REGULATORY RNAs IN THE NUCLEUS

EDITED BY: Xiao Li, Liang Chen, Wenbo Li, Yuanchao Xue and Bing Zhou
PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-740-8
DOI 10.3389/978-2-88971-740-8

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

REGULATORY RNAs IN THE NUCLEUS

Topic Editors:
Xiao Li, Texas Heart Institute, United States
Liang Chen, Wuhan University, China
Wenbo Li, University of Texas Health Science Center at Houston, United States
Yuanchao Xue, Chinese Academy of Sciences (CAS), China
Bing Zhou, Chinese Academy of Sciences (CAS), China

Topic Editors Xiao Li, Bing Zhou, and Wenbo Li hold patents related to the Research Topic subject. All other Topic Editors declare no competing interests.

Citation: Li, X., Chen, L., Li, W., Xue, Y., Zhou, B., eds. (2021). Regulatory RNAs in the Nucleus. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-740-8

Table of Contents

04 ***Genome-Wide Transcriptional Regulation of the Long Non-coding RNA Steroid Receptor RNA Activator in Human Erythroblasts***
Waritta Sawaengdee, Kairong Cui, Keji Zhao, Suradej Hongeng, Suthat Fucharoen and Patompon Wongtrakoongate

19 ***Deregulation of a* Cis*-Acting lncRNA in Non-small Cell Lung Cancer May Control HMGA1 Expression***
Greg L. Stewart, Adam P. Sage, Katey S. S. Enfield, Erin A. Marshall, David E. Cohn and Wan L. Lam

28 ***LncRNA H19 Suppresses Osteosarcomagenesis by Regulating snoRNAs and DNA Repair Protein Complexes***
An Xu, Mo-Fan Huang, Dandan Zhu, Julian A. Gingold, Danielle A. Bazer, Betty Chang, Donghui Wang, Chien-Chen Lai, Ihor R. Lemischka, Ruiying Zhao and Dung-Fang Lee

39 ***Effects of Circular RNA of Chicken Growth Hormone Receptor Gene on Cell Proliferation***
Haidong Xu, Qiying Leng, Jiahui Zheng, Patricia Adu-Asiamah, Shudai Lin, Ting Li, Zhang Wang, Lilong An, Zhuihui Zhao and Li Zhang

49 ***Functional Network of the Long Non-coding RNA Growth Arrest-Specific Transcript 5 and Its Interacting Proteins in Senescence***
Siqi Wang, Shengwei Ke, Yueming Wu, Duo Zhang, Baowei Liu, Yao-hui He, Wen Liu, Huawei Mu and Xiaoyuan Song

61 ***Comprehensive Identification and Alternative Splicing of Microexons in* Drosophila**
Ting-Lin Pang, Zhan Ding, Shao-Bo Liang, Liang Li, Bei Zhang, Yu Zhang, Yu-Jie Fan and Yong-Zhen Xu

71 ***Complex RNA Secondary Structures Mediate Mutually Exclusive Splicing of Coleoptera* Dscam1**
Haiyang Dong, Lei Li, Xiaohua Zhu, Jilong Shi, Ying Fu, Shixin Zhang, Yang Shi, Bingbing Xu, Jian Zhang, Feng Shi and Yongfeng Jin

83 ***LncRNAs: Architectural Scaffolds or More Potential Roles in Phase Separation***
Jie Luo, Lei Qu, Feiran Gao, Jun Lin, Jian Liu and Aifu Lin

96 ***7SK Acts as an Anti-tumor Factor in Tongue Squamous Cell Carcinoma***
Bowen Zhang, Sainan Min, Qi Guo, Yan Huang, Yuzhu Guo, Xiaolin Liang, Li-ling Wu, Guang-yan Yu and Xiangting Wang

106 ***Epigenetic Regulation of the Vascular Endothelium by Angiogenic LncRNAs***
Noeline Subramaniam, Ranju Nair and Philip A. Marsden

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 August 2020
doi: 10.3389/fgene.2020.00850

Genome-Wide Transcriptional Regulation of the Long Non-coding RNA Steroid Receptor RNA Activator in Human Erythroblasts

*Waritta Sawaengdee[1], Kairong Cui[2], Keji Zhao[2], Suradej Hongeng[3], Suthat Fucharoen[4] and Patompon Wongtrakoongate[1,5]**

[1] Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand, [2] Laboratory of Epigenome Biology, Systems Biology Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, United States, [3] Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, [4] Thalassemia Research Center, Institute of Molecular Biosciences, Mahidol University, Bangkok, Thailand, [5] Center for Neuroscience, Faculty of Science, Mahidol University, Bangkok, Thailand

OPEN ACCESS

Edited by:
Wenbo Li,
The University of Texas Health Science Center at Houston, United States

Reviewed by:
Argyris Papantonis, University Medical Center Göttingen, Germany
Punit Prasad, Institute of Life Sciences (ILS), India

***Correspondence:**
Patompon Wongtrakoongate
patompon.won@mahidol.ac.th

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: 09 May 2020
Accepted: 13 July 2020
Published: 11 August 2020

Citation:
Sawaengdee W, Cui K, Zhao K, Hongeng S, Fucharoen S and Wongtrakoongate P (2020) Genome-Wide Transcriptional Regulation of the Long Non-coding RNA Steroid Receptor RNA Activator in Human Erythroblasts. Front. Genet. 11:850. doi: 10.3389/fgene.2020.00850

Erythropoiesis of human hematopoietic stem cells (HSCs) maintains generation of red blood cells throughout life. However, little is known how human erythropoiesis is regulated by long non-coding RNAs (lncRNAs). By using ChIRP-seq, we report here that the lncRNA steroid receptor RNA activator (SRA) occupies chromatin, and co-localizes with CTCF, H3K4me3, and H3K27me3 genome-wide in human erythroblast cell line K562. CTCF binding sites that are also occupied by SRA are enriched for either H3K4me3 or H3K27me3. Transcriptome-wide analyses reveal that SRA facilitates expression of erythroid-associated genes, while repressing leukocyte-associated genes in both K562 and CD36-positive primary human proerythroblasts derived from HSCs. We find that SRA-regulated genes are enriched by both CTCF and SRA bindings. Further, silencing of SRA decreases expression of the erythroid-specific markers TFRC and GYPA, and down-regulates expression of globin genes in both K562 and human proerythroblast cells. Taken together, our findings establish that the lncRNA SRA occupies chromatin, and promotes transcription of erythroid genes, therefore facilitating human erythroid transcriptional program.

Keywords: steroid receptor RNA activator, erythroblasts, histone modification, epigenetics, stem cells

INTRODUCTION

Adult erythropoiesis is a cellular physiological process in the bone marrow which produces red blood cells (RBCs) to maintain homeostasis of the body. Through the proerythroblast stage producing transit amplifying cells, billions of RBCs are spatiotemporally generated by hematopoietic stem cells (HSCs). Intrinsic, extrinsic, and environmental factors play crucial roles in this process to precisely control a sufficient quantity of the oxygen-carrying cells that are required for oxygen transport. Among key intrinsic factors regulating erythropoiesis are chromatin binding proteins including transcriptional and epigenetic machineries. At the onset of erythropoiesis, a sequential activation of DNA-binding transcription factors such as GATA1, TAL1, and KLF1 delineates the gradual development of erythroid cells (Wickrema and Crispino, 2007).

The histone modifying complexes trithorax group (TrxG) and polycomb repressive complex 2 (PRC2), which methylate H3K4, and H3K27, respectively, are also critical for normal erythropoiesis (Majewski et al., 2008; Gan et al., 2010; Mochizuki-Kashio et al., 2011). Yet, little is known about how distinct transcription and epigenetic factors are recruited or tethered to chromatin. Thus characterization of mechanisms involved in genetic-epigenetic crosstalk is essential to understand erythropoiesis.

A significant role has been discovered for long non-coding RNAs (lncRNAs) in transcriptional control (Rinn and Chang, 2012). The lncRNA steroid receptor RNA activator (SRA) was identified as a non-coding transcript which promotes transcriptional activation of the estrogen receptors (Lanz et al., 1999, 2002). A role of SRA has been reported in regulation of imprinted gene expression via the chromatin architectural transcription factor CTCF and SRA-associated RNA helicase DDX5 (Yao et al., 2010). Moreover, we have also shown that SRA physically and directly interacts with NANOG, CTCF, TrxG, and PRC2, and that SRA is important for maintenance of pluripotency and transition into induced pluripotent stem cells (Wongtrakoongate et al., 2015). Whether SRA participates in regulation of erythropoiesis has been elusive. In the present study, we report a novel function of the lncRNA SRA in regulation of global gene expression through direct chromatin binding in human erythroleukemia cell line K562 and in primary human proerythroblasts derived from HSCs. We demonstrate that SRA, together with CTCF, H3K4me3, and H3K27me3, occupies various genomic regions in K562. Further, SRA facilitates transcriptome-wide expression of erythroid program and expression of erythroid markers in K562 and in primary human proerythroblasts. Hence, a possible function of the lncRNA SRA is to promote transcription of erythroid-associated genes.

MATERIALS AND METHODS

Cell Culture

The cell line K562 (ATCC) was cultured in RPMI 1640 medium with GlutaMAX^{TM} (Invitrogen) supplemented with 10% fetal bovine serum at 37°C with 5% CO_2 and passaged every 3 days. CD36-positive human proerythroblasts were derived from bone marrow CD34-positive cells, which were purchased from Stem Cell Technologies (70002.1) and cultured in erythroid differentiation condition as previous described (Wong et al., 2008). Briefly, the CD34-positive cells at 10^4 cells/ml were grown in the serum-free erythroid expansion medium containing Alpha minimum essential medium (AMEM; Mediatech) and 20% BIT9500 (Stem Cell Technologies) to achieve bovine serum albumin, recombinant human insulin and iron-saturated human transferrin at 10 mg/ml, 10 μg/ml, and 200 μg/ml, respectively. In addition, 900 ng/ml ferrous sulfate (Sigma), 90 ng/ml ferric nitrate (Sigma), 1 μM hydrocortisone (Sigma), 100 ng/ml of recombinant human stem cell factor (SCF; Stem Cell Technologies), 5 ng/ml of recombinant human interleukin-3 (IL-3; R&D Systems), and 3 IU/ml of recombinant human EPO were also included. Fresh medium was added into the culture to maintain cells at 2×10^6 cells/ml. The cells were cultured for 7 days to obtain CD36-positive cells.

RNA Silencing

shRNA templates including luciferase shRNA scramble control (shLuc) and shRNA sequences targeting *SRA* transcript (shSRA-1: 5′ CCACAAGTTTCCCAGTCGAGT 3′, shSRA-2: 5′ TGCAGCCACAGCTGAGAAGAA 3′, and shSRA-3: 5′ ACTGAGGTCAGTCAGTGGAT 3′) were individually cloned into the lentiviral vector pGreenPuro (System Biosciences) at *BamH*1/*Eco*RI restriction sites according to the manufacturer's instruction. The plasmids were transformed into *Escherichia coli* strain *stbl3* (Invitrogen) via heat shock method and propagated in LB broth supplemented with carbenicillin. All plasmids were purified by using PureLink^{TM} HiPure Plasmid Maxiprep Kit (Invitrogen).

Lentiviral particles were produced by co-transfecting LentiX-293T cells (Clontech) with a packaging vector (psPAX2), an envelope vector (pLP/VSVG), and an shRNA plasmid (shLuc, shSRA-1, shSRA-2, or shSRA-3) using Lipofectamine 2000 (Invitrogen) as previous described (Kidder et al., 2017). Twenty four hour after the transfection, the medium was changed to the target cell medium. Then the medium containing lentiviral particles were collected and filtered through 0.22 μM filter at 48 h post-transfection. Transduction was performed by adding the medium containing lentiviral particles with 10 μg/ml polybrene into either K562 or CD36-positive proerythroblasts. The cells were then centrifuged at 1,000 *g* at room temperature for 2 h and incubated at 37°C with 5% CO_2 overnight before changing medium. Expression of GFP was examined under fluorescent microscope to validate transduction efficiency at 48 h post-transduction. The GFP-positive cells were then sorted by FACS at 96 h post-transduction and maintained in the presence of 0.5 μg/ml puromycin for further analysis.

RNA-Sequencing Analysis

Total RNA was extracted from the sorted cells and purified using QIAzol Lysis Reagent and miRNeasy Micro Kit (Qiagen). RNA samples for sequencing were prepared according to Smart-seq2 method (Picelli et al., 2014) with some modifications as previously described (Hu et al., 2018). RNA-seq libraries were prepared with an End-It DNA End-repair Kit (Epicenter) and a Multiplexing Sample Preparation Oligonucleotide Kit (Illumina), and the libraries submitted for single-end sequencing on the Illumina HiSeq2500 (Hu et al., 2013). Sequencing data was annotated to the human reference genome GRch38 by Tophat2 (Kim et al., 2013) with Bowtie2 (Langmead and Salzberg, 2012), and the raw read counts and FPKMs were acquired by HTseq (Anders et al., 2015) and Cufflinks (Trapnell et al., 2010), respectively. TPMs were then calculated according to their FPKMs. Using raw read count from HTseq as input, differentially expressed genes (DEGs) were determined by DEseq package (Anders and Huber, 2010) with p-value < 0.01 and with TPM > 2. Three different SRA silencing samples targeted for individual shRNA target sites were used as three biological replicates. Accession number of sequencing data associated

with RNA-seq is GSE151926. Volcano plots and heatmaps were generated by R studio (RStudio Team, 2016) using EnhancedVolcano and gplots package, respectively (Blighe et al., 2020). Gene ontology (GO) enrichment and KEGG pathways were determined by DIVID software (Huang et al., 2008, 2009) using the DEGs from DEseq. Network analysis of coding genes from DEGs was performed using STRING or Search Tool for the Retrieval of Interacting Genes/Proteins (Szklarczyk et al., 2015).

Quantitative Real-Time PCR Analysis

RNA was extracted and purified using QIAzol Lysis Reagent and miRNeasy Micro Kit (Qiagen). Reverse transcription was carried on with 1 μg RNA using iScript™ Reverse Transcription Supermix for RT-qPCR (Biorad). qRT-PCR was performed by using KAPA SYBR® FAST qPCR Master Mix (2X) Kit (Kapa Biosystems) with LightCycler® 96 system (Roche). *ACTB* gene was used for normalization of gene expression and the ΔΔCt method was used for analysis of relative expression level. Primer sequences are available upon requested.

Flow Cytometry Analysis

One million cells and five million cells were used for surface marker analysis and cell sorting, respectively. Cells were collected and resuspended in 100 μl PBS with 2% FBS. Fluorescent conjugated antibodies were added into the cell suspension and incubated at 4°C for 30 min in dark. The stained cells were washed once and resuspended in the PBS/FBS buffer before analysis or sorting with FACSAria II cell sorter (BD Biosciences). Data were analyzed with FlowJo software. Unstained wild-type cells and the Fluorescence Minus One (FMO) controls were used as negative control for gating population. The antibodies using in the experiment include APC-conjugated anti-human CD235a (eBioscience, 17-9987-41), PE-conjugated anti-human CD34 (eBioscience, 12-0349-41), and PerCP-eFluor710-conjugated anti-human CD36 (eBioscience, 46-0369-41).

Chromatin Isolation by RNA Purification (ChIRP)

Chromatin isolation by RNA purification (ChIRP) analysis was performed as previously described with minor modifications (Chu et al., 2011, 2012; Wongtrakoongate et al., 2015). The cell line K562 harvested at 3×10^7 cells were fixed with 1% glutaraldehyde for 10 min at room temperature with a rotator and then stopped by adding glycine solution at 125 mM of its final concentration. Crosslinked cells were washed with PBS, and resuspended in 1 ml swelling buffer. Samples were incubated at 4°C for 30 min with a rotator. Cell pellet was collected by centrifugation and resuspended with 350 μL of ChIRP lysis buffer. Cell sonication was performed using a Bioruptor (Diagenode) at maximum power, 30 s ON and 30 s OFF for 7.5 min of 6 cycles to obtain chromatin fragments ranging from 100–1000 bp. Chromatin fragments was then collected by centrifugation. Two hundred micrograms of sheared chromatin samples were pre-cleared using 100 μL of Ultralink-streptavidin beads (Pierce) for 1 h at room temperature with a rotator, and supernatant was collected. The pre-cleared chromatin was used per hybridization reaction with 10 μL of 100 μM pooled 3′ Biotin TEG oligonucleotide probes (Integrated DNA Technologies) against SRA transcript (Wongtrakoongate et al., 2015). LacZ probes were employed as negative control (Chu et al., 2011). The sample and the probes were hybridized at 37°C for 4 h with a rotator. Once the hybridization was completed, 100 μL of C-1 magnetic beads (Invitrogen) was mixed with the sample to pull down the biotinylated probes. DNA was eluted in the presence of 12.5 mM D-Biotin (Invitrogen). DNA was ethanol precipitated and subjected to library preparation, which was performed using MicroPlex Library Preparation Kit (Diagenode) according to manufacturer's instruction. Three biological triplicates were used for ChIRP-seq. Briefly, 5–10 ng of DNA starting material, which was quantified by Qubit (Invitrogen), was used for each biological sample. The DNA was end-repaired, 3′ adenylated, and ligated with adapters. Then the ligated DNA was size-selected to obtain DNA fragments at 250–300 bp by agarose gel electrophoresis. The purified DNA was amplified to enrich the library. The final PCR product was purified by Agencourt AMPure XP beads (Beckman Coulter) and was submitted for high-throughput sequencing using Illumina HiSeq2500. The sequencing was performed with the run type of single-end, 50 bp read. Data were aligned against the human genome version human_hg19, and were exported into BAM file format. Accession number of sequencing data associated with ChIRP-seq is GSE153004.

The associated-binding regions of SRA identified from ChIRP-seq and ChIP-seq data for CTCF, H3K4me3, and H3K27me3 in K562 were identified by ChIPpeakAnno package (Zhu et al., 2010; Zhu, 2013) with max gap equals to 500 bp in R studio. The ChIP-seq data were taken from the ENCODE project of K562 cells. The region-associated genes were identified by GREAT (McLean et al., 2010) using two nearest genes' TSS within 50 kb up- and down-stream of the regulatory binding sites including curated regulatory domains. The Fisher's Exact test to measure peak enrichment was taken from the Fisher's exact function from the R package for statistical computing (R Core Team, 2013).

RNA Pull Down

RNA pull down experiments were performed as previously described (Tsai et al., 2010; Wongtrakoongate et al., 2015). The plasmid pLITMUS28i (New England Biolabs) containing full length SRA was linearized by *Stu*I or *Bgl*I to generate antisense or sense SRA transcripts, respectively (Wongtrakoongate et al., 2015). Biotinylated SRA and a maltose-binding protein transcript were *in vitro* transcribed using HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) in the presence of biotin-14-CTP (Invitrogen). Transcribed RNA products were DNase-treated (Roche), and purified by ethanol precipitation. 3 μg of sense SRA, antisense SRA, and MBP RNA was individually prepared in RNA structure buffer (Tris–Cl pH 7.5, 0.1 M KCl, and 10 mM $MgCl_2$) and incubated at 78°C for 3 min. The RNA was then gradually cooled down to 37°C. Five hundred micrograms of K562 nuclear extract, which was prepared by Nuclear Protein Extraction Kit (Pierce), was mixed with the RNA in immunoprecipitation buffer (PBS plus 0.1% Triton X-100, 1 mM DTT, protease inhibitor cocktail, PMSF, and 80 U RNase inhibitor) in a total volume of 500 μL. The reaction was incubated

for 4 h at 4°C with rotation. The RNA-beads complex isolated by MyOne Streptavidin C1 beads was further incubated overnight. Beads were washed five times with immunoprecipitation buffer and boiled with SDS loading buffer for western blot analysis.

Statistical Analysis

For qPCR and FACS, data was analyzed using two-tailed unpaired Student *t*-test and shown as mean with standard deviation of three independent replicates. The significance values were determined at 95%, with $p \leq 0.05$. Differential expressed genes were determined by DEseq package based on the negative binomial distribution with significant at $p < 0.01$. Fisher Exact test was used to determine significant enriched GO terms with $p < 0.05$ using DAVID software.

RESULTS

SRA Co-localizes With CTCF, H3K4me3, or H3K27me3 Genome-Wide

The lncRNA SRA has been shown to mediate transcriptional regulation in several cellular contexts (Colley and Leedman, 2011). We have previously reported that SRA possesses genome-wide binding regions of human pluripotent stem cells (Wongtrakoongate et al., 2015). Yet, little is known for the role of SRA in transcriptional regulation in erythropoiesis. To identify SRA-binding sites of human erythroblasts, we performed SRA ChIRP-seq (Chu et al., 2011; Wongtrakoongate et al., 2015) for the human erythroblast cell line K562. Biotin-conjugated deoxyoligonucleotide probes tiling along SRA (Wongtrakoongate et al., 2015) were hybridized with the lncRNA using sheared chromatin from K562 cells. Using next generation sequencing, we identified 2,790 SRA-binding sites genome-wide; most of which are located within 50 kb upstream or downstream of transcription start site (**Supplementary Figure S1**). Up- or down-stream nearest genes within this 50 kb were queried. Among these 2,790 SRA-bound genomic regions, 1,742 and 1,048 regions representing 62.4 and 37.6% of total SRA binding sites are associated with 2,170 genes and not associated with any nearby genes, respectively. Gene classification analysis reveals that SRA-bound regions are associated with genes involved in cell proliferation, Wnt signaling, NF-κB signaling, regulation of protein phosphorylation, cell differentiation, and metabolisms (**Figure 1A** and **Supplementary Table S1**). Seven genes bound by SRA were also identified as oxygen transport including *HBA2*, *HBB*, *HBD*, *HBG1*, *HBE1*, *HBM*, and *CYGB*. We validated the occupancy of the lncRNA SRA at alpha and beta globin loci using ChIRP followed by real-time PCR analysis, which reveals an association of the lncRNA along both alpha and beta globin loci (**Supplementary Figure S2**). Specifically, at the alpha locus SRA occupies the regulatory element HS40, a site upstream of *HBA2* and a site downstream of the locus (**Supplementary Figures S2A,B**). At the beta locus, SRA occupies *HBB*, *HBG*, *HBE* as well as the locus control region (LCR; **Supplementary Figures S2C,D**). The ChIRP-seq and ChIRP-PCR results therefore indicate the direct association of the lncRNA SRA at chromatin level of human erythroblasts.

We and others have previously reported that SRA directly forms complexes with the chromatin architectural transcription factor CTCF (Yao et al., 2010; Wongtrakoongate et al., 2015), the histone H3 lysine 4 (H3K4) methyltransferase TrxG, and the histone H3 lysine 27 (H3K27) methyltransferase PRC2 (Wongtrakoongate et al., 2015). RNA pull down in K562 cells reveals that sense SRA, but not antisense SRA or MBP transcripts, can pull down the RNA helicase DDX5, the chromatin architectural protein CTCF, the TrxG component WDR5, and the PRC2 member EZH2 (**Supplementary Figure S3**). This result suggests that SRA might interact with TrxG and PRC2 in the cells, supporting our previous finding of a direct physical interaction of SRA/TrxG/PRC2 *in vitro* (Wongtrakoongate et al., 2015). Using ChIP-seq data from ENCODE, we show here that SRA and CTCF co-occupy 540 sites representing 19.3% of SRA binding sites (**Figures 1B–D**). Comparing SRA with profiles of H3K4me3 and H3K27me3 in K562, we find that 465 and 789 sites representing 16.7 and 28.3% of total SRA binding sites possess, respectively, either the H3K4me3 or H3K27me3 modification (**Figures 1C,D**). When comparing SRA, CTCF and the histone modifications, 245 and 192 sites representing about 8.8% and 6.9% of SRA binding regions are also co-occupied by CTCF plus H3K4me3 and CTCF plus H3K27me3, respectively.

Since SRA has been proposed to deliver TrxG or PRC2 to SRA-associated transcription factors including CTCF (Wongtrakoongate et al., 2015), we then asked whether sites of H3K4me3 or H3K27me3 modifications might be enriched at genomic regions occupied by both CTCF and SRA relative to those occupied by CTCF alone. We observe a higher proportion (45.4% versus 20.8%) of CTCF binding sites carrying the H3K4me3 modification at genomic regions occupied by both CTCF and SRA compared with those occupied by CTCF but lacking SRA (**Figure 1C**). Similarly, the presence of SRA at CTCF binding sites correlates with a higher proportion of H3K27me3 modification (35.6% versus 11.5%) (**Figure 1D**). Thus the genome-wide occupancy of H3K4me3 or H3K27me3 at SRA-associated CTCF binding sites suggest a possible role for the lncRNA SRA in transcriptional control of human erythroblast cells.

SRA Regulates Hematopoiesis-Related Genes Transcriptome-Wide in K562

To ascertain whether the lncRNA SRA globally regulates genes of the erythroblasts K562, a lentiviral transduction carrying an shRNA cassette was introduced into the cells. The lncRNA SRA transcript was successfully depleted (**Supplementary Figure S4**). RNA-seq analysis of K562 was then performed upon depletion of SRA using an Illumina HiSeq platform. DEGs with at least two fold-change were subsequently identified by using DEseq. Three individual shRNA knockdown samples, which were transduced with target site-specific shRNA targeting *SRA* transcript, were used as biological replicates for the analysis. Silencing of SRA led to differential expression of 675 genes, with 322 and 353 genes were down- and up-regulated by SRA knockdown, respectively (**Figure 2** and **Supplementary Table S2**). Gene ontology analysis of genes positively controlled by SRA shows

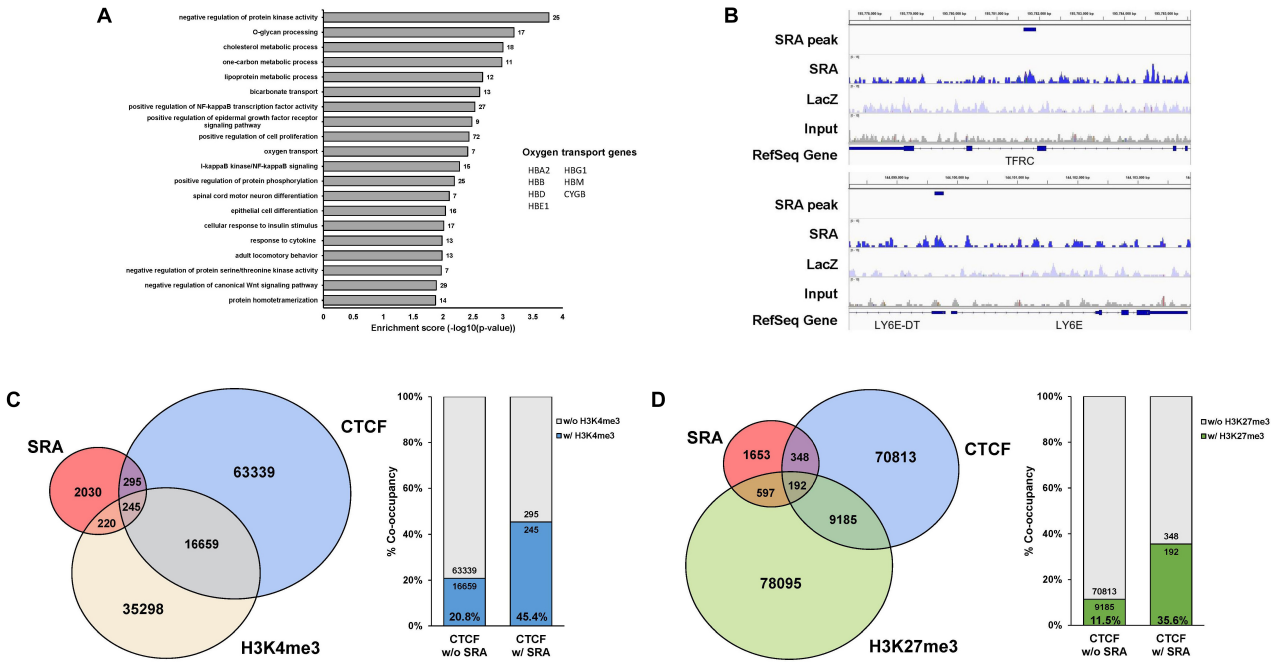


FIGURE 1 | The lncRNA SRA occupies chromatin genome-wide with CTCF, H3K4me3, and H3K27me3. ChIRP-seq analysis of SRA was performed using human erythroblast cells K562. Publicly available ChIP-seq data for CTCF, H3K4me3, and H3K27me3 were derived from the ENCODE project. **(A)** Categories of enriched gene ontologies of SRA-associated genes (p-value < 0.05) and their enrichment scores [−log(p-value)] were analyzed using DAVID. **(B)** Examples of genes with the enrichment signal of SRA, LacZ, and the input. **(C)** Genomic regions occupied by SRA and containing CTCF binding site tend to associate with H3K4me3. Left; Genome-wide SRA-binding sites were compared with CTCF and H3K4me3. Right; Percentage of co-occupancy of H3K4me3 at SRA binding sites without or with CTCF occupancy. **(D)** Genomic regions occupied by SRA and containing CTCF binding site tend to associate with H3K27me3. Left; Genome-wide SRA-binding sites were compared with CTCF and H3K27me3. Right; Percentage of co-occupancy of H3K27me3 at SRA binding sites without or with CTCF occupancy.

that erythroblast-associated pathways such as heme biosynthesis (e.g., *CPOX*, *PPOX,* and *ALAS2*), iron homeostasis (e.g., *TF*, *TFRC,* and *SLC11A2*), cell proliferation (e.g., *TAL1*, *BMX,* and *ERBB3*), and erythrocytes (e.g., *GYPA*, *AQP1,* and *AHSP*) are enriched in the gene groups induced by SRA (**Figure 2B** and **Supplementary Table S3**). Genes belonging to these pathways are clustered as shown by a functional network analysis (**Supplementary Figure S5**). On the other hand, the classification analysis of genes negatively controlled by SRA shows that leukocyte-associated pathways such as immune response (e.g., *FTH1*, *IL1RN,* and *LCP2*), inflammatory response (e.g., *FOS*, *JUN*, *CCL2,* and *AIM2*), and chemotaxis (e.g., *CCL2*, *CCR7,* and *CXCL2*) are enriched in the gene groups repressed by SRA (**Figure 2B** and **Supplementary Table S4**). Genes associated with these pathways are well clustered (**Supplementary Figure S6**).

We and others have previously shown that SRA can form complexes with the chromatin architectural protein CTCF, RNA helicases, TrxG, and PRC2, to control transcription of SRA target genes (Yao et al., 2010; Wongtrakoongate et al., 2015). To further elucidate whether genes differentially expressed by SRA knockdown are occupied by SRA and CTCF at the chromatin level, an association analysis among the DEGs, SRA-bound genes and CTCF-bound genes was performed. There are 221 genes regulated by SRA without CTCF binding compared to 454 genes regulated by SRA with CTCF binding (**Figure 3A**). In addition, we also asked whether SRA-regulated genes might be enriched by both CTCF and SRA binding relative to those with SRA alone. Even though we do not observe genes bound by SRA alone, there is a 14.8% (14.8% versus 0.0%) increase of SRA-regulated genes occupied by both CTCF and SRA compared with those without CTCF binding (**Figure 3A**), indicating a contribution of the transcription factor CTCF in regulation of genes controlled by SRA. To determine to what extent SRA-occupied DEGs also contain H3K4me3 or H3K27me3, we also compared DEGs occupied by SRA in association with H3K4me3 or H3K27me3 (**Figure 3B**). Among genes occupied by SRA, we found that genes up-regulated or down-regulated by SRA tend to associate with H3K4me3 together with H3K27me3 (**Figure 3C**). To a lesser extent, genes up-regulated or down-regulated by SRA also harbor H3K4me3 but not H3K27me3. Collectively, these results suggest the role of SRA in transcriptome-wide regulation of human erythroblast cells.

Next, we confirmed whether SRA silencing affects expression of erythroblast markers of K562. Upon the silencing of SRA, expression of committed erythroid genes *TFRC* and *GYPA* was reduced as determined by real-time PCR (**Figures 4A,D**). Flow cytometry analysis of the two erythroid markers reveals that depletion of SRA led to a decrease in the antigen expression (**Figures 4B,E**) and in the number of cells positive for the markers (**Figures 4C,F**). Further, expression of globin genes including *HBA1/2*, *HBE*, *HBG1/2*, and *HBD* is reduced in K562 cells following SRA depletion (**Figure 4G**). Since the beta globin gene

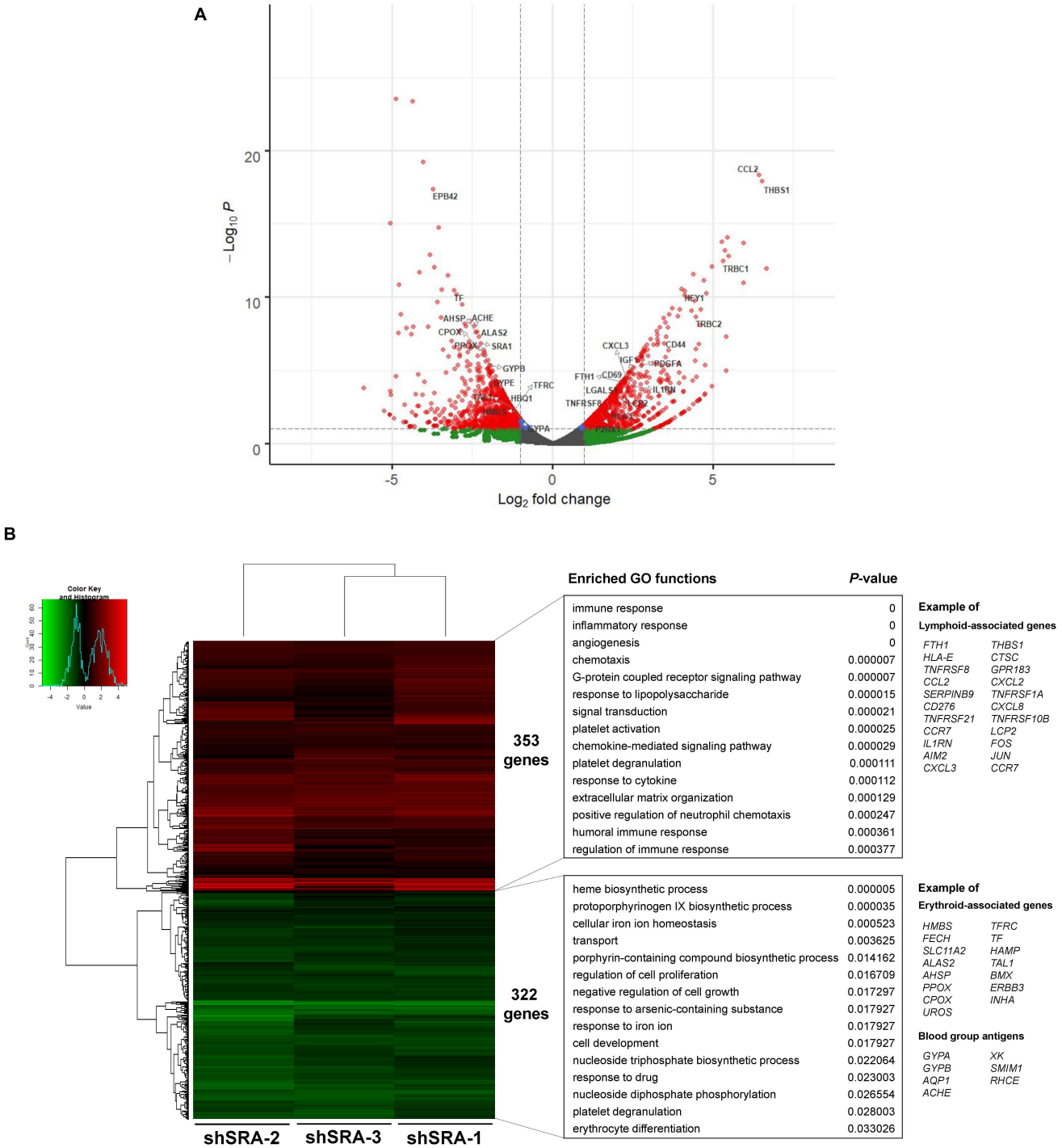


FIGURE 2 | SRA regulates transcription of hematopoiesis-related genes in the human erythroblast cells K562. The lncRNA SRA controls expression of 675 genes in the human erythroblast cells K562. RNA-seq was performed for cells depleted for SRA using individual shRNA constructs, i.e., shSRA-1, shSRA-2, or shSRA-3, and the control knockdown shLuc. Following silencing of SRA in K562 cells, 322, and 353 genes are down- and up-regulated in SRA knockdown cells, respectively. **(A)** Volcano plot illustrating changes in gene expression upon SRA silencing. The plot represents statistical significance vs expression fold change between the two conditions. Results from three biological replicates using different shRNA targets are shown. Genes with log2 fold change > 1 or < -1 and p-value < 0.01 are considered to be differentially expressed genes (DEGs) and are shown in red dots. **(B)** Heat map of DEGs between SRA knockdown samples and the control was analyzed. Pseudocount of TPMs was employed for calculation of fold-change using TPM + 1. The fold-change was calculated from TPM values of knockdown per TPM values of control, and the graph was plotted by representing log2(fold-change) of DEGs. Categories of enriched gene ontologies (GO) of genes up- and down-regulated by SRA in K562 (p-value < 0.05) were analyzed using DAVID. Examples of genes within the GO terms are shown on the right.

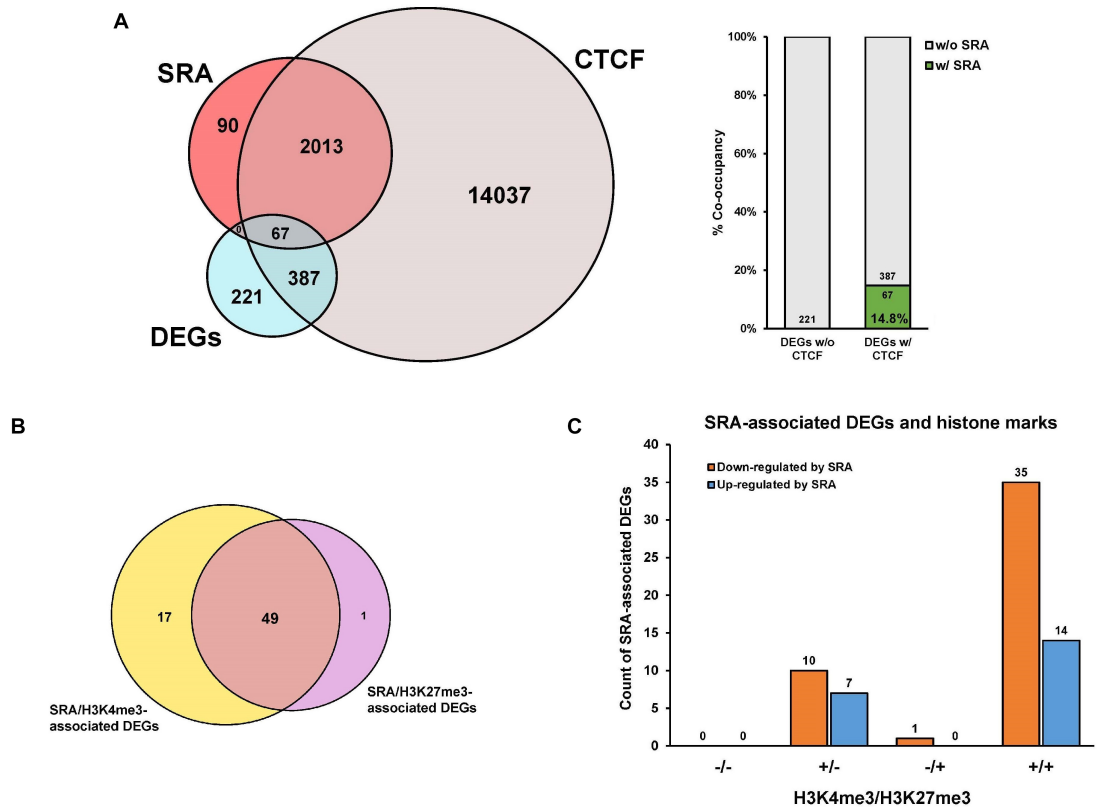


FIGURE 3 | Genes differentially expressed by SRA knockdown are occupied by SRA, CTCF and histone marks at the chromatin level. Publicly available ChIP-seq data for CTCF, H3K4me3, and H3K27me3 were derived from the ENCODE project. **(A)** Genes transcriptionally regulated by SRA and containing SRA binding site tend to associate with CTCF. Numbers of differentially expressed genes (DEGs) were compared with numbers of SRA-associated genes or CTCF-associated genes as shown by the Venn diagram. Percentage of DEGs with both SRA and CTCF occupancy is higher than that without CTCF. **(B)** A comparison of differentially expressed genes occupied by SRA in association with H3K4me3 or H3K27me3. **(C)** Genes down-regulated or up-regulated by SRA containing SRA occupancy were grouped into four categories depending on the presence (+) or absence (–) of H3K4me3 or H3K27me3.

is not expressed in this cell line, it was not included in our analysis. These results indicate a supportive role of the lncRNA SRA in erythroid-specific transcriptional regulation of the human erythroblasts K562.

SRA Regulates Hematopoiesis-Related Genes Transcriptome-Wide in HSC-Derived Primary Human Proerythroblast Cells

To translate whether SRA controls transcriptome-wide expression of primary human proerythroblasts, CD34-positive HSCs were induced to CD36-positive proerythroblast cells. qPCR analysis revealed that expression of the lncRNA SRA is induced upon erythroblast differentiation of HSCs, although only marginally (**Supplementary Figure S7**). The CD36-positive proerythroblast cells were then transduced with lentiviruses to silence SRA expression. As shown in **Supplementary Figure S8**, SRA transcript was successfully knocked down in CD36-positive proerythroblasts. RNA-seq analysis of CD36-positive proerythroblast cells was performed upon depletion of SRA using three different shRNA constructs. Silencing of SRA led to differential expression of 515 genes, with 233 and 282 genes were down- and up-regulated by SRA knockdown, respectively (**Figure 5** and **Supplementary Table S5**). Gene ontology analysis of genes positively controlled by SRA shows that cell division- and erythroblast-associated pathways such as cell cycle (e.g., *CDK1*, *CDC6,* and *MCM2*), telomere maintenance (e.g., *RFC3, RFC4,* and *RPA2*), heme biosynthesis (e.g., *ALAS2, CPOX,* and *UROS*), and erythrocytes (e.g., *KLF1*, *TFRC*, *GYPA*, *HBG2*, *HBD,* and *AHSP*) are enriched in the gene groups induced by SRA (**Figure 5B** and **Supplementary Table S6**). Genes belonging to these pathways are clustered in two nodes as shown by a functional network analysis (**Supplementary Figure S9**). On the other hand, the classification analysis of genes negatively controlled by SRA shows that leukocyte-associated pathways such as inflammatory response (e.g., *CD14*, *PYCARD*, *SERPINE1*, *PF4*, *IL1B*, *IL2RA,* and *LYZ*), immune response (e.g., *FTH1*, *CSF2,* and *IL7R*), and chemokine (e.g., *CCL1*, *CCL2*, *CCL4*, *CXCL8*, and *CXCR4*) are enriched in the gene groups repressed by SRA (**Figure 5B** and **Supplementary Table S7**). Genes associated with these pathways are also well clustered (**Supplementary Figure S10**). By comparing genes down- or up-regulated upon SRA knockdown, we find that 30 and 80 genes

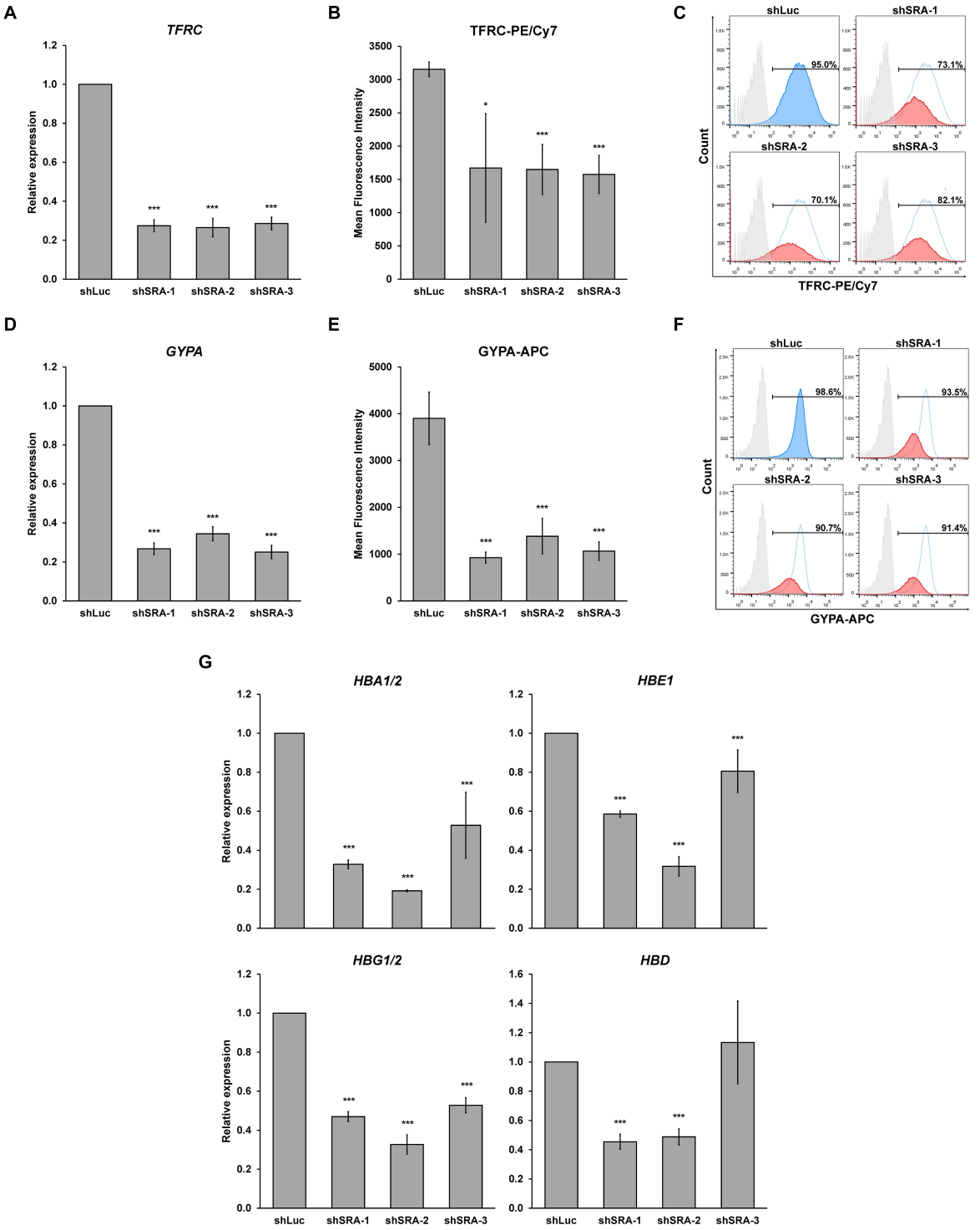


FIGURE 4 | The lncRNA SRA maintains expression of erythroid-specific genes in K562. In the human erythroblast cell line, depletion of SRA decreased *TFRC* **(A)** and *GYPA* **(D)** gene expression. *ATCB* was utilized as an internal control. Error bars represent SD. ($n = 3$; $^{*}p < 0.05$; $^{***}p < 0.01$). **(B,C,E,F)** Flow cytometry analysis shows that both expression level and the number of cells positive for the two markers are reduced by SRA knockdown. The histograms are shown to compare percentage of positive populations and expression level. Gray: negative control staining; Blue: control knockdown; Red: SRA knockdown of three different constructs; and White: control knockdown shown as a background. **(G)** SRA also facilitates expression of globin genes *HBA1/2*, *HBE*, *HBG1/2,* and *HBD* in K562. *ATCB* was utilized as an internal control. Error bars represent SD. ($n = 3$; $^{***}p < 0.01$).

