g., colony formation assay, matrigel invasiveness assay, xenograft mouse model, and metastasis nude mouse model). As the category of cancer unrelated lncRNA is difficult to define, and for consistency with other classifiers, we located a large number of single-nucleotide polymorphisms (SNPs) associated with phenotypes derived from the NHGRI-EBI GWAS Catalog (Welter et al., 2014) in the sequences of lncRNAs, and only those lncRNAs which had no phenotype-related SNPs detected within its 10 kb up/down stream were selected as cancer non-related lncRNAs. Finally, we identified 158 positive lncRNAs (**Data Sheet 1**) and 4,533 negative lncRNAs (**Data Sheet 2**).

We downloaded lncRNA feature data from CRlncRC; CRlncRC retrieves 85 features and groups them into four categories: genomic features, expression features, epigenetic features, and network features. Feature category, name, source database, and description information are detailed in **Data Sheet 3**.

Feature Engineering

Features play an essential role in classification, and appropriate features can improve classification performance significantly. In cancer genomic research, the currently known cancer-related lncRNA (positive) set are only available because they were identified by humans. It is possible that some samples in

the negative set may be considered to belong to the positive set in the future. Hence, we employed Laplacian scoring (He et al., 2005), which is designed to select features without labels, as a criterion to evaluate the correlations of each feature. The basic idea of Laplacian score is to evaluate the features according to their locality preserving power, which is from the Laplacian Eigenmaps (Chung, 1997) and Locality Preserving Projection (He and Niyogi, 2003).

In detail, we applied the scikit-feature (Li et al., 2017) to calculate Laplacian scores; the parameters for the affinity matrix used for the calculation are as follows: metric = euclidean, neighbor mode = knn, and k = 5. Calculated scores range from 0 to 1, with smaller values indicating more important features. The distribution of calculated Laplacian scores is presented in **Figure 2** and clearly shows that there are large margins in each category of features. In this case, we can determine the difference between the sorted Laplacian scores (asc) and use the first two differential values to set a threshold. Specifically, we set the margins in "Epigenetic" to the second and third largest differential values, because these appeared to be the inflection points. Hence, the features were split into three parts, and the features located in the lower part (i.e., those with scores indicating that the features are more important) retained immediately. Nevertheless it is not advisable to simply remove those features located in the other parts, as these also contain some information. Therefore, we merged the features according to the mean in each part and retained the merged features to preserve the information. For example, the middle scoring part of "Expression" contains two features, and we removed these two features, while retaining their mean value. The mean-merged feature obtained from the high scoring parts were also retained. Finally, generated training and validation sets by concatenating the processed category features. Changes in the feature number in each category are summarized in **Table 1**. After feature selection, we obtained 51 features, eight of which are synthetic. A "Bigtable", containing 11194 lncRNAs from CRlncRC, with 85 features, is included in **Data Sheet 4**.

Model Training

The machine learning method, XGBoost (Chen and Guestrin, 2016), was tuned to search for an optimal prediction solution. XGBoost is a type of gradient boosting decision tree method; its objective function is defined in equation (1).

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} loss(y_i \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \qquad \text{(equation 1)}$$

where loss is the training loss, $\Omega(f)$ is the complexity of the tree, and K is the number of trees in the model. This model can be optimized by minimizing this objective function. To this end, an additive training method was employed for training loss, and prediction at the additive t^{th} training round could be quickly optimized using Taylor expansion. The greedy algorithm [31] was used to determine optimal tree complexity.

In our study, we used the dmlc XGBoost library (https://xgboost.ai/) for implementation of the XGBoost model. To tune the hyper-parameters, we first adopted Bayesian optimization to

FIGURE 1 | Pipeline for the experiment. **(A)** Designment of experimental workflow. **(B)** Details of feature selection. **(C)** Details of 10-fold cross validation with over-sampling.

FIGURE 2 | Laplacian score distribution. Right, sorted scores (asc). Red dotted line and dashed line, assumed thresholds.
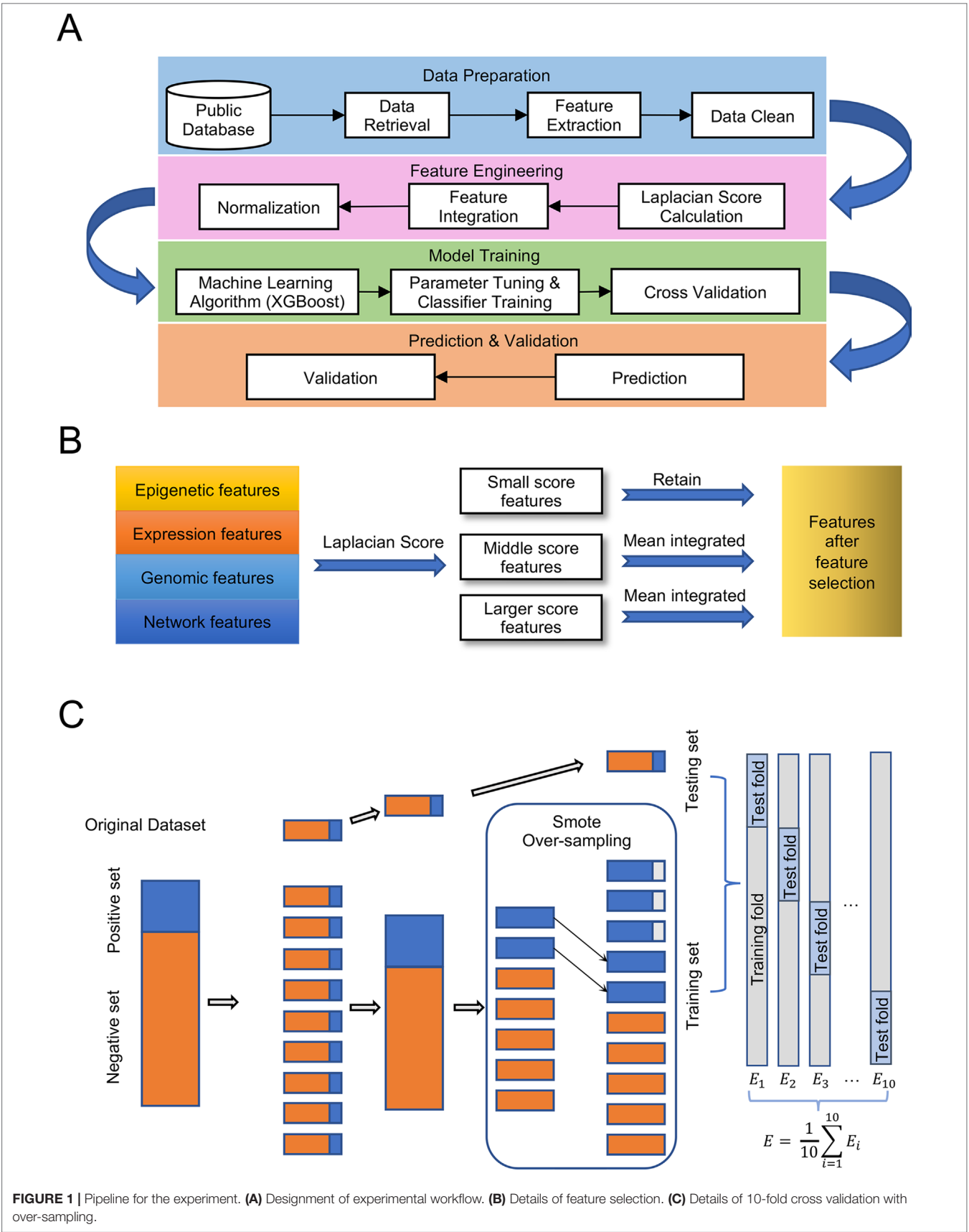
TABLE 1 | Changes in feature number for each type of feature.

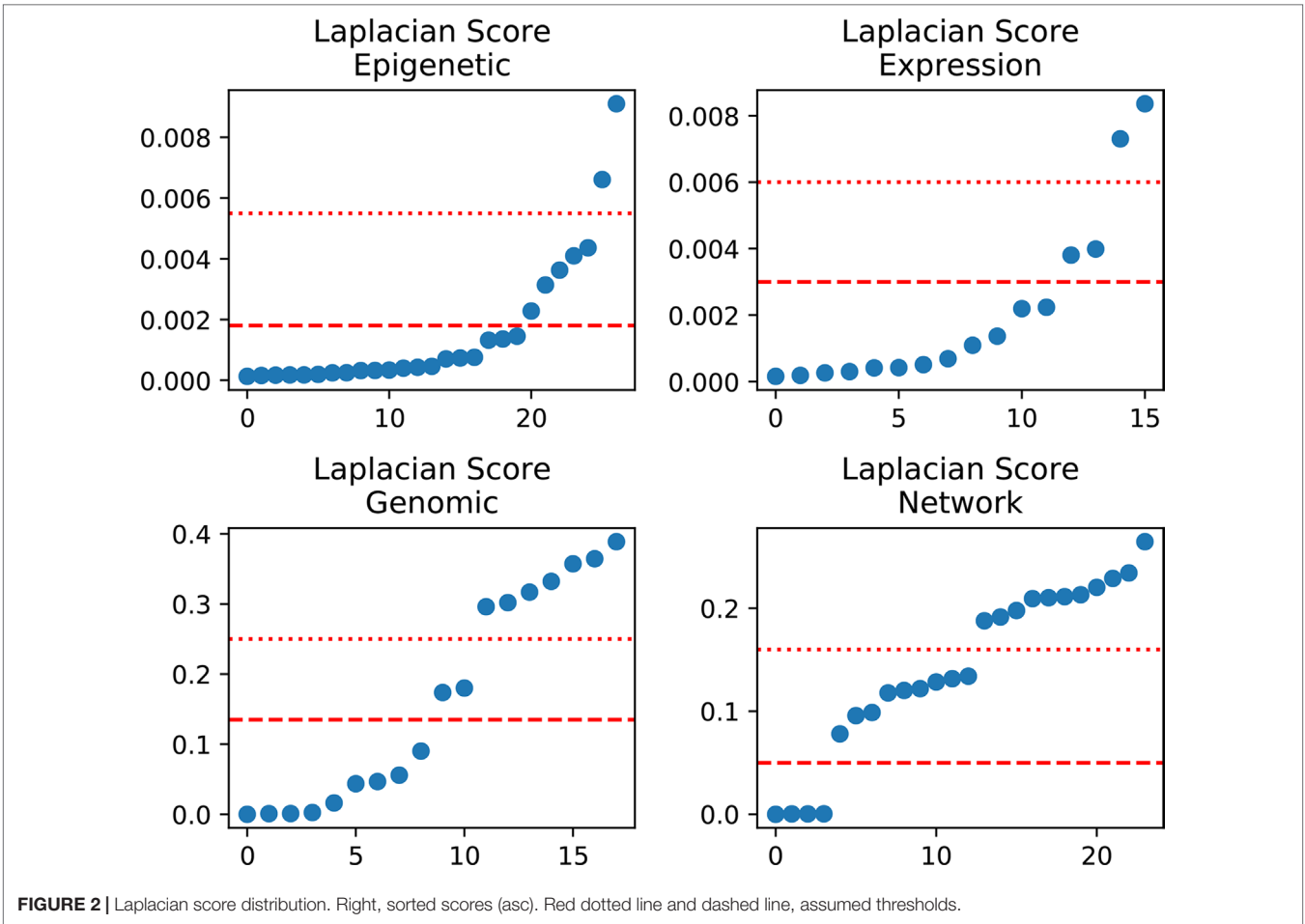| Epigenetic			Expression			Genomic			Network		
Feature number before feature selection											
LP*	MP*	UP*	LP*	MP*	UP*	LP*	MP*	UP*	LP*	MP*	UP*
20	5	2	12	2	2	9	2	7	4	9	11
Feature number after feature selection											
22			14			11			6		

*LP, lower part of Laplacian Score; MP, middle part of Laplacian Score; UP, higher part of Laplacian Score.

search for potential hyper-parameters and then manually fine-tuned those hyper-parameters to improve the performance of the model. The hyper-parameters for XGBoost primarily control the growth and the robustness of the model:

- Growth: n estimators, max depth, and learning rate
- Robustness: colsample bytree, subsample, and gamma

In addition, as our sample was unbalanced (the ratio of the minority positive class versus majority negative class was approximately 1/30), we adopted SMOTE (Nakamura et al., 2013) to re-sample our training set by Bayesian optimization, which reduces the impact of data imbalance. The final tuning result for

this model is n estimator = 546, max depth = 10, learning rate = 0.01, colsample bytree = 0.7, subsample = 0.826, and gamma = 0.036.

Ten-fold cross validation was adopted to evaluate the model trained by parameters obtained using Bayesian optimization. The algorithm stratified shuffles the total samples into 10 folds, and begins an iteration: each time 9 folds are initially over-sampled, and then assigned for training. The single remaining fold is adopted as the pair for validation. Subsequently, the over-sampled training set was used to fit the model, while the validation set was utilized to evaluate the model's performance. Note that the validation set in each iteration is not re-sampled and does not include any data used for training. Further, the models trained

on each iteration are independent of one another. To rigorously evaluate the performance of our model, we measured the AUC scores using the abovementioned 10-fold cross-validation (**Figure 1C**).

Further, to rigorously evaluate the model's performance, we measured the recall, precision, and F1 score, using the 10-fold cross-validation process described above.

The recall is the ratio of correctly predicted positive observations to all observations in a specific class, and was calculated using equation 3:

$$Recall = \frac{TP + FN}{TP},$$ (equation 2)

The precision is the ratio of correctly predicted positive observations to total predicted positive observations, and was calculated using equation 4:

$$Precision = \frac{TP + FP}{TP},$$ (equation 3)

The F1 Score is the weighted average of Precision and Recall, and was calculated using equation 5:

$$F1\ score = 2 * \frac{(Recall * Precision)}{(Recall + Precision)},$$ (equation 4)

Prediction and Evaluation

To predict novel cancer-related lncRNAs, we used our pre-trained model to predict 7,253 unknown lncRNAs from TANRIC [33]. To evaluate the accuracy of our model, we used various methods to test the reliability of our predictions. First, predictions were searched against the Lnc2Cancer v2.0 database. Next, the Kolmogorov-Smirnov test was used to examine whether there were significant differences among the different sets (positive, negative, and predictive) in the distance to cancer protein-coding genes, mutation numbers, and numbers of samples differentially expressed between tumor and normal tissues. Mutation data and cancer protein-coding gene sets were download from COSMIC [34]. Tumor and normal tissue expression profiles were downloaded from TANRIC. Further, survival analysis for the top 10 predictions was conducted using TANRIC.

RESULTS

Data Collection

We collected 158 highly trusted cancer-related lncRNAs from CRlncRC as our positive data set. All have been reported in the literature with the support of strict experimental validation and are involved with cancer hallmarks. lncRNAs (n = 4,553) in CRlncRC without phenotype-related SNPs within 10 kb up- or down-stream were used as our negative data set. In CRlncRC, we collected 85 features that could potentially facilitate the

recognition of cancer-related lncRNAs and grouped them into four different categories (see **Data Sheet 3** for details): Genomic features (such as GC content and sequence conservation score), Expression features (the expression profiles of lncRNAs in 16 different tissue types), Epigenetic features (different types of epigenetic signals in different types of cell lines), and Network features (the interactions between lncRNAs and cancer-related protein-coding genes and miRNAs). After feature selection using Laplacian scores, we reduced the feature number from 85 to 51. Cumulative curves were plotted and showed that the distribution of the feature values between the positive and negative sets was significantly different (Kolmogorov-Smirnov test, p-value < 0.05) (**Data Sheet 5**). The number in each feature category before and after feature selection is shown in **Table 1**.

Performance Evaluation

The results of 10-fold cross-validation are presented in **Figure 3**. We drew 10 ROC curves, which had minimum and maximum AUC values of 0.73 and 0.93, respectively, and an average value of 0.86 ± 0.6. In addition to AUC values, additional evaluation indicators were used to assess our results, including precision, recall, and F1-Score (**Table 2**). The average precision, recall, and F1-Score values were 0.72, 0.62, and 0.65, respectively. Overall, these data demonstrate that CRlncRC2 is an efficient tool for identification of lncRNAs related with cancer, with high accuracy and stable performance.

Compared with other methods, CRlncRC2 has superior performance. Relative to CRlncRC, CRlncRC2 reduced features number from 85 to 51 and the mean AUC value reached 0.86, which is 0.04 higher than that achieved using CRlncRC (**Figure 4A**). Further, we compared the prediction performance of CRlncRC2 with that of the method described by Zhao et al. (2015). To ensure a fair comparison, we retrained our CRlncRC2 method using the dataset reported by Zhao et al. Compared with the method of Zhao et al., the resulting mean AUC value for CRlncRC2 was much higher (0.90 vs. 0.79) (**Figure 4B**).

To determine why CRlncRC2 performed better than CRlncRC, we analyzed the feature importance (XGBoost importance weight) in CRlncRC2 (**Data Sheet 6**). Compared with the features used in CRlncRC, it is clear that the epigenetic and expression feature numbers in CRlncRC2 were almost unchanged, while those of genomic features were reduced by half, while network features were decreased by two thirds (**Figure 5A**). Expression features were two among the top ten most important features in CRlncRC2, while they were not among the top ten in CRlncRC (**Figure 5B**). In addition, there are four types of features in the top 20 features of CRlncRC2, indicating that CRlncRC2 can make better use of different features (**Figure 5C**). Furthermore, as illustrated in **Figures 5C**, **D**, the proportions of epigenetic features among the first 20 and 50 features for CRlncRC2 were much larger than those for CRlncRC. Surprisingly, although genomic and network features accounted for a small proportion, the three synthetic genomic and network features (Gen_LevelTwo, Gen_LevelOne, and Net_LevelTwo) ranked the highest, indicating that synthetic features generated by combining low Laplacian score features may contribute substantially to the model (**Figure 5E**). Two
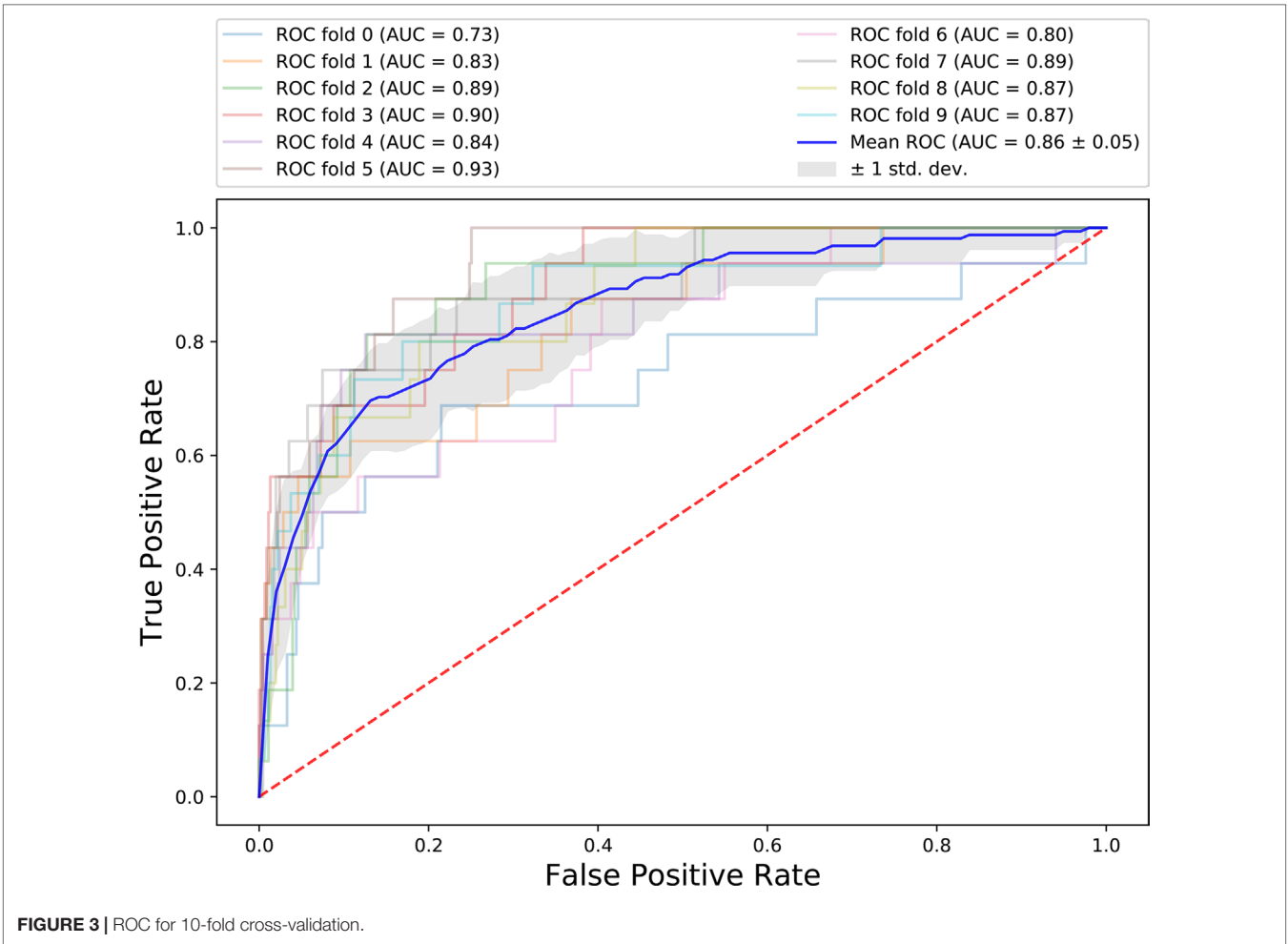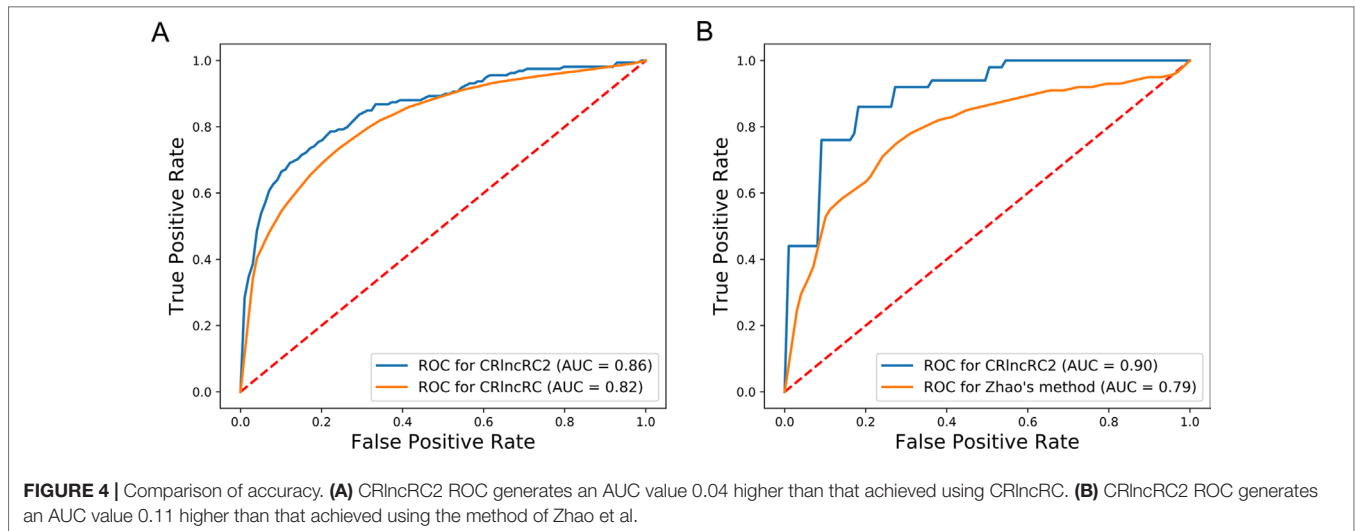
FIGURE 3 | ROC for 10-fold cross-validation.

TABLE 2 | Performance of 10-fold cross-validation.

Ten-fold cross-validation	Precision	Recall	F1-Score
macro avg fold 0	0.74	0.56	0.59
macro avg fold 1	0.90	0.66	0.72
macro avg fold 2	0.57	0.55	0.56
macro avg fold 3	0.77	0.62	0.67
macro avg fold 4	0.71	0.62	0.65
macro avg fold 5	0.78	0.71	0.74
macro avg fold 6	0.67	0.65	0.66
macro avg fold 7	0.74	0.65	0.68
macro avg fold 8	0.65	0.59	0.62
macro avg fold 9	0.70	0.60	0.63

repeat features, short interspersed nuclear element (SINE) and long interspersed nuclear element (LINE) signals on gene bodies, ranked No. 4 and No. 5, followed by gene expression level in colon tissue (No. 6), prostate gland (No. 8), "H3k4me1" epigenetic modification signals within the Transcription Start Site upstream and downstream 5k (TSS5k) region in GM12878 (No. 7), and "H3k4me3" epigenetic modification signals within lncRNA gene body/TSS1k region in H1hesc/GM12878 cell line (No. 9 and No. 10).

We further evaluated the effectiveness of our approach to dealing with the available imbalanced data. The SMOTE over-sampling method was used to balance the imbalanced data, and it contributed to an increase of 0.01 in the AUC value, relative to non-SMOTE adjusted data (**Figure 6A**). In addition, to compare the performance of different machine learning algorithms, several models were compared using the non-SMOTE adjusted over-sampling data. ROC curve analysis showed that the XGBoost-based method performed better than Decision tree (DT) (0.85 vs. 0.60) and Support Vector Machine (SVM) (0.85 vs. 0.74) -based approaches (**Figure 6B**). These results indicate that our new method facilitated superior performance relative to previous methods. XGBoost contributed substantially to the AUC values, while data over-sampling was also very important.

Statistical Analysis of Candidate Cancer-Related lncRNA Candidates

We used the pre-trained model to predict novel candidate cancer-related lncRNAs from 7,253 unknown lncRNAs, which were not in our training or testing sets. Finally, we predicted 439 cancer-related lncRNA candidates (**Data Sheet 7**). First, we used the data from the newly updated database, Lnc2Cancer v2.0, to test our

FIGURE 4 | Comparison of accuracy. **(A)** CRlncRC2 ROC generates an AUC value 0.04 higher than that achieved using CRlncRC. **(B)** CRlncRC2 ROC generates an AUC value 0.11 higher than that achieved using the method of Zhao et al.

predictions, since we did not collect our positive dataset from this database. We studied the intersection of our predictions and their collections. Among our top 10, 20, and 50 predictions, 2, 5, 8 lncRNAs, respectively, were also collected by Lnc2Cancer, and were functionally validated as cancer-related (**Figure 7A**). In total, 47 candidate cancer-related lncRNAs were found in Lnc2Cancer (**Data Sheet 7**). According to the tag information provided in the Lnc2Cancer database, these lncRNAs can be classified into several categories: drug-resistant, methylation, circulating, transcription factor (TF), and variant (**Data Sheet 8, Figure A**). Further, we selected the top 10 among these 47 cancer-related lncRNAs and evaluated their expression in cancers. Surprisingly, almost all lncRNAs exhibited inconsistent changes in expression in various tissues (**Data Sheet 8, Figure B**), confirming their functional diversity and reflecting the strong tissue specificity of lncRNAs. In addition, the 47 predicted lncRNAs had roles in numerous malignant tumors, including 17 involved in colorectal cancer, 10 in gastric cancer, and 10 in hepatocellular carcinoma (**Data Sheet 8, Figure C**).

Using statistical methods and multigroup data, we further analyzed the reliability of our predictions. First, we hypothesized that the potential cancer-related lncRNAs were likely to have more somatic mutations in cancer genomes, since many previous studies have demonstrated that mutations in functional genes are a primary cause of carcinogenesis. To validate this assumption, we compared the number of somatic mutations (documented in COSMIC) between different lncRNA sets and a cancer-related protein-coding gene set (**Figure 7B**). The results showed that the cancer-related protein-coding gene set, as the positive control, contained far more somatic mutations than the cancer-unrelated lncRNA set (negative control, Kolmogorov-Smirnov test, p-value = 6.10e-33). The somatic mutation numbers in both the positive and predicted cancer-related lncRNA sets were between those of cancer-unrelated lncRNAs and cancer-related protein-coding genes, with a significantly higher quantity than those in cancer-unrelated lncRNAs (Kolmogorov-Smirnov test, p-value 2.35e-07 and 8.27e-06, respectively).

As a number of lncRNAs exert their function in cis, by influencing neighboring genes, we assumed that these potential cancer-related lncRNAs were likely closer to cancer-related protein-coding genes than cancer-unrelated lncRNAs. Therefore, we calculated the distances of different lncRNA sets to their closest cancer-related proteins, and compared them with the random background (i.e., the distance between cancer-related protein-coding genes and random positions in genome) (**Figure 7C**). We found that the distances between cancer-unrelated lncRNAs and cancer-related protein-coding genes were significantly larger than those between cancer-related lncRNAs and cancer-related protein-coding genes (Kolmogorov-Smirnov test, p-value = 4.1e-4). Similarly, the distance of predicted cancer-related lncRNAs from cancer-related protein-coding genes was far shorter than that of cancer-unrelated lncRNAs (Kolmogorov-Smirnov test, p-value = 4.9e-06). Moreover, no significant difference in distance was detected between background and the cancer-unrelated lncRNA set, as expected.

Next, we examined whether the expression levels of cancer-related lncRNAs differed from those of cancer-unrelated lncRNAs in cancer samples (**Figure 7D**). Using lncRNA expression data from the TANRIC database, we calculated the percentage of lncRNAs that were differentially expressed (absolute $\log_2$-fold change > 1) between cancer and paracancerous tissue sample pairs, to determine whether this differed among the lncRNA sets. We found that lncRNAs in the positive set had the highest percentage of differentially expressed genes (approximately 40%), while the value for the negative set was only approximately 20%. Among predicted cancer-related lncRNAs, > 35% of them showed differential expression. These results further support the association of our prediction products with cancer, and also reveal that simple dependence on differential expression to identify cancer-related lncRNAs is far from sufficient.

Case Study
Although functional identification of lncRNAs is very challenging, using bioinformatics analysis, database searches, and literature
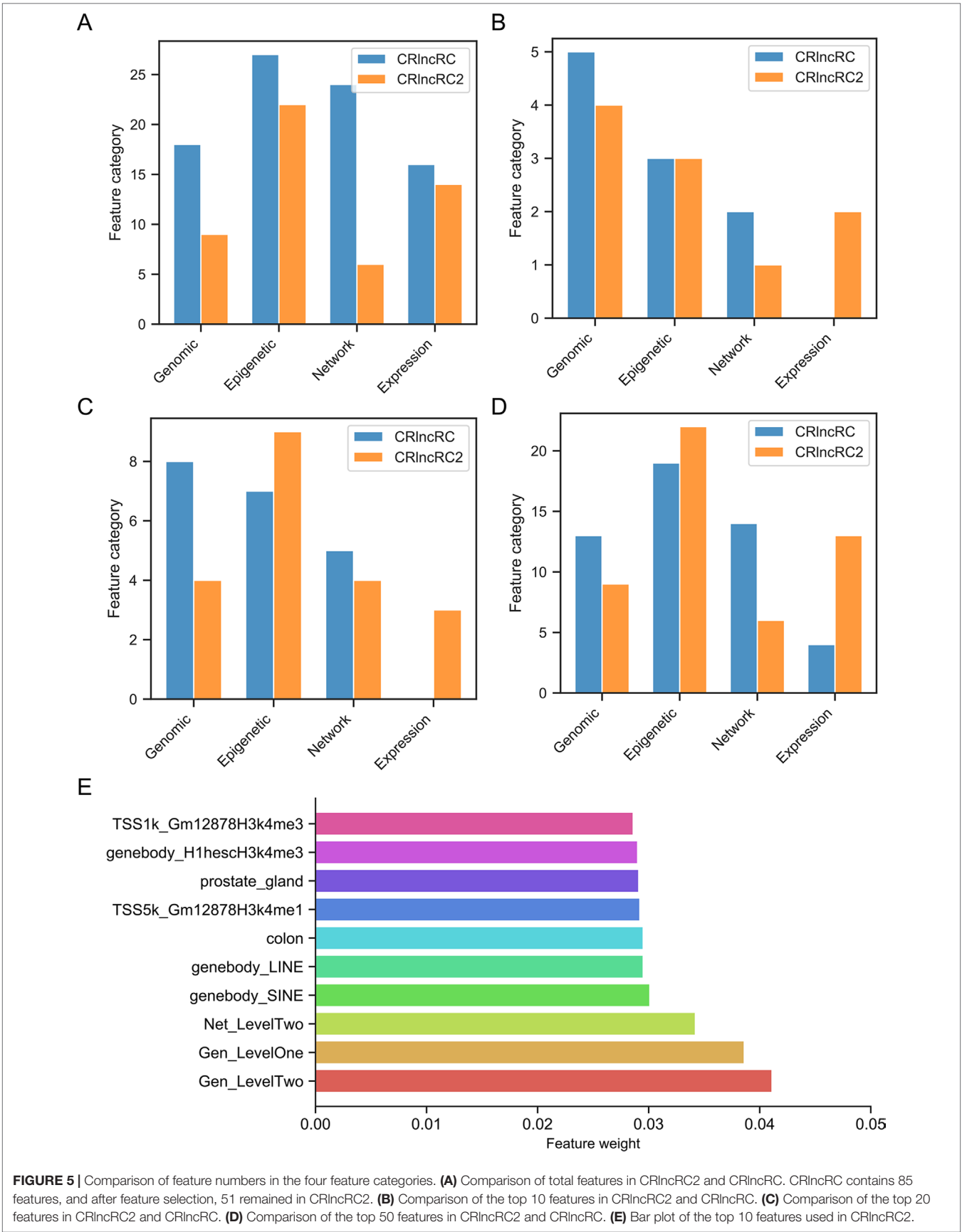
FIGURE 5 | Comparison of feature numbers in the four feature categories. **(A)** Comparison of total features in CRlncRC2 and CRlncRC. CRlncRC contains 85 features, and after feature selection, 51 remained in CRlncRC2. **(B)** Comparison of the top 10 features in CRlncRC2 and CRlncRC. **(C)** Comparison of the top 20 features in CRlncRC2 and CRlncRC. **(D)** Comparison of the top 50 features in CRlncRC2 and CRlncRC. **(E)** Bar plot of the top 10 features used in CRlncRC2.
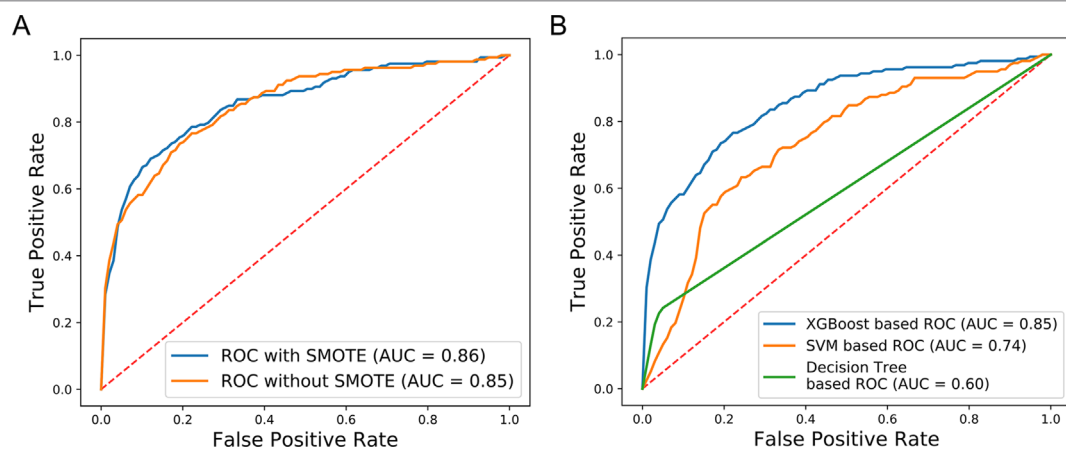
FIGURE 6 | Comparison of SMOTE and non-SMOTE, non-SMOTE XGBoost, and others. **(A)** The ROC generated using SMOTE XGBoost has AUC value 0.01 higher than that achieved using non-SMOTE XGBoost. **(B)** The XGBoost-based ROC without SMOTE generated AUC values 0.11 and 0.25 higher than the SVM-based and Decision Tree-based ROC curves, respectively.
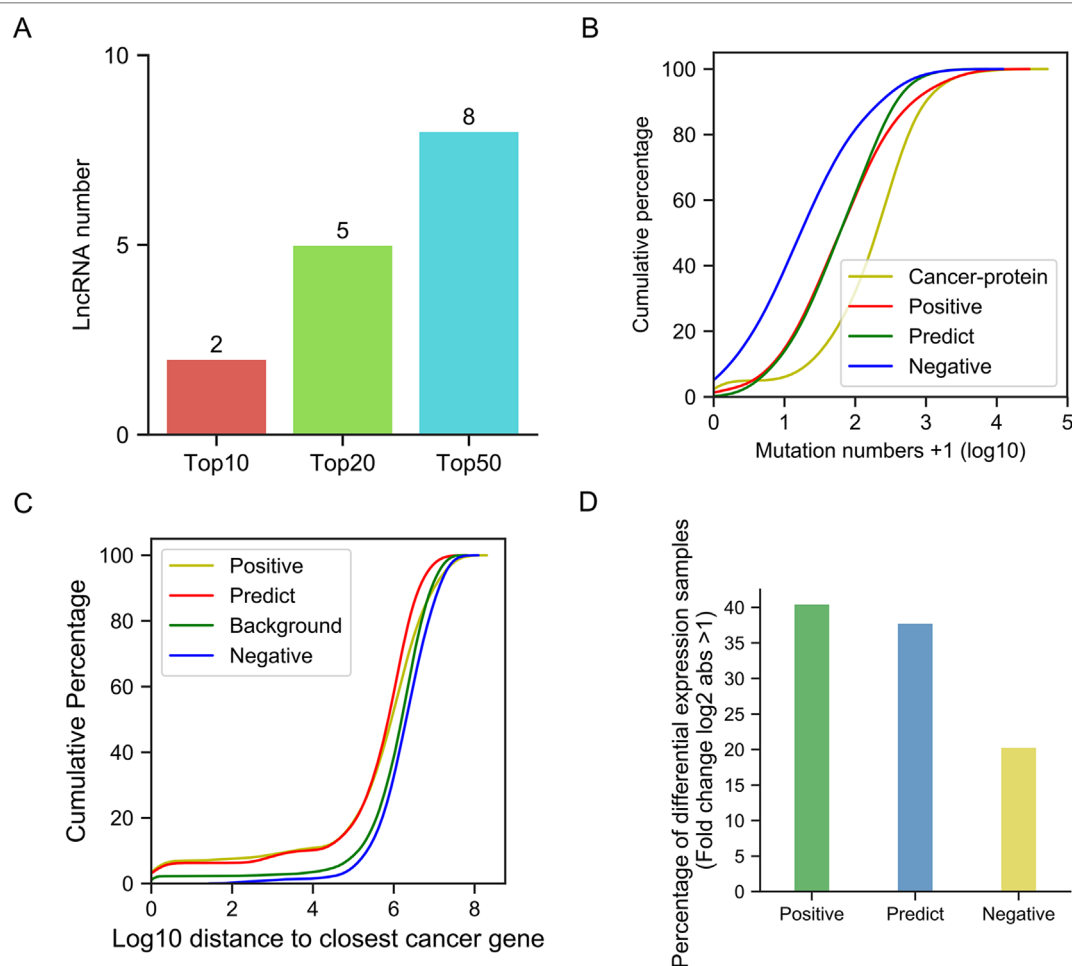
FIGURE 7 | Validation of our predictions. **(A)** Bar plots of cancer-related lncRNA numbers confirmed by Lnc2Cancer in the top 10, top 20, and top 50. **(B)** Cumulative distribution of mutation number. **(C)** Cumulative distribution of the closest distance to cancer-related proteins. **(D)** Bar plot of the percentage of differentially expressed lncRNAs.

review, we can uncover evidence that our predictions represent lncRNAs with functions in cancer. For the top 10 candidate genes we used the TANRIC database to generate Kaplan-Meier survival curves for each cancer type. The results showed that there was a significant difference in the overall survival rate between low and high lncRNA expression groups for all genes in at least one tumor tissues (**Figure 8A**).