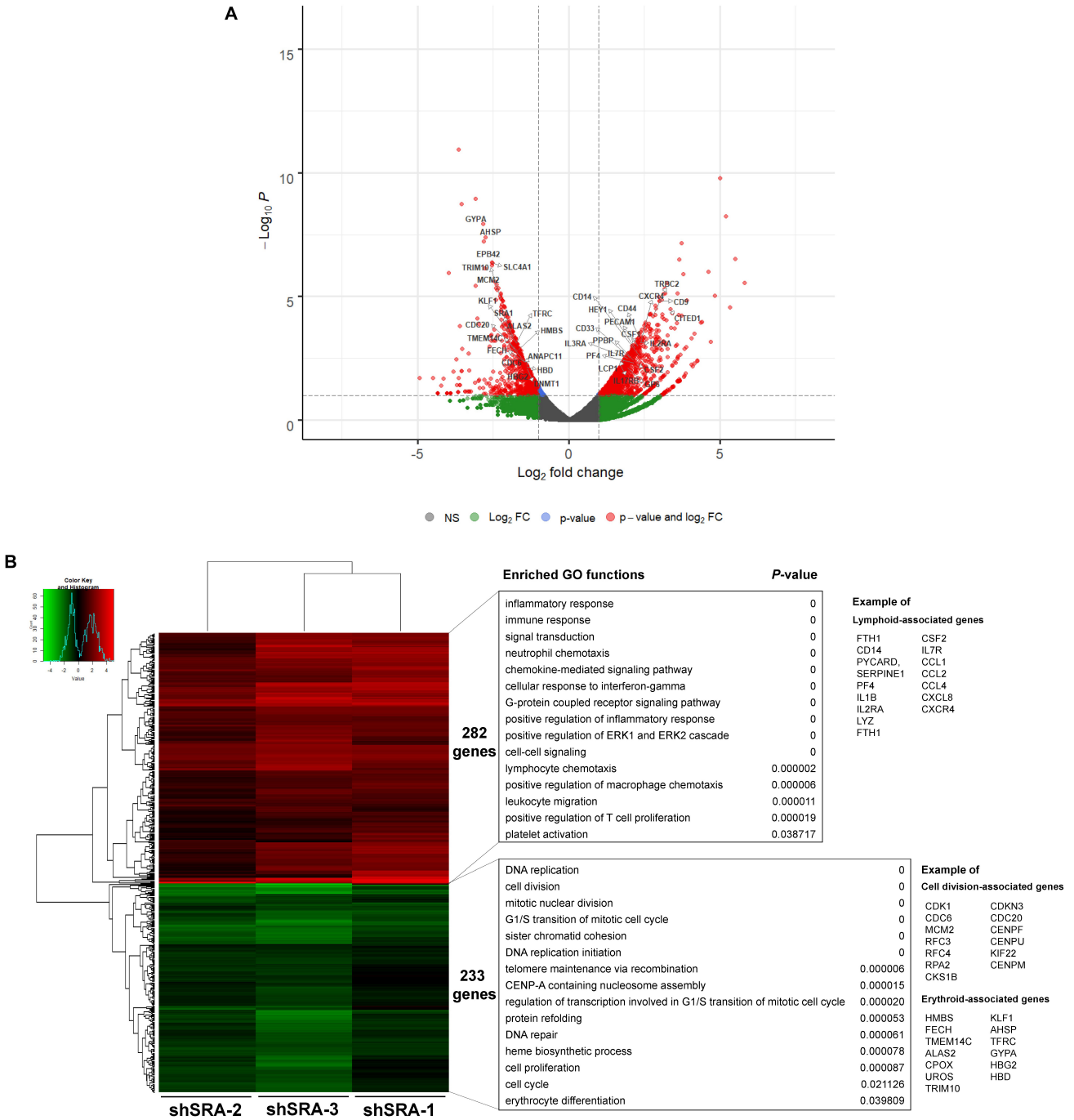


FIGURE 5 | SRA regulates transcription of hematopoiesis-related genes in the human CD36-positive erythroid progenitor cells. The lncRNA SRA controls expression of 515 genes in the human CD36-positive erythroid progenitor cells. RNA-seq was performed for cells depleted for SRA using individual shRNA constructs, i.e., shSRA-1, shSRA-2, or shSRA-3, and the control knockdown shLuc. Following silencing of SRA in the CD36-positive cells, 233 and 282 genes are down- and up-regulated in SRA knockdown cells, respectively. **(A)** Volcano plot illustrating changes in gene expression upon SRA silencing. The plot represents statistical significance vs expression fold change between the two conditions. Results from three biological replicates using different shRNA targets are shown. Genes with log2 fold change > 1 or < -1 and p-value < 0.01 are considered to be differentially expressed genes (DEGs) and are shown in red dots. **(B)** Heat map of DEGs between SRA knockdown samples and the control was analyzed. Pseudocount of TPMs was employed for calculation of fold-change using TPM + 1. The fold-change was calculated from TPM values of knockdown per TPM values of control, and the graph was plotted by representing log2(fold-change) of DEGs. Categories of enriched gene ontologies of genes up- and down-regulated by SRA in CD36-positive proerythroblasts (p-value < 0.05) and their enrichment scores [−log(p-value)] were analyzed using DAVID.

are overlapped between K562 and primary erythroblasts, respectively (**Supplementary Figure S11**). Examples of gene categories commonly induced by SRA between the two cell types include those involved in heme biosynthesis, hemoglobin, and erythrocyte differentiation such as *TFRC*, *GYPA*, *ALAS2*, *AHSP,* and *EPB42*. Examples of gene categories commonly suppressed by SRA between the two cell types include those involved in immune response, angiogenesis, and lipopolysaccharide response such as *FTH1*, *CCL2*, *IL1RN*, *THBS1,* and *CXCL8*. Nonetheless, the majority of DEGs are cell type-specific, either found exclusively in K562 or in primary erythroblasts. This cell type-specific gene expression might reflect differences of their origins, i.e., K562 as erythroleukemia cells and CD36-positive cells as primary erythroblasts. Specifically, 292 genes induced by SRA exclusively in K562 such as *DNAJB2*, *GDF2*, *SLC11A2*, *TF,* and *LGALS3BP* are functionally grouped as negative regulation of cell growth, cellular iron ion homeostasis, and platelet degranulation. Further, 203 genes suppressed by SRA exclusively in K562 such as *GNG12*, *JUN*, *TNFRSF8*, *DGKG,* and *RAP2B* are functionally grouped as response to lipopolysaccharide and platelet activation. For CD36-positive proerythroblasts, 273 genes induced by SRA exclusively in the cells but not in K562 such as *CENPW*, *CENPX*, *MCM2*, *RPA2,* and *CDC6* can be grouped within cell division and DNA replication. In addition, 202 genes suppressed by SRA exclusively in proerythroblasts such as *GBP5*, *CCL4*, *CCL17*, *CD14,* and *ITGB2* are functionally grouped as inflammatory response and neutrophil chemotaxis. Together, this result indicates the role of SRA in global gene regulation of erythroleukemia and in primary human proerythroblasts.

Next, we substantiated whether SRA silencing reduces expression of erythroblast markers of the primary proerythroblast cells. At the early stage of differentiation, we found that SRA depletion led to an increase in CD34-positive population, while CD36-positive population was reduced (**Supplementary Figure S12**). Expression of the committed erythroid marker *TFRC* and *GYPA* is reduced in erythroid-induced differentiating HSCs (**Figures 6A,D**). However, unlike K562, flow cytometry analysis of TFRC in proerythroblasts shows that both antigen expression level and number of cells positive of TFRC are not different between the control and SRA-depleted cells (**Figures 6B,C**). Nonetheless, depletion of SRA led to a decrease in the antigen expression and in the number of cells positive for GYPA (**Figures 6E,F**). We also tested whether expression of the globin genes is transcriptionally controlled by SRA. Real-time PCR analysis reveals that silencing of SRA led to a decrease in globin gene expression (**Figure 6G**). Taken together with the RNA-seq data, these results suggest that the lncRNA SRA facilitates transcriptional expression of erythroid-associated genes of primary human proerythroblast cells.

DISCUSSION

Delineating molecular mechanisms underlying erythroblast gene regulation is critical for understanding RBC disorders. A vast arrays of molecular and cellular pathways have been discovered to control this process (Oburoglu et al., 2016). For example, crosstalk between signal transductions and transcription factors modulates erythropoiesis in both mice and men (Perreault and Venters, 2018). Epigenetic regulators such as enzymes that modify DNA and histones also participate in regulation of erythropoiesis at chromatin level (Gnanapragasam and Bieker, 2017). In this work, we find that the lncRNA SRA occupies chromatin genome-wide in the human erythroblast cell line K562 (**Figure 1**), and controls expression of erythroblast-associated genes transcriptome-wide in both K562 and HSC-derived primary erythroblast cells (**Figures 2**, **5**). Moreover, expression of the erythroid marker GYPA, and the number of GYPA-positive cells are decreased in K562 and primary erythroblasts depleted for SRA (**Figures 4**, **6**) suggesting that the lncRNA SRA facilitates erythroid transcriptional program. However, we observed a reduction in TFRC expression only in K562 but not in the primary erythroblasts depleted for SRA. The relatively high level of TFRC expression of primary erythroblasts is consistent with the maintenance of TFCR expression in erythroblasts *in vitro* (Fajtova et al., 2013; Mao et al., 2016).

Recently, more than 9,000 genes encoding lncRNAs have been identified as being transcribed from the human genome (Derrien et al., 2012). They can participate in transcriptional regulation by acting as scaffold machineries for transcription factors and epigenetic modifying enzymes (Rinn and Chang, 2012). Using ChIRP-seq and ChIRP-PCR, we show here that SRA occupies at the alpha and beta globin loci, and facilitates the expression of the globin genes including *HBA1/2*, *HBG1/2*, *HBE,* and *HBD* in K562 cells and human proerythroblasts (**Figures 4**, **6**). SRA can form a complex with the chromatin architectural transcription factor CTCF, whose function in transcriptional control of genes at the beta globin locus has long been appreciated (Wallace and Felsenfeld, 2007; Ghirlando et al., 2012). Specifically, CTCF has been shown to facilitate expression of the gamma globin gene (Hou et al., 2010). In addition, the DNA binding transcription factor ATF2 which interacts with SRA also induces expression of the gamma globin gene (Liu et al., 2013). In contrast to SRA, the lncRNA HMI has been shown to suppress expression of the gamma globin gene (Morrison et al., 2018). At the alpha globin locus, lncRNA-αGT controls chicken globin expression (Arriaga-Canon et al., 2014). Since reactivation of *HBG* is a promising strategy for sickle cell anemia (Vinjamur et al., 2018) and accumulating evidence have suggested the role of lncRNAs in transcriptional regulation of globin genes, it is pivotal to determine which chromatin-associated factor(s) brings SRA and other lncRNAs to their target sites to induce expression of the globin genes.

The lncRNA SRA has been reported to promote cell fate transition including myogenesis (Caretti et al., 2006; Hube et al., 2011) and adipogenesis (Xu et al., 2010) as well as a transition into the pluripotent state (Wongtrakoongate et al., 2015). Yet, it has been elusive whether SRA is involved in transcriptional control during erythropoiesis. Estrogen receptor and glucocorticoid receptor, which are SRA-associated nuclear receptors, have been suggested to attenuate erythroid lineage (Schroeder et al., 1993; Blobel et al., 1995; Wessely et al., 1997; Lanz et al., 1999). On the other hand, transcription factors involved in SRA-mediated transcriptional regulation such as

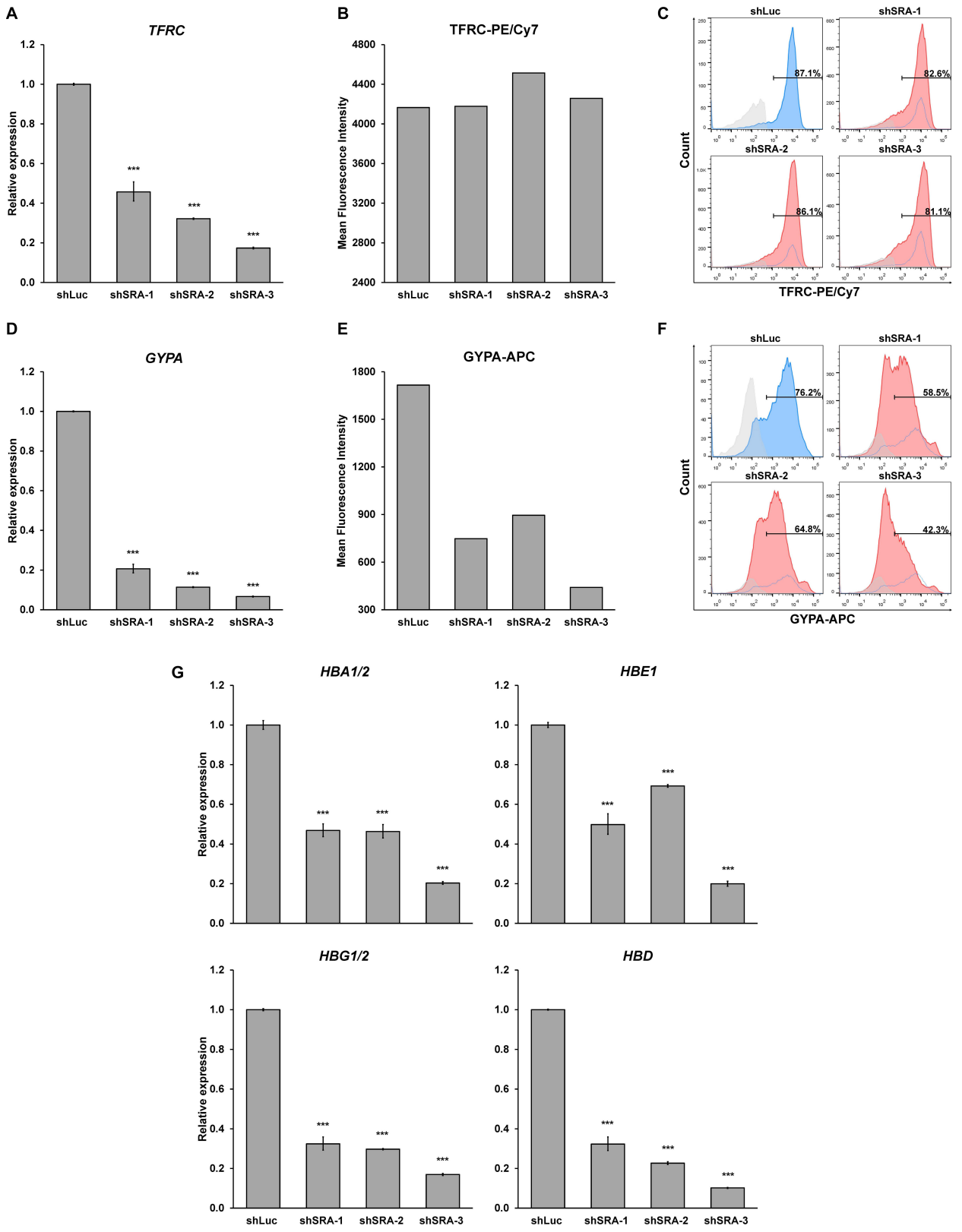


FIGURE 6 | Continued

FIGURE 6 | SRA facilitates expression of erythroid-specific genes of CD36-positive erythroid progenitor cells. During erythroid-induced differentiation, depletion of SRA decreased *TFRC* **(A)** and *GYPA* **(D)** gene expression. *ATCB* was utilized as an internal control. Error bars represent SD. ($n = 3$; $^{***}p < 0.01$). However, flow cytometry analysis shows that expression level of the erythroid marker TFRC **(B)** and number of TFRC-positive cells **(C)** are not affected by SRA knockdown, whereas those of GYPA **(E,F)** are reduced by SRA knockdown. The histograms are shown to compare percentage of positive populations and expression level. Gray: negative control staining; Blue: control knockdown; Red: SRA knockdown of three different constructs; and White: control knockdown shown as a background. **(G)** SRA facilitates expression of globin genes *HBA1/2*, *HBE*, *HBG1/2,* and *HBD* in CD36-positive erythroid progenitor cells. *ATCB* was utilized as an internal control. Error bars represent SD. ($n = 3$; $^{***}p < 0.01$).

CTCF and thyroid hormone receptor have been shown to facilitate generation of erythroid cells (Bartunek and Zenke, 1998; Xu and Koenig, 2004; Torrano et al., 2005; Yao et al., 2010; Gao et al., 2017). Apart from being associated with transcription factors, the role of SRA in supporting cell fate transition and plasticity might be in part due to its interaction with epigenetic machineries (Wongtrakoongate et al., 2015). We have previously reported that SRA interacts with CTCF, TrxG, and PRC2 (Yao et al., 2010; Wongtrakoongate et al., 2015). In addition, CTCF tends to localize nearby H3K4me3, and H3K27me3, which are established by TrxG and PRC2, respectively (Barski et al., 2007; Li et al., 2008; Cuddapah et al., 2009). This could explain the enrichment of the two histone marks at CTCF binding sites containing SRA (**Figure 1**). However, the majority of H3K4me3 or H3K27me3 sites are associated neither with SRA nor CTCF, supporting the existence of multiple mechanisms for establishing these histone modifications. Further studies will be required to uncover possible synergistic regulation by lncRNAs and their protein binding partners in erythropoiesis.

There are growing evidence of various functions of lncRNAs in blood cells (Li et al., 2018; Dahariya et al., 2019). During mouse embryonic hematopoiesis, the lncRNA H19 promotes gene expression program of hematopoiesis transcriptome-wide via regulation of promoter DNA methylation of key HSC genes, and is therefore critical for embryonic endothelial-to-hematopoietic transition and generating embryonic HSCs in aorta-gonads-mesonephros (Zhou et al., 2019). The mouse lncRNA EC2, which is conserved in human, has been reported to facilitate expression of the erythroid marker Ter119 and enucleation of mouse erythroblasts (Alvarez-Dominguez et al., 2014). In human, a transcriptome-wide analysis of erythroid-induced human HSCs has revealed expression of approximately 1,100 genes encoding lncRNAs. Of these, the expression level of 34 lncRNAs is correlated with that of protein coding genes involved in hematopoiesis, leukocyte activation and DNA repair in erythroblasts suggesting a possible function of these lncRNAs in transcriptional regulation of the associated genes (Ding et al., 2016). Heme biosynthesis is erythroblasts mediated by the lncRNA UCA1, which is upregulated at the proerythroblast stage interacts with the ribonucleoprotein PTBP1 (Liu et al., 2018). In another study, the lncRNA HMI (also called HMI-lncRNA), which is transiently induced during human erythropoiesis, is a negative regulator of gamma globin expression (Morrison et al., 2018). Intriguingly, using RNA-seq and ChIRP-seq approaches, the enhancer-associated lncRNA Bloodlinc has been reported to facilitate enucleation of mouse RBCs by inducing erythroid-related genes and repressing non-erythroid genes through direct binding to chromatin. Interestingly, similar to SRA, Bloodlinc also forms a complex with the RNA helicase DDX5 (Alvarez-Dominguez et al., 2017), which is important for establishment of H3K4me3 (Wongtrakoongate et al., 2015). Therefore, these findings have shed light on the function of these lncRNA transcripts in regulation of erythropoiesis. In conclusion, we have suggested the role of SRA in human erythropoiesis, as well as a direct transcriptional control of SRA in regulation of proerythroblast-associated genes. Together, our work supports the roles of lncRNAs in erythroblast gene regulation.

DATA AVAILABILITY STATEMENT

All relevant sequencing data has been uploaded to NCBI: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153004 and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151926.

AUTHOR CONTRIBUTIONS

WS and PW conceived and designed the research, analyzed the data, and wrote the manuscript. WS, KC, and PW conducted the experiments. KZ, SH, and SF contributed to reagents and analytical tools. All authors read and approved the manuscript.

FUNDING

This research project was supported by Mahidol University and CIF Grant, Faculty of Science, Mahidol University. WS was supported by Science Achievement Scholarship of Thailand. KZ was supported by the Intramural Research Program of the NIH, National Heart, Lung, and Blood Institute. PW was supported by the Thailand Research Fund and Faculty of Science, Mahidol University (grant no. TRG5880105).

ACKNOWLEDGMENTS

We are indebted to Gary Felsenfeld, Gregory Riddicks, Peerapat Khamwachirapitak, and PW lab members for their suggestions and comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00850/full#supplementary-material

FIGURE S1 | The lncRNA SRA preferentially occupies within 50 kb upstream or downstream of transcription start site. **(A)** Amount region-gene associations at different distance to TSS. Since lncRNAs, such as SRA, can function as RNA scaffolds for chromatin regulators such as CTCF, the chromatin spanning regions from TSS were therefore set into different bins ranging from 500 kb up- or down-stream of TSS to allow discovery of potential SRA occupancy beyond TSS. **(B)** Percent genomic regions occupied by SRA. White and red bar represents genomic regions associated with one or more genes and not associated with any genes, respectively.

FIGURE S2 | The lncRNA SRA occupies at alpha and beta globin gene loci. Chromatin Isolation by RNA Purification (ChIRP) of SRA in K562 cells was performed using deoxyoligonucleotide probes tiling along the lncRNA SRA followed by PCR. Genomic regions for PCR primer binding were selected according to DNaseI hypersensitive sites (HS), and occupancy of CTCF, H3K4me1 and H3K4me3 as shown in the histograms, where *x*-axis and *y*-axis represent physical map and occupancy levels, respectively. The numbers of primer pairs are indicated under the histograms. **(A)** Thirteen pairs of primers for ChIRP-PCR were designed to determine association of SRA at the alpha globin chromatin locus. **(B)** ChIRP-PCR analysis at the alpha locus revealed that SRA occupies the sites HS40 and numbers 7 and 12. **(C)** Twenty two pairs of primers for ChIRP-PCR were designed to determine association of SRA at the beta globin chromatin locus. **(D)** ChIRP-PCR analysis revealed that SRA occupies site numbers 3, 6, 10, 12, 13, 14, 17, and 19. Enrichment signals of SRA-occupied chromatin fragments were normalized to those of the negative control probes LacZ. Error bars represent SD. ($n = 3$; $^*p < 0.05$).

FIGURE S3 | RNA pull down in K562 cells. Sense and antisense of biotinylated SRA transcripts were incubated with K562 nuclear extract. Western blot were performed with pulled down proteins. Sense SRA, but not antisense SRA or MBP transcripts, can pull down the RNA helicase DDX5, the chromatin architectural protein CTCF, the TrxG component WDR5, and the PRC2 member EZH2.

FIGURE S4 | Silencing of SRA by shRNA led to a reduction of SRA lncRNA transcript. *ATCB* was utilized as an internal control. Error bars represent SD. ($n = 5$; $^{***}p < 0.01$).

FIGURE S5 | Functional network analysis of genes induced by SRA in K562 was identified by the STRING (Search Tool for the Retrieval of Interacting Genes) with connected lines representing genetic/physical interaction of connected genes.

FIGURE S6 | Functional network analysis of genes repressed by SRA in K562 was identified by the STRING (Search Tool for the Retrieval of Interacting Genes) with connected lines representing genetic/physical interaction of connected genes.

FIGURE S7 | The lncRNA SRA is marginally induced during erythroblast differentiation of human HSCs. *ATCB* was utilized as an internal control. Error bars represent SD. ($n = 3$; $^*p < 0.05$).

FIGURE S8 | Silencing of SRA by shRNA led to a reduction of SRA in human CD36-positive cells. *ATCB* was utilized as an internal control. Error bars represent SD. ($n = 3$; $^{***}p < 0.01$).

FIGURE S9 | Functional network analysis of genes induced by SRA in primary erythroblasts was identified by the STRING (Search Tool for the Retrieval of Interacting Genes) with connected lines representing genetic/physical interaction of connected genes.

FIGURE S10 | Functional network analysis of genes repressed by SRA in primary erythroblasts was identified by the STRING (Search Tool for the Retrieval of Interacting Genes) with connected lines representing genetic/physical interaction of connected genes.

FIGURE S11 | Comparative analysis of differential expressed genes controlled by SRA between K562 and CD36 + proerythroblasts. Group of erythroid- and lymphoid-associated genes are consistency down-regulated **(A)** and up-regulated **(B)**, upon SRA silencing in both K562 and CD36 + proerythroblasts.

FIGURE S12 | The lncRNA SRA facilitates differentiation of $CD34^+$ human HSCs into $CD36^+$ human erythroid progenitors. Erythroid differentiation of HSCs was performed before transducing with shRNA-containing lentiviruses at day 7 of differentiation. The lentiviral-transduced cells were collected at 96 h post-transduction, and were subjected to flow cytometry analysis co-stained for the human HSC surface marker CD34 and the erythroid progenitor marker CD36. **(A)** Flow cytometry histograms of CD34. **(B)** Flow cytometry histograms of CD36. **(C)** Flow cytometry histograms of double staining cells. **(D)** Silencing of SRA increased $CD34^+$ $CD36^+$ double-positive and $CD34^-$ $CD36^-$ double-negative populations at the expense of $CD34^-$ $CD36^+$ double-positive population.

REFERENCES

Alvarez-Dominguez, J. R., Hu, W., Yuan, B., Shi, J., Park, S. S., Gromatzky, A. A., et al. (2014). Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood* 123, 570–581. doi: 10.1182/blood-2013-10-530683

Alvarez-Dominguez, J. R., Knoll, M., Gromatzky, A. A., and Lodish, H. F. (2017). The Super-Enhancer-Derived alncRNA-EC7/Bloodlinc Potentiates Red Blood Cell Development in trans. *Cell Rep.* 19, 2503–2514. doi: 10.1016/j.celrep.2017.05.082

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Arriaga-Canon, C., Fonseca-Guzman, Y., Valdes-Quezada, C., Arzate-Mejia, R., Guerrero, G., and Recillas-Targa, F. (2014). A long non-coding RNA promotes full activation of adult gene expression in the chicken alpha-globin domain. *Epigenetics* 9, 173–181. doi: 10.4161/epi.27030

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837. doi: 10.1016/j.cell.2007.05.009

Bartunek, P., and Zenke, M. (1998). Retinoid X receptor and c-cerbA/thyroid hormone receptor regulate erythroid cell growth and differentiation. *Mol. Endocrinol.* 12, 1269–1279. doi: 10.1210/me.12.9.1269

Blighe, K., Rana, S., and Lewis, M. (2020). *EnhancedVolcano: Publication-Ready Volcano Plots with Enhanced Colouring and Labeling. R package version 1.6.0.*

Blobel, G. A., Sieff, C. A., and Orkin, S. H. (1995). Ligand-dependent repression of the erythroid transcription factor GATA-1 by the estrogen receptor. *Mol. Cell. Biol.* 15, 3147–3153. doi: 10.1128/mcb.15.6.3147

Caretti, G., Schiltz, R. L., Dilworth, F. J., Di Padova, M., Zhao, P., Ogryzko, V., et al. (2006). The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. *Dev. Cell* 11, 547–560. doi: 10.1016/j.devcel.2006.08.003

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., and Chang, H. Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell.* 44, 667–678. doi: 10.1016/j.molcel.2011.08.027

Chu, C., Quinn, J., and Chang, H. Y. (2012). Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* 25:3912.

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

Cuddapah, S., Jothi, R., Schones, D. E., Roh, T. Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* 19, 24–32. doi: 10.1101/gr.082800.108

Dahariya, S., Paddibhatla, I., Kumar, S., Raghuwanshi, S., Pallepati, A., and Gutti, R. K. (2019). Long non-coding RNA: classification, biogenesis and functions in blood cells. *Mol. Immunol.* 112, 82–92. doi: 10.1016/j.molimm.2019.04.011

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

Ding, N., Xi, J., Li, Y., Xie, X., Shi, J., Zhang, Z., et al. (2016). Global transcriptome analysis for identification of interactions between coding and noncoding RNAs during human erythroid differentiation. *Front. Med.* 10:297–310. doi: 10.1007/s11684-016-0452-0

Fajtova, M., Kovarikova, A., Svec, P., Kankuri, E., and Sedlak, J. (2013). Immunophenotypic profile of nucleated erythroid progenitors during

maturation in regenerating bone marrow. *Leuk. Lymphoma* 54, 2523–2530. doi: 10.3109/10428194.2013.781167

Gan, T., Jude, C. D., Zaffuto, K., and Ernst, P. (2010). Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis. *Leukemia* 24, 1732–1741. doi: 10.1038/leu.2010.171

Gao, X., Lee, H. Y., Li, W., Platt, R. J., Barrasa, M. I., Ma, Q., et al. (2017). Thyroid hormone receptor beta and NCOA4 regulate terminal erythrocyte differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10107–10112. doi: 10.1073/pnas.1711058114

Ghirlando, R., Giles, K., Gowher, H., Xiao, T., Xu, Z., Yao, H., et al. (2012). Chromatin domains, insulators, and the regulation of gene expression. *Biochim. Biophys. Acta* 1819, 644–651. doi: 10.1016/j.bbagrm.2012.01.016

Gnanapragasam, M. N., and Bieker, J. J. (2017). Orchestration of late events in erythropoiesis by KLF1/EKLF. *Curr. Opin. Hematol.* 24, 183–190. doi: 10.1097/moh.0000000000000327

Hou, C., Dale, R., and Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3651–3656. doi: 10.1073/pnas.0912087107

Hu, G., Cui, K., Fang, D., Hirose, S., Wang, X., Wangsa, D., et al. (2018). Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. *Immunity* 48:227–242.e8. doi: 10.1016/j.immuni.2018.01.013

Hu, G., Tang, Q., Sharma, S., Yu, F., Escobar, T. M., Muljo, S. A., et al. (2013). Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat. Immunol.* 14, 1190–1198. doi: 10.1038/ni.2712

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.

Hube, 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

Kidder, B. L., Hu, G., Cui, K., and Zhao, K. (2017). SMYD5 regulates H4K20me3-marked heterochromatin to safeguard ES cell self-renewal and prevent spurious differentiation. *Epigenet. Chromatin* 10:8.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

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

Lanz, R. B., Razani, B., Goldberg, A. D., and O'Malley, B. W. (2002). Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc. Natl. Acad. Sci. U.S.A.* 99, 16081–16086. doi: 10.1073/pnas.192571399

Li, T., Hu, J. F., Qiu, X., Ling, J., Chen, H., Wang, S., et al. (2008). CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol. Cell. Biol.* 28, 6473–6482. doi: 10.1128/mcb.00204-08

Li, W., Ren, Y., Si, Y., Wang, F., and Yu, J. (2018). Long non-coding RNAs in hematopoietic regulation. *Cell Regen.* 7, 27–32.

Liu, J., Li, Y., Tong, J., Gao, J., Guo, Q., Zhang, L., et al. (2018). Long non-coding RNA-dependent mechanism to regulate heme biosynthesis and erythrocyte development. *Nat. Commun.* 9:4386.

Liu, L., Karmakar, S., Dhar, R., Mahajan, M., Choudhury, A., Weissman, S., et al. (2013). Regulation of Ggamma-globin gene by ATF2 and its associated proteins through the cAMP-response element. *PLoS One* 8:e78253. doi: 10.1371/journal.pone.0078253

Majewski, I. J., Blewitt, M. E., de Graaf, C. A., McManus, E. J., Bahlo, M., Hilton, A. A., et al. (2008). Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *PLoS Biol.* 6:e93. doi: 10.1371/journal.pbio.0060093

Mao, B., Huang, S., Lu, X., Sun, W., Zhou, Y., Pan, X., et al. (2016). Early development of definitive erythroblasts from human pluripotent stem cells defined by expression of Glycophorin A/CD235a, CD34, and CD36. *Stem Cell Rep.* 7, 869–883. doi: 10.1016/j.stemcr.2016.09.002

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., et al. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501.

Mochizuki-Kashio, M., Mishima, Y., Miyagi, S., Negishi, M., Saraya, A., Konuma, T., et al. (2011). Dependency on the polycomb gene Ezh2 distinguishes fetal from adult hematopoietic stem cells. *Blood* 118, 6553–6561. doi: 10.1182/blood-2011-03-340554

Morrison, T. A., Wilcox, I., Luo, H. Y., Farrell, J. J., Kurita, R., Nakamura, Y., et al. (2018). A long noncoding RNA from the HBS1L-MYB intergenic region on chr6q23 regulates human fetal hemoglobin expression. *Blood Cells Mol. Dis.* 69, 1–9. doi: 10.1016/j.bcmd.2017.11.003

Oburoglu, L., Romano, M., Taylor, N., and Kinet, S. (2016). Metabolic regulation of hematopoietic stem cell commitment and erythroid differentiation. *Curr. Opin. Hematol.* 23, 198–205. doi: 10.1097/moh.0000000000000234

Perreault, A. A., and Venters, B. J. (2018). Integrative view on how erythropoietin signaling controls transcription patterns in erythroid cells. *Curr. Opin. Hematol.* 25, 189–195. doi: 10.1097/moh.0000000000000415

Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181. doi: 10.1038/nprot.2014.006

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166.

RStudio Team (2016). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc.

Schroeder, C., Gibson, L., Nordstrom, C., and Beug, H. (1993). The estrogen receptor cooperates with the TGF alpha receptor (c-erbB) in regulation of chicken erythroid progenitor self-renewal. *EMBO J.* 12, 951–960. doi: 10.1002/j.1460-2075.1993.tb05736.x

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452.

Torrano, V., Chernukhin, I., Docquier, F., D'Arcy, V., Leon, J., Klenova, E., et al. (2005). CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. *J. Biol. Chem.* 280, 28152–28161. doi: 10.1074/jbc.m501481200

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

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* 329, 689–693. doi: 10.1126/science.1192002

Vinjamur, D. S., Bauer, D. E., and Orkin, S. H. (2018). Recent progress in understanding and manipulating haemoglobin switching for the haemoglobinopathies. *Br. J. Haematol.* 180, 630–643. doi: 10.1111/bjh.15038

Wallace, J. A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* 17, 400–407. doi: 10.1016/j.gde.2007.08.005

Wessely, O., Deiner, E. M., Beug, H., and von Lindern, M. (1997). The glucocorticoid receptor is a key regulator of the decision between self-renewal and differentiation in erythroid progenitors. *EMBO J.* 16, 267–280. doi: 10.1093/emboj/16.2.267

Wickrema, A., and Crispino, J. D. (2007). Erythroid and megakaryocytic transformation. *Oncogene* 26, 6803–6815. doi: 10.1038/sj.onc.1210763

Wong, S., Zhi, N., Filippone, C., Keyvanfar, K., Kajigaya, S., Brown, K. E., et al. (2008). Ex vivo-generated CD36+ erythroid progenitors are highly permissive to human parvovirus B19 replication. *J. Virol.* 82, 2470–2476. doi: 10.1128/jvi.02247-07

Wongtrakoongate, P., Riddick, G., Fucharoen, S., and Felsenfeld, G. (2015). Association of the Long Non-coding RNA Steroid Receptor RNA Activator (SRA) with TrxG and PRC2 Complexes. *PLoS Genet.* 11:e1005615. doi: 10.1371/journal.pgen.1005615

Xu, B., Gerin, I., Miao, H., Vu-Phan, D., Johnson, C. N., Xu, R., et al. (2010). Multiple roles for the non-coding RNA SRA in regulation of adipogenesis

and insulin sensitivity. *PLoS One* 5:e14199. doi: 10.1371/journal.pone.0014199

Xu, B., and Koenig, R. J. (2004). An RNA-binding domain in the thyroid hormone receptor enhances transcriptional activation. *J. Biol. Chem.* 279, 33051–33056. doi: 10.1074/jbc.m404930200

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R. D., and Felsenfeld, G. (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 24, 2543–2555. doi: 10.1101/gad.1967810

Zhou, J., Xu, J., Zhang, L., Liu, S., Ma, Y., Wen, X., et al. (2019). Combined Single-Cell Profiling of lncRNAs and Functional Screening Reveals that H19 Is Pivotal for Embryonic Hematopoietic Stem Cell Development. *Cell Stem Cell* 24, 285-298.e5.

Zhu, L. J. (2013). "Integrative analysis of ChIP-Chip and ChIP-Seq dataset," in *Tiling Arrays: Methods and Protocols*, eds T.-L. Lee, and A. C. Shui Luk (Totowa, NJ: Humana Press), 105–124.

Zhu, L. J., Gazin, C., Lawson, N. D., Pagès, H., Lin, S. M., Lapointe, D. S., et al. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* 11:237. doi: 10.1186/1471-2105-11-237

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

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 11 January 2021
doi: 10.3389/fgene.2020.615378

Deregulation of a *Cis*-Acting lncRNA in Non-small Cell Lung Cancer May Control HMGA1 Expression

Greg L. Stewart†, Adam P. Sage†, Katey S. S. Enfield‡, Erin A. Marshall, David E. Cohn and Wan L. Lam*

Department of Integrative Oncology, BC Cancer Research Centre, Vancouver, BC, Canada

OPEN ACCESS

Edited by:
Liang Chen,
Wuhan University, China

Reviewed by:
Matthias S. Leisegang,
Goethe University Frankfurt, Germany
Rui Xiao,
Medical Research Institute, Wuhan University, China

***Correspondence:**
Greg L. Stewart
gstewart@bccrc.ca

†These authors have contributed equally to this work

‡Present address:
Katey S. S. Enfield,
The Francis Crick Institute, London, United Kingdom

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *09 October 2020*
Accepted: *11 December 2020*
Published: *11 January 2021*

Citation:
Stewart GL, Sage AP, Enfield KSS, Marshall EA, Cohn DE and Lam WL (2021) Deregulation of a Cis-Acting lncRNA in Non-small Cell Lung Cancer May Control HMGA1 Expression. Front. Genet. 11:615378. doi: 10.3389/fgene.2020.615378

Background: Long non-coding RNAs (lncRNAs) have long been implicated in cancer-associated phenotypes. Recently, a class of lncRNAs, known as *cis*-acting, have been shown to regulate the expression of neighboring protein-coding genes and may represent undiscovered therapeutic action points. The chromatin architecture modification gene *HMGA1* has recently been described to be aberrantly expressed in lung adenocarcinoma (LUAD). However, the mechanisms mediating the expression of *HMGA1* in LUAD remain unknown. Here we investigate the deregulation of a putative *cis*-acting lncRNA in LUAD, and its effect on the oncogene *HMGA1*.

Methods: LncRNA expression was determined from RNA-sequencing data of tumor and matched non-malignant tissues from 36 LUAD patients. Transcripts with significantly deregulated expression were identified and validated in a secondary LUAD RNA-seq dataset (TCGA). SiRNA-mediated knockdown of a candidate *cis*-acting lncRNA was performed in BEAS-2B cells. Quantitative real-time PCR was used to observe the effects of lncRNA knockdown on the expression of HMGA1.

Results: We identified the lncRNA RP11.513I15.6, which we refer to as HMGA1-*lnc,* neighboring *HMGA1* to be significantly downregulated in both LUAD cohorts. Conversely, we found *HMGA1* significantly overexpressed in LUAD and anticorrelated with HMGA1-*lnc*. *In vitro* experiments demonstrated siRNA-mediated inhibition of HMGA1-*lnc* in immortalized non-malignant lung epithelial cells resulted in a significant increase in *HMGA1* gene expression.

Conclusion: Our results suggest that HMGA1-*lnc* is a novel *cis*-acting lncRNA that negatively regulates *HMGA1* gene expression in lung cells. Further characterization of this regulatory mechanism may advance our understanding of the maintenance of lung cancer phenotypes and uncover a novel therapeutic intervention point for tumors driven by *HMGA1*.

Keywords: long non-coding RNA, HMGA1, gene regulation, *cis*-acting, non-coding RNA, lung cancer LncRNA-Mediated Control of HMGA1 Expression

INTRODUCTION

Long non-coding RNAs (lncRNAs) are a previously-underappreciated class of transcripts with a wide variety of now-recognized functions in gene regulation. Since their functional roles have been uncovered, numerous lncRNAs have been implicated in the onset of many cancer-associated phenotypes, such as progression, tumorigenesis, and metastasis (Gibb et al., 2011; Schmitt and Chang, 2016). Further, as RNA represents the functional unit for lncRNAs, rather than an intermediate as is the case with mRNAs, these transcripts are promising targets for the development of future RNA-based therapies (Amodio et al., 2018). Despite these recognized roles in tumor development, a key challenge in the translational utility of lncRNA-based research is the effective identification of their downstream target genes. In order to harness their potential for disease-specific markers and potential therapeutic targets, a better understanding of lncRNA mechanisms of action in disease is required.

An emerging class of lncRNAs – *cis*-acting – has been shown to regulate the expression of neighboring protein-coding genes, frequently including protein-coding genes with oncogenic or tumor-suppressive functions. Through a variety of mechanisms primarily occurring in the nucleus, these *cis*-acting non-coding transcripts can activate or repress transcription of neighboring genes. For example, the lncRNA *ANRIL* (antisense non-coding RNA in the *INK4* locus) interacts with polycomb repression complex 2 (PRC2), recruiting the complex to condense active chromatin and silence the genes around its transcriptional loci, including the well-known tumor suppressor INK4B (p15) (Yap et al., 2010).

Thus, *cis*-acting lncRNAs may represent novel mechanisms of cancer-gene regulation as well as potentially actionable intervention points in known cancer-driving pathways. Despite the prevalence of genetic and epigenetic deregulation events in lung adenocarcinoma (LUAD), the extent and consequences of aberrant *cis*-acting lncRNA expression on known cancer-driving genes is unknown.

High mobility group A1 (HMGA1) is part of a family of proteins involved in maintenance of chromatin architecture within the nucleus and as such, are implicated in tumorigenesis (Sgarra et al., 2018). Specifically, HMGA1 regulates a wide variety of genes through direct interactions with target sequences in promoter and enhancer regions. These downstream regulatory functions have been shown to lead to tumor development, particularly in breast cancer where the HMGA family has been shown to contribute, in part, to nearly every hallmark of cancer (Sgarra et al., 2018) (Hanahan and Weinberg, 2011). HMGA1 is enriched in several aggressive cancer types, including non-small cell lung cancer (NSCLC), where both mRNA and protein expression are substantially increased (Zhang et al., 2015). Further, overexpression of the *HMGA1* gene has been shown to be a key factor driving lung metastasis (Lin and Peng, 2016; Fu et al., 2018). In LUAD, high *HMGA1* gene expression has been associated with poor overall survival and chemotherapy resistance (Zhang et al., 2015). The oncogenic role of HMGA proteins stems from chromatin-mediated activation of cancer-driving genes such as *E2F1*, *AP1*, and *CCNA1*, as well as the repression of tumor suppressive genes such as *TP53* (Fusco and Fedele, 2007). Interestingly, overexpression of HMGA proteins has also been shown to alter the expression of non-coding genes, such as miRNAs, leading to lung development through dysregulation of the cell cycle (Pallante et al., 2015). However, the role of lncRNAs in HMGA1 regulation has not been investigated. As such we hypothesized lncRNA expression changes may be a mechanism of *HMGA1* deregulation in NSCLC. Our search for regulatory lncRNA has led to the identification of HMGA1-*lnc*, a *cis*-acting lncRNA neighboring the *HMGA1* gene at *6p21.31* that acts as alternative mechanisms of *HMGA1* deregulation.

RESULTS

Cis-acting LncRNAs are deregulated in LUAD

We first sought to identify *cis*-acting lncRNAs with potential biological relevance in LUAD. We hypothesized that a biologically relevant *cis*-acting lncRNA would (i) be significantly deregulated in two LUAD datasets (Fold Change, FC: 1.5), (ii) would overlap or closely neighbor (within 1.5 Kb) protein coding genes, that (iii) were also deregulated in LUAD.

Putative *cis*-acting lncRNAs that fit these parameters were identified, and literature searches on PubMed were performed to identify neighboring protein coding genes with experimental evidence of roles in cancer biology (**Supplementary Table 1**). This gave us a list of deregulated putative *cis*-acting lncRNAs that may function to regulate known cancer driving genes. For example, we find a lncRNA RP11-122M14.1 neighboring protein-coding gene *NEK2*. *NEK2* expression is associated with poor survival and inhibition results in anticancer effects in many cancer types, including lung cancer (Xia et al., 2015; Yao et al., 2019).

While many of these putative *cis*-acting lncRNAs are interesting, we decided to focus on *cis*-lncRNA-gene pairs that were deregulated in the opposite direction (discordant). We took this approach to avoid the false positives that can occur from regional "passenger" effects such as non-specific DNA copy number changes affecting multiple neighboring genes. Specifically, we decided to focus on a particular deregulated lncRNA, *RP11.513I15.6,* because of: (i) proximity to the known oncogene *HMGA1*, and (ii) the discordant expression relationship between the lncRNA and neighboring protein-coding gene. For simplicity we refer to this lncRNA as *HMGA1-lnc* (**Figure 1**).

Expression of HMGA1-lnc and HMGA1 are deregulated in LUAD

We observed the lncRNA *HMGA1-lnc* to be significantly downregulated in tumors when compared to adjacent non-malignant tissues, which holds true in both datasets (**Figure 2A** and **Supplementary Figure 1**). In contrast, the neighboring protein-coding gene *HMGA1* was found to be significantly overexpressed in both tumor datasets relative to matched non-malignant tissue (**Figure 2B** and **Supplementary Figure 1**). As

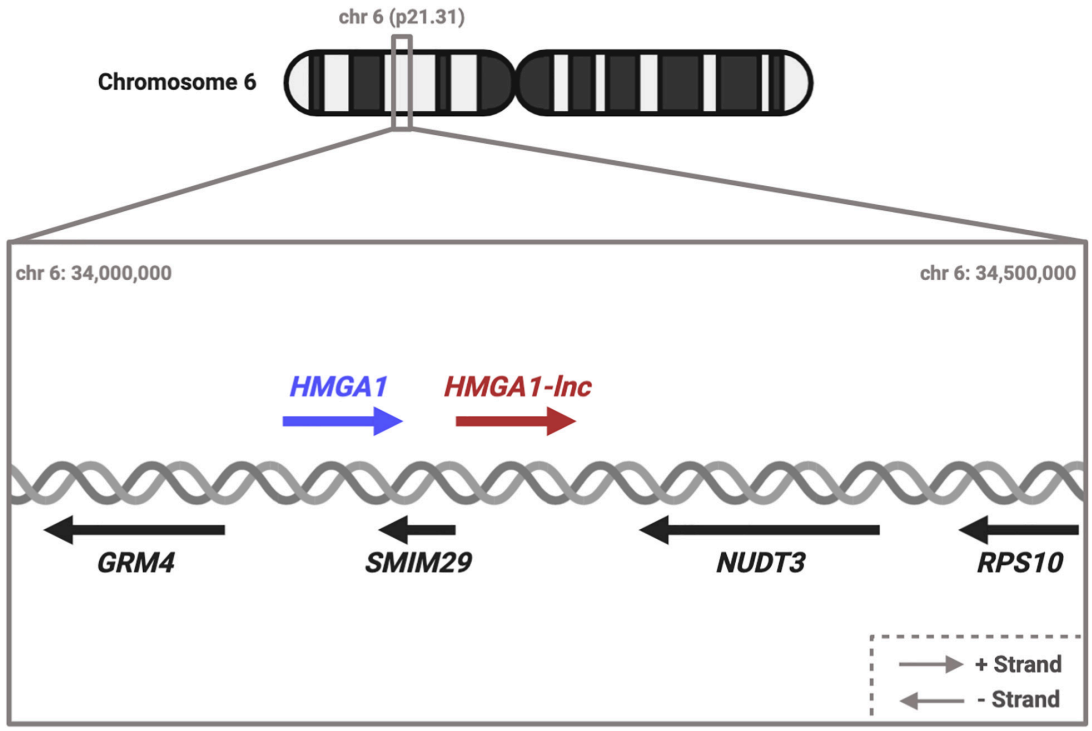


FIGURE 1 | Diagram of HMGA1 transcriptional locus. HMGA1 and HMGA1-lnc are neighbors on the + strand. Located on the opposite strand are genes SMIM29, NUDT3, GRM4, and RPS10. #Dereg: Deregulation.

mentioned above, the discordant expression in lung tumors between these genes lends further evidence to the existence of a possible regulatory relationship rather than a genomic-locus level alteration. To highlight this difference in expression, we compared the levels of *HMGA1* between tertiles of tumors with the highest and lowest expression of *HMGA1-lnc*. Interestingly we found that levels of *HMGA1* were significantly greater (p = 0.0326) in the low lncRNA expressing tumors (**Figures 2C,D**), and that *HMGA1* and *HMGA1-lnc* were negatively correlated (p = 0.0153). To determine if HMGA1-lnc was affecting other genes in its transcriptional locus we performed correlation analysis between HMGA1-lnc expression and other neighboring genes. Interestingly we found SMIM29, NUDT3, and RPS10 to be significantly positively correlated with HMGA1-lnc expression in the BCCA dataset, however, we did not observe significant correlations with these genes in the TCGA dataset (**Supplementary Table 2**). We also tested whether genes known to regulated by HMGA1 were affected by HMGA1-lnc expression. Interestingly we found two genes known to be downregulated by HMGA1 (CAV1 and FOXP1) to be positively correlated with HMGA1-lnc expression (**Supplementary Figure 2**). As the expression of HMGA1 and HMGA1-lnc appears to be anti-correlated, this lncRNA may be involved in the inhibition of *HMGA1* expression in normal lung contexts.

As expression of *HMGA1* has been previously described to increase with tumor stage and cancer aggressiveness, we examined whether expression of *HMGA1-lnc* was inversely associated with stage (Zhang et al., 2015). As the majority of our tumor samples were Stage | and || we performed a Mann-Whitney *U*-test between these two groups in our larger dataset (TCGA) to identify significant associations (**Figures 3A,B**). Interestingly, while *HMGA1* was associated with increased tumor stage (p = 0.0011), the opposite was true for *HMGA1-lnc*, where expression of the lncRNA is significantly decreased in more advanced tumors (p = 0.0125). Further, as the TCGA dataset has paired DNA methylation data in the same tumors, we were able to investigate whether expression of *HMGA1-lnc* was associated with changes in DNA methylation to the *HMGA1* locus. When we compared tumors with high and low expression of *HMGA1-lnc* (tertiles), we found that tumors with high *HMGA1-lnc* had significantly higher methylation of *HMGA1* (**Figure 3C**). We also find that *HMGA1* expression is significantly higher in tumors with low levels of HMGA1 methylation (**Figure 3D**). This may indicate that *HMGA1-lnc* plays a role in regulating the DNA methylation state of *HMGA1*.

HMGA1-lnc affects HMGA1 expression

To determine whether loss of *HMGA1-lnc* expression is a mechanism of *HMGA1* overexpression in the lung, we performed a siRNA-mediated knockdown of *HMGA1-lnc* in the BEAS-2Bs using a pool of siRNAs specific to *HMGA1-lnc*. As HMGA1-*lnc* is downregulated in tumors, we used a non-malignant lung cell line (BEAS-2Bs) to demonstrate the effect of downregulation of *HMGA1-lnc* on *HMGA1*. We then quantified expression changes using qRT-PCR as described in Methods. From this, we observed

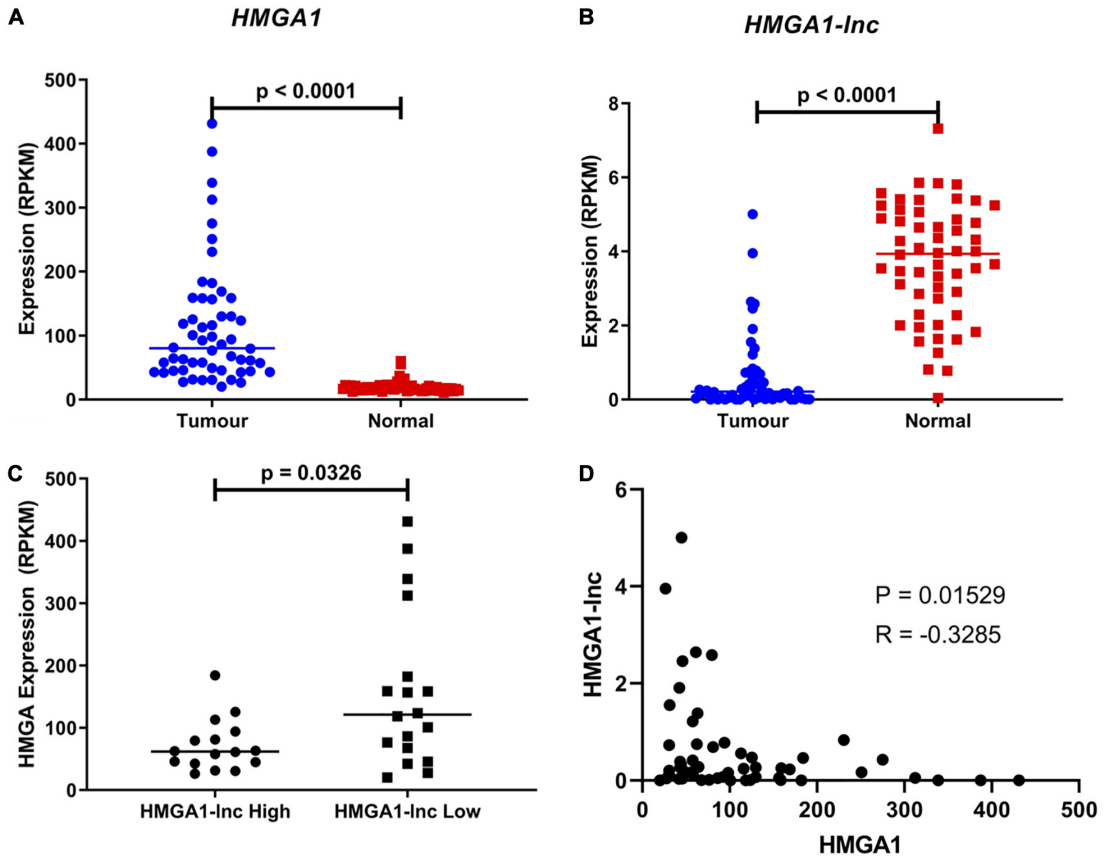


FIGURE 2 | Expression of *HMGA1* and *HMGA1-lnc* in LUAD. Expression of *HMGA1* is upregulated in LUAD compared to adjacent non-malignant tissue (Students *T*-test) **(A)** while conversely, expression of *HMGA1-lnc* is downregulated in tumors (Students *T*-test) **(B)**. Additionally, tumors with high levels of *HMGA1-lnc*, have significantly lower levels of *HMGA1*, when compared to tumors with low levels of the lncRNA (Mann-Whitney *U*-test) **(C)**, and expression of HMGA1 and HMGA1-lnc are negatively correlated (Spearman's Correlation) **(D)**. All data shown above is from the TCGA cohort (*n* = 108, 54 paired samples).

a 3.42-fold reduction of the lncRNA after 48 h. Strikingly, in the cells with reduced *HMGA1-lnc* expression, the mRNA expression levels of *HMGA1* were increased by 1.57-fold compared with cells transfected with non-targeting control siRNAs (**Figure 4**). This observed increase in *HMGA1* levels in the lncRNA-inhibited cell lines, suggests that *HMGA1-lnc* may act to negatively regulate the expression of *HMGA1*, and that downregulation of this lncRNA in LUAD may be a mechanism for whereby this well-known cancer-driving gene becomes overexpressed in certain LUAD tumors.

DISCUSSION

The role of protein-coding genes in the onset and progression of LUAD is well-established; however, there remains a lack of treatment options for patients who do not harbor one of the few clinically-actionable driver-gene alterations. LncRNAs have been shown to have important roles in the regulation of cancer-associated genes, but complex folding patterns and unknown binding motifs make lncRNAs particularly challenging to functionally characterize. Here, we used an approach that considered the genomic location, as well as the known function of neighboring oncogenic protein-coding genes, to find and characterize a novel *cis*-acting lncRNA deregulated in cancer, which we refer to as *HMGA1-lnc*.

Further, we found that this approach could be applied to other lncRNAs, finding several deregulated lncRNAs neighboring cancer associated protein-coding genes. This included well-known lung-cancer-associated genes such as *NEK2,* suggesting that there may be a selective pressure for the deregulation of these lncRNAs in order to release these cancer-promoting genes from negative regulation. Thus, alterations in lncRNA expression may consequently disrupt coding-gene expression as a means of promoting tumor development. This may be a useful methodology for researchers to use to identify other *cis*-acting lncRNAs that may be regulating genes of interest. However, it is worth noting that using a genomics-based approach to identify these *cis*-gene relationships is not without potential pitfalls. Previous studies have shown that many *cis*-acting lncRNAs have positive expression relationships with their neighbors in several tissue types (Balbin et al., 2015). However, genes in the same vicinity are often subject to regulation that can affect whole genomic regions, such as silencing through chromatin

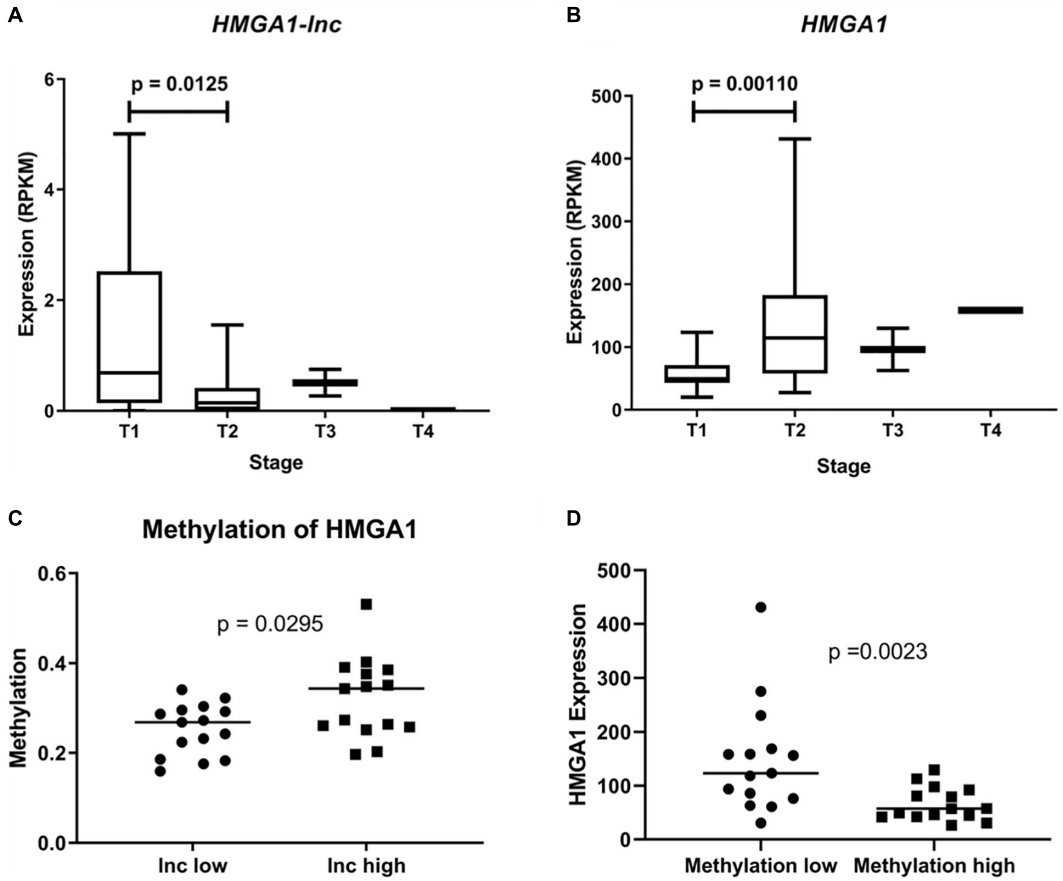


FIGURE 3 | Expression of *HMGA1* and *HMGA1-lnc* is associated with tumor stage. Expression of *HMGA1-lnc* decreases with increasing tumor stage **(A)**, whereas *HMGA1* expression increases with higher grade tumors **(B)**. Tumors with high levels of HMGA1-lnc expression (Tertiles) had significantly higher levels of HMGA1 methylation than tumors with low levels of HMGA1-*lnc* **(C)**. Expression of HMGA1 is higher in tumors with low levels of HMGA1 DNA methylation compared to tumors with high levels of DNA methylation (tertiles $n = 30$) **(D)**.

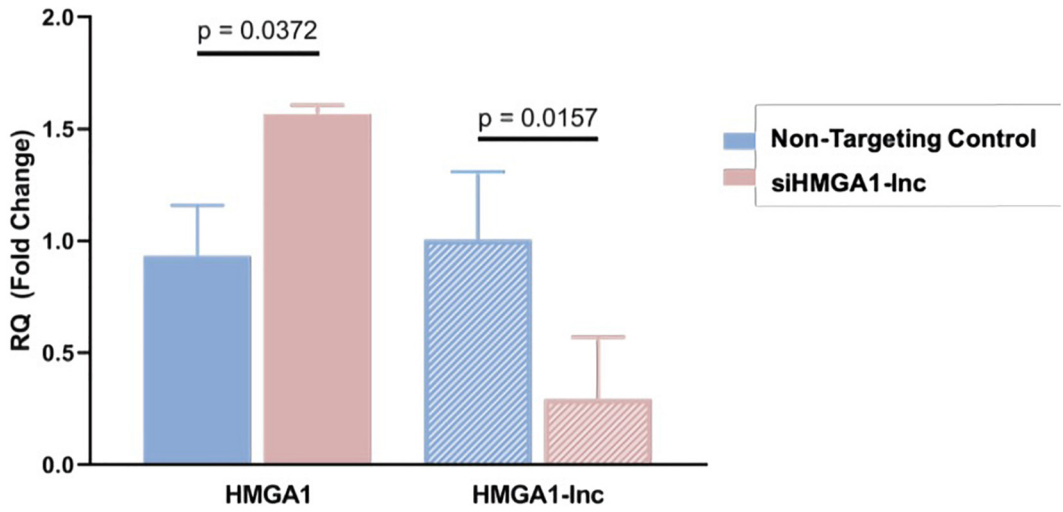


FIGURE 4 | Inhibition of HMGA1-lnc results in increases of HMGA1 expression. SiRNA mediated inhibition of *HMGA1-lnc* was performed in normal bronchial epithelial cells and resulted in significant reduction of the lncRNA HMGA1-lnc. Conversely, cells where the lncRNA was inhibited showed significant increases in protein-coding *HMGA1* expression.

condensation. In particular, tumors often have significantly elevated levels of these broad genomic alterations to the DNA, which enables tumor suppressor gene silencing or oncogene activation. Genes neighboring these oncogenes and tumor suppressors are often caught in these regions of alteration, and display concordant expression with these genes, a phenomenon known as the passenger effect (Sheltzer and Amon, 2011; Pon and Marra, 2015). For example, frequent DNA amplification of the *MET* oncogene occurs in 5–20% of LUAD, leading the surrounding genes to display significantly increased DNA copy number (Zack et al., 2013; Presutti et al., 2015). While it is difficult to separate *cis*-acting concordant regulatory relationships from oncogene passengers without further verifying direct interactions via *in vitro* expression modulation, genes displaying discordant expression relationships with their neighbors are less susceptible to this effect.

We sought to avoid the passenger effect pitfall of concordant regulatory relationships by focusing the majority of our analysis on discordant relationships, particularly that between *HMGA1-lnc* and *HMGA1*. We found *HMGA1-lnc* to be significantly downregulated in LUAD, where its expression level is decreased 15-fold in tumors, compared to *HMGA1* which has expression levels 5-fold greater in tumors (of the TCGA cohort). These observations in tandem with anti-correlated expression relationships within tumor samples led to our hypothesis that *HMGA1-lnc* acts to repress *HMGA1* expression in non-malignant samples. HMGA1 is known to interact with large transcriptional networks in order to drive cancers (Sumter et al., 2016). Interestingly we found that FOXP1 and CAV1, genes that are commonly repressed by HMGA1 in cancer, were positively associated with HMGA1-lnc expression (Treff et al., 2004; Schuldenfrei et al., 2011). This may indicate that HMGA1-lnc expression is able to have the opposite effect on these cancer-associated transcriptional networks (**Supplementary Figure 2**). Consequently, the finding that *HMGA1-lnc* was downregulated with increasing stage, while *HMGA1* expression increased with more advanced stages strengthened this putative regulatory relationship.

Methylation data for the TCGA samples allowed us to query whether expression of *HMGA1-lnc* affected the DNA methylation of *HMGA1*. Indeed, we found *HMGA-lnc* expression to be significantly correlated with DNA methylation of *HMGA1*. Further, other well-known *cis*-acting lncRNAs are known to function through regulation of the methylation state of their neighboring genes. For example, the lncRNA *TARID* which recruits GADD45A to actively demethylate the tumor suppressor TCF21 (Arab et al., 2014). This hints at one possible mechanism of *HMGA1-lnc* action, where expression may result in active methylation of *HMGA1*. We also tested the expression relationship between HMGA1-lnc and neighboring genes SMIM29, NUDT3, GRM4, and RPS10 to see if they were potentially affected by expression of this lncRNA. While we saw positive correlations for three of these genes in the BCCA dataset, these gene relationships were not consistent with the TCGA dataset (**Supplementary Table 2**). Further studies will be required to fully elucidate the effect of HMGA1-lnc on its neighboring genes, and the potential mechanism of its action.

To verify that *HMGA1-lnc* was able to affect *HMGA1* expression levels, we performed siRNA-mediated knockdown of the lncRNA in cells derived from normal lung epithelium (BEAS-2B cells). We chose to use an inhibitory model to best recapitulate the phenotype we seen in our RNA sequencing datasets where HMGA1-lnc is downregulated in tumor samples. Additionally, some common methods of gene modulation such as exogenous over-expression may not work well to activate *cis*-acting lncRNAs. For example, the function of some *cis*-acting lncRNAs is tied to the act of transcription, rather than the produced transcript itself (Margaritis et al., 2012; Pelechano and Steinmetz, 2013). When the lncRNA was inhibited we noted significant increases in *HMGA1* mRNA, confirming that this lncRNA directly regulates the expression of *HMGA1 in vitro*. Previous studies modulating expression levels of *HMGA1* in these same normal lung epithelial cells (BEAS-2B) have shown that increased *HMGA1* expression leads to transformed phenotypes and increases in anchorage-independent cell growth (Hillion et al., 2009). These results suggest that downregulation of this previously-uncharacterized lncRNA may lead to *HMGA1* upregulation, potentially driving the onset of these same cancer phenotypes in normal human lung epithelial cells.

Our study shows that a novel *cis*-acting lncRNA, *HMGA1-lnc*, is deregulated in LUAD, which represents an alternative mechanism of activation of the oncogene *HMGA1*. The methodology used in this work was able to identify candidate *cis*-acting lncRNAs that may regulate cancer driving genes. This approach may be useful in the study of other cancer types, where with different genetic backgrounds, other *cis*-acting lncRNAs may be regulating other known oncogenes or tumor suppressors specific to other cancer types. Interestingly, this lncRNA has been described previously in hepatocellular carcinoma (HCC), where its expression was used in an RNA-based biomarker panel used to differentiate HCC patients from patients with chronic hepatitis C virus, and healthy controls (Abd El Gwad et al., 2018). As HCC is a malignancy known to be driven by HMGA1 it would be interesting to determine if HMGA1-lnc was similarly downregulated in these tumors in order to deregulate expression of HMGA1. Further, as *HMGA1* is a known oncogene in other forms of malignancy, particularly breast cancer, it would be useful to determine if this lncRNA based mechanism is common across cancer types. If so, this interaction could represent a novel clinical intervention point in *HMGA1*-driven cancers.

MATERIALS AND METHODS

Sample collection and processing

Collection and sequencing of both cohorts were performed in congruent manners as described in a previous publication (Stewart et al., 2019). Two separate cohorts of raw RNA sequencing reads from LUAD tumors with matched adjacent non-malignant tissues were used in this study: an in-house microdissected cohort collected at the BC Cancer Research Centre (BCCA, n = 72, or 36 pairs) (**Supplementary Table 3**), and a secondary set of LUAD tumors and matched non-malignant tissue (TCGA, n = 108, or 54 pairs) were downloaded

from The Cancer Genome Atlas (TCGA) Data Portal[1]. The BCCA cohort was composed of fresh-frozen LUAD tumors and matched non-malignant lung parenchymal tissue collected from 36 patients at the Vancouver General Hospital. Consent obtained from the tissue donors of this study was both informed and written, and sample collection was approved from the University of British Columbia-BCCA Research Ethics Board. Matched non-malignant samples were collected from areas >2 cm away from the tumor. In order to reduce contaminating sequences derived from alternative cell types, tissue microdissection was guided by a pathologist. Samples used in this study contained >80% tumor cell or >80% non-malignant cell content. Total RNA was extracted using Trizol reagent and standard procedures.

RNA sequencing and processing

Libraries were constructed at Canada's Michael Smith Genome Sciences Center using total RNA extracted from tumors and matched normal tissue (GSC, Vancouver, Canada). RNA quality was assessed using the Agilent Bioanalyzer RNA nanochip, and arrayed into a 96 well plate. RNA containing poly A sequences were then purified using the MultiMACS mRNA isolation kit 2 μg total RNA with on-column DNaseI-treatment (Miltenyi Biotec, Germany). The Superscript Double-Stranded cDNA Synthesis kit was used to synthesize DS cDNA (Life Technologies, United States). Library construction was done in a paired end, strand-specific manner following the GSC library preparation protocol, and the Illumina HiSeq 2,000 platform was used for RNA sequencing. Raw sequencing reads were subject to quality control based on length (<50 nt, under two thirds of maximum read length of 75 nt) and quality level (Phred < 20 were discarded). STAR aligned (version 2.4.1d) was then used to align reads (.fastq) to the human genome (NCBI GRCh37) (Dobin et al., 2013). Quantification of aligned reads (.bam) was performed using the Ensemble Transcripts reference (Release 75) (Flicek et al., 2014). As described in a previous manuscript, raw sequencing reads for the BCCA dataset were deposited at the Bioproject http://www.ncbi.nlm.nih.gov/bioproject/516232 (Stewart et al., 2019). RNA sequencing data was downloaded for TCGA dataset from The Cancer Genome Atlas (TCGA) Data Portal[1]. Sequencing data (.bam files) was then processed as described above. Quantification of both datasets was performed using the Partek Flow platform as reads per kilobase million (RPKM).

Gene expression analyses

To identify genes deregulated in LUAD, we performed a Wilcoxon signed-rank test on both lncRNA and protein-coding gene expression between tumor and matched non-malignant tissues. lncRNAs significantly deregulated in the same direction in both cohorts [Benjamini Hochberg-corrected p-value < 0.05; Fold Change (FC) > 1.5] were considered for further analyses (**Supplementary Table 4**). To assess potential cancer-relevant *cis*-acting lncRNAs, we identified deregulated lncRNAs neighboring protein-coding genes that were also deregulated. We then queried the literature (using PubMed) for experimental evidence of deregulated protein-coding genes with tumor biology. LncRNAs close enough to enact transcriptional or epigenetic changes (within 1.5 Kb) to protein-coding genes were considered as putative *cis*-acting lncRNAs (**Supplementary Table 1**; Gil and Ulitsky, 2020). To reduce the effect of passenger effects (such as DNA copy number alterations) on lncRNA:protein-coding-gene relationships, we focused on gene pairs with discordant (negatively correlated) expression patterns. Significant associations between *HMGA1-lnc* and neighboring protein-coding *HMGA1* expression were determined using a Spearman's correlation. These results were confirmed and visualized using a Mann-Whitney *U*-Test between the upper and lower tertiles of samples based on lncRNA expression. To determine if *HMGA1-lnc* had protein coding potential we used the coding potential assessment tool (CPAT[2]) which identified the gene as non-coding (**Supplementary Figure 3**; Wang et al., 2013).

Methylation analysis of HMGA1

DNA methylation data (HM450 beta-values) of *HMGA1* for the TCGA LUAD dataset was downloaded from the cBioPortal for Cancer Genomics (www.cbioportal.org). Samples were then ranked by expression of HMGA1-*lnc* and separated into high and low tertiles. The methylation of *HMGA1* was then compared between the high and low lncRNA expressing tertiles (Mann-Whitney *U*-Test). To determine if DNA methylation of HMGA1 had an effect on HMGA1 expression we compared tumors with high levels of methylation to those with low levels of methylation (tertiles). We found that tumors with low level of HMGA1 DNA methylation had significantly higher levels of HMGA1 expression (Mann-Whitney *U*-Test).

In vitro analyses

The immortalized non-malignant epithelial lung cell line BEAS-2Bs was used to assess the effect of inhibition of the candidate lncRNA identified in the above analysis (referred to as: *HMGA1-lnc*) on *HMGA1* expression *in vitro*. Cells were cultured in serum-free medium: K-SFM supplemented with 30 μg/mL bovine pituitary extract (BPE) and 0.0002 ng/μL epidermal growth factor (EGF); maintained in an incubator at 37°C and 5% CO_2. Once confluent, 2 mL of cell solution was seeded into each well of a 6 × 2 cm-well plate at a concentration of 50,000 cells/mL. DharmaFECT siRNAs were prepared for transfection as per manufacturer's instructions in five conditions: (i) untreated control; (ii) a positive control siRNA targeting GAPDH (25 nM); (iii) a non-targeting control siRNA (25 nM); (iv) siRNA targeting *HMGA1-lnc* at a concentration of 12.5 nM; and (v) siRNA targeting *HMGA1-lnc* at a concentration of 25 nM (**Supplementary Tables 4, 5**). Non-targeting control was designed to target no known human genes and provide a baseline response for cellular exposure to siRNAs (Dharmacon, D-001210-01-D001210-05). RNA was harvested after both 48 and 72 h using the Quick-RNA™ MiniPrep Kit (Zymo Research, Catalog number R1055). Total RNA was converted to cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Catalogue number 4374967). Gene expression was assessed using real-time quantitative PCR with

[1]https://portal.gdc.cancer.gov/

[2]http://lilab.research.bcm.edu/cpat/index.php

custom primers specific to *HMGA1-lnc* generated by Thermo Fisher, as well as established primers for the *18S* ribosomal RNA (endogenous control), *GAPDH* (positive siRNA control), and *HMGA1*. RT-qPCR reactions were performed in triplicate as per Thermo Fisher recommended settings (denature 95°C for 15 s, anneal 60°C for 60 s, 40 PCR cycles). Relative expression was determined using the $2^{-\Delta\Delta Ct}$ method.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of British Columbia-BCCA Research Ethics Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GS and AS were responsible for manuscript conceptualization, study design, and performance of *in silico* and *in vitro* analyses. KE, EM, and DC contributed to data collection, interpretation, and manuscript preparation. WL was the principle investigator and was involved in manuscript conceptualization, study design, and manuscript preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Canadian Institutes for Health Research (CIHR FDN-143345). GS, AS, KE, EM, and DC are supported by Frederick Banting and Charles Best Canada Graduate Scholarships, and EAM is a Vanier Canada Graduate Scholar.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of Victor Martinez, Roland Hubaux, and Dave Rowbotham for helpful discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.615378/full#supplementary-material

REFERENCES

Abd El Gwad, A., Matboli, M., El-Tawdi, A., Habib, E. K., Shehata, H., Ibrahim, D., et al. (2018). Role of exosomal competing endogenous RNA in patients with hepatocellular carcinoma. *J. Cell. Biochem.* 119, 8600–8610. doi: 10.1002/jcb.27109

Amodio, N., Stamato, M. A., Juli, G., Morelli, E., Fulciniti, M., Manzoni, M., et al. (2018). Drugging the lncRNA MALAT1 via LNA gapmeR ASO inhibits gene expression of proteasome subunits and triggers anti-multiple myeloma activity. *Leukemia* 32, 1948–1957. doi: 10.1038/s41375-018-0067-3

Arab, K., Park, Y. J., Lindroth, A. M., Schäfer, A., Oakes, C., Weichenhan, D., et al. (2014). Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol. Cell* 55, 604–614. doi: 10.1016/j.molcel.2014.06.031

Balbin, O. A., Malik, R., Dhanasekaran, S. M., Prensner, J. R., Cao, X., Wu, Y. M., et al. (2015). The landscape of antisense gene expression in human cancers. *Genome Res.* 25, 1068–1079. doi: 10.1101/gr.180596.114

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

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., et al. (2014). Ensembl 2014. *Nucleic Acids Res.* 42, D749–D755. doi: 10.1093/nar/gkt1196

Fu, F., Wang, T., Wu, Z., Feng, Y., Wang, W., Zhou, S., et al. (2018). HMGA1 exacerbates tumor growth through regulating the cell cycle and accelerates migration/invasion via targeting miR-221/222 in cervical cancer. *Cell Death Dis.* 9:594. doi: 10.1038/s41419-018-0683-x

Fusco, A., and Fedele, M. (2007). Roles of HMGA proteins in cancer. *Nat. Rev. Cancer* 7, 899–910. doi: 10.1038/nrc2271

Gibb, E. A., Vucic, E. A., Enfield, K. S., Stewart, G. L., Lonergan, K. M., Kennett, J. Y., et al. (2011). Human cancer long non-coding RNA transcriptomes. *PLoS One* 6:e25915. doi: 10.1371/journal.pone.0025915

Gil, N., and Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* 21, 102–117. doi: 10.1038/s41576-019-0184-5

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

Hillion, J., Wood, L. J., Mukherjee, M., Bhattacharya, R., Di Cello, F., Kowalski, J., et al. (2009). Upregulation of MMP-2 by HMGA1 promotes transformation in undifferentiated, large-cell lung cancer. *Mol. Cancer Res.* 7, 1803–1812. doi: 10.1158/1541-7786.MCR-08-0336

Lin, S. Y., and Peng, F. (2016). Association of SIRT1 and HMGA1 expression in non-small cell lung cancer. *Oncol. Lett.* 11, 782–788. doi: 10.3892/ol.2015.3914

Margaritis, T., Oreal, V., Brabers, N., Maestroni, L., Vitaliano-Prunier, A., Benschop, J. J., et al. (2012). Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3′-end antisense transcription. *PLoS Genet.* 8:e1002952. doi: 10.1371/journal.pgen.1002952

Pallante, P., Sepe, R., Puca, F., and Fusco, A. (2015). High mobility group a proteins as tumor markers. *Front. Med.* 2:15. doi: 10.3389/fmed.2015.00015

Pelechano, V., and Steinmetz, L. M. (2013). Gene regulation by antisense transcription. *Nat. Rev. Genet.* 14, 880–893. doi: 10.1038/nrg3594

Pon, J. R., and Marra, M. A. (2015). Driver and passenger mutations in cancer. *Annu Rev. Pathol.* 10, 25–50. doi: 10.1146/annurev-pathol-012414-040312

Presutti, D., Santini, S., Cardinali, B., Papoff, G., Lalli, C., Samperna, S., et al. (2015). MET Gene Amplification and MET Receptor Activation Are Not Sufficient to Predict Efficacy of Combined MET and EGFR Inhibitors in EGFR TKI-Resistant NSCLC Cells. *PLoS One* 10:e0143333. doi: 10.1371/journal.pone.0143333

Schmitt, A. M., and Chang, H. Y. (2016). Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* 29, 452–463. doi: 10.1016/j.ccell.2016.03.010

Schuldenfrei, A., Belton, A., Kowalski, J., Talbot, C. C. Jr., Di Cello, F., Poh, W., et al. (2011). HMGA1 drives stem cell, inflammatory pathway, and cell cycle progression genes during lymphoid tumorigenesis. *BMC Genom.* 12:549. doi: 10.1186/1471-2164-12-549

Sgarra, R., Pegoraro, S., Ros, G., Penzo, C., Chiefari, E., Foti, D., et al. (2018). High Mobility Group A (HMGA) proteins: Molecular instigators of breast cancer onset and progression. *Biochim. Biophys. Acta Rev. Cancer* 1869, 216–229. doi: 10.1016/j.bbcan.2018.03.001

Sheltzer, J. M., and Amon, A. (2011). The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends Genet.* 27, 446–453. doi: 10.1016/j.tig.2011.07.003

Stewart, G. L., Enfield, K. S. S., Sage, A. P., Martinez, V. D., Minatel, B. C., Pewarchuk, M. E., et al. (2019). Aberrant Expression of Pseudogene-Derived lncRNAs as an Alternative Mechanism of Cancer Gene Regulation in Lung Adenocarcinoma. *Front. Genet.* 10:138. doi: 10.3389/fgene.2019.00138

Sumter, T. F., Xian, L., Huso, T., Koo, M., Chang, Y. T., Almasri, T. N., et al. (2016). The High Mobility Group A1 (HMGA1) Transcriptome in Cancer and Development. *Curr. Molecular Med.* 16, 353–393. doi: 10.2174/1566524016666160316152147

Treff, N. R., Pouchnik, D., Dement, G. A., Britt, R. L., and Reeves, R. (2004). High-mobility group A1a protein regulates Ras/ERK signaling in MCF-7 human breast cancer cells. *Oncogene* 23, 777–785. doi: 10.1038/sj.onc.1207167

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

Xia, J., Franqui Machin, R., Gu, Z., and Zhan, F. (2015). Role of NEK2A in human cancer and its therapeutic potentials. *Biomed. Res. Int.* 2015:862461. doi: 10.1155/2015/862461

Yao, Y., Su, J., Zhao, L., Luo, N., Long, L., and Zhu, X. (2019). NIMA-related kinase 2 overexpression is associated with poor survival in cancer patients: a systematic review and meta-analysis. *Cancer Manag. Res.* 11, 455–465. doi: 10.2147/CMAR.S188347

Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., et al. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* 38, 662–674. doi: 10.1016/j.molcel.2010.03.021

Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., et al. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* 45, 1134–1140. doi: 10.1038/ng.2760

Zhang, Z., Wang, Q., Chen, F., and Liu, J. (2015). Elevated expression of HMGA1 correlates with the malignant status and prognosis of non-small cell lung cancer. *Tumour Biol.* 36, 1213–1219. doi: 10.1007/s13277-014-2749-4

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

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 15 January 2021
doi: 10.3389/fgene.2020.611823

LncRNA H19 Suppresses Osteosarcomagenesis by Regulating snoRNAs and DNA Repair Protein Complexes

An Xu[1†], Mo-Fan Huang[1†], Dandan Zhu[1], Julian A. Gingold[2], Danielle A. Bazer[3], Betty Chang[4,5,6], Donghui Wang[1], Chien-Chen Lai[7,8], Ihor R. Lemischka[4,5,6], Ruiying Zhao[1] and Dung-Fang Lee[1,9,10,11]**

[1] Department of Integrative Biology and Pharmacology, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, United States, [2] Department of Obstetrics and Gynecology and Women's Health, Einstein/Montefiore Medical Center, Bronx, NY, United States, [3] Department of Neurology, Renaissance School of Medicine at Stony Brook University, Stony Brook, NY, United States, [4] Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [5] The Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [6] The Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [7] Institute of Molecular Biology, National Chung Hsing University, Taichung, Taiwan, [8] Graduate Institute of Chinese Medical Science, China Medical University, Taichung, Taiwan, [9] The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX, United States, [10] Center for Stem Cell and Regenerative Medicine, The Brown Foundation Institute of Molecular Medicine for the Prevention of Human Diseases, The University of Texas Health Science Center at Houston, Houston, TX, United States, [11] Center for Precision Health, School of Biomedical Informatics and School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

Osteosarcoma is one of the most frequent common primary malignant tumors in childhood and adolescence. Long non-coding RNAs (lncRNAs) have been reported to regulate the initiation and progression of tumors. However, the exact molecular mechanisms involving lncRNA in osteosarcomagenesis remain largely unknown. Li-Fraumeni syndrome (LFS) is a familial cancer syndrome caused by germline p53 mutation. We investigated the tumor suppressor function of lncRNA H19 in LFS-associated osteosarcoma. Analyzing H19-induced transcriptome alterations in LFS induced pluripotent stem cell (iPSC)-derived osteoblasts, we unexpectedly discovered a large group of snoRNAs whose expression was significantly affected by H19. We identified SNORA7A among the H19-suppressed snoRNAs. SNORA7A restoration impairs H19-mediated osteogenesis and tumor suppression, indicating an oncogenic role of SNORA7A. TCGA analysis indicated that SNORA7A expression is associated with activation of oncogenic signaling and poor survival in cancer patients. Using an optimized streptavidin-binding RNA aptamer designed from H19 lncRNA, we revealed that H19-tethered protein complexes include proteins critical for DNA damage response and repair, confirming H19's tumor suppressor role. In summary, our findings demonstrate a critical role of H19-modulated SNORA7A expression in LFS-associated osteosarcomas.

Keywords: osteosarcoma, H19 lncRNA, iPSCs, Li-Fraumeni syndrome, snoRNA, p53

OPEN ACCESS

Edited by:
Liang Chen,
Wuhan University, China

Reviewed by:
Matthias S. Leisegang,
Goethe University Frankfurt, Germany
Dandan Li,
Nanyang Normal University, China

***Correspondence:**
Dung-Fang Lee
dung-fang.lee@uth.tmc.edu
Ruiying Zhao
ruiying.zhao@uth.tmc.edu

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *29 September 2020*
Accepted: *15 December 2020*
Published: *15 January 2021*

Citation:
Xu A, Huang M-F, Zhu D, Gingold JA, Bazer DA, Chang B, Wang D, Lai C-C, Lemischka IR, Zhao R and Lee D-F (2021) LncRNA H19 Suppresses Osteosarcomagenesis by Regulating snoRNAs and DNA Repair Protein Complexes. Front. Genet. 11:611823. doi: 10.3389/fgene.2020.611823

INTRODUCTION

Li-Fraumeni syndrome (LFS) (OMIM #151623) is a rare familial cancer syndrome characterized by early onset of various tumors, soft-tissue sarcomas, osteosarcomas, breast cancers, brain tumors, adrenocortical carcinomas, and leukemia (Zhou et al., 2017). Germline mutations in the p53 tumor suppressor gene are responsible for LFS. Patient-derived iPSCs have been used to model various diseases (Lee et al., 2009b; Carvajal-Vergara et al., 2010; Itzhaki et al., 2011; Yagi et al., 2011; Mulero-Navarro et al., 2015; Gingold et al., 2016; Liu et al., 2018; Zhu et al., 2018). The development of various refined differentiation protocols utilizing induced pluripotent stem cells (iPSCs) has enabled the production of large quantities of differentiated cells from individual patients. In our previous studies, using a LFS patient-derived iPSC model, we demonstrated the tumor suppressor role of lncRNA H19 and the oncogenic role of SFRP2 during the formation of osteosarcoma in LFS patients (Lee et al., 2015; Lin et al., 2017b; Kim et al., 2018; Zhou et al., 2018). Although H19-mediated tumor suppression has been demonstrated in LFS-associated osteosarcomagenesis, the underlying mechanisms of its tumor suppressor activity remain unclear.

Long non-coding RNAs (lncRNAs) represent a class of nonprotein-coding RNAs longer than 200 nucleotides (Sanchez Calle et al., 2018). Despite not being translated into proteins, lncRNAs are important regulators of diverse biological processes and pathologies. LncRNAs are proposed to function via multiple mechanisms, including co-transcriptional regulation, bridging proteins and chromatin, offering cytoplasmic scaffolding, pairing with other RNAs, and serving as molecular decoys (Ulitsky and Bartel, 2013).

One of the most studied lncRNAs is the imprinted lncRNA H19, located on human chromosome 11 and expressed exclusively from the maternal allele (Gabory et al., 2009). Numerous functional studies have assessed the role of H19 in the pathogenesis of human cancers and yielded conflicting results. On one hand, H19 is a precursor of miR-675 (Keniry et al., 2012) as well as a "molecular sponge" for soaking up microRNAs let-7 (Kallen et al., 2013; Li et al., 2020) and miR-138 (Liang et al., 2015), supporting a role in promoting cancer cell proliferation and migration. On the other hand, *in vivo* mouse (Yoshimizu et al., 2008) and LFS iPSC-derived osteoblast (Lee et al., 2015) studies demonstrated that H19 displays a tumor-suppressive effect. Evidence of numerous human tumors displaying either overexpression or lack of expression of H19 (Hao et al., 1993; Lustig-Yariv et al., 1997; Yoshimizu et al., 2008; Lee et al., 2015; Li et al., 2020) suggests the possibility of context-dependent oncogenic and tumor-suppresive roles.

Small nucleolar RNAs (snoRNAs) are small non-coding RNAs 60–300 nucleotides in length, primarily located in the nucleolus. The main functions of snoRNAs are to assist with post-transcriptional modification and maturation of ribosomal RNAs (rRNAs), small nuclear RNA (snRNAs), and other cellular RNAs. SnoRNAs can interact with RNA binding proteins to form small nuclear ribonucleoproteins. SnoRNAs are divided into two classes according to their catalytic activity: C/D box snoRNAs, catalyzing 2-O-ribose methylation; and H/ACA box snoRNAs, catalyzing pseudouridylation (Bachellerie et al., 2002). Emerging evidence indicates that snoRNAs are widely involved in various cancer-related signaling pathways. For example, SNORA42 is overexpressed in non-small cell lung cancer (NSCLC) and plays an oncogenic role through suppressing p53 function and/or expression (Mei et al., 2012). SNORD76 leads to the activation of the WNT/β-Catenin pathway to promote hepatocellular carcinoma (HCC) tumorigenicity (Wu et al., 2018).

The DNA damage/repair response plays a key role in maintaining genome integrity and stability, and its dysfunction girds the development and progression of various cancer types. Recently, studies revealed that lncRNA also regulates DNA damage response and DNA repair through the ATM/ATR (Wan et al., 2013), p53 regulatory network (Hu et al., 2018), and DNA double-strand break repair pathway (Gazy et al., 2015). However, it remains unclear whether H19 is actively involved in DNA damage/repair response.

Our current study reveals that H19 functions as a tumor suppressor through negatively regulating the expression of oncogenic SNORA7A in the osteoblast context. Furthermore, by exploring its interactions with multiple essential DNA damage/repair response factors, we demonstrate how H19 executes its tumor-suppressive role.

MATERIALS AND METHODS

Cell Cultures

U2OS cells were maintained in DMEM (Invitrogen, USA) supplemented with 10% FBS and 1% penicillin/streptomycin. iPSCs were maintained on Matrigel (Corning, USA)-coated plates in StemMACS™ iPS-Brew XF medium (MiltenyiBiotec, USA). All cells were cultured in a humidified incubator at 37°C and 5% CO2 and were tested to exclude mycoplasma contamination.

Differentiation of iPSCs to Mesenchymal Stem Cells (MSCs) and Then to Osteoblasts

In vitro differentiation of LFS iPSCs to MSCs was performed by a PDGF-AB-based method described previously (Lian et al., 2007; Zhou et al., 2018). Appropriately differentiated CD105+/CD166+/CD24- MSCs were plated in a 6-well plate at a density of 2×10^4 cells per well and cultured in an osteogenic differentiation medium (α-MEM supplemented with 10% FBS, 10 mM β-glycerol phosphate, 200 μM ascorbic acid, and 0.1 μM dexamethasone) to induce osteogenic differentiation as previously described (Lee et al., 2015).

RNA Isolation and RNA-Sequencing

LFS osteoblasts with ectopic H19 expression and vector control osteoblasts were collected and compared for H19-mediated tumor suppressor effects. Total mRNA was isolated using TRIzol reagent (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. All RNA sample preparation and RNA-sequencing (RNA-seq) data analyses were performed as described previously (Lee et al., 2015). RNA was aggregated

from biological triplicate experiments prior to running as a single sample for RNA-seq. The FPKM (fragments per kilobase of exon model per million reads mapped) of genes, lncRNAs, snoRNAs, and miRNAs were calculated and summarized in **Supplementary Table 1**.