For example, survival analysis of the No. 1 lncRNA, AC074117.1, indicated significant differences in survival time between low and high expression groups in individuals with invasive breast carcinoma (BRCA) and kidney renal clear cell carcinoma (KIRC), with p-values of 1.5e-2 and 4.0e-5, respectively (**Figure 8B**, **C**). To study the regulatory function of AC074117.1, we downloaded data on cancer-related small
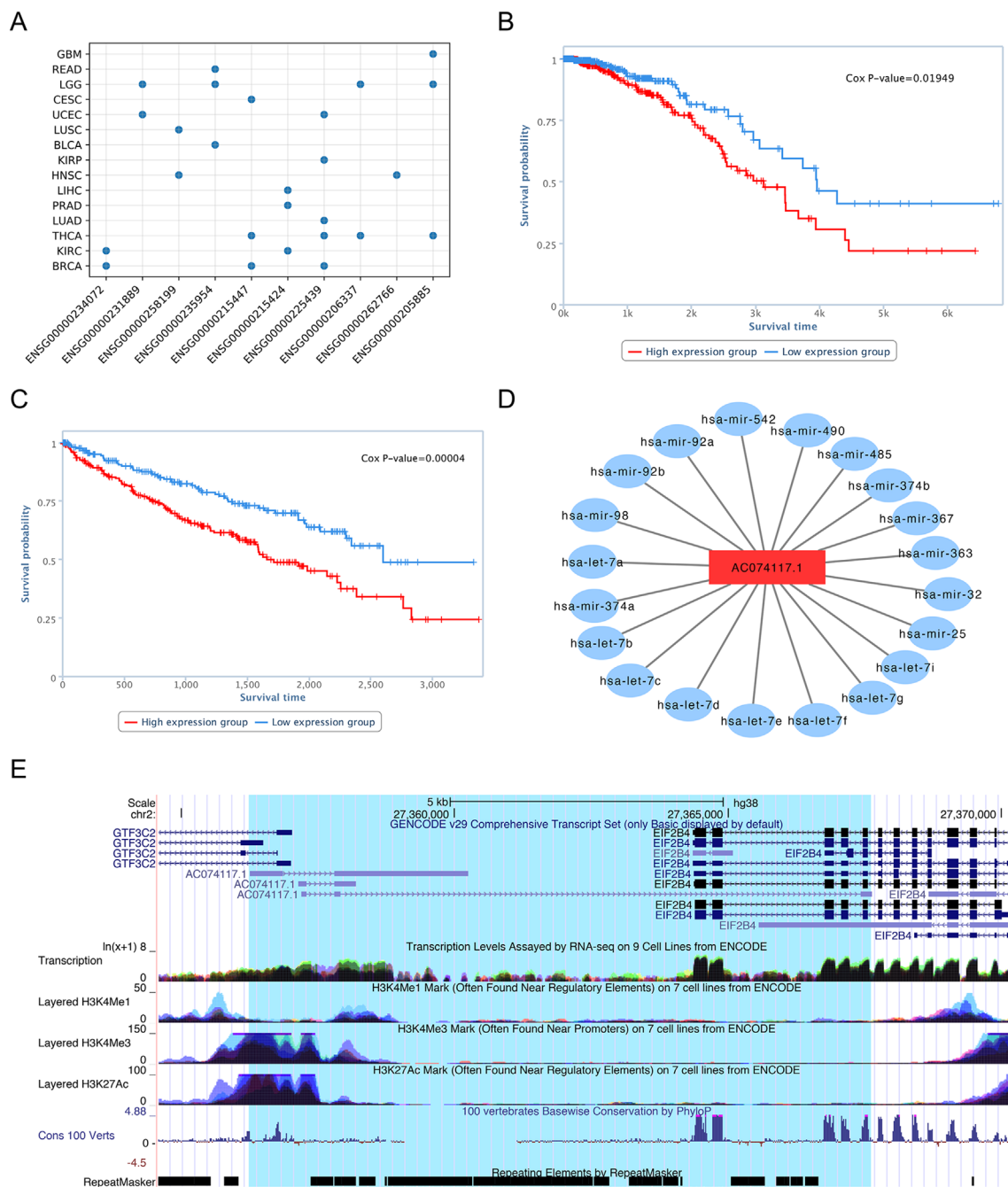

FIGURE 8 | Characterization of lncRNA AC074117.1. **(A)** Survival statistics for the top 10 lncRNA predictions. **(B)** Survival analysis of AC07117.1 in BRCA. **(C)** Survival analysis of AC07117.1 in KIRC. **(D)** Sub-network of AC07117.1 and cancer-related miRNAs. **(E)** Gene structure, epigenetic features, conservation, and repeats of AC07117.1 in the UCSC genome browser.

RNA molecules from The Human microRNA Disease Database (HMDD) (Huang et al., 2019), and the interaction network between lncRNAs and miRNAs from StarBase (Li et al., 2014). Subsequently, we constructed an interaction network between AC074117.1 and cancer-related microRNAs (**Figure 8D**). In addition, according to predictions using the LncRNA and Disease Database (version 2.0), AC074117.1 likely targets a gene cluster on chromosome 2, and is associated with a variety of cancers (Bao et al., 2019). Together, all these clues suggest that AC074117.1 may be involved in cancer and act as a ceRNA. As shown in **Figure 8E** (data from the UCSC genome browser), AC074117.1 is highly expressed in almost all tissues. Further, there is histone methylation signal in the AC074117.1 transcription start site. The H3K4Me3 and H3K27Ac signals in the first exon were high, while the H3K4Me1 signals were relatively weak. Moreover, high conservation signals (100 vertebrates basewise conservation scores generated using PlyloP) were found in its exon regions. Notably, there are a large number of repetitive elements in the whole body region of AC074117.1. These methylation signals and repeat elements may contribute to the mechanism by which this lncRNA is involved in cancer progression (Anwar et al., 2017; Di Ruocco et al., 2018; Solovyov et al., 2018). In conclusion, our predictions indicate that lncRNA AC074117.1 has a strong potential correlation with cancer.

In addition, recent literature reports support some of the predicted lncRNAs in the top 10 list; for example, the *TRAF3IP2-AS1* lncRNA ranked second (No. 2) among our predictions and is a hub gene in a lncRNA-mediated ceRNA network that competes with the onco-lncRNAs, PVT1 and XIST, and could be a clinically relevant biomarker in glioblastoma (Zan and Li, 2019). TTC28-AS1 (No. 4) is an antisense RNA of TTC28 which is associated with colorectal cancer (Pitkanen et al., 2014). Further, C1RL-AS1 (No. 10) has been linked to angiogenesis, as predicted in the ANGIOGENES database (Muller et al., 2016).

DISCUSSION

Accumulating reports demonstrate that lncRNAs have significant roles in human cancers. Using experimental methods to study the relationships between lncRNA and cancer is time consuming and costly. In contrast, computational methods enable integration of multi-omics data and provide additional information for data mining. In this study, we developed a new method, CRlncRC2, based on a powerful machine learning algorithm — XGBoost, Laplacian score feature selection, and SMOTE over-sampling, to predict associations of lncRNAs with cancer. Compared with CRlncRC, CRlncRC2 improves the performance while requires fewer features (see **Table 3** for a detailed comparison). The results

show that CRlncRC2 is much more sensitive and specific than the previous version (CRlncRC), primarily due to the selected algorithm model, as the results show huge differences between results generated using other methods and those from application of XGBoost. XGBoost has also been used in numerous other projects, achieving good results. For example, Zheng et al. developed a scalable, flexible approach, BiXGBoost, to reconstruct gene regulatory networks (GRNs), and tested it on DREAM4 and *Escherichia coli* datasets, demonstrating good performance of BiXGBoost in different scale networks (Zheng et al., 2018).

Machine learning algorithms have important roles in bioinformatics, where they facilitate the solution of problems, such as classification, clustering, regression, and prediction; however, the machine learning approach still faces a number of obstacles in predicting cancer-related lncRNAs. First, for biological data, frequently, only small positive sets are available, due to the difficulty of collecting information, such as patient data and experimental verification of functional genes, which greatly impedes the practical application of machine learning. Further, machine learning models require optimization for high performance, according to the specific data and situation. To address these problems, in this study, we selected the most stringent criteria to select the positive and negative sets, and used the latest histological information for feature extraction. We chose over-sampling in our new algorithm because it enables use of more information relative to under-sampling, and the results confirmed that it can improve accuracy and specificity. Moreover, we merged features with high Laplacian scores and got eight synthesis features, which had a highest feature importance rank. Our findings suggest that high Laplacian score features still contain useful information and is not good practice to simply discard them.

LncRNAs have been applied in clinical practice as new biomarkers and prognostic indicators. Research on the relationships between lncRNAs and cancer is attractive and progressing very rapidly. Machine learning methods have the power to discover novel lncRNAs, including disease associated lncRNAs (Kang et al., 2017; Bao et al., 2019). Efforts should continue to improve the ability of machine learning algorithms to predict cancer associations. Moreover, with increasing research into lncRNAs, greater quantities of relevant high-throughput data are becoming easier to obtain. The development of functional research into lncRNAs has revealed additional functional elements and mechanisms (Zhang et al., 2014; Brockdorff, 2018). Further, numerous new tools for evaluating the similarity of non-linear sequences, using k-mer content (Kirk et al., 2018) and a new evolutionary classification perspective (Chen et al., 2016), have been developed, which can be used to extract new features, such as lncRNA conservation. These can facilitate better application of bioinformatics methods to predict cancer-related lncRNAs and assist in cancer diagnosis and treatment.

TABLE 3 | Comparison of CRlncRC and CRlncRC2.

Method	Algorithm model	Number of Features	Feature selection	Sampling strategy	AUC
CRlncRC	Random Forest	85	No	Under-sampling	0.82
CRlncRC2	XGBoost	51	Yes	Over-sampling	0.86

CONCLUSIONS

In this study, we upgraded CRlncRC to CRlncRC2, using a powerful machine learning algorithm (XGBoost), Laplacian score feature selection, and an advanced over-sampling method (SMOTE). The results show that both XGBoost and SMOTE can help to improve model accuracy and specificity. After feature engineering, most of the expressed and methylated features are retained, indicating their importance for predicting lncRNAs with potential functions in cancer. Using much fewer features, CRlncRC2 has a mean AUC value 0.04 higher than that of CRlncRC. In addition, our predicted top-ranking cancer-related lncRNA candidates are supported by Inc2Cancer v2.0, literature reports, and statistical data. In summary, CRlncRC2 is an effective and useful method for lncRNA-cancer association identification.

DATA AVAILABILITY

The datasets analyzed for this study can be found at https://github.com/xuanblo/CRlncRC2.

AUTHOR CONTRIBUTIONS

CL and LC conceived, designed, and supervised this study. JW and XZ collected and compiled data from the literature and public databases. XZ, TL, and JW designed and developed the data analysis. JL participated in discussion of the project. XZ, JW, TJ, and CL compiled the manuscript draft. CL, LC, and JL revised the manuscript. All authors reviewed, edited, and approved the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00735/full#supplementary-material.

DATA SHEET 1 | Positive dataset. XLSX 11KB

DATA SHEET 2 | Negative dataset. XLSX 66KB

DATA SHEET 3 | Feature categories. DOCX 16KB

DATA SHEET 4 | Bigtable. CSV 6.9M

DATA SHEET 5 | Cumulative curves of positive and negative lncRNAs for all features. PDF 538KB

DATA SHEET 6 | Feature importance. CSV 1KB

DATA SHEET 7 | Predicted positives. XLSX 27KB

DATA SHEET 8 | Statistics of the interactions with Lnc2Cancer v2.0. PDF 207KB

REFERENCES

Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albuquerque, I. F., et al. (2016). Measurement of the radiation energy in the radio signal of extensive air showers as a universal estimator of cosmic-ray energy. *Phys. Rev. Lett.* 116, 241101. doi: 10.1103/PhysRevLett.116.241101

Anwar, S. L., Wulaningsih, W., and Lehmann, U. (2017). Transposable elements in human cancer: causes and consequences of deregulation. *Int. J. Mol. Sci.* 18 (5), 974. doi: 10.3390/ijms18050974

Balas, M. M., and Johnson, A. M. (2018). Exploring the mechanisms behind long noncoding RNAs and cancer. *Noncoding RNA Res.* 3, 108–117. doi: 10.1016/j.ncrna.2018.03.001

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* 47, D1034–D1037. doi: 10.1093/nar/gky905

Bhan, A., and Mandal, S. S. (2015). LncRNA HOTAIR: a master regulator of chromatin dynamics and cancer. *Biochim. Biophys. Acta* 1856, 151–164. doi: 10.1016/j.bbcan.2015.07.001

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long noncoding RNA and cancer: a new paradigm. *Cancer Res.* 77, 3965–3981. doi: 10.1158/0008-5472.CAN-16-2634

Brockdorff, N. (2018). Local tandem repeat expansion in Xist RNA as a Model for the Functionalisation of ncRNA. *Non-coding RNA,* 4 (4), 28. doi: 10.3390/ncrna4040028

Chen, J., Shishkin, A. A., Zhu, X., Kadri, S., Maza, I., Guttman, M., et al. (2016). Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* 17, 19. doi: 10.1186/s13059-016-0880-9

Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* (San Francisco, California, USA: ACM). doi: 10.1145/2939672.2939785

Chung, F. R. K. (1997). *Spectral Graph Theory*. United States: American Mathematical Society.

De Kok, J. B., Verhaegh, G. W., Roelofs, R. W., Hessels, D., Kiemeney, L. A., Aalders, T. W., et al. (2002). DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *Cancer Res.* 62, 2695–2698.

Di Ruocco, F., Basso, V., Rivoire, M., Mehlen, P., Ambati, J., De Falco, S., et al. (2018). Alu RNA accumulation induces epithelial-to-mesenchymal transition by modulating miR-566 and is associated with cancer progression. *Oncogene* 37, 627–637. doi: 10.1038/onc.2017.369

Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., et al. (2019). Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.* 47, D1028–D1033. doi: 10.1093/nar/gky1096

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013

He, X., Cai, D., and Niyogi, P. (2005). "Laplacian score for feature selection," in *Proceedings of the 18th International Conference on Neural Information Processing Systems.* (Vancouver, British Columbia, Canada: MIT Press).

He, X., and Niyogi, P. (2003). "Locality Preserving Projections," in *Proceedings of the 16th International Conference on Neural Information Processing Systems* (Whistler, British Columbia, Canada: MIT Press).

Hessels, D., Klein Gunneweik, J. M., Van Oort, I., Karthaus, H. F., Van Leenders, G. J., Van Balken, B., et al. (2003). DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur. Urol.* 44, 8–15discussion 15–16. doi: 10.1016/S0302-2838(03)00201-X

Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019). HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* 47, D1013–D1017. doi: 10.1093/nar/gky1010

Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Kanwal, R., and Gupta, S. (2010). Epigenetics and cancer. *J. Appl. Physiol.* 109 (1985), 598–605. doi: 10.1152/japplphysiol.00066.2010

Kirk, J. M., Kim, S. O., Inoue, K., Smola, M. J., Lee, D. M., Schertzer, M. D., et al. (2018). Functional classification of long non-coding RNAs by k-mer content. *Nat. Genet.* 50, 1474–1482. doi: 10.1038/s41588-018-0207-8

Kogure, T., Yan, I. K., Lin, W. L., and Patel, T. (2013). Extracellular Vesicle-Mediated Transfer of a Novel Long Noncoding RNA TUC339: A Mechanism of Intercellular Signaling in Human Hepatocellular Cancer. *Genes Cancer* 4, 261–272. doi: 10.1177/1947601913499020

Larranaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., et al. (2006). Machine learning in bioinformatics. *Brief Bioinform.* 7, 86–112. doi: 10.1093/bib/bbk007

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., et al. (2017). Feature Selection: A Data Perspective. *ACM Comput. Surv.* 50, 1–45. doi: 10.1145/3136625

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97. doi: 10.1093/nar/gkt1248

Mercer, T. R., and Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* 20, 300–307. doi: 10.1038/nsmb.2480

Muller, R., Weirick, T., John, D., Militello, G., Chen, W., Dimmeler, S., et al. (2016). ANGIOGENES: knowledge database for protein-coding and noncoding RNA genes in endothelial cells. *Sci. Rep.* 6, 32475. doi: 10.1038/srep32475

Nakamura, M., Kajiwara, Y., Otsuka, A., and Kimura, H. (2013). LVQ-SMOTE - learning vector quantization based synthetic minority over-sampling Technique for biomedical data. *BioData Min.* 6 (1), 16.

Pitkanen, E., Cajuso, T., Katainen, R., Kaasinen, E., Valimaki, N., Palin, K., et al. (2014). Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer. *Oncotarget* 5, 853–859. doi: 10.18632/oncotarget.1781

Ransohoff, J. D., Wei, Y., and Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell. Biol.* 19, 143–157. doi: 10.1038/nrm.2017.104

Renganathan, A., and Felley-Bosco, E. (2017). Long noncoding RNAs in cancer and therapeutic potential. *Adv. Exp. Med. Biol.* 1008, 199–222. doi: 10.1007/978-981-10-5203-3_7

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer statistics, 2018. *CA Cancer J. Clin.* 68, 7–30. doi: 10.3322/caac.21442

Solovyov, A., Vabret, N., Arora, K. S., Snyder, A., Funt, S. A., Bajorin, D. F., et al. (2018). Global cancer transcriptome quantifies repeat element polarization between immunotherapy responsive and T cell suppressive classes. *Cell Rep.* 23, 512–521. doi: 10.1016/j.celrep.2018.03.042

Tinzl, M., Marberger, M., Horvath, S., and Chypre, C. (2004). DD3PCA3 RNA analysis in urine–a new perspective for detecting prostate cancer. *Eur. Urol.* 46, 182–186; discussion 187. doi: 10.1016/j.eururo.2004.06.004

Wang, J., Zhang, X., Chen, W., Hu, X., Li, J., and Liu, C. (2019). Regulatory roles of long noncoding RNAs implicated in cancer hallmarks. *Int. J. Cancer* doi: 10.1002/ijc.32277

Wang, J., Zhang, X., Chen, W., Li, J., and Liu, C. (2018). CRlncRNA: a manually curated database of cancer-related long non-coding RNAs with experimental proof of functions on clinicopathological and molecular features. *BMC Med. Genomics* 11, 114. doi: 10.1186/s12920-018-0430-2

Welter, D., Macarthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006. doi: 10.1093/nar/gkt1229

Yang, W., Hong, L., Xu, X., Wang, Q., Huang, J., and Jiang, L. (2017). LncRNA GAS5 suppresses the tumorigenesis of cervical cancer by downregulating miR-196a and miR-205. *Tumour Biol.* 39, 1010428317711315. doi: 10.1177/1010428317711315

Yang, Y. R., Zang, S. Z., Zhong, C. L., Li, Y. X., Zhao, S. S., and Feng, X. J. (2014). Increased expression of the lncRNA PVT1 promotes tumorigenesis in non-small cell lung cancer. *Int. J. Clin. Exp. Pathol.* 7, 6929–6935.

Yarmishyn, A. A., and Kurochkin, I. V. (2015). Long noncoding RNAs: a potential novel class of cancer biomarkers. *Front. Genet.* 6, 145. doi: 10.3389/fgene.2015.00145

You, J. S., and Jones, P. A. (2012). Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* 22, 9–20. doi: 10.1016/j.ccr.2012.06.008

Zan, X. Y., and Li, L. (2019). Construction of lncRNA-mediated ceRNA network to reveal clinically relevant lncRNA biomarkers in glioblastomas. *Oncol. Lett.* 17, 4369–4374. doi: 10.3892/ol.2019.10114

Zhang, B., Gunawardane, L., Niazi, F., Jahanbani, F., Chen, X., and Valadkhan, S. (2014). A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. *Mol. Cell Biol.* 34, 2318–2329. doi: 10.1128/MCB.01673-13

Zhang, X., Wang, J., Li, J., Chen, W., and Liu, C. (2018). CRlncRC: a machine learning-based method for cancer-related long noncoding RNA identification using integrated features. *BMC Med. Genomics* 11, 120. doi: 10.1186/s12920-018-0436-9

Zhao, T., Xu, J., Liu, L., Bai, J., Xu, C., Xiao, Y., et al. (2015). Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features. *Mol. Biosyst.* 11, 126–136. doi: 10.1039/C4MB00478G

Zheng, R., Li, M., Chen, X., Wu, F. X., Pan, Y., and Wang, J. (2018). BiXGBoost: a scalable, flexible boosting based method for reconstructing gene regulatory networks. *Bioinformatics.* 35 (11), 1893-1900. doi: 10.1093/bioinformatics/bty908

Predicting lncRNA-miRNA Interaction *via* Graph Convolution Auto-Encoder

Yu-An Huang[1†], Zhi-An Huang[2†], Zhu-Hong You[1*], Zexuan Zhu[3*], Wen-Zhun Huang[1], Jian-Xin Guo[1] and Chang-Qing Yu[1]

[1] College of Electronics and Information Engineering, Xijing University, Xi'an, China, [2] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, [3] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China

The interaction of miRNA and lncRNA is known to be important for gene regulations. However, the number of known lncRNA-miRNA interactions is still very limited and there are limited computational tools available for predicting new ones. Considering that lncRNAs and miRNAs share internal patterns in the partnership between each other, the underlying lncRNA-miRNA interactions could be predicted by utilizing the known ones, which could be considered as a semi-supervised learning problem. It is shown that the attributes of lncRNA and miRNA have a close relationship with the interaction between each other. Effective use of side information could be helpful for improving the performance especially when the training samples are limited. In view of this, we proposed an end-to-end prediction model called GCLMI (Graph Convolution for novel lncRNA-miRNA Interactions) by combining the techniques of graph convolution and auto-encoder. Without any preprocessing process on the feature information, our method can incorporate raw data of node attributes with the topology of the interaction network. Based on a real dataset collected from a public database, the results of experiments conducted on k-fold cross validations illustrate the robustness and effectiveness of the prediction performance of the proposed prediction model. We prove the graph convolution layer as designed in the proposed model able to effectively integrate the input data by filtering the graph with node features. The proposed model is anticipated to yield highly potential lncRNA-miRNA interactions in the scenario that different types of numerical features describing lncRNA or miRNA are provided by users, serving as a useful computational tool.

Keywords: LncRNA–miRNA interactions, graph convolution network, computational prediction model, regulation network, system biology model

INTRODUCTION

In recent years, the knowledge of the role of RNA in gene regulation has emerged from the advances in next-generation sequencing technologies, allowing a deeper and more comprehensive study on full transcriptomes of organisms. It is demonstrated by the ENCODE project that in mammals noncoding RNA could constitute a substantial majority of transcripts within the genome (Science, 2004). There is as much as 98% of the whole human genome encoding for noncoding transcripts, most of which are processed to generate small noncoding RNA such as miRNA, or long noncoding RNA (lncRNA).

Even though the current understanding of lncRNA functions is still limited, it is revealed that they are key regulators of multiple biological processes through a complex mechanism in which their modular structure permits them to interact with specific proteins, RNA, and DNA (Wapinski and Chang, 2011). On the other hand, miRNAs post-transcriptionally regulate the expression of their target genes. Accumulating studies are showing that, similar to the protein-coding genes, both of these two types of noncoding RNA influence almost all aspects of biology (Lu et al., 2005). The aberrant expression level of noncoding RNAs appears to be one of the initiating factors of different types of disease including cancers (Lewis et al., 2003; Calin and Croce, 2006).

A number of studies have begun to uncover the interactions between miRNA and lncRNA and more and more details about the influence of miRNA on lncRNA function is now coming into view (Tay et al., 2014). In some cases, miRNA triggers lncRNA decay. In other cases, lncRNA acts as miRNA sponges/decoys, or competes with miRNA for binding mRNAs or generate miRNAs. Recently, the hypothesis of competing endogenous RNA (ceRNA) has been proposed and become a mainstream view for explaining the interaction between lncRNA and miRNA (Salmena et al., 2011). Specifically, lncRNA competes with pseudogenes, circular RNAs and messenger RNAs for binding or sequestering microRNAs from the same pool through matching the miRNA response elements (MREs). Considering that both lncRNA and miRNA are keys to regulate gene expression and they interact with each other, it is not unexpected that their relationship in interaction network is firmly regulated. Understanding the lncRNA-miRNA interactions networks governing the initiation and development of diverse diseases is essential but remains largely uncompleted (Karreth and Pandolfi, 2013).

LncRNAs and miRNAs interact with each other forming a huge and complex regulation network for controlling gene expression on transcriptional, post-transcriptional, and post-translational levels. Through this multi-level regulation, these two vast families of noncoding RNAs are involved in almost all aspects of cell cycles including cell division, senescence, differentiation, stress response, immune activation, and apoptosis (Shi et al., 2013). In view of this, interactions of noncoding RNAs on the regulation network have attracted widespread attention in medical research (Huang et al., 2016a). A comprehensive understanding of the molecular and cellular effects of such noncoding interaction can offer great insight into the disease mechanism at a molecular level. Noncoding RNAs in those interactions newly discovered to be associated with a specific disease can be regarded as potential diagnostic markers and therefore is of high value in therapeutic approaches.

Some efforts have been made to design a computational method to meet the emerging need for an accurate prediction of lncRNA-miRNA interactions on a large scale. One popular direction is to do statistical analysis on the data collected from biological experiments. For example, Sumazin et al. attempted to construct a miRNA-mediated network of coding and non-coding RNA interactions for inferring the key dysregulation of ncRNA expression in pathogenesis (Sumazin et al., 2011). The algorithm of Hermes they proposed for such network calculates the statistical significance of each RNA-miRNA-RNA triplet

by matching the expression profiles of gene and miRNAs in glioblastoma. Similarly, Paci et al. and Conte et al. construct lncRNA-miRNA-RNA interaction network by calculating so-called sensitivity correlation which denotes the difference between Pearson correlation coefficient and partial correlation coefficient for each triplet obtained from the breast cancer data (Paci et al., 2014; Conte et al., 2017). To investigate the underlying roles of lncRNA in the diseases of prostate cancer and lung adenocarcinoma, Du et al. and Sui et al. integrate different types of attribute data of RNA to construct a regulatory network in which lncRNAs centrally mediate miRNAs (Du et al., 2016; Sui et al., 2016). All of these methods are designed based on statistics measure and their statistics analysis is for a specific type of disease. To identify the noncoding RNA-mediated sponge regulatory network in various diseases recorded in TCGA and UCEC, Wang et al. construct lncRNA-miRNA-gene triplet networks yielded by prediction algorithms. Based on such constructed networks, hypothesis testing approach is implemented for predicting those triplets associated with diseases (Wang et al., 2015).

Another direction for predicting lncRNA-miRNA is based on matching seed sequences. Most computational tools of such type, such like TargetScan, miRanda and RNAhybrid, aim at predicting miRNA targets selecting evolutionarily conserved microRNA binding sites (Zheng et al., 2017). However, it is pointed out by Natalia et al. that prediction using these methods could be of high false positives and often biologically irrelevant (Pinzón et al., 2017). They show that the interaction between lncRNAs and miRNAs is dose-sensitivity. In view of this, it is hardly to predict miRNA target only using the sequence information as they are not always dose-sensitive enough to be functionally regulated by miRNAs.

The past decade has witnessed the exponential growth of noncoding RNA expression profiling data in cancers but the number lncRNA-miRNA interactions underlying such big data is still limited (Zheng et al., 2016b). Considering that different attributes of noncoding RNAs are being continuously updated, the big data about noncoding RNAs poses significant challenge for data analysis and integration, which is important for predicting new links on the current sparse lncRNA-miRNA interaction network. In taking forward this area of work, some methods based on machine learning have been proposed. Huang et al. propose the first prediction model for inferring lncRNA-miRNA on a large scale. Specifically, the EPLMI model uses a network diffusion method on weighted networks associated with expression profiles, sequence information and biological function (Huang et al., 2018). The basic assumption of this method is based on the finding that miRNAs of similar patterns tend to interact with similar lncRNA and vice versa. However, how to define the similarity among noncoding RNAs based on their expression profile is still an open problem. EPLMI model use Person correlation coefficients to compute such similarity, which means it assumes each element in the noncoding RNA features equally contributes to the similarity score. However, it would be inappropriate for the nature of its mechanism.

In recent years, the advance of deep learning fuels the widespread use of data mining in many different science areas including bioinformatics (Li et al., 2016). Specially, graph

convolution comes to be a powerful and popular technique in data mining for graph-based data. It proves to be powerful for its ability to automatically learn latent features from an end-to-end model structure. The hidden layers within the model thus are able to extract meaningful information from the raw input data. In this work, we introduce the technique of graph convolution into the model of autoencoder for building an end-to-end deep learning prediction model called GCLMI for inferring new lncRNA-miRNA interaction on a large scale. Specifically, two different layers are respectively designed to encode and decode the raw feature of each nodes on the input graph. As a result, the decoder can yield a fully-connected network in which the predicted score of each link represent the confidence coefficient of it to be true. Different from the sequence-based algorithms which only consider the sequence information, GCLMI is a network-based algorithm which considers the known lncRNA-miRNA interactions along with the expression levels of lncRNA and miRNA. In addition, GCLMI aims to compute the possibility of a lncRNA-miRNA pair to be interactive in biological processes while sequence-based tools aim to predict the binding sites of miRNA in transcripts.

To evaluate the prediction performance of the proposed model, we implement it in a real dataset of lncRNA-miRNA interactions. By using the frameworks of 2-fold, 5-fold and 10-fold cross validation, the prediction model yielded average AUCs of 0.8492+/−0.0013, 0.8567+/−0.0009 and 0.8590+/−0.0005, respectively. The results of a series of comparison experiments show that the model we present is superior to some methods previously proposed. In addition, the results also illustrate the ability of graph convolution to integrate the raw features of nodes and the topology of graph. The experimental results overall prove that the deep learning-based model we proposed is reliable to yield accurate results and robust to parameter settings. It is anticipated that the proposed model could be served as a useful computational tool for predicting large-scale lncRNA-miRNA interactions in the scenario that know lncRNA-miRNA interactions along with their expression profile are given by users.

METHOD

Materials

The number of known lncRNA-miRNA interactions is still limited and expression profiles of lncRNA and miRNA are often be used for inferring those lncRNA-miRNA pairs of high correlations. Although the number such results is huge, but they are not truly confirmed by the experiments based on CLIP-Seq techniques and therefore would negatively affect the prediction results (Zheng et al., 2016a). To obtain the ground true data resource for our prediction, we collected a dataset of lncRNA-miRNA interactions that are experimentally confirmed from the lncRNASNP database (version v1.0). lncRNASNP is a comprehensive database for lncRNA and provides different kinds of relevant data resource including lncRNA expression profiling, expanded lncRNA-associated diseases, and noncoding variants in lncRNAs (available at http://bioinfo.life.hust.edu.cn/lncRNASNP). The database matches the IDs of lncRNAs and

integrates data from different public databases including that of lncRNA-miRNA interactions from starBase. Eight thousand ninety-one pairwise interactions including 780 types of lncRNA and 275 types of miRNA are totally recorded (Gong et al., 2014). Such interactions have already been verified *via* laboratory examination and therefore are of high confidence.

There are different types of data able to be used as the features of lncRNA and miRNA, such as sequence information of nucleotides, expression profiles, target genes and predicted functional annotations (Zheng et al., 2018). Sequence information is complete for both of lncRNA and miRNA but is too complicated for model to learn as it is nominal and of different length for different types. The links between noncoding RNA and target genes would be meaningful for inferring their interactions but such information is scarce and incomplete for many types of them. The functional annotation of noncoding RNA is important for understanding the characters of one noncoding RNA and many works have been made to inferring them by considering different types of complementary data. However, such information is yielded by prediction algorithms with additional assumptions and therefore is possible to cause computation bias on further prediction models. The superiority of expression profile data to others has been illustrated in our previous work by the experimental results (Huang et al., 2018). For such reason, we only focus on the expression profiles of noncoding RNAs in this work.

To collect the expression profile data of lncRNAs, we match the ids of lncRNAs from two different databases of lncRNASNP and NONCODE (http://www.noncode.org/) (Bu et al., 2011). For 780 types of lncRNA recorded in the lncRNASNP database, 450 of them are successfully matched with their expression profiles. The data of expression profile for each type of lncRNA present the expression level of it in 22 different human tissues or cell lines. For the features of miRNAs, we collect them from the microRNA.org database (http://www.microrna.org/) (Betel et al., 2008). As a result, the ids of 230 out of 275 types of miRNA are successfully converted from lncRNASNP into microRNA.org database. Each entry of miRNA expression profiles consists of 172 values describing the expression level of such miRNA in 172 various tissues and cell lines in human body.

Graph Convolution

It is still an open problem to define the convolution operator on a graph and generalizing convolutional neural networks (CNNs) to arbitrary graphs comes to be a recent area of interest (Kearnes et al., 2016). So far, the approaches with graph convolution could be categorized into two types: i) one is based on definitions of spatial convolution and ii) the other is based on the graph spectral theory. The latter is more popular and is elegantly defined as a multiplication in the graph Fourier domain. The spectral framework was first to be introduce in the context of graph CNNs by Bruna et al. (Bruna et al., 2013). Along this direction, Kipf et al. propose an optimization strategy based on approximal first-order on the spectral filters, reducing its complexity from $O(n^2)$ to $O(|\varepsilon|)$ (see **Figure 1**) (Kipf and Welling, 2016).

To formulate the operator of spectral convolution on graph, given an adjacent matrix A of graph G with its Laplacian $L :=$

FIGURE 1 | The diagram of spectral graph convolution.
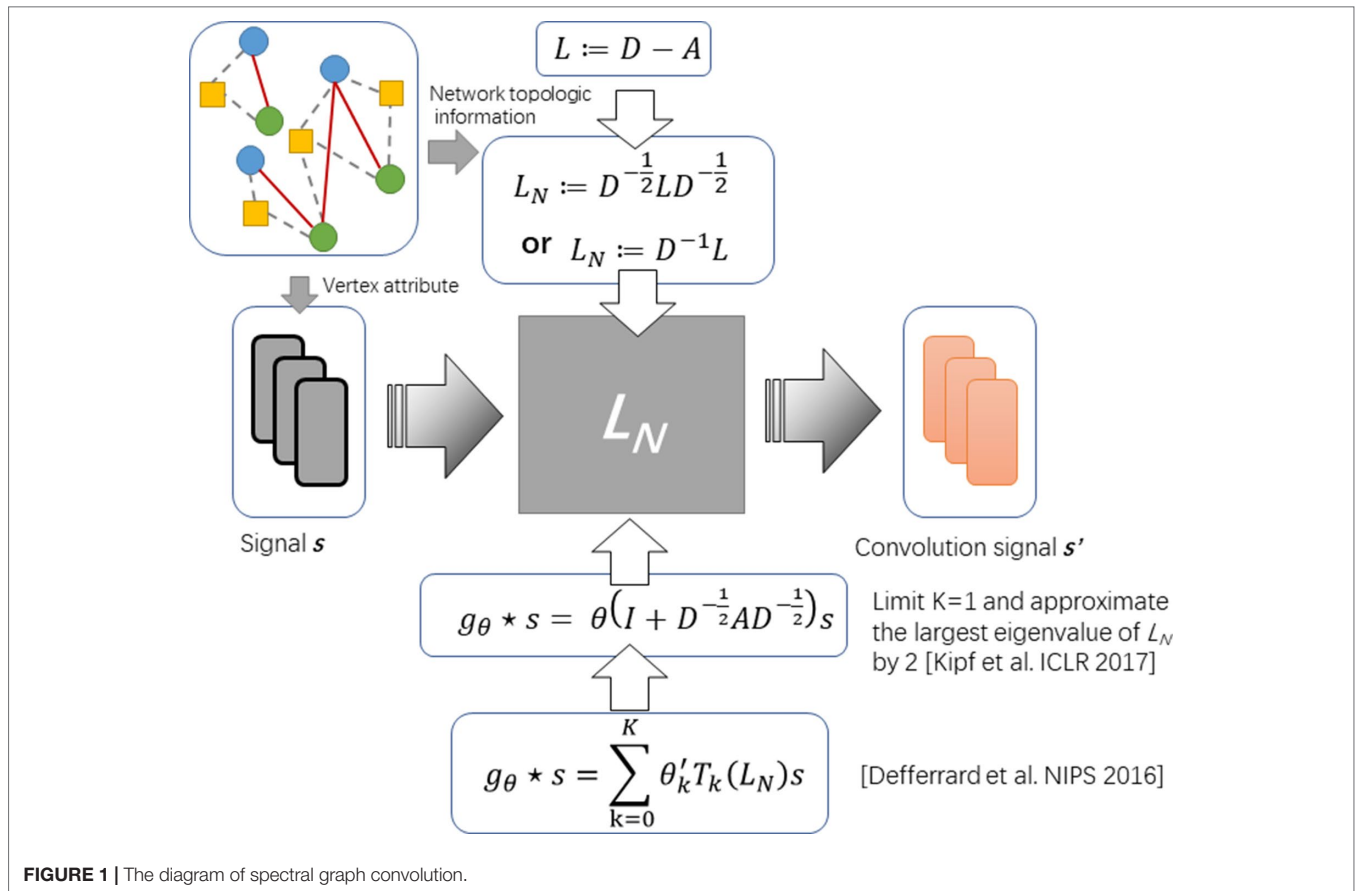
$D – A$ and attributes of each node on graph (say s), Defferrard et al. (2016) propose spectral graph convolution to filter s by a non-parametric kernel $g_\theta(\Lambda) = diag(\theta)$, where θ is a vector of Fourier coefficients. Given L can be decomposed by $L = U\Lambda U^T$, where Λ is the diagonal matrix of eigenvalues and U is eigenvector matrix, such operator could be defined as

$$g_\theta \star s = U g_\theta U^T s \qquad (1)$$

Approximating the spectral filter by using a truncated expansion in terms of Chebyshev polynomials $T_k(s)$ up to K^{th} order, the definition is as follows:

$$g_\theta \star s \approx \sum_{k=0}^{K} \theta'_k T_k(L_N) s \qquad (2)$$

where T_k denotes Chebyshev polynomials and θ' is a vector of Chebyshev coefficients. Considering the complexity of computing L is as large as $O(n^2)$, Kipf and Welling. (2016) further simplified this definition by limiting $K = 1$ and approximating the largest eigenvalue of L by 2. The convolution operator comes to be:

$$g_\theta \star s = \theta(I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) s \qquad (3)$$

By introducing the renormalization tricks: $I_n + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \rightarrow \widetilde{D}^{-\frac{1}{2}} \tilde{A} \widetilde{D}^{-\frac{1}{2}}$ with $\tilde{A} = A + I_N$ and $\widetilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, formula (3) can be simplified as:

$$g_\theta \star s = \theta \widetilde{D}^{-\frac{1}{2}} \tilde{A} \widetilde{D}^{-\frac{1}{2}} s \qquad (4)$$

In this work, we follow this definition as formula 4 for design our deep learning model based on the graph convolution.