Enrichr Analysis

GO Biological Process (GO_BP) and Wikipathway analyses were performed using Enrichr (https://amp.pharm.mssm.edu/Enrichr/) (Chen et al., 2013; Kuleshov et al., 2016) to identify enriched biological processes and pathways in H19-expressing LFS osteoblasts. Processes and pathways enriched with a $p < 0.05$ were considered significant. These genes were analyzed in Enrichr using Wikipathway and GO_BP pathway datasets to identify enriched GO terms, pathways, and functions. The TRANSFAC and JASPAR PWMs databases of transcription factors were used in Enrichr to identify transcription factors positively and negatively correlated with SNORA7A expression.

snoRNA Analysis

To determine pathological functions in osteosarcomagenesis significantly associated with SNORA7A, we identified genes positively and negatively correlated with SNORA7A expression in the TCGA-SARC dataset from SNORic (snoRNA in cancers; http://bioinfo.life.hust.edu.cn/SNORic/basic/) (Gong et al., 2017). Correlation of SNORA7A and mRNA expression of significant genes was run at default settings. Clinical analysis of 5-year survival in different cancers was analyzed from TCGA.

Plasmid Construction

A modified 1x streptavidin-binding RNA aptamer (S1m) (Leppek and Stoecklin, 2014) was synthesized (Integrated DNA Technologies, USA) from DNA oligo pairs (S1m_Forward: AATTGgtagaaaATGCGGCCGCCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGTCGGCGGCCGCATctgctgggG; S1m_Reverse: AATTCcccagcagATGCGGCCGCCGACCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGCGGCCGCATtttctacC), annealed, and ligated into the multiple cloning sites of the tetracycline-inducible TetO-FUW vector (Lee et al., 2015) 4 times sequentially. The TetO-FUW vector containing 4×streptavidin-binding RNA aptamer was named TetO-FUW-4S1m. H19 was then cloned into the N terminal side of 4S1m to form TetO-FUW-4S1m-H19. For the SNORA7A expression construct, the synthesized full-length SNORA7A (Ensembl Transcript ID: ENST00000384765.1) was cloned into the pLKO.pig plasmid (Lee et al., 2012a,b) within EcoRI-AgeI cloning sites and confirmed by Sanger sequencing.

qRT-PCR

Total RNA was isolated using TRIzol reagent (Thermo Fisher Scientific, USA) and snoRNA was isolated using mirVana miRNA isolation kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. The qRT-PCR primers for GAPDH (internal control), BGLAP, MEPE, FGF23, H19, and SNORA7A were described previously (Lee et al., 2015; Zhang et al., 2017).

Virus Packaging and Infection

TetO-FUW-4S1m or TetO-FUW-4S1m-H19 plasmids were co-transfected with packaging plasmids psPAX2 (Addgene, plasmid # 12260) and pMD2.G (Addgene, plasmid # 12259) into HEK-293T cells, and the virus collected from the cell culture medium 48 hours later. Osteosarcoma cell line U2OS was infected with viral particles together with the M2rtTA virus (Addgene, plasmid # 20342), which was similarly produced in HEK-293T cells, in the presence of 8ug/ml polybrene (Sigma-Aldrich, USA). Thirty-six hours post-infection, U2OS cells were treated with 1μg/ml doxycycline (Sigma-Aldrich, USA) for 24 hours to induce the expression of 4S1m or 4S1m-H9.

In vitro Anchorage-Independent Growth (AIG) Assay

AIG assay was performed as described previously (Lee et al., 2015). Briefly, LFS MSCs were transduced with H19 and/or SNORA7A. 10,000 LFS MSCs were mixed with 0.5% UltraPure low-melting-point agarose (Thermo Fisher Scientific, USA) and cultured in an osteogenic differentiation medium for 1 month. Colonies (≥50 μm) were counted under a Leica microscope DMi8.

H19-Interacting Protein Complex Purification and Mass Spectrometry Analysis

H19-interacting proteins were purified via the streptavidin-binding aptamers S1m as described previously (Leppek and Stoecklin, 2014). Briefly, S1m-H19 expressing U2OS cells were resuspended in 500 μl ice-cold lysis buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1.5 mM $MgCl_2$, 2 mM DTT, 2 mM RNase inhibitor, and complete protease inhibitors cocktail). Lysates were subjected to centrifugation for 5 min at 12,000 rpm at 4°C. The supernatants were incubated with streptavidin agarose beads (Thermo Fisher Scientific, USA) overnight at 4°C under rotation. The streptavidin agarose beads were then washed five times for 5 min at 4°C with wash buffer [20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5 mM$MgCl_2$, and 2 mM DTT]. The 4S1m-H19 associated protein complex samples were subjected to SDS-PAGE electrophoresis and micro-liquid chromatography/tandem mass spectrometry as described previously (Lee et al., 2007, 2009a, 2012a). Mass spectrometry analysis results are summarized in **Supplementary Table 2**.

Statistical Analyses

Results were presented as mean ± standard error of the mean (SEM). Error bars in figures represent SEM. Differences between two groups were examined by the two-tailed unpaired Student *t*-test. $*p < 0.05$; $**p < 0.01$; and $***p < 0.001$.

Data Availability

The data supporting the findings are available within the manuscript text, figures, and **Supplementary Tables 1, 2**. The RNA-seq data are available at the sequencing read archive (SRA) under accession number PRJNA673185.

RESULTS

H19 Modulates snoRNA Expression in LFS Osteoblasts

To explore the suppressive effects of H19 on osteosarcomagenesis, we analyzed the genome-wide transcriptomes of LFS iPSC-derived osteoblasts by RNA-seq following ectopic overexpression of H19. The RNA-seq experiment was performed with either vector- or H19-expressing LFS osteoblasts ($n = 1$) after pooling RNA from biological triplicate experiments. Transcriptome analysis confirmed ectopic H19 overexpression and identified a set of significantly altered genes [107 upregulated genes (FPKM $\geq$1 and fold change $\geq$5) and 81 downregulated genes (FPKM $\geq$1 and fold change $\geq$5)]. These H19-regulated genes were analyzed for enriched Gene Ontology biological processes (GO_BP) and Wikipathways using the comprehensive gene set enrichment analysis web tool Enrichr. GO_BP analyses showed that the GO_BPs enriched in H19-expressing LFS osteoblasts compared with vector-expressing (H19-depleted) LFS osteoblasts included ketone pathway, regulation of complement activation, skeletal system development, and positive regulation of cell death. In contrast, H19-depleted LFS osteoblasts demonstrated enrichment of GO_BPs including pre-mRNA cleavage required for polyadenylation, negative regulation of the cellular process, negative regulation of transport, and extracellular matrix organization (**Figure 1A**). Wikipathway analyses indicated that H19-overexpressing LFS osteoblasts were enriched for genes in the complement and coagulation cascade, complement activation, oxidative damage, prostaglandin synthesis, and regulation, TYROBP causal network, benzopyrene metabolism, apoptosis, and CCK2R signaling. In contrast, genes enriched in H19-depleted LFS osteoblasts were involving in methylation pathways, biogenic amine synthesis, estrogen metabolism, and hypertrophy model (**Figure 1B**).

In analyzing the transcriptome data, we classified RNA alterations into three groups by their fold-change: less than 2, between 2 and 5, and higher than 5. We unexpectedly found that a much larger share of miRNAs and snoRNAs increased in expression following H19 expression in LFS osteoblasts compared with protein-coding mRNAs, lncRNAs, etc. (i.e all transcripts except miRNAs and snoRNAs). H19 restoration in LFS osteoblasts led to a more than 5-fold change (increased or decreased) in expression in 0.96% of protein-coding mRNAs, lncRNAs, etc. with the majority of mRNAs and lncRNAs changing <2-fold (FPKM<1) (**Figure 1C**). In contrast, 16% of miRNAs and 28.4% of snoRNAs changed greater than 5-fold in expression (**Figures 1D,E**). Scatter plots comparing miRNA and snoRNA expression between H19-expressing and vector control LFS osteoblasts showed that miRNAs (e.g., MIR570, MIR574, MIR943, MIR635, MIR7-1, and MIR145) and snoRNAs (e.g., SNORA6, SNORA7A, SNORA9, SNORA15B, SNORA19B, and SNORA36C) were significantly altered in H19-expressing LFS osteoblasts (**Figures 1D,E**). These findings suggest that H19 suppresses osteosarcomagenesis in LFS patients by primarily altering the expression of miRNAs and snoRNAs rather than mRNAs.

SnoRA7A Is an onco-snoRNA Involved in Tumor Progression and Associated With Poor Prognosis

Among the identified H19-regulated miRNAs and snoRNAs, SNORA7A was chosen for further study in light of its previously recognized role in controlling the self-renewal of MSCs (Zhang et al., 2017). To validate our transcriptome results, we ectopically expressed H19 in LFS iPSC-derived osteoblasts and found that H19 significantly inhibits SNORA7A expression (**Figure 2A**). H19-induced osteogenic gene expression (e.g., BGLAP, MEPE, and FGF23) was inhibited by SNORA7A (**Figure 2B**), indicating that SNORA7A inhibits osteogenesis. To investigate whether the inhibition of SNORA7A plays a role in H19-mediated tumor suppression, we performed an *in vitro* AIG assay and found retarded clonal growth of H19-transduced LFS osteoblasts upon SNORA7A ectopic expression (**Figure 2C**). These results suggest that SNORA7A is negatively regulated by H19, and that SNORA7A functions as an onco-snoRNA by antagonizing H19 tumor suppressor function.

To further explore the role of SNORA7A in tumorigenesis, we analyzed the TCGA-SARC dataset and identified 780 and 945 protein-coding genes whose mRNA levels were positively and negatively correlated with SNORA7A expression, respectively. Wikipathway analysis indicated that genes positively correlated with SNORA7A expression are mainly involved in cytoplasmic ribosomal protein function and cell metabolism (e.g., pentose phosphate metabolism, Cori cycle, pyrimidine metabolism, and glycolysis and gluconeogenesis), while genes negatively correlated with SNORA7A expression are associated with focal adhesion, integrin-mediated cell adhesion, and G protein signaling pathways, the latter of which is involved in osteoblast differentiation (Wu et al., 2010) (**Figure 2D**).

Similarly, GO_BP analysis demonstrated that genes positively correlated with SNORA7A expression are involved in rRNA metabolic process and processing, ribosome biogenesis, translation, glyceraldehyde-3-phosphate metabolic process, and ribose phosphate metabolic process, while genes negatively correlated with SNORA7A expression are associated with glycogen metabolic process and cytosolic calcium ion transport (**Figure 2E**). The molecules controlling pentose phosphate metabolism (G6PD), protein translation (RPS9), and cell cycle progression (CDK4) were positively correlated with SNORA7A expression. In contrast, cyclin-dependent kinase inhibitor (CDKN1B), adenylyl cyclase (ADCY9), and downstream effector of Rho (ROCK1) were negatively correlated with SNORA7A expression (**Figure 2F**). These systems analyses suggest that SNORA7A may transcriptionally and/or post-transcriptionally regulate multiple oncogenic features, including cellular metabolism, ribosome biogenesis, cell cycle, etc., culminating in osteosarcoma development in LFS patients.

The regulation of RNA splicing is increasingly recognized to be an essential mechanism underlying cancer development, and dysregulation of RNA splicing machinery has been found to contribute to tumorigenesis in various human cancers, including glioma, lymphoma, and breast cancer (David et al., 2010; Hsu et al., 2015; Koh et al., 2015). Given the well-recognized

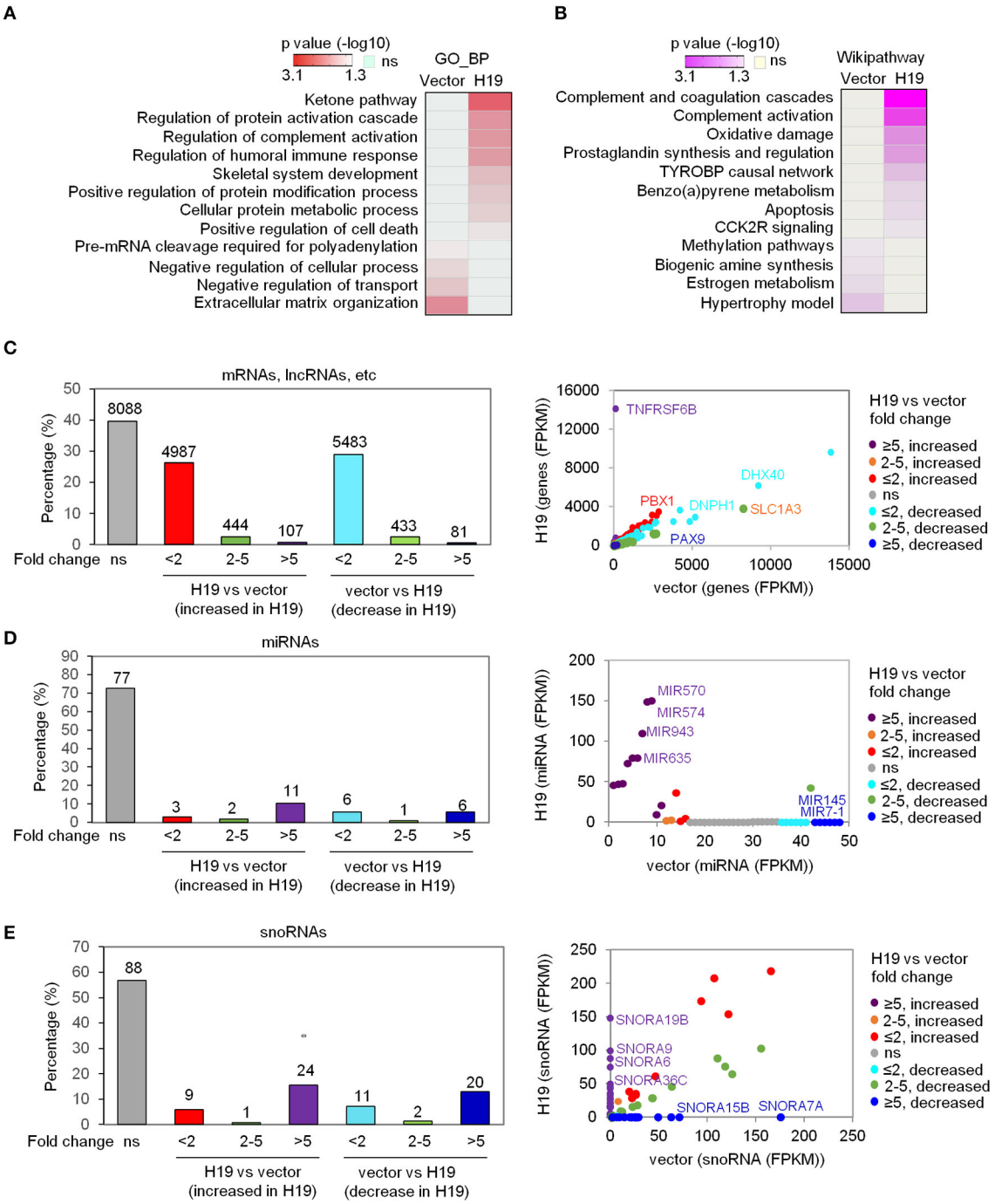


FIGURE 1 | H19 significantly affects miRNA and snoRNA expression. **(A,B)** Heatmaps show the GO_BP **(A)** and Wikipathway **(B)** categories enriched in either H19-expressing LFS osteoblasts (right column) or vector control LFS osteoblasts (left column) compared with each other based on mRNA expression of respective samples. Colors represent *p* values of enrichment. **(C-E)** H19 overexpression induces small changes in overall mRNA expression but large changes in miRNA and snoRNA expression in LFS osteoblasts. The percentage of mRNAs, lncRNAs, etc. **(C)**, miRNAs **(D)**, and snoRNAs **(E)** with different fold-changes (<2, 2-5, and >5 increase or decrease) are shown in bar plots. RNA expression between H19-overexpressing and vector control cells is presented in a scatter plot (right). Dot colors are used to represent the fold change of RNA expression. ns (not significant) indicates mRNAs, lncRNAs, miRNAs, snoRNAs, etc. with fold-change <2 and/or FPKM<1. The RNA-seq experiment was done with one sample pooled from 3 biological replicates. miRNAs and snoRNAs are defined based on the miRbase and snoDB database, respectively. "mRNA, lncRNA, etc." indicates all transcripts except miRNAs and snoRNAs.

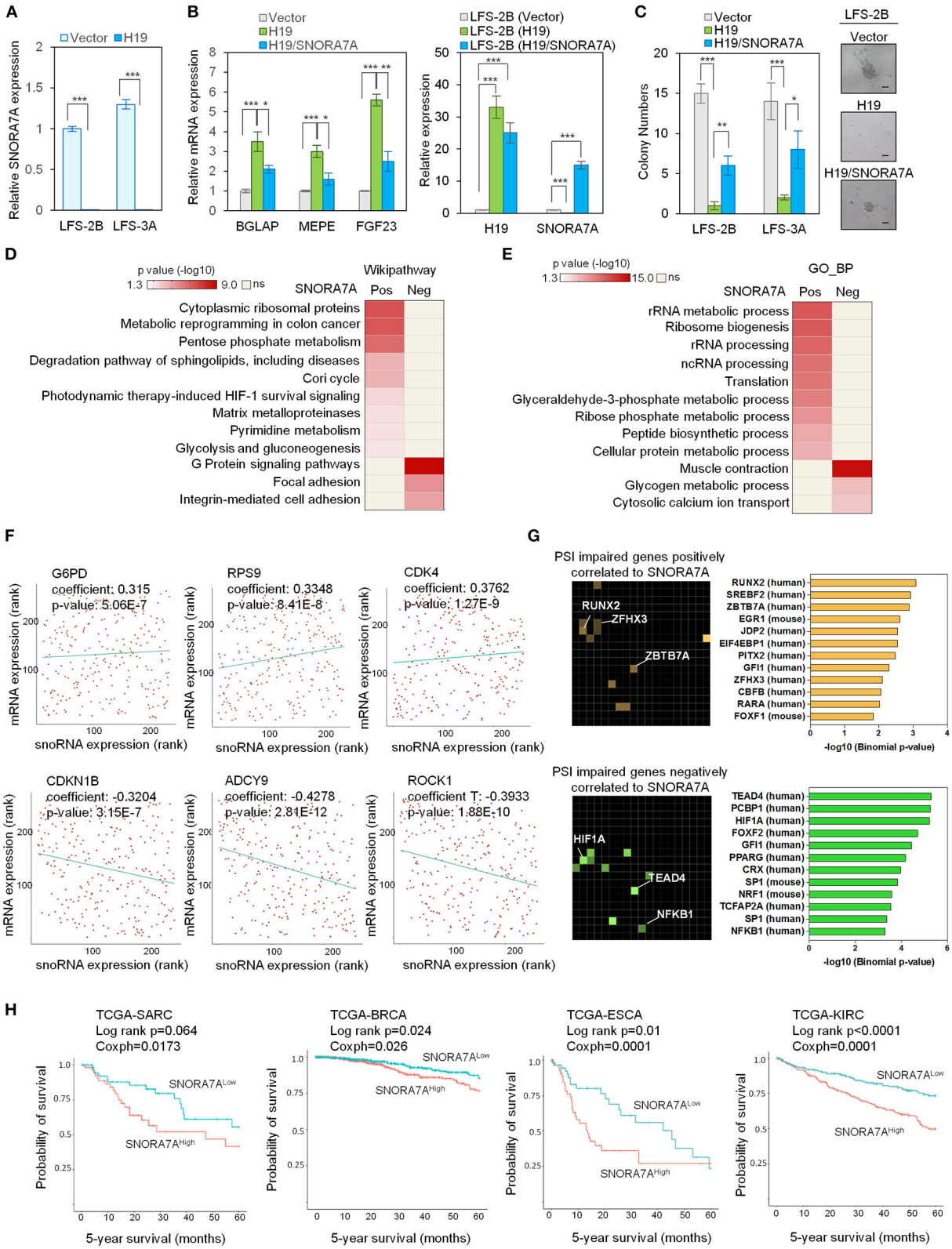


FIGURE 2 | Systems analyses of the oncogenic role of SNORA7A from TCGA datasets. **(A)** Ectopic expression of H19 downregulates SNORA7A expression in LFS osteoblasts. qRT-PCR data are represented as mean ±SEM; $n = 3$ biological replicates; statistical significance is determined using two-tailed Student's *t*-test;

(Continued)

FIGURE 2 | ***$p < 0.001$. **(B)** Left panel, ectopic expression of H19 increases osteogenic gene expression in LFS osteoblasts; in contrast, restoration of SNORA7A expression impairs H19-upregulated osteogenic gene expression in H19-transduced LFS osteoblasts. Right panel, qRT-PCR results demonstrate the expression of H19 and SNORA7A upon their ectopic expression for assays in the left panel. qRT-PCR data are represented as mean ± SEM; $n = 3$ biological replicates; statistical significance is determined using two-tailed Student's *t*-test; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. **(C)** AIG assay for *in vitro* tumorigenicity demonstrates that H19 impairs the colony numbers of LFS osteoblasts and SNORA7A expression rescues the H19-suppressed tumorigenicity of LFS osteoblasts. H19 or H19/SNORA7A-transduced LFS osteoblasts were grown for 1 month and then assayed. Positive colonies are considered those larger than 50μm diameter. Data are represented as mean ± SEM; $n = 6$ biological replicates; statistical significance is determined using two-tailed Student's *t*-test; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. **(D,E)** Genes positively ($n = 780$) and negatively ($n = 945$) correlated with SNORA7A expression in TCGA-SARC are identified. Pathway analysis by Wikipathway **(A)** and GO_BP **(B)** is performed on these gene sets using Enrichr to identify pathways significantly enriched or depleted ($p \leq 0.05$) in association with SNORA7A expression. **(F)** Scatterplots of G6PD, RPS9, CDK4, CDKN1B, ADCY9, and ROCK1 mRNA expression correlation with SNORA7A expression. **(G)** Genes whose abnormal splicing (including exon skip, mutually exclusive splicing, or intron retention) by PSI was positively ($n = 91$, upper panel) or negatively ($n = 182$, lower panel) correlated with SNORA7A expression in sarcomas are identified by TRANSFAC and JASPAR PWWMs in Enrichr. PSI: percent spliced in index. **(H)** High SNORA7A expression is associated with poor cancer survival. Five-year overall survival is analyzed according to SNORA7A expression from TCGA in various cancers. SARC: sarcoma. BRCA: breast invasive carcinoma. ESCA: esophageal carcinoma. KIRC: kidney renal clear cell carcinoma.

potential for snoRNAs to regulate pre-mRNA splicing (Falaleeva et al., 2016; Liang et al., 2019), we investigated the effects of SNORA7A levels toward potentially pathological mRNA splicing events including skipped exons, mutually exclusive alternative splicing, and intron retention. The percent spliced in index (PSI) was calculated for all identified genes in the TCGA-SARC dataset to identify genes whose alternative splicing was correlated with SNORA7A expression. These genes were then mapped to transcription factors using TRANSFAC and JASPAR analysis. Genes whose abnormal splicing was positively correlated to SNORA7A expression were mainly downstream targets of transcription factors such as the osteoblastic lineage regulator RUNX2 (Komori, 2019) and the tumor suppressors ZBTB7A (Liu et al., 2014) and ZFHX3 (Hu et al., 2019). Moreover, genes whose abnormal splicing was negatively correlated to SNORA7A expression were downstream targets of transcription factors such as the tumor angiogenesis regulator HIF1A (Koukourakis et al., 2002), the Hippo pathway transcription factor TEAD4 (Lin et al., 2017a; Shi et al., 2017) and the tumorigenic transcription factor NF-KB1 (Concetti and Wilson, 2018) (**Figure 2G**). These results imply that oncogenic effects of SNORA7A occur through regulation of RNA splicing of tumor suppressor genes and oncogenes.

We next analyzed the expression of SNORA7A in multiple cancers using TCGA datasets and correlated its expression with patient survival data. The five-year overall survival analyses showed that high SNORA7A levels were significantly associated with poorer survival in multiple cancer types including sarcoma (SARC, Log-rank $p = 0.064$, coxph $= 0.017$), breast cancer (BRCA, Log-rank $p = 0.024$, coxph $= 0.026$), esophageal carcinoma (ESCA, Log-rank $p = 0.01$, coxph $= 0.0001$), and kidney carcinoma (KIRC, Log-rank $p \leq 0.0001$, coxph $= 0.0001$) (**Figure 2H**). Taken together, these results suggest a potential role of SNORA7A in promoting cancer progression and emphasize that H19 tumor suppressor functions may be mediated through the repression of SNORA7A expression.

H19 Potentially Interacts With DNA Damage/Repair Response Protein Complexes

Previous studies revealed that H19 is capable of interacting with EZH2 (Luo et al., 2013) to promote WNT/β-Catenin activation and subsequent downregulation of E-cadherin and that H19 also associates with KSRP (Giovarelli et al., 2014) to control mRNA decay. However, the protein complexes tethered by H19 in osteosarcoma remain largely unknown. To elucidate H19-associated protein complexes and gain insights into the potential biological functions controlled by the H19 interactome, we ligated H19 with an optimized 4×streptavidin-binding RNA aptamer (S1m) and expressed H19-S1m in U2OS cells. H19-bound protein complexes could then be pulled down with streptavidin beads and identified by mass spectrometry (**Supplementary Table 2**). Compared with the S1m vector control, S1m-H19 pull-down identified 15 putatively interacting cellular proteins: TUBA1A, KHSRP, ALDH18A1, HIST1H1A, GANAB, TARS2, LRRFIP1, GPANK1, ZNF326, BLM, ABCC12, DMD, CLCN5, RBBP8, and OR8U1 (**Figure 3A**, left panel).

We applied the Biogrid human database to identify proteins previously demonstrated to bind to at least two of these 15 potential H19-binding proteins. This bioinformatics approach yielded 47 proteins strongly suspected to be connected to the 15 H19-associated proteins (**Figure 3A**, right panel). Wikipathway and GO_BP analyses revealed that these putative H19-associated protein complexes mainly function in DNA damage response, DNA repair, the ATM/ATR signaling pathway, the WNT signaling pathway, and the cell cycle (**Figures 3B,C**), emphasizing the potential role of H19 in regulating DNA damage response and repair. Particularly, BLM, an ATP-dependent DNA helicase, participates in DNA replication and DNA repair by modulating DNA double-strand break resection (Gravel et al., 2008). Mutations in the BLM gene are associated with Bloom syndrome, which carries a greatly increased risk of cancers including squamous cell carcinoma, leukemia, lymphoma, and gastrointestinal cancer (Lin et al., 2017b). Another H19-interacting protein is RBBP8 (also known as CtIP), an endonuclease that cooperates with the MRE11-RAD50-NBN (MRN) complex in DNA-end resection, the first step of homologous recombination (HR)-mediated double-strand break repair. RBBP8 mutations are commonly detected in various human cancer cell lines (You and Bailis, 2010). Inactivation of one *RBBP8* allele predisposed mice to multiple types of cancers (e.g., lymphoma) suggesting that RBBP8 functions as a tumor suppressor (Chen et al., 2005). Based on these findings, we speculate that H19 tethers DNA repair

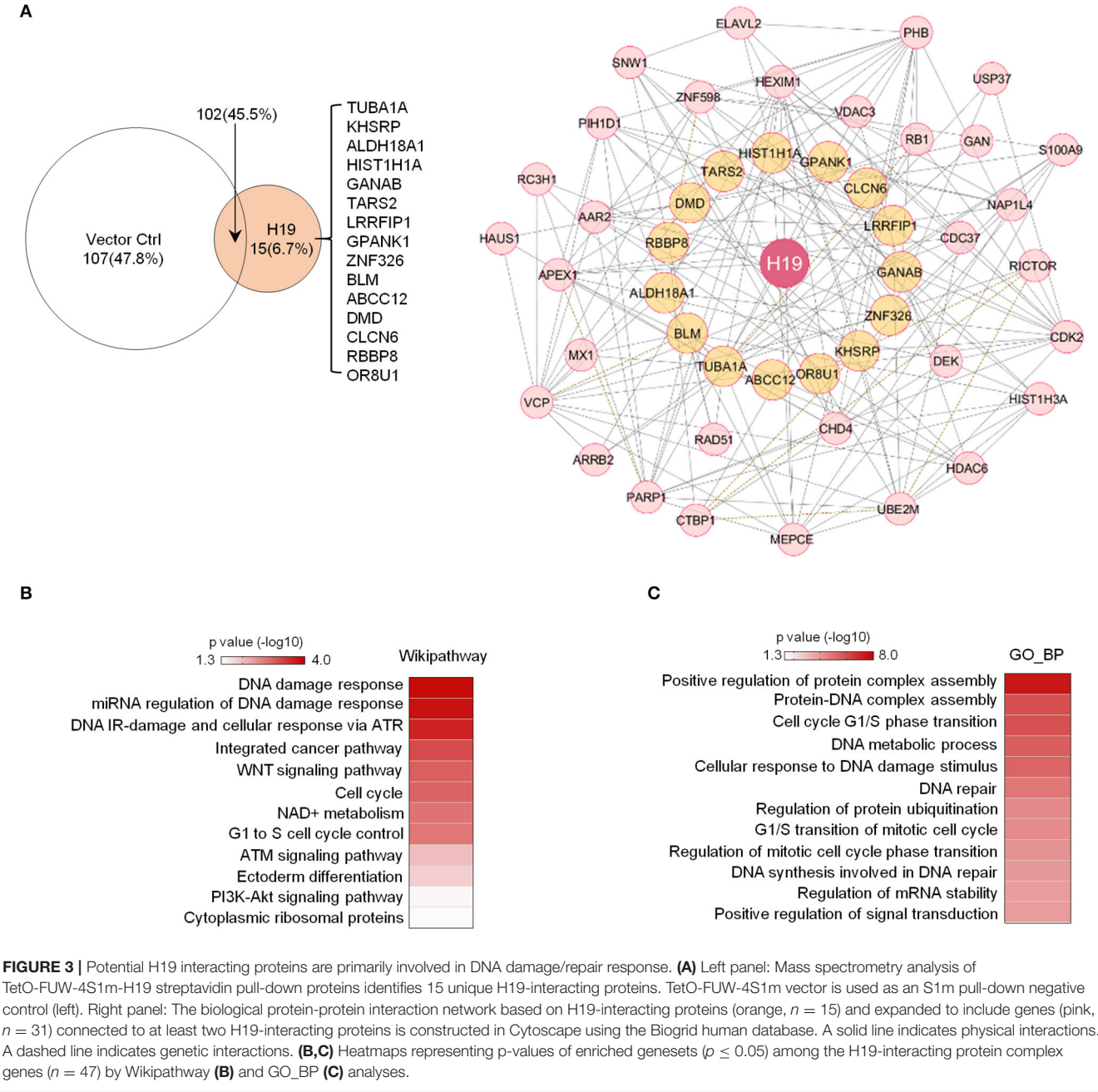


FIGURE 3 | Potential H19 interacting proteins are primarily involved in DNA damage/repair response. **(A)** Left panel: Mass spectrometry analysis of TetO-FUW-4S1m-H19 streptavidin pull-down proteins identifies 15 unique H19-interacting proteins. TetO-FUW-4S1m vector is used as an S1m pull-down negative control (left). Right panel: The biological protein-protein interaction network based on H19-interacting proteins (orange, $n = 15$) and expanded to include genes (pink, $n = 31$) connected to at least two H19-interacting proteins is constructed in Cytoscape using the Biogrid human database. A solid line indicates physical interactions. A dashed line indicates genetic interactions. **(B,C)** Heatmaps representing p-values of enriched genesets ($p \leq 0.05$) among the H19-interacting protein complex genes ($n = 47$) by Wikipathway **(B)** and GO_BP **(C)** analyses.

proteins to other functional factors and helps maintain genomic integrity by suppressing oncogenic events such as DNA double-stranded breaks.

DISCUSSION

The essential biological functions of non-coding RNAs (e.g., miRNAs, snoRNAs, and lncRNAs) and their pathologic roles in tumorigenesis are becoming increasingly appreciated. In our previous work, we reported that H19, a lncRNA, is downregulated by mutant p53s (mutp53s) in LFS iPSC-derived osteoblasts and modulates osteoblastic differentiation and oncogenic repression through the imprinted gene network (IGN) (Gabory et al., 2009; Lee et al., 2015). While H19 has been demonstrated in various contexts to act as either a tumor suppressor or an oncogene, our current work further demonstrates the important regulatory role of H19 in osteosarcoma initiation and progression.

We discovered considerable H19-mediated regulation of snoRNAs. We confirmed that SNORA7A, a snoRNA known to promote MSC proliferation and self-renewal (Zhang et al.,

2017), is suppressed by H19 and that restoration of SNORA7A expression impaired H19-mediated osteogenesis and tumor suppression. These findings suggest that SNORA7A functions as an onco-snoRNA in LFS-associated osteosarcoma. SNORNA7 expression in TCGA datasets is associated with numerous oncogenic pathways (ribosome biogenesis, pentose phosphate pathway, glycolysis, and HIF signaling) and RNA splicing events, suggesting a link between snoRNA expression and oncogenic events in clinical tumors.

p53 mutations have been widely linked to increased tumor oncogenesis through both gain-of-function interactions, for example by promoting HIF1A-regulated angiogenic genes (Amelio et al., 2018), and loss-of-function mutations, for example by abrogating wild-type p53 repression of FBL expression, culminating in increased ribosome biogenesis (Marcel et al., 2013). Upregulation of SNORA7A by mutp53-mediated H19 suppression provides yet another mechanism for mutp53 downstream pathways to cooperate in promoting bone malignancies.

In addition, mass spectrometry analysis of potential H19-interacting proteins revealed multiple DNA damage response and repair molecules directly associated with H19. H19 was previously proposed to regulate DNA damage response (Zheng et al., 2016; Zhu et al., 2017; Ma et al., 2018; Cheng et al., 2019), but the underlying mechanisms have yet to be defined. Our identification of potential interactions between H19 and BLM as well as RBBP8, both of which are recognized to regulate DNA damage response and repair, may provide another angle to elucidate these mechanisms. Importantly, gene mutations in both BLM and RBBP8 are associated with increased risks of tumor formation, including osteosarcoma. The evidence of physical interactions between H19 and DNA damage response genes leads us to speculate that H19 indirectly maintains genomic integrity, explaining its tumor-suppressor activity.

Finally, H19 expression was significantly increased in cancer cells treated with chemotherapeutic drugs (e.g., doxorubicin and cisplatin) (Zheng et al., 2016; Zhu et al., 2017). We speculate that upregulation of H19 following DNA damage events occurs in order to accelerate the repair process and prevent oncogenic effects of DNA damage.

In conclusion, our study indicated that lncRNA H19 plays a vital regulatory role in inhibiting osteosarcomagenesis and provides mechanistic insights for improving our understanding of H19-mediated tumor suppression in LFS patient-associated osteosarcomas.

DATA AVAILABILITY STATEMENT

The transcriptome data has been deposited into the sequencing read archive (SRA; PRJNA673185).

AUTHOR CONTRIBUTIONS

AX, M-FH, DZ, BC, DW, C-CL, and RZ conducted all experiments. AX, M-FH, DZ, DW, C-CL, IL, RZ, and D-FL conceived and designed the study and interpreted results. AX, M-FH, DZ, JG, DB, RZ, and D-FL wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

DZ was supported by DoD Horizon Award (W81XWH-20-1-0389). RZ was supported by the CPRIT UTHealth Cancer Genomic Core pilot grant (CGC-FY20-1). D-FL was a CPRIT Scholar in Cancer Research and is supported by CPRIT Award RR160019, the Rolanette and Berdon Lawrence Research Award, the Pablove Foundation Childhood Cancer Research Seed Grant Award, and CPRIT UTHealth Cancer Genomic Core pilot grant (CGC-FY20-1). We thank the Cancer Prevention and Research Institute of Texas (CPRIT RP180734) for supporting this work.

ACKNOWLEDGMENTS

In memory of our mentor and friend IL (1953-2017), who inspired us with his lifelong passion to understand stem cell biology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.611823/full#supplementary-material

REFERENCES

Amelio, I., Mancini, M., Petrova, V., Cairns, R. A., Vikhreva, P., Nicolai, S., et al. (2018). p53 mutants cooperate with HIF-1 in transcriptional regulation of extracellular matrix components to promote tumor progression. *Proc. Natl. Acad. Sci. U.S.A* 115, E10869–E10878. doi: 10.1073/pnas.1808314115

Bachellerie, J. P., Cavaille, J., and Huttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* 84, 775–790. doi: 10.1016/s0300-9084(02)01402-5

Carvajal-Vergara, X., Sevilla, A., D'Souza, S. L., Ang, Y. S., Schaniel, C., Lee, D. F., et al. (2010). Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature* 465, 808–812. doi: 10.1038/nature09005

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* 14:128. doi: 10.1186/1471-2105-14-128

Chen, P. L., Liu, F., Cai, S., Lin, X., Li, A., Chen, Y., et al. (2005). Inactivation of CtIP leads to early embryonic lethality mediated by G1 restraint and to tumorigenesis by haploid insufficiency. *Mol. Cell Biol.* 25, 3535–3542. doi: 10.1128/MCB.25.9.3535-3542.2005

Cheng, T., Xu, M., Qin, B., Wu, J., Tu, Y., Kang, L., et al. (2019). lncRNA H19 contributes to oxidative damage repair in the early age-related cataract by regulating miR-29a/TDG axis. *J. Cell Mol. Med.* 23, 6131–6139. doi: 10.1111/jcmm.14489

Concetti, J., and Wilson, C. L. (2018). NFKB1 and cancer: friend or foe? *Cells* 7:9. doi: 10.3390/cells7090133

David, C. J., Chen, M., Assanah, M., Canoll, P., and Manley, J. L. (2010). HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* 463, 364–368. doi: 10.1038/nature08697

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, E1625–1634. doi: 10.1073/pnas.1519292113

Gabory, A., Ripoche, M. A., Le Digarcher, A., Watrin, F., Ziyyat, A., Forne, T., et al. (2009). H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development* 136, 3413–3421. doi: 10.1242/dev.036061

Gazy, I., Zeevi, D. A., Renbaum, P., Zeligson, S., Eini, L., Bashari, D., et al. (2015). TODRA, a lncRNA at the RAD51 locus, is oppositely regulated to RAD51, and enhances RAD51-dependent DSB (double strand break) repair. *PLoS ONE* 10:e0134120. doi: 10.1371/journal.pone.0134120

Gingold, J., Zhou, R., Lemischka, I. R., and Lee, D. F. (2016). Modeling cancer with pluripotent stem cells. *Trends Cancer* 2, 485–494. doi: 10.1016/j.trecan.2016.07.007

Giovarelli, M., Bucci, G., Ramos, A., Bordo, D., Wilusz, C. J., Chen, C. Y., et al. (2014). H19 long noncoding RNA controls the mRNA decay promoting function of KSRP. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5023–5028. doi: 10.1073/pnas.1415098111

Gong, J., Li, Y., Liu, C. J., Xiang, Y., Li, C., Ye, Y., et al. (2017). A pan-cancer analysis of the expression and clinical relevance of small nucleolar RNAs in human cancer. *Cell Rep.* 21, 1968–1981. doi: 10.1016/j.celrep.2017.10.070

Gravel, S., Chapman, J. R., Magill, C., and Jackson, S. P. (2008). DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes Dev.* 22, 2767–2772. doi: 10.1101/gad.503108

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

Hsu, T. Y., Simon, L. M., Neill, N. J., Marcotte, R., Sayad, A., Bland, C. S., et al. (2015). The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* 525, 384–388. doi: 10.1038/nature14985

Hu, Q., Zhang, B., Chen, R., Fu, C., Fu, A. J., et al. (2019). ZFHX3 is indispensable for ERbeta to inhibit cell proliferation via MYC downregulation in prostate cancer cells. *Oncogenesis* 8:28. doi: 10.1038/s41389-019-0138-y

Hu, W. L., Jin, L., Xu, A., Wang, Y. F., Thorne, R. F., Zhang, X. D., et al. (2018). GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. *Nat. Cell Biol.* 20, 492–502. doi: 10.1038/s41556-018-0066-7

Itzhaki, I., Maizels, L., Huber, I., Zwi-Dantsis, L., Caspi, O., Winterstern, A., et al. (2011). Modelling the long QT syndrome with induced pluripotent stem cells. *Nature* 471, 225–229. doi: 10.1038/nature09747

Kallen, A. N., Zhou, X. B., Xu, J., Qiao, C., Ma, J., Yan, L., et al. (2013). The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol. Cell* 52, 101–112. doi: 10.1016/j.molcel.2013.08.027

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., et al. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat. Cell Biol.* 14, 659–665. doi: 10.1038/ncb2521

Kim, H., Yoo, S., Zhou, R., Xu, A., Bernitz, J. M., Yuan, Y., et al. (2018). Oncogenic role of SFRP2 in p53-mutant osteosarcoma development via autocrine and paracrine mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 115, E11128–E11137. doi: 10.1073/pnas.1814044115

Koh, C. M., Bezzi, M., Low, D. H., Ang, W. X., Teo, S. X., Gay, F. P., et al. (2015). MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature* 523, 96–100. doi: 10.1038/nature14351

Komori, T. (2019). Regulation of proliferation, differentiation and functions of osteoblasts by Runx2. *Int. J. Mol. Sci.* 20:7. doi: 10.3390/ijms20071694

Koukourakis, M. I., Giatromanolaki, A., Sivridis, E., Simopoulos, C., Turley, H., Talks, K., et al. (2002). Hypoxia-inducible factor (HIF1A and HIF2A), angiogenesis, and chemoradiotherapy outcome of squamous cell head-and-neck cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 53, 1192–1202. doi: 10.1016/s0360-3016(02)02848-1

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–97. doi: 10.1093/nar/gkw377

Lee, D. F., Kuo, H. P., Chen, C. T., Hsu, J. M., Chou, C. K., Wei, Y., et al. (2007). IKK beta suppression of TSC1 links inflammation and tumor angiogenesis via the mTOR pathway. *Cell* 130, 440–455. doi: 10.1016/j.cell.2007.05.058

Lee, D. F., Kuo, H. P., Liu, M., Chou, C. K., Xia, W., Du, Y., et al. (2009a). KEAP1 E3 ligase-mediated downregulation of NF-kappaB signaling by targeting IKKbeta. *Mol. Cell* 36, 131–140. doi: 10.1016/j.molcel.2009.07.025

Lee, D. F., Su, J., Ang, Y. S., Carvajal-Vergara, X., Mulero-Navarro, S., Pereira, C. F., et al. (2012a). Regulation of embryonic and induced pluripotency by aurora kinase-p53 signaling. *Cell Stem Cell* 11, 179–194. doi: 10.1016/j.stem.2012.05.020

Lee, D. F., Su, J., Kim, H. S., Chang, B., Papatsenko, D., Zhao, R., et al. (2015). Modeling familial cancer with induced pluripotent stem cells. *Cell* 161, 240–254. doi: 10.1016/j.cell.2015.02.045

Lee, D. F., Su, J., Sevilla, A., Gingold, J., Schaniel, C., and Lemischka, I. R. (2012b). Combining competition assays with genetic complementation strategies to dissect mouse embryonic stem cell self-renewal and pluripotency. *Nat. Protoc.* 7, 729–748. doi: 10.1038/nprot.2012.018

Lee, G., Papapetrou, E. P., Kim, H., Chambers, S. M., Tomishima, M. J., Fasano, C. A., et al. (2009b). Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* 461, 402–406. doi: 10.1038/nature08320

Leppek, K., and Stoecklin, G. (2014). An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* 42:e13. doi: 10.1093/nar/gkt956

Li, A., Mallik, S., Luo, H., Jia, P., Lee, D. F., and Zhao, Z. (2020). H19, a long non-coding, RNA, mediates transcription factors and target genes through interference of MicroRNAs in pan-cancer. *Mol. Ther. Nucleic Acids* 21, 180–191. doi: 10.1016/j.omtn.2020.05.028

Lian, Q., Lye, E., Suan Yeo, K., Khia Way Tan, E., Salto-Tellez, M., Liu, T. M., et al. (2007). Derivation of clinically compliant MSCs from CD105+, CD24- differentiated human ESCs. *Stem Cells* 25, 425–436. doi: 10.1634/stemcells.2006-0420

Liang, J., Wen, J., Huang, Z., Chen, X. P., Zhang, B. X., and Chu, L. (2019). Small nucleolar RNAs: insight into their function in cancer. *Front. Oncol.* 9:587. doi: 10.3389/fonc.2019.00587

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

Lin, K. C., Park, H. W., and Guan, K. L. (2017a). Regulation of the hippo pathway transcription factor TEAD. *Trends Biochem. Sci.* 42, 862–872. doi: 10.1016/j.tibs.2017.09.003

Lin, Y. H., Jewell, B. E., Gingold, J., Lu, L., Zhao, R., Wang, L. L., et al. (2017b). Osteosarcoma: molecular pathogenesis and iPSC modeling. *Trends Mol. Med.* 23, 737–755. doi: 10.1016/j.molmed.2017.06.004

Liu, M., Tu, J., Gingold, J. A., Kong, C. S. L., and Lee, D. F. (2018). Cancer in a dish: progress using stem cells as a platform for cancer research. *Am. J. Cancer Res.* 8, 944–954.