GCLMI: An Auto-Encoder Prediction Model for lncRNA-miRNA Interactions

In this work, we cast the prediction task for lncRNA-miRNA interactions as a link prediction problem on a heterogeneous bipartite graph. Consider an adjacent matrix of such graph M of shape $N_l \times N_m$, where N_l is the number of lncRNA nodes and Nm is the number of miRNA nodes. Entry M_{ij} in this matrix encode either the interaction between i-th type of lncRNA and j-th type of miRNA is identified by biological experiments or not. The task of prediction can be considered as referring the value of unobserved entries in M using semi-supervised learning on the observed ones.

In an equivalent picture, we can also represent the interaction data by an undirected graph $G = (\nu, \varepsilon, X_l, X_m)$, where X_l and X_m are the feature matrices for the lncRNA nodes and miRNA nodes,

respectively. The goal is to learn embedding features for lncRNAs and miRNAs E by building a graph-based encoder $[E_l, E_m] = f_{en}(v, \varepsilon, X_l, X_m)$ and predicting new links by building a decoder $M' = f_{de}(E_l, E_m)$. E_l and E_m are the feature matrices for lncRNAs and miRNA with shapes of $N_l \times L$ and $N_m \times L$, respectively (see **Figure 2**).

To this aim, our proposed model is composed by two layers of different types: i) an encoder layer for filtering node features of lncRNA and miRNA on the graph of their interaction network and ii) a decoder layer for predicting fully-collected interaction network using the embedding features learned from the former layers.

The inputs of encoder layer include the feature matrixes of lncRNA and miRNA (i.e. F_l and F_m) and the adjacent matrix of known lncRNA-miRNA interaction network (i.e. M). In order to integrate the features of lncRNA and miRNA into one input matrix, an expanded matrix X is constructed based on F_l and F_m as follows:

$$X = \begin{bmatrix} F_l & 0 \\ 0 & F_m \end{bmatrix} \tag{5}$$

Accordingly, the adjacent matrix of known lncRNA-miRNA interaction network is expanded as:

$$A = \begin{bmatrix} 0 & M \\ M^T & 0 \end{bmatrix} \tag{6}$$

Based on the above two input matrixes, we compute a graph convolution matrix G according to formula 4:

$$G = X_{rw}\left(I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) \tag{7}$$

The hidden layer is then built based on G by introducing its weight matrix We and bias matrix Be. With ReLU as the activation function, the output E of the encoder layer would be as follows:

$$E = \text{ReLU}\left(G \cdot W_e + B_e\right) = \begin{bmatrix} E_l \\ E_m \end{bmatrix} \tag{8}$$

where the trainable weight matrix $W_e \in (D_l + D_m) \times N_e$ transforms the convolution matrix G into a hidden matrix E. N_e denotes the number of latent factors and is set manually. The

output layer learned from the encoder layer is a projection from the space of raw features into a hidden space with lower rank. As lncRNA and miRNA are known to interact with each through the MRE on transcripts, the design of hidden accords with the nature that lncRNA, MRE, miRNA is associated in a three-layer relation network.

The output of the encoder layer has two components, which are the matrix of the embedding feature matrix of lncRNA E_l and that of miRNA E_m. Introducing a trainable weight matrix Wd, the decoder layer is then built based on these separate matrixes of the same raw dimension as follows:

$$M' = E_l W_d E_m^T \tag{9}$$

The output matrix M' clearly has the same shape out the input matrix M. As matrix M' is numerical, it describes the weight of links in a fully-connected network. All lncRNA-miRNA pairs with value of 0 in matrix M would be assigned a predicted value by the decoder. Those pairs with high predicted scores are anticipated to more possibly be connected.

To train the model of GCLMI in a semi-supervised learning manner, we use the strategy of negative sampling. Specifically, in each epoch of training process, we randomly select a fixed number of negative samples from the unlabeled lncRNA-miRNA pairs. The loss function of our training is defined as follows:

$$\mathcal{L} = \sqrt{\frac{\sum ij; \Omega_{p,ij} = 1 or \Omega_{n,ij} = 1\left(M'_{ij} - M_{ij}\right)^2}{\sum ij\left(\Omega_{p,ij} + \Omega_{n,ij}\right)}} + \frac{1}{2}\|W_e\|^2 + \frac{1}{2}\|W_d\|^2 + \frac{1}{2}\|B_e\|^2 \tag{10}$$

where the matrices $\Omega_p \in \{0,1\}^{N_l \times N_m}$ and $\Omega_n \in \{0,1\}^{N_l \times N_m}$ denote the masks for positive samples and the negative samples from random sampling, respectively. The first term in equation (10) aims to minimize the prediction error and the second and the third term define the constraint on the weight matrix in encoder and decoder, respectively. As negative sampling is implemented for training, in each epoch the Ω_n would be randomly generated

FIGURE 2 | The flowchart of GCLMI.
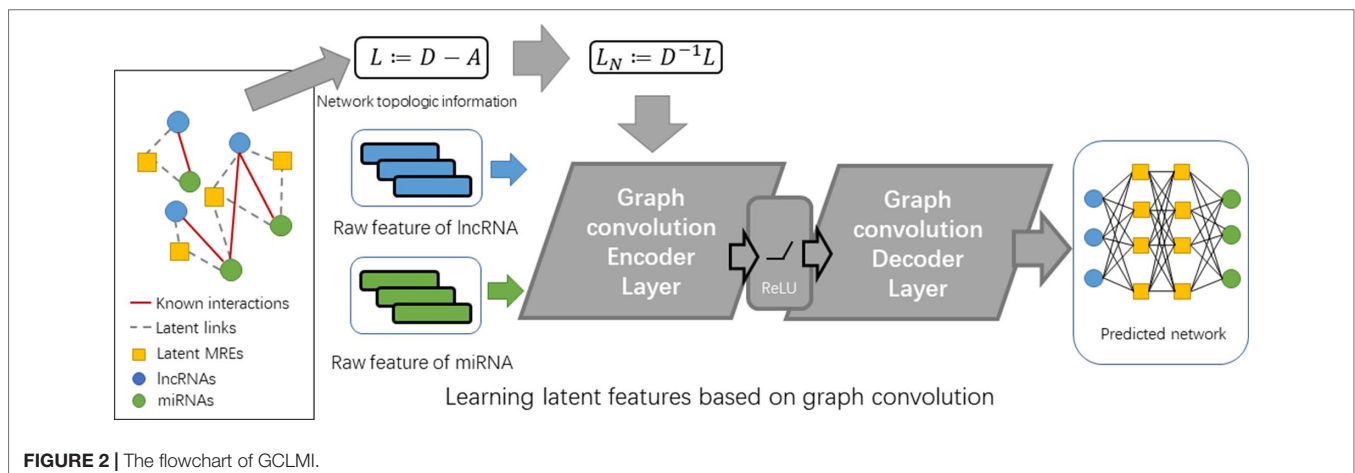
in which the number of "1" would be fixed as a specific percentage of the number of positive samples. Hence, we would only optimize over the positive samples if we set this percentage as 0 or optimize over the positive samples and partial negative samples otherwise.

RESULTS AND DISCUSSION

Evaluation of Graph Convolution's Effectiveness

Using techniques of graph convolution, spectral filter function integrates the information of attribute feature of input node with that of its neighbor nodes on the graph. GCLMI model uses graph convolution to build a data pre-processing module so that it can train the embedding features of nodes in an end-to-end learning manner. In this section, we evaluate the effectiveness of graph convolution with regard to its ability to integrate the raw data of input feature. Specifically, we compare the standard pipeline of GCLMI with the case that the input features are removed. To this aim, each entry of the input feature matrix A in formula 7 is replaced with the value of 1. In this case, the operator of graph convolution would be meaningless as all node features are the same. We implemented such modified computation process in 5-fold cross validation. As a result, without any input of node feature, the GCLMI model yield an AUC of 0.8483 on the 5-fold cross volition experiment, significantly lower than AUC of 0.8567 yield by the standard computational pipeline (see **Figure 3**). The result shows that the graph convolution designed in the model of GCLMI is feasible and effectively to integrate the raw data of feature inputs (see **Figure 4**).

Evaluation of the Impact of Negative Sampling

There is still no biological experiment confirming any lncRNA-miRNA pair that are definitely not interactive so that no database can provide the data of negative samples for our training. For

this reason, the prediction task in this work can be considered as a semi-supervised learning problem. Considering the known lncRNA-miRNA network is sparse, sampling on the unlabeled samples could generate a data source in which underlying negative samples are involved. Information of unlabeled data can be properly leveraged to push the limits of poor data resource for training (Zheng et al., 2014). To do so, we implement negative sampling on the unlabeled samples in each training epoch to construct negative sample set for training. However, the number of samples from negative sampling can have an effect on the prediction performance of the proposed model. A larger amount of negative sample can provide data resource for training and good performance could be achieved with more information for model to learn. However, it can also cause the problem of unbalanced training data. In this view, the choice of the size of negative sample set is important for an accurate prediction of GCLMI model. In each training epoch, the size of negative sample set is fixed as a ratio p of that of positive samples. In this section, we explore the prediction performance of GCLMI with different values of p (i.e. 0, 0.5, 1.0, 3.0, 5.0, 10.0).

Figure 5 shows the training loss and training error along with increased training epoch in this series of experiments. We calculated the training loss and training error whose definitions are as the Equation 10 and the first term of Equation 10, respectively. The curves of **Figure 5(A)** and **Figure 5(B)** show that the training processes of GCLMI with different sizes of negative sample set are similar. For most of experiments, the corresponding training loss and training error could be convergent to their lower bounds before the 250th epoch and 150th epoch, respectively, illustrating the computational process is robust to different negative sampling. The prediction performance of GCLMI with different negative sampling is also evaluated. As shown in the **Figure 5**, the prediction performance varies with different sizes of negative sample set in term of the AUC value. Specifically, when the number of negative samples is set as 3 times of positive samples,
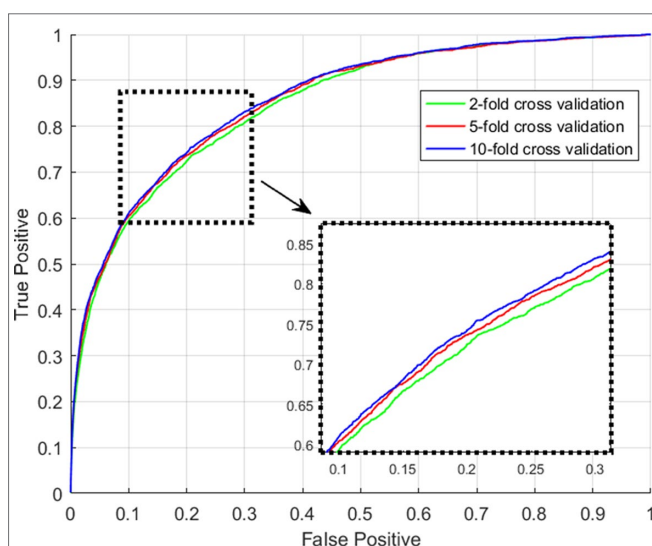

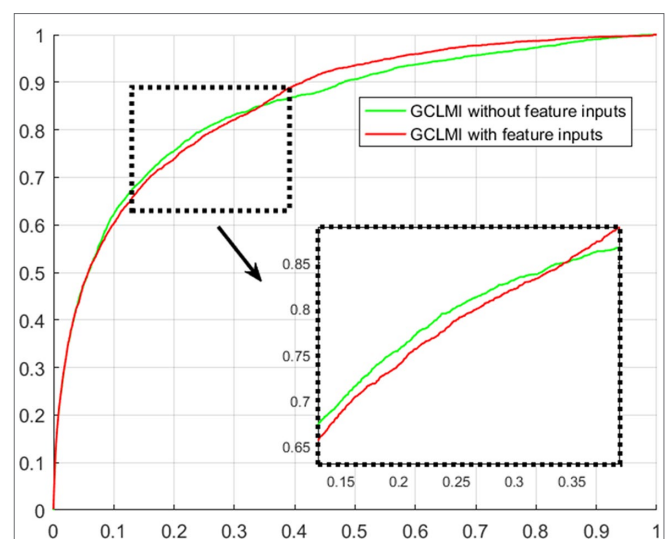**FIGURE 3 |** The ROC curves yielded by GCLMI on 2-fold, 5-fold and 10-fold cross validation.

FIGURE 4 | Evaluation of graph convolution layer w.r.t ROC curves on 5-fold cross validation.

FIGURE 5 | Training process of GCLMI in different training epochs with different negative sample sets. **(A)** and **(B)** illustrate the training loss and training error in training process, respectively.

the model achieves its highest prediction performance with AUC of 0.8567. It also should be noted that the prediction performance of GCLMI declines greatly with p set as 0. As setting p = 0 means that no negative sample is used for training, this result illustrates that negative sampling is effective and necessary for an accurate prediction of large-scale lncRNA-miRNA interactions.

Prediction Performance of GCLMI on k-Fold Cross Validation

For the evaluation of the performance of our proposed model with regard to the prediction accuracy on lncRNA-miRNA interactions, we adopt the evaluation frameworks of 2-fold, 5-fold and 10-fold cross validation. All experiments in this work are conducted on a real dataset involving experimentally-confirmed lncRNA-miRNA interactions. Specifically, in the k-fold cross validation, all known lncRNA-miRNA interactions are roughly divided into k parts, each of which is used as testing sample set in turn and the rest is used as training sample set. After implement the prediction process of GCLMI with training set as input, each testing sample obtain its prediction score presenting the confidence coefficient about the link existence. We consider all the 209,152 unlabeled lncRNA-miRNA pairs as candidate samples and compute the ranks of prediction scores among the candidates.

We consider those testing samples with a higher rank than a given threshold as positive. By setting different threshold in the experiments, we compute the corresponding true positive rates (TPRs, sensitivity) and FPRs (1-specificity) for each threshold. Specifically, given a threshold, sensitivity denotes the percentage of testing samples with higher ranks and specificity is the percentage of testing sample with lower ranks. Based on TPRs and FPRs, the corresponding ROC curve (receiver operating characteristic curve) is plotted and the area under the curve (AUC) is computed as a main evaluation criterion for the performance. The value of AUC

lies between 0.5 and 1, where 0.5 means a purely random guess and 1 denotes a perfect prediction. As some of known lncRNA-miRNA interactions take turns to be used as testing samples and assumed to be unknown in the prediction process, if they obtained a high rank among those unlabeled samples in general, it means the prediction performance is good and prediction model is feasible. In addition, as the division of sample sets is random, we repeat the sampling implement GCLMI model with different sample division 20 times to avoid the bias caused by such partition. The standard deviation is also calculated for each cross validation. As a result, conducting GCLMI on the collected dataset, we obtain good prediction performance with average AUCs of 0.8492+/−0.0013, 0.8567+/−0.0009 and 0.8590+/−0.0005 in 2-fold, 5-fold and 10-fold cross validation, respectively (see **Figure 6**). As shown in **Table 1**, the increase of fold number in cross validation boosts the performance of GCLMI because more data resource in training set would benefit the prediction performance. In this view, we
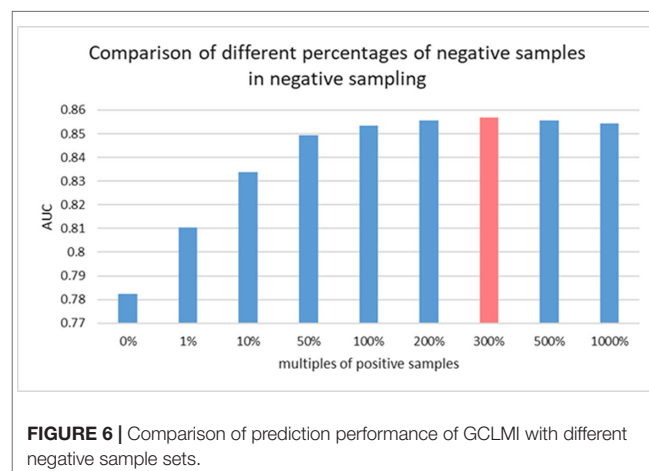

FIGURE 6 | Comparison of prediction performance of GCLMI with different negative sample sets.

TABLE 1 | Prediction performance w.r.t. AUC in 2-fold, 5-fold and 10-fold cross validation.

Cross validation	2-fold CV	5-fold CV	10-fold CV
Average AUC	0.8492 + /-0.0013	0.8567 + /-0.0009	0.8590+/-0.0005

TABLE 2 | Performance comparison among different methods by using RNA expression profile-based similarity in the framework of 5-fold cross validation.

Method	5-fold cross validation
lncRNA-based CF	0.6359 + /−0.0024
miRNA-based CF	0.8235 + /−0.0015
SVD-based CF	0.4967 + /−0.0340
Katz-based method	0.7439 + /−0.0017
Basic latent factor model	0.8253 + /−0.0024
EPLMI (Huang et al., 2018)	0.8447 + /−0.0017
The proposed model	0.8567 + /−0.0009

anticipate that GCLMI model is able to yield more reliable results with more ground true input data in the future. The results of high AUCs illustrate the reliable performance for predicting lncRNA-miRNA interaction on a large scale.

Performance Comparison With Other Similarity-Based Methods

Current approaches to predict new links on biological bipartite networks are mainly based on similarity-based assumption (Sun et al., 2018). Given a network in which two types of nodes representing two kinds of research objects are involved, most of previous prediction model assumes that similar objects of one type tend to be associated with those of another type (Huang et al., 2016b). Therefore, their prediction performance could be greatly influenced by the measurement they adopt to calculate the similarity scores among object of the same types (Huang et al., 2017b). For example, KATZHMDA model calculates the similarity of microbes using the Gaussian kernel and EPLMI model uses Person correlation coefficient for the similarity of lncRNA and miRNA based on their expression profiles (Huang et al., 2016a; Huang et al., 2018). However, such linear computation method may so simple that it fails to describe the general similarity of lncRNA or miRNA with regards to their roles in regulation network based on their expression profile. To bypass such barrier, we propose an end-to-end prediction model using graph convolution technique and therefore the prediction is free of any calculation for similarity.

To further evaluate the prediction performance, several similarity-based methods are implemented on the same dataset for performance evaluation, using the same similarity matrices of lncRNA and miRNA based on Person correlation coefficients of expression profiles. The comparison methods include two types of neighbor-based collaborative filtering (i.e. lncRNA-based CF and miRNA-based CF), matrix factorization-based method (i.e. SVD-based CF and basic latent factor model) and EPLMI. Using 5-fold cross validation on the same dataset, the comparison result shows that the proposed model has the best prediction ability among five comparison methods with highest AUC values of 0.8567+/−0.0009

(see **Table 2**). We consider such superior in performance on link prediction is benefited from the end-to-end learning approach as GCLMI model was designed. It is anticipated that such end-to-end prediction model as we proposed would yield more accurate prediction results with a larger amount of high-dimension data as inputs in the future.

CONCLUSION

Increasing evidence show that lncRNA and miRNA collaborate to form a regulation network for gene regulation. Interactions between lncRNA and miRNA thus provide great insights into understanding the molecular mechanism of the initiation and development of various types of complex diseases. However, little effort has been made to develop computational approach to predict lncRNA-miRNA interaction on a large scale. The main challenge comes from the small number of known interactions between lncRNA and miRNA (i.e. the sparsity of lncRNA-miRNA interaction network) and the limited understanding on the underlying pattern on lncRNA-miRNA interaction.

To address this issue, we proposed a deep learning-based prediction model named GCLMI which can effectively predict large-scale lncRNA-miRNA interactions. Given raw data as RNA attribute features, the GCLMI model is able to extract meaningful embedding features for both miRNA and lncRNA in an end-to-end training manner. The results of a series of experiments show that the low-dimension embedding learned from the proposed model is of good representation ability with regards to their relation on the interaction network. Benefited from the deep learning structure as GCLMI is designed, we anticipate that the proposed model could be used as a useful tool for an accurate prediction of large-scale lncRNA-miRNA interactions in the scenario that additional information describing features of lncRNA and miRNA is offered by the users. In the current version of GCLMI, other types of data relevant to intrinsic features of lncRNA and miRNA, such like ncRNA sequence information and structural data are still inapplicable for GCLMI to handle with, as the graph convolution operator needs numerical data as inputs. In the future, we will investigate solutions about this limitation.

AUTHOR CONTRIBUTIONS

Y-AH and Z-AH conceived the algorithm, developed the program, and wrote the manuscript. Z-HY and ZZ helped with manuscript editing, designed and performed experiments. Y-AH and Z-AH prepared the data sets, carried out analyses and helped with program design. All authors read and approved the final manuscript.

FUNDING

REFERENCES

Betel, D., Wilson, M., Gabow, A., Marks, D. S., and Sander, C. J. N. A. R. (2008). The microRNA. org resource: targets and expression. *Nucleic Acids Res.* 36, D149–D153. doi: 10.1093/nar/gkm995

Bruna, J., Zaremba, W., Szlam, A., and Lecun, Y. J. a P. A. (2013). Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*.

Bu, D., Yu, K., Sun, S., Xie, C., Skogerbø, G., Miao, R., et al. (2011). NONCODE v3. 0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res.* 40, D210–D215. doi: 10.1093/nar/gkr1175

Calin, G. A., and Croce, C. M. J. N. R. C. (2006). MicroRNA signatures in human cancers. *Nat. Rev. Cancer.* 6, 857. doi: 10.1038/nrc1997

Conte, F., Fiscon, G., Chiara, M., Colombo, T., Farina, L., and Paci, P. J. P. O. (2017). Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer. *PLoS One* 12, e0171661. doi: 10.1371/journal.pone.0171661

Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering In: Advances in neural information processing systems. 3844–3852.

Du, Z., Sun, T., Hacisuleyman, E., Fei, T., Wang, X., Brown, M., et al. (2016). Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat. Commun.* 7, 10982. doi: 10.1038/ncomms10982

Gong, J., Liu, W., Zhang, J., Miao, X., and Guo, A.-Y. J. N. a. R. (2014). lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res.* 43, D181–D186. doi: 10.1093/nar/gku1000

Huang, Y.-A., Chan, K. C., and You, Z.-H. (2017a). Constructing prediction models from expression profiles for large scale lncRNA-mirna interaction profiling. *Bioinformatics*. doi: 10.1093/bioinformatics/btx672

Huang, Y.-A., You, Z.-H., Chen, X., Huang, Z.-A., Zhang, S., and Yan, G.-Y.J.J.O.T.M. (2017b). Prediction of microbe–disease association from the integration of neighbor and graph with collaborative recommendation model. *J. Transl. Med.* 15, 209. doi: 10.1186/s12967-017-1304-7

Huang, Y.-A., You, Z.-H., Chen, X., and Yan, G.-Y. J. B. S. B. (2016a). Improved protein-protein interactions prediction via weighted sparse representation model combining continuous wavelet descriptor and PseAA composition. *BMC Syst Biol* 10, 120. doi: 10.1186/s12918-016-0360-6

Huang, Y.-A., You, Z.-H., Li, X., Chen, X., Hu, P., Li, S., et al. (2016b). Construction of reliable protein–protein interaction networks using weighted sparse representation based classifier with pseudo substitution matrix representation features. *Neurocomputing* 218, 131–138. doi: 10.1016/j.neucom.2016.08.063

Huang, Z.-A., Huang, Y.-A., You, Z.-H., Zhu, Z., and Sun, Y. J. B. M. G. (2018). Novel link prediction for large-scale miRNA-lncRNA interaction network in a bipartite graph. *BMC Med Genomics* 11, 113. doi: 10.1186/s12920-018-0429-8

Karreth, F. A., and Pandolfi, P. P. J. C. D. (2013). ceRNA cross-talk in cancer: when ce-bling rivalries go awry. 3, 1113–1121. doi: 10.1158/2159-8290.CD-13-0202

Kearnes, S., Mccloskey, K., Berndl, M., Pande, V., and Riley, P. J. J. O. C.-a. M. D. (2016). Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.* 30, 595–608. doi: 10.1007/s10822-016-9938-8

Kipf, T. N., and Welling, M.J.a.P.A. (2016). Variational graph auto-encoders. arXiv preprint arXiv: 1611.07308.

Lewis, B. P., Shih, I.-H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. J. C. (2003). Prediction of mammalian microRNA targets. *Cell* 115, 787–798. doi: 10.1016/S0092-8674(03)01018-3

Li, Z., Wang, S., Cheng, J., Su, C., Zhong, S., Liu, Q., et al. (2016). Intron lariat RNA inhibits microRNA biogenesis by sequestering the dicing complex in Arabidopsis. *PLoS Genet.* 12, e1006422. doi: 10.1371/journal.pgen.1006422

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435, 834. doi: 10.1038/nature03702

Paci, P., Colombo, T., and Farina, L. J. B. S. B. (2014). Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Syst. Biol.* 8, 83. doi: 10.1186/1752-05090-8-83

Pinzón, N., Li, B., Martinez, L., Sergeeva, A., Presumey, J., Apparailly, F., et al. (2017). microRNA target prediction programs predict many false positives. *Genome Res.* 27, 234–245. doi: 10.1101/gr.205146.116

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. J. C. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014

Science, E. P. C. J. (2004). The ENCODE (ENCyclopedia of DNA elements) project. 306, 636–640. doi: 10.1126/science.1105136

Shi, X., Sun, M., Liu, H., Yao, Y., and Song, Y. J. C. L. (2013). Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett.* 339, 159–166. doi: 10.1016/j.canlet.2013.06.013

Sui, J., Li, Y.-H., Zhang, Y.-Q., Li, C.-Y., Shen, X., Yao, W.-Z., et al. (2016). Integrated analysis of long non-coding RNA-associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. *Int. J. Oncol.* 49, 2023–2036. doi: 10.3892/ijo.2016.3716

Sumazin, P., Yang, X., Chiu, H.-S., Chung, W.-J., Iyer, A., Llobet-Navas, D., et al. (2011). An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* 147, 370–381. doi: 10.1016/j.cell.2011.09.041

Sun, Y., Zhu, Z., You, Z.-H., Zeng, Z., Huang, Z.-A., and Huang, Y.-a. J. B. S. B. (2018). FMSM: a novel computational model for predicting potential miRNA biomarkers for various human diseases. *BMC Syst Biol.* 12, 121. doi: 10.1186/s12918-018-0664-9

Tay, Y., Rinn, J., and Pandolfi, P. P. J. N. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505, 344. doi: 10.1038/nature12986

Wang, P., Ning, S., Zhang, Y., Li, R., Ye, J., Zhao, Z., et al. (2015). Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic Acids Res.* 43, 3478–3489. doi: 10.1093/nar/gkv233

Wapinski, O., and Chang, H. Y. J. T. I. C. B. (2011). Long noncoding RNAs and human disease. *Cell Biol.* 21, 354–361. doi: 10.1016/j.tcb.2011.04.001

Zheng, Y., Guo, J., Wai, C. M., Ming, R., and Sunkar, R. (2018). "MicroRNAs, tasiRNAs, phasiRNAs, and their potential functions in pineapple." In *Genetics and Genomics of Pineapple.* Springer p. 167–182. doi: 10.1007/978-3-030-00614-3_12

Zheng, Y., Ji, B., Song, R., Wang, S., Li, T., Zhang, X., et al. (2016a). Accurate detection for a wide range of mutation and editing sites of microRNAs from small RNA high-throughput sequencing profiles. *Nucleic Acids Res.* 44, e123–e123. doi: 10.1093/nar/gkw471

Zheng, Y., Li, T., Ren, R., Shi, D., and Wang, S. J. B. G. (2014). Revealing editing and SNPs of microRNAs in colon tissues by analyzing high-throughput sequencing profiles of small RNAs. *BMC Genomics* 15, S11. doi: 10.1186/1471-2164-15-S9-S11

Zheng, Y., Li, T., Xu, Z., Wai, C. M., Chen, K., Zhang, X., et al. (2016b). Identification of microRNAs, phasiRNAs and their targets in pineapple. *Trop. Plant Biol.* 9, 176–186. doi: 10.1007/s12042-016-9173-4

Zheng, Y., Liu, L., and Shukla, G. C. J. C. L. (2017). A comprehensive review of web-based non-coding RNA resources for cancer research. *Cancer Lett* 407, 1–8. doi: 10.1016/j.canlet.2017.08.015

Purification and Identification of miRNA Target Sites in Genome Using DNA Affinity Precipitation

Yu Xun[1], Yingxin Tang[1], Linmin Hu[1], Hui Xiao[1], Shengwen Long[1], Mengting Gong[1], Chenxi Wei[1], Ke Wei[1,2]* and Shuanglin Xiang[1]*

[1] Key Laboratory of Protein Chemistry and Developmental Biology of Education Ministry of China, College of Life Science, Hunan Normal University, Changsha, China, [2] Medical School, Hunan University of Chinese Medicine, Changsha, China

Combination with genomic DNA is one of the important ways for microRNAs (miRNAs) to perform biological processes. However, because of lack of an experimental method, the identified genomic sites targeted by microRNA were only located in the promoter and enhancer regions. In this study, based on affinity purification of labeled biotin at the 3′-end of miRNAs, we established an efficiently experimental method to screen miRNA binding sequences in the whole genomic regions *in vivo*. Biotinylated miR-373 was used to test our approach in MCF-7 cells, and then Sanger and next-generation sequencing were used to screen miR-373 binding sequences. Our results demonstrated that the genomic fragments precipitated by miR-373 were located not only in promoter but also in intron, exon, and intergenic. Eleven potentially miR-373 targeting genes were selected for further study, and all of these genes were significantly regulated by miR-373. Furthermore, the targeting sequences located in E-cadherin, cold-shock domain-containing protein C2 (CSDC2), and PDE4D genes could interact with miR-373 in MCF-7 cells rather than HeLa cells, which is consistent with our data that these three genes can be regulated by miR-373 in MCF-7 cells while not in HeLa cells. On the whole, this is an efficient method to identify miRNA targeting sequences in the whole genome.

Keywords: miRNA, target sites, genome, DNA, affinity precipitation

INTRODUCTION

MicroRNAs (miRNAs) are a class of endogenous small non-coding RNAs that are processed from pre-miRNAs by Dicer into 21- to 25-nt double-stranded sequences (Bartel, 2004; He and Hannon, 2004). Through regulating gene expression at the post-transcriptional level, miRNAs can take part in many biological processes including development, cell proliferation, apoptosis, organogenesis, and tumorigenesis (Carrington and Ambros, 2003; Bartel, 2004; Filipowicz et al., 2008). It has been shown clearly that miRNAs regulate gene expression on the post-transcriptional level *via* RNA-induced silencing complex (RISC) pathway in the cytoplasm (Liu et al., 2005a; Dong et al., 2013). However, with the development of new techniques, numerous miRNAs were found enriched in nucleus,

Abbreviations: CSDC2, cold-shock domain-containing C2; ITSN2, intersectin 2; Hsp60, heat shock protein family D (Hsp60) member 1; ALOX5, arachidonate 5-lipoxygenase; ARID2, AT-rich interaction domain 2; PDE4D, phosphodiesterase 4D; SUN1, Sad1 and UNC84 domain-containing 1; ZNF76, zinc finger protein 76; ZNF385B, ZNF385B; TTC34, tetratricopeptide repeat domain 34; TPM1, tropomyosin 1; EVI5, ecotropic viral integration site 5; RPL37, ribosomal protein L37; FANCC, FA complementation group C.

which suggests that microRNAs play important roles in nucleus (Liu et al., 2018). Several studies have proved that miRNA can regulate gene expression *via* interacting with genomic sequences. In 2008, Place et al. reported that miR-373 can up-regulate cold-shock domain-containing protein C2 (CSDC2) and E-cadherin *via* sequence complementarity with the promoter of these genes. MiR-223 can combine with the promoter of NF1A and down-regulates the expression of NF1A (Place et al., 2008).

Based on the existing rule of interaction between miRNA and mRNA, some software tools for predicting miRNA binding sites in genome are developed. However, it is hard to accurately predict microRNA target sites in genome, for the mechanism of gene regulation by miRNAs *via* combination with genomic DNA remains to be elucidated. First, the location of miRNA binding sites in genome should be further studied. Janowski et al. found that small dsRNAs, which are completely complimentary with the sequence in the region −56 to +17 of the promoter, can up-regulate the expression of genes (Janowski et al., 2007). Then Meng et al. reported that the siRNA binding position can be located around −1611 from the transcription start site (Meng et al., 2016). Moreover, it was also reported that miRNA can bind in the enhancer region and increase the transcription activity of neighboring genes (Xiao et al., 2017). Second, the mechanism of interaction between genome and miRNA has not been fully illustrated. Some papers suggested that the 2–8 nt from the 5′-end of the antisense is the key to transcription activation (Xiao et al., 2017). However, it is also reported that let-7i can interact with promoter TATA-box motifs of interleukin (IL)-2 because of low minimal free energy (MFE) value (−27.6 kcal/mol), while the "seed region" of let-7i is not completely complementary with IL-2 promoter, which suggests that the complement of 5′-end of miRNA with target sequence is not the only principle for microRNA target prediction (Zhang et al., 2014). Finally, the prediction based on bioinformatics is insufficient to reflect the real condition *in vivo*, for the epigenetic modification of genome may affect the interaction of miRNA with targeting site (Liu et al., 2018).

Our recent study reported a convenient experimental approach for the isolation and identification of binding miRNAs for messenger RNA by applying short biotinylated DNA anti-sense oligonucleotides mix to enhanced green fluorescent protein (EGFP) mRNA, which was fused to target gene mRNA (Wei et al., 2014). We wonder whether this affinity assay could be used to screen miRNA binding sequence in genomic regions *via* biotinylated miRNA of interest. In the present study, based on biotinylated miRNA capture affinity technique, we have developed an experimental procedure for searching miRNA targeting sequences in the promoter and even in whole genomic regions (**Figure 1**). MiR-373 has been used to test our method in MCF-7. First, we proved that biotinylated miR-373, with the same function as miR-373, can up-regulate the expression of E-cadherin, which have been reported to be up-regulated by miR-373 *via* targeting its promoter. Then using the method as described in this paper, we have collected DNA fragments precipitated by biotinylated miR-373 or negative control RNA. Semi-quantitative polymerase chain reaction (PCR) and real-time PCR showed the E-cadherin promoter and CSDC2

promoter, in the previously reported miR-373 binding site, can be pulled down by biotinylated miR-373 rather than negative control RNA, which suggests that our approach is feasible. Then to find the unknown miR-373 binding sequence, the DNA fragments were inserted into pGEM-T vectors (Promega) and sequenced. Ten unreported miR-373 binding sequences were identified. Interestingly, six identified sequences were located in intron of genes and two sequences in intergenic. Only two of the rest sequences were in the promoter of gene. Western blot and real-time PCR demonstrated that six of seven identified genes can be up-regulated by miR-373 in MCF-7. Interestingly, our results shown that miR-373 cannot improve the expression of E-cadherin, CSDC2, and PDE4D in HeLa cells, which is consistent with our data that miR-373 targeting sequences of these genes cannot be precipitated by miR-373 in HeLa cells. Finally, to efficiently screen miRNA targeting genomic sequence, next-generation sequencing experiment was used to detect the samples precipitated by miR-373 and numerous miR-373 targeting sites were sequenced. On the whole, we developed an efficient approach to screen miRNA targeting genomic sequence and provided a new perspective for studying the interaction of miRNA and genome.

MATERIALS AND METHODS

Materials
The biotinylated miR-373 and biotinylated negative control miRNA (biotinylated NC miRNA) were synthesized from TaKaRa Biotechnology (Dalian, China) *via* labeling with biotin at the 3′-end of the miRNA (**Table 1**). pGEM-T vectors were purchased from Promega (Madison, USA). ARID2, SUN1, E-cadherin, and ZNF76 antibodies were purchased from ABclonal, Inc. (Wuhan, China).