Liu, X. S., Haines, J. E., Mehanna, E. K., Genet, M. D., Ben-Sahra, I., Asara, J. M., et al. (2014). ZBTB7A acts as a tumor suppressor through the transcriptional repression of glycolysis. *Genes Dev.* 28, 1917–1928. doi: 10.1101/gad.245910.114

Luo, M., Li, Z., Wang, W., Zeng, Y., Liu, Z., and Qiu, J. (2013). Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. *Cancer Lett.* 333, 213–221. doi: 10.1016/j.canlet.2013.01.033

Lustig-Yariv, O., Schulze, E., Komitowski, D., Erdmann, V., Schneider, T., de Groot, N., et al. (1997). The expression of the imprinted genes H19 and IGF-2 in choriocarcinoma cell lines. Is H19 a tumor suppressor gene? *Oncogene* 15, 169–177. doi: 10.1038/sj.onc.1201175

Ma, H., Yuan, L., Li, W., Xu, K., and Yang, L. (2018). The LncRNA H19/miR-193a-3p axis modifies the radio-resistance and chemotherapeutic tolerance of hepatocellular carcinoma cells by targeting PSEN1. *J. Cell Biochem.* 119, 8325–8335. doi: 10.1002/jcb.26883

Marcel, V., Ghayad, S. E., Belin, S., Therizols, G., Morel, A. P., Solano-Gonzalez, E., et al. (2013). p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. *Cancer Cell* 24, 318–330. doi: 10.1016/j.ccr.2013.08.013

Mei, Y., Liao, J., Shen, J., Yu, L., Liu, B., Liu, L., et al. (2012). Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene* 31, 2794–2804. doi: 10.1038/onc.2011.449

Mulero-Navarro, S., Sevilla, A., Roman, A. C., Lee, D. F., D'Souza, S. L., Pardo, S., et al. (2015). Myeloid dysregulation in a human induced pluripotent stem cell

model of PTPN11-associated juvenile myelomonocytic leukemia. *Cell Rep.* 13, 504–515. doi: 10.1016/j.celrep.2015.09.019

Sanchez Calle, A., Kawamura, Y., Yamamoto, Y., Takeshita, F., and Ochiya, T. (2018). Emerging roles of long non-coding RNA in cancer. *Cancer Sci.* 109, 2093–2100. doi: 10.1111/cas.13642

Shi, Z., He, F., Chen, M., Hua, L., Wang, W., Jiao, S., et al. (2017). DNA-binding mechanism of the Hippo pathway transcription factor TEAD4. *Oncogene* 36, 4362–4369. doi: 10.1038/onc.2017.24

Ulitsky, I., and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46. doi: 10.1016/j.cell.2013.06.020

Wan, G., Hu, X., Liu, Y., Han, C., Sood, A. K., Calin, G. A., et al. (2013). A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. *EMBO J* 32, 2833–2847. doi: 10.1038/emboj.2013.221

Wu, L., Chang, L., Wang, H., Ma, W., Peng, Q., and Yuan, Y. (2018). Clinical significance of C/D box small nucleolar RNA U76 as an oncogene and a prognostic biomarker in hepatocellular carcinoma. *Clin. Res. Hepatol. Gastr.* 42, 82–91. doi: 10.1016/j.clinre.2017.04.018

Wu, M., Deng, L., Zhu, G., and Li, Y. P. (2010). G Protein and its signaling pathway in bone development and disease. *Front. Biosci.* 15:957–985. doi: 10.2741/3656

Yagi, T., Ito, D., Okada, Y., Akamatsu, W., Nihei, Y., Yoshizaki, T., et al. (2011). Modeling familial Alzheimer's disease with induced pluripotent stem cells. *Hum. Mol. Genet.* 20, 4530–4539. doi: 10.1093/hmg/ddr394

Yoshimizu, T., Miroglio, A., Ripoche, M. A., Gabory, A., Vernucci, M., Riccio, A., et al. (2008). The H19 locus acts *in vivo* as a tumor suppressor. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12417–12422. doi: 10.1073/pnas.0801540105

You, Z., and Bailis, J. M. (2010). DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints. *Trends Cell Biol.* 20, 402–409. doi: 10.1016/j.tcb.2010.04.002

Zhang, Y., Xu, C., Gu, D., Wu, M., Yan, B., Xu, Z., et al. (2017). H/ACA box small nucleolar RNA 7A promotes the self-renewal of human umbilical cord mesenchymal stem cells. *Stem Cells* 35, 222–235. doi: 10.1002/stem.2490

Zheng, Z. G., Xu, H., Suo, S. S., Xu, X. L., Ni, M. W., Gu, L. H., et al. (2016). The essential role of H19 contributing to cisplatin resistance by regulating glutathione metabolism in high-grade serous ovarian cancer. *Sci. Rep.* 6:26093. doi: 10.1038/srep26093

Zhou, R., Xu, A., Gingold, J., Strong, L. C., Zhao, R., and Lee, D. F. (2017). Li-fraumeni syndrome disease model: a platform to develop precision cancer therapy targeting oncogenic p53. *Trends Pharmacol. Sci.* 38, 908–927. doi: 10.1016/j.tips.2017.07.004

Zhou, R., Xu, A., Tu, J., Liu, M., Gingold, J. A., Zhao, R., et al. (2018). Modeling osteosarcoma using li-fraumeni syndrome patient-derived induced pluripotent stem cells. *J. Vis. Exp.* 13:57664. doi: 10.3791/57664

Zhu, D., Kong, C. S. L., Gingold, J. A., Zhao, R., and Lee, D. F. (2018). Induced pluripotent stem cells and induced pluripotent cancer cells in cancer disease modeling. *Adv. Exp. Med. Biol.* 1119, 169–183. doi: 10.1007/5584_2018_257

Zhu, Q. N., Wang, G., Guo, Y., Peng, Y., Zhang, R., Deng, J. L., et al. (2017). LncRNA H19 is a major mediator of doxorubicin chemoresistance in breast cancer cells through a cullin4A-MDR1 pathway. *Oncotarget* 8, 91990–92003. doi: 10.18632/oncotarget.21121

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 February 2021
doi: 10.3389/fgene.2021.598575

Effects of Circular RNA of Chicken Growth Hormone Receptor Gene on Cell Proliferation

*Haidong Xu[1], Qiying Leng[1], Jiahui Zheng[1], Patricia Adu-Asiamah[1], Shudai Lin[2], Ting Li[1], Zhang Wang[1], Lilong An[1], Zhuihui Zhao[1] and Li Zhang[1]**

[1] College of Coastal Agricultural Sciences, Guangdong Ocean University, Zhanjiang, China, [2] Guangdong Provincial Key Laboratory of Agro-Animal Genomics and Molecular Breeding, Key Laboratory of Chicken Genetics, Breeding and Reproduction, Ministry of Agriculture, South China Agricultural University, Guangzhou, China

Animal growth and development are regulated by neural and endocrine growth axes, in which cell proliferation plays key roles. Recently, many research showed that circular RNAs were involved in hepatocyte and myoblast proliferation. Previously, we identified a circular RNA derived from the chicken *GHR* gene, named circGHR. However, the function of circGHR is unclear. The objective of this study was to investigate circGHR expression pattern and its roles in cell proliferation. Results indicated that circGHR was a closed-loop structure molecule, and it was richer in the nucleus of hepatocytes and myoblast. Real-time PCR showed that circGHR was increased from E13 to the 7th week in the liver but decreased in the thigh and breast muscle. The CCK-8 assay displayed that circGHR promoted cell proliferation. Simultaneously, the biomarker genes *PCNA*, *CCND1*, and *CDK2* and the linear transcripts *GHR* and *GHBP* were upregulated when circGHR was overexpressed. Altogether, these data exhibited that circGHR could promote cell proliferation possibly by regulating *GHR* mRNA and *GHBP* expression.

Keywords: chicken, growth hormone receptor, circular RNA, cell proliferation, linear transcript

OPEN ACCESS

Edited by:
Wenbo Li,
University of Texas Health Science Center at Houston, United States

Reviewed by:
Marton Doleschall, Independent Researcher, Budapest, Hungary
Yun Cheng, The University of Hong Kong, Hong Kong

****Correspondence:***
Li Zhang
zhangli761101@163.com

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *25 August 2020*
Accepted: *11 January 2021*
Published: *11 February 2021*

Citation:
Xu H, Leng Q, Zheng J, Adu-Asiamah P, Lin S, Li T, Wang Z, An L, Zhao Z and Zhang L (2021) Effects of Circular RNA of Chicken Growth Hormone Receptor Gene on Cell Proliferation. Front. Genet. 12:598575. doi: 10.3389/fgene.2021.598575

INTRODUCTION

Animal growth and development are regulated by the neural and endocrine growth axes, including the hypothalamus, pituitary, growth hormone (GH), and target organs (Picard et al., 2002). GH is transported to the target organ by growth hormone binding protein (GHBP) in body fluid (Picard et al., 2002). Then, it can combine with the growth hormone receptor (GHR) to activate the intracellular insulin-like growth factor (IGF) pathway and involve in cell proliferation and differentiation (Sperling, 2016).

Cell proliferation is the starting point for tissue and organ development. Hepatocyte proliferation is the primary biological process in liver development during the early embryonic stage (Hamada et al., 1995). Also, muscle development begins with myoblast proliferation, which depends on the increased number of muscle fibers (Picard et al., 2002; Welle et al., 2007). Once the growth phase is completed, the fiber numbers do not increase, but the fiber volume increases (Buckingham et al., 2003). Recently, many researchers identified circular RNAs (circRNAs) regulating hepatocyte and myoblast proliferation, such as circMark14 (Li L. et al., 2017) and circFUT10 (Li H. et al., 2017).

CircRNAs, a class of endogenous molecules with covalent-formed closed-loop structure and abundant in eukaryotic organisms (Conn et al., 2015), functioned in the skeletal muscle development of animals by regulating gene expression at multiple levels. Most researchers exhibited

that circRNAs functioned by multiple mechanisms, such as miRNA sponges, translation, and regulating their parental genes (Patricia et al., 2020). Specifically, bovine circFUT10 regulated myoblast differentiation and cell survival by directly binding to miR-133a and inhibiting miR-133a activity (Li H. et al., 2017), while human and mouse circ-ZNF609 consisted of direct reading scaffolds, which can translate into protein and play roles in skeletal muscle disorders (Legnini et al., 2017). Another interesting circRNA was circMbl. Findings confirmed that muscleblind (MBL/MBNL1) could bind to its pre-mRNA and encouraged circMbl production, which may be associated with myotonic dystrophy initiation and progression (Ashwal-Fluss et al., 2014).

Previously, we identified a circRNA divided from the chicken *GHR* gene, named circGHR (Zhang et al., 2019). Yet, the function and regulation mechanisms of circGHR are uncertain. This study mainly analyzed its expression patterns and its effects on cell proliferation (the summary technical pipeline is in **Supplementary Figure 1**). We found that chicken circGHR could promote cell proliferation, possibly by regulating *GHR* mRNA and *GHBP* expression.

MATERIALS AND METHODS

Ethics Statement

Animals involved in this research were humanely sacrificed as necessary to ameliorate suffering. The study was approved by the Animal Care Committee of Guangdong Ocean University (Zhangjiang, China).

Experimental Animal and Sample Collection

Body tissues including breast muscle, thigh muscle, liver, heart, and small intestine were collected from four male Huaixiang chickens (a local Chinese breed) at every age from embryonic 13 (E13) to 7 weeks (7W) (E13, E16, E19, 1D, 1W, 2W, 3W, 4W, 5W, 6W, and 7W). Besides, the tissue samples were snap-frozen in liquid nitrogen and stored in a freezer at −80°C until analyzed. All the fertile chicken eggs were obtained from the HuaiXiang Chicken Breeding Farm (Yingfu Company, Xinyi City, Guangdong Province). They were incubated in an automatic incubator (Baihui, Shandong Province) at 37.8°C, with 50–60% humidity. Chicken myoblast and hepatocytes were individually isolated at E11 and E16 during incubation. The broilers were fed with the normal basal diet, including 11.51% metabolic energy, 15.23% crude protein, 3.10% calcium, 0.45% available phosphorus, and a few non-essential amino acids.

Identification of Chicken CircGHR Molecular Characteristics

In our previous study, we characterized circGHR by RNase R treatment and sequencing (Zhang et al., 2019). Here, we compared the reverse transcriptional efficiency with two kinds of primers. Firstly, the cDNA was synthesized by PrimeScript RT Reagent Kit (TaKaRa, Kyoto, Japan) with random hexamer or oligo$(dT)_{18}$ primers, respectively. Then, qRT-PCR was conducted to compare the relative level between the two groups. The *GAPDH* gene was used as an internal control. Additionally, the amplification products were detected using agarose gel electrophoresis and Sanger sequencing. The primers used in this experiment are listed in **Table 1**.

Cell Culture

Two types of chicken primary cells (myoblast and hepatocyte) and two types of chicken cell lines (LMH and DF-1 cell lines) were cultured for analyzing the roles of circGHR.

Myoblast was isolated from chicken thigh muscles at E11 according to previously described methods (Zhang et al., 2017). Briefly, a differential attachment was performed three times every 40 min after the thigh muscles were cut into pieces. During the third time, the adherent cells were collected and directly grown on a plastic plate. The cells were cultured in growth medium (GM) with DMEM-F12 (Gibco, New York, United States), 15% fetal bovine serum (Gibco, New York, United States), and 1% penicillin/streptomycin (Gibco, New York, United States) in 5% CO_2 incubator at 37°C.

Hepatocytes were isolated from the chicken liver at E16 according to previously described methods (Wang et al., 2019). Briefly, the liver was washed with phosphate buffer saline (PBS) three times to remove the impurities. It was then cut into pieces and digested using collagenase type V at 34°C for 7 min. The cellular suspension was filtered through 200 and 500 mesh sieves, respectively, and washed with PBS three times at 1,000 rpm for 5 min. The obtained cells were isolated by non-continuous density Percoll gradient centrifugation (Sigma, MO, United States) and diluted with Williams' E uncomplete medium (Sigma, no fetal bovine serum) to the concentration of 1×10^6 cells/ml, followed by plating into 12-well dishes. After incubation in 5% CO_2 incubator at 37°C for 4 h, the medium was changed with Williams' E complete medium containing 10% fetal bovine serum and 1% penicillin/streptomycin solution (Gibco, New York, United States) until treatment.

Chicken LMH and DF-1 cell lines, provided by South China Agricultural University, were cultured with DMEM containing 10% fetal bovine serum and 1% penicillin/streptomycin solution (Gibco, New York, United States) in 5% CO_2 incubator at 37°C.

Nucleus and Cytoplasm Separation

To analyze the location of circGHR in the cells, the nucleus and cytoplasm RNA were extracted from three wells of chicken myoblast or hepatocytes by PARIS^{TM} Kit Protein and RNA Isolation System (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's instructions. The PARIS method was based on differential lysis of plasma and nuclear membranes by non-ionic detergents. Briefly, cells were first separated into nuclear and cytoplasmic fractions, and then RNA was isolated. The cell pellet was suspended in a buffer for RNA purification and then centrifuged twice at 4°C for 5 min. The supernatant was the cytoplasmic fraction, and the pellet was the nucleus fraction. RNAs were extracted from both fractions using TRIzol (Mange, Guangzhou, China). The relative levels of circGHR between the

TABLE 1 | Primers used in this study.

Primer name	Primer sequence (5′ → 3′)	Transcript or Gene ID	Tm/°C	Product/bp	Application
circGHR-F	GTCCCTCAGCTCAACTGC		53.0	115	circGHR quantitative PCR
circGHR-R	ATCTTCGGCATCTGCTGT				
GHR-F	AGTCCGATCAAGACAACGTAC	XM_015272680.1	59.4	128	*GHR* mRNA quantitative PCR
GHR-R	CTAAGAACCAGGGAAACTCG				
GHBP-F	TGATGAAATAGTACTACCTGATCC	DQ138367	56.3	208	*GHBP* quantitative PCR
GHBP-R	TAAATATTTCCTCCATACCTCC				
circGHR-V-F	gg*GGTACC*tgaaatatgctatcttacagGTGCTGGGATGGCTGGAGAAGG		70.0	221	Construct circGHR expression vector
circGHR-V-R	cg*GGATCC*tcaagaaaaaatatattcacCATCACTTGCAGAAAGTGAGTCATT				
GHR-luc-F	*tcttacgcgtgctag**CCCGGG**CAGTTAGGCAAGTAAATGTATATTGGA	408184	60.0	2,998	Construct GHR promoter reporter vector
GHR-luc-R	*acttagatcgcagat**CTCGAG**CTGCTAACCTCCTTCTCTAGGTATGC				
GHBP-luc-F	*tcttacgcgtgctag**CCCGGG**CTTAGATTAAAACCTCTGCGAG	408184	60.0	1,430	Construct GHBP promoter reporter vector
GHBP-luc-R	*acttagatcgcagat**CTCGAG**GGGAAGGAGGGGATGAGGGA				
PCNA-F	CTCTGAGGGCTTCGACACCT	NM_204170.2	58.0	133	*PCNA* quantitative PCR
PCNA-R	ATCCGCATTGTCTTCTGCTCT				
CCND1-F	AACCCACCTTCCATGATCGC	NM_205381.1	57.8	159	*CCND1* quantitative PCR
CCND1-R	CTGTTCTTGGCAGGCTCGTA				
CDK2-F	GTACAAGGCCCGGAACAAGG	NM_001199857.1	58.2	168	*CDK2* quantitative PCR
CDK2-R	TTCTCCGTGTGGATCACGTC				
GAPDH-F	AGGACCAGGTTGTCTCCTGT	NM_204305.1	60.0	153	Internal control in qPCR analysis
GAPDH-R	CCATCAAGTCCACAACACGG				
Sno-U6-F	CTCGCTTCGGCAGCACA	X07425.1	53.0	94	
Sno-U6-R	AACGCTTCACGAATTTGCGT				

The underlined and italic type is the enzyme site of KpnI and BamHI and the lowercase part is the filling sequence and protecting base for circGHR-V primers. The underlined and italic type is the enzyme site of SmaI and XhoI and the lowercase part is the homologous sequences with pGL3-basic for GHR/GHBP-luc primers.

nucleus and cytoplasm were compared by qRT-PCR. Both *sno-U6* and *GAPDH* were used as control.

Bioinformatics Analysis

The transcription start sites and promoter regions of *GHR* and *GHBP* were analyzed by referencing the thesis of Lau and Joanna (2005). The conserved transcription factor binding sites in the vertebrate database were predicted using the JASPAR website tool[1] (Ovcharenko et al., 2005).

CircGHR Plasmid Construction and Cell Transfection

For the circGHR circular overexpression vector (pCD2.1-circGHR), the full length of circGHR was amplified by PCR with forward and reverse primers, including *Kpn*I and *Bam*HI restriction enzyme sites at 5′-ends, respectively (**Table 1**). The PCR was performed with a 20-μl reaction mixture containing 1 μl of cDNA transcribed from chicken liver RNA, 10 μl of TaKaRa Taq Version 2.0 plus dye (TaKaRa, Kyoto, Japan), 1 μl each of the forward and reverse primers (10 μM), and double-distilled water. The PCR reaction procedure was as follows: 94°C 2 min, 40 cycles at 94°C 30 s, 57°C 30 s, and 72°C 30 s, followed by 72°C for 5 min. The PCR products were digested with *Kpn*I and *Bam*HI restriction enzymes. The purified fragment was ligated into pCD2.1-ciR(+) plasmid (with green fluorescent protein tag, Geneseed Company, Guangzhou, China) to obtain the pCD2.1-circGHR recombinant plasmid. pCD2.1-circGHR overexpression experiment was carried out in a 12-well plate by transient transfecting pCD2.1-circGHR into the cell using Lipofectamine 3000 (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's protocol.

Genomic DNA was isolated from chicken liver samples ($n = 3$) using the phenol–chloroform method (Green and Sambrook, 2014) and stored at −20°C. For the construction of *GHR* and *GHBP* promoter-luciferase reporter vector, chicken *GHR* and *GHBP* promoter fragments were amplified with primers GHR/GHBP-Luc (**Table 1**) using genomic DNA. Subsequently, the amplified *GHR* and *GHBP* promoter fragments were forwardly inserted into the promoter-less luciferase reporter vector pGL3-basic (Promega, United States) between *Sma*I

[1]http://jaspar.genereg.net/

and *Xho*I sites and named pGL3-GHR(−2,730/ +226) and pGL3-GHBP(−1,322/ +66), respectively (Lau and Joanna, 2005). DF-1 cells were transiently co-transfected with either luciferase report vector [pGL3-GHR(−2,730/ +226) or pGL3-GHBP(−1,322/ +66)] and pCD2.1-circGHR or pCD2.1-ciR plus pRL-TK Renilla luciferase vector as above.

All primers were synthesized by Sangon Biotech (Shanghai, Chain), and all plasmids were confirmed by DNA sequencing (Sangon, Shanghai, Chain).

Cell Counting Kit-8 Assay

The Cell Counting Kit-8 (CCK-8) assay was used to examine cell proliferation. Briefly, the cells were plated into 96-well culture plates at a density of 1×10^4 cells/well in 100 μl of culture medium per well. In the beginning, the CCK-8 reagent was used to select the uniform well for transfection, and each group had 12 independent replicates. After transfecting pCD2.1-circGHR recombinant plasmid or control pCD2.1-ciR(+) plasmid for 12 h, 10 μl of CCK-8 reagent (Dojindo Laboratories, Kumamoto, Japan) was added to each well and incubated at 37°C for 2 h. Each sample's absorbance was detected at 450 nm using a microplate reader (Thermo Fisher, MA, United States). Cell proliferation was then monitored every 12 h according to the manufacturer's protocol.

Dual-Luciferase Reporter Assay

Briefly, the DF-1 cells (2×10^5 cells/well) were seeded in a 24-well plate and cultured in the medium. After reaching 70–80% confluence, the cells were washed with PBS, and transient transfection was performed using Lipofectamine 3000 (Invitrogen, United States). Dual-luciferase reporter assays were performed 48 h post-transfection using the dual-luciferase reporter assay system (Promega, United States) according to the manufacturer's instructions, in which Firefly luciferase (*Fluc*) activity was normalized to Renilla luciferase (*Rluc*) activity.

RNA Extraction and Quantitative Real-Time PCR

Following the manufacturer's protocol, total RNA was extracted from the cells by TRIzol reagent (Mange, Guangzhou, China). The RNA integration and concentration were detected using electrophoresis at 1.5% agarose gel and Nanodrop 2000 spectrophotometry. cDNA was synthesized with PrimeScript RT Reagent Kit (TaKaRa, Toyoto, Japan). Besides, cDNA was diluted four times with RNase-free water and stored at −20°C. Real-time PCR was conducted in a Bio-Rad CFX96 Real-Time Detection System (Bio-Rad, Hercules, CA, United States) using TransStart Green qPCR SuperMix (Transgen Co., Ltd., Beijing, China). The primers of circGHR, *GHR* mRNA, and *GHBP* and biomarker genes, like *PCNA*, *CCND1*, and *CDK2* (Fukami-Kobayashi and Mitsui, 1999), are listed in **Table 1**. The expression was normalized with *GAPDH*. We performed qRT-PCR with 20 μl mixture containing 1 μl of cDNA, 10 μl of 2× TransStart Green qPCR SuperMix, 0.5 μl each of the forward and reverse primers (10 μM), and double-distilled water. The qRT-PCR reaction procedure is as follows: 95°C for 30 s, 40 cycles at 95°C for 5 s, annealing for 30 s, 72°C for 30 s, followed by 72°C for 5 min. The expressed gene was quantified using the comparative threshold cycle ($2^{-\Delta\Delta Ct}$) methods.

Statistical Analysis

Each experiment was repeated three times and all the data were expressed as the mean ± SE and processed using the statistical software SAS 9.1.3 (SAS Institute Inc., NC, United States). Unpaired Student's *t* test was used for *P* value calculations. GraphPad Prism 8.3 was used for the creation of boxplot figures. A single asterisk (*) was considered significant ($P < 0.05$), whereas double asterisks (**) ($P < 0.01$) were considered extremely significant between groups.

RESULTS

Chicken CircGHR Characteristics

In our previous study, we found that the chicken *GHR* gene may possess a circular transcript (circGHR, GenBank accession number is MW145198) transcribed from its 5′ UTR, exon1, and exon2. The circGHR was resistant to RNase R treatment, confirming its circularized characteristics (Zhang et al., 2019). In this study, we further compared its reverse transcriptional efficiency with different primers. The results showed that random primers had higher efficiency than the oligo$(dT)_{18}$ primers, further demonstrating that circGHR was a circular molecule without poly(A) sequences. Meanwhile, the linear transcript *GHR* mRNA had no significant differences in both groups (**Figure 1A**). In the following experiment, all the RNA was transcribed with oligo$(dT)_{18}$ and random mixture primers. To understand the possible roles of circGHR in hepatocytes and myoblast, we analyzed the location of circGHR, and the results showed that circGHR was richer in the nucleus than in the cytoplasm in both hepatocytes and myoblast (**Figure 1B**).

Expression Patterns of CircGHR in Chicken Growth

To investigate the expression patterns of circGHR, chicken liver, thigh, and breast muscle were collected from the age of E13 to 7W. The qRT-PCR analysis showed that circGHR expressed increased from 3W to 7W in the liver (**Figure 2A**) and decreased from E13 to 7W of age in the thigh and breast muscle (**Figures 2B,C**). Comparing different tissues, we found that a higher level of circGHR expression was in the small intestine, breast, and thigh muscle than in the other tissues at the age of E13 (**Figure 2D**), while in the heart, liver, and small intestine, the level of circGHR expression was higher than in the other tissues at the age of 7W (**Figure 2D**). These results showed that circGHR was always highly expressed in the small intestine at both E13 and 7W, which inspired us to resolve whether circGHR regulates

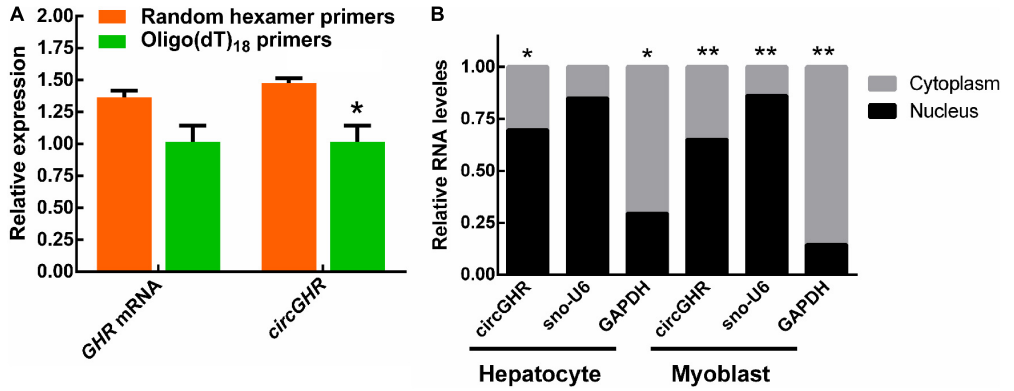


FIGURE 1 | CircGHR characteristic confirmation and its expression pattern in chicken cells. **(A)** The relative expression of *GHR* mRNA and circGHR after specific RT-PCR using oligo(dT)$_{18}$ or random hexamer primers. Fold change was relative to the expression of the random hexamer primers group. **(B)** CircGHR subcellular localization in chicken hepatocytes and myoblast. The RNA relative levels in the nucleus and cytoplasm were calculated by the 2^{-Ct} methods. Fold change was relative to the expression of the nucleus. All data are representative of three independent experiments and are shown as the mean ± SEM. $^{*}P < 0.05$ and $^{**}P < 0.01$.

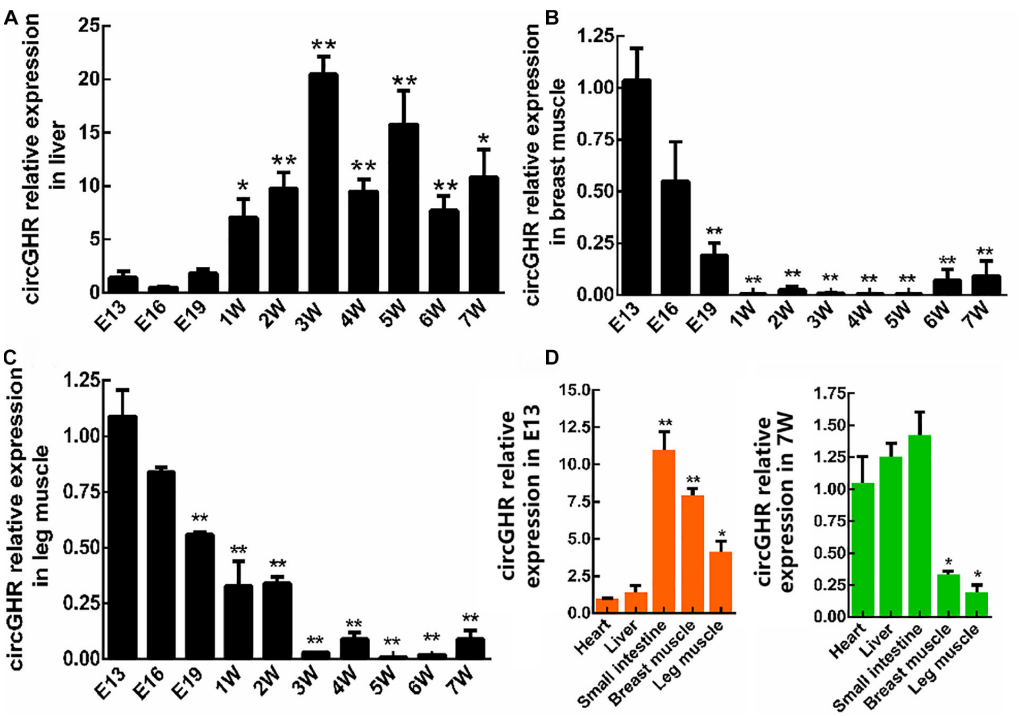


FIGURE 2 | CircGHR relative expression in tissue at different times. **(A–C)** CircGHR time expression profile in the liver, thigh, and breast muscle. Fold change was relative to the expression of E13. **(D)** CircGHR tissue expression profile in E13 and 7W. The RNA relative levels were calculated by the $2^{-\Delta\Delta Ct}$ methods. Fold change was relative to the expression of heart tissues. All data are representative of three independent experiments and are shown as the mean ± SEM. $^{*}P < 0.05$ and $^{**}P < 0.01$.

cell proliferation or is involved in material absorption and tissue development.

Effects of CircGHR on Cell Proliferation

To evaluate the roles of circGHR in cell proliferation, chicken hepatocytes and myoblast were transfected with pCD2.1-ciR or pCD2.1-circGHR, respectively. The expression of proliferation marker genes (*PCNA*, *CCND1*, and *CDK2*) was detected. Quantitative RT-PCR showed that overexpression of circGHR in hepatocytes and myoblast (**Figures 3A,B**, **4A,B**) significantly promoted *PCNA*, *CCND1*, and *CDK2* expression (**Figures 3C–E**, **4C–E**). Consistent with real-time PCR results, the CCK-8 assay results showed that the absorbance of hepatocytes and myoblast transfected with pCD-2.1-circGHR was significantly higher than that of cells transfected with the empty vector pCD2.1-ciR at 48 h (**Figures 3F**, **4F**). Furthermore, the experimental condition

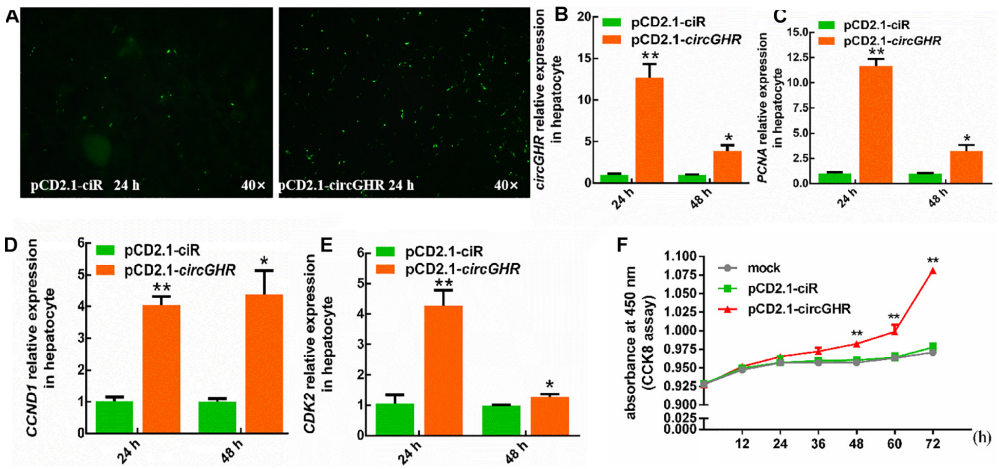


FIGURE 3 | The cell status and the expression profile of the proliferation gene after circGHR overexpression in hepatocytes. **(A)** The status of hepatocytes after transfection vectors 24 h. **(B–E)** The expression profile of circGHR, *PCNA*, *CCND1*, and *CDK2* after circGHR overexpression in hepatocytes. The RNA relative levels were calculated by the $2^{-\Delta\Delta Ct}$ methods. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR at the corresponding time. **(F)** Hepatocyte growth curves following transfection of pCD2.1-ciR and pCD2.1-circGHR. Fold change was relative to the initial value. All data are representative of three independent experiments and are shown as the mean ± SEM. $^{*}P < 0.05$ and $^{**}P < 0.01$.

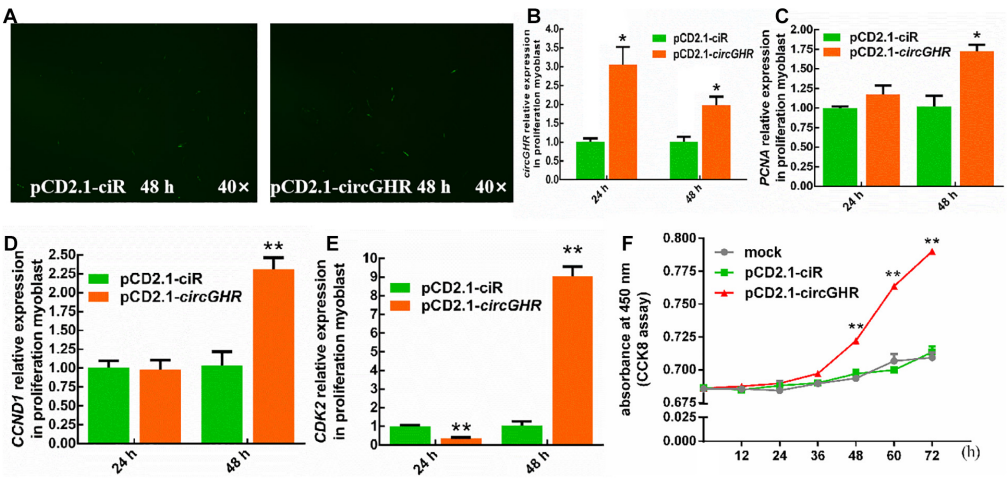


FIGURE 4 | The cell status and the expression profile of the proliferation gene after circGHR overexpression in proliferation phase myoblast. **(A)** The status of myoblast after transfection vectors 48 h. **(B–E)** The expression profile of circGHR, *PCNA*, *CCND1*, and *CDK2* after circGHR overexpression in myoblast. The RNA relative levels were calculated by the $2^{-\Delta\Delta Ct}$ methods. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR at the corresponding time. **(F)** Myoblast growth curves following the transfection of pCD2.1-ciR and pCD2.1-circGHR. Fold change was relative to the initial value. All data are representative of three independent experiments and are shown as the mean ± SEM. $^{*}P < 0.05$ and $^{**}P < 0.01$.

was completed in chicken DF-1 and LMH cell lines, and similar results were displayed (**Supplementary Figures 2**, **3**).

CircGHR Might Promote Cell Proliferation *via* Regulating GHR and GHBP

To explore the potential mechanism of circGHR in promoting cell proliferation, the expression of *GHR* and *GHBP*, transcribed from the *GHR* gene, was detected after circGHR overexpression in the four types of cells. Various results were shown. Chicken *GHR* increased in hepatocytes (**Figure 5A**) and DF-1 (**Supplementary Figure 4A**) but decreased in myoblast (**Figure 6A**) and LMH cell (**Supplementary Figure 5A**), while *GHBP* significantly increased in all the cells (**Figures 5B**, **6B** and **Supplementary Figure 5B**), except in the DF-1 cells (**Supplementary Figure 4B**). The proofs hinted that circGHR might affect cell proliferation by regulating *GHR* and *GHBP* production in different mechanisms that existed in different cells.

To evaluate whether circGHR regulates GHR and GHBP expression, we constructed the *GHR* and *GHBP* promoter-reporter plasmid pGL3-GHR(−2,730/ +226) and pGL3-GHBP(−1,322/ +66), respectively. As expected, both

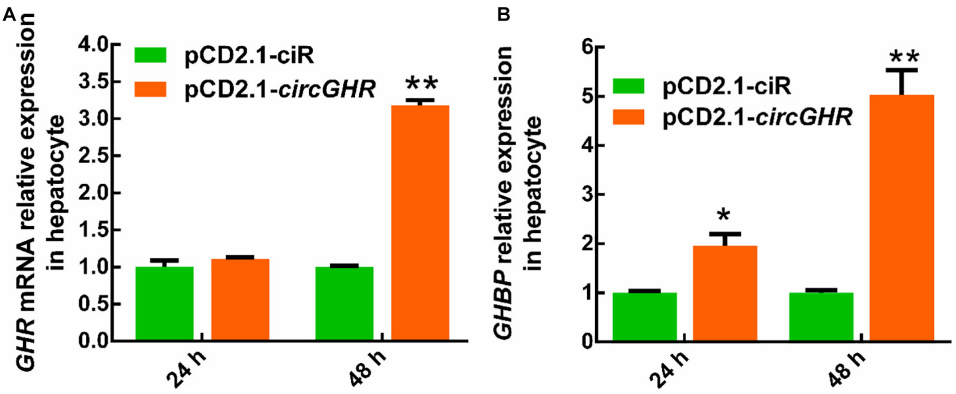


FIGURE 5 | The expression profile of *GHR* gene linear transcripts after circGHR overexpression in hepatocytes. **(A,B)** The expression profile of *GHR* mRNA and *GHBP* after circGHR overexpression in hepatocytes. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR at the corresponding time. All data are representative of three independent experiments and are shown as the mean ± SEM. $^{*}P < 0.05$ and $^{**}P < 0.01$.

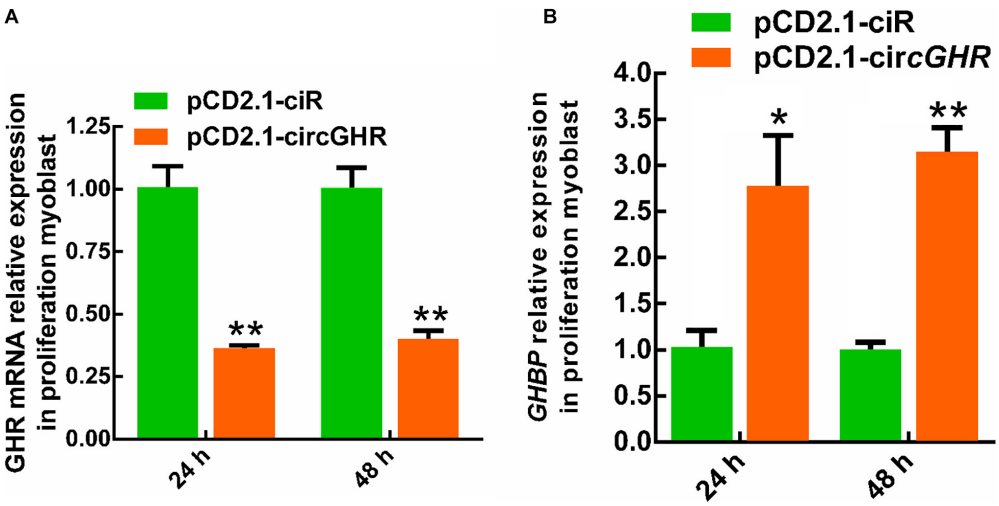


FIGURE 6 | The expression profile of *GHR* gene linear transcripts after circGHR overexpression in proliferation phase myoblast. **(A,B)** The expression profile of *GHR* mRNA and *GHBP* after circGHR overexpression in myoblast during the proliferation phase. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR at the corresponding time. All data are representative of three independent experiments and are shown as the mean ± SEM. $^{*}P < 0.05$ and $^{**}P < 0.01$.

plasmids' luciferase activity was 2.28-fold and 94.74-fold higher than that of the pGL3-basic vector ($P < 0.05$, **Figure 7A**), suggesting that the −2,730/ +226 *GHR* and −1,322/ +66 *GHBP* had promoter activity. Then, we performed co-transfection and reporter gene assay to test whether circGHR regulates *GHR* and *GHBP* promoter activity. The reporter gene assay showed that the luciferase activity of pGL3-GHBP(−1,322/ +66) and pGL3-GHR(−2,730/ +226) was not significantly changed after circGHR overexpression (**Figure 7B**).

Various conserved transcription factor binding sites were predicted in the sequences of circGHR and the promoter regions of *GHR* and *GHBP* transcripts (**Supplementary Table 2**), like aristaless-like homeobox 3 (ALX3), AT-rich interaction domain 3A (Arid3a), nuclear factor erythroid 2 like 1 (NFE2L1), and lin-54 DREAM MuvB core complex component (Lin54) (more detailed information listed in **Supplementary Table 2**). The Lin54 is especially an essential regulator of cell cycle genes by binding to the promoters of G2/M genes whose products are required for mitosis, and participates in their cell cycle-dependent activation (Schmit et al., 2007).

DISCUSSION

Until now, several chicken circRNAs have been found, involved in liver and muscle development (Li et al., 2019). In a previous study, we identified a circRNA transcribed from the *GHR* gene and strongly resisted the exonuclease RNase R (Zhang et al., 2019). Also, reverse transcription showed to be less effective in circGHR using oligo$(dT)_{18}$ primers than random primers. These experiments not only confirmed that circGHR was a circRNA but also laid the foundation for the following experiments. All the

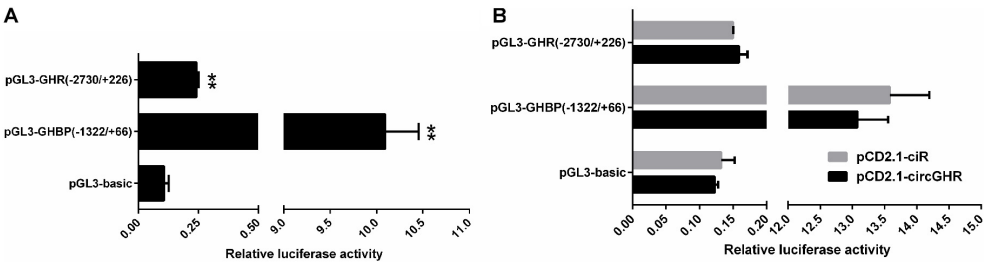


FIGURE 7 | Effects of circGHR overexpression on the promoter activities of *GHR* and *GHBP* transcripts. **(A)** Effects of the promoter activities of the *GHR* and *GHBP* transcripts in DF-1 cells. **(B)** Effects of circGHR on the promoter activities of *GHR* and *GHBP* transcripts in DF-1 cells. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR. All data are representative of three independent experiments and are shown as the mean ± SEM. $^{*}P < 0.05$ and $^{**}P < 0.01$.

RNA should be transcribed with oligo$(dT)_{18}$ and random mixture primers before qRT-PCR.

Myofiber ontogenesis begins very early during embryonic life, with the total number of fibers fixed during fetal life (Xu et al., 2019). In this study, the circGHR expression increased in the liver and decreased in the breast and thigh muscle with chicken development, suggesting that circGHR plays an essential role in chicken embryonic muscle development. In a previous study, we found similar expression trends of *GHR* mRNA in the liver (Wang et al., 2019) and myoblast (Zhang et al., 2016). The metabolic function is gradually improved during chicken liver development, impacting myoblast proliferation in the embryo and early postnatal stage (Welle et al., 2007; Zhou et al., 2013). The positive correlations between circGHR and *GHR* mRNA hinted that circGHR was closely involved in chicken development.

Animal growth and development were affected by the GH–GHR–IGF1 pathway, in which cell proliferation is a meaningful and influential process (Luo et al., 2018). We analyzed the effects of circGHR on cell proliferation in four types of cells and found that circGHR not only can promote cell proliferation but also can increase marker gene expression, indicating that circGHR was closely related to chicken development. Many researchers concluded that circRNAs are involved in animal development *via* regulating their parent's gene expression (Li et al., 2015; Qin et al., 2018), especially for the circRNAs rich in the nucleus. Chicken circGHR was richer in the nucleus than in the cytoplasm in both hepatocytes and myoblast. Thus, we inferred that circGHR might promote cell proliferation *via* regulating *GHR* mRNA and *GHBP* expression. However, dual-luciferase reporter gene assay revealed that the luciferase activity of pGL3-GHBP(−1,322/ +66) and pGL3-GHR(−2,730/ +226) was not significantly changed after circGHR overexpression (**Figure 7B**). Hence, the complex concrete regulation mechanisms need further investigation.

CircRNAs regulate gene transcription at the initiation and elongation steps as well as during post-transcription. Li et al. reported that circEIF3J and circPAIP2 performed a scaffold for combining with U1 snRNPs and polymerase II in increasing parental gene *EIF3J* and *PAIP2* mRNA expression, respectively (Li et al., 2015). Additionally, ci-ankrd5, ci-sirt7, and ci-mcm5, transcribed from intron sequences, regulated homologous linear RNA transcript expression by associating with phosphorylated polymerase II, which is pivotal to the transcription elongation process (Zhang et al., 2013). Furthermore, circMbl contains several MBL protein-binding sites and, thus, can act as an RBP sponge by binding and sequestering the MBL protein, lowering its free cellular concentration so that is can no longer produce circMbl but manufacture mbl linear transcripts, suggesting that circMbl regulates circMbl and *Mbl* linear transcript expression in posttranscriptional patterns (Ashwal-Fluss et al., 2014). Some circRNAs sequester the translation start site on their linear transcripts to monitor their protein expression level, like "mRNA trapping" competing with linear transcript expression (Chao et al., 1998).

The sequence of circGHR contained 5′ UTR of the *GHR* gene, reminding us that both 5′ UTR and the intron cannot be translated and that 5′ UTR can be considered as the intron original circRNA. Coincidentally, *GHR* mRNA and *GHBP* are transcribed from different transcription initiation sites and polyadenylation sites of the *GHR* gene, thus forming two different alternative splicings during mRNA maturation. The expression of *GHR* and *GHBP* transcripts was not always increased in circGHR, which is overexpressed among different cells, and the effects and mechanisms were partly variant in different cells, even though false-positive results may exist. Therefore, we hypothesized that circGHR is involved in the formation of two transcripts and regulated their expressions like circEIF3J and circPAIP2; however, this requires further research.

Abundance indications describing circRNAs interacting with transcription factors existed. Circ-CTNNB1 bound EDAD-box polypeptide 3 (DDX3) to facilitate its physical interaction with the transcription factor Yin Yang 1 (YY1), resulting in the transactivation of YY1 and transcriptional alternation of downstream genes (Yang et al., 2019). circ-DONSON recruits the NURF complex to the SOX4 promoter and initiates its transcription (Ding et al., 2019). Interestingly, numerous conserved transcription factor binding sites were predicted in the sequences of circGHR and the promoter regions of *GHR* and *GHBP* transcripts, like ALX3, Arid3a, NFE2L1, and Lin54 (**Supplementary Table 2**). Especially, lin54 is an essential regulator of cell cycle genes, which may be the reason to explain that circGHR promoted cell proliferation in this study (Schmit et al., 2007, 2010).

CONCLUSION

In summary, we proved that circGHR positively regulated chicken cell proliferation. CircGHR was a closed-loop structure molecule and richer in the nucleus in hepatocytes and myoblast. Quantitative real-time PCR exhibited that circGHR was increased in the liver from E13 to 7W but decreased in thigh and breast muscle. Further analysis showed that circGHR overexpression promoted cell proliferation and biomarker gene expression, as well as regulated *GHR* and *GHBP* expression in various types. Additionally, circGHR had no significant effects on the promoter activity of *GHR* and *GHBP* in the DF-1 cell line.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The animal study was reviewed and approved by Animal Care Committee of Guangdong Ocean University. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

HX and LZ participated in the study design and coordination and wrote the first draft of the manuscript. HX, QL, JZ, PA-A, TL, and ZW carried out the experiments and statistical analysis. SL was involved in data interpretation and revision. LA, ZZ, and LZ participated in writing the final versions of the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation (31672412 and 31972550) and the Natural Science Foundation of Guangdong Province (2020A1515011576).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.598575/full#supplementary-material

Supplementary Figure 1 | The experiment summary technical pipeline about this research paper.

Supplementary Figure 2 | The cell status and the expression profile of the proliferation gene after circGHR overexpression in DF-1 cell. **(A)** The status of DF-1 cell after transfection vectors 48 h. **(B–D,E)** The expression profile of circGHR, *PCNA*, *CCND1*, and *CDK2* after circGHR overexpression in DF-1 cell. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR at the corresponding time. **(F)** DF-1 cell growth curves following the transfection of pCD2.1-ciR and pCD2.1-circGHR. Fold change was relative to the initial value. All data are representative of three independent experiments and are shown as the mean ± SEM. $*P < 0.05$ and $**P < 0.01$.

Supplementary Figure 3 | The cell status and the expression profile of the proliferation gene after circGHR overexpression in LMH cell. **(A)** The status of LMH cell after transfection vectors 48 h. **(B–D,E)** The expression profile of circGHR, *PCNA*, *CCND1*, and *CDK2* after circGHR overexpression in LMH cell. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR at the corresponding time. **(F)** LMH cell growth curves following the transfection of pCD2.1-ciR and pCD2.1-circGHR. Fold change was relative to the initial value. All data are representative of three independent experiments and are shown as the mean ± SEM. $*P < 0.05$ and $**P < 0.01$.

Supplementary Figure 4 | The expression profile of *GHR* gene linear transcripts after circGHR overexpression in DF-1 cell. **(A,B)** The expression profile of *GHR* mRNA and *GHBP* after circGHR overexpression in DF-1 cell. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR at the corresponding time. All data are representative of three independent experiments and are shown as the mean ± SEM. $*P < 0.05$ and $**P < 0.01$.

Supplementary Figure 5 | The expression profile of GHR gene linear transcripts after circGHR overexpression in LMH cell. **(A,B)** The expression profile of *GHR* mRNA and *GHBP* after circGHR overexpression in LMH cell. Fold change was relative to the expression of the cells transfected with empty vector pCD2.1-ciR at the corresponding time. All data are representative of three independent experiments and are shown as the mean ± SEM. $*P < 0.05$ and $**P < 0.01$.

Supplementary Table 1 | Original data.

Supplementary Table 2 | The conserved transcription factors in circGHR and the promoter regions of *GHR* and *GHBP* transcripts.

REFERENCES

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

Buckingham, M., Bajard, L., Chang, T., Daubas, P., Hadchouel, J., Meilhac, S., et al. (2003). The formation of skeletal muscle: from somite to limb. *J. Anat.* 202, 59–68. doi: 10.1046/j.1469-7580.2003.00139.x

Chao, C. W., Chan, D. C., Kuo, A., and Leder, P. (1998). The mouse formin (Fmn) gene: abundant circular RNA transcripts and gene-targeted deletion analysis. *Mol. Med.* 4, 614–628. doi: 10.1007/BF03401761

Conn, S. J., Pillman, K. A., Toubia, J., Conn, V. M., Salmanidis, M., Phillips, C. A., et al. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125–1134. doi: 10.1016/j.cell.2015.02.014

Ding, L., Zhao, Y., Dang, S., Wang, Y., Li, X., Yu, X., et al. (2019). CircularRNA circ-DONSON facilitates gastric cancer growth and invasion via NURF complex dependent activation otranscription factor SOX4. *Mol. Cancer* 18:45. doi: 10.1186/s12943-019-1006-2

Fukami-Kobayashi, J., and Mitsui, Y. (1999). Cyclin D1 inhibits cell proliferation through binding to *PCNA* and *CDK2*. *Exp. Cell Res.* 246, 338–347. doi: 10.1006/excr.1998.4306

Green, M. R., and Sambrook, J. (2014). *Molecular Cloning: A Laboratory Manual*, 4 th Edn. New York, NY: Cold Spring Harbor Laboratory Press.

Hamada, M., Kihira, T., Takase, K., Nakano, T., Tameda, Y., and Kosaka, Y. (1995). Hepatocyte regeneration in chronic hepatitis C and interferon treatment: analysis of immunohistological identification of proliferating cell nuclear antigen (PCNA). *J. Gastroenterol.* 30, 372–378. doi: 10.1007/BF02347514

Lau, S. L., and Joanna (2005). *Molecular Characterization of the Chicken Growth Hormone Receptor Gene*. Hong Kong: Hong Kong University.

Legnini, I., Timoteo, G. D., 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. doi: 10.1016/j.molcel.2017.02.017

Li, H., Yang, J., Wei, X., Song, C., Dong, D., Huang, Y., et al. (2017). CircFUT10 reduces proliferation and facilitates differentiation of myoblasts by sponging miR-133a. *J. Cell. Physiol.* 233, 4643–4651. doi: 10.1002/jcp.26230

Li, L., Guo, J., Chen, Y., Chang, C., and Xu, C. (2017). Comprehensive circRNA expression profile and selection of key circRNAs during priming phase of rat liver regeneration. *BMC Genomics* 18:80. doi: 10.1186/s12864-016-3476-6

Li, P., Shan, K., Liu, Y., Zhang, Y., Xu, L., and Xu, L. (2019). CircScd1 promotes fatty liver disease via the Janus kinase 2/signal transducer and activator of transcription 5 pathway. *Dig. Dis. Sci.* 64, 113–122. doi: 10.1007/s10620-018-5290-2

Li, Z., Huang, C., Chun, B., Liang, C., 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

Luo, W., Chen, J., Li, L., Ren, X., Cheng, T., Lu, S., et al. (2018). c-Myc inhibits myoblast differentiation and promotes myoblast proliferation and muscle fibre hypertrophy by regulating the expression of its target genes, miRNAs and lincRNAs. *Cell Death Differ.* 26, 426–442. doi: 10.1038/s41418-018-0129-0

Ovcharenko, I., Loots, G. G., Giardine, B. M., Hou, M. M., Ma, J., Hardison, R. C., et al. (2005). Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res.* 15, 184–194. doi: 10.1101/gr.3007205

Patricia, A.-A., Leng, Q., Xu, H., Zheng, J., and Zhang, L. (2020). Functions of circular RNAs involved in animal skeletal muscle development – a review. *Ann. Anim. Sci.* 20, 3–10. doi: 10.2478/aoas-2019-0053

Picard, B., Lefaucheur, L. C., and Duclos, M. J. (2002). Muscle fibre ontogenesis in farm animal species. *Reprod. Nutr. Dev.* 42, 415–431. doi: 10.1051/rnd:2002035PMIDPMID:NOPMID

Qin, M., Wei, G., and Sun, X. (2018). Circ-UBR5: an exonic circular RNA and novel small nuclear RNA involved in RNA splicing. *Biochem. Biophys. Res. Commun.* 503, 1027–1034. doi: 10.1016/j.bbrc.2018.06.112

Schmit, F., Cremer, S., and Gaubatz, S. (2010). LIN54 is an essential core subunit of the DREAM/LINC complex that binds to the cdc2 promoter in a sequence−specific manner. *FEBS J.* 276, 5703–5716. doi: 10.1111/j.1742-4658.2009.07261.x

Schmit, F., Korenjak, M., Mannefeld, M., Schmitt, K., Franke, C., Von Eyss, B., et al. (2007). LINC, a human complex that is related to pRB-containing complexes in invertebrates regulates the expression of G2/M genes. *Cell Cycle* 6, 1903–1913. doi: 10.4161/cc.6.15.4512

Sperling, M. A. (2016). Traditional and novel aspects of the metabolic actions of growth hormone. *Growth Horm. IGF Res.* 28, 69–75. doi: 10.1016/j.ghir

Wang, Z., Xu, H. D., Li, T., Wu, J., An, L. L., Zhao, Z. H., et al. (2019). Chicken GHR antisense transcript regulates its sense transcript in hepatocytes. *Gene* 682, 101–110. doi: 10.1016/j.gene.2018.10.001

Welle, S., Bhatt, K., Pinkert, C. A., Tawil, R., and Thornton, C. A. (2007). Muscle growth after postdevelopmental myostatin gene knockout. *Am. J. Physiol. Endocrinol. Metab.* 292, E985–E991. doi: 10.1152/ajpendo.00531.2006

Xu, H., Li, T., Zhang, W., Adu-Asiamah, P., Leng, Q., Zheng, J., et al. (2019). Roles of chicken growth hormone receptor antisense transcript in chicken muscle development and myoblast differentiation. *Poult. Sci.* 98, 6980-6988. doi: 10.3382/ps/pez416

Yang, F., Fang, E., Mei, H., Chen, Y., Li, H., Li, D., et al. (2019). Cis-Acting circ-CTNNB1 promotes b-Catenin signaling and cancer progression via DDX3-Mediated transactivation of YY1. *Cancer Res.* 79, 557–571. doi: 10.1158/0008-5472.CAN-18-1559

Zhang, L., Lin, S. D., An, L. L., Ma, J. E., Qiu, F. F., Jia, R. M., et al. (2016). Chicken GHR natural antisense transcript regulates GHR mRNA in LMH cells. *Oncotarget* 7, 73607–73617. doi: 10.18632/oncotarget.12437

Zhang, L., Wang, Z., Li, T., Xu, H., Xiao, M., Liu, M., et al. (2017). The effect of GH protein on the expression of IGF1 in chicken(*Gallus gallus*) myoblast cells. *J. Agric. Biotechnol.* 25, 511–516.

Zhang, L., Xu, H., Wang, Z., Li, T., Guo, J., Patricia, A.-A., et al. (2019). Identification and characterization of circular RNAs in chicken hepatocytes. *Growth Horm. IGF Res.* 46, 16–23. doi: 10.1016/j.ghir.2019.05.002

Zhang, Y., Zhang, X., Chen, T., Xiang, J., Ying, Q., Xing, Y., et al. (2013). Circular intronic long noncoding RNAs. *Mol. Cell* 51, 792–806. doi: 10.1016/j.molcel.2013.08.017

Zhou, Z., Xiao, H., Zhang, J., Peng, S., and Zhang, H. (2013). Postembryonic developmental microstructure of liver in Taihe black-bone silky fowl. *Chin. J. Vet. Sci.* 33, 1744–1749.

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

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 10 March 2021
doi: 10.3389/fgene.2021.615340

Functional Network of the Long Non-coding RNA Growth Arrest-Specific Transcript 5 and Its Interacting Proteins in Senescence

Siqi Wang[1†], Shengwei Ke[1†], Yueming Wu[1], Duo Zhang[1,2], Baowei Liu[1], Yao-hui He[3], Wen Liu[3], Huawei Mu[1] and Xiaoyuan Song[1]**

[1] Hefei National Laboratory for Physical Sciences at the Microscale, CAS Key Laboratory of Brain Function and Disease, Division of Life Sciences and Medicine, School of Life Sciences, University of Science and Technology of China, Hefei, China, [2] CAS Key Laboratory of Mechanical Behavior and Design of Materials, Department of Modern Mechanics, University of Science and Technology of China, Hefei, China, [3] School of Pharmaceutical Sciences, Fujian Provincial Key Laboratory of Innovative Drug Target Research, Xiamen University, Xiamen, China

OPEN ACCESS

Edited by:
Yuanchao Xue,
Institute of Biophysics (CAS), China

Reviewed by:
Gary S. Stein,
University of Vermont, United States
Simranjeet Kaur,
Steno Diabetes Center Copenhagen (SDCC), Denmark

***Correspondence:**
Huawei Mu
huaweimu@ustc.edu.cn
Xiaoyuan Song
songxy5@ustc.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: 08 October 2020
Accepted: 01 February 2021
Published: 10 March 2021

Citation:
Wang S, Ke S, Wu Y, Zhang D, Liu B, He Y-H, Liu W, Mu H and Song X (2021) Functional Network of the Long Non-coding RNA Growth Arrest-Specific Transcript 5 and Its Interacting Proteins in Senescence. Front. Genet. 12:615340. doi: 10.3389/fgene.2021.615340

Increasing studies show that long non-coding RNAs (lncRNAs) play essential roles in various fundamental biological processes. Long non-coding RNA growth arrest-specific transcript 5 (GAS5) showed differential expressions between young and old mouse brains in our previous RNA-Seq data, suggesting its potential role in senescence and brain aging. Examination using quantitative reverse transcription-polymerase chain reaction revealed that GAS5 had a significantly higher expression level in the old mouse brain hippocampus region than the young one. Cellular fractionation using hippocampus-derived HT22 cell line confirmed its nucleoplasm and cytoplasm subcellular localization. Overexpression or knockdown of GAS5 in HT22 cell line revealed that GAS5 inhibits cell cycle progression and promotes cell apoptosis. RNA-Seq analysis of GAS5-knockdown HT22 cells identified differentially expressed genes related to cell proliferation (e.g., DNA replication and nucleosome assembly biological processes). RNA pull-down assay using mouse brain hippocampus tissues showed that potential GAS5 interacting proteins could be enriched into several Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and some of them are involved in senescence-associated diseases such as Parkinson's and Alzheimer's diseases. These results contribute to understand better the underlying functional network of GAS5 and its interacting proteins in senescence at brain tissue and brain-derived cell line levels. Our study may also provide a reference for developing diagnostic and clinic biomarkers of GAS5 in senescence and brain aging.

Keywords: long non-coding RNA, Gas5, RNA pull down, mouse brain tissue, brain-derived cell line, RNA-seq

INTRODUCTION

Cellular senescence refers to the cessation of normal cell division under various conditions such as cellular stress and DNA damage. Senescent cells have several characteristics, including constitutive DNA damage response, increased activity of senescence-related galactosidase, higher expression of p16 (*CDKN2A*) and p21 (*CDKN1A*), and formation of senescence-associated

secretory phenotype and senescence-associated heterochromatin aggregation (Wei and Ji, 2018). Cell cycle arrest is also a key feature of cellular senescence (Calcinotto et al., 2019). Studies have shown that nuclear fiber layer forms can be used as an identifier of phenotypic senescent cells (Raz et al., 2008), at the same time, mitochondrial DNA damage causes mitochondrial dysfunction and upregulation of reactive oxygen species, which may be induced by telomere dysfunction caused cellular senescence (Passos et al., 2007). Higher oxidative status caused by mitochondrial dysfunction was reported to contribute to senescence acceleration and the age-dependent alterations in cell structure and function (Hosokawa, 2002, 2004).

The hippocampus, a brain region critically involved in learning and memory, is particularly susceptible to dysfunction during senescence (Kumar and Foster, 2019). It has been reported that senescent glutamatergic synapses contributed to the age-related cognitive impairment based on the mechanisms for age-associated changes in Ca^{2+}-dependent synaptic plasticity (Kumar and Foster, 2019). Astrocytes, which are star-shaped glial cells in the brain, play critical roles in maintaining normal brain physiology during development and in adulthood (Wu et al., 2005). Studies showed that the expression level of glial fibrillary acidic protein (GFAP), a protein marker of astrocytes in the brain, was upregulated significantly in old senescence-accelerated-prone 8 mice, indicating an important role of GFAP in the age-related deficits in learning and memory (Wu et al., 2005). In addition, upregulation of senescence-associated proteins p16 and p21 was found in granule cells of the dentate gyrus in the irradiation-induced mouse hippocampus, showing a similar senescence phenotype (Cheng et al., 2017).

Long non-coding RNAs (lncRNAs), or RNAs > 200 nt, which remain untranslated although sometimes generate short peptides (Payre and Desplan, 2016), have a variety of functions, such as acting as a scaffold, bait, or signaling molecule, and playing a role through targeting in the genome by *cis* or *trans* regulation and antisense interference (Quinn and Chang, 2016). LncRNAs are widely distributed in cells, both in the nucleus and cytoplasm and even in mitochondria. In recent years, functional nuclear lncRNA has been widely reported. Xist, as a classical lncRNA in the nucleus, can cause random silencing of an X chromosome in female mammals (Plath et al., 2002). Metastasis-associated lung adenocarcinoma transcript 1 is widely expressed in normal mammalian tissues and abnormally expressed in many human malignant tumors, which plays an important role in varying degrees of tumor proliferation, apoptosis, invasion and metastasis, and drug resistance (Gutschner et al., 2013). Telomeres are transcribed to produce a lncRNA called TERRA (telomere repeating RNA). It can promote homologous recombination between telomeres, delay cell senescence, and maintain genomic instability (Bunch et al., 2019). Recently, more and more studies have found that lncRNAs play important roles in cellular senescence (Abdelmohsen et al., 2013); for example, lncRNA GUARDIN suppresses cellular senescence through upregulation of p21 (Sun et al., 2020), whereas lncRNA SENEBLOC is shown to block the induction of cellular senescence through dual mechanisms that converge to repress the expression of p21 (Xu et al., 2020), and lncRNA OVAAL blocks cell senescence by regulating the expression of CDK inhibitor p27 (Sang et al., 2018). LncRNA-OIS1 modulates oncogenic RAS-induced senescence through regulating the activation of dipeptidyl peptidase 4, which has established a role in tumor suppression (Li et al., 2018). LncRNA ANRIL inhibits senescence in vascular smooth muscle cells, possibly through regulating miR-181a/Sirt1 and inhibiting the p53-p21 pathway (Tan et al., 2019).

Growth arrest-specific transcript 5 (GAS5), a member of the lncRNA family, including many isoforms and some of which were localized in the nucleus, plays important roles in various processes. GAS5 can be folded into a secondary RNA structure, exposing the sequence similar to the glucocorticoid-receptor binding element, and inhibits glucocorticoid-mediated physiological functions by interacting with the DNA binding domain of the glucocorticoid receptor (Kino et al., 2010). Studies in human renal epithelial cell HEK293T have demonstrated that GAS5 can affect the expression of glucocorticoid-induced protein kinases by their downstream signaling molecules cIAP2 and SGK1 through competitive binding of glucocorticoid receptors and participates in the regulation of cell proliferation and cell growth (Tani et al., 2013). GAS5 can also be used as a decoy for microRNA or shear factors, simulating the adsorption of molecular sponge, and directly binds to microRNAs to regulate their downstream target genes (Huang, 2018). GAS5 can adsorb miR-106a-5p to regulate the Akt/mTOR pathway, thus participating in the development of gastric cancer (Dong et al., 2019). GAS5 can also participate in the regulation of gene translation. In retinal ganglion cells, the downregulation of GAS5 can reduce the expression of adenosine triphosphate-binding cassette transporter A1, thus inhibiting the apoptosis of retinal ganglion cells (Zhou et al., 2019).

Our previous RNA-Seq data showed that GAS5 exhibits differential expression patterns in young and old mice brain tissues, which suggested that it may play a role in brain aging (Wang et al., 2019). Thus, we conducted various experiments on mice brain tissues and brain-derived cell lines to explore the role and underlying mechanism of GAS5 in senescence and brain aging. In this study, we found that GAS5 inhibits cell cycle progression and promotes cell apoptosis in the hippocampus-derived HT22 cell line. Protein expression of p21, a senescence-associated marker protein, exhibited a negative correlation with GAS5 RNA level, indicating an antiaging effect of GAS5. RNA-Seq analysis of GAS5 KD HT22 cells identified differentially expressed genes (DEGs) related to cell proliferation biological processes (for example, DNA replication, nucleosome assembly, and DNA replication-dependent nucleosome assembly), which is consistent with our cell cycle analysis results. We also performed an RNA pull-down assay which suggested that GAS5 may participate in hippocampus senescence through interacting with protein groups that are involved in mitochondrial membrane respiratory chain reaction and regulation of the synaptic structure of neurons, which is associated with neurological diseases (such as Parkinson's disease and Alzheimer's disease).

RESULTS

Growth Arrest-Specific Transcript 5 Is Highly Expressed in the Hippocampus Region of Old Mice

To obtain the spatial expression landscape of GAS5 in mice brain regions, we performed quantitative reverse transcription PCR (RT-qPCR) on various brain regions. Young and old male adult mice brains were dissected, and four brain regions were separated: hypothalamus (HT), olfactory bulb (OB), cerebellum (CB), and hippocampus (HC) (**Figure 1A**). As the GAS5 gene encodes numerous alternatively spliced lncRNA isoforms (Goustin et al., 2019), we designed two primers to cover as many GAS5 isoforms as possible for the RT-qPCR experiment: GAS5-E234 and GAS5-E456 (**Figure 1B**). Primer lists are shown in **Supplementary Table 1**. Results revealed that GAS5 showed significantly higher expression levels in old mice HC region than in young ones (**Figure 1C**), whereas the other three brain regions (HT, OB, and CB) did not have significant differential expression levels (**Figures 1D–F**). This result was consistent with the high expression level of GAS5 in a growth-arrested cell line (Plath et al., 2002). Our result also suggested that GAS5 may participate in senescence in the brain, as it showed higher expression in the aging mouse brain and also had a brain region-specific expression pattern. We also examined the expression level of GAS5 in brain-derived cell lines—C8D1A (cerebellum) and HT22 (hippocampus) (**Figure 1G**). Two pairs of GAS5 primers showed all high expression levels in these two cell lines, and we chose the HT22 cells to conduct the KD and OE in the following experiments because the two GAS5 primers all showed relatively consistent and high expression in HT22 cells.

To elucidate the subcellular localization of GAS5, cell fractionation was performed on HT22 cells, and we got three fractions: chromatin binding, nucleoplasm, and cytoplasm. The relative expression level (in percentage) of GAS5 and the messenger RNAs (mRNAs) of two reference genes (*GAPDH* and *U6*) in each cellular fraction was also calculated (**Figure 1H**). Results suggested that most of the GAS5 isoforms were mainly localized in nucleoplasm and cytoplasm.

Growth Arrest-Specific Transcript 5 Inhibits Cell Growth and Promotes Apoptosis in Hippocampus-Derived HT22 Cells

To further identify the molecular function of GAS5 in hippocampus-derived HT22 cells, we used GAS5-specific small interfering RNAs (siRNAs) (si210 and si596) to perform the KD experiment. The RT-qPCR result showed that GAS5 RNA was efficiently knocked down through siRNA compared with the control group (**Figure 2A**). Cell cycle analysis was performed after GAS5 KD, and results showed that depletion of GAS5 led to decreased cell number in the G1 phase (although only one siRNA group showed significance, the other siRNA group showed the same downregulation trend), whereas the cell number in S/G2/M phase was upregulated, suggesting accelerated cell cycle transition from G1 phase into the S/G2/M phase upon GAS5 KD (**Figure 2B**, quantification in **Figure 2C**). We further used annexin V–fluorescein isothiocyanate and propidium iodide (PI) to detect the apoptosis level in GAS5 KD HT22 cells. The percentage of late-stage apoptosis cells decreased after GAS5 KD (**Figure 2D**, quantification in **Figure 2E**), indicating loss of function of GAS5 inhibits cell apoptosis.

We cloned the complementary DNA (cDNA) of GAS5 into the pcDNA3.1(+) vector, and the GAS5 + pcDNA3.1(+) OE vector was introduced into HT22 cells with transfection reagent to overexpress GAS5 RNA. RT-qPCR was performed to confirm that GAS5′s RNA expression was around six times higher in the GAS5 OE group than the control group (**Figure 2F**). OE of GAS5 led to the opposite effect of GAS5 KD. In cell cycle analysis, cell numbers in the S/G2/M phase decreased in the GAS5 OE group, with around 10% cell arresting in the G1 phase (**Figure 2G**). In apoptosis assay, the portion of early- and late-stage apoptosis cells together increased upon GAS5 OE (**Figure 2H**).

Western blotting was performed to further detect the senescence-associated maker proteins (p16 and p21) in HT22 cells upon GAS5 KD or OE. In GAS5-KD HT22 cells, we found that p21 protein expression was upregulated (**Figure 2I**, quantification in **Figure 2J**). In the GAS5 OE group, p21 protein showed an opposite expression pattern with a lower expression level (**Figure 2K**, quantification in **Figure 2L**). However, p16 protein expression did not change significantly upon GAS5 KD but decreased in the GAS5 OE group (**Figures 2I,K**, quantification in **Figures 2J,L**). Overall, our data suggested that GAS5 inhibits cell cycle progression and promotes the apoptosis process in HT22 cells.

RNA-Seq Revealed That Genes Affected by Downregulation of Growth Arrest-Specific Transcript 5 Might Participate in Cell Cycle

To further explore the genes that GAS5 may regulate, we performed RNA-Seq analysis using GAS5 KD HT22 cells with siRNA NC as a control. The data showed that GAS5 RNA level indeed decreased in both siRNA 210 and siRNA 596 KD HT22 cells. Using the Pearson correlation analysis method, we found that the reproducibility of the data of two siRNA KD groups was very good (R^2 = 0.9965 for KD group; R^2 = 0.9707 for NC group). We thus combined the data of siRNA 210 and siRNA 596 KD groups together as KD group and performed a comparison between KD and control groups.

The result showed that a total of 58 DEGs were identified upon GAS5 KD in HT22 cells using the criteria of P-value < 0.05 and fold change > 2 (**Figure 3A**). The heatmap of hierarchical clustering analysis indicated that all of the differentially expressed mRNAs were clustered in several groups (**Figure 3B** and **Supplementary Table 2**). We further performed gene ontology (GO) enrichment analysis to explore the biological process of the DEGs involved in. We found that 27 GO terms were significantly enriched for the upregulated DEGs ($P < 0.05$), and the top 10 enriched terms included nucleosome assembly, protein-DNA complex assembly, and DNA replication-dependent nucleosome

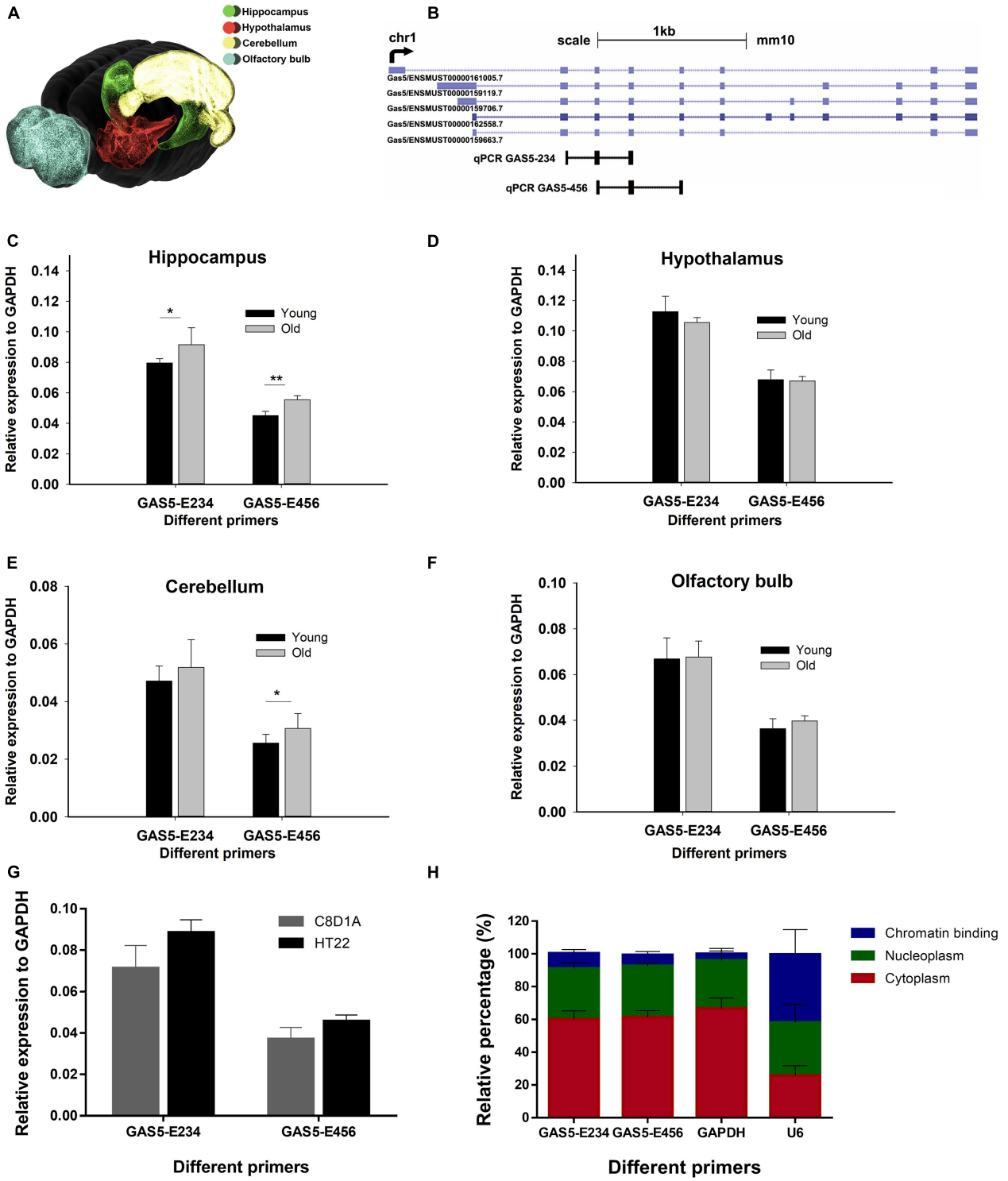


FIGURE 1 | GAS5 exhibits differential expressions in different mouse brain regions of different ages. **(A)** A three-dimensional model of mouse brain showing the four regions selected for RT-qPCR. Color with green, red, yellow, and blue represent hippocampus (HC), hypothalamus (HT), cerebellum (CB), and olfactory bulb (OB), respectively. Brain model was adopted from Blue Brain Cell Atlas. Representative isoforms of GAS5 transcript were shown in **(B)** (data from University of California, Santa Cruz), and two qPCR amplicons of GAS5 (E234 and E456) were indicated below. Expression level GAS5 (represented by two pairs of primers: GAS5-E234 and GAS5-E456) in four different brain regions of young and old mice: hippocampus **(C)**, hypothalamus **(D)**, cerebellum **(E)**, and olfactory bulb **(F)**. Relative expression level of GAS5 in two brain-derived cell lines **(G)**. Subcellular localization of GAS5 in HT22 cell line **(H)**. $N = 4$ for young mice brain sample; $N = 5$ for old mice sample; $N = 3$ for **(G,H)**. $^{*}P < 0.05$; $^{**}P < 0.01$.

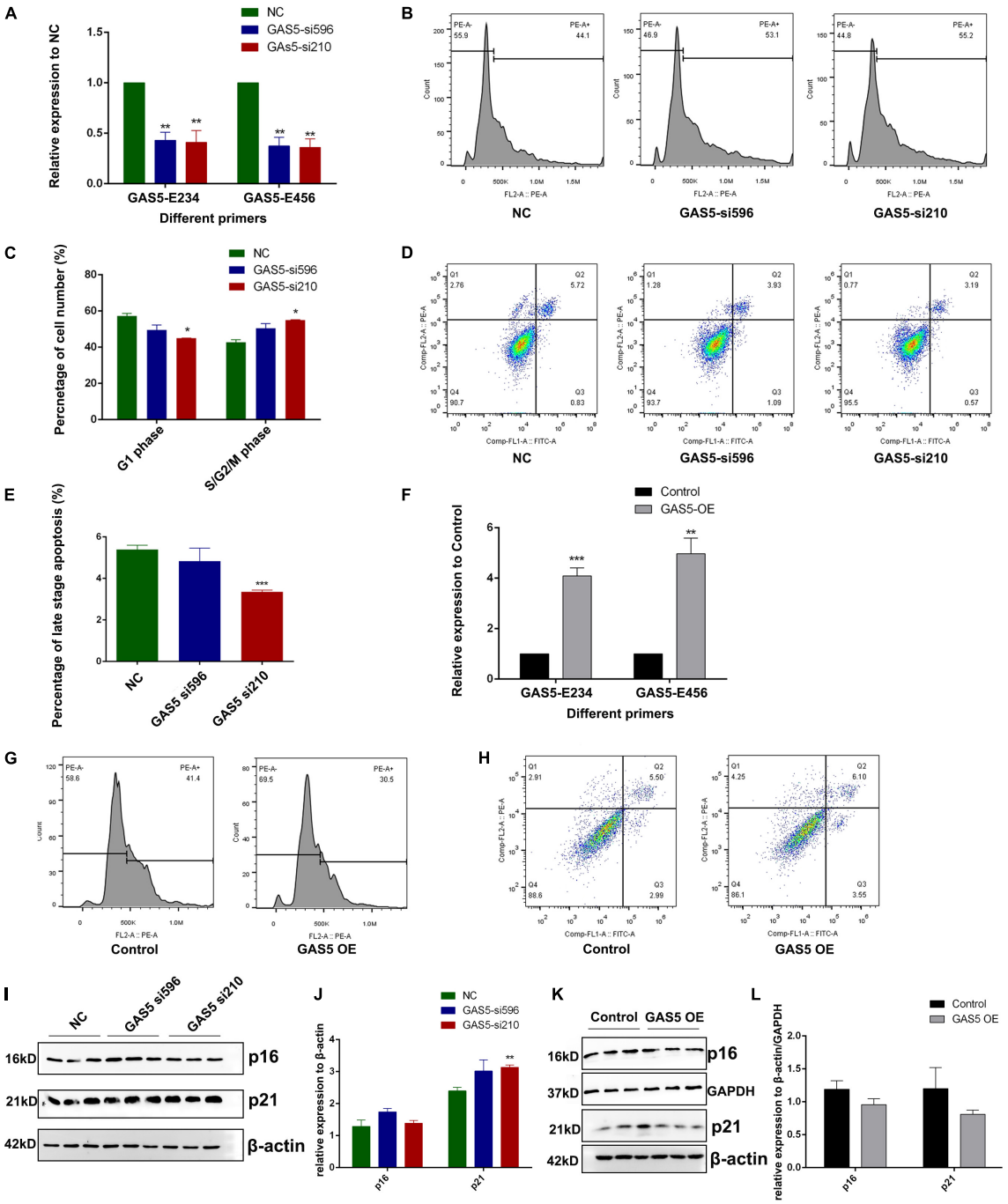


FIGURE 2 | GAS5 inhibits cell growth and promotes apoptosis level in HT22 cells. **(A)** RT-qPCR was performed to quantify the relative expression of GAS5 using two different primers (GAS5-E234 and GAS5-E456) that can amplify different GAS5 isoforms in GAS5-knockdown (KD) HT22 cells. Two different siRNAs (GAS5-si596 and GAS5-si210) were used to KD GAS5. **(B)** Cell cycle analysis of GAS5 KD HT22 cell. NC, GAS5-si596, GAS5-si210 represents HT22 cells transfected with negative control siRNA, GAS5 siRNA 596, and GAS5 siRNA 210, respectively. **(C)** Percentage of cell number at G1 phase and S/G2/M phase during cell cycle was quantified. Columns in different colors represent control and GAS5 KD groups. **(D)** Cell apoptosis analysis of GAS5 KD in HT22 cells. Cells stained with negative or positive annexin V–fluorescein isothiocyanate and propidium iodide signal were separated into four parts. Top right part represents late-stage apoptosis cell. **(E)** Percentage of late-stage apoptosis cell number was quantified. Columns in different colors represent control and KD groups. **(F)** pcDNA3.1(+) vector contains GAS5 cDNA was introduced to overexpress GAS5 in HT22 cells. RT-qPCR was performed to quantify the relative expression of GAS5. **(G)** Cell cycle analysis of GAS5 overexpression (OE) in HT22 cells. **(H)** Cell apoptosis analysis of GAS5 OE HT22 cells. **(I)** Immunoblot result of p16 and p21 in GAS5-KD HT22 cells. β-actin was used as a loading control. Intensity quantification was showed in **(J)**. **(K)** Immunoblot result of p16 and p21 in GAS5 OE HT22 cells. β-actin and *GAPDH* were used as loading control. Intensity quantification was showed in **(L)**. Three biological repeats were performed in each experiment, except for **(G,H)** ($N = 1$). $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

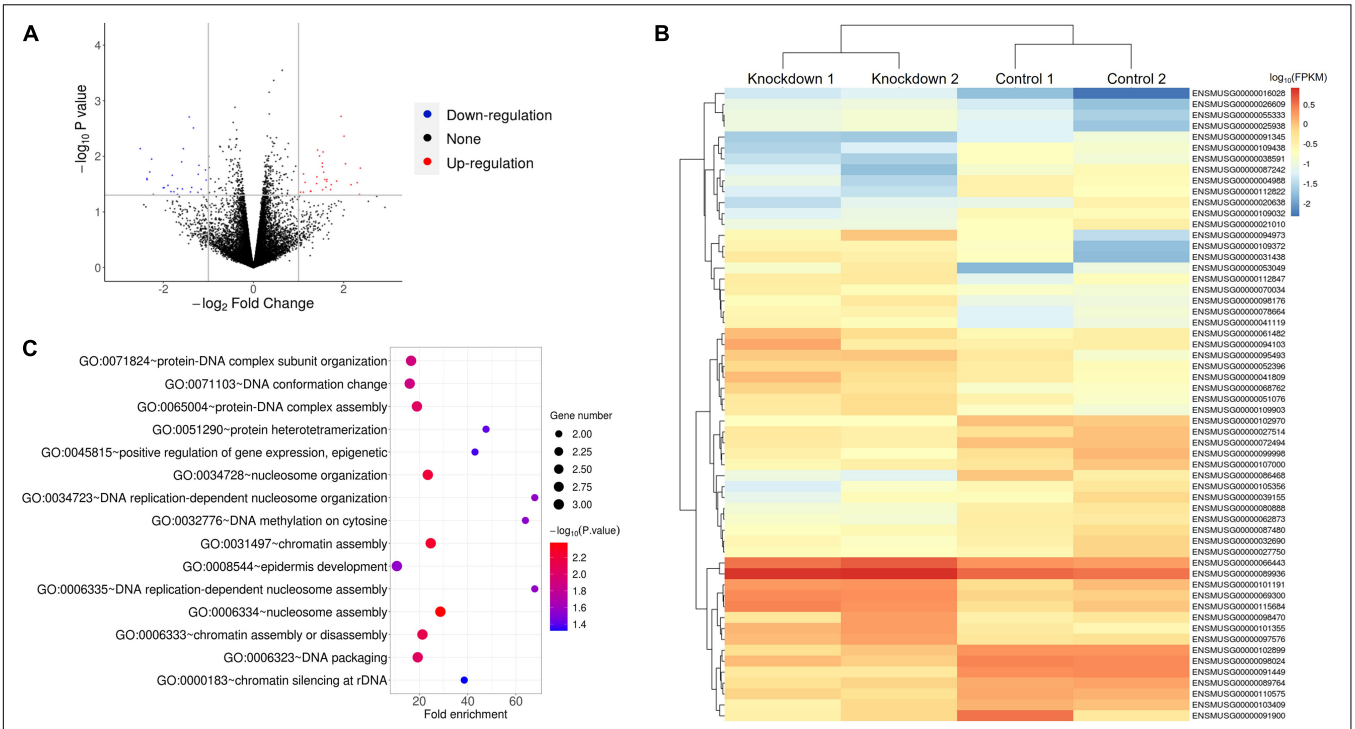


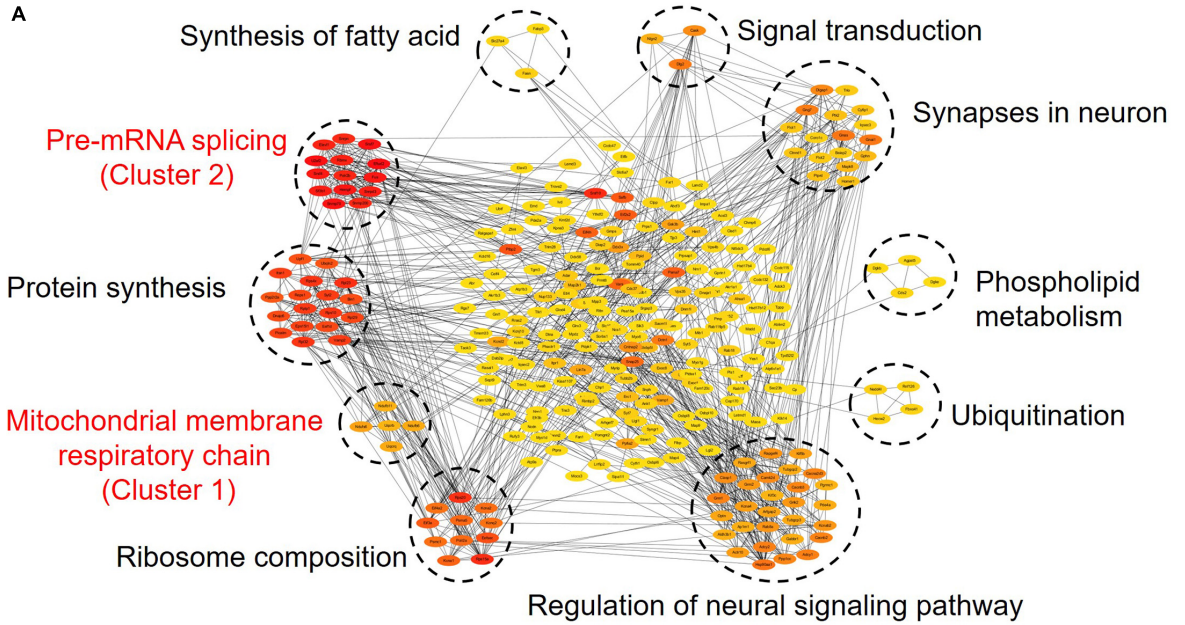
FIGURE 3 | RNA-Seq analysis of HT22 cells upon GAS5 KD revealed that the affected genes might participate in cell cycle. **(A)** Volcano plot for differentially expressed genes (DEGs) between control and GAS5 KD groups. **(B)** Heatmap of the 58 DEGs with the highest significance levels between control and GAS5 KD groups. Key of heatmap represented the $\log_{10}$(FPKM) of 58 DEGs. **(C)** Gene ontology enrichment analysis for "Biological Process" of upregulated DEGs in GAS5 KD group.

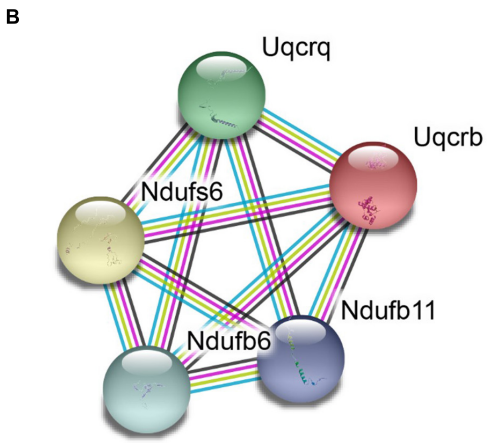
assembly (**Figure 3C** and **Supplementary Table 3**). A previous study reported that DNA replication plays a vital role in the S phase during the cell cycle, and it is highly correlated with the protein–DNA complex during nucleosome assembly (Krude and Keller, 2001). Chromatin assembly factor-1 also plays essential roles in nucleosome assembly, which is associated with DNA replication and cell proliferation, through interacting with polymerase sliding clamp proliferating cell nuclear antigen (Takami et al., 2007). In addition to that, cell cycle machinery also showed a strong link with the nucleosome assembly activity by chromatin assembly factor-1 during DNA replication (Krude and Keller, 2001). Consistently, our data suggested that GAS5 inhibits cell cycle progression, indicating that GAS5 affects the HT22 cell cycle by regulating the mRNA level of these DEGs.

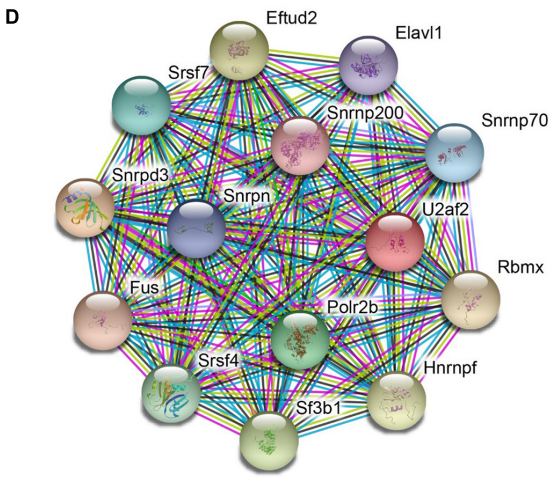
Growth Arrest-Specific Transcript 5 May Interact With Proteins Associated With Hippocampus Function in Young Mice

Growth arrest-specific transcript 5 was reported to play a very important role in tumors, and it can inhibit migration and invasion of gastric cancer *via* interacting with p53 protein (Liu et al., 2020), and it can also regulate the expression of p21 to enhance G1 cell cycle arrest through binding to Y-box binding protein 1 (Liu et al., 2015). To explore the role of GAS5 in the mouse hippocampus, we conducted an RNA pull-down assay in the mouse hippocampus, exploring its potential interacting proteins. Our data showed that GAS5 may interact with 290 proteins, which can form approximately 10 clusters, and they are involved in a variety of physiological processes in hippocampal brain tissue, indicating that GAS5 has multiple functions in the brain (**Figure 4A** and **Supplementary Table 4**), for example, the pre-mRNA splicing and extension of protein synthesis, mitochondrial membrane respiratory chain and ribosome composition, regulation of neural signaling pathways and ubiquitination, phospholipid metabolism and synapses in neurons, signal transduction, and synthesis of fatty acids.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis indicated that two clusters showed relatively strong interaction with GAS5, and they were taken into account for further analysis. In the first cluster, the proteins encoded by these five genes are mainly involved in the mitochondrial membrane respiratory chain (**Figure 4B**). KEGG pathway enrichment analysis showed that they are not only involved in oxidative phosphorylation but also participate in some important neurological diseases, such as Parkinson's disease and Alzheimer's disease (**Figure 4C**). We thus speculated that GAS5 might play a role in some neurological diseases. Proteins of the second cluster interact strongly with each other (**Figure 4D**). They mainly function in the shearing process of pre-mRNA (**Figure 4C**). Taken together, our results showed that GAS5 potentially regulates mitochondrial function, neurological diseases, and synapse function through interacting with specific protein clusters in the mouse brain hippocampus.