Cell Culture and miRNA Transfection
MCF-7 and HeLa cells were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China) and cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco-BRL, Carlsbad, USA) supplemented with glutamine, antibiotics, and 10% fetal bovine serum (Gibco-BRL, Carlsbad, USA) in a humidified atmosphere of 5% CO_2 at 37°C. Plasmid DNA or miRNA was transfected into cells using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

Western Blot
Cells were harvested 24, 48, or 72 h post-transfection. Then cells were lysed in radioimmunoprecipitation assay (RIPA) buffer [150-mM NaCl, 1-M Tris-HCl (pH 7.2), 1% (v/v) Triton X-100, 1% (w/v) sodium deoxycholate, 0.1% (w/v) sodium dodecyl sulfate (SDS)] with protease inhibitors. Proteins were separated on 10% or 15% SDS–polyacrylamide gel and transferred to poly(vinylidene difluoride) (PVDF) membranes. The resulting blots were blocked with 5% non-fat dry milk, and specific proteins were detected with appropriate antibodies. The proteins were detected using horseradish

FIGURE 1 | Strategy for screening microRNA binding sequences in the whole genomic regions.
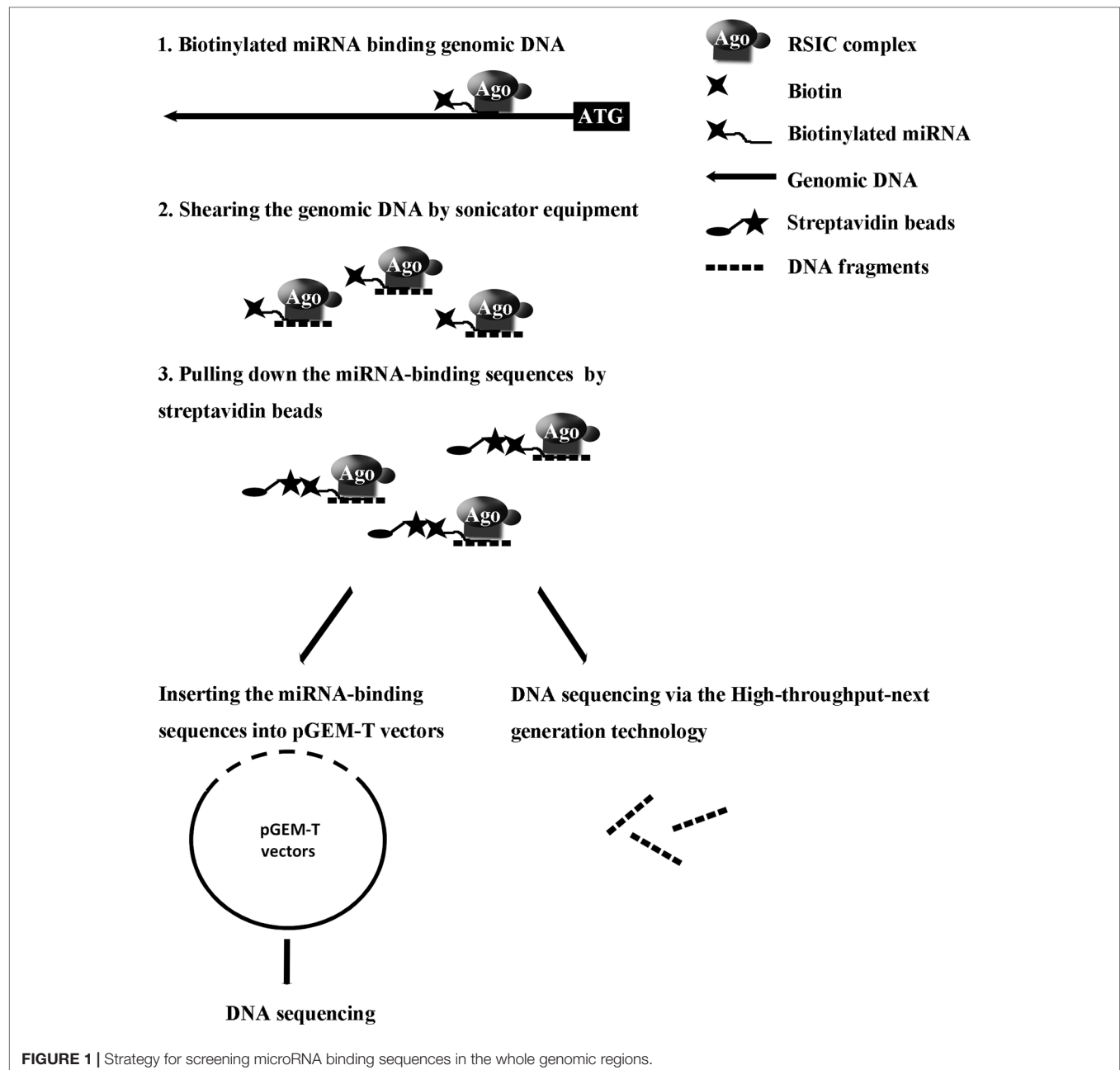
peroxidase (HRP)-conjugated secondary antibody and Super Signal West Pico Chemiluminescent substrate kits (Pierce).

DNA–miRNA Pull-Down Assay

The procedures used for affinity purification of biotinylated miRNAs were partly in reference to those previously described by Tidi Hassan and colleagues (Liu et al., 2005b). Cells were transfected by biotinylated miRNA or biotinylated NC miRNA for 24 h. Then cells were treated with 37% formaldehyde to a final concentration of 1% and incubated at room temperature for 15 min for cross-linking. The cross-linking reaction was stopped by the addition of 100-mM

glycine. Next, cells were collected; lysed in lysis buffer that contains 1% SDS, 1-mM EDTA, 50-mM HEPES (pH 7.5), 140-mM NaCl, and 1% Triton X-100; and supplemented with 100× protease inhibitor (Boehringer cocktail) and 1-U/μl RNase inhibitor (Invitrogen). The genomic DNA was sheared by sonicator equipment. This step should be performed on ice to avoid the denaturation of chromatin and miRNA. The supernatants were recovered by 12,000*g* centrifugation for 10 min and incubated with equilibrium streptavidin beads for 1 h at room temperature. Streptavidin beads were washed four times by washing buffer, which contains 10-mM Tris-HCl (pH 7.5), 1-mM EDTA, 0.15-mM LiCl, and 10-mM Tris-HCl. Proteinase K (Roche Applied Science) and RNase A (Roche Applied Science) were be used

TABLE 1 | miRNA sequences and real-time quantitative PCR primers.

Name		Sequence (5′–3′)
miR-373	Sense	ACUCAAAAUGGGGGCGCUUUCC
miR-373	Antisense	GAAGUGCUUCGAUUUUGGGGUGU-biotin
miRNA-NC	Sense	UCACAACCUCCUAGAAAGAGUAGA
miRNA-NC	Antisense	UCUACUCUUUCUAGGAGGUUGUGA-biotin
ARID2	Forward	ACACAGTGGTACCAGGACAG
ARID2	Reverse	TGAAGTTTGCACAGGTTGGG
PDE4D	Forward	GACCAATGTCTCAGATCAGTGG
PDE4D	Reverse	GTCAAGGGCCGGTTACCAG
E-cadherin	Forward	CCTGGGACTCCACCTACAGA
E-cadherin	Reverse	GGATGACACAGCGTGAGAGA
ZNF76	Forward	CAGGTGACGGTACAGAAAGAAGC
ZNF76	Reverse	TGATGAGCGGTGGTGTAGAGAC
SUN1	Forward	CAGAAGCACAAACAAATC
SUN1	Reverse	CACCATCATCATCAAGAC
ZNF385B	Forward	GGGCTAACCTGGAAACCGA
ZNF385B	Reverse	TCAGCTGACAGGAATTTGGACA
ALOX5	Forward	TCATCGTGGACTTTGAGCTG
ALOX5	Reverse	AGAAGGTGGGTGATGGTCTG
KIAA1958	Forward	TAGCCCTTTCTCCCTCA
KIAA1958	Reverse	TGCTCTTTCATCCTGGTC
CSDC2	Forward	GGCCCACCACATAAAATCTG
CSDC2	Reverse	ACCAACAAGCCCTCTCTCAA
β-Actin	Forward	CGTGGACATCCGCAAAGAC
β-Actin	Reverse	TCGTCATACTCCTGCTTGCTG
TTC34	Forward	TCAAAATCGACTCAGGGCAAC
TTC34	Reverse	CAGGGCTTTTTCCAGGTGG
TPM1	Forward	TTGAGAGTCGAGCCCAAAAG
TPM1	Reverse	CATATTTGCGGTCGGCATCTT
KIAA1377	Forward	GGGCACTGAATCATCGGACAA
KIAA1377	Reverse	TTTACGTGCTCGATTTCGACA
EVI5	Forward	AGAAACCCTAGTGGGAAACAGG
EVI5	Reverse	TGACTGTATGCGATACTGTGTTC
RPL37	Forward	TCGCAATAAGACGCACACGTT
RPL37	Reverse	CTCATTCGACCAGTTCCGGT
FANCC	Forward	CTGCCATATTCCGGGTTGTTG
FANCC	Reverse	AGCACTGCGTAAACACCTGAA
ITSN2	Forward	ATACGGTGGCGGCTTGAGTT
ITSN2	Reverse	GGAAGGTGGGAAGGAGGTTGA
Hsp60	Forward	GCCACGCGGGACTCACCATT
Hsp60	Reverse	CGCTCGGTTCCAGAACTTTCCA
GAPDH	Forward	CTCACCCTGCCCTCAATATCCC
GAPDH	Reverse	AGCCCTGTAGCCTGGACCTGAT

to degrade protein and RNA. Then the DNA was separated from streptavidin beads after treating beads at 80°C for 5 min. The eluted DNA was recovered using Chromatin Immunoprecipitation (ChIP) Kit (Millipore, USA) according to manufacturer's instructions.

Illumina HiSeq 2000 Next-Generation Sequencing and Bioinformatics Analysis

The PCR products were fragmented to an average length of 150 bp. After DNA-end repair, 3′dA overhang, and ligation of methylated sequencing have been performed, the DNA samples were sent to Beijing Genomics Institute (BGI, China) for sequencing by Illumina Genome Analyzer. Bioinformatics analysis steps for ChIP-Seq libraries are presented below. First, the original image data are transferred into sequence data *via* base calling, which is defined as raw data or raw reads, and saved as FASTQ file. Second, quality control was performed to detect whether the data are qualified. In

addition, filtering of raw data was used to decrease data noise. As a result, "dirty" raw reads which contain the sequence of adapter, more than 10% unknown bases, or low-quality bases have been removed in this step. Third, the clean reads were mapped to the *Homo sapiens* genome reference, and only the alignments within two mismatches and unique mapping reads were considered in further analyses. Then genome-wide peak scanning was performed in UCSC Genome Browser to get the information of peak location and peak sequence. Peaks were classified based on the location (UCSC annotation data) and showed in the following genome regions: intergenic, introns, downstream, upstream, and exons. Furthermore, after peak scanning, all the related genes related to miR-373 or NC RNA can be listed. Last, to predict potential functions of the putative miRNA targets in different cellular components, biological processes, and molecular functions, we used gene ontology (GO) categories (http://www.geneontology.org/) to classify the identified target genes. Besides, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (ftp://ftp.genome.jp/pub/kegg/pathway/) was applied for KEGG pathway analyses. We also submitted the HiSeq 2000 next-generation sequencing data to National Center for Biotechnology Information (NCBI).

Validation of miRNA Targets *via* qRT–PCR

Total RNA was isolated from cells that were transfected with synthetic miRNAs using TRIzol reagent (TaKaRa) according to the manufacturer's instructions. For quantification of mRNA, 1 μg of total RNA was reversely transcribed using the Reverse Transcription System (Promega, Madison, USA). The resulting cDNA was used as template for semi-quantitative PCR or quantitative real-time PCR. β-Actin served as an endogenous control used to normalized expression data. Each sample was analyzed in triplicate. Relative expression and standard error were calculated by the supplied ABI 7900HT Real-Time System software. All primers used in the qRT–PCR experiments are listed in **Table 1**.

Statistical Analysis

Data were expressed as means ± SD from three to four independent experiments. Data were analyzed using Student's t test for two groups or analysis of variance (ANOVA) with Tukey–Kramer tests for multiple group comparisons. $P < 0.05$ was considered statistically significant.

RESULTS

E-Cadherin Is Up-Regulated by Both miR-373 and Biotinylated miR-373 in MCF-7

Previous reports described that miR-373 can increase the expression of E-cadherin *via* targeting to its promoter in PC-3 cells (**Figure 2A**), while it has no impact on E-cadherin expression in HCT-116 and LNCaP cells (Place et al., 2008). To ensure whether miR-373 and biotinylated miR-373 could regulate E-cadherin in MCF-7 cells, biotinylated miR-373, non-biotinylated miR-373, or NC miRNA was transfected into MCF-7 cells for 48 h. Both semi-quantitative PCR and real-time PCR showed that E-cadherin increases over four times in the mRNA level after transfection with miR-373 or
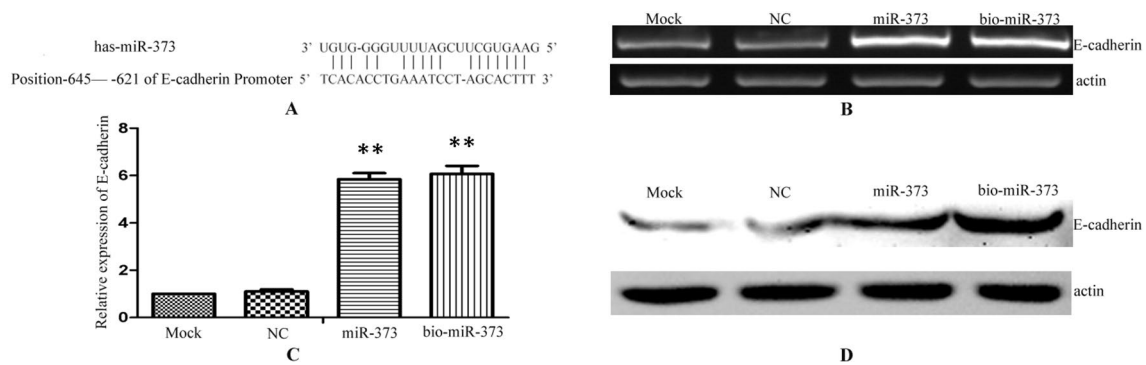
FIGURE 2 | Biotinylated miR-373 can regulate the expression of E-cadherin. **(A)** The identified miR-373 binding site in E-cadherin promoter. MCF-7 cells were transfected with miR-373, NC miRNA, or bio-miR-373 for 48 h. Then semi-quantitative PCR **(B)**, real-time PCR **(C)**, and Western blotting **(D)** were performed to detect the expression of E-cadherin. $n = 3$, **$P < 0.01$ compared with the sample transfected with NC miRNA.

biotinylated miR-373 than does NC miRNA (**Figures 2B**, **C**). We also confirmed that both miR-373 and biotinylated miR-373 can up-regulate E-cadherin protein levels in MCF-7 cells (**Figure 2D**). These data demonstrated that miR-373 increases the expression of E-cadherin in MCF-7 cells. Furthermore, being labeled with biotin at the 3′-end of miR-373 would not significantly affect the function of miR-373 to regulate E-cadherin.

The Promoters of E-Cadherin and CSDC2 Can be Pulled Down by Biotinylated miR-373 in MCF-7

After having confirmed that biotinylated miR-373 could increase the expression of E-cadherin in MCF-7, the miR-373 targeting sequence

in E-cadherin promoter was regarded as a positive control to test whether our method worked. Biotinylated miR-373 or biotinylated NC miRNA was transfected into MCF-7 cells. After 24 h post-transfection, miR-373 targeting sequences were isolated through DNA–miRNA pull-down assay described in the Materials and Methods section. Then semi-quantitative PCR and quantitative RT–PCR were performed to detect the enrichment of E-cadherin and CSDC2 promoters. As shown in **Figures 3A**, **B**, both E-cadherin and CSDC2 promoters can be amplified by semi-quantitative PCR from the sample transfected with biotinylated miR-373 rather than biotinylated NC miRNA. GAPDH promoter, which did not contain potential target site of miR-373, has no detectable signal when the sample transfected with biotinylated miR-373 was used. As shown in **Figure 3C**, the quantitative RT–PCR results were consistent
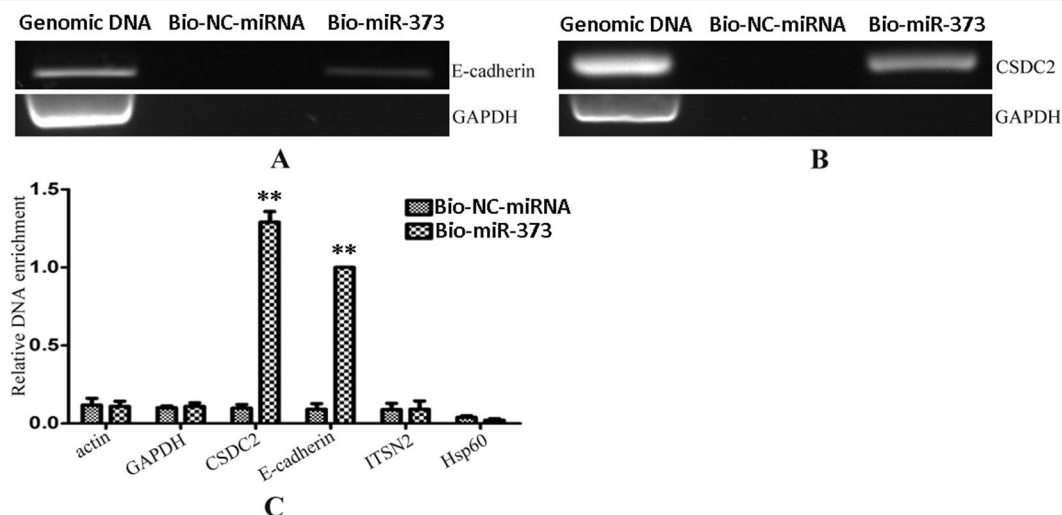
FIGURE 3 | Promoters of E-cadherin and CSDC2 can be pulled down by biotinylated miR-373. Bio-miR-373 or bio-NC-miRNA was transfected into MCF-7 cells for 24 h. Then cells were harvested, and the DNA fragments were enriched *via* DNA–miRNA pull-down assay. **(A)** The enrichments of E-cadherin promoter were measured by semi-quantitative PCR. **(B)** The enrichments of CSDC2 promoter were measured by semi-quantitative PCR. **(C)** Real-time PCR was performed to detect the enrichments of E-cadherin and CSDC2 promoters. U6 fragments added in the samples were taken as an endogenous control. $n = 3$, **$P < 0.01$ compared with the sample transfected with bio-NC-miRNA.

with those of semi-quantitative PCR. The amount of E-cadherin promoter and CSDC2 promoter in the sample transfected with biotinylated miR-373 was as over 10 times as the sample transfected with biotinylated NC miRNA, while the amount of promoters of actin, GAPDH, ITSN2, and Hsp60, which were regarded as negative control, is almost the same in the sample transfected with biotinylated miR-373, compared with the sample transfected with biotinylated NC miRNA. These results suggested that our method can be used to enrich miR-373 binding DNA sequences.

Identifying the Potential miR-373 Targeting DNA Sequences in Purified DNA Products

To identify unknown miR-373 targeting DNA sequences, we inserted the precipitated DNA into pGEM-T vectors (Promega, USA) and sequenced the vectors using the primer combining with T7 promoter. The specific procedure is shown in **Figure 1**. First, the Quick Blunting Kit (NEB, USA) has been used to convert DNA with incompatible 5′ or 3′ overhangs to blunt-ended DNA, because bio-miR-373-precipitated DNA fragments sheared by sonicator equipment contain fragments with incompatible 5′ or 3′ overhangs, which were hard to insert into pGEM-T vectors. Second, we tailed the blunt-ended DNA with an adenine in the 3′-end *via* using Ex Taq DNA Polymerase. Third, to improve the efficiency to insert DNA fragments into pGEM-T vectors, the A-tailing fragments have been condensed by 20% PEG-8000 and purified by 75% alcohol. Fourth, the purified DNA fragments have been inserted into pGEM-T vectors (Promega, USA). Fifth, the vectors were transformed into Top10 (Invitrogen, USA), and blue-white screening was used to select positive clones. Over 40 clones have been sequenced, and 10 sequences have been identified. Last, the sequences were analyzed *via* UCSC Genome Browser and NCBI Map Viewer. As shown in **Table 2**, six identified sequences are located in the introns of ZNF76, PDE4D, ALOX5, KIAA1959, and ZNF385B. Two sequences are in intergenic. Two sequences are in promoters of ARID2 and SUN1.

The Regulation of the Identified Genes by miR-373

The identified miR-373 targeting sequences are not only located in promoters but also located in introns. We wonder whether miR-373 could regulate these genes *via* directly binding with its promoter or intron. First, quantitative RT–PCR and Western blotting were performed to detect the regulation of the identified

genes by miR-373. As shown in **Figure 4A**, the mRNA levels were not observed to be significantly changed at 24 h post-transfection with miR-373. MiR-373 increases the mRNAs of ALOX, ARID2, CSDS2, KIAA1958, PDE4D, SUN1, and ZNF358B only 30% to 80% times at 48 h, while it up-regulates the mRNAs of E-cadherin and ZNF76 over four times at 48 h. All genes were up-regulated over two times in the mRNA level by miR-373 at 72 h, excepting ZNF358B. Then we purchased ARID2, SUN1, and ZNF76 antibodies to detect the expression of these genes at protein levels. As shown in **Figure 4B**, miR-373 can significantly increase these genes at 72 h. These data indicated that ARID2, SUN1, and ZNF76 can be obviously up-regulated by miR-373.

To investigate whether the regulation by miR-373 to the identified genes is a common phenomenon, RT–PCR was performed to measure the expression of the identified genes in HeLa cells. Interestingly, ALOX5, ARID2, KIAA1958, SUN-1, ZNF76, and ZNF385B can be up-regulated by miR-373, while CSDC2, E-cadherin, and PDE4D have not significantly increased after being transfected with miR-373 (**Figure 5A**). We also found that miR-373 cannot bind with the targeting sequence in CSDC2, E-cadherin, and PDE4D genes (**Figure 5B**). Some papers reported that some cell lines were resistant to specific miRNAs or dsRNAs-reduced transcriptional activation while sensitive to others. Our results provide an evidence that the direct interaction between miRNA and genomic sequences is key to miRNA-induced regulation of genes.

Screening miR-373 Binding Sequences *via* the High-Throughput Next-Generation Sequencing Technology

We have successfully established a method to identify unknown miRNA targeting DNA sites, but only 10 sequences have been identified in over 40 clones (data not shown). To improve efficiency to screen unknown miRNA target sequences, two genomic DNA fragment libraries were constructed and subjected to next-generation sequencing: one was constructed from miR-373-precipitated sample in MCF-7 and named HM-7-DNA, and the other was from NC RNA-precipitated sample and named HM-7-DNA-NC. As shown in **Supplementary Figure 1**, the main peak of HM-DNA sample was distributed at 360 bp and the main peak of HM-DNA-NC sample was distributed at 276 bp. So both of the samples were qualified and suitable for further sequencing. Then the quality control was used to analyze the quality of raw data obtained from Illumina HiSeq 2000 sequencing. As shown in **Supplementary Figures 1A, C**, both HM-7-DNA and HM-7-DNA-NC represented good-quality sequences, because the base ratios are mostly higher than 20. The raw data also had satisfactory base composition, for four bases of A, T, G, and C were distributed uniformly, and the AT content exceeded the GC content (**Supplementary Figures 2B, D**). The raw data have been submitted into SRA database, and the accession number is PRJNA547356.

The information of the peak location and sequence has been identified by genome-wide peak scanning in UCSC Genome Browser (**Supplementary Tables 1, 2**). Then we analyzed the distribution of the sequences from HM-7-DNA. As shown in **Figure 6A**, 49.7% of the sequences are located in intergenic, 25.5% of the sequences in intron, 11.7% of the sequences in promoter (Up2k),

TABLE 2 | miR-373 binding sequences identified by Sanger sequencing.

BLAT location	Located gene	GenBank accession no.	Location
Chr12: 46122709-46123286	ARID2	NC_000012.11	Promoter
Chr6: 35234579-35234951	ZNF76	NC_000006.11	Intron
Chr5: 59660052-59660485	PDE4D	NC_000005.9	Intron
Chr7: 814338-814698	SUN1	NC_000007.14	Promoter
Chr2: 180706031-180706314	ZNF385B	NC_000002.11	Intron
Chr10: 45883501-45883757	ALOX5	NC_000010.10	Intron
Chr9: 115377923-115378596	KIAA1958	NC_000009.11	Intron
Chr 21: 42657214-42657541	Not found	Not found	Intergenic
Chr 10: 2979740-2980027	Not found	Not found	Intergenic
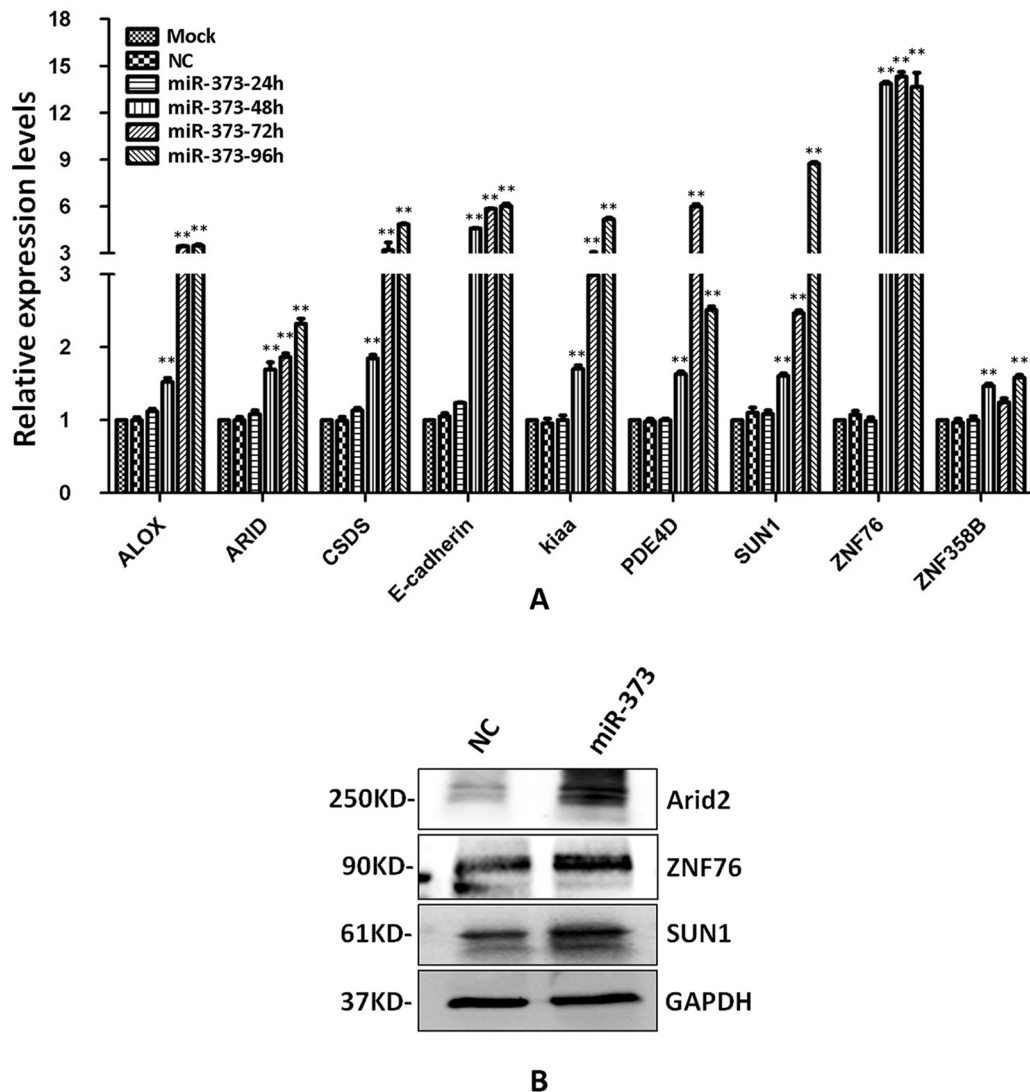
FIGURE 4 | The identified genes can be regulated by miR-373. **(A)** MCF-7 cells were transfected with miR-373 or NC-miRNA. Cells were harvested after 24, 28, 72, or 96 h post-transfection. Real-time PCR was used to measure the mRNA level of identified genes. The expression of β-actin was used as the internal reference. $n = 3$, **$P < 0.01$ compared with the sample transfected with NC miRNA. **(B)** MiR-373 or NC-miRNA was transfected into MCF-7 cell for 72 h. The protein levels of identified genes were detected by Western blotting.

10.3% of the sequences in exon, and 2.8% of the sequences in down2k. Meanwhile, we also analyzed the chromosomal location of miR-373 targeting sequences (**Figure 6B**). The results showed that the candidate targets of miR-373 were mainly distributed in 5th, 9th, 10th, and 20th chromosomes (**Figure 6B**).

MiRNAs can perform its biological functions *via* targeting genomic DNA and regulating gene expression, so pathway-based analysis of miR-373 targeting gene helps us to better understand the role of miR-373 in cells. On the one hand, GO analysis was performed to annotate the function of genes. **Figure 6C** shows the classification of the peak-related gene of HM-7-DNA based on the GO analysis. Biological process, cellular component, and molecular function, respectively, included 17, 6, and 5 categories. On the other hand, based on KEGG analysis, we found that miR-373 targeting genes were related with hypertrophic cardiomyopathy, dilated

cardiomyopathy, tight junction, cardiac muscle contraction, and viral myocarditis (**Table 3**).

We also compared the differences between HM-7-DNA and HM-7-DNA-NC. As shown in **Figure 6D**, 1,966 genes containing miR-373 targeting sequences have been found. Interestingly, 443 genes containing NC miRNA targeting sequences also have been identified. It cannot be denied that all designed NC miRNAs have the ability to combine with certain DNAs, so it is a possibility to have NC miRNA binding sites in genomic DNA. According to our results, there are 169 genes containing both miR-373 and NC miRNA target sites. These results suggest that the NC miRNA used in our paper is not suitable for studying the regulation of these 169 genes by miR-373, because NC miRNA also has a possibility to regulate these genes. Hence, to better study the genes identified by

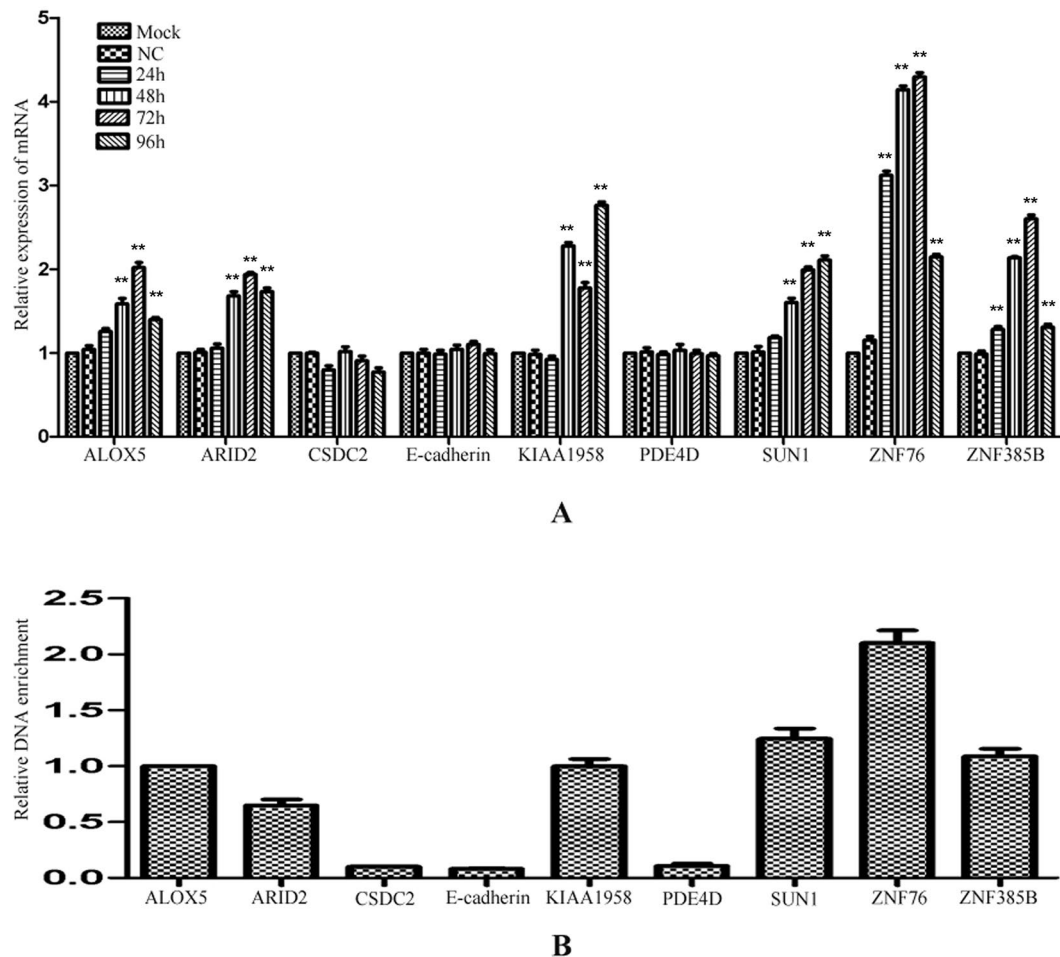**FIGURE 5 |** The regulation of identified genes by miR-373 in HeLa cells. **(A)** HeLa cells were transfected with miR-373 or NC-miRNA. Cells were harvested after 24, 48, 72, or 96 h post-transfection. Real-time PCR was used to measure the mRNA level of ALOX5, ARID2, E-cadherin, KIAA1958, PDE4D, SUN1, ZNF76, and ZNF385B genes. The expression of β-actin was used as the internal reference. $n = 3$, *$P < 0.05$, **$P < 0.01$ compared with the sample transfected with NC miRNA. **(B)** Bio-miR-373 or bio-NC-miRNA was transfected into HeLa cells. After 24 h post-transfection, DNA–miRNA pull-down assays were performed to enrich the miR-373 binding sequences. Then real-time PCR was performed to detect the enrichments of the genomic sequences in ALOX5, ARID2, E-cadherin, KIAA1958, PDE4D, SUN1, ZNF76, and ZNF385B genes. U6 fragments added in the samples were taken as an endogenous control.

our method, it is necessary to use biotinylated NC miRNA as negative control to prove that the studied genes have no potential NC miRNA binding sites.

So we selected six sequences from 1,827 genes that only contain miR-373 targeting sites to do further study. These sequences are located in exon or intron (**Table 4**). Semi-quantitative PCR has been performed to detect the enrichment of the six sequences in miR-373-precipitated DNA. As shown in **Figure 7A**, all sequences can be pulled down by miR-373. Then quantitative RT–PCR results demonstrated that the mRNAs of TTC34 and FANCC significantly increase after transfection by miR-373 for 48 and 72 h. TPM1, KIAA1377, EVI15, and RPL37 can be down-regulated by miR-373 after 24 h post-transfection, while these genes can be significantly up-regulated by miR-373 at 72 h (**Figure 7B**). We also randomly selected 35 potential miR-373 target genes and analyzed the changes in these gene expression after transfecting miR-373 for 48 h. As shown in **Table 5**, 21 gene expression changed more than two-fold

and only four gene expression changed less than quarter-fold after transfecting with miR-373 than did NC miRNA.

DISCUSSION

Although it has been proved that binding with promoters is an important way for miRNA to regulate gene expression, the mechanism of miRNA target recognition in genome should be further illuminated. Similar to the miRNA–mRNA interaction model, some papers suggested the "seed sequence" in miRNA is the key to binding with promoter (Xu et al., 2014; Xiao et al., 2017). However, Zhang et al. reported that let-7i can bind with TATA-box motifs in IL-2 promoter and the seed sequence of let-7i is not completely complementary with IL-2 promoter (Zhang et al., 2014). Another type of prediction tools, such as RNA hybrid, evaluates the interaction ability between miRNA and genomic sequences
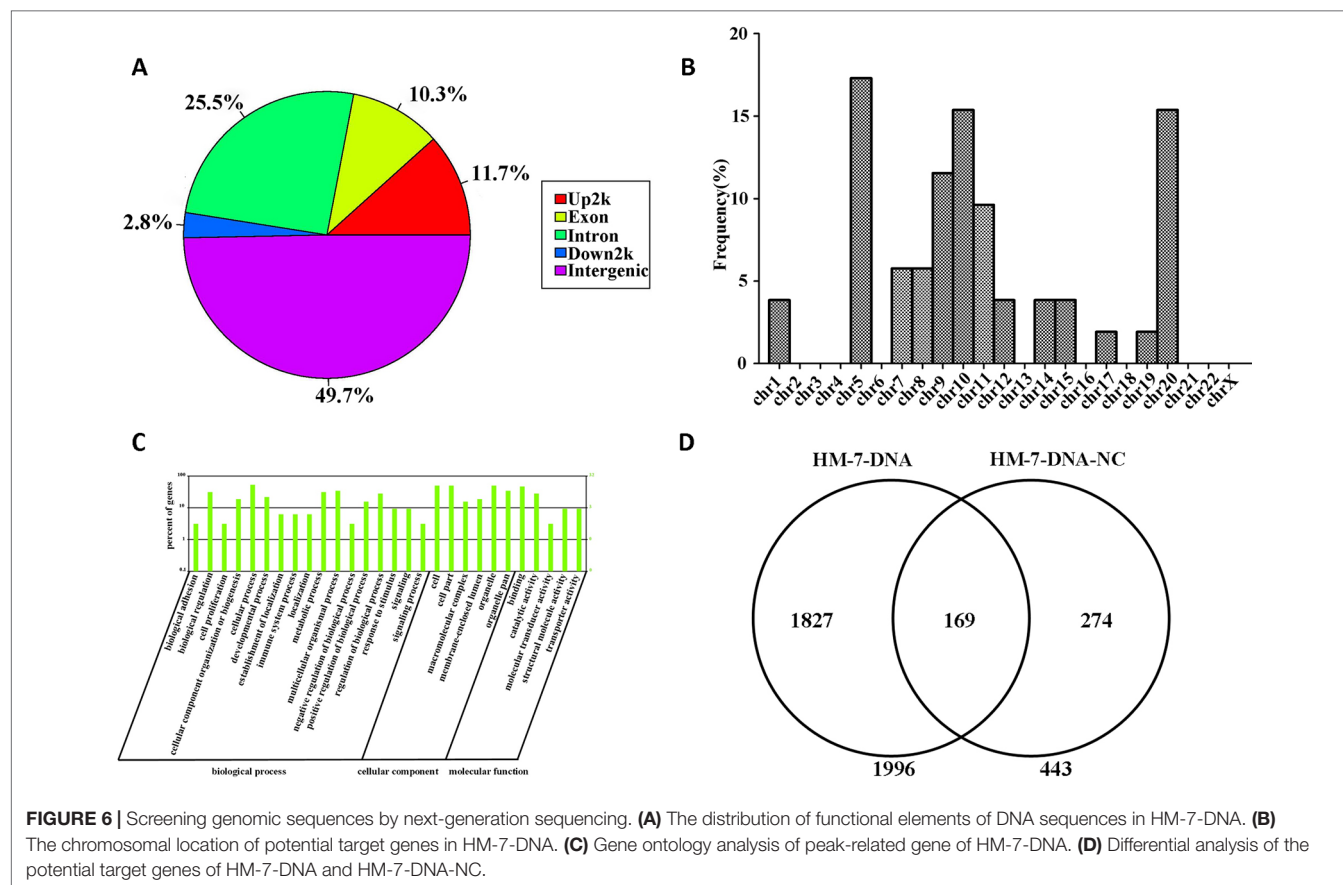
FIGURE 6 | Screening genomic sequences by next-generation sequencing. **(A)** The distribution of functional elements of DNA sequences in HM-7-DNA. **(B)** The chromosomal location of potential target genes in HM-7-DNA. **(C)** Gene ontology analysis of peak-related gene of HM-7-DNA. **(D)** Differential analysis of the potential target genes of HM-7-DNA and HM-7-DNA-NC.