C

Cluster No.	KEGG term	Fold Enrichment	FDR	*P* value
Cluster 1	Oxidative phosphorylation	55.3	5.27E-05	1.02E-07
	Parkinson's disease	51.6	6.98E-05	1.35E-07
	Non-alcoholic fatty liver disease	48.9	8.62E-05	1.67E-07
	Alzheimer's disease	43.4	1.40E-04	2.71E-07
	Huntington's disease	38.8	2.20E-04	4.26E-07
Cluster 2	Spliceosome	42.1	2.73E-08	4.54E-11

FIGURE 4 | GAS5 RNA pull-down results revealed regulatory GAS5-associated protein–protein interaction network related to aging-associated diseases. **(A)** Overall protein–protein interaction network of GAS5 interacting proteins. Network nodes of different colors present proteins. Red color represents strong interaction, whereas yellow suggests a relatively weak connection. Lines between proteins represent protein–protein associations. **(B)** Interaction network diagram of proteins in first cluster, which were involved in the formation of mitochondrial membrane respiratory chain reaction. **(C)** KEGG pathway enrichment analysis of clusters 1 and 2. **(D)** Interaction network diagram of proteins in second cluster, which participated in the shearing process of pre-mRNA.

DISCUSSION

Growth arrest-specific transcript 5 was first discovered in 1988 by Schneider et al. (1988) in mouse NIH/3T3 fibroblast due to its increased expression after cell growth stagnation. As the research on GAS5 going further, more and more studies have shown that lncRNA GAS5 is low-expressed in most of the tumor samples, such as glioma, non-small cell lung cancer, breast carcinoma, etc. As a tumor suppressor, GAS5 is involved in the regulation of tumor occurrence and progression (Cao et al., 2017; Gu et al., 2018; Liu et al., 2018). However, it was not clear whether it plays a role in senescence or aging.

Our previous RNA-Seq data showed that GAS5 had differential expression patterns in young and old mice brains, which suggested that it may participate in brain aging (Wang et al., 2019). Further examination using RT-qPCR on four different brain tissues confirmed that GAS5 is expressed significantly higher in the old mouse brain hippocampus region comparing with that of a young mouse. To explore the role and underlying mechanism of GAS5 in senescence and brain aging, we performed a series of experiments on mice brain tissues and brain-derived cell lines. GAS5 showed nucleoplasm and cytoplasm subcellular localization.

Cellular senescence is an effective barrier to tumorigenesis *in vivo* (Braig et al., 2005; Collado et al., 2005; Michaloglou et al., 2005). GAS5 can inhibit bladder cancer cells proliferation and promote apoptosis (Chen et al., 2020) and also inhibits migration of gastric carcinoma cell *via* interacting with p53 protein (Liu et al., 2020). Our data showed that declined expression of GAS5 could activate cell cycle progression and reduce cell apoptosis level, and OE of GAS5 led to the opposite effect. p16 is a senescence-associated marker, the expression of which increases exponentially with aging in most tissues (He and Sharpless, 2017). p21 is a cyclin-dependent kinase inhibitor, which can also be used to measure the levels of aging (Wang et al., 2017). Upregulation of senescence-associated proteins p16 and p21 was found in granule cells of the dentate gyrus in the irradiation-induced mouse hippocampus, showing a similar senescence phenotype (Cheng et al., 2017). In our immunoblot result, the expression of p21 was upregulated upon GAS5 KD in HT22 cells and decreased in the presence of GAS5, indicating the antiaging effect of GAS5 in HT22 cells probably through the p21 pathway, although p16 protein expression did not change due to alteration of GAS5 RNA.

Growth arrest-specific transcript 5 can competitively bind Smad3 proteins through a variety of RNA SMAD-binding elements to inhibit transforming growth factor/Smad3-mediated smooth muscle cell differentiation (Tang et al., 2017), and subsequent experiments have found that GAS5 can inhibit the expression of adenosine triphosphate-binding cassette transporter A1 by binding to the enhancer of zeste homolog 2 (Meng et al., 2020). Our RNA pull-down assay of mouse hippocampus tissue identified a lot of new GAS5 interacting proteins. Mitochondrial DNA damage causes mitochondrial dysfunction and upregulation of reactive oxygen species, which may be induced by telomere dysfunction-caused cellular senescence (Passos et al., 2007). Mitochondrial dysfunction has been implicated in the pathogenesis of many neurodegenerative diseases, such as Parkinson's disease, Huntington's disease, and Alzheimer's disease (Aliev et al., 2004). It is also believed that higher oxidative status caused by mitochondrial dysfunction contributes to senescence acceleration and the age-dependent alterations in cell structure and function (Hosokawa, 2002, 2004). In our study, in addition to one cluster with the strongest interactions involved in pre-mRNA splicing and extension of protein synthesis, one cluster (Uqcrq, Uqcrb, Ndufb6, Ndufb11, and Ndufs6) is involved in mitochondrial membrane respiratory chain reaction. KEGG pathway enrichment analysis showed that they are involved in some important neurological diseases, such as Parkinson's disease, Alzheimer's disease, Huntington's disease, and so on. These data imply that GAS5 may have a potential role in brain aging through regulating the mitochondrial function in the hippocampus with specific mitochondrial proteins.

Growth arrest-specific transcript 5 has been reported to participate in several pathways such as apoptosis and cell cycle in T cell line and tumor suppression in many neoplasms. Our study reports the potential roles of GAS5 in senescence and brain aging. It may interfere with cell cycle, progression, and proliferation through interacting with various proteins and affecting mRNAs expression at brain tissue and brain-derived cell line levels. This study could provide a reference for further studies on molecular functions of GAS5 as well as its interacting proteins and affected mRNAs in aging and also contribute to better understand the multiple roles of non-coding RNAs in various biological systems.

MATERIALS AND METHODS

Animal Experiments

All the animal experiments performed in this study were conducted in adherence with guidelines of University of Science and Technology of China Animal Resources Center and University Animal Care and Use Committee. Male C57BL/6 mice were housed in a specific-pathogen-free and temperature-controlled room with a 12-h light/dark cycle and unrestricted access to food and water. Mouse experiment was approved by the Committee on the Ethics of Animal Experiments of the USTC (Permit Number: USTCACUC1801051).

Young and old adult male mice were anesthetized and transcardially perfused with 1 × phosphate-buffered saline (PBS). Brain tissues from the HT, OB, CB, and HC were carefully dissected under a stereomicroscope and then quickly immersed into TRIzol Reagent (Ambion, 15596018, Life Technology, United States) on ice, which will be stored in −80°C and used for RT-qPCR experiment.

Cell Culture

HT22 (hippocampal neuronal cell line) was routinely cultured in Dulbecco's modified Eagle medium (Gibco, 12100046, Thermo Fisher Scientific, United States) with 10% fetal bovine serum (ExCell Bio, FSS500, China), antibiotic–antimycotic (Wisent, 450-115-EL, Nanjing, China), and incubated in a humidified

atmosphere of 5% CO_2 and 95% air at 37°C. The cells were subcultured every 2 days.

Quantitative Reverse Transcription PCR

Total RNA from four brain regions (HT, OB, CB, and HC) were extracted using TRIzol Reagent (Ambion, 15596018, Life Technology, United States). RNA quality and quantity were checked by agarose gel electrophoresis and Nano-300 Micro Spectrophotometer (ALLSHENG, China). The first-strand cDNA was reverse transcribed using a HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, R212-02, Nanjing, China) according to the manufacturer's instructions. qPCR was then performed on a Bio-Rad CFX96 qPCR machine using the AceQ qPCR SYBR Green Master Mix (Vazyme, Q111-03, Nanjing, China). The expression levels of target genes were normalized to a housekeeping gene (*GAPDH*). Negative controls and technical replicates were also conducted. Primer lists are shown in **Supplementary Table 1**.

Fractionation of Cells

Cells with a density of 10^6 were rinsed three times with 1 × PBS and resuspended with RSB-100 solution [10-mM Tris–hydrochloric acid, pH = 7.4; 100-mM sodium chloride (NaCl); 25-mM magnesium chloride ($MgCl_2$); 40 μg/ml digitonin]. Gently mix the solution on ice and centrifuge at 2,000 × *g* for 8 min, 4°C. The supernatant was the cytoplasm part. Resuspend the pellets with 500 μl of RSB-100 solution containing 0.5% Triton X-100, incubated the samples on ice for 5 min, and then centrifuged at 2,000 × *g* for 8 min, 4°C. The supernatant was the nucleoplasm part. The left pellets, which was a chromatin-binding fraction, were dissolved using 500 μl of RSB-100 solution containing 0.5% Triton X-100. Total RNAs from three fractions were extracted using TRIzol Reagent, and RT-qPCR was conducted as mentioned above.

Growth Arrest-Specific Transcript 5 Knockdown and Overexpression

Small interfering RNA (siRNA) was used to knock down the expression of GAS5 non-coding RNA in HT22 cells and were synthesized by GenePharma (Shanghai, China). The forward and reverse sequences of GAS5-siRNA210 were 5′-CCU CUG UGA UGG GAC AUC UTT-3′ and 5′-AGA UGU CCC AUC ACA GAG GTT-3′ (Sun et al., 2017), respectively. The forward and reverse sequences of GAS5 siRNA596 were 5′-CCG GUC CUU CAU UCU GAA UTT-3′ and 5′-AUU CAG AAU GAA GGA CCG GTT-3′, respectively. The forward and reverse sequences of negative-control siRNA (NC siRNA) were 5′-UUC UCC GAA CGU GUC ACG UTT-3′ and 5′-ACG UGA CAC GUU CGG AGA ATT-3′, respectively. Cells were seeded in 12-well culture plates and transfected with siRNA when cells had reached 30 to 40% confluences. Then, cells were transfected with 10-nM GAS5 siRNA or NC siRNA using INTERFERin reagent (Polyplus, 409-10, France) at 1 of 250 final dilutions for 48 h.

cDNA of GAS5 isoform 210 was cloned into pcDNA3.1(+) (Invitrogen) OE vector. pcDNA3.1(+) empty vector was used as a negative control. Cells were seeded in 12-well culture plates and transfected with an OE vector when cells had reached 60 to 80% confluences. Then, cells were transfected with 2-μg GAS5-iso210 pcDNA3.1(+) or empty pcDNA3.1(+) using jetOPTIMUS reagent (Polyplus, 117-01, France) at 1 of 500 final dilutions for 48 h.

Cell Apoptosis Assay by Flow Cytometry

Cells were collected from 12-well plates at 48 h after transfection, washed with PBS, suspended with 100 μf binding buffer, and then, 5-μb annexin V–fluorescein isothiocyanate and 5-μa PI (Vazyme, A211-02, Nanjing, China) were added. The liquid was incubated in binding buffer at room temperature for 10 min, and then a 400-μl binding buffer was added to the tube. Finally, cell apoptosis was examined by flow cytometry (CytoFLEX, United States). Data were analyzed using the Flowjo VX. At least three independent experiments were conducted in the analysis.

Cell Cycle Analysis

Cells were collected from 12-well plates at 48 h after transfection, washed with 1 × PBS, fixed in 70% ethanol, and incubated at 4°C for 30 min. Fixed cells were washed twice and resuspended in PBS containing RNase A (100 μg/ml) and PI (50 μg/ml), incubated at 4°C for 30 min. At last, cell cycle analysis was performed using flow cytometry (CytoFLEX, United States), and then data were analyzed using the Modfit. At least three independent experiments were conducted in the analysis.

Western Blot Analysis

Cells were collected from 12-well plates at 48 h after transfection, washed with 1 × PBS, lysed in radioimmunoprecipitation assay buffer [50-mM Tris–hydrochloric acid, pH = 7.4, 150-mM NaCl, 1-mM ethylenediaminetetraacetic acid, 1% Triton X-100, 1% sodium dodecyl sulfate (SDS), 1% sodium deoxycholate supplemented with 1 × protease inhibitor cocktail, and 1-mM phenylmethylsulfonyl fluoride] at 4°C for 30 min. Protein supernatants were collected through a centrifuge at 12,000 rcf for 15 min at 4°C and mixed with protein loading buffer and denatured at 100°C for 10 min. Protein samples were separated through 12% SDS-polyacrylamide gel electrophoresis (PAGE) and transferred onto polyvinylidene difluoride membranes. After blocking non-specific binding with 3% bovine serum albumin (Sangon, China) in Tris–buffered saline with 0.05% Tween 20, the blots were incubated with specific antibodies. Commercial antibody used for blot includes p16 antibody (Abcam, ab108349), beta-actin antibody (Proteintech, 60008-1-Ig), and p21 antibodies (Millipore, 05-345). The secondary antibodies used for IB (1:5,000) were horseradish peroxidase-conjugated Affinipure Goat Anti-Mouse IgG(H + L) antibody (Proteintech, China), and the signals were detected with Supersensitive ECL chemiluminescence solution (P&Q Science Technology, Shanghai, China). The intensity of bands was assessed using Image J.

RNA Pull-Down

Yong mice (2.5 months old) were anesthetized and transcardially perfused with 1 × PBS. Hippocampus tissue from the brain was collected for homogenization on ice for 90 s and centrifuged (700 rcf) at 4°C for 5 min. After removing the supernatant,

the precipitates were resuspended in the lysis buffer containing 10 mmol/L 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, 10 mmol/L KCl, 1.5 mmol/L $MgCl_2$, and 1% NP-40 and incubated on ice for 1.5 h. During lysis, *in vitro* transcribed and biotinylated GAS5 sense and antisense RNA probes were heated in RNA structure buffer containing 10 mmol/L Tris–Cl (pH = 8.0), 0.1 mol/L KCl, 1 nmol/L, and 10 mmol/L $MgCl_2$ at 95°C for 2 min and was incubated on ice for 3 min. Then sense and antisense (negative control) GAS5 RNA probes were incubated with lysate from the previous discussion for 12 h at 4°C with rotation. Next, 20-μl M-280 Streptavidin magnetic beads (Invitrogen, 11205D, Thermo Fisher Scientific, United States) were added into the mixture mentioned earlier and incubated for 3 h at 4°C with rotation. Beads were washed at least three times, and RNA–protein complex was collected through a magnetic base. Beads were resuspended in elution buffer containing 7.5 mmol/L 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, 15 mmol/L ethylenediaminetetraacetic acid, 0.15% SDS, 75 mmol/L NaCl, 0.02% sodium deoxycholate. Finally, the eluted proteins were separated by SDS-PAGE and then identified by mass spectrometry (MS). All processes were performed under RNase-free conditions.

Mass Spectrometry

The proteins that were separated by SDS-PAGE were reduced using 20-mM dithiothreitol (Sigma) at 95°C for 5 min and alkylated in 50-mM iodoacetamide (Sigma) for 30 min (in the dark). Then, the samples were digested into peptides using sequencing grade trypsin (Promega) and recovered from the SDS-PAGE gel. Peptides were desalted using StageTips and resuspended using 0.1% formic acid with 2% acetonitrile. Mass spectrometry experiments were performed on a nanoscale ultra-high-pressure liquid chromatography system (EASY-nLC1200, Thermo Fisher Scientific) connected to an Orbitrap Fusion Lumos equipped with a nanoelectrospray source (Thermo Fisher Scientific). The peptides were separated on a reversed-phase high-performance liquid chromatography analytical column (75 μm × 25 cm) packed with 2-μm C18 beads (Thermo Fisher Scientific) using a linear gradient ranging from 9 to 28% acetonitrile for 90 min and followed by a linear increase to 45% B for 20 min at a flow rate of 300 nl/min. The Orbitrap Fusion Lumos acquired data in a data-dependent manner alternating between full-scan MS and MS2 scans. The MS spectra (350–1,800 m/z) were collected with 120,000 resolution, automatic gain control of 4×10^5, and 50-ms maximal injection time. Selected ions were sequentially fragmented in a 3-s cycle by higher-energy C-trap dissociation with 30% normalized collision energy, specified isolated windows 1.6 m/z, 30,000 resolution. Automatic gain control of 5×10^4 and 100-ms maximal injection time were used. Dynamic exclusion was set to 30 s. Raw data were processed using Proteome Discoverer (PD, version 2.2), and MS/MS spectra were searched against the reviewed Swiss-Prot mouse proteome database. The cutoff values are shown later: a precursor mass tolerance of 10 ppm; fragment mass tolerance of 0.02 Da; only peptides longer than six amino acids were kept; false discovery rate <1%; and at least one unique peptide was required for protein identification. The MS data have been submitted to ProteomeXchange and available *via* ProteomeXchange with identifier PXD023363.

RNA Sequencing

The mRNA-Seq library was prepared according to the manufacturer's protocol. Reagents used for mRNA-seq library preparation were from New England BioLabs Company. Briefly, total RNAs were extracted by TRIzol reagent. RNA quality and concentration were checked by agarose gel electrophoresis and Nanodrop. RNA integrity was validated by Agilent 2100. The mRNAs were purified by oligo-dT attached beads and fragmented into small pieces. Cleaved RNA fragments were reverse-transcribed into first-strand cDNAs. DNA polymerase I was used to synthesizing second-strand cDNAs. End repair and ligation of adapters were performed to add a single "A" base and adapters to cDNA fragments. Purified cDNA fragments were amplified through PCR. The final cDNA library was then sequenced by high throughput sequencing platform Illumina Novaseq 6000. The RNA-Seq data have been deposited in the Gene Expression Omnibus (accession number: GSE163237).

Bioinformatics Analysis

The RNA-Seq reads were initially subject to adapter removal and filtered using Trimmomatic (Bolger et al., 2014) (version 0.39). Then, we use FastQC (version 0.11.8) to do some quality control checks. Clean reads were mapped using HISAT2 (Kim et al., 2015) (version 2.1.0) to the GRCM38 Ensembl genome. Sort and convert the SAM files to BAM files using SAMtools (Li et al., 2009) (version 1.7). The mapped reads were then counted using featureCounts (Liao et al., 2014) (version 2.0.1) with Mus_musculus.GRCm38.101.gtf annotations[1]. DEGs were determined using R package DESeq2 (Love et al., 2014) (version 1.26.0) with a P-value < 0.05 and an absolute value of $\log_2 FC > 1$. Pearson correlation analysis was conducted in R (version 3.6.3). Volcano map and heatmap generation were conducted, respectively, using R packages "ggplot2" (version 3.3.2) and "pheatmap" [Kolde R, Kolde MR. Package "pheatmap"[2] (Oct 12; 2015)] (version 1.0.12). The interaction network was analyzed in STRING (Szklarczyk et al., 2019)[3] and visualized in Cytoscape (Otasek et al., 2019) (version 3.8.0). We used DAVID (Huang et al., 2009)[4] to perform functional annotation clustering with the DEGs.

Statistical Analysis

Each experiment was repeated two or three times. Student's t-tests analyzed statistical differences. Data are presented as the mean ± standard errors of the means. The statistical significance was established at 0.05 ($^*P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$) using SPSS® software V.21 (Chicago, IL, United States). GraphPad Prism 7.0 was used for statistical illustrations.

[1] http://asia.ensembl.org/index.html

[2] https://cran.r-project.org/web/packages/pheatmap/

[3] https://string-db.org/

[4] https://david.ncifcrf.gov/

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by the University of Science and Technology of China Animal Resources Center and University Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

XS and HM conceived the experiment. SW and SK performed the gene KD and OE part. SW, HM, and BL performed the RNA pull-down experiment. Y-HH and WL conducted the protein MS part. YW conducted the bioinformatic analyses. HM, SW, SK, and DZ finished RT-qPCR and cellular localization. DZ and BL took care of the animals. All authors read, wrote, and approved the article.

FUNDING

This work was supported by the National Key Scientific Program of China (2016YFA0100502 to XS), National Natural Science Foundation of China (31671490 to XS), Anhui Provincial Natural Science Foundation (Grant number 1908085QC102 to HM), and Fundamental Research Funds for the Central Universities (WK2070000133 and WK2070000116 to HM).

ACKNOWLEDGMENTS

We thank Prof. Cheng He (Second Military Medical University, China) for providing us with plasmids and primers, Prof. Jiangning Zhou (University of Science and Technology of China, China) for providing the HT22 cell line, Prof. Qiangfeng Cliff Zhang and Tuanlin Xiong (Tsinghua University, China) for valuable advice on GAS5 structure, and Leilei Shi and Qianqian Wang (University of Science and Technology of China, China) for their kind help in RNA pull-down experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.615340/full#supplementary-material

Supplementary Table 1 | The sequence of RT-qPCR primer for mouse GAS5.

Supplementary Table 2 | The expression level, $\log_2$FoldChange, and adjusted *P*-values of 58 DEGs.

Supplementary Table 3 | The detailed information of GO terms showed in **Figure 3C**.

Supplementary Table 4 | The detailed information of identified proteins in RNA pull-down experiment.

REFERENCES

Abdelmohsen, K., Panda, A., Kang, M. J., Xu, J., Selimyan, R., Yoon, J. H., et al. (2013). Senescence-associated lncRNAs: senescence-associated long noncoding RNAs. *Aging Cell* 12, 890–900. doi: 10.1111/acel.12115

Aliev, G., Smith, M. A., de la Torre, J. C., and Perry, G. (2004). Mitochondria as a primary target for vascular hypoperfusion and oxidative stress in Alzheimer's disease. *Mitochondrion* 4, 649–663. doi: 10.1016/j.mito.2004.07.018

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Braig, M., Lee, S., Loddenkemper, C., Rudolph, C., Peters, A. H. F. M., Schlegelberger, B., et al. (2005). Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature* 436, 660–665. doi: 10.1038/nature03841

Bunch, H., Choe, H., Kim, J., Jo, D. S., Jeon, S., Lee, S., et al. (2019). P-TEFb Regulates Transcriptional Activation in Non-coding RNA Genes. *Front. Genet.* 10:342. doi: 10.3389/fgene.2019.00342

Calcinotto, A., Kohli, J., Zagato, E., Pellegrini, L., Demaria, M., and Alimonti, A. (2019). Cellular Senescence: Aging, Cancer, and Injury. *Physiol. Rev.* 99, 1047–1078. doi: 10.1152/physrev.00020.2018

Cao, L., Chen, J., Ou, B. Q., Liu, C. Z., Zou, Y., and Chen, Q. (2017). GAS5 knockdown reduces the chemo-sensitivity of non-small cell lung cancer (NSCLC) cell to cisplatin (DDP) through regulating miR-21/PTEN axis. *Biomed. Pharmacother.* 93, 570–579. doi: 10.1016/j.biopha.2017.06.089

Chen, D., Guo, Y. H., Chen, Y. Q., Guo, Q. N., Chen, J. Y., Li, Y. N., et al. (2020). LncRNA growth arrest-specific transcript 5 targets miR-21 gene and regulates bladder cancer cell proliferation and apoptosis through PTEN. *Cancer Med.* 9, 2846–2858. doi: 10.1002/cam4.2664

Cheng, Z., Zheng, Y. Z., Li, Y. Q., and Wong, C. S. (2017). Cellular Senescence in Mouse Hippocampus After Irradiation and the Role of p53 and p21. *J. Neuropathol. Exp. Neurol.* 76, 260–269. doi: 10.1093/jnen/nlx006

Collado, M., Gil, J., Efeyan, A., Guerra, C., Schuhmacher, A. J., Barradas, M., et al. (2005). Tumour biology - Senescence in premalignant tumours. *Nature* 436, 642–642. doi: 10.1038/436642a

Dong, S. J., Zhang, X. F., and Liu, D. C. (2019). Overexpression of long noncoding RNA GAS5 suppresses tumorigenesis and development of gastric cancer by sponging miR-106a-5p through the Akt/mTOR pathway. *Biol. Open* 8:bio041343. doi: 10.1242/bio.041343

Goustin, A. S., Thepsuwan, P., Kosir, M. A., and Lipovich, L. J. N. C. R. (2019). The growth-arrest-specific (gas)-5 long non-coding RNA: a fascinating lncRNA widely expressed in cancers. *Noncoding RNA* 5:46. doi: 10.3390/ncrna5030046

Gu, J., Wang, Y., Wang, X., Zhou, D., Wang, X., Zhou, M., et al. (2018). Effect of the LncRNA GAS5-MiR-23a-ATG3 Axis in Regulating Autophagy in Patients with Breast Cancer. *Cell Physiol. Biochem.* 48, 194–207. doi: 10.1159/000491718

Gutschner, T., Hämmerle, M., and Diederichs, S. (2013). MALAT1 — a paradigm for long noncoding RNA function in cancer. *J. Mole. Med.* 91, 791–801. doi: 10.1007/s00109-013-1028-y

He, S., and Sharpless, N. E. (2017). Senescence in Health and Disease. *Cell* 169, 1000–1011. doi: 10.1016/j.cell.2017.05.015

Hosokawa, M. (2002). A higher oxidative status accelerates senescence and aggravates age-dependent disorders in SAMP strains of mice. *Mech. Ageing Dev.* 123, 1553–1561. doi: 10.1016/S0047-6374(02)00091-X

Hosokawa, M. (2004). *Mitochondrial dysfunction and an impaired response to higher oxidative status accelerate cellular aging in SAMP strains of mice. Senescence-Accelerated Mouse (Sam): An Animal Model of Senescence*. Amsterdam: Elsevier. 1260, 47–52. doi: 10.1016/S0531-5131(03)01729-1

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* 4, 44–57. doi: 10.1038/nprot.2008.211

Huang, Y. (2018). The novel regulatory role of lncRNA-miRNA-mRNA axis in cardiovascular diseases. *J. Cell Mol. Med.* 22, 5768–5775. doi: 10.1111/jcmm.13866

Kim, D., Landmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–U121. doi: 10.1038/nmeth.3317

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010). Noncoding RNA Gas5 Is a Growth Arrest- and Starvation-Associated Repressor of the Glucocorticoid Receptor. *Sci. Signal.* 3:ra8. doi: 10.1126/scisignal.2000568

Krude, T., and Keller, C. (2001). Chromatin assembly during S phase: contributions from histone deposition, DNA replication and the cell division cycle. *Cell Mol. Life Sci.* 58, 665–672. doi: 10.1007/pl00000890

Kumar, A., and Foster, T. C. (2019). Alteration in NMDA Receptor Mediated Glutamatergic Neurotransmission in the Hippocampus During Senescence. *Neurochem. Res.* 44, 38–48. doi: 10.1007/s11064-018-2634-4

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, L., van Breugel, P. C., Loayza-Puch, F., Ugalde, A. P., Korkmaz, G., Messika-Gold, N., et al. (2018). LncRNA-OIS1 regulates DPP4 activation to modulate senescence induced by RAS. *Nucleic Acids Res.* 46, 4213–4227. doi: 10.1093/nar/gky087

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, Q., Yu, W., Zhu, S., Cheng, K., Xu, H., Lv, Y., et al. (2018). Long noncoding RNA GAS5 regulates the proliferation, migration, and invasion of glioma cells by negatively regulating miR-18a-5p. *J. Cell Physiol.* 234, 757–768. doi: 10.1002/jcp.26889

Liu, Y., Yin, L., Chen, C., Zhang, X., and Wang, S. (2020). Long non-coding RNA GAS5 inhibits migration and invasion in gastric cancer via interacting with p53 protein. *Dig. Liver Dis.* 52, 331–338. doi: 10.1016/j.dld.2019.08.012

Liu, Y., Zhao, J., Zhang, W., Gan, J., Hu, C., Huang, G., et al. (2015). lncRNA GAS5 enhances G1 cell cycle arrest via binding to YBX1 to regulate p21 expression in stomach cancer. *Sci. Rep.* 5:10159. doi: 10.1038/srep10159

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

Meng, X. D., Yao, H. H., Wang, L. M., Yu, M., Shi, S., Yuan, Z. X., et al. (2020). Knockdown of GAS5 Inhibits Atherosclerosis Progression via Reducing EZH2-Mediated ABCA1 Transcription in ApoE(-/-) Mice. *Mol. Ther. Nucleic Acids* 19, 84–96. doi: 10.1016/j.omtn.2019.10.034

Michaloglou, C., Vredeveld, L. C., Soengas, M. S., Denoyelle, C., Kuilman, T., van der Horst, C. M., et al. (2005). BRAFE600-associated senescence-like cell cycle arrest of human naevi. *Nature* 436, 720–724. doi: 10.1038/nature03890

Otasek, D., Morris, J. H., Boucas, J., Pico, A. R., and Demchak, B. (2019). Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* 20:185. doi: 10.1186/s13059-019-1758-4

Passos, J. F., Saretzki, G., and von Zglinicki, T. (2007). DNA damage in telomeres and mitochondria during cellular senescence: is there a connection? *Nucleic Acids Res.* 35, 7505–7513. doi: 10.1093/nar/gkm893

Payre, F., and Desplan, C. (2016). RNA. Small peptides control heart activity. *Science* 351, 226–227. doi: 10.1126/science.aad9873

Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* 36, 233–278. doi: 10.1146/annurev.genet.36.042902.092433

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

Raz, V., Vermolen, B. J., Garini, Y., Onderwater, J. J., Mommaas-Kienhuis, M. A., Koster, A. J., et al. (2008). The nuclear lamina promotes telomere aggregation and centromere peripheral localization during senescence of human mesenchymal stem cells. *J. Cell Sci.* 121(Pt 24), 4018–4028. doi: 10.1242/jcs.034876

Sang, B., Zhang, Y. Y., Guo, S. T., Kong, L. F., Cheng, Q., Liu, G. Z., et al. (2018). Dual functions for OVAAL in initiation of RAF/MEK/ERK prosurvival signals and evasion of p27-mediated cellular senescence. *Proc. Natl. Acad. Sci. U S A* 115, E11661–E11670. doi: 10.1073/pnas.1805950115

Schneider, C., King, R. M., and Philipson, L. (1988). Genes Specifically Expressed at Growth Arrest of Mammalian-Cells. *Cell* 54, 787–793. doi: 10.1016/S0092-8674(88)91065-3

Sun, D., Yu, Z., Fang, X., Liu, M., Pu, Y., Shao, Q., et al. (2017). LncRNA GAS5 inhibits microglial M2 polarization and exacerbates demyelination. *EMBO Rep.* 18, 1801–1816. doi: 10.15252/embr.201643668

Sun, X. D., Thorne, R. F., Zhang, X. D., He, M., Li, J. M., Feng, S. S., et al. (2020). LncRNA GUARDIN suppresses cellular senescence through a LRP130-PGC1 alpha-FOXO4-p21-dependent signaling axis. *Embo Rep.* 21:201948796. doi: 10.15252/embr.201948796

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Takami, Y., Ono, T., Fukagawa, T., Shibahara, K., and Nakayama, T. (2007). Essential role of chromatin assembly factor-1-mediated rapid nucleosome assembly for DNA replication and cell division in vertebrate cells. *Mol. Biol. Cell* 18, 129–141. doi: 10.1091/mbc.e06-05-0426

Tan, P., Guo, Y. H., Zhan, J. K., Long, L. M., Xu, M. L., Ye, L., et al. (2019). LncRNA-ANRIL inhibits cell senescence of vascular smooth muscle cells by regulating miR-181a/Sirt1. *Biochem. Cell Biol.* 97, 571–580. doi: 10.1139/bcb-2018-0126

Tang, R., Zhang, G., Wang, Y. C., Mei, X., and Chen, S. Y. J. J. O. B. C. (2017). The long non-coding RNA GAS5 regulates TGF-β-induced smooth muscle cell differentiation via RNA-Smad binding element. *J. Biol. Chem.* 292:jbc.M117.790030.

Tani, H., Torimura, M., and Akimitsu, N. (2013). The RNA Degradation Pathway Regulates the Function of GAS5 a Non-Coding RNA in Mammalian Cells. *PLoS One* 8:e55684.

Wang, F., Ren, D., Liang, X., Ke, S., Zhang, B., Hu, B., et al. (2019). A long noncoding RNA cluster-based genomic locus maintains proper development and visual function. *Nucleic Acids Res.* 47, 6315–6329. doi: 10.1093/nar/gkz444

Wang, M. J., Chen, F., Lau, J. T. Y., and Hu, Y. P. (2017). Hepatocyte polyploidization and its association with pathophysiological processes. *Cell Death Dis.* 8:e2805. doi: 10.1038/cddis.2017.167

Wei, W., and Ji, S. (2018). Cellular senescence: Molecular mechanisms and pathogenicity. *J. Cell Physiol.* 233, 9121–9135. doi: 10.1002/jcp.26956

Wu, Y., Zhang, A. Q., and Yew, D. T. (2005). Age related changes of various markers of astrocytes in senescence-accelerated mice hippocampus. *Neurochem. Int.* 46, 565–574. doi: 10.1016/j.neuint.2005.01.002

Xu, C. L., Sang, B., Liu, G. Z., Li, J. M., Zhang, X. D., Liu, L. X., et al. (2020). SENEBLOC, a long non-coding RNA suppresses senescence via p53-dependent and independent mechanisms. *Nucleic Acids Res.* 48, 3089–3102. doi: 10.1093/nar/gkaa063

Zhou, R. R., Li, H. B., You, Q. S., Rong, R., You, M. L., Xiong, K., et al. (2019). Silencing of GAS5 Alleviates Glaucoma in Rat Models by Reducing Retinal Ganglion Cell Apoptosis. *Human Gene Ther.* 30, 1505–1519. doi: 10.1089/hum.2019.056

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

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 30 March 2021
doi: 10.3389/fgene.2021.642602

Comprehensive Identification and Alternative Splicing of Microexons in *Drosophila*

*Ting-Lin Pang[1,2,3], Zhan Ding[1,2,3], Shao-Bo Liang[3], Liang Li[1,2,3], Bei Zhang[1,2], Yu Zhang[1,2], Yu-Jie Fan[3] and Yong-Zhen Xu[3]**

[1] University of Chinese Academy of Sciences, Beijing, China, [2] Key Laboratory of Insect Developmental and Evolutionary Biology, Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China, [3] RNA Institute, State Key Laboratory of Virology, Hubei Key Laboratory of Cell Homeostasis, College of Life Science, Wuhan University, Wuhan, China

Interrupted exons in the pre-mRNA transcripts are ligated together through RNA splicing, which plays a critical role in the regulation of gene expression. Exons with a length $\leq$ 30 nt are defined as microexons that are unique in identification. However, microexons, especially those shorter than 8 nt, have not been well studied in many organisms due to difficulties in mapping short segments from sequencing reads. Here, we analyzed mRNA-seq data from a variety of *Drosophila* samples with a newly developed bioinformatic tool, ce-TopHat. In addition to the Flybase annotated, 465 new microexons were identified. Differentially alternatively spliced (AS) microexons were investigated between the *Drosophila* tissues (head, body, and gonad) and genders. Most of the AS microexons were found in the head and two AS microexons were identified in the sex-determination pathway gene *fruitless*.

Keywords: microexon, *Drosophila*, alternative splicing, *ce-Tophat*, sex determination

OPEN ACCESS

Edited by:
Yuanchao Xue,
Institute of Biophysics (CAS), China

Reviewed by:
Haihong Shen,
Gwangju Institute of Science and Technology, South Korea
Yang Yu,
Chinese Academy of Sciences, China

****Correspondence:***
Yong-Zhen Xu
Yongzhen.Xu@whu.edu.cn

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *16 December 2020*
Accepted: *05 March 2021*
Published: *30 March 2021*

Citation:
Pang T-L, Ding Z, Liang S-B, Li L, Zhang B, Zhang Y, Fan Y-J and Xu Y-Z (2021) Comprehensive Identification and Alternative Splicing of Microexons in Drosophila. Front. Genet. 12:642602. doi: 10.3389/fgene.2021.642602

INTRODUCTION

Ribonucleic acid splicing removes intronic sequences and ligates exonic sequences in eukaryotic cells. This is catalyzed by the spliceosome, a large and dynamic RNA–protein complex composed of five small nuclear RNAs (U1, U2, U4, U5, and U6 snRNAs) and over 100 proteins (Will and Luhrmann, 2011). On the Ensembl website, 595,500 and 83,529 exons can be retrieved from human and *Drosophila* transcriptomes, respectively. The length of exons varies from two to 11,555 nts (Sakharkar et al., 2004), usually longer in the lower eukaryotes than in the higher eukaryotes (Berget, 1995).

Definition of microexons, the range of microexon length, varies in different reports, such as $\leq$ 51 nt (Volfovsky et al., 2003; Irimia et al., 2014; Yang and Chen, 2014), 3–27 nt (Li et al., 2015), and 3–30 nt (Ustianenko et al., 2017). Here, we defined exons with length $\leq$ 30 nt are microexons. It was first reported that two 5-nt exons were found in the *Drosophila Ubx* gene, and a 6-nt exon was in the chicken cardiac troponin T (*TnT*) gene (Beachy et al., 1985; Cooper and Ordahl, 1985);

and soon after a 30-nt and a 3-nt exon were identified from the rat and mouse *Ncam* genes, respectively (Small et al., 1988; Santoni et al., 1989). In a mammal, microexons play a crucial role in the development and maintenance of neuronal functions (Irimia et al., 2014; Li et al., 2015), and there are 15,688 microexons in humans according to transcriptome annotation (version: GRCH38.96).

Early annotation tools had a poor ability to find microexons until a systematical alignment between cDNA and genomic DNA sequences identified 319 microexons from four genomes (Volfovsky et al., 2003). In the past decade, many tools have been developed to predict microexons, including the commonly used TopHat (Trapnell et al., 2009) and HISAT (Kim et al., 2015) and customized software GMAP (Wu and Watanabe, 2005), Olego (Wu et al., 2013), VAST-TOOLS (Irimia et al., 2014), and ATMap (Li et al., 2015). However, there are still limitations in the prediction of microexons. First, due to their alignment principles, TopHat and HISAT have difficulty accurately aligning and predicting the 3–7 nt very short microexons. Second, although Olego, VAST-TOOLS, and ATMap can partly predict the 3–7 nt microexons, their principles ignore the ends of sequencing reads that could not be mapped but may contribute to microexons identification. Third, most of these tools were originally developed based on mammalian transcriptomes, which could undervalue microexons in other organisms.

Alternative splicing plays critical roles in the regulation of gen = −0987 \e expression, generating multiple RNA isoforms from one gene and thereby extended proteome in eukaryotes. More than 95% of genes are alternatively spliced (Pan et al., 2008). The first reported alternatively spliced (AS) microexon is a 9-nt exon in the *Drosophila fasciclin I*, resulting in altered binding specificity of Fasciclin I (McAllister et al., 1992). Another interesting early example is a 3-nt microexon in the *Drosophila TnT* gene, which is specifically skipped in the larva's subcutaneous muscular system, resulting in a protein with a missed lysine residue (Benoist et al., 1998). Later studies in mice have revealed that spatial structures and the domains of proteins were changed due to AS of microexons, such as splicing of microexon 20 in the *ITSN1* gene (Tsyba et al., 2008) and microexon L in the *Protrudin* gene (Ohnishi et al., 2014). AS of microexons could also introduce a phosphorylation site that alters the original protein function. For example, the microexon E8a in the *LSD1* gene encodes a phosphorylated site, Thr369b (Toffolo et al., 2014). On the other hand, retained microexons may introduce a premature stop codon, leading the transcript into an NMD-mediated degradation pathway (Li et al., 2015).

Systematic studies of microexons are mainly related to the mammalian neural systems, and little is known about the development of other organisms. To extensively identify microexons and their alternative splicing, we used in-house RNA deep sequencing data from a variety of *Drosophila* samples and developed an improved computational model ce-TopHat. In total, we identified 985 reliable microexons, of which 465 are novel. Subsequent analyses revealed that the microexons in coding regions are more prone to be alternatively spliced than exons with length > 30 nt. Differentially AS microexons have been investigated between tissues and genders. Over 65% of the AS microexons are found in the head, and two AS microexons in the sex-determination pathway gene *fruitless* are investigated.

MATERIALS AND METHODS

Fly Strain, Culture, and Samples

Stocks and crosses of the wild-type w^{1118} isogenic *Drosophila* strain (BDSC 5905) were maintained and cultured on standard cornmeal agar medium (Qiu et al., 2019). Embryos of 18 h wandering third instar larvae and adults of 24 h were collected, respectively. Heads and abdomen-thorax (bodies) of adults were dismembered using liquid nitrogen; the gonads were dissected in ice-cold PBS from fresh flies and collected, respectively, for further RNA extraction.

RNA-Seq

Ribonucleic acid samples were prepared as described (Li et al., 2020), and the construction of cDNA libraries and sequencing were performed using Illumina Hi-Seq 2000 (Stark et al., 2019). The head, body, and gonads of the female and male *Drosophila* were sequenced in the form of 150 bp fr-firststrand pair-end reads, and the embryo, 3L larvae, and adults were sequenced in the form of 100 bp fr-untstranded pair-end reads. Most of these sequencing data are provided by members of Xu Lab from their unpublished projects.

Reads Mapping and Remapping

All raw data were trimmed and filtered by FastQC v0.11.5 and cutadapt v1.15, and the clean reads were then mapped to the *Drosophila* genome (dmel_r6.15 version) using TopHat (v2.1.1), in which 2-nt mismatches were allowed with other default parameters (**Supplementary Figure 1A**).

Unmapped reads were then remapped by an improved computational model, ce-TopHat (cut-end TopHat). First, the ce-TopHat removed 3–7 nt from the two ends of each unmapped read, respectively. These new reads were defined as the cut-end reads, which were remapped to the genome using TopHat (mismatch = 0 and anchor ≥ 8). Second, for the successfully remapped cut-end reads, ce-TopHat will find their genomic locations, which must meet three conditions: (1) mismatch = 0, (2) the cut-end part is flanked by classical splicing site sequences AG/____/GT, and (3) it is located within the gene range and the gap is between 20 and 10,000 nt. The finally met cut-end parts were then recombined to form complete and unique genomic localized reads for the construction of transcripts.

Identification of Microexons

After mapping and remapping, all junction reads were retrieved from the reconstructed transcriptome and divided into two groups: multi-junction reads and single-junction reads (**Supplementary Figure 1B**). These junction reads, if containing a ≤30 nt junction part, would have a potential sequence of microexon. If the middle junction part in a multi-junction read is ≤30 nt and not annotated as an exon, it will be defined as a novel microexon. For the two ends of the single-junction reads,

if it is ≤30 nt but cannot be correctly mapped to an annotated exon, a further extension is performed by other sequencing reads.

Determination of Reliable Microexons

The microexons predicted by TopHat or ce-TopHat were defined as microexon candidates. To improve the reliability, the following screening processes were used (**Supplementary Figure 1C**): (1) unique exon junction reads ≥2; (2) total exon junction reads ≥3; (3) at least one multi-junction reads support; and (4) the length of both sides of the novel multi-junction reads must be ≥8 nt. If the candidates of 8–30 nt in length satisfied the first two conditions or the candidates of 3–7 nt in length satisfied all the four conditions, they were defined as reliable microexons.

RT-PCR

Total RNAs from *Drosophila* samples were isolated by TRIzol (Ambion) and treated with RNase-free DNase I (Invitrogen). For RT-PCR, cDNAs were reverse transcribed using RevertAid Reverse Transcriptase (Thermo) and amplified by Ex-Taq (TaKaRa). All used primers are listed in **Supplementary Table 1**.

Alternative Splicing Analysis

Alternative splicing events were determined by the reconstructed *Drosophila* transcriptome as described above. The PSI and Differential PSI (ΔPSI) of each alternative splicing event were calculated by the Perl scripts, in which events with a difference >0.05 were considered as significantly different between samples (Shen et al., 2014; Hartley and Mullikin, 2016; Wu et al., 2018). In the "exon-microexon-exon" model, microexons with 1 > PSI > 0 were defined as alternatively spliced (AS) microexons, and microexons with PSI = 1 were defined as constitutively spiced (CS) microexons. For comparison, longer exons (> 30 nt) were retrieved and analyzed using the same procedure.

Splice Sites and GO Analyses

Consensus and conservation of splice sites were analyzed by Weblogo as described (Crooks et al., 2004). Briefly, the 5′SS and 3′SS were extracted by Perl, in which the 5′SS sequences contained the last 3 nt in the upstream exon and the first 6 nt in the intron, and the 3′SS sequences contained the last 20 nt in the intron and the first 3 nt in the downstream exon. Those sequences were also analyzed for the strength of splicing sites (binding ability to the spliceosome) by MaxEntScan online software as described (Yeo and Burge, 2004).

The Gene Ontology (GO) enrichment analysis on microexon genes is calculated by the clusterprofiler (Yu et al., 2012) and visualized with an R language package.

RESULTS

Identification of Microexons Using Multiple RNA-Seq Data by TopHat

To comprehensively identify microexons in *Drosophila*, we collected a variety of our lab's published and unpublished RNA-seq data of the WT fruit fly, including the head, body, and gonads from both females and males, as well as the embryo (18 h), 3L larva (wandering stage) and adults (2 h). In total, we obtained > 1.3 billion reads (∼430 G data) for subsequent analysis (**Supplementary Table 2**). Using the common alignment software TopHat and after performing the microexon identification and filtering process (**Supplementary Figure 1**), we identified 37,083 microexon candidates in *Drosophila*, of which 881 are highly reliable and 361 are novel microexons that were not annotated by the Flybase (**Figure 1A**). The length distribution of those reliable microexons, with and without novel ones, is obviously higher at the integral multiple of 3 nt than their nearby non-integrals (**Figure 1B**). To validate, we selected 12 novel microexons and performed RT-PCRs followed by individual Sanger sequencings; all of them showed proper bands and right sequences (**Figure 1C**), demonstrating that the above bioinformatic identifications are reliable.

ce-TopHat, a Modified TopHat, Identifies More Microexons

Due to difficulties in mapping short segments to unique genomic locations, the above analyses did not find novel microexons that are shorter than 8 nt. To identify more microexons, especially in the lengths of 3–7 nts, we improved TopHat and developed a new computing model, named ce-TopHat (**Figure 2A** and Materials and Methods section). Briefly, ce-TopHat focused on mapping the 3–7 nt segments from the TopHat-unmapped reads to the *Drosophila* genome. There were two groups of unmapped reads with short segments according to their locations: (i) at the two ends and (ii) in the middle of the unmapped reads.

For the first group, each unmapped read could generate as many as 10 derivative reads by cutting 3, 4, 5, 6, and 7 nts at its 5′- and 3′-ends. Those cutting-end reads that could be remapped to the fly genome were collected, and sequences of short segments at their ends were used for searching locations in the nearby introns, sometimes in exons. We defined those short segments uniquely located in the genome and with flanking splice sites (AG/___/GT) as microexon candidates (**Figure 2A** right). After three rounds of remapping and transcriptome reconstruction, the recovered unmapped reads were decreased to near zero (**Figure 2B** and **Supplementary Figures S2A,B**), and in total, we remapped 7,369,109 reads from all the *Drosophila* RNA-seq samples (**Supplementary Table 3**). For the second group of unmapped reads, the 3–7 nt segments in the middle of remapped reads were also defined as microexon candidates if they had unique locations in the genome and flanking AG/___/GT sequences (**Figure 2A** left).