TABLE 3 | The significant enrichment analysis of target gene's function in the pathway.

Pathway ID	Pathway	Q value
ko05410	Hypertrophic cardiomyopathy (HCM)	0.001193353
ko05414	Dilated cardiomyopathy	0.001193353
ko04530	Tight junction	0.001193353
ko04260	Cardiac muscle contraction	0.001361812
ko05416	Viral myocarditis	0.019827399

$Q \leq 0.05$ is a pathway that was significantly enriched in the peak-related gene.

TABLE 4 | MiR-373 binding sequences identified by next-generation sequencing.

BLAT location	Located gene	GenBank accession no.	Location
Chr1: 2585008-2585961	TTC34	NC_000001.10	Intron
Chr15: 63335003-63335190	TPM1	NC_000015.9	Exon
Chr11: 101785594-101785989	KIAA1377	NC_000011.9	Intron/exon
Chr1: 93143762-93143908	EVI5	NC_000001.10	Intron
Chr5: 40835156-40835380	RPL37	NC_000005.9	Intron/ promoter
Chr9: 97937920-97938060	FANCC	NC_000009.11	Intron

via measuring thermo-dynamic stability of miRNA and dsDNA or ssDNA (Rehmsmeier et al., 2004). Furthermore, Paugh et al. proved that miRNAs can form triplexes with dsDNA in genome

and regulate gene expression (Paugh et al., 2016). Because of lacking support by mechanism, computational prediction of miRNA targeting site in genome is in an initial step. In this paper, based on miRNA targeting–mRNA purification technique, which has been reported previously (Hassan et al., 2013), we have established an effective biochemical procedure to screen the potential miRNA targeting genes *via* pulling down the genomic sequences, which directly combined by miRNA. As a result, the putative target DNA sequences that were bound by biotinylated miRNAs can be easily isolated from cell extracts. These isolated DNA sequences can be analyzed through cloning and sequencing, and then the potential target genes may be found using the bioinformatics analysis. As described in this article, we successfully identified the known target genes of miR-373; moreover, we also detected unreported target genes. Therefore, we demonstrated that the target genes of miRNA complementary to DNA sequences can be efficiently obtained through our biochemical procedure directly from cultured cells.

Another limitation of bioinformatic prediction is that the prediction cannot reflect the real situation *in vivo*. When Li et al. have studied the regulation of gene expression by dsRNA *via* binding with promoter, they found that some specific dsRNAs can increase target gene expression in some cell lines, but not in others (Li et al., 2006). They also reported that E-cadherin expression was up-regulated by miR-373 in PC-3 and LNCaP cells, while not in HCT-116 cells (Place et al., 2008). Meanwhile, our results demonstrated that the expression of E-cadherin, CSCD2, and PDE4D, which can be
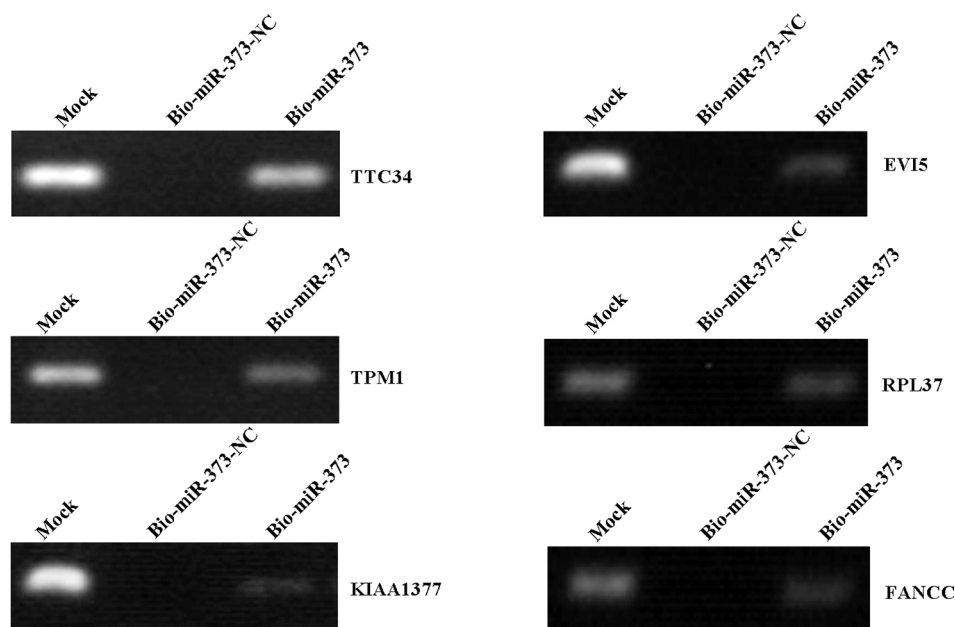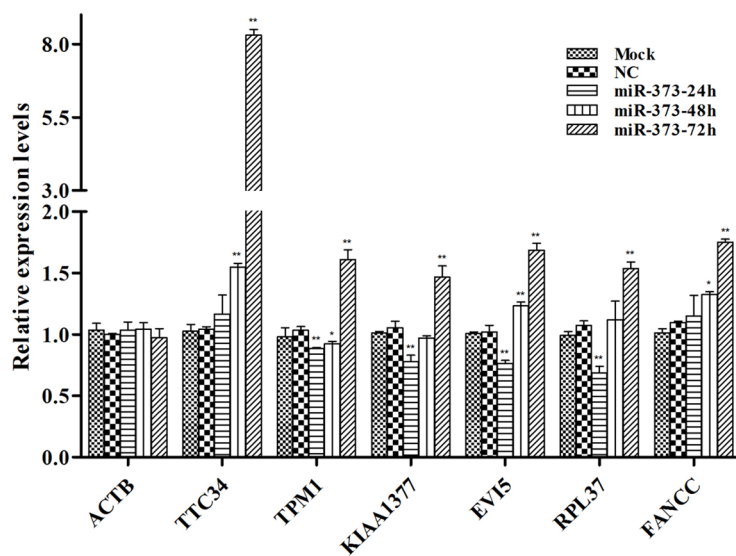
FIGURE 7 | The identified genes *via* next-generation sequencing can be regulated by miR-373. **(A)** Bio-miR-373 or bio-NC-miRNA was transfected into MCF-7 cells for 24 h. Then cells were harvested, and the DNA fragments were enriched *via* DNA–miRNA pull-down assay. The enrichments of TTC34, EVI5, TPM1, RPL37, KIAA1377, and FANCC promoters were measured by semi-quantitative PCR. **(B)** MCF-7 cells were transfected with miR-373 or NC-miRNA. Cells were harvested after 24, 48, or 72 h post-transfection. Real-time PCR was used to measure the mRNA level of identified genes. The expression of β-actin was used as the internal reference. $n = 3$, *$P < 0.05$, **$P < 0.01$ compared with the sample transfected with NC miRNA.

up-regulated by miR-373 in MCF-7 cells, was not significantly increased after transfecting miR-373 in HeLa cells. One of the reasons affecting miRNA-mediated gene activation is the epigenetic state of genome, for it is proved that the promoter of E-Cadherin is hypermethylated in HeLa cells, which prevented saRNA-induced

E-cadherin up-regulation (Li et al., 2006). Furthermore, our results demonstrated miR-373 can interact with the sequences located in E-cadherin, CSCD2, and PDE4D genes in MCF-7 cells but cannot in HeLa cells, which indicated that the direct interaction of miRNAs and targeting sequences is key to regulation of targeting genes. On

TABLE 5 | The identified genes affected by miR-373.

Gene name	Gene ID	Fold change
RPL37A	6168	−2.98
tnni3	7137	−5.34
KATNAL2	83473	−2.35
DOCK8	81704	1.85
COL23A1	91522	1.83
TNNT2	7139	4.03
GPR75	10936	−1.89
ASB3	51130	1.22
ATP5G3	518	−1.33
IVNS1ABP	10625	−4.81
WASH3P	374666	1.58
ARMC9	80210	−1.03
BAIAP2L1	55971	2.02
CRYAB	1410	2.73
NBAS	51594	1.61
GALNT9	50614	1.66
TMOD2	29767	−4.03
mapk6	5597	−3.86
Ybx1	4904	−1.67
rpl13	23521	−2.81
ralgps2	55103	−3.71
ITGB3BP	23421	−9.89
RHBDL2	54933	−3.31
PRMT8	56341	−2.38
COX11P1	140468	−6.28
DNAH14	127602	2.32
HEATR1	55127	−4.05
R3HCC1L	27291	−2.73
TMEM99	147184	1.08
ENKUR	219670	−5.13
WIPF1	7456	−1.44
SPTSSB	165679	−9.34
LGR4	55366	−2.28
LTBP1	4052	−1.33
LRCH3	84859	−1.22

the whole, our method can measure the direct interaction between miRNA and genomic DNA, which can avoid false positives caused by ignoring the modification of genome.

A very noteworthy finding in the present study is that some genomic fragments precipitated by miR-373 were located in intron. The roles of miRNAs in intron have not been widely studied. Meng et al. reported that some miRNAs binding sites are located in intron in plants (Meng et al., 2013). It has also been reported that siRNA targeting intronic sequences near alternative exons regulate splicing of mRNA and that Ago1 is essential for RNAi-mediated alternative splicing (Allo et al., 2009). It has been reported that some miRNAs and long noncoding RNAs are transcribed from the intron *via* sharing of the promoters with their host genes (Bosia et al., 2012; Kung et al., 2013; Chamorro-Jorganes et al., 2014; Ramalingam et al., 2014), so miRNAs (e.g., miR-373) targeting intron may play a role in regulating miRNAs and long noncoding RNAs, which are located in intron. Our results demonstrated that miR-373 can interact with the sequences located in intron. Then our results showed that ZNF76, PDE4D, ALOX5, KIAA1958, ZNF385B, TTC34, EVI5, and FANCC, which contain miR-373 binding sites in intron, can be regulated by miR-373. So there is existing interaction among miR-373 and intron sequences, which might affect gene expression. Taken together, the results suggested that the interaction of miRNAs and intron may

play some biological functions in cells, though we have not provided direct evidence that miRNAs regulate gene expression *via* binding with intron. The expression level of ZNF76 mRNA regulated by miR-373 was dramatic increased, but this regulated mechanism need to study in further study should focus on the mechanism of ZNF76 regulation by miR-373. We will investigate whether miR-373 binding sequence in ZNF76 intron is key to regulate expression of ZNF76 by knocking out the sequence.

Although we successfully identified several miR-373 binding sequences located in promoter, exon, or intron, there were some purified DNA fragments located in genomic DNA region far away from any known gene (over 80 kb, data not show). These DNA fragments may be located near the uncharacterized genes or may be a trigger for mediating the long-range regulation as described in a previous report (Zhao et al., 2009).The biological function of interaction between miRNAs and these target sites need to be further researched. However, our experimental procedure provides a way to find this kind of target sites and uncover new regulated mechanism of miRNAs.

Strangely, there is lack of connection between Sanger sequencing and next-generation sequencing results. Because Sanger sequencing and next-generation sequencing results were obtained from two independent experiments, many factors, including cell culture conditions, ChIP, and library construction, may contribute to variability between datasets. Another reason that may contribute to variability between two kinds of sequencing is that the samples have been prepared *via* different procedures. Preparing the samples for next-generation sequencing, compared with the samples for Sanger sequencing, have an extra step in that samples need to be amplified by PCR. In this step, some CG-rich sequences may be lost because the sequences are hard to be amplified by PCR.

In conclusion, this is a suitable method for identifying miRNA target genes that are complementary to genomic DNA. With the use of this method, the interaction of miRNA and putative miRNA targets can be confirmed by quantitative PCR with specific primers. So this method can be used to confirm the regulation mechanism of miRNAs to genes *via* binding genomic DNA. Furthermore, the experimental procedure can be applied to screen potential miRNA targets. On the whole, the method can improve the miRNA research enormously.

DATA AVAILABILITY

The raw data was submitted to NCBI SRA database and the accession number is PRJNA547356.

AUTHOR CONTRIBUTIONS

Designed the experiments: SX, KW. Performed the experiments: YX, KW, YT, LH, HX, SL, MG, CW. Wrote the paper: YX, KW, SX.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00778/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | Distribution of the main peak of genomic DNA. The genomic DNA was sheared by sonicator equipment. Then 1.5% agarose gel was used to measure the quality of samples. Results from **(A)** HM-DNA and **(B)** HM-DNA-NC are; both the main peaks of genomic DNAs were distributed between 100 and 500 bp. Agilent 2100 was used to detect specific distribution of sample fragments. HM-DNA fragment distributed at 360 bp **(C)**; HM-DNA-NC fragment distributed at 276 bp **(D)**.

SUPPLEMENTARY FIGURE 2 | Quality distribution and base distribution of HM-7-DNA and HM-DNA-NC. Quality distribution of **(A)** HM-7-DNA and **(C)** HM-7-DNA-NC are shown; the X-axis corresponds to the base site of the read. The Y-axis is quality value. Each dot in the image represents the quality value of the corresponding position along reads. Base distribution of **(B)** HM-7-DNA and **(D)** HM-7-DNA-NC is shown; both show a balanced base composition. The X-axis was the base position on the reads, and the Y-axis was the percentage of the corresponding base at each position. A, C, G, T, and N represent different bases.

SUPPLEMENTARY TABLE 1 | The peak information of HM-7-DNA.

SUPPLEMENTARY TABLE 2 | The peak information of HM-7-DNA-NC.

REFERENCES

Allo, M., Buggiano, V., Fededa, J. P., Petrillo, E., Schor, I., de la Mata, M., et al. (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.* 16 (7), 717–724. doi: 10.1038/nsmb.1620

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116 (2), 281–297.

Bosia, C., Osella, M., Baroudi, M. E., Cora, D., and Caselle, M. (2012). Gene autoregulation *via* intronic microRNAs and its functions. *BMC Syst. Biol.* 6, 131. doi: 10.1186/1752-0509-6-131

Carrington, J. C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science* 301 (5631), 336–338. doi: 10.1126/science.1085242

Chamorro-Jorganes, A., Araldi, E., Rotllan, N., Cirera-Salinas, D., and Suarez, Y. (2014). Autoregulation of glypican-1 by intronic microRNA-149 fine tunes the angiogenic response to FGF2 in human endothelial cells. *J. Cell Sci.* 127 (Pt 6), 1169–1178. doi: 10.1242/jcs.130518

Dong, H., Lei, J., Ding, L., Wen, Y., Ju, H., and Zhang, X. (2013). MicroRNA: function, detection, and bioanalysis. *Chem. Rev.* 113 (8), 6207–6233. doi: 10.1021/cr300362f

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 (2), 102–114. doi: 10.1038/nrg2290

Hassan, T., Smith, S. G., Gaughan, K., Oglesby, I. K., O'Neill, S., McElvaney, N. G., et al. (2013). Isolation and identification of cell-specific microRNAs targeting a messenger RNA using a biotinylated anti-sense oligonucleotide capture affinity technique. *Nucleic Acids Res.* 41 (6), e71. doi: 10.1093/nar/gks1466

He, L., and Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5 (7), 522–531. doi: 10.1038/nrg1379

Janowski, B. A., Younger, S. T., Hardy, D. B., Ram, R., Huffman, K. E., and Corey, D. R. (2007). Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat. Chem. Biol.* 3 (3), 166–173. doi: 10.1038/nchembio860

Kung, J. T., Colognori, D., and Lee, J. T. (2013). Long noncoding RNAs: past, present, and future. *Genetics* 193 (3), 651–669. doi: 10.1534/genetics.112.146704

Li, L. C., Okino, S. T., Zhao, H., Pookot, D., Place, R. F., Urakami, S., et al. (2006). Small dsRNAs induce transcriptional activation in human cells. *Proc. Natl. Acad. Sci. U. S. A.* 103 (46), 17337–17342. doi: 10.1073/pnas.0607015103

Liu, H., Lei, C., He, Q., Pan, Z., Xiao, D., and Tao, Y. (2018). Nuclear functions of mammalian microRNAs in gene regulation, immunity and cancer. *Mol. Cancer* 17 (1), 64. doi: 10.1186/s12943-018-0765-5

Liu, J., Valencia-Sanchez, M. A., Hannon, G. J., and Parker, R. (2005a). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat. Cell Biol.* 7 (7), 719–723. doi: 10.1038/ncb1274

Liu, J., Valencia-Sanchez, M. A., Hannon, G. J., and Parker, R. (2005b). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat. Cell Biol.* 7 (7), 719–723. doi: 10.1038/ncb1274

Meng, X., Jiang, Q., Chang, N., Wang, X., Liu, C., Xiong, J., et al. (2016). Small activating RNA binds to the genomic target site in a seed-region-dependent manner. *Nucleic Acids Res.* 44 (5), 2274–2282. doi: 10.1093/nar/gkw076

Meng, Y., Shao, C., Ma, X., and Wang, H. (2013). Introns targeted by plant microRNAs: a possible novel mechanism of gene regulation. *Rice (N Y)* 6 (1), 8. doi: 10.1186/1939-8433-6-8

Paugh, S. W., Coss, D. R., Bao, J., Laudermilk, L. T., Grace, C. R., Ferreira, A. M., et al. (2016). MicroRNAs form triplexes with double stranded dna at sequence-specific binding sites; a eukaryotic mechanism *via* which microRNAs could directly alter gene expression. *PLoS Comput. Biol.* 12 (2), e1004744. doi: 10.1371/journal.pcbi.1004744

Place, R. F., Li, L. C., Pookot, D., Noonan, E. J., and Dahiya, R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. U. S. A.* 105 (5), 1608–1613. doi: 10.1073/pnas.0707594105

Ramalingam, P., Palanichamy, J. K., Singh, A., Das, P., Bhagat, M., Kassab, M. A., et al. (2014). Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA* 20 (1), 76–87. doi: 10.1261/rna.041814.113

Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA* 10 (10), 1507–1517. doi: 10.1261/rna.5248604

Wei, K., Yan, F., Xiao, H., Yang, X., Xie, G., Xiao, Y., et al. (2014). Affinity purification of binding miRNAs for messenger RNA fused with a common tag. *Int. J. Mol. Sci.* 15 (8), 14753–14765. doi: 10.3390/ijms150814753

Xiao, M., Li, J., Li, W., Wang, Y., Wu, F., Xi, Y., et al. (2017). MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol.* 14 (10), 1326–1334. doi: 10.1080/15476286.2015.1112487

Xu, W., San Lucas, A., Wang, Z., and Liu, Y. (2014). Identifying microRNA targets in different gene regions. *BMC Bioinform.* 15 Suppl 7, S4. doi: 10.1186/1471-2105-15-S7-S4

Zhang, Y., Fan, M., Zhang, X., Huang, F., Wu, K., Zhang, J., et al. (2014). Cellular microRNAs up-regulate transcription *via* interaction with promoter TATA-box motifs. *RNA* 20 (12), 1878–1889. doi: 10.1261/rna.045633.114

Zhao, J., Ding, J., Li, Y., Ren, K., Sha, J., Zhu, M., et al. (2009). HnRNP U mediates the long-range regulation of Shh expression during limb development. *Hum. Mol. Genet.* 18 (16), 3090–3097. doi: 10.1093/hmg/ddp250

Non-Coding RNAs in Pediatric Solid Tumors

Christopher M. Smith[1], Daniel Catchpoole[2,3]* and Gyorgy Hutvagner[1]*

[1] School of Biomedical Engineering, University of Technology Sydney, Sydney, Australia [2] School of Software, University of Technology Sydney, Sydney, Australia [3] The Tumour Bank–CCRU, Kids Research, The Children's Hospital at Westmead, Sydney, Australia

Pediatric solid tumors are a diverse group of extracranial solid tumors representing approximately 40% of childhood cancers. Pediatric solid tumors are believed to arise as a result of disruptions in the developmental process of precursor cells which lead them to accumulate cancerous phenotypes. In contrast to many adult tumors, pediatric tumors typically feature a low number of genetic mutations in protein-coding genes which could explain the emergence of these phenotypes. It is likely that oncogenesis occurs after a failure at many different levels of regulation. Non-coding RNAs (ncRNAs) comprise a group of functional RNA molecules that lack protein coding potential but are essential in the regulation and maintenance of many epigenetic and post-translational mechanisms. Indeed, research has accumulated a large body of evidence implicating many ncRNAs in the regulation of well-established oncogenic networks. In this review we cover a range of extracranial solid tumors which represent some of the rarer and enigmatic childhood cancers known. We focus on two major classes of ncRNAs, microRNAs and long non-coding RNAs, which are likely to play a key role in the development of these cancers and emphasize their functional contributions and molecular interactions during tumor formation.

Keywords: pediatric tumors, miRNA, long noncoding RNA, cancer biology, gene expression

Pediatric cancers are often categorized as hematologic, intracranial, or extracranial (Chen et al., 2015). Hematologic cancers include those derived from the blood or blood forming tissues, including bone marrow and the lymph nodes. Intracranial cancers are tumors that develop inside the brain, whereas extracranial solid tumors, often referred to as pediatric solid tumors, arise outside the brain. Collectively, pediatric solid tumors represent approximately 40% of all pediatric cancers and commonly form in the developing sympathetic nervous system (neuroblastoma), retina (retinoblastoma), kidneys (Wilms tumor), liver (hepatoblastoma), bones (osteosarcoma, Ewing sarcoma), or muscles (rhabdomyosarcoma) (Kline and Sevier, 2003; Allen-Rhoades et al., 2018). Solid tumors can originate from cells of any of the three germ layers, the ectoderm, mesoderm, or endoderm, and likely arise due to disruptions in the developmental processes of these precursor cells, leading them to develop cancerous phenotypes (Chen et al., 2015). This contrasts with most adult cancers, which tend to be of epithelial origin and are believed to develop over time due to exposure to toxins and environmental stress. As a result, adult cancers often display a high occurrence of genetic mutations, whereas pediatric solid tumors tend to feature a relatively low number of genetic mutations. This has led to investigations into alternative forms of gene regulation that may contribute to the emergence and development of cancerous cells in pediatric cancers.

Non-coding RNAs (ncRNAs) form a group of functional RNAs lacking protein-coding potential, which play a crucial role in the regulation of gene expression at every level, from epigenetic regulation via methylation and chromatin packaging to post-transcriptional regulation (Cech and Steitz, 2014; Zhao et al., 2016). The most widely studied ncRNAs are the microRNAs (miRNAs), small 20- to 25-nucleotide-long RNAs that play an important role in regulating translation and messenger RNA stability via complementary base pairing (Huang et al., 2011). Other classes of small ncRNAs include small interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), small transfer RNAs (tRFs), small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), and small cytoplasmic RNAs (scRNAs). Additionally, long non-coding RNAs (lncRNAs) are a loosely defined group of RNAs normally larger than 200 nucleotides long that lack protein-coding potential and do not fall into any of the other categories but nonetheless play key roles in the regulation of gene expression (Mercer et al., 2009). Increasing data suggest that ncRNAs play a role in regulating all biological processes, and it is no surprise that studies have observed widespread dysregulation of ncRNAs in nearly all forms of cancer (Prensner and Chinnaiyan, 2011; Leichter et al., 2017). Interestingly, dysregulated RNA patterns are often specific to the type of cancer or even subtype and can provide insight into the mechanisms underlying phenotypic differences between tumors or cells within a tumor, such as their aggressiveness or resistance to certain types of treatments (Blenkiron et al., 2007; Li et al., 2014). Additionally, genome-wide association studies have suggested that over 80% of single nucleotide polymorphisms found associated with cancer are outside of coding regions (Carninci et al., 2005; Cheetham et al., 2013). In this review, we will discuss how two major classes of ncRNAs, miRNAs and lncRNAs, may contribute to pediatric solid tumors by participating in the regulation of established oncogenic networks known to drive these cancers.

MICRORNAS AND GENE REGULATION

Not long after the first human miRNA, *let-7*, was discovered in 2002 by the Ruvkun lab, miRNAs began to emerge as key participants in tumorigenesis (Pasquinelli et al., 2000). In 2002, two miRNAs were identified as potential tumor suppressors due to their frequent downregulation or deletion in chronic lymphocytic leukemia (Calin et al., 2002). Calin et al. (2004) later showed that many miRNA genes are located close to fragile sites or common breakpoints that frequently occur in cancers, suggesting that their loss of function was a key event in oncogenesis. Since then, oncomiRs—cancer-associated miRNAs—have become a major research focus (Esquela-Kerscher and Slack, 2006). A better understanding of the mechanisms behind miRNA regulation in cancer is invaluable to researchers and clinicians alike, not only to aid in the identification of new drug targets but also for the development of promising RNA-based therapies and their potential use as early detection biomarkers.

miRNAs: Biogenesis and Functions

The life cycle of a miRNA typically begins with its transcription into a primary miRNA (pri-miRNA) by RNA polymerase II (Ha and Kim, 2014). pri-miRNAs share several similarities with messenger RNAs (mRNAs) in that they are 5' capped, are 3' polyadenylated, and can be several hundreds or thousands of nucleotides long. In many cases, the pri-miRNA encodes for one miRNA species; however, in humans, a substantial number are polycistronic and encode several different miRNAs together. pri-miRNAs must be processed in the nucleus by the RNAse III enzyme Drosha, which releases shorter ~65-nucleotide-long precursor RNAs (pre-miRNAs) with a secondary hairpin structure. This hairpin is recognized by the Exportin-5/Ran-GTP transporter, which transports the pre-miRNA from the nucleus to the cytoplasm. In the cytoplasm, the pre-miRNA is further processed by Dicer, another RNAse III enzyme, which cleaves the loop and releases a double-stranded miRNA duplex containing the 5 prime (5p) and 3 prime (3p) sequences. The duplex is then recognized by one of the four human Argonaute proteins, which loads one of the strands and discards the other.

miRNAs carry out their functions by binding to Argonaute and associating with various other proteins to form the RNA induced silencing complex (RISC). As part of this complex, miRNAs serve as guides by binding via complementary base pairing to target sites that are normally found in the 3'-untranslated region (3'UTR) of mRNAs. RISCs can regulate gene expression by direct cleavage of transcripts, transcript destabilization, or blocking translation. In a broader sense, miRNAs play a role in globally "fine-tuning" gene expression and are particularly important in inducing and maintaining differentiated cell states. In cancer, this finely tuned expression is often impaired, enabling gene networks that are normally switched on or off to reverse and begin influencing cellular behavior in a deleterious manner.

miRNAs: Drivers or Passengers in Cancer?

Microarrays and next-generation sequencing technologies enabled global measurements of miRNA expression changes and have revealed miRNA dysregulation to be a hallmark in nearly all cancers. miRNA expression profiles often correlate with cancer subtypes and have been effective at classifying cancer samples for risk stratification (De Preter et al., 2011). However, understanding the contribution of specific miRNAs can prove difficult. miRNAs are predicted to regulate hundreds to thousands of genes; however, their influence may be minor, and often, they must act in concert with other miRNAs. Current miRNA target prediction algorithms are imperfect and do not capture the true range of regulatory targets; therefore, biological validation is still needed (Riffo-Campos et al., 2016). Additionally, opposing behavior is seen with many miRNAs, where the same miRNA may be considered an oncogene in one cancer and a tumor suppressor in another. Because of their integration within complex gene networks, it is often not obvious whether a dysregulated miRNA actively participates in the maintenance of a cancerous state or whether it is simply a
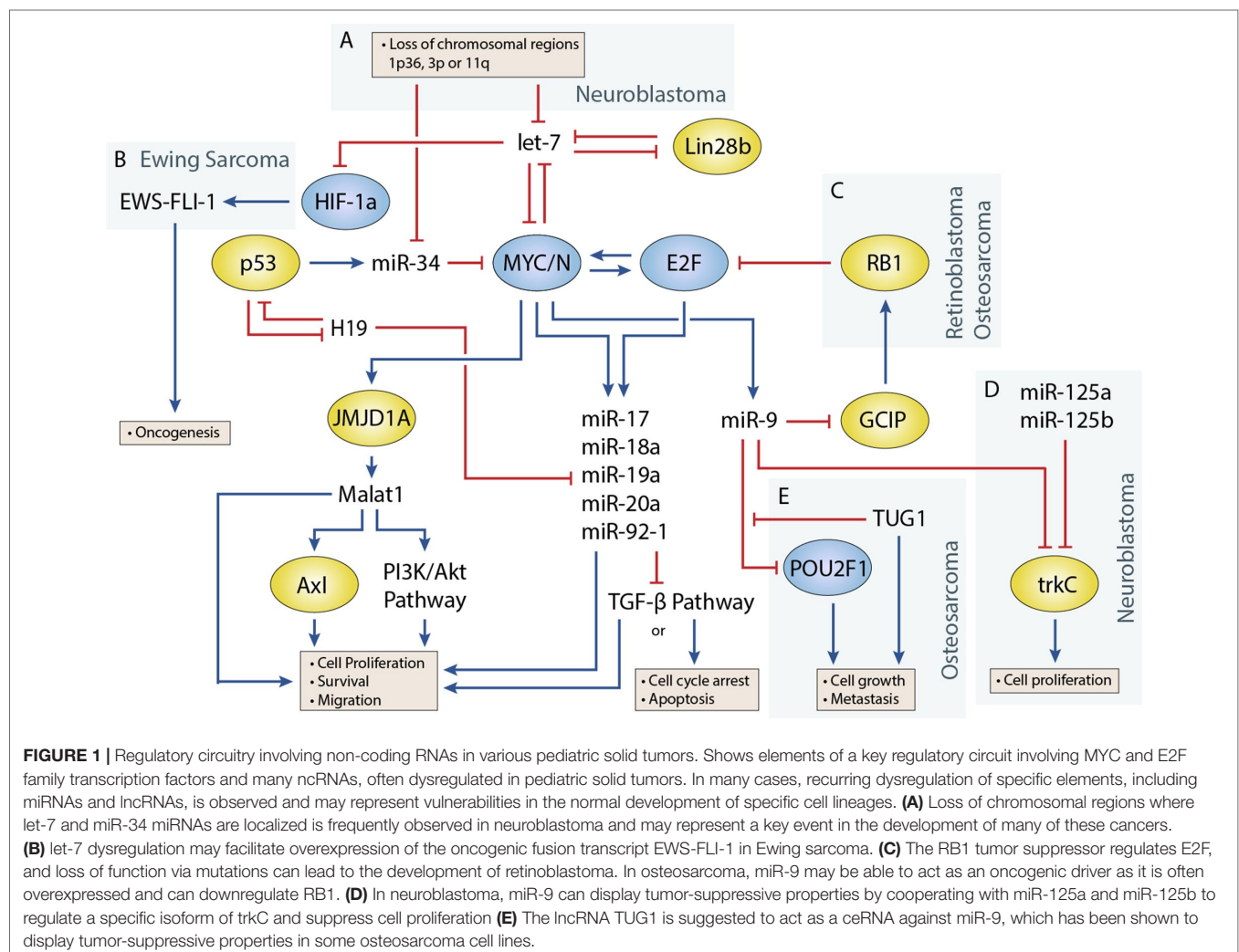
bystander. Therefore, it is important to examine how miRNAs participate in oncogenic networks on a functional level in order to properly understand their role.

Transcription factors that play an important role in regulating cell proliferation, migration, and apoptosis are commonly perturbed in pediatric solid tumors. One of the best examples of this is in neuroblastoma, where *MYCN* amplification is present in approximately 25% of neuroblastoma patients and disproportionally represents high-risk cases (Huang and Weiss, 2013). MYCN upregulation is also observed at a higher frequency in several other pediatric solid tumors including Wilms tumor, rhabdomyosarcoma (Williamson et al., 2005), and retinoblastoma, although generally not to the extent seen in neuroblastoma. Germline inactivation of the Wilms Tumor 1 (WT1) transcription factor has been linked to a genetic predisposition towards Wilms tumor. Several transcription factors, including Twist, Snails, and Zebs, involved in the epithelial-to-mesenchymal transition have also been implicated in the development of osteosarcoma (Yang et al., 2013). miRNAs are often closely tied to transcription factors, either as regulators or as transcriptional targets (**Figure 1**) (Sin-Chan et al.,

2019). One of the earliest studies linking miRNAs to an oncogenic transcription factor was by O'Donnell et al. in 2005 (O'Donnell et al., 2005). In this study, they demonstrated that c-Myc could induce expression of the miR-17~92 cluster and that several of these miRNAs could in turn regulate E2F1 transcription to control cell proliferation.

Disruptions in miRNA Processing
Recent studies have shown that impairments of the miRNA processing machinery are common in Wilms tumor and likely contribute to this disease. For example, a study by Torrezan et al. (2014) found mutations in miRNA processing genes in 33% of tumors, most commonly occurring in the *Drosha* gene, with other mutations in *DICER1*, *XPO5*, *DGCR8* and *TARBP2*. These results are supported by several other studies by Wu et al. (2013), Rakheja et al. (2014), Walz et al. (2015), Wegert et al. (2015), and Gadd et al. (2017). In Rakheja et al.'s study, they further examined the potential consequences of several of these mutations and found that Drosha mutations often led to a loss of RNAse IIIB activity, which prevented processing of

FIGURE 1 | Regulatory circuitry involving non-coding RNAs in various pediatric solid tumors. Shows elements of a key regulatory circuit involving MYC and E2F family transcription factors and many ncRNAs, often dysregulated in pediatric solid tumors. In many cases, recurring dysregulation of specific elements, including miRNAs and lncRNAs, is observed and may represent vulnerabilities in the normal development of specific cell lineages. **(A)** Loss of chromosomal regions where let-7 and miR-34 miRNAs are localized is frequently observed in neuroblastoma and may represent a key event in the development of many of these cancers. **(B)** let-7 dysregulation may facilitate overexpression of the oncogenic fusion transcript EWS-FLI-1 in Ewing sarcoma. **(C)** The RB1 tumor suppressor regulates E2F, and loss of function via mutations can lead to the development of retinoblastoma. In osteosarcoma, miR-9 may be able to act as an oncogenic driver as it is often overexpressed and can downregulate RB1. **(D)** In neuroblastoma, miR-9 can display tumor-suppressive properties by cooperating with miR-125a and miR-125b to regulate a specific isoform of trkC and suppress cell proliferation **(E)** The lncRNA TUG1 is suggested to act as a ceRNA against miR-9, which has been shown to display tumor-suppressive properties in some osteosarcoma cell lines.

pri-miRNAs, leading to a global reduction in mature miRNAs. *DICER1* mutations also frequently affected the RNAse IIIB domain; however, this mutation only affected processing of 5p miRNAs from precursors, as *DICER1* contains a second RNAse domain for 3p processing. As a result, this mutation led to a shift towards 3p miRNA maturation. These mutations have interesting consequences for global miRNA expression and most likely favor expression of oncogenic miRNAs or reduce expression of miRNAs with tumor-suppressive effects. In line with this, the let-7 family is predominantly 5p-derived, and lower expression of several of its 5p members was found in both *Drosha* and *DICER1* mutants in two of these studies (Rakheja et al., 2014; Walz et al., 2015). Additionally, the miR-200 family was found downregulated in Wilms tumors with mutated miRNA processing genes, which is known to regulate the mesenchymal-to-epithelial transition and has been associated with highly aggressive forms of cancer (Ceppi and Peter, 2014; Walz et al., 2015). The functional role of several oncomiRs has been investigated in detail within the context of pediatric solid tumors and is discussed in the following section.

The miR-17~92 Cluster is a Downstream Effector of Oncogenic Transcription Factors

The miR-17~92 cluster is expressed during normal development of the brain, heart, lungs, and immune system (Koralov et al., 2008; Ventura et al., 2008; Bian et al., 2013; Chen et al., 2013) and is known to regulate critical genes involved in cell growth, proliferation, and apoptosis. This cluster is comprised of six different miRNAs that are co-expressed, including miR-17, miR-18a, miR-19a, miR-19b-1, miR-20a, and miR-92-1. Dysregulation of the miR-17~92 cluster has been shown in several pediatric solid tumors including neuroblastoma, Wilms tumor, retinoblastoma, and osteosarcoma, where a higher expression generally correlates with a poorer prognosis (Chen and Stallings, 2007; Baumhoer et al., 2012; Li et al., 2014). The miR-17~92 cluster is particularly interesting due to its regulation by the transcription factor MYC and its homologue MYCN, where it seems to act as a mediator for some of MYC/MYCN's oncogenic effects (Schulte et al., 2008). Other transcription factors known to target the miR-17~92 cluster include members of the E2F family and STAT3 (Mogilyansky and Rigoutsos, 2013).

Several studies have demonstrated that the miR-17~92 cluster regulates many downstream components of the transforming growth factor beta (TGF-β) pathway, which is known to participate in a variety of cellular process such as differentiation, proliferation, and immune cell activation. A study by Fontana et al. (2008) demonstrated that in neuroblastoma, miR-17 and miR-20a downregulate the cyclin-dependent kinase inhibitor p21, which is activated by TGF-β. p21 plays a key role in the inhibition of cell cycle progression by blocking the transition from G1 to S phase, and its deregulation leads to uncontrolled cell growth. Additionally, Fontana et al. (2008) showed that miR-17-5p regulated another downstream component of TGF-β, the pro-apoptotic factor Bcl-2 interacting mediator (BIM). Mestdagh et al. (2010) later investigated miR-17~92 regulation of the TGF-β pathway in more depth and identified miR-17 and miR-20a as regulators of TGFBR2 and miR-18a as a regulator of SMAD2 and SMAD4, both signal transducers for TGF-β

receptors. miR-18a and miR-19a have also been shown to repress estrogen receptor (ESR1) expression, and prolonged knockdown of miR-18a induced morphological differentiation of SK-N-BE neuroblastoma cells. Interestingly, the TGF-β pathway interacts with ESR1 signaling via several of the SMADs (Band and Laiho, 2011), suggesting a complex interplay between miR-17~92 and its targeted pathways necessary for fine-tuning differentiation during neuronal development—a balance that is disrupted when miR-17~92 is overexpressed. While no studies have investigated in detail the interaction between the miR-17~92 cluster and TGF-β pathway in Wilms tumor, the TGF-β pathway has been implicated in Wilms tumor development. In contrast with neuroblastoma, the TGF-β pathway appears to function as a promoter of Wilms tumor progression, and TGF-β is highly expressed in primary tumors, even more so in metastatic tumors. This multifaceted behavior of the TGF-β pathway has been shown in other cancers and implies that the pathway's influence is specific to the tumor it is activated in.

The E2F family of transcription factors serve an important role in cell cycle control as their expression can cause cells to enter the G1 phase to initiate cell division (Chen et al., 2009). Several members, including E2F1, E2F2, and E2F3, all regulate miR-17~92 expression. In a study by Kort et al. (2008) a member of the E2F family of transcription factors, E2F3, was shown to be exclusively expressed in Wilms tumor and not in other types of kidney tumors. In line with this, they compared expression of the miR-17~92 miRNAs in Wilms tumor samples to other renal tumor subtypes and found them all to be upregulated. They were also able to show a correlation between E2F3 expression and the stage of Wilms tumor, where it was highest in late-stage metastatic tissues. In retinoblastoma, an early study investigating the miR-17~92 cluster identified that one of its members, miR-20a, participates in an autoregulatory feedback loop with E2F2 and E2F3 (Sylvestre et al., 2007), as they found both transcription factors are themselves downregulated by miR-20a. The authors suggested that this autoregulation was critical in preventing expression of excessive amounts of E2F transcription factors. Given that MYC/MYCN and E2F have previously been shown to induce each other's expression, miR-20a appears to play an important role in keeping this positive feedback loop in check (Leone et al., 1997; Strieder and Lutz, 2003). Therefore, it is easy to see how disruption in one or more of these regulatory elements could lead to uncontrolled expression of these proliferative and anti-apoptotic signals.

A later study by Conkrite et al. (2011) investigated miR-17~92 in retinoblastoma and revealed that this cluster was capable of driving retinoblastoma formation in *RB1/p107*-deficient mice. *RB1* plays a key role in inhibiting cell cycle progression, and germline mutations of this gene can lead to familial retinoblastoma formation (Friend et al., 1986; Classon and Harlow, 2002). *RB1*'s protein product, pRB, inhibits E2F transcription factors by binding and inactivating them, and its absence enables miR-17~92–driven tumor formation.

The miR-17~92 cluster also plays a role in driving tumor progression and metastasis in osteosarcoma (Li et al., 2014). A recent study by Yang et al. (2018) identified QKI2 as a regulatory target of the miR-17~92 cluster. QKI proteins have previously

been shown to inhibit β-catenin and induce differentiation in colon cancer. Yang at el. demonstrated that miR-17~92 downregulated QKI2, causing upregulation of β-catenin, leading to increased proliferation, invasion, and migration in osteosarcoma (Yang et al., 2018). Additionally, miR-20a has previously been shown to downregulate Fas expression, which is a cell surface marker that interacts with FasL to induce apoptosis in the lungs, where osteosarcoma almost exclusively metastasizes to (Huang et al., 2012).

The miR-17~92 cluster plays a tumorigenic role in a number of pediatric solid tumors including neuroblastoma, Wilms tumor, retinoblastoma, and osteosarcoma. The use of the miRNA pathway by transcription factors such as the MYC and E2F families enables them to target a wide range of genes and immediately effect gene expression at the post-transcriptional level. Continued research into how miRNAs may operate as oncogenic drivers will likely expand the repertoire of potential drug targets available to us.

Let-7 Dysregulation is a Feature in Many Pediatric Solid Tumors

The let-7 family of miRNAs are among the most well-characterized tumor suppressors due to their frequent downregulation in cancers. In total, there are 12 members of the let-7 family located across eight different chromosomes; however, in most cells, only a selection of these miRNAs will be expressed (Balzeau et al., 2017). let-7 miRNAs are important in regulating the cell cycle and maintaining cells' differentiated state by targeting a wide range of genes with known roles in cancer biogenesis such as *MYC/MYCN*, *RAS*, *CDK6*, and *HMGA2* (Buechner et al., 2011; Wu et al., 2015).

let-7 is regulated by the LIN28 proteins, LIN28A and LIN28B, which mediate uridylation, prevent processing of the let-7 precursor, and are important for maintaining pluripotency in cells (Lehrbach et al., 2009; Balzeau et al., 2017). Both *Lin28* genes contain let-7 target sites and participate in a double-negative feedback loop with let-7 (Yin et al., 2017). Overexpression of *Lin28* tends to drive cells towards oncogenesis and is a common feature in cancers. In a study by Urbach et al. (2014), *Lin28b* overexpression was found in approximately 30% of Wilms tumors. Additionally, they found overexpression of *Lin28* could induce tumor formation in specific renal intermediates and that restoration of let-7 activity could reverse this effect in mice. Similar examples have been shown in mouse models, where *Lin28b* overexpression can drive hepatoblastoma and hepatocellular carcinoma in the liver and neuroblastoma in the neural crest (Molenaar et al., 2012; Nguyen et al., 2014). Molenaar et al. (2012) investigated Lin28b in neuroblastoma and demonstrated that Lin28b could enhance MYCN protein levels via let-7 regulation. However, a later study by Powers et al. (2016) showed that Lin28b expression was redundant in certain MYCN-amplified neuroblastoma cells, as overexpression of the MYCN transcript could function as a miRNA sponge for let-7, thereby negating their effect regardless of expression level. Powers et al. showed that most neuroblastomas were characterized by a loss of let-7 with either MYCN overexpression or chromosomal loss of arm 3p or 11q, where several let-7 miRNAs are located

(**Figure 1A**). The authors noted that these events were generally mutually exclusive and suggested that the presence of one event alleviated selective pressure for the other.

A study by Di Fiore et al. (2016) revealed that let-7d could promote and suppress tumor formation within the same system. In this study, they found that let-7d overexpression in osteosarcoma cells reduced several stemness genes, including *Lin28b*, *HMGA2*, *Oct3/4*, and *SOX2*, and could elicit the mesenchymal-to-epithelial transition with upregulation of the epithelial marker E-cadherin and downregulation of mesenchymal markers N-cadherin and vimentin. However, they also found that let-7d enhanced cell migration and invasion, presumably by acting via the TGF-β pathway, which is known to promote this behavior. let-7d strongly increased versican VI expression, which has previously been shown to activate the TGF-β pathway in osteosarcoma (Li S. et al., 2014).

In Ewing sarcoma, Hameiri-Grossman et al. (2015) found that let-7 downregulated the *Ras* oncogene, as well as the transcription factor HIF-1a, to reduce EWS-FLI-1 expression (**Figure 1B**). EWS-FLI-1 is a hybrid transcript that results from a translocation event involving EWS and FLI1, and translocation events such as this are present in nearly all Ewing sarcoma cases and are believed to drive the disease (Delattre et al., 1994).

Loss of Let-7 plays a key role in many pediatric solid tumors as its loss enables expression of transcription factors and other genes that participate in oncogenesis. This has been emphasized in neuroblastoma, where it has been suggested that loss of let-7 function is an essential event in tumor development and positions the miRNA pathway as a central player in pediatric solid tumors.

miR-9 Has Been Shown to Play Oncogenic and Tumor-Suppressive Roles in Different Pediatric Tumors

miR-9 is a highly conserved miRNA involved in several different cellular processes including cell proliferation, differentiation, and migration. Early studies revealed miR-9 to be highly expressed in the brain and play a role both during development and in the adult brain; however, miR-9 has also been associated with many cancers outside the brain, acting as an oncogene or tumor suppressor (Coolen et al., 2013). Mir-9 is upregulated by MYC/MYCN and plays a role in promoting tumor growth and metastasis in several cancers including breast cancer, osteosarcoma, and rhabdomyosarcoma, where it is often overexpressed (Iorio et al., 2005; Luo et al., 2017) However, in other cancers such as neuroblastoma, miR-9's role is less clear, and studies have argued for oncogenic and tumor suppressor functions (Laneve et al., 2007; Zhi et al., 2014).

The role of miR-9 in osteosarcoma appears to be in promoting cell growth and metastasis (Zhu et al., 2015; Qi et al., 2016). In a study by Zhu et al. (2015), miR-9 knockdown suppressed cell growth and migration of osteosarcoma cells. They were also able to show that miR-9 downregulated RB1 via the Grap2 and cyclin D interacting protein (GCIP), thereby promoting E2F-mediated cell division (**Figure 1C**). Similar behavior has been observed in the alveolar subtype of rhabdomyosarcoma, where miR-9 contributes to increased cell proliferation and migration (Missiaglia et al., 2017). In this study by Missiaglia

et al. (2017), miR-9 was shown to be induced by the PAX3/FOXO1 fusion gene via MYCN, which is specific to this subtype of rhabdomyosarcoma.

In neuroblastoma, miR-9 expression has been shown to be both up- and downregulated in different studies. An early study by Laneve et al. (2007) showed that miR-9 was downregulated in 50% of primary neuroblastoma samples, and follow-up experiments demonstrated that miR-9 could act together with miR-125a and miR-125b to suppress cell proliferation by targeting a truncated isoform of the neurotrophin receptor tropomyosin-related kinase C (trkC) (**Figure 1D**). However, a later study by Ma et al. (2010) found miR-9 to be a target of MYCN and that miR-9 expression correlated with MYCN and metastatic status in neuroblastoma tumors. In this same study (albeit in breast cancer cells), Ma et al. also demonstrated that miR-9 suppressed E-cadherin to activate β-catenin and promote the epithelial-to-mesenchymal transition. Mir-9 is frequently involved in promoting cell migration; however, its absence has also been shown to produce different responses such as cell cycle arrest or apoptosis in neurons depending on their origin (Bonev et al., 2011). The contradictory behavior seen with studies of miR-9 highlight the diverse roles that individual miRNAs can play, and more comprehensive studies are needed to identify the relevant contextual influences on miRNA behavior.

miR-34 Is a Key Regulator of the Cell Cycle and Drug Resistance in Pediatric Solid Tumors

The miR-34 family has garnered significant interest since its members were discovered to be direct transcriptional targets of the tumor suppressor and transcription factor p53 (Hermeking, 2010). The miR-34 family consists of three miRNAs encoded by two genes, *mir-34a* and *mir-34b/c*. All three miRNAs play a key role in regulating apoptosis and the cell cycle by inducing G1 phase arrest. One of the more interesting facts about *miR-34a* and *miR-34b/c* is their genomic locations, which are located on chromosomes 1p36 and 11q23, respectively, regions that are frequently lost in pediatric solid tumors (Ruteshouser et al., 2005; Wittmann et al., 2007). In particular, loss of 1p36 occurs in 20–30% of neuroblastoma cases and correlates with MYCN amplification (Caron et al., 1993; Maris et al., 1995), whereas loss of 11q23 in occurs in approximately 40% of cases but almost never occurs with MYCN amplification (**Figure 1A**) (Guo et al., 1999; Attiyeh et al., 2005). miR-34 members are also regulators of the MYC family, as miR-34a is known to regulate MYCN and miR-34b and mir-34c to regulate c-MYC (Wei et al., 2008).

Studies on mir-34a expression have identified frequent downregulation in neuroblastoma, osteosarcoma, and hepatoblastoma (Jiao et al., 2016). miR-34a is itself considered a tumor suppressor due to its involvement in cell cycle arrest and apoptosis (De Antonellis et al., 2014). In neuroblastoma, Cole et al. (2008) investigated the growth-inhibitory effects of several miRNAs mapping to common chromosomal aberrations by overexpressing them in cell lines. In most cases, overexpression did not lead to a noticeable change in phenotype; however, miR-34a and miR-34c induced

significant growth inhibition in cell lines with 1p36 deletion. Growth inhibition and suppression of metastasis by miR-34a have also been shown in osteosarcoma by several studies, where members of key proliferative signal transduction pathways such as c-Met, DUSP1, and Eag1 were identified as regulatory targets (Yan et al., 2012; Wu X. et al., 2013; Gang et al., 2017). The miR-34 family also targets several members of the Notch signaling pathway, which has been linked to both oncogenic and tumor-suppressive roles depending on the cellular context. In osteosarcoma, activation of the Notch pathway is known to contribute to tumor growth, and miR-34a–mediated downregulation of this pathway likely contributes to its tumor-suppressive role. However, in Ewing sarcoma, a recent study investigating miR-34b suggested that it could act as an oncogene, promoting proliferation, migration, and invasion through Notch1 repression (Lu Q. et al., 2018). Prior studies have shown correlations between high mir-34a expression and patient survival, which would indicate a tumor-suppressive role for mir-34a (Nakatani et al., 2012; Marino et al., 2014). It is unclear why miR-34a and miR-34b would display contrasting effects given their shared targets, and further investigation is needed.

Several studies by Pu et al. (2016) and Pu et al. (2017) have suggested that miR-34a may also play a role in promoting multidrug resistance in osteosarcoma. In these studies, they found that miR-34a-5p enhanced multidrug resistance through downregulation of the *CD117* and *AGTR1* genes *in vitro*. CD117 is often highly expressed in drug-resistant tumors and is commonly used as a marker for stemness (Adhikari et al., 2010). In contrast, Nakatani et al. found that miR-34a increased chemosensitivity in Ewing sarcoma (Nakatani et al., 2012).

Other miRNAs Involved in Multiple Pediatric Solid Tumors

A substantial number of other miRNAs have been discovered with functional implications in multiple pediatric solid tumors. One such miRNA is miR-125b, which typically exhibits tumor-suppressive properties in cancers such as neuroblastoma, osteosarcoma, and Ewing sarcoma, where it is commonly dysregulated (Laneve et al., 2007; Li J. et al., 2014; Xiao et al., 2019). Previously, it was mentioned that miR-125b participates in a network with miR-125a and miR-9, regulating expression of a truncated trkC isoform to control neuroblastoma growth and differentiation (Laneve et al., 2007; Le et al., 2009). In osteosarcoma, miR-125b was found to regulate STAT3 by downregulating MAP kinase kinase 7 (MKK7), which inactivates STAT3 via dephosphorylation (Xiao et al., 2019). Loss of miR-125b and consequent overexpression of MKK7 led to increased tumor formation and poorer prognosis. In Ewing sarcoma, miR-125b is involved in regulating the PI3K signaling pathway; could inhibit cell proliferation, migration, and invasion; and induce apoptosis through suppression of PIK3CD (Li J et al., 2014). Conversely, in retinoblastoma, miR-125b is overexpressed and has shown oncogenic properties by promoting cell proliferation and migration and inhibiting apoptosis (Bai et al., 2016). Conflicting behavior with miR-125b

has been observed in many other cancers, which suggests that its role is highly dependent on cell identity (Sun et al., 2013).

miR-124 has been widely reported to act as a tumor suppressor by inhibiting cell growth and metastasis and acts as a key mediator of differentiation in several pediatric solid tumors (Peng et al., 2014; Feng et al., 2015; Zhao et al., 2017). In neuroblastoma, miR-124a increased the proportion of differentiated cells possessing neurite outgrowths (Le et al., 2009). In retinoblastoma, miR-124 participates in a regulatory network with lncRNAs Malat1 and XIST, which both function as oncogenes by enhancing growth and metastasis through downregulation of miR-124 (Liu S. et al., 2017; Hu et al., 2018). miR-124 itself was shown to target STAT3 to inhibit cell proliferation, migration, and invasion (Liu S. et al., 2016). In Ewing sarcoma, miR-124 expression is suppressed, and expression was found to reduce growth and metastasis *via* downregulation of mesenchymal genes such as *SLUG* and cyclin D2 (*CCND2*) (Li et al., 2017). Finally, in osteosarcoma, retinoblastoma, and Ewing sarcoma, miR-143 has been found to be dysregulated (De Vito et al., 2012; Li S. et al., 2014; Wang et al., 2016; Sun et al., 2018). For example, Li et al. investigated miR-143 function in osteosarcoma and showed that miR-143 participated in the TGF-β pathway by targeting versican, and TGF-β could reduce miR-143 expression to promote cell migration and invasion (Li S. et al., 2014). FOS-like antigen 2 (FOSL2) was also identified as a miR-143 target, which enhanced cell proliferation, migration, and invasion in the absence of miR-143 (Sun et al., 2018). Additional miRNA studies have been listed in **Table 1**.

TABLE 1 | miRNAs that have been shown to exhibit oncogenic or tumor-suppressive effects through functional studies in various pediatric solid tumors.

Cancer	Oncogenic miRNAs/clusters	Comments
Neuroblastoma	mir-15 (Xin et al., 2013), mir-17~92 (Fontana et al., 2008; Loven et al., 2010; Mestdagh et al., 2010), miR-93 (Chakrabarti et al., 2012), miR-380 (Swarbrick et al., 2010), miR-558 (Shohet et al., 2011; Qu et al., 2015)	mir-17~92 dysregulation is common in MYCN-amplified neuroblastomas.
Osteosarcoma	let-7d (Di Fiore et al., 2016), miR-9 (Zhu et al., 2015; Qi et al., 2016), miR-17~92 (Huang et al., 2012; Li X. et al., 2014; Lu et al., 2018a; Yang et al., 2018b), miR-34a (Pu et al., 2016; Pu et al., 2017), miR-214 (Rehei et al., 2018)	
Retinoblastoma	miR-17~92 (Conkrite et al., 2011; Nittner et al., 2012; Jo et al., 2014), miR-125b (Bai et al., 2016)	Loss of RB1 function may enable mir-17~92–mediated oncogenicity.
Wilms tumor	miR-19b (Liu G.-L. et al., 2017), miR-483 (Veronese et al., 2010; Liu M. et al., 2013), miR-1180 (Jiang and Li, 2018)	
Hepatoblastoma	miR-492 (von Frowein et al., 2018)	
Ewing sarcoma	mir-17~92 (Schwentner et al., 2017), miR-20b (Kawano et al., 2017), miR-34b, b (Lu et al., 2018b) miR-130b (Satterfield et al., 2017)	EWS-FLI-1 may upregulate mir-17~92 and miR-34b.
Cancer	Tumor-suppressive miRNAs/clusters	Comments
Neuroblastoma	let-7 (Buechner et al., 2011; Molenaar et al., 2012; Hennchen et al., 2015; Powers et al., 2016), mir-7-1 (Chakrabarti et al., 2012), mir-9 (Laneve et al., 2007), mir-10 (Foley et al., 2011), miR-34a (Welch et al., 2007; Cole et al., 2008; Tivnan et al., 2011), miR-34c (Cole et al., 2008), miR-101 (Buechner et al., 2011), miR-124a (Le et al., 2009), miR-125 (Laneve et al., 2007; Le et al., 2009), miR-145 (Zhang et al., 2014), miR-184 (Chen and Stallings, 2007; Foley et al., 2010; Tivnan et al., 2010), miR-193b (Roth et al., 2018), miR-202 (Buechner et al., 2011), miR-203 (Zhao et al., 2015), miR-449 (Buechner et al., 2011), miR-542 (Bray et al., 2011), miR-584 (Xiang et al., 2015), miR-591 (Shohet et al., 2011)	let-7 and mir-34 are regulators of the MYCN oncogene.
Osteosarcoma	let-7d (Di Fiore et al., 2016), miR-1 (Novello et al., 2013), miR-34 (Yan et al., 2012; Wu et al., 2013b; Gang et al., 2017; Wen et al., 2017), miR-125b (Le et al., 2009; Xiao et al., 2019), miR-133b (Novello et al., 2013), miR-134 (Thayanithy et al., 2012), miR-138 (Zhu et al., 2016), miR-143 (Li et al., 2014c; Sun et al., 2018), miR-195 (Han et al., 2015), miR-223 (Dong et al., 2016), miR-363 (Wang K. et al., 2018), miR-369 (Thayanithy et al., 2012), miR-382 (Thayanithy et al., 2012), miR-451 (Yuan et al., 2015), miR-454 (Niu et al., 2015), miR-485 (Du et al., 2018), miR-544 (Thayanithy et al., 2012), miR-590 (Wang W.T. et al., 2018), miR-708 (Chen and Zhou, 2018), miR-2682 (Zhang et al., 2018b)	miR-34 suppresses tumor growth in osteosarcoma but may also contribute to drug resistance.
Retinoblastoma	miR-101 (Lei et al., 2014), miR-124 (Liu S. et al., 2016), miR-143 (Wang et al., 2016)	STAT3 is a target of miR-124.
Rhabdomyosarcoma	miR-1 (Rao et al., 2010; Li et al., 2012), mir-22 (Bersani et al., 2016), miR-29 (Li et al., 2012), miR-133a (Rao et al., 2010), miR-206 (Li et al., 2012; Missiaglia et al., 2010)	miR-1 appears to play a key role in differentiation of several sarcomas.
Wilms tumor	let-7 (Urbach et al., 2014), miR-16 (Chen W. et al., 2018), miR-34a (Chen W. et al., 2018), mir-92a (Zhu et al., 2018b), miR-613 (Wang et al., 2017a)	miR-92a was shown to act as a tumor suppressor unlike what is observed in other pediatric solid tumors.
Hepatoblastoma	miR-26 (Zhang et al., 2018d)	miR-26 was shown to repress LIN28B in hepatoblastoma.
Ewing sarcoma	let-7 (Hameiri-Grossman et al., 2015; Kawano et al., 2015), miR-16 (Kawano et al., 2015), miR-22 (Parrish et al., 2015), miR-29b (Kawano et al., 2015), miR-30a (Franzetti et al., 2013), miR-30d (Ye et al., 2018), miR-31 (Karnuth et al., 2014), miR-34a (Nakatani et al., 2012; Ventura et al., 2016), miR-124 (Li et al., 2017), miR-125b (Li et al., 2014b), miR-143 (De Vito et al., 2012), miR-145 (Riggi et al., 2010; De Vito et al., 2012), miR-185 (Zhang et al., 2018c), miR-193b (Moore et al., 2017), miR-708 (Robin et al., 2012)	Several miRNAs such as let-7 and miR-145 are downregulated by EWS-FLI-1.

miRNAs Regulate All Aspects of Tumorigenesis

Widespread dysregulation of miRNAs is observed in many pediatric solid tumors, and functional studies have demonstrated that many of these miRNAs can drive or repress oncogenic pathways responsible for cell proliferation, apoptosis, angiogenesis, metastasis, and drug resistance. Importantly, miRNAs such as let-7 and miR-34 play a vital role in pediatric solid tumors by regulating established oncogenic transcription factors such as the MYC and E2F families (Wei et al., 2008; Buechner et al., 2011). Other miRNAs, such as the miR-17~92 cluster and miR-9, serve as downstream effectors for these transcription factors, although their exact role in tumorigenesis seems to depend on the overall transcriptional landscape (Schulte et al., 2008; Ma et al., 2010). In some cases, viewing miRNAs as oncogenes or tumor suppressors likely represents an oversimplification of their role in cancer, and a better understanding of their participation in oncogenic networks will be needed to clarify their exact contributions.

LONG NON-CODING RNAS REGULATE ONCOGENIC PATHWAYS IN PEDIATRIC SOLID TUMORS

For a long time, it was believed that the human genome was mostly comprised of "junk" DNA, despite pervasive transcription of much of the genome outside of protein-coding genes and other known RNAs at the time (Prensner and Chinnaiyan, 2011). Originally thought of as transcriptional noise, lncRNAs have now emerged as functional regulators of nearly all essential cellular processes including growth, differentiation, cell state maintenance, apoptosis, splicing, and epigenetic regulation. The first lncRNA, H19, was discovered in 1990 where an RNA molecule was found spliced and polyadenylated in a manner typical of mRNAs; however, it lacked an open reading frame and was believed to function as an untranslated RNA molecule (Brannan et al., 1990).

Often, lncRNAs participate within protein complexes and can operate as scaffolds, guides, decoys, or allosteric regulators. Many lncRNAs function as epigenetic regulators by interacting with proteins involved in chromatin remodeling and DNA methylation. Frequently, these lncRNAs will be cis-acting and regulate the regions near their transcribed location; however, some are trans-acting. Other lncRNAs function as competing RNAs (ceRNAs), which contain miRNA binding sites in a similar manner to mRNAs in order to compete and reduce the activity of miRNAs.

Several studies have investigated lncRNA expression in pediatric tumors and have successfully identified unique expression profiles in different cancers and tumor subtypes (Mitra et al., 2012; Brunner et al., 2012; Dong et al., 2014; Sahu et al., 2018). For example, Dong et al. (2014) compared hepatoblastoma samples to normal liver tissue in patients and found 2,736 differentially expressed lncRNAs. A study by Pandey et al. (2014) found 24 lncRNAs that could distinguish low- and high-risk neuroblastoma tumors. In a more recent study, Sahu et al. (2018) identified 16 differentially expressed lncRNAs that could be used to predict event-free

survival with greater accuracy than other commonly used clinical risk factors. Mechanistic studies into many of these lncRNAs have revealed that they frequently act as an additional layer of regulation within established oncogenic networks involving protein-coding genes and miRNAs. While the field of lncRNAs is still relatively young, many studies have emerged that suggest that lncRNAs are far more integrated into existing gene networks than what has previously been appreciated (**Figure 1**). In the following section, the roles of some of the better-characterized lncRNAs in pediatric solid tumors will be discussed.

Malat1 Is Induced by MYCN in Neuroblastoma and Competes With Many miRNAs

One of the earliest lncRNAs to be associated with disease was Malat1 (metastasis-associated lung adenocarcinoma transcript 1), which was shown to associate with metastatic tumors in non–small cell lung cancer patients (Ji et al., 2003). Malat1 is abundantly expressed and highly conserved across species, unlike many other lncRNAs, and displays remarkably diverse functions in cellular processes including alternative splicing, nuclear organization, and epigenetic modulation. Studies have suggested an important role for Malat1 in brain development, as it is highly expressed in neurons and its depletion has been shown to affect synapse and dendrite development (Bernard et al., 2010; Chen et al., 2016). However, its importance has been questioned as other studies have found that Malat1-KO mice are viable with no discernable change in phenotype (Nakagawa et al., 2012; Zhang et al., 2012).

In addition to lung cancer, Malat1 is known to contribute to metastasis in other common types of cancer including hepatocellular carcinoma and bladder cancer, with evidence that it acts through induction of the epithelial-to-mesenchymal transition (Ying et al., 2012; Li G. et al., 2014; Yang et al., 2017). The role of Malat1 in several pediatric cancers has also been explored in recent studies. In neuroblastoma, Tee et al. (2014) recently identified a regulatory network involving N-Myc, Malat1, and the histone demethylase JMJD1A. They found that N-Myc upregulated JMJD1A via direct binding of its promoter region and that JMJD1A could demethylate histone H3K9 near the promoter region of Malat1, leading to its upregulation. MYCN-mediated upregulation of Malat1 provides one mechanism in which its amplification can lead to increased metastasis in neuroblastoma patients. Another study by Bi et al. (2017) also demonstrated that Malat1 regulated Axl expression, a transmembrane receptor tyrosine kinase, which is known to activate pathways involved in cell proliferation, survival, and migration. In osteosarcoma, Dong et al. (2015) demonstrated that Malat1 was highly expressed and could activate the PI3K/Akt pathway to promote proliferation and invasion.

Malat1 is known to interact with many miRNAs implicated in cancer. In osteosarcoma, several studies have shown Malat1 can function as a ceRNA for different miRNAs (Wang et al., 2017b; Liu K. et al., 2017b; Sun and Qin, 2018). miR-140-5p is a tumor suppressor that downregulates HDAC4, a histone deacetylase that contributes to tumorigenesis, and competitive binding by Malat1 with miR-140-5p was shown to increase HDAC4 activity (Sun

and Qin, 2018). Malat1 was also shown to compete with miR-144-3p binding to ROCK1/ROCK2, promoting proliferation and metastasis (Wang et al., 2017b). In a similar manner, Liu K. et al. (2017) found that Malat1 could regulate cell growth through high-mobility group protein B1 (HMGB1) via ceRNA activity with miR-142-3p and miR-129-5p. Finally, in retinoblastoma, Malat1 downregulated miR-124 activity, leading to activation of the transcription factor SLUG, which is also targeted by miR-124 (Liu S. et al., 2017). SLUG has a known role in the epithelial-to-mesenchymal transition by suppressing E-cadherin via the Wnt/B-catenin pathway (Prasad et al., 2009).

In addition to interactions with miRNAs, Malat1 has also been shown to be processed directly by the Drosha–DGCR8 microprocessor complex through binding sites in the 5' end of the transcript (MacIas et al., 2012). lncRNAs such as Malat1 cooperate with the miRNA pathway and a number of transcription factors and epigenetic factors to form a complex network responsible for regulating tumorigenesis. The capacity for Malat1 to drive proliferation and metastasis in pediatric solid tumors suggests that dysregulation of any of these regulatory components can be sufficient for the development of cancer and highlights the value of further research into the relatively new field of lncRNAs.

H19: lncRNA Dysregulation via Loss of Imprinting may Contribute to Tumorigenesis

H19 is a paternally imprinted gene that is typically expressed exclusively from the maternal allele. Early reports suggested that H19 functioned as a tumor suppressor capable of inhibiting cell growth (Hao et al., 1993; Zhang et al., 1993; Casola et al., 1997; Fukuzawa et al., 1999). Studies in childhood solid tumors such as hepatoblastoma, Wilms tumor, and embryonic rhabdomyosarcoma supported this idea, as all three cancers often exhibited reduced H19 expression and had frequently lost the maternal 11p15 chromosomal region housing this gene (Fukuzawa et al., 1999). Other studies, in osteosarcoma and retinoblastoma, suggested an oncogenic role for H19, as its upregulation and loss of imprinting were commonly seen (Chan et al., 2014; Li L. et al., 2018). This observation was also seen in many other cancers including breast cancer (Lottin et al., 2002). Recently, the Hedgehog signaling pathway, a regulator of differentiation known to participate in cancer development and metastasis, was shown to induce H19 expression (Chan et al., 2014).

Understanding the exact function of H19 has proved difficult; however, it was known to sit downstream of the insulin growth factor 2 (IGF2) gene, a growth factor known to play a role in tumorigenesis. Early reports suggested interactions between IGF2 and H19, as loss of imprinting of either gene caused biallelic expression of the other gene (Ulaner et al., 2003). Ulaner et al. (2003) proposed a model for H19 and IGF2 involving a CCTF-binding site seated between the two genes, which could facilitate the blocking of IGF2 or transcription of H19 depending on its methylation status. However, this model suggested that H19 may simply serve as a marker for epigenetic disruptions and left open the question of what H19's actual function is.

More recent studies have demonstrated a role for H19 in epigenetic regulation. H19 binds to several epigenetic regulators including S-adenosylhomocysteine hydrolase (SAHH), methyl-CpG–binding domain protein 1 (MBD1), and enhancer of zeste homolog 2 (EZH2) (Raveh et al., 2015; Zhou et al., 2015). H19 was found to inhibit SAHH, which led to downregulation of DNMT3B-mediated methylation. MBD1 binds methylated DNA and recruits other proteins to mediate transcriptional repression or histone methylation, and H19 was shown to recruit this protein to several genes including IGF2 (Monnier et al., 2013). Finally, EZH2 is a histone methyltransferase that forms part of the Polycomb repressive complex 2 (PRC2) (Sauliere et al., 2006; Zhou et al., 2015).

H19 also plays a role in maintaining cells in an undifferentiated state by associating with the KH-type splicing regulatory protein (KSRP). When multipotent mesenchymal cells were induced, H19 was found to dissociate with KSRP to promote several of its functions including the decay of unstable mRNAs and increasing the expression of specific miRNAs involved in proliferation and differentiation though association with Drosha and Dicer.

H19's role in cancer has been emphasized by studies highlighting its relationship to the tumor suppressor p53. The H19 locus reciprocally regulates p53, as p53 suppresses H19 transcription and H19 can inactivate p53 by directly interacting with it (Yang et al., 2012). Notably, H19 also encodes for a miRNA in its first exon, miR-675, which suppresses p53 and several other targets including Rb, Igf1r, and several SMAD and cadherin genes. In the absence of functional p53, H19 was shown to promote tumor proliferation and survival under hypoxic conditions. Later studies in colorectal cancer showed that H19 could induce EMT by acting as a ceRNA (Liang et al., 2015). ceRNA function has recently been shown in a retinoblastoma study, targeting the mir-17~92 cluster (Zhang A. et al., 2018a). In this study, they found that H19 contained seven functional binding sites for mir-17~92 and was able to sponge mir-17~92 activity. This led to a de-repression of genes such as *p21* and STAT3 targets *BCL2*, *BCL2L1*, and *BIRC5*.

In a review by Raveh et al. (2015) it was proposed that H19 may behave differently in a manner that was dependent on the developmental stage of the cell, which could explain the evidence suggesting both oncogenic and tumor-suppressive roles. Here, the authors found that H19 functioned as a promoter of differentiation during the embryonic period and that absence of H19 at this stage could leave cells vulnerable to forming cancer, thereby seemingly acting as a tumor suppressor. However, in adult cells, where it is not normally expressed, H19 could function as an oncogene by promoting tumor survival and metastasis (Matouk et al., 2015).

TUG1 Regulates Transcription Factors Through Competition With miRNAs in Osteosarcoma

Recent studies have investigated the role of lncRNA TUG1 as a prognostic factor and ceRNA in osteosarcoma. Ma et al. (2016) identified a correlation between upregulation of TUG1 and poor

prognosis and metastasis, which was also evident in plasma, and suggested a potential use as a biomarker for patients with osteosarcoma. TUG1 is known to act through ceRNA activity against a number of miRNAs including miR-9, miR-132, miR-144, miR-153, miR-212, and miR-335 (Xie et al., 2016; Cao et al., 2017; Wang et al., 2017a; Li G. et al., 2018; Li H. et al., 2018). These miRNAs are known to regulate pathways involved in proliferation, cell cycle control, migration, and apoptosis. For example, TUG1 was shown to mediate de-repression of the transcription factor POU class 2 homeobox1 (POU2F1) via downregulation of mir-9 (**Figure 1E**) (Xie et al., 2016). POU2F1 itself participates in various cellular processes including growth, metabolism, stem cell identity, and metastasis (Vázquez-Arreguín and Tantin, 2016). In another example by Cao et al. (2017) they found that TUG1 also regulates migration and the epithelial-to-mesenchymal transition via ceRNA action on miRNA-144-3p. miR-144-3p is a regulator of EZH2, and upregulation of EZH2 induced cell migration through the Wnt/β-catenin pathway (Cao et al., 2017). Studies have also demonstrated direct interactions between TUG1 and the Polycomb repressor complex; however, to our knowledge, this has not been investigated in pediatric solid tumors (Yang et al., 2011).

Other Long Non-Coding RNAs in Pediatric Solid Tumors

In addition to those mentioned above, there are a number of other lncRNAs that have been identified as potential oncogenes or tumor suppressors involved in the pathogenesis of pediatric solid tumors (see **Table 2**) (Chen et al., 2017; Pandey et al., 2015).

For example, in osteosarcoma, lncRNAs HOTAIR, SNHG16, SNHG12, THOR, PACER, MFI2, and HOTTIP have all been shown to promote tumor or cell growth (Li et al., 2016; Qian et al., 2016; Ruan et al., 2016; Yin et al., 2016; Chen W. et al., 2018; Su et al., 2019; Wang et al., 2019). HOTAIR is known to play a role in chromatin regulation by acting as a scaffold for PRC2 and lysine-specific histone demethylase 1 (LSD1), and can also act as a ceRNA for miR-217 (Gupta et al., 2010; Tsai et al., 2010; Wang et al., 2019). SNHG16 has been shown to act as a ceRNA for several miRNAs including miR-205 and miR-340 (Zhu C. et al., 2018; Su et al., 2019). Additionally, several lncRNAs are downregulated in osteosarcoma with potential tumor-suppressive activity such as loc285194, MEG3, and TUSC7 (Pasic et al., 2010; Cong et al., 2016; Shi et al., 2018). loc285194 has been identified as a transcriptional target of p53 and can downregulate miR-211 (Liu Q. et al., 2013). In another study, increased p53

TABLE 2 | lncRNAs that play a role in pediatric solid tumors. OS, osteosarcoma; RB, retinoblastoma; NB, neuroblastoma; WT, Wilms tumor; HB, hepatoblastoma; RMS, rhabdomyosarcoma; ES, Ewing sarcoma.

Long Non-coding RNA	Cancer	Cellular Functions	References
Malat1	OS, RB, NB	Upregulates—proliferation, survival, migration, invasion.	(Tee et al., 2014; Dong et al., 2015; Bi et al., 2017; Liu K. et al., 2017; Liu S. et al., 2017; Wang K. et al., 2017; Sun and Qin, 2018)
H19	OS, RB, WT, HB, RMS	Upregulates—proliferation, survival. Regulates cell fate/differentiation.	(Zhang et al., 1993; Casola et al., 1997; Fukuzawa et al., 1999; Chan et al., 2014; Matouk et al., 2015; Raveh et al., 2015; Li L. et al., 2018)
TUG1	OS	Upregulates—proliferation, survival, migration.	(Ma et al., 2016; Xie et al., 2016; Cao et al., 2017; Wang S. et al., 2017; Li G. et al., 2018; Li H. et al., 2018)
HOTAIR	OS, RB	Upregulates—proliferation, survival, migration, invasion.	(Yang G. et al., 2018; Wang et al., 2019)
HOTTIP	OS	Upregulates—proliferation, resistance.	(Li et al., 2016)
SNHG12	OS	Upregulates—proliferation, migration.	(Ruan et al., 2016)
SNHG16	OS	Upregulates—proliferation, survival, migration, invasion.	(Su et al., 2019; Zhu et al., 2018a)
THOR	OS, RB	Upregulates—proliferation, migration.	(Chen K.S. et al., 2018; Shang, 2018)
PACER	OS	Upregulates—proliferation, invasion.	(Qian et al., 2016)
MFI2	OS	Upregulates—proliferation, survival, migration, invasion.	(Yin et al., 2016)
loc285194	OS	Downregulates—proliferation.	(Pasic et al., 2010)
TUSC7	OS	Downregulates—proliferation.	(Cong et al., 2016)
MEG3	OS, RB	Downregulates—proliferation, survival, invasion.	(Gao et al., 2017; Shi et al., 2018)
EWSAT1	OS, ES	Upregulates—proliferation, metastasis.	(Howarth et al., 2014; Sun et al., 2016)
XIST	RB	Upregulates—proliferation, survival.	(Hu et al., 2018)
DANCR	RB	Upregulates—proliferation, migration, invasion.	(Wang J. X. et al., 2018)
HOXA11-AS	RB	Upregulates—proliferation, survival.	(Han et al., 2019)
PANDAR	RB	Upregulates—survival.	(Sheng et al., 2018)
lncUSMYcN	NB	Upregulates—proliferation.	(Liu et al., 2014; Liu S. et al., 2016)
NBAT-1	NB	Downregulates—proliferation, invasion. Regulates cell fate/differentiation.	(Pandey et al., 2014)
CASC15-S	NB	Downregulates—proliferation, migration.	(Russell et al., 2015)
LINC00473	WT	Upregulates—proliferation, survival.	(Zhu et al., 2018b)
CRNDE	HB	Upregulates—proliferation, angiogenesis.	(Dong et al., 2017)
LINC01314	HB	Downregulates—proliferation, migration.	(Lv et al., 2018)

expression and a decrease in cell proliferation and invasion were observed when MEG3 was overexpressed (Shi et al., 2018). Furthermore, MEG3 was found to be downregulated by another lncRNA, EWSAT1, which had previously been shown to enhance cell proliferation and metastasis in both osteosarcoma and Ewing sarcoma (Howarth et al., 2014; Sun et al., 2016).