Remapped reads by ce-TopHat extended the fly transcriptome information and allowed us to identify an additional 104 highly reliable microexons (**Figure 2C** and **Supplementary Table 4**), of which 19 were 3–7 nts in length. The pattern of length distribution was not changed after adding those new microexons by ce-TopHat (**Supplementary Figure 2C**). Similarly, we then selected six microexons with a length of 3–7 nts for validation. Due to their short lengths, two sets of PCR primers were used for validation of each microexon. One set was primers both located in

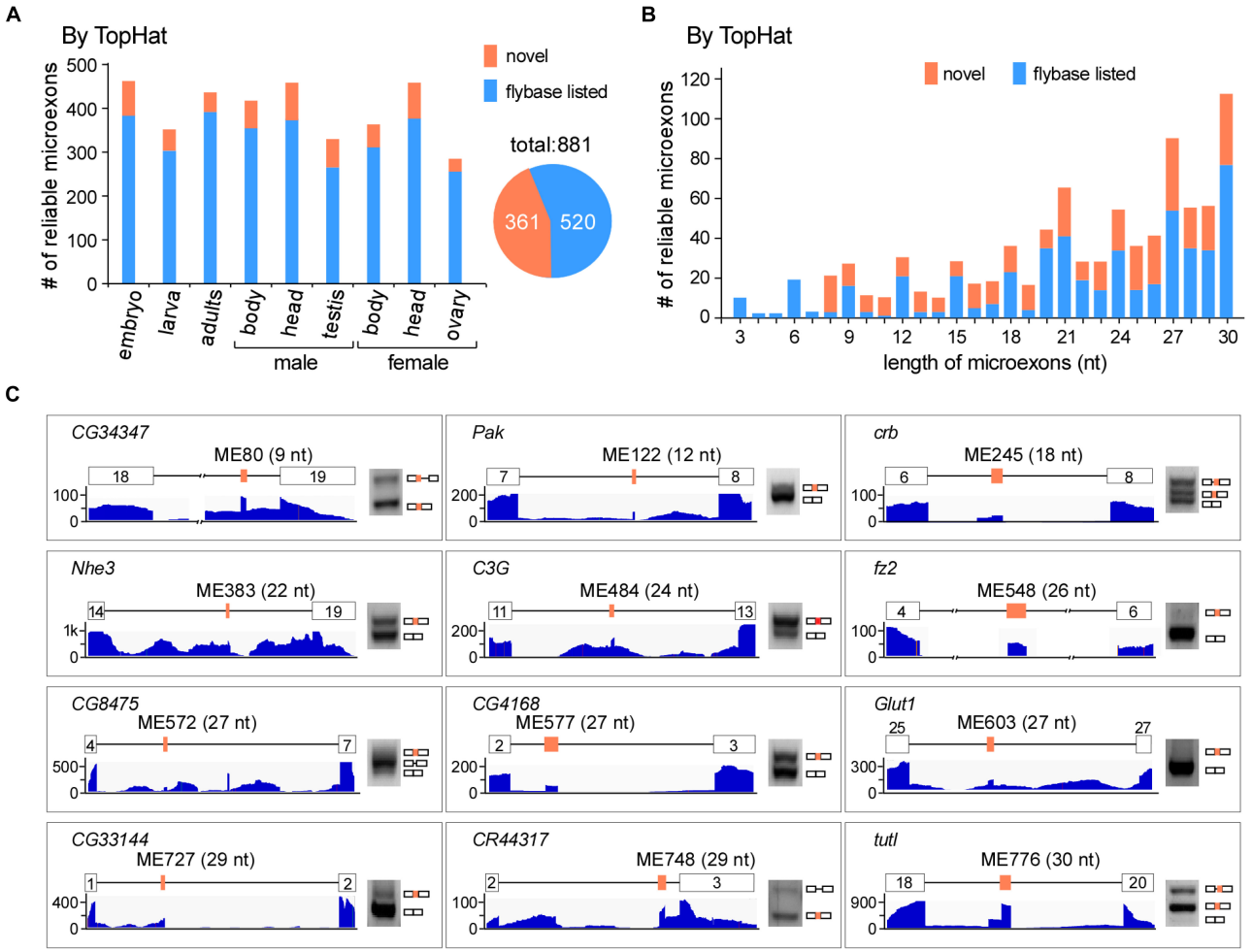


FIGURE 1 | Identification of microexons in *Drosophila* by TopHat. **(A)** Identified reliable microexons from RNA-seq data of a variety of *Drosophila* samples by TopHat. Orange: novel microexons; blue: annotated microexons in Flybase. **(B)** Length distribution of the identified microexons by TopHat. **(C)** Validation of novel microexons by RT-PCR. Each microexon was validated by amplification of cDNA using primers located in its flanking exons. Lengths and numbering of microexons and the genes to which they belong are indicated. Blank boxes: flanking exons, orange boxes: microexons. Primers information is listed in **Supplementary Table 1**.

flanking exons; the other set contained one primer wholly located in one of the flanking exons and the other primer located in the other flanking exon with extended new microexon sequence. All of the six tested novel microexons were correct in RT-PCR bands and sequencing (**Figure 2D**).

Taken together, we identified 985 reliable microexons by ce-TopHat, of which 465 are novel microexons that were not annotated by the Flybase (**Figure 2C**).

Microexons Are Enriched in Cell Morphogenesis and Neural-Related Genes

We then performed GO analysis and found that microexon-containing genes in *Drosophila* are highly enriched in two aspects, the cell structure and morphogenesis (93 genes) and the neural-related signaling and development (80 genes), including pathways such as actin filament-based process and cytoskeleton organization, supramolecular fiber organization and cell part morphogenesis, neuron projection development and morphogenesis, synaptic signaling, neuromuscular junction development and synapse organization (**Figure 3A**).

To further characterize the microexon-containing genes, we split them into two groups based on the length of microexons, the 3–21 nt, and 22–30 nt groups. GO analyses revealed that the top 5 enrichments were different between these two groups. The 3–21 nt group was enriched in the aspect of cell morphogenesis, including cell part morphogenesis, cell morphogenesis involved in neuron differentiation, plasma membrane bounded cell projection morphogenesis, and neuron projection development and morphogenesis (**Supplementary Figure 3A**). However, genes in the 22–30 nt group were mostly enriched in the aspect of neural signaling and morphogenesis, including synaptic signaling, neuromuscular junction development, neuromuscular synaptic transmission, and synaptic signaling, as well as genes in the actin filament-based process (**Supplementary Figure 3B**). This

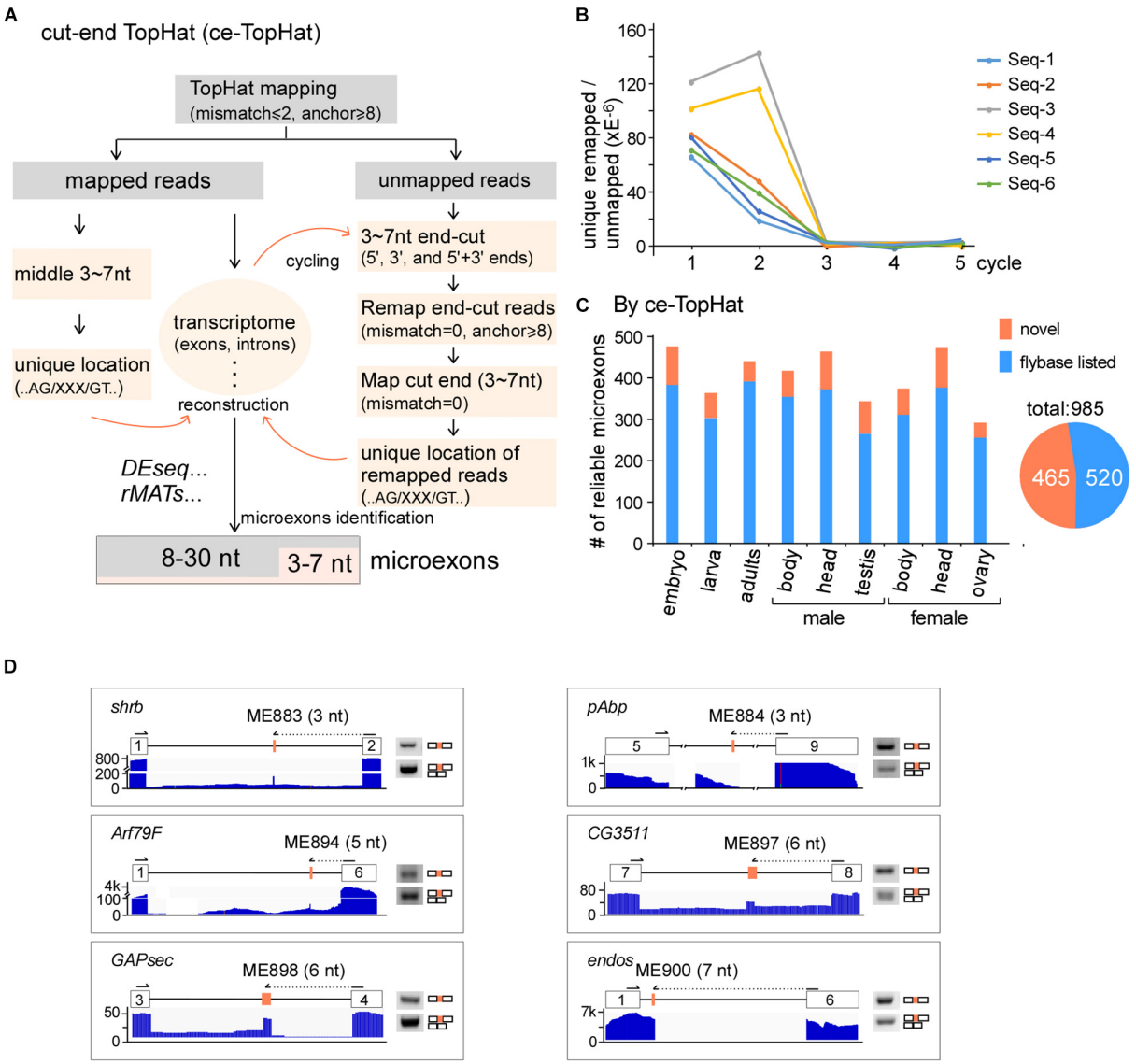


FIGURE 2 | Identification of microexons in *Drosophila* by ce-TopHat. **(A)** Strategy for the identification of microexons by ce-TopHat. **(B)** Unique genomic locations of the remapped reads reach zero after three cycles of ce-TopHat. Six RNA-seq samples are indicated here, and other tested samples are shown in **Supplementary Figure 2**. **(C)** Identified reliable microexons from a variety of *Drosophila* samples using ce-TopHat. Orange: novel microexons; blue: annotated microexons on Flybase. **(D)** Validation of 3–7 nt novel microexons by RT-PCR. Arrows indicate the location of one set of PCR primers. All the PCR products were confirmed by Sanger sequencing.

observation implies that splicing would be different for various lengths of microexons, and this might be due to delicate splicing regulation in different environments. This is also consistent with previous studies, showing that alternatively spliced microexons play a major role in the development of the nervous system (Tsyba et al., 2004, 2008; Zibetti et al., 2010; Ohnishi et al., 2014; Toffolo et al., 2014).

Alternative Splicing of Microexons

We further found that most *Drosophila* microexons were in coding genes, while only 5.8% were in non-coding genes. Among the coding genes, 58.1 and 34.2% of the microexons were in the CDS and 5′-UTR regions, respectively; few were found in the 3′-UTR region (**Figure 3B** light blue). Interestingly, microexons with the length of integer multiples of 3 nt (3n) exhibited significantly higher ratios (71.7%) in the CDS region (**Figure 3B** dark blue), suggesting that skipping or inclusion of most *Drosophila* microexons in CDS does not change much of their ORFs, but results in a difference in amino acids in a small number of proteins.

We then focused on the 411 exon-microexon-exon mode microexons, which have flanking exons on both sides (**Supplementary Table 5**). Among them, 68.9% (283) were alternatively spliced (AS) exons and 31.1% (128) were

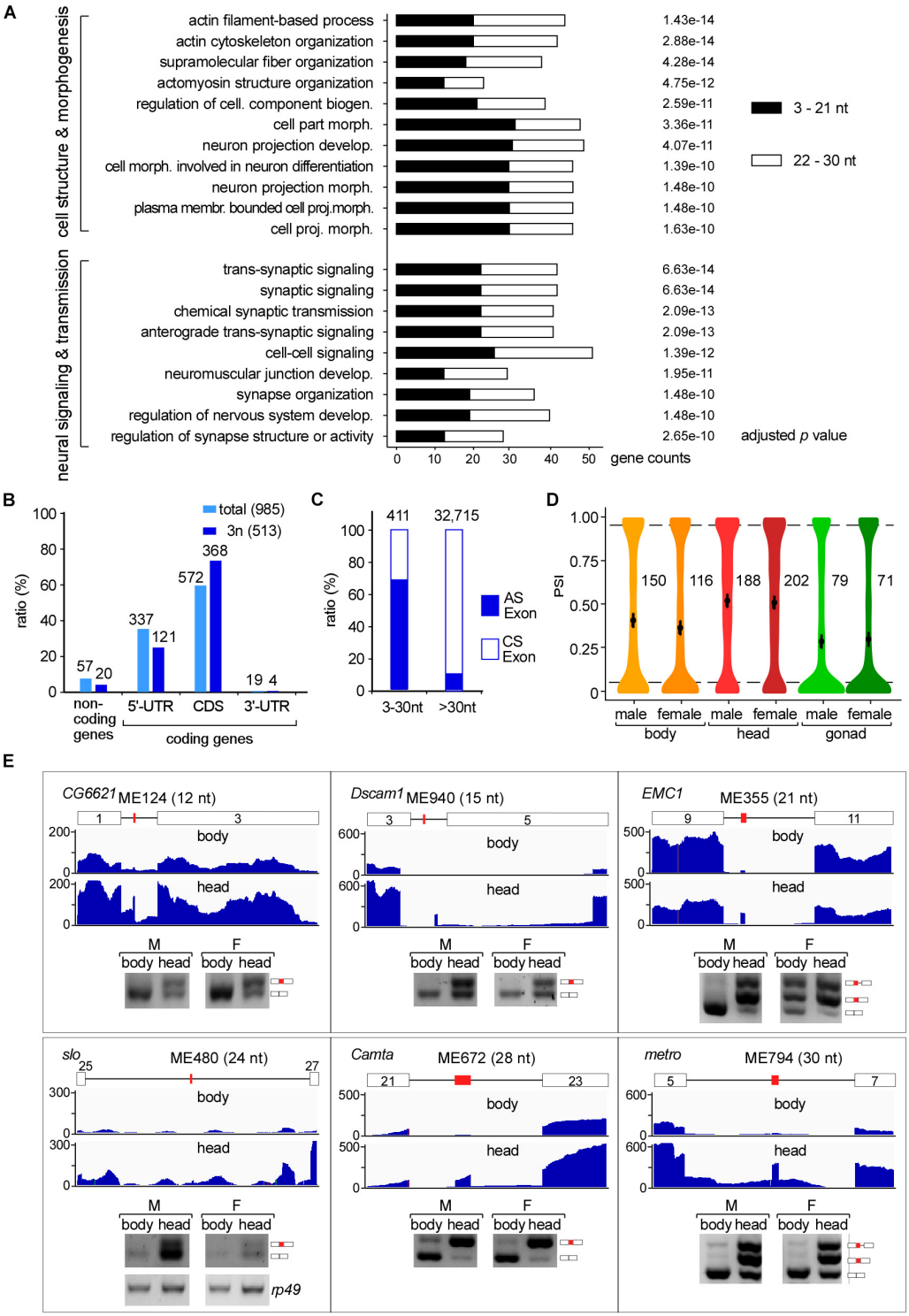


FIGURE 3 | Tissues specific AS of Drosophila microexons. **(A)** Microexon-containing genes in *Drosophila* are highly enriched in two aspects of cellular functions by GO analysis. Enriched microexons are also divided into two groups based on length. **(B)** Distribution of microexons in gene regions. 3n: microexons in the length of integer multiples of 3. **(C)** Ratio of the AS exons is much higher in microexons than in longer exons. Microexons and longer exons that are located in the CDS region and have flanking exons were retrieved and analyzed. AS and CS exons were determined by transcriptomic information. **(D)** Distribution of the PSI values of AS microexons in *Drosophila* part/tissues. Average PSIs are indicated for each sample. **(E)** Validation of head-specific AS microexons by RT-PCR. Each microexon was validated by amplification of cDNA using two primers located in its flanking exons. Lengths, belonging genes, and names of microexons are indicated. Blank boxes: flanking exons, orange boxes: microexons. Primers information is listed in **Supplementary Table 1**.

constitutive spliced (CS) exons. The ratio of AS in microexons is significantly higher than in the longer *Drosophila* exons, in which only 9.8% were alternatively spliced (**Figure 3C**). Analyses of the MAXENT scores, representing the strength of splice site signals, revealed that the strength of 5′SSs from AS microexons was at similar levels as their upstream 5′SSs, but the strength of 5′SSs from CS microexons was significantly higher than their upstream 5′SSs. A similar pattern was also found for the strength of 3′SSs (**Supplementary Figure 4A**). Consistent with this, consensus sequences of the upstream 5′SSs and downstream 3′SSs of CS microexons were less conserved than their counterparts of the AS microexons (**Supplementary Figure 4B**).

Many AS Microexons Only Spliced in the Head

The alternative spliced 283 microexons were involved in 388 AS events (**Supplementary Table 5**). Percent-spliced-in (PSI) analysis revealed that AS events in the head from both females and males exhibited significantly higher average PSI values than in the bodies and gonads (**Figure 3D**), suggesting that a large portion of the AS microexons are actively spliced and included in transcripts in the *Drosophila* head. Comparison between the three parts of *Drosophila* using ΔPSI, an indicator for differential AS, also showed that the head was significantly different from the body and gonad in the splicing of microexons, while the difference between the bodies and gonads was much less (**Supplementary Figure 4C**). Interestingly, we found that 31 microexons were specifically spliced/included in the head and skipped in the bodies and gonads, while only four microexons were specific in the bodies and no microexons were specific in the gonads (**Supplementary Table 6**). These results strongly suggested that AS microexons have tissue specificities and splicing regulation of microexon is important for development and differentiation, especially in the head or the neuron systems. We also performed similar splicing analyses between our *Drosophila* developmental stage samples, including the embryo, larva, and adults. However, the results showed much less difference between the developmental stages than between the tissues.

To verify, we performed RT-PCR and analyzed splicing of differentially AS microexons between the heads and bodies, including the ME124 (12 nt in *CG6621*), ME940 (15 nt in *Dscam1*), ME335 (21 nt in *EMC1*), ME480 (24 nt in *slo*), ME672 (28 in *Camta*), and ME794 (30 nt in *metro*). The six tested microexons exhibited different AS splicing patterns between the two parts of *Drosophila*, in which more were spliced/included in the head samples (**Figure 3E**).

Sexually AS Microexons in *Drosophila*

We then analyzed differentially AS microexons in the three parts of *Drosophila* females and males. Unlike the above analyses, the female and male head samples showed few differentially AS microexons, while the sexually AS microexons mainly occurred in the bodies and gonads (**Figure 4A**). GO analyses showed that 61 genes containing the sexually AS microexons are mainly involved in the sensory system morphogenesis, cell structure organization, and the development of muscle cells, ovarian follicle cells, and epithelial cells (**Supplementary Figure 5** and **Supplementary Table 7**).

To verify this, we tested five genes that contain sexually AS microexons, including the ME18 (6 nt in *UGP*), ME133 (12 nt in *shi*), ME340 (21 nt in *Spn*), ME451 (24 nt in *didum*), and ME482 (24 nt in *sdk*). Confirmed by RT-PCR, more ME18, ME133 and ME451 were spliced/included in transcripts in females than in males, and more ME340 and ME482 were spliced/included in males than in females (**Figure 4B**).

The *fruitless* Gene Has Two AS Microexons

Among the genes in the sex determination pathway and sexual development, one of them, the *fruitless*, has two microexons, ME756 and ME177. ME756 (29 nt) is a novel head-specific microexon. Transcripts containing this microexon resulted in a premature termination codon and would decrease the protein level of Fruitless in the head. ME177 was previously annotated in the Flybase, here we found that it was specifically spliced in the male samples, especially in testis (**Figure 4C**).

DISCUSSION

Microexons are unique in both identification and function because of their short lengths. In this study, combined with deep sequencing of multiple *Drosophila* RNA samples and a new tool that recovers unmapped reads, we found hundreds of new microexons from the fruit fly transcriptomes and analyzed their differential AS between tissues and genders. These results demonstrate that more microexons could be identified through broader investigation of developmental and tissue/cell-specific transcriptomes, as well as the optimized and developed bioinformatic tools.

Most AS of microexons in previous studies were analyzed in the neural systems of mammals and fruit flies. In this study, we provide additional evidence that the AS of microexons are common in the *Drosophila* developmental stages, tissues, and gonads. These results imply that splicing regulation of microexons would be critical for the development of organs in *Drosophila*.

Lengths of exons and introns vary greatly in most species. The mode lengths of exons are 96 and 135 nt in humans and *Drosophila*, respectively (**Supplementary Figure 6**). Splicing of long-intron-flanking exons is usually considered in a mode of exon-definition in mammals, which defines exons by recognition of exonic splicing enhancers (ESE) as well as sequences of splice sites (Berget, 1995; De Conti et al., 2013). However, due to the short length, it has been speculated that the definition of microexons was also facilitated by sequences in the flanking introns (Scheckel and Darnell, 2015). Our identification of AS microexons in *Drosophila*, especially those 3–7 nts, would be useful to address such mechanisms in the future.

Another issue with microexons is their conservation across species. We performed evolutionary analyses of the *fruitless*

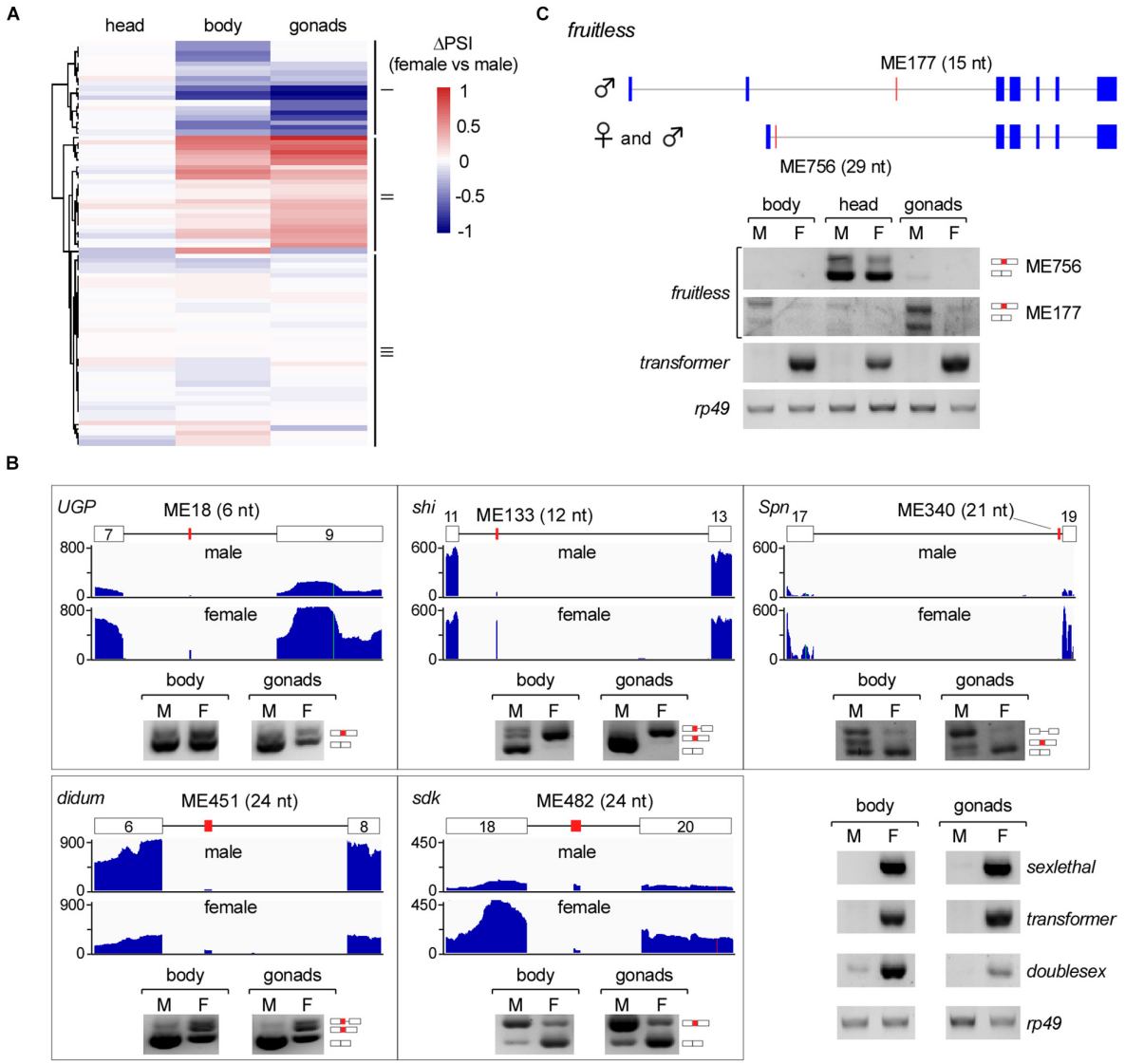


FIGURE 4 | Sex-specific AS of *Drosophila* microexons. **(A)** Comparison of differential AS between samples from two *Drosophila* genders. **(B)** Validation of sexually AS microexons. Each microexon was validated by amplification of cDNA using two primers located in its flanking exons. Lengths, belonging genes, and names of microexons are indicated. Blank boxes: flanking exons, orange boxes: microexons. Primers information is listed in **Supplementary Table 1**. **(C)** Two microexons in *fruitless* are alternatively spliced.

ortholog genes and found that the sequence of the ME756 in *Drosophila* is highly conserved from zebrafish to human (**Supplementary Figure 7**). All of the sequences are located in the middle region for coding a BTB domain in the protein orthologs, suggesting an important function of this microexon in the fruit fly.

DATA AVAILABILITY STATEMENT

Next-generation sequencing has been submitted to the Gene Expression Omnibus (accession number GSE163147). All other data will be made available on request.

AUTHOR CONTRIBUTIONS

T-LP and Y-ZX designed the methodology. T-LP and ZD performed the bioinformatic analyses with help from S-BL. YZ, BZ, and LL carried out the investigation. T-LP and ZD performed the visualization. T-LP and Y-ZX wrote the original draft. Y-JF and Y-ZX acquired the funding and supervised the study. All authors contributed to the review and editing.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (NSFC 31570821, 31525022,

31971225, and 91440109) to Y-JF and Y-ZX and from the Science and Technology Department of Hubei Province, China (2020CFA017) to Y-ZX.

ACKNOWLEDGMENTS

We thank other members in the Xu Lab for data entry and discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.642602/full#supplementary-material

Supplementary Figure 1 | TopHat mapping, identification, and filtering of microexons.

Supplementary Figure 2 | Remapping of the TopHat-unmapped reads by ce-TopHat.

Supplementary Figure 3 | The top five GO enrichments of microexon-containing genes.

Supplementary Figure 4 | Analyses of alternatively spliced microexons in *Drosophila*.

Supplementary Figure 5 | Distribution and GO analysis of sexually AS microexons.

Supplementary Figure 6 | Length distribution of exons in human, mouse, zebrafish, and fruit fly.

Supplementary Figure 7 | The evolutionary analysis of BTB domain-containing orthologs of the *Drosophila fruitless* gene.

Supplementary Table 1 | Primers used in this study.

Supplementary Table 2 | Analyses of RNA-seq reads from fruit fly samples.

Supplementary Table 3 | Remapping of unmapped-reads by ce-TopHat.

Supplementary Table 4 | List of the identified reliable microexons in this study.

Supplementary Table 5 | List of alternatively and constitutively spliced CDS microexons in *Drosophila.*

Supplementary Table 6 | List of AS microexons that are specifically spliced and included in tissues.

Supplementary Table 7 | List of significantly sexually AS microexons in *Drosophila.*

REFERENCES

Beachy, P. A., Helfand, S. L., and Hogness, D. S. (1985). Segmental distribution of bithorax complex proteins during Drosophila development. *Nature* 313, 545–551. doi: 10.1038/313545a0

Benoist, P., Mas, J. A., Marco, R., and Cervera, M. (1998). Differential muscle-type expression of the Drosophila troponin T gene. A 3-base pair microexon is involved in visceral and adult hypodermic muscle specification. *J. Biol. Chem.* 273, 7538–7546. doi: 10.1074/jbc.273.13.7538

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

Cooper, T. A., and Ordahl, C. P. (1985). A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. *J. Biol. Chem.* 260, 11140–11148. doi: 10.1016/s0021-9258(17)39158-5

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

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

Hartley, S. W., and Mullikin, J. C. (2016). Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res.* 44:e127.

Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi: 10.1016/j.cell.2014.11.035

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

Li, L., Ding, Z., Pang, T. L., Zhang, B., Li, C. H., Liang, A. M., et al. (2020). Defective minor spliceosomes induce SMA-associated phenotypes through sensitive intron-containing neural genes in Drosophila. *Nat. Commun.* 11:5608.

Li, Y. I., Sanchez-Pulido, L., Haerty, W., and Ponting, C. P. (2015). RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* 25, 1–13. doi: 10.1101/gr.181990.114

McAllister, L., Rehm, E. J., Goodman, G. S., and Zinn, K. (1992). Alternative splicing of micro-exons creates multiple forms of the insect cell adhesion molecule fasciclin I. *J. Neurosci.* 12, 895–905. doi: 10.1523/jneurosci.12-03-00895.1992

Ohnishi, T., Shirane, M., Hashimoto, Y., Saita, S., and Nakayama, K. I. (2014). Identification and characterization of a neuron-specific isoform of protrudin. *Genes Cells* 19, 97–111. doi: 10.1111/gtc.12109

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415. doi: 10.1038/ng.259

Qiu, C., Zhang, Y., Fan, Y. J., Pang, T. L., Su, Y., Zhan, S., et al. (2019). HITS-CLIP reveals sex-differential RNA binding and alterative splicing regulation of SRm160 in Drosophila. *J. Mole. Cell Biol.* 11, 170–181. doi: 10.1093/jmcb/mjy029

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

Santoni, M. J., Barthels, D., Vopper, G., Boned, A., Goridis, C., and Wille, W. (1989). Differential exon usage involving an unusual splicing mechanism generates at least eight types of NCAM cDNA in mouse brain. *EMBO J.* 8, 385–392. doi: 10.1002/j.1460-2075.1989.tb03389.x

Scheckel, C., and Darnell, R. B. (2015). Microexons–tiny but mighty. *EMBO J.* 34, 273–274. doi: 10.15252/embj.201490651

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

Small, S. J., Haines, S. L., and Akeson, R. A. (1988). Polypeptide variation in an N-CAM extracellular immunoglobulin-like fold is developmentally regulated through alternative splicing. *Neuron* 1, 1007–1017. doi: 10.1016/0896-6273(88)90158-4

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

Toffolo, E., Rusconi, F., Paganini, L., Tortorici, M., Pilotto, S., Heise, C., et al. (2014). Phosphorylation of neuronal Lysine-Specific Demethylase 1LSD1/KDM1A impairs transcriptional repression by regulating interaction with CoREST and histone deacetylases HDAC1/2. *J. Neurochem.* 128, 603–616. doi: 10.1111/jnc.12457

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Tsyba, L., Gryaznova, T., Dergai, O., Dergai, M., Skrypkina, I., Kropyvko, S., et al. (2008). Alternative splicing affecting the SH3A domain controls the binding

properties of intersectin 1 in neurons. *Biochem. Biophy. Res. Commun.* 372, 929–934. doi: 10.1016/j.bbrc.2008.05.156

Tsyba, L., Skrypkina, I., Rynditch, A., Nikolaienko, O., Ferenets, G., Fortna, A., et al. (2004). Alternative splicing of mammalian Intersectin 1: domain associations and tissue specificities. *Genomics* 84, 106–113. doi: 10.1016/j.ygeno.2004.02.005

Ustianenko, D., Weyn-Vanhentenryck, S. M., and Zhang, C. (2017). Microexons: discovery, regulation, and function. Wiley interdisciplinary reviews. *RNA* 2017:8.

Volfovsky, N., Haas, B. J., and Salzberg, S. L. (2003). Computational discovery of internal micro-exons. *Genome Res.* 13, 1216–1221. doi: 10.1101/gr.677503

Will, C. L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 2011:3.

Wu, J., Anczukow, O., Krainer, A. R., Zhang, M. Q., and Zhang, C. (2013). OLego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. *Nucleic Acids Res.* 41, 5149–5163. doi: 10.1093/nar/gkt216

Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. doi: 10.1093/bioinformatics/bti310

Wu, W., Zong, J., Wei, N., Cheng, J., Zhou, X., Cheng, Y., et al. (2018). CASH: a constructing comprehensive splice site method for detecting alternative splicing events. *Briefings Bioinform.* 19, 905–917. doi: 10.1093/bib/bbx034

Yang, L., and Chen, L. L. (2014). Microexons go big. *Cell* 159, 1488–1489. doi: 10.1016/j.cell.2014.12.004

Yeo, G., and Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* 11, 377–394. doi: 10.1089/1066527041410418

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

Zibetti, C., Adamo, A., Binda, C., Forneris, F., Toffolo, E., Verpelli, C., et al. (2010). Alternative splicing of the histone demethylase LSD1/KDM1 contributes to the modulation of neurite morphogenesis in the mammalian nervous system. *J. Neurosci.* 30, 2521–2532. doi: 10.1523/jneurosci.5500-09.2010

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 March 2021
doi: 10.3389/fgene.2021.644238

Complex RNA Secondary Structures Mediate Mutually Exclusive Splicing of Coleoptera *Dscam1*

*Haiyang Dong, Lei Li, Xiaohua Zhu, Jilong Shi, Ying Fu, Shixin Zhang, Yang Shi, Bingbing Xu, Jian Zhang, Feng Shi and Yongfeng Jin**

MOE Laboratory of Biosystems Homeostasis, Protection and Innovation Center for Cell Signaling Network, College of Life Sciences, Zhejiang University, Hangzhou, China

Mutually exclusive splicing is an important mechanism for expanding protein diversity. An extreme example is the Down syndrome cell adhesion molecular (*Dscam1*) gene of insects, containing four clusters of variable exons (exons 4, 6, 9, and 17), which potentially generates tens of thousands of protein isoforms through mutually exclusive splicing, of which regulatory mechanisms are still elusive. Here, we systematically analyzed the variable exon 4, 6, and 9 clusters of *Dscam1* in Coleoptera species. Through comparative genomics and RNA secondary structure prediction, we found apparent evidence that the evolutionarily conserved RNA base pairing mediates mutually exclusive splicing in the *Dscam1* exon 4 cluster. In contrast to the fly exon 6, most exon 6 selector sequences in Coleoptera species are partially located in the variable exon region. Besides, bidirectional RNA–RNA interactions are predicted to regulate the mutually exclusive splicing of variable exon 9 of *Dscam1*. Although the docking sites in exon 4 and 9 clusters are clade specific, the docking sites-selector base pairing is conserved in secondary structure level. In short, our result provided a mechanistic framework for the application of long-range RNA base pairings in regulating the mutually exclusive splicing of Coleoptera *Dscam1*.

Keywords: clade-specific, Coleoptera, mechanism, RNA secondary structure, *Dscam1*, alternative splicing

OPEN ACCESS

Edited by:
Liang Chen,
Wuhan University, China

Reviewed by:
Qiangfeng Zhang,
Tsinghua University, China
Lucie Grodecká,
Center of Cardiovascular and Transplant Surgery (Czechia), Czechia

****Correspondence:***
Yongfeng Jin
jinyf@zju.edu.cn

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *20 December 2020*
Accepted: *23 February 2021*
Published: *30 March 2021*

Citation:
Dong H, Li L, Zhu X, Shi J, Fu Y, Zhang S, Shi Y, Xu B, Zhang J, Shi F and Jin Y (2021) Complex RNA Secondary Structures Mediate Mutually Exclusive Splicing of Coleoptera Dscam1. Front. Genet. 12:644238. doi: 10.3389/fgene.2021.644238

INTRODUCTION

Alternative splicing is an important precursor RNA processing method to increase protein diversity in eukaryotes (Nilsen and Graveley, 2010; Pandey et al., 2020; Suresh et al., 2020). Alternative splicing is ubiquitous in various processes such as human nerve development, spermatogenesis, muscle contraction, and immune defense (Gallego-Paez et al., 2017). Abnormal alternative splicing events might be associated with diseases, e.g., cancers and neurodegenerative diseases (Kim et al., 2018; Montes et al., 2019; Bonnal et al., 2020; Wang et al., 2020). Pre-messenger RNA (pre-mRNA) alternative splicing has recently been thought to be related to the aging process and longevity (Bhadra et al., 2020). There are five main types of alternative splicing, including intron retention, exon skipping, alternative 3′ splice sites, alternative 5′ splice sites, and mutually exclusive splicing (Nilsen and Graveley, 2010; Zhang et al., 2016; Hatje et al., 2017; Jin et al., 2018). Mutually exclusive splicing is a specific type of alternative splicing; in a tandem exon array, only one variable

exon can be spliced into the mature mRNA at a time (Smith, 2005). Mutually exclusive exons originate from exon duplication events (Graveley et al., 2004; Chen et al., 2011; Brites et al., 2013; Hatje and Kollmar, 2013; Yue et al., 2017). An extreme case of mutually exclusive splicing event is *Dscam1* in arthropods (Lee et al., 2010). In *Drosophila melanogaster*, *Dscam1* contains four clusters of variable exons 4, 6, 9, and 17 with 12, 48, 33, and 2 variable exons, respectively, and potentially produce 38,016 protein isoforms *via* mutually exclusive splicing (Schmucker et al., 2000). Due to the fact that homologous Dscam1 protein isoforms mediate self-avoidance (Wojtowicz et al., 2004; Soba et al., 2007; Zipursky and Grueber, 2013), such a staggering number of Dscam1 protein isoforms are functional for *D. melanogaster* neurons to identify self or non-self (Hattori et al., 2007; Hughes et al., 2007; Matthews et al., 2007; Kise and Schmucker, 2013). Dscam1 also plays an important role in the neuron circuit as an axon guidance receptor (Schmucker et al., 2000; Cvetkovska et al., 2013). Besides, evidence has revealed that Dscam1 is required for the immune function as the Ig superfamily member (Dong et al., 2006; Armitage et al., 2015; Ng and Kurtz, 2020).

An attractive regulatory mechanism of alternative splicing is the competitive RNA secondary structure mediating the splicing of exon variants (Graveley, 2005; Anastassiou et al., 2006; Yang et al., 2011; Xu et al., 2020). The most typical gene of this model is the variable exon 6 cluster of *Dscam1* in *D. melanogaster*. In the exon 6 cluster, two types of conserved intron elements participate in the alternative splicing of variable exon 6. The first intron element was located in the intron between the constitutive exon 5 and variable exon 6.1 and was referred to as the docking site. The docking site was the most conserved intron element in the entire *Dscam1* gene. Another type of intron element is the selector sequence; 48 selector sequences were located upstream of 48 variable exon 6s and were relatively conserved. Moreover, all 48 selector sequences were complementary to the only one docking site (Graveley, 2005). Besides, there is a class of heterogeneous nuclear ribonucleoprotein protein (hrp36) that uniformly covers the entire variable exon 6 cluster to maintain the fidelity of the mutually exclusive splicing (Olson et al., 2007). When the docking site pairs with the selector sequence of a specific exon to form an RNA secondary structure, the hrp36 protein on this exon will fall off, thereby promoting the splicing of this exon 6 (Graveley, 2005; Xu et al., 2019). Only the variable exon that forms secondary structures can release the inhibition proteins and trigger splicing. Moreover, an RNA locus control region (LCR) exists between constitutive exon 5 and the exon 6 docking site of *Dscam1* to promote the splicing of the adjacent downstream variable exon that forms the RNA secondary structure (Wang et al., 2012). Besides, similar docking site-selector base pairings also exist in vertebrate genes (Pervouchine et al., 2012; Suyama, 2013).

The mechanism by which competitive RNA secondary structure regulates the mutually exclusive splicing of variable exon 6 had been widely recognized (May et al., 2011). However, there are still some obstacles and doubts for the complete cognition of the variable exon 4 and 9 clusters of *Dscam1*. In our previous studies, downstream RNA pairings have been identified to regulate the splicing of exons 4 and 9 variants of *Dscam1* in *Drosophila* (Yang et al., 2011). Bidirectional competitive RNA secondary structure regulated the inclusion of variable exons in the exon 4 cluster of Hymenopteran *Dscam1* and the exon 9 clusters of Lepidopteran and Hymenopteran *Dscam1* (Yue et al., 2016). However, some other researchers questioned the regulatory mechanisms by which long-range competitive RNA secondary structure regulates the splicing of exons 4 and 9 due to the lack of apparent conserved intron elements (Haussmann et al., 2019; Ustaoglu et al., 2019). Recently, a unique evolutionary midge-specific docking site has been found in the exon 6 cluster, which regulates the process of alternative splicing *via* base pairing (Hong et al., 2020). However, the splicing of exon 4 and 9 clusters has still not been well explained.

Whether clade- or species-specific but RNA secondary structure conserved docking site can mediate alternative splicing of exons 4 and 9 of *Dscam1*? We focus on Coleoptera to further explore that. Coleoptera, roughly 360,000 described species make up about 40% of all insect species (Bouchard et al., 2017), is the largest order in Insecta (Woodcock et al., 2013; Zhang et al., 2018), and make up almost 25% of all animals (Hunt et al., 2007). Thus, many species provide convenience for evolutionary analysis. Moreover, the rapid development of public databases has enabled the genomic data of multiple species of Coleoptera to be found in GenBank (Bocak et al., 2014), providing us with a rich source of sequence alignment. These characteristics make Coleoptera a suitable material for studying alternative splicing of *Dscam1*.

Through sequence alignment and secondary structure prediction, we found that the clade-specific docking site can mediate the selection of exon 4 *via* the formation of RNA secondary structure with the selector sequences in a base-pairing manner. Moreover, bidirectional competitive RNA secondary structures were also discovered in the exon 9 cluster. Although the primary sequence of exon 4 and 9 docking sites were clade specific or species specific, the docking site-selector base pairing was conserved in the RNA secondary structure level. In addition, due to the short intron of the exon 6 cluster in Coleoptera, most selector sequences were partially located in exon regions. Taken together, our findings provided a mechanistic framework that competitive RNA secondary structure regulates mutually exclusive splicing of *Dscam1* exon 4, 6, and 9 clusters in Coleoptera.

MATERIALS AND METHODS

Identification and Annotation of *Dscam1* Gene Structure

The *Dscam1* genome sequences of Coleoptera species were obtained by using the *Dscam1* of *D. melanogaster* as the query sequences and performing TBLASTN search in the NCBI WGS database[1]. Annotation of the *Dscam1* was performed by comparative genomics with cross-species or intraspecies. The identification and the numbers of variable exons 4, 6, 9,

[1]https://blast.ncbi.nlm.nih.gov/Blast.cgi

and 17 were confirmed by nucleic acid or protein sequence alignment of variable exons between different species or within species. Combined with the existing RNA sequencing data, the boundaries of the variable exons can be further confirmed (**Supplementary Table 1**).

Sequence Alignment and Secondary Structure Analysis

Clustal Omega[2] was applied to sequence alignment. The docking site-selector sequences base pairings were predicted by the Mfold project[3] (Zuker, 2003). The conserved selector sequences were derived *via* the WebLogo[4] (Crooks et al., 2004).

The Drawing of the Evolutionary Tree

The amino acid sequence was composed of constitutive exons and randomly selected variable exons in each cluster, and the amino acid sequences of 14 Coleopteran *Dscam1* were imported into MEGA X[5]. Evolutionary relationships of taxa were drawn based on the Minimum Evolution method (Kumar et al., 2018).

RESULTS

Dscam1 Gene Structure and Molecular Diversity in Coleoptera Species

Sitophilus oryzae, a representative species of Coleoptera, has a similar gene structure to *D. melanogaster Dscam1*, containing 26 constitutive exons and 4 clusters of variable exons. However, the number of variable exons in exon 4, 6, and 9 clusters were different from those in *D. melanogaster* (Schmucker et al., 2000). In *S. oryzae*, exon 4, 6, 9, and 17 clusters contain 10, 38, 36, and 2 variable exons, respectively. It potentially produces 27,360 ($10 \times 38 \times 36 \times 2$) protein isoforms through mutually exclusive splicing. Dscam1 protein contains 10 immunoglobulin (Ig) domains and six fibronectin type III (FNIII) domains, a transmembrane domain, and a C-terminal intracellular region. Variable exons 4 and 6 encode half Ig2 and Ig3 domains, respectively, while exons 9 and 17 encode the whole Ig7 and transmembrane domains, respectively (**Figure 1A**).

After annotation of *Dscam1* genes in other 12 species (*Rhynchophorus ferrugineus*, *Dendroctonus ponderosae*, *Hypothenemus hampei*, *Callosobruchus maculatus*, *Anoplophora glabripennis*, *Leptinotarsa decemlineata*, *Aethina tumida*, *Oryzaephilus surinamensis*, *Coccinella septempunctata*, *Harmonia axyridis*, *Onthophagus taurus*, and *Nicrophorus vespilloides*), we found that the transmembrane domain of each species contains two variable exons (exon 17). However, the number of variable exons in exon 4, 6, and 9 clusters of the Coleoptera species vary. The number of variable exon 4s ranges from eight to 11, mostly with 9 exon variants, and does not change as much as exons 6 and 9. In the exon 4 cluster, the variable exon 4.4 was missing during evolution, resulting in only eight variants in *L. decemlineata*. On the contrary, 10 or 11 variable exons can be identified due to the duplication of variable exons in the *S. oryzae*, *R. ferrugineus*, *D. ponderosae*, and *H. hampei*, which all belong to the same superfamily (**Supplementary Figures 1, 2**). Correspondingly, the number of variable exon 9 ranges to a staggering 53 in *N. vespilloides*, more than twice to that in *D. ponderosae*, which only have 24. In the exon 6 cluster, unfortunately, due to the genomic sequence break in the database, we failed to determine the number of exon 6 variants of *Dscam1* in *C. maculatus* and *R. ferrugineus*. However, an interesting phenomenon was that the number of variable exons of *S. oryzae* was