In retinoblastoma, HOTAIR, THOR, and MEG3 appear to have a similar influence as seen in osteosarcoma, where they also acted as oncogenes (HOTAIR and THOR) or tumor suppressors (MEG3) (Gao et al., 2017; Shang, 2018; Yang G. et al., 2018). In the study examining HOTAIR in retinoblastoma, HOTAIR was shown to be engaged in a reciprocal regulatory loop with miR-613 and promoted cell proliferation and activation of the EMT, potentially through upregulation of N-cadherin, vimentin, and α-SMA (Yang G. et al., 2018). Several lncRNAs have also been found acting as oncogenic ceRNAs including XIST, DANCR, and HOXA11-AS (Hu et al., 2018; Wang J. X. et al., 2018; Han et al., 2019). Finally, PANDAR is upregulated in retinoblastoma and may regulate cell proliferation and apoptosis via the Bcl-2/caspase-3 pathway (Sheng et al., 2018).

A number of studies have also suggested an important role for lncRNAs in neuroblastoma. For example, lncUSMYcN is an lncRNA that is frequently co-amplified alongside MYCN (Liu P. Y. et al., 2016). Liu et al. found that in neuroblastoma, lncUSMycN could upregulate MYCN through transcriptional activation of NCYM (a.k.a. MYCNOS), which codes for a protein that stabilizes MYCN (Suenaga et al., 2014). NCYM RNA has also been suggested to bind to the RNA-binding protein NonO, which is also known to upregulate MYCN expression (Liu et al., 2014; Liu P. Y. et al., 2016). Neuroblastoma associated transcript-1 (NBAT-1) is an epigenetic regulator that interacts with EZH2, and functions as a tumor suppressor due to its important role in neuronal differentiation (Pandey et al., 2014). Loss of NBAT-1 expression was found to increase cell proliferation and invasion (Pandey et al., 2014). Finally, an isoform of lncRNA CASC15, CASC15-S, was also implicated as a key element in neuronal differentiation, and low expression was associated with a poor outcome in patients (Russell et al., 2015).

In Wilms tumor, a study by Zhu et al. identified LINC00473 as an oncogenic lncRNA that is upregulated in unfavorable tumors (Zhu et al., 2018b). LINC00473 was shown to promote tumor growth and metastasis by acting as a ceRNA for the tumor suppressor miR-195 (Zhu et al., 2018b).

A study by Dong et al. identified 1757 upregulated and 979 downregulated lncRNAs comparing hepatoblastoma and normal tissues, suggesting that lncRNAs play a key role in this disease as well (Dong et al., 2014). The lncRNAs Colorectal Neoplasia Differentially Expressed (CRNDE) and LINC01314 have been investigated in more detail in hepatoblastoma (Dong et al., 2017; Lv et al., 2018). CRNDE is known to be frequently upregulated in hepatoblastoma, and knockdown of CRNDE activated the mTOR pathway and inhibited tumor growth and angiogenesis with a corresponding decrease in VEGFA and Ang-2 levels (Dong et al., 2017). LINC01314 was identified as a tumor suppressor, reducing proliferation and migration via downregulation of cell cycle proteins MCM7 and cyclin D1 (Lv et al., 2018).

CONCLUDING REMARKS

It is now clear that both miRNAs and lncRNAs form integral parts of the biological networks known to be impaired in pediatric solid tumors. miRNAs such as let-7 and mir-34 are key regulators of many pediatric oncogenes including *MYC*, *MYCN*, *RAS*, and *MET* (Johnson et al., 2005; Wei et al., 2008; Buechner et al., 2011; Yan et al., 2012). Additionally, ncRNAs such as the miR-17~92 cluster, mir-9, and Malat1 also serve as downstream effectors of MYC and MYCN (Schulte et al., 2008; Ma et al., 2010; Tee et al., 2014). Many more ncRNAs participate in these and other pathways to form a highly complex regulatory network essential for maintaining an optimal cell state (See **Tables 1** and **2**). ncRNA dysregulation offers an alternative mechanism to genetic mutations and DNA methylation whereby cell development and differentiation can be disturbed. Despite the relatively rare occurrence of mutations in pediatric solid tumors, copy number variations are common and often occur at regions of the genome that harbor ncRNAs with tumor-suppressive roles (Wei et al., 2008; Powers et al., 2016). Gene expression is often imprecise; however, miRNAs provide a layer of robustness, which helps ensure that biological networks respond appropriately to signals and remain functional despite an ever-increasing cellular disorder (Ebert and Sharp, 2012). lncRNAs, too, play a vital role in maintaining order by forming RNA–protein complexes and serving as ceRNA antagonists against miRNA-mediated repression, although much more work is needed in this field to fully comprehend their range of biological roles. Functional studies have revealed that dysregulation of ncRNAs is capable of driving progenitor cells towards oncogenesis. For example, this has been shown in retinoblastoma, where overexpression of the mir-17~92 cluster could drive tumor formation in *RB/p107*-deficient mice (Conkrite et al., 2011).

While genome-wide association studies have revealed that miRNA processing is frequently disrupted in Wilms tumor, this has not been shown to the same extent in other pediatric solid tumors. However, genetic mutations of protein-coding genes are only one way in which disruptions of miRNA processing can be revealed. Most miRNA studies ignore the fact that a high proportion of expressed miRNAs are isoforms (isomiRs). isomiRs originating from the same miRNA gene can possess a great deal of functional variability, with differences in target acquisition or turnover rate that can have a significant impact on overall gene regulation. Studies focusing on isomiR expression will provide an additional layer of resolution to our understanding of miRNA dysregulation.

Recent developments in single-cell technology have revealed heterogeneity in gene expression profiles among individual cells in many cancers such as glioblastoma and neuroblastoma (Patel et al., 2014; Boeva et al., 2017). Such studies suggest that many tumors comprise different cellular subtypes with unique phenotypes such as growth rate, drug resistance, and metastatic potential, which demand a new way of approaching cancer treatments. miRNA expression in pediatric solid tumors may also be heterogenous; however, limitations in single-cell technologies have left this avenue relatively unexplored, and further developments are needed.

So far, ncRNA research has played a key role in advancing our understanding of the mechanisms behind pediatric solid tumor development. Evidence supports an active role for ncRNAs in cancer that extends beyond mere passengers. However, continued research is needed to fully comprehend the molecular events leading to the development of cancer and unlock new possibilities for drug targets and biomarkers, which will ultimately lead to a better outcome for patients afflicted by these diseases.

REFERENCES

Adhikari, A. S., Agarwal, N., Wood, B. M., Porretta, C., Ruiz, B., Pochampally, R. R., et al. (2010). CD117 and Stro-1 identify osteosarcoma tumor-initiating cells associated with metastasis and drug resistance. *Cancer Res.* 70, 4602–4612. doi: 10.1158/0008-5472.CAN-09-3463

Allen-Rhoades, W., Whittle, S. B., and Rainusso, N. (2018). Pediatric solid tumors of infancy: an overview. *Pediatr. Rev.* 39, 57–67. doi: 10.1542/pir.2017-0057

Attiyeh, E. F., London, W. B., Mossé, Y. P., Wang, Q., Winter, C., Khazi, D., et al. (2005). Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N. Engl. J. Med.* 353, 2243–2253. doi: 10.1056/NEJMoa052399

Bai, S., Tian, B., Li, A., Yao, Q., Zhang, G., and Li, F. (2016). MicroRNA-125b promotes tumor growth and suppresses apoptosis by targeting DRAM2 in retinoblastoma. *Eye (Lond.)* 30, 1630–1638. doi: 10.1038/eye.2016.189

Balzeau, J., Menezes, M. R., Cao, S., and Hagan, J. P. (2017). The LIN28/let-7 pathway in cancer. *Front. Genet.* 8, 1–16. doi: 10.3389/fgene.2017.00031

Band, A. M., and Laiho, M. (2011). Crosstalk of TGF-β and estrogen receptor signaling in breast cancer. *J. Mammary Gland Biol. Neoplasia* 16, 109–115. doi: 10.1007/s10911-011-9203-7

Baumhoer, D., Zillmer, S., Unger, K., Rosemann, M., Atkinson, M. J., Irmler, M., et al. (2012). MicroRNA profiling with correlation to gene expression revealed the oncogenic miR-17-92 cluster to be up-regulated in osteosarcoma. *Cancer Genet.* 205, 212–219. doi: 10.1016/j.cancergen.2012.03.001

Bernard, D., Prasanth, K. V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., et al. (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* 29, 3082–3093. doi: 10.1038/emboj.2010.199

Bersani, F., Lingua, M. F., Morena, D., Foglizzo, V., Miretti, S., Lanzetti, L., et al. (2016). Deep sequencing reveals a novel miR-22 regulatory network with therapeutic potential in rhabdomyosarcoma. *Cancer Res.* 76, 6095–6106. doi: 10.1158/0008-5472.CAN-16-0709

Bi, S., Wang, C., Li, Y., Zhang, W., Zhang, J., Lv, Z., et al. (2017). LncRNA-MALAT1–mediated Axl promotes cell invasion and migration in human neuroblastoma. *Tumor Biol.* 39, 1–7. doi: 10.1177/1010428317699796

Bian, S., Hong, J., Li, Q., Schebelle, L., Pollock, A., Knauss, J. L., et al. (2013). MicroRNA Cluster miR-17-92 regulates neural stem cell expansion and transition to intermediate progenitors in the developing mouse neocortex. *Cell Rep.* 3, 1398–1406. doi: 10.1016/j.celrep.2013.03.037

Blenkiron, C., Goldstein, L. D., Thorne, N. P., Spiteri, I., Chin, S.-F., Dunning, M. J., et al. (2007). MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol.* 8, R214. doi: 10.1186/gb-2007-8-10-r214

Boeva, V., Louis-Brennetot, C., Peltier, A., Durand, S., Pierre-Eugène, C., Raynal, V., et al. (2017). Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* 49, 1408–1413. doi: 10.1038/ng.3921

Bonev, B., Pisco, A., and Papalopulu, N. (2011). MicroRNA-9 reveals regional diversity of neural progenitors along the anterior–posterior axis. *Dev. Cell.* 20, 19–32. doi: 10.1016/j.devcel.2010.11.018

Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* 10, 28–36. doi: 10.1128/MCB.10.1.28

Bray, I., Tivnan, A., Bryan, K., Foley, N. H., Watters, K. M., Tracey, L., et al. (2011). MicroRNA-542-5p as a novel tumor suppressor in neuroblastoma. *Cancer Lett.* 303, 56–64. doi: 10.1016/j.canlet.2011.01.016

Brunner, A. L., Beck, A. H., Edris, B., Sweeney, R. T., Zhu, S. X., Li, R., et al. (2012). Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol.* 13, R75. doi: 10.1186/gb-2012-13-8-r75

Buechner, J., Tømte, E., Haug, B. H., Henriksen, J. R., Løkke, C., Flægstad, T., et al. (2011). Tumour-suppressor microRNAs let-7 and mir-101 target the proto-oncogene MYCN and inhibit cell proliferation in MYCN-amplified neuroblastoma. *Br. J. Cancer* 105, 296–303. doi: 10.1038/bjc.2011.220

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. U.S A.* 99, 15524–15529. doi: 10.1073/pnas.242606799

Calin, G. A., Sevignani, C., Dumitru, C. D., Hyslop, T., Noch, E., Yendamuri, S., et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci.* 101, 2999–3004. doi: 10.1073/pnas.0307323101

Cao, J., Han, X., Qi, X., Jin, X., and Li, X. (2017). TUG1 promotes osteosarcoma tumorigenesis by upregulating EZH2 expression via MIR-144-3p. *Int. J. Oncol.* 51, 1115–1123. doi: 10.3892/ijo.2017.4110

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2005). Molecular biology: the transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563. doi: 10.1126/science.1112014

Caron, H., van Sluis, P., van Hoeve, M., de Kraker, J., Bras, J., Slater, R., et al. (1993). Allelic loss of chromosome 1p36 in neuroblastoma is of preferential maternal origin and correlates with N-myc amplification. *Nat. Genet.* 4, 187–190. doi: 10.1038/ng0693-187

Casola, S., Pedone, P. V., Cavazzana, A. O., Basso, G., Luksch, R., D'Amore, E. S. G., et al. (1997). Expression and parental imprinting of the H19 gene in human rhabdomyosarcoma. *Oncogene* 14, 1503–1510. doi: 10.1038/sj.onc.1200956

Cech, T. R., and Steitz, J. A. (2014). The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 157, 77–94. doi: 10.1016/j.cell.2014.03.008

Ceppi, P., and Peter, M. E. (2014). MicroRNAs regulate both epithelial-to-mesenchymal transition and cancer stem cells. *Oncogene* 33, 269–278. doi: 10.1038/onc.2013.55

Chakrabarti, M., Khandkar, M., Banik, N. L., and Ray, S. K. (2012). Alterations in expression of specific microRNAs by combination of 4-HPR and EGCG inhibited growth of human malignant neuroblastoma cells. *Brain Res.* 1454, 1–13. doi: 10.1016/j.brainres.2012.03.017

Chan, L. H., et al. (2014). Hedgehog signaling induces osteosarcoma development through Yap1 and H19 overexpression. *Oncogene* 33, 4857–4866. doi: 10.1038/onc.2013.433

Cheetham, S. W., Gruhl, F., Mattick, J. S., and Dinger, M. E. (2013). Long noncoding RNAs and the genetics of cancer. *Br. J. Cancer* 108, 2419–2425. doi: 10.1038/bjc.2013.233

Chen, Y., and Stallings, R. L. (2007). Differential patterns of microRNA expression in neuroblastoma are correlated with prognosis, differentiation, and apoptosis. *Cancer Res.* 67, 976–983. doi: 10.1158/0008-5472.CAN-06-3667

Chen, G., and Zhou, H. (2018). MiRNA-708/CUL4B axis contributes into cell proliferation and apoptosis of osteosarcoma. *Eur. Rev. Med. Pharmacol. Sci.* 22, 5452–5459. doi: 10.26355/eurrev_201809_15805

Chen, H. Z., Tsai, S. Y., and Leone, G. (2009). Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat. Rev. Cancer* 9, 785–797. doi: 10.1038/nrc2696

Chen, J., Huang, Z.-P., Seok, H. Y., Ding, J., Kataoka, M., Zhang, Z., et al. (2013). mir-17–92 cluster is required for and sufficient to induce cardiomyocyte

proliferation in postnatal and adult hearts. *Circ. Res.* 112, 1557–1566. doi: 10.1161/CIRCRESAHA.112.300658

Chen, X., Pappo, A., and Dyer, M. A. (2015). Pediatric solid tumor genomics and developmental pliancy. *Oncogene* 34, 5207–5215. doi: 10.1038/onc.2014.474

Chen, L., Feng, P., Zhu, X., He, S., Duan, J., and Zhou, D. (2016). Long non-coding RNA Malat1 promotes neurite outgrowth through activation of ERK/MAPK signalling pathway in N2a cells. *J. Cell. Mol. Med.* 20, 2102–2110. doi: 10.1111/jcmm.12904

Chen, R., Wang, G., Zheng, Y., Hua, Y., and Cai, Z. (2017). Long non-coding RNAs in osteosarcoma. *Oncotarget* 8, 20462–20475. doi: 10.18632/oncotarget.14726

Chen, K. S., Stroup, E. K., Budhipramono, A., Rakheja, D., Nichols-Vinueza, D., Xu, L., et al. (2018). Mutations in microRNA processing genes in Wilms tumors derepress the IGF2 regulator PLAG1. *Genes Dev.* 32, 996–1007. doi: 10.1101/gad.313783.118

Chen, W., Chen, M., Xu, Y., Chen, X., Zhou, P., Zhao, X., et al. (2018). Long non-coding RNA THOR promotes human osteosarcoma cell growth in vitro and in vivo. *Biochem. Biophys. Res. Commun.* 499, 913–919. doi: 10.1016/j.bbrc.2018.04.019

Classon, M., and Harlow, E. (2002). The retinoblastoma tumour suppressor in development and cancer. *Nat. Rev. Cancer* 2, 910–917. doi: 10.1038/nrc950

Cole, K. A., Attiyeh, E. F., Mosse, Y. P., Laquaglia, M. J., Diskin, S. J., Brodeur, G. M., et al. (2008). A Functional screen identifies miR-34a as a candidate neuroblastoma tumor suppressor gene. *Mol. Cancer Res.* 6, 735–742. doi: 10.1158/1541-7786.MCR-07-2102

Cong, M., Li, J., Jing, R., and Li, Z. (2016). Long non-coding RNA tumor suppressor candidate 7 functions as a tumor suppressor and inhibits proliferation in osteosarcoma. *Tumor Biol.* 37, 9441–9450. doi: 10.1007/s13277-015-4414-y

Conkrite, K., Sundby, M., Mukai, S., Michael Thomson, J., Mu, D., Hammond, S. M., et al. (2011). Mir-17~92 cooperates with RB pathway mutations to promote retinoblastoma. *Genes Dev.* 25, 1734–1745. doi: 10.1101/gad.17027411

Coolen, M., Katz, S., and Bally-Cuif, L. (2013). miR-9: a versatile regulator of neurogenesis. *Front. Cell. Neurosci.* 7, 1–11. doi: 10.3389/fncel.2013.00220

De Antonellis, P., Carotenuto, M., Vandenbussche, J., De Vita, G., Ferrucci, V., Medaglia, C., et al. (2014). Early targets of miR-34a in neuroblastoma. *Mol. Cell. Proteomics* 13, 2114–2131. doi: 10.1074/mcp.M113.035808

De Preter, K., Mestdagh, P., Vermeulen, J., Zeka, F., Naranjo, A., Bray, I., et al. (2011). miRNA expression profiling enables risk stratification in archived and fresh neuroblastoma tumor samples. *Clin. Cancer Res.* 17, 7684–7692. doi: 10.1158/1078-0432.CCR-11-0610

De Vito, C., Riggi, N., Cornaz, S., Suvà, M.-L., Baumer, K., Provero, P., et al. (2012). A TARBP2-dependent miRNA expression profile underlies cancer stem cell properties and provides candidate therapeutic reagents in Ewing sarcoma. *Cancer Cell* 21, 807–821. doi: 10.1016/j.ccr.2012.04.023

Delattre, O., Zucman, J., Melot, T., Garau, X. S., Zucker, J.-M., Lenoir, G. M., et al. (1994). The Ewing family of tumors — a subgroup of small-round-cell tumors defined by specific chimeric transcripts. *N. Engl. J. Med.* 331, 294–299. doi: 10.1056/NEJM199408043310503

Di Fiore, R., Drago-Ferrante, R., Pentimalli, F., Di Marzo, D., Forte, I. M., Carlisi, D., et al. (2016). Let-7d miRNA shows both antioncogenic and oncogenic functions in osteosarcoma-derived 3AB-OS cancer stem cells. *J. Cell. Physiol.* 231, 1832–1841. doi: 10.1002/jcp.25291

Dong, R., Jia, D., Xue, P., Cui, X., Li, K., Zheng, S., et al (2014). Genome-wide analysis of long noncoding RNA (lncRNA) expression in hepatoblastoma tissues. *PLoS One* 9, 1–9. doi: 10.1371/journal.pone.0085599

Dong, Y., Liang, G., Yuan, B., Yang, C., Gao, R., and Zhou, X. (2015). MALAT1 promotes the proliferation and metastasis of osteosarcoma cells by activating the PI3K/Akt pathway. *Tumour Biol.* 36, 1477–1486. doi: 10.1007/s13277-014-2631-4

Dong, J., Liu, Y., Liao, W., Liu, R., Shi, P., and Wang, L. (2016). miRNA-223 is a potential diagnostic and prognostic marker for osteosarcoma. *J. Bone Oncol.* 5, 74–79. doi: 10.1016/j.jbo.2016.05.001

Dong, R., Liu, X.-Q., Zhang, B.-B., Liu, B.-H., Zheng, S., and Dong, K.-R. (2017). Long non-coding RNA-CRNDE: a novel regulator of tumor growth and angiogenesis in hepatoblastoma. *Oncotarget* 8, 42087–42097. doi: 10.18632/oncotarget.14992

Du, Z., Zhang, X., Lou, Z., Guo, P., Zhang, F., Wang, B., et al. (2018). MicroRNA485-3p negatively regulates the transcriptional co-repressor CtBP1

to control the oncogenic process in osteosarcoma cells. *Int. J. Biol. Sci.* 14, 1445–1456. doi: 10.7150/ijbs.26335

Ebert, M. S., and Sharp, P. A. (2012). Roles for MicroRNAs in conferring robustness to biological processes. *Cell* 149, 505–524. doi: 10.1016/j.cell.2012.04.005

Esquela-Kerscher, A., and Slack, F. J. (2006). Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* 6, 259–269. doi: 10.1038/nrc1840

Feng, T., Xu, D., Tu, C., Li, W., Ning, Y., Ding, J., et al. (2015). miR-124 inhibits cell proliferation in breast cancer through downregulation of CDK4. *Tumor Biol.* 36, 5987–5997. doi: 10.1007/s13277-015-3275-8

Foley, N. H., Bray, I. M., Tivnan, A., Bryan, K., Murphy, D. M., Buckley, P. G., et al. (2010). MicroRNA-184 inhibits neuroblastoma cell survival through targeting the serine/threonine kinase AKT2. *Mol. Cancer* 9, 83. doi: 10.1186/1476-4598-9-83

Foley, N. H., Bray, I., Watters, K. M., Das, S., Bryan, K., Bernas, T., et al. (2011). MicroRNAs 10a and 10b are potent inducers of neuroblastoma cell differentiation through targeting of nuclear receptor corepressor 2. *Cell Death Differ.* 18, 1089–1098. doi: 10.1038/cdd.2010.172

Fontana, L., Fiori, M. E., Albini, S., Cifaldi, L., Giovinazzi, S., Forloni, M., et al. (2008). Antagomir-17-5p abolishes the growth of therapy-resistant neuroblastoma through p21 and BIM. *PLoS One* 3, e2236. doi: 10.1371/journal.pone.0002236

Franzetti, G.-A., Laud-Duval, K., Bellanger, D., Stern, M.-H., Sastre-Garau, X., and Delattre, O. (2013). MiR-30a-5p connects EWS-FLI1 and CD99, two major therapeutic targets in Ewing tumor. *Oncogene* 32, 3915–3921. doi: 10.1038/onc.2012.403

Friend, S. H., Bernards, R., Rogelj, S., Weinberg, R. A., Rapaport, J. M., Albert, D. M., et al. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323, 643–646. doi: 10.1038/323643a0

Fukuzawa, R., Umezawa, A., Ochi, K., Urano, F., Ikeda, H., and Hata, J. I. (1999). High frequency of inactivation of the imprinted H19 gene in 'sporadic' hepatoblastoma. *Int. J. Cancer* 82, 490–497. doi: 10.1002/(SICI)1097-0215(19990812)82:4<490::AID-IJC4>3.0.CO;2-I

Gadd, S., Huff, V., Walz, A. L., Ooms, A. H.A.G., Armstrong, A. E., Gerhard, D. S., et al. (2017). A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat. Genet.* 49, 1487–1494. doi: 10.1038/ng.3940

Gang, L., Qun, L., Liu, W.-D., Li, Y.-S., Xu, Y.-Z., and Yuan, D.-T. (2017). MicroRNA-34a promotes cell cycle arrest and apoptosis and suppresses cell adhesion by targeting DUSP1 in osteosarcoma. *Am. J. Transl. Res.* 9, 5388–5399.

Gao, Y., Huang, P., and Zhang, J. (2017). Hypermethylation of MEG3 promoter correlates with inactivation of MEG3 and poor prognosis in patients with retinoblastoma. *J. Transl. Med.* 15, 1–10. doi: 10.1186/s12967-017-1372-8

Guo, C., White, P. S., Weiss, M. J., Hogarty, M. D., Thompson, P. M., Stram, D. O., et al. (1999). Allelic deletion at 11q23 is common in MYCN single copy neuroblastomas. *Oncogene* 18, 4948–4957. doi: 10.1038/sj.onc.1202887

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. doi: 10.1038/nature08975

Ha, M., and Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509–524. doi: 10.1038/nrm3838

Hameiri-Grossman, M., Porat-Klein, A., Yaniv, I., Ash, S., Cohen, I. J., Kodman, Y., et al. (2015). The association between let-7, RAS and HIF-1α in Ewing sarcoma tumor growth. *Oncotarget* 6, 33834–33848. doi: 10.18632/oncotarget.5616

Han, K., Chen, X., Bian, N., Ma, B., Yang, T., Cai, C., et al. (2015). MicroRNA profiling identifies MiR-195 suppresses osteosarcoma cell metastasis by targeting CCND1. *Oncotarget* 6, 8875–8889. doi: 10.18632/oncotarget.3560

Han, N., Zuo, L., Chen, H., Zhang, C., He, P., and Yan, H. (2019). Long non-coding RNA homeobox A11 antisense RNA (HOXA11-AS) promotes retinoblastoma progression via sponging miR-506-3p. *Onco. Targets Ther.* 12, 3509–3517. doi: 10.2147/OTT.S195404

Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E., and Tycko, B. (1993). Tumour-suppressor activity of H19 RNA. *Nature* 365, 764–767. doi: 10.1038/365764a0

Hennchen, M., Stubbusch, J., Abarchan-El Makhfi, I., Kramer, M., Deller, T., Pierre-Eugene, C., et al. (2015). Lin28B and Let-7 in the control of sympathetic neurogenesis and neuroblastoma development. *J. Neurosci.* 35, 16531–16544. doi: 10.1523/JNEUROSCI.2560-15.2015

Hermeking, H. (2010). The miR-34 family in cancer and apoptosis. *Cell Death Differ.* 17, 193–199. doi: 10.1038/cdd.2009.56

Howarth, M. M., Simpson, D., Ngok, S. P., Nieves, B., Chen, R., Siprashvili, Z., et al. (2014). Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing sarcoma oncogenesis. *J. Clin. Invest.* 124, 5275–5290. doi: 10.1172/JCI72124

Hu, C., Liu, S., Han, M., Wang, Y., and Xu, C. (2018). Knockdown of lncRNA XIST inhibits retinoblastoma progression by modulating the miR-124/STAT3 axis. *Biomed. Pharmacother.* 107, 547–554. doi: 10.1016/j.biopha.2018.08.020

Huang, M., and Weiss, W. A. (2013). Neuroblastoma and MYCN. *Cold Spring Harb. Perspect. Med.* 3, a014415. doi: 10.1101/cshperspect.a014415

Huang, Y., Shen, X. J., Zou, Q., Wang, S. P., Tang, S. M., and Zhang, G. Z. (2011). Biological functions of microRNAs: a review. *J. Physiol. Biochem.* 67, 129–139. doi: 10.1007/s13105-010-0050-6

Huang, G., Nishimoto, K., Zhou, Z., Hughes, D., and Kleinerman, E. S. (2012). miR-20a encoded by the miR-17-92 cluster increases the metastatic potential of osteosarcoma cells by regulating Fas expression. *Cancer Res.* 72, 908–916. doi: 10.1158/0008-5472.CAN-11-1460

Iorio, M. V., Ferracin, M., Liu, C. G., Veronese, A., Spizzo, R., Sabbioni, S., et al. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* 65, 7065–7070. doi: 10.1158/0008-5472.CAN-05-1783

Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P. M., et al. (2003). MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 22, 8031–8041. doi: 10.1038/sj.onc.1206928

Jiang, X., and Li, H. (2018). MiR-1180-5p regulates apoptosis of Wilms' tumor by targeting p73. *Onco. Targets. Ther.* 11, 823–831. doi: 10.2147/OTT.S148684

Jiao, C., Zhu, A., Jiao, X., Ge, J., and Xu, X. (2016). Combined low miR-34s are associated with unfavorable prognosis in children with hepatoblastoma: a Chinese population-based study. *J. Pediatr. Surg.* 51, 1355–1361. doi: 10.1016/j.jpedsurg.2016.02.091

Jo, D. H., Kim, J. H., Cho, C. S., Cho, Y.-L., Jun, H. O., Yu, Y. S., et al. (2014). STAT3 inhibition suppresses proliferation of retinoblastoma through down-regulation of positive feedback loop of STAT3/miR-17-92 clusters. *Oncotarget* 5, 11513–11525. doi: 10.18632/oncotarget.2546

Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., et al. (2005). RAS is regulated by the let-7 microRNA family. *Cell* 120, 635–647. doi: 10.1016/j.cell.2005.01.014

Karnuth, B., Dedy, N., Spieker, T., Lawlor, E. R., Gattenlöhner, S., Ranft, A., et al. (2014). Differentially expressed miRNAs in Ewing sarcoma compared to mesenchymal stem cells: low miR-31 expression with effects on proliferation and invasion. *PLoS One* 9, e93067. doi: 10.1371/journal.pone.0093067

Kawano, M., Tanaka, K., Itonaga, I., Iwasaki, T., and Tsumura, H. (2015). c-Myc represses tumor-suppressive microRNAs, let-7a, miR-16 and miR-29b, and induces cyclin D2–mediated cell proliferation in Ewing's sarcoma cell line. *PLoS One* 10, e0138560. doi: 10.1371/journal.pone.0138560

Kawano, M., Tanaka, K., Itonaga, I., Iwasaki, T., and Tsumura, H. (2017). MicroRNA-20b promotes cell proliferation via targeting of TGF-β receptor II and upregulates MYC expression in Ewing's sarcoma cells. *Int. J. Oncol.* 51, 1842–1850. doi: 10.3892/ijo.2017.4155

Kline, N. E., and Sevier, N. (2003). Solid tumors in children. *J. Pediatr. Nurs.* 18, 96–102. doi: 10.1053/jpdn.2003.12

Koralov, S. B., Muljo, S. A., Galler, G. R., Krek, A., Chakraborty, T., Kanellopoulou, C., et al. (2008). Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. *Cell* 132, 860–874. doi: 10.1016/j.cell.2008.02.020

Kort, E. J., Farber, L., Tretiakova, M., Petillo, D., Furge, K. A., Yang, X. J., et al. (2008). The E2F3–oncomir-1 axis is activated in Wilms' tumor. *Cancer Res.* 68, 4034–4038. doi: 10.1158/0008-5472.CAN-08-0592

Laneve, P., Di Marcotullio, L., Gioia, U., Fiori, M. E., Ferretti, E., Gulino, A., et al. (2007). The interplay between microRNAs and the neurotrophin receptor tropomyosin-related kinase C controls proliferation of human neuroblastoma cells. *Proc. Natl. Acad. Sci.* 104, 7957–7962. doi: 10.1073/pnas.0700071104

Le, M. T. N., Xie, H., Zhou, B., Chia, P. H., Rizk, P., Um, M., et al. (2009). MicroRNA-125b promotes neuronal differentiation in human cells by repressing multiple targets. *Mol. Cell. Biol.* 29, 5290–5305. doi: 10.1128/MCB.01694-08

Lehrbach, N. J., Armisen, J., Lightfoot, H. L., Murfitt, K. J., Bugaut, A., Balasubramanian, S., et al. (2009). LIN-28 and the poly(U) polymerase PUP-2

regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* 16, 1016–1020. doi: 10.1038/nsmb.1675

Lei, Q., Shen, F., Wu, J., Zhang, W., Wang, J., and Zhang, L. (2014). miR-101, downregulated in retinoblastoma, functions as a tumor suppressor in human retinoblastoma cells by targeting EZH2. *Oncol. Rep.* 32, 261–269. doi: 10.3892/or.2014.3167

Leichter, A. L., Sullivan, M. J., Eccles, M. R., and Chatterjee, A. (2017). MicroRNA expression patterns and signalling pathways in the development and progression of childhood solid tumours. *Mol. Cancer* 16, 1–17. doi: 10.1186/s12943-017-0584-0

Leone, G., DeGregori, J., Sears, R., Jakoi, L., and Nevins, J. R. (1997). Myc and Ras collaborate in inducing accumulation of active cyclin E/Cdk2 and E2F. *Nature* 387, 422–426. doi: 10.1038/387422a0

Li, L., Sarver, A. L., Alamgir, S., and Subramanian, S. (2012). Downregulation of microRNAs miR-1, -206 and -29 stabilizes PAX3 and CCND2 expression in rhabdomyosarcoma. *Lab. Investig.* 92, 571–583. doi: 10.1038/labinvest.2012.10

Li, G., Zhang, H., Wan, X., Yang, X., Zhu, C., Wang, A., et al. (2014). Long noncoding RNA plays a key role in metastasis and prognosis of hepatocellular carcinoma. *Biomed Res. Int.* 2014, 1–8. doi: 10.1155/2014/780521

Li, J., You, T., and Jing, J. (2014). MiR-125b inhibits cell biological progression of Ewing's sarcoma by suppressing the PI3K/Akt signalling pathway. *Cell Prolif.* 47, 152–160. doi: 10.1111/cpr.12093

Li, S., Li, F., and Cheng, T. (2014). TGF-β1 promotes osteosarcoma cell migration and invasion through the miR-143–versican pathway. *Cell. Physiol. Biochem.* 34, 2169–2179. doi: 10.1159/000369660

Li, X., Yang, H., Tian, Q., Liu, Y., and Weng, Y. (2014). Upregulation of microRNA-17-92 cluster associates with tumor progression and prognosis in osteosarcoma. *Neoplasma* 61, 453–460. doi: 10.4149/neo_2014_056

Li, Z., Zhao, L., and Wang, Q. (2016). Overexpression of long non-coding RNA HOTTIP increases chemoresistance of osteosarcoma cell by activating the Wnt/β-catenin pathway. *Am. J. Transl. Res.* 8, 2385–2393.

Li, Y., Shao, G., Zhang, M., Zhu, F., Zhao, B., He, C., et al. (2017). miR-124 represses the mesenchymal features and suppresses metastasis in Ewing sarcoma. *Oncotarget* 8, 10274–10286. doi: 10.18632/oncotarget.14394

Li, G., Liu, K., and Du, X. (2018). Long non-coding RNA TUG1 promotes proliferation and inhibits apoptosis of osteosarcoma cells by sponging miR-132-3p and upregulating SOX4 expression. *Yonsei Med. J.* 59, 226–235. doi: 10.3349/ymj.2018.59.2.226

Li, H., Tian, G., Tian, F., and Shao, L. (2018). Long non-coding RNA TUG1 promotes osteosarcoma cell proliferation and invasion through inhibition of microrna-212-3p expression. *Exp. Ther. Med.* 16, 779–787. doi: 10.3892/etm.2018.6216

Li, L., Chen, W., Wang, Y., Tang, L., and Han, M. (2018). Long non-coding RNA H19 regulates viability and metastasis, and is upregulated in retinoblastoma. *Oncol. Lett.* 8424–8432. doi: 10.3892/ol.2018.8385

Liang, W.-C., Fu, W.-M., Wong, C.-W., Wang, Y., Wang, W.-M., Hu, G.-X., et al. (2015). The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. *Oncotarget* 6, 22513–22525. doi: 10.18632/oncotarget.4154

Liu, M., Roth, A., Yu, M., Morris, R., Bersani, F., Rivera, M. N., et al. (2013). The IGF2 intronic miR-483 selectively enhances transcription from IGF2 fetal promoters and enhances tumorigenesis. *Genes Dev.* 27, 2543–2548. doi: 10.1101/gad.224170.113

Liu, Q., Huang, J., Zhou, N., Zhang, Z., Zhang, A., Lu, Z., et al. (2013). LncRNA loc285194 is a p53-regulated tumor suppressor. *Nucleic Acids Res.* 41, 4976–4987. doi: 10.1093/nar/gkt182

Liu, P. Y., Erriquez, D., Marshall, G. M., Tee, A. E., Polly, P., Wong, M., et al. (2014). Effects of a novel long noncoding RNA, lncUSMycN, on N-Myc expression and neuroblastoma progression. *J. Natl. Cancer Inst.* 106, dju113–dju113. doi: 10.1093/jnci/dju113

Liu, P. Y., Atmadibrata, B., Mondal, S., Tee, A. E., and Liu, T. (2016). NCYM is upregulated by lncUSMycN and modulates N-Myc expression. *Int. J. Oncol.* 49, 2464–2470. doi: 10.3892/ijo.2016.3730

Liu, S., Hu, C., Wang, Y., Shi, G., Li, Y., and Wu, H. (2016). miR-124 inhibits proliferation and invasion of human retinoblastoma cells by targeting STAT3. *Oncol. Rep.* 36, 2398–2404. doi: 10.3892/or.2016.4999

Liu, G.-L., Yang, H.-J., Liu, B., and Liu, T. (2017). Effects of microRNA-19b on the proliferation, apoptosis, and migration of Wilms' tumor cells via the PTEN/

PI3K/AKT signaling pathway. *J. Cell. Biochem.* 118, 3424–3434. doi: 10.1002/jcb.25999

Liu, K., Huang, J., Ni, J., Song, D., Ding, M., Wang, J., et al. (2017). MALAT1 promotes osteosarcoma development by regulation of HMGB1 via miR-142–3p and miR-129–5p. *Cell Cycle* 16, 578–587. doi: 10.1080/15384101.2017.1288324

Liu, S., Yan, G., Zhang, J., and Yu, L. (2017). Knockdown of long noncoding RNA (lncRNA) metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) inhibits proliferation, migration, and invasion and promoted apoptosis by targeting miR-124 in retinoblastoma. *Oncol. Res. Featur. Preclin. Clin. Cancer Ther.* 26, 581–591. doi: 10.3727/096504017X149539 48675403

Lottin, S., Adriaenssens, E., Dupressoir, T., Berteaux, N., Montpellier, C., Coll, J., et al. (2002). Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. Carcinogenesis 23, 1885–1895. doi: 10.1093/carcin/23.11.1885

Loven, J., Zinin, N., Wahlstrom, T., Muller, I., Brodin, P., Fredlund, E., et al. (2010). MYCN-regulated microRNAs repress estrogen receptor-α (ESR1) expression and neuronal differentiation in human neuroblastoma. *Proc. Natl. Acad. Sci.* 107, 1553–1558. doi: 10.1073/pnas.0913517107

Lu, C., Peng, K., Guo, H., Ren, X., Hu, S., Cai, Y., et al. (2018). miR-18a-5p promotes cell invasion and migration of osteosarcoma by directly targeting IRF2. *Oncol. Lett.* 16, 3150–3156. doi: 10.3892/ol.2018.9032

Lu, Q., Lu, M., Li, D., and Zhang, S. (2018). MicroRNA-34b promotes proliferation, migration and invasion of Ewing's sarcoma cells by downregulating Notch1. *Mol. Med. Rep.* 18, 3577–3588. doi: 10.3892/mmr.2018.9365

Luo, T., Yi, X., and Si, W. (2017). Identification of miRNA and genes involving in osteosarcoma by comprehensive analysis of microRNA and copy number variation data. *Oncol. Lett.* 14, 5427–5433. doi: 10.3892/ol.2017.6845

Lv, B., Zhang, L., Miao, R., Xiang, X., Dong, S., Lin, T., et al. (2018). Comprehensive analysis and experimental verification of LINC01314 as a tumor suppressor in hepatoblastoma. *Biomed. Pharmacother.* 98, 783–792. doi: 10.1016/j.biopha.2018.01.013

Ma, L., Young, J., Prabhala, H., Pan, E., Mestdagh, P., Muth, D., et al. (2010). MiR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat. Cell Biol.* 12, 247–256. doi: 10.1038/ncb2024

Ma, B., Li, M., Zhang, L., Huang, M., Lei, J.-B., Fu, G.-H., Liu, C.-X., et al. (2016). Upregulation of long non-coding RNA TUG1 correlates with poor prognosis and disease status in osteosarcoma. *Tumor Biol.* 37, 4445–4455. doi: 10.1007/s13277-015-4301-6

MacIas, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Cáceres, J. F. (2012). DGCR8 HITS-CLIP reveals novel functions for the microprocessor. *Nat. Struct. Mol. Biol.* 19, 760–766. doi: 10.1038/nsmb.2344

Marino, M. T., Grilli, A., Baricordi, C., Manara, M. C., Ventura, S., Pinca, R. S., et al. (2014). Prognostic significance of miR-34a in Ewing sarcoma is associated with cyclin D1 and ki-67 expression. *Ann. Oncol.* 25, 2080–2086. doi: 10.1093/annonc/mdu249

Maris, J. M., White, P. S., Beltinger, C. P., Sulman, E. P., Castleberry, R. P., Shuster, J. J., et al. (1995). Significance of chromosome 1p loss of heterozygosity in neuroblastoma. *Cancer Res.* 55, 4664–4669.

Matouk, I. J., Halle, D., Raveh, E., Gilon, M., Sorin, V., and Hochberg, A. (2015). The role of the oncofetal H19 lncRNA in tumor metastasis: orchestrating the EMT–MET decision. *Oncotarget* 7, 3748–3765. doi: 10.18632/oncotarget.6387

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159. doi: 10.1038/nrg2521

Mestdagh, P., Boström, A.-K., Impens, F., Fredlund, E., Van Peer, G., De Antonellis, P., et al. (2010). The miR-17-92 microRNA cluster regulates multiple components of the TGF-β pathway in neuroblastoma. *Mol. Cell* 40, 762–773. doi: 10.1016/j.molcel.2010.11.038

Missiaglia, E., Shepherd, C. J., Patel, S., Thway, K., Pierron, G., Pritchard-Jones, K., et al. (2010). MicroRNA-206 expression levels correlate with clinical behaviour of rhabdomyosarcomas. *Br. J. Cancer* 102, 1769–1777. doi: 10.1038/sj.bjc.6605684

Missiaglia, E., Shepherd, C. J., Aladowicz, E., Olmos, D., Selfe, J., Pierron, G., et al. (2017). MicroRNA and gene co-expression networks characterize biological and clinical behavior of rhabdomyosarcomas. *Cancer Lett.* 385, 251–260. doi: 10.1016/j.canlet.2016.10.011

Mitra, S. A., Mitra, A. P., and Triche, T. J. (2012). A central role for long non-coding RNA in cancer. *Front. Genet.* 3, 1–9. doi: 10.3389/fgene.2012.00017

Mogilyansky, E., and Rigoutsos, I. (2013). The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ.* 20, 1603–1614. doi: 10.1038/cdd.2013.125

Molenaar, J. J., Domingo-Fernández, R., Ebus, M. E., Lindner, S., Koster, J., Drabek, K., et al. (2012). LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. *Nat. Genet.* 44, 1199–1206. doi: 10.1038/ng.2436

Monnier, P., Martinet, C., Pontis, J., Stancheva, I., Ait-Si-Ali, S., and Dandolo, L. (2013). H19 lncRNA controls gene expression of the imprinted gene network by recruiting MBD1. *Proc. Natl. Acad. Sci.* 110, 20693–20698. doi: 10.1073/pnas.1310201110

Moore, C., Parrish, J. K., and Jedlicka, P. (2017). MiR-193b, downregulated in Ewing sarcoma, targets the ErbB4 oncogene to inhibit anchorage-independent growth. *PLoS One* 12, e0178028. doi: 10.1371/journal.pone.0178028

Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., et al. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA* 18, 1487–1499. doi: 10.1261/rna.033217.112

Nakatani, F., Ferracin, M., Manara, M. C., Ventura, S., Del Monaco, V., Ferrari, S., et al. (2012). miR-34a predicts survival of Ewing's sarcoma patients and directly influences cell chemo-sensitivity and malignancy. *J. Pathol.* 226, 796–805. doi: 10.1002/path.3007

Nguyen, L. H., Robinton, D. A., Seligson, M. T., Wu, L., Li, L., Rakheja, D., et al. (2014). Lin28b is sufficient to drive liver cancer and necessary for its maintenance in murine models. *Cancer Cell* 26, 248–261. doi: 10.1016/j.ccr.2014.06.018

Nittner, D., Lambertz, I., Clermont, F., Mestdagh, P., Köhler, C., Nielsen, S. J., et al. (2012). Synthetic lethality between Rb, p53 and Dicer or miR-17-92 in retinal progenitors suppresses retinoblastoma formation. *Nat. Cell Biol.* 14, 958–965. doi: 10.1038/ncb2556

Niu, G., Li, B., Sun, J., and Sun, L. (2015). miR-454 is down-regulated in osteosarcomas and suppresses cell proliferation and invasion by directly targeting c-Met. *Cell Prolif.* 48, 348–355. doi: 10.1111/cpr.12187

Novello, C., Pazzaglia, L., Cingolani, C., Conti, A., Quattrini, I., Manara, M. C., et al. (2013). MiRNA expression profile in human osteosarcoma: role of miR-1 and miR-133b in proliferation and cell cycle control. *Int. J. Oncol.* 42, 667–675. doi: 10.3892/ijo.2012.1717

O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., and Mendell, J. T. (2005). c-Myc–regulated microRNAs modulate E2F1 expression. *Nature* 435, 839–843. doi: 10.1038/nature03677

Pandey, G. K., Mitra, S., Subhash, S., Hertwig, F., Kanduri, M., Mishra, K., et al. (2014). The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer Cell* 26, 722–737. doi: 10.1016/j.ccell.2014.09.014

Pandey, G. K., and Kanduri, C. (2015). Long noncoding RNAs and neuroblastoma. *Oncotarget* 6, 18265–18275. doi: 10.18632/oncotarget.4251

Parrish, J. K., Sechler, M., Winn, R. A., and Jedlicka, P. (2015). The histone demethylase KDM3A is a microRNA-22-regulated tumor promoter in Ewing sarcoma. *Oncogene* 34, 257–262. doi: 10.1038/onc.2013.541

Pasic, I., Shlien, A., Durbin, A. D., Stavropoulos, D. J., Baskin, B., Ray, P. N., et al. (2010). Recurrent focal copy-number changes and loss of heterozygosity implicate two noncoding RNAs and one tumor suppressor gene at chromosome 3q13.31 in osteosarcoma. *Cancer Res.* 70, 160–171. doi: 10.1158/0008-5472.CAN-09-1902

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408, 86–89. doi: 10.1038/35040556

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-.)* 344, 1396–1401. doi: 10.1126/science.1254257

Peng, X. H., Huang, H. R., Lu, J., Liu, X., Zhao, F. P., Zhang, B., et al. (2014). MiR-124 suppresses tumor growth and metastasis by targeting Foxq1 in nasopharyngeal carcinoma. *Mol. Cancer* 13, 1–13. doi: 10.1186/1476-4598-13-186

Powers, J. T., Tsanov, K. M., Pearson, D. S., Roels, F., Spina, C. S., Ebright, R., et al. (2016). Multiple mechanisms disrupt the let-7 microRNA family in neuroblastoma. *Nature* 535, 246–251. doi: 10.1038/nature18632

Prasad, C. P., Rath, G., Mathur, S., Bhatnagar, D., Parshad, R., and Ralhan, R. (2009). Expression analysis of E-cadherin, Slug and GSK3beta in invasive ductal carcinoma of breast. *BMC Cancer* 9, 325. doi: 10.1186/1471-2407-9-325

Prensner, J. R., and Chinnaiyan, A. M. (2011). The emergence of lncRNAs in cancer biology. *Cancer Discov.* 1, 391–407. doi: 10.1158/2159-8290.CD-11-0209

Pu, Y., Zhao, F., Wang, H., Cai, W., Gao, J., Li, Y., et al. (2016). MiR-34a-5p promotes the multi-drug resistance of osteosarcoma by targeting the CD117 gene. *Oncotarget* 7, 28420–28434. doi: 10.18632/oncotarget.8546

Pu, Y., Zhao, F., Li, Y., Cui, M., Wang, H., Meng, X., et al. (2017). The miR-34a-5p promotes the multi-chemoresistance of osteosarcoma via repression of the AGTR1 gene. *BMC Cancer* 17, 45. doi: 10.1186/s12885-016-3002-x

Qi, X. J., Wang, J. F., Wang, G. D., Xu, Q., and Sun, H. L. (2016). Pivotal role of microRNA-9 in osteosarcoma tumorigenesis and tumor progression. *Genet. Mol. Res.* 15, 1–10. doi: 10.4238/gmr.15017318

Qian, M., Yang, X., Li, Z., Jiang, C., Song, D., Yan, W., et al. (2016). P50-associated COX-2 extragenic RNA (PACER) overexpression promotes proliferation and metastasis of osteosarcoma cells by activating COX-2 gene. *Tumor Biol.* 37, 3879–3886. doi: 10.1007/s13277-015-3838-8

Qu, H., Zheng, L., Pu, J., Mei, H., Xiang, X., Zhao, X., et al. (2015). miRNA-558 promotes tumorigenesis and aggressiveness of neuroblastoma cells through activating the transcription of heparanase. *Hum. Mol. Genet.* 24, 2539–2551. doi: 10.1093/hmg/ddv018

Rakheja, D., Chen, K. S., Liu, Y., Shukla, A. A., Schmid, V., Chang, T. C., et al. (2014). Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. *Nat. Commun.* 2, 1–10. doi: 10.1038/ncomms5802

Rao, P. K., Missiaglia, E., Shields, L., Hyde, G., Yuan, B., Shepherd, C. J., et al. (2010). Distinct roles for miR-1 and miR-133a in the proliferation and differentiation of rhabdomyosarcoma cells. *FASEB J.* 24, 3427–3437. doi: 10.1096/fj.09-150698

Raveh, E., Matouk, I. J., Gilon, M., and Hochberg, A. (2015). The H19 long non-coding RNA in cancer initiation, progression and metastasis—a proposed unifying theory. *Mol. Cancer* 14, 1–14. doi: 10.1186/s12943-015-0458-2

Rehei, A.-L., Zhang, L., Fu, Y.-X., Mu, W.-B., Yang, D.-S., Liu, Y., et al. (2018). MicroRNA-214 functions as an oncogene in human osteosarcoma by targeting TRAF3. *Eur. Rev. Med. Pharmacol. Sci.* 22, 5156–5164. doi: 10.26355/eurrev_201808_15711

Riffo-Campos, Á. L., Riquelme, I., and Brebi-Mieville, P. (2016). Tools for sequence-based miRNA target prediction: what to choose? *Int. J. Mol. Sci.* 17, E1987. doi: 10.3390/ijms17121987

Riggi, N., Suvà, M.-L., De Vito, C., Provero, P., Stehle, J.-C., Baumer, K., et al. (2010). EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes Dev.* 24, 916–932. doi: 10.1101/gad.1899710

Robin, T. P., Smith, A., McKinsey, E., Reaves, L., Jedlicka, P., and Ford, H. L. (2012). EWS/FLI1 Regulates EYA3 in Ewing sarcoma via modulation of miRNA-708, resulting in increased cell survival and chemoresistance. *Mol. Cancer Res.* 10, 1098–1108. doi: 10.1158/1541-7786.MCR-12-0086

Roth, S. A., Hald, Ø. H., Fuchs, S., Løkke, C., Mikkola, I., Flægstad, T., et al. (2018). MicroRNA-193b-3p represses neuroblastoma cell growth via downregulation of Cyclin D1, MCL-1 and MYCN. *Oncotarget* 9, 18160–18179. doi: 10.18632/oncotarget.24793

Ruan, W., Wang, P., Feng, S., Xue, Y., and Li, Y. (2016). Long non-coding RNA small nucleolar RNA host gene 12 (SNHG12) promotes cell proliferation and migration by upregulating angiomotin gene expression in human osteosarcoma cells. *Tumor Biol.* 37, 4065–4073. doi: 10.1007/s13277-015-4256-7

Russell, M. R., Penikis, A., Oldridge, D. A., Alvarez-Dominguez, J. R., McDaniel, L., Diamond, M., et al. (2015). CASC15-S is a tumor suppressor lncRNA at the 6p22 neuroblastoma susceptibility locus. *Cancer Res.* 75, 3155–3166. doi: 10.1158/0008-5472.CAN-14-3613

Ruteshouser, E. C., Hendrickson, B. W., Colella, S., Krahe, R., Pinto, L., and Huff, V. (2005). Genome-wide loss of heterozygosity analysis of WT1-wild-type and WT1-mutant Wilms tumors. *Genes Chromosom. Cancer* 43, 172–180. doi: 10.1002/gcc.20169

Sahu, D., Ho, S.-Y., Juan, H.-F., and Huang, H.-C. (2018). High-risk, expression-based prognostic long noncoding RNA signature in neuroblastoma. *JNCI Cancer Spectr.* 2, 1–12. doi: 10.1093/jncics/pky015

Satterfield, L., Shuck, R., Kurenbekova, L., Allen-Rhoades, W., Edwards, D., Huang, S., et al. (2017). miR-130b directly targets ARHGAP1 to drive activation of a metastatic CDC42–PAK1–AP1 positive feedback loop in Ewing sarcoma. *Int. J. Cancer* 141, 2062–2075. doi: 10.1002/ijc.30909

Sauliere, J., Sureau, A., Expert-Bezancon, A., and Marie, J. (2006). The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol. Cell. Biol.* 26, 8755–8769. doi: 10.1128/MCB.00893-06

Schulte, J. H., Horn, S., Otto, T., Samans, B., Heukamp, L. C., Eilers, U. C., et al. (2008). MYCN regulates oncogenic microRNAs in neuroblastoma. *Int. J. Cancer* 122, 699–704. doi: 10.1002/ijc.23153

Schwentner, R., Herrero-Martin, D., Kauer, M. O., Mutz, C. N., Katschnig, A. M., Sienski, G., et al. (2017). The role of miR-17-92 in the miRegulatory landscape of Ewing sarcoma. *Oncotarget* 8, 10980–10993. doi: 10.18632/oncotarget.14091

Shang, Y. (2018). LncRNA THOR acts as a retinoblastoma promoter through enhancing the combination of c-myc mRNA and IGF2BP1 protein. *Biomed. Pharmacother.* 106, 1243–1249. doi: 10.1016/j.biopha.2018.07.052

Sheng, L., Wu, J., Gong, X., Dong, D., and Sun, X. (2018). SP1-induced upregulation of lncRNA PANDAR predicts adverse phenotypes in retinoblastoma and regulates cell growth and apoptosis in vitro and in vivo. *Gene* 668, 140–145. doi: 10.1016/j.gene.2018.05.065

Shi, Y., Lv, C., Shi, L., and Tu, G. (2018). MEG3 inhibits proliferation and invasion and promotes apoptosis of human osteosarcoma cells. *Oncol. Lett.* 15, 1917–1923. doi: 10.3892/ol.2017.7463

Shohet, J. M., Ghosh, R., Coarfa, C., Ludwig, A., Benham, A. L., Chen, Z., et al. (2011). A genome-wide search for promoters that respond to increased MYCN reveals both new oncogenic and tumor suppressor microRNAs associated with aggressive neuroblastoma. *Cancer Res.* 71, 3841–3851. doi: 10.1158/0008-5472.CAN-10-4391

Sin-Chan, P., Mumal, I., Suwal, T., Ho, B., Fan, X., Singh, I., et al. (2019). A C19MC–LIN28A–MYCN oncogenic circuit driven by hijacked super-enhancers is a distinct therapeutic vulnerability in ETMRs: a lethal brain tumor. *Cancer Cell.* 36, 51–67. doi: 10.1016/j.ccell.2019.06.002

Strieder, V., and Lutz, W. (2003). E2F proteins regulate MYCN expression in neuroblastomas. *J. Biol. Chem.* 278, 2983–2989. doi: 10.1074/jbc.M207596200

Su, P., Mu, S., and Wang, Z. (2019). Long noncoding RNA SNHG16 promotes osteosarcoma cells migration and invasion via sponging miRNA-340. *DNA Cell Biol.* 38, 170–175. doi: 10.1089/dna.2018.4424

Suenaga, Y., Islam, S. M., Alagu, J., Kaneko, Y., Kato, M., Tanaka, Y., et al. (2014). NCYM, a cis-antisense gene of MYCN, encodes a de novo evolved protein that inhibits GSK3β resulting in the stabilization of MYCN in human neuroblastomas. *PLoS Genet.* 10, e1003996. doi: 10.1371/journal.pgen.1003996

Sun, Y., and Qin, B. (2018). Long noncoding RNA MALAT1 regulates HDAC4-mediated proliferation and apoptosis via decoying of miR-140-5p in osteosarcoma cells. *Cancer Med.* 7, 4584–4597. doi: 10.1002/cam4.1677

Sun, Y. M., Lin, K. Y., and Chen, Y. Q. (2013). Diverse functions of miR-125 family in different cell contexts. *J. Hematol. Oncol.* 6, 1–8. doi: 10.1186/1756-8722-6-6

Sun, L., Yang, C., Xu, J., Feng, Y., Wang, L., and Cui, T. (2016). Long noncoding RNA EWSAT1 promotes osteosarcoma cell growth and metastasis through suppression of MEG3 expression. *DNA Cell Biol.* 35, 812–818. doi: 10.1089/dna.2016.3467

Sun, X., Dai, G., Yu, L., Hu, Q., Chen, J., and Guo, W. (2018). MiR-143-3p inhibits the proliferation, migration and invasion in osteosarcoma by targeting FOSL2. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-017-18739-3

Swarbrick, A., Woods, S. L., Shaw, A., Balakrishnan, A., Phua, Y., Nguyen, A., et al. (2010). miR-380-5p represses p53 to control cellular survival and is associated with poor outcome in MYCN-amplified neuroblastoma. *Nat. Med.* 16, 1134–1140. doi: 10.1038/nm.2227

Sylvestre, Y., De Guire, V., Querido, E., Mukhopadhyay, U. K., Bourdeau, V., Major, F., et al. (2007). An E2F/miR-20a autoregulatory feedback loop. *J. Biol. Chem.* 282, 2135–2143. doi: 10.1074/jbc.M608939200

Tee, A. E., Ling, D., Nelson, C., Atmadibrata, B., Dinger, M. E., Xu, N., et al. (2014). The histone demethylase JMJD1A induces cell migration and invasion by up-regulating the expression of the long noncoding RNA MALAT1. *Oncotarget* 5, 10000. doi: 10.18632/oncotarget.1785

Thayanithy, V., Sarver, A. L., Kartha, R. V., Li, L., Angstadt, A. Y., Breen, M., et al. (2012). Perturbation of 14q32 miRNAs-cMYC gene network in osteosarcoma. *Bone* 50, 171–181. doi: 10.1016/j.bone.2011.10.012

Tivnan, A., Foley, N. H., Tracey, L., Davidoff, A. M., and Stallings, R. L. (2010). MicroRNA-184-mediated inhibition of tumour growth in an orthotopic murine model of neuroblastoma. *Anticancer Res.* 30, 4391–4395.

Tivnan, A., Tracey, L., Buckley, P. G., Alcock, L. C., Davidoff, A. M., and Stallings, R. L. (2011). MicroRNA-34a is a potent tumor suppressor molecule in vivo in neuroblastoma. *BMC Cancer* 11, 33. doi: 10.1186/1471-2407-11-33

Torrezan, G. T., Ferreira, E. N., Nakahata, A. M., Barros, B. D. F., Castro, M. T. M., Correa, B. R., et al. (2014). Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. *Nat. Commun.* 5, 4039. doi: 10.1038/ncomms5039

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., et al. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science (80-.)* 329, 689–693. doi: 10.1126/science.1192002

Ulaner, G. A., Vu, T. H., Li, T., Hu, J. F., Yao, X. M., Yang, Y., et al. (2003). Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site. *Hum. Mol. Genet.* 12, 535–549. doi: 10.1093/hmg/ddg034

Urbach, A., Yermalovich, A., Zhang, J., Spina, C. S., Zhu, H., Perez-Atayde, A. R., et al. (2014). Lin28 sustains early renal progenitors and induces Wilms tumor. *Genes Dev.* 28, 971–982. doi: 10.1101/gad.237149.113

Vázquez-Arreguín, K., and Tantin, D. (2016). The Oct1 transcription factor and epithelial malignancies: old protein learns new tricks. *Biochim. Biophys. Acta* 1859, 792–804. doi: 10.1016/j.bbagrm.2016.02.007

Ventura, A., Young, A. G., Winslow, M. M., Lintault, L., Meissner, A., Erkeland, S. J., et al. (2008). Targeted deletion reveals essential and overlapping functions of the miR-17~92 family of miRNA clusters. *Cell* 132, 875–886. doi: 10.1016/j.cell.2008.02.019

Ventura, S., Aryee, D. N., Felicetti, F., De Feo, A., Mancarella, C., Manara, M. C., et al. (2016). CD99 regulates neural differentiation of Ewing sarcoma cells through miR-34a-Notch–mediated control of NF-κB signaling. *Oncogene* 35, 3944–3954. doi: 10.1038/onc.2015.463

Veronese, A., Lupini, L., Consiglio, J., Visone, R., Ferracin, M., Fornari, F., et al. (2010). Oncogenic role of miR-483-3p at the IGF2/483 locus. *Cancer Res.* 70, 3140–3149. doi: 10.1158/0008-5472.CAN-09-4456

von Frowein, J., Hauck, S. M., Kappler, R., Pagel, P., Fleischmann, K. K., Magg, T., et al. (2018). MiR-492 regulates metastatic properties of hepatoblastoma via CD44. *Liver Int.* 38, 1280–1291. doi: 10.1111/liv.13687

Walz, A. L., Ooms, A., Gadd, S., Gerhard, D. S., Smith, M. A., Guidry Auvil, J. M., et al. (2015). Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. *Cancer Cell* 27, 286–297. doi: 10.1016/j.ccell.2015.01.004

Wang, L.-L., Hu, H.-F., and Feng, Y.-Q. (2016). Suppressive effect of microRNA-143 in retinoblastoma. *Int. J. Ophthalmol.* 9, 1584–1590. doi: 10.18240/ijo.2016.11.08

Wang, H.-F., Zhang, Y.-Y., Zhuang, H.-W., and Xu, M. (2017). MicroRNA-613 attenuates the proliferation, migration and invasion of Wilms' tumor via targeting FRS2. *Eur. Rev. Med. Pharmacol. Sci.* 21, 3360–3369.

Wang, Y., Zhang, Y., Yang, T., Zhao, W., Wang, N., Li, P., et al. (2017a). Long non-coding RNA MALAT1 for promoting metastasis and proliferation by acting as a ceRNA of miR-144-3p in osteosarcoma cells. *Oncotarget* 8, 59417–59434. doi: 10.18632/oncotarget.19727

Wang, Y., Yang, T., Zhang, Z., Lu, M., Zhao, W., Zeng, X., et al. (2017b). Long non-coding RNA TUG1 promotes migration and invasion by acting as a ceRNA of miR-335-5p in osteosarcoma cells. *Cancer Sci.* 108, 859–867. doi: 10.1111/cas.13201

Wang, J. X., Yang, Y., and Li, K. (2018). Long noncoding RNA DANCR aggravates retinoblastoma through miR-34c and miR-613 by targeting MMP-9. *J. Cell. Physiol.* 233, 6986–6995. doi: 10.1002/jcp.26621

Wang, K., Yan, L., and Lu, F. (2018). MiR-363-3p inhibits osteosarcoma cell proliferation and invasion via targeting SOX4. *Oncol. Res.* 5, 157–163. doi: 10.3727/096504018X15190861873459

Wang, W.-T., Qi, Q., Zhao, P., Li, C.-Y., Yin, X.-Y., and Yan, R.-B. (2018). miR-590-3p is a novel microRNA which suppresses osteosarcoma progression by targeting SOX9. *Biomed. Pharmacother.* 107, 1763–1769. doi: 10.1016/j.biopha.2018.06.124

Wang, B., Qu, X. L., Liu, J., Lu, J., and Zhou, Z. Y. (2019). HOTAIR promotes osteosarcoma development by sponging miR-217 and targeting ZEB1. *J. Cell. Physiol.* 234, 6173–6181. doi: 10.1002/jcp.27394

Wegert, J., Ishaque, N., Vardapour, R., Geörg, C., Gu, Z., Bieg, M., et al. (2015). Mutations in the SIX1/2 Pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type Wilms tumors. *Cancer Cell* 27, 298–311. doi: 10.1016/j.ccell.2015.01.002

Wei, J. S., Song, Y. K., Durinck, S., Chen, Q. R., Cheuk, A. T. C., Tsang, P., et al. (2008). The MYCN oncogene is a direct target of miR-34a. *Oncogene* 27, 5204–5213. doi: 10.1038/onc.2008.154

Welch, C., Chen, Y., and Stallings, R. L. (2007). MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene* 26, 5017–5022. doi: 10.1038/sj.onc.1210293

Wen, J., Zhao, Y. K., Liu, Y., and Zhao, J. F. (2017). MicroRNA-34a inhibits tumor invasion and metastasis in osteosarcoma partly by effecting C-IAP2 and Bcl-2. *Tumor Biol.* 39, 101042831770567. doi: 10.1177/1010428317705761

Williamson, D., Lu, Y. J., Gordon, T., Sciot, R., Kelsey, A., Fisher, C., et al. (2005). Relationship between MYCN copy number and expression in rhabdomyosarcomas and correlation with adverse prognosis in the alveolar subtype. *J. Clin. Oncol.* 23, 880–888. doi: 10.1200/JCO.2005.11.078

Wittmann, S., Zirn, B., Alkassar, M., Ambros, P., Graf, N., and Gessler, M. (2007). Loss of 11q and 16q in Wilms tumors is associated with anaplasia, tumor recurrence, and poor prognosis. *Genes Chromosom. Cancer* 46, 163–170. doi: 10.1002/gcc.20397

Wu, M. K., Sabbaghian, N., Xu, B., Addidou-Kalucki, S., Bernard, C., Zou, D., et al. (2013). Biallelic DICER1 mutations occur in Wilms tumours. *J. Pathol.* 230, 154–164. doi: 10.1002/path.4196

Wu, X., Zhong, D., Gao, Q., Zhai, W., Ding, Z., and Wu, J. (2013). MicroRNA-34a inhibits human osteosarcoma proliferation by downregulating ether à go-go 1 expression. *Int. J. Med. Sci.* 10, 676–682. doi: 10.7150/ijms.5528

Wu, L., Nguyen, L. H., Zhou, K., De Soysa, T. Y., Li, L., Miller, J. B., et al. (2015). Precise let-7 expression levels balance organ regeneration against tumor suppression. *Elife* 4, 1–16. doi: 10.7554/eLife.09431

Xiang, X., Mei, H., Qu, H., Zhao, X., Li, D., Song, H., et al. (2015). miRNA-584-5p exerts tumor suppressive functions in human neuroblastoma through repressing transcription of matrix metalloproteinase 14. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1852, 1743–1754. doi: 10.1016/j.bbadis.2015.06.002

Xiao, T., Zhou, Y., Li, H., Xiong, L., Wang, J., Wang, Z.-H., et al. (2019). MiR-125b suppresses the carcinogenesis of osteosarcoma cells via the MAPK–STAT3 pathway. *J. Cell. Biochem.* 120, 2616–2626. doi: 10.1002/jcb.27568

Xie, C.-H., Cao, Y.-M., Huang, Y., Shi, Q.-W., Guo, J.-H., et al. (2016). Long non-coding RNA TUG1 contributes to tumorigenesis of human osteosarcoma by sponging miR-9-5p and regulating POU2F1 expression. *Tumor Biol.* 37, 15031–15041. doi: 10.1007/s13277-016-5391-5

Xin, C., Buhe, B., Hongting, L., Chuanmin, Y., Xiwei, H., Hong, Z., et al. (2013). MicroRNA-15a promotes neuroblastoma migration by targeting reversion-inducing cysteine-rich protein with Kazal motifs (RECK) and regulating matrix metalloproteinase-9 expression. *FEBS J.* 280, 855–866. doi: 10.1111/febs.12074

Yan, K., Gao, J., Yang, T., Ma, Q., Qiu, X., Fan, Q., et al. (2012). MicroRNA-34a inhibits the proliferation and metastasis of osteosarcoma cells both in vitro and in vivo. *PLoS One* 7, e33778. doi: 10.1371/journal.pone.0033778

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., et al. (2011). ncRNA- and Pc2 methylation–dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* 147, 773–788. doi: 10.1016/j.cell.2011.08.054

Yang, F., Bi, J., Xue, X., Zheng, L., Zhi, K., Hua, J., et al. (2012). Up-regulated long non-coding RNA H19 contributes to proliferation of gastric cancer cells. *FEBS J.* 279, 3159–3165. doi: 10.1111/j.1742-4658.2012.08694.x

Yang, G., Yuan, J., and Li, K. (2013). EMT transcription factors: implication in osteosarcoma. *Med. Oncol.* 30, 1–5. doi: 10.1007/s12032-013-0697-2

Yang, L., Li, Y., Wei, Z., and Chang, X. (2017). Coexpression network analysis identifies transcriptional modules associated with genomic alterations in neuroblastoma. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1864, 2341–2348. doi: 10.1016/j.bbadis.2017.12.020

Yang, G., Fu, Y., Lu, X., Wang, M., Dong, H., and Li, Q. (2018). LncRNA HOTAIR/miR-613/c-met axis modulated epithelial–mesenchymal transition of retinoblastoma cells. *J. Cell. Mol. Med.* 22, 5083–5096. doi: 10.1111/jcmm.13796

Yang, H., Peng, Z., Liang, M., Zhang, Y., Wang, Y., Jiang, Y., et al. (2018). The miR-17-92 cluster/QKI2/β-catenin osteosarcoma progression axis promotes 9, 25285–25293. doi: 10.18632/oncotarget.23935

Ye, C., Yu, X., Liu, X., Dai, M., and Zhang, B. (2018). miR-30d inhibits cell biological progression of Ewing's sarcoma by suppressing the MEK/ERK and PI3K/Akt pathways in vitro. *Oncol. Lett.* 15, 4390–4396. doi: 10.3892/ol.2018.7900

Yin, Z., Ding, H., He, E., Chen, J., and Li, M. (2016). Overexpression of long non-coding RNA MFI2 promotes cell proliferation and suppresses apoptosis in human osteosarcoma. *Oncol. Rep.* 36, 2033–2040. doi: 10.3892/or.2016.5013

Yin, J., Zhao, J., Hu, W., Yang, G., Yu, H., Wang, R., et al. (2017). Disturbance of the let-7/LIN28 double negative feedback loop is associated with radio- and chemo-resistance in non–small cell lung cancer. *PLoS One* 12, 1–16. doi: 10.1371/journal.pone.0172787

Ying, L., Chen, Q., Wang, Y., Zhou, Z., Huang, Y., and Qiu, F. (2012). Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition. *Mol. Biosyst.* 8, 2289–2294. doi: 10.1039/c2mb25070e

Yuan, J., Lang, J., Liu, C., Zhou, K., Chen, L., and Liu, Y. (2015). The expression and function of miRNA-451 in osteosarcoma. *Med. Oncol.* 32, 324. doi: 10.1007/s12032-014-0324-x

Zhang, Y., Shields, T., Crenshaw, T., Hao, Y., Moulton, T., and Tycko, B. (1993). Imprinting of human H19: allele-specific CpG methylation, loss of the active allele in Wilms tumor, and potential for somatic allele switching. *Am. J. Hum. Genet.* 53, 113–124.

Zhang, B., Arun, G., Mao, Y. S., Lazar, Z., Hung, G., Bhattacharjee, G., et al. (2012). The lncRNA malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.* 2, 111–123. doi: 10.1016/j.celrep.2012.06.003

Zhang, H., Pu, J., Qi, T., Qi, M., Yang, C., Li, S., et al. (2014). MicroRNA-145 inhibits the growth, invasion, metastasis and angiogenesis of neuroblastoma cells through targeting hypoxia-inducible factor 2 alpha. *Oncogene* 33, 387–397. doi: 10.1038/onc.2012.574

Zhang, A., Shang, W., Nie, Q., Li, T., and Li, S. (2018). Long non-coding RNA H19 suppresses retinoblastoma progression via counteracting miR-17-92 cluster. *J. Cell. Biochem.* 119, 3497–3509. doi: 10.1002/jcb.26521

Zhang, F., Zhu, Y., Fan, G., and Hu, S. (2018). MicroRNA-2682-3p inhibits osteosarcoma cell proliferation by targeting CCND2, MMP8 and Myd88. *Oncol. Lett.* 16, 3359–3364. doi: 10.3892/ol.2018.9029

Zhang, S., Li, D., Jiao, G.-J., Wang, H.-L., and Yan, T.-B. (2018). miR-185 suppresses progression of Ewing's sarcoma via inhibiting the PI3K/AKT and Wnt/β-catenin pathways. *Onco. Targets. Ther.* 11, 7967–7977. doi: 10.2147/OTT.S167771

Zhang, Y., Zhao, Y., Wu, J., Liangpunsakul, S., Niu, J., and Wang, L. (2018). MicroRNA-26-5p functions as a new inhibitor of hepatoblastoma by repressing lin-28 homolog B and aurora kinase a expression. *Hepatol. Commun.* 2, 861–871. doi: 10.1002/hep4.1185

Zhao, D., Tian, Y., Li, P., Wang, L., Xiao, A., Zhang, M., et al. (2015). MicroRNA-203 inhibits the malignant progression of neuroblastoma by targeting Sam68. *Mol. Med. Rep.* 12, 5554–5560. doi: 10.3892/mmr.2015.4013

Zhao, Y., Sun, H., and Wang, H. (2016). Long noncoding RNAs in DNA methylation: new players stepping into the old game. *Cell Biosci.* 6, 1–6. doi: 10.1186/s13578-016-0109-3

Zhao, Y., Ling, Z., Hao, Y., Pang, X., Han, X., Califano, J. A., et al. (2017). MiR-124 acts as a tumor suppressor by inhibiting the expression of sphingosine kinase 1 and its downstream signaling in head and neck squamous cell carcinoma. *Oncotarget* 8, 25005–25020. doi: 10.18632/oncotarget.15334

Zhi, F., Wang, R., Wang, Q., Xue, L., Deng, D., Wang, S., et al. (2014). MicroRNAs in neuroblastoma: small-sized players with a large impact. *Neurochem. Res.* 39, 613–623. doi: 10.1007/s11064-014-1247-9

Zhou, J., Yang, L., Zhong, T., Mueller, M., Men, Y., Zhang, N., et al. (2015). H19 lncRNA alters DNA methylation genome wide by regulating S-adenosylhomocysteine hydrolase. *Nat. Commun.* 6, 1–13. doi: 10.1038/ncomms10221

Zhu, S. W., Li, J. P., Ma, X. L., Ma, J. X., Yang, Y., Chen, Y., et al. (2015). miR-9 modulates osteosarcoma cell growth by targeting the GCIP tumor suppressor. *Asian Pacific J. Cancer Prev.* 16, 4509–4513. doi: 10.7314/APJCP.2015.16.11.4509

Zhu, Z., Tang, J., Wang, J., Duan, G., Zhou, L., and Zhou, X. (2016). MIR-138 acts as a tumor suppressor by targeting EZH2 and enhances cisplatin-induced apoptosis in osteosarcoma cells. *PLoS One* 11, 1–12. doi: 10.1371/journal.pone.0150026

Zhu, C., Cheng, D., Qiu, X., Zhuang, M., and Liu, Z. (2018). Long noncoding RNA SNHG16 promotes cell proliferation by sponging microRNA-205 and upregulating ZEB1 expression in osteosarcoma. *Cell. Physiol. Biochem.* 51, 429–440. doi: 10.1159/000495239

Zhu, S., Fu, W., Zhang, L., Fu, K., Hu, J., Jia, W., et al. (2018a). LINC00473 antagonizes the tumour suppressor miR-195 to mediate the pathogenesis of Wilms tumour via IKKα. *Cell Prolif.* 51, 1–10. doi: 10.1111/cpr.12416

Zhu, S., Zhang, L., Zhao, Z., Fu, W., Fu, K., Liu, G., et al. (2018b). MicroRNA-92a-3p inhibits the cell proliferation, migration and invasion of Wilms tumor by targeting NOTCH1. *Oncol. Rep.* 40, 571–578. doi: 10.3892/or.2018.6458

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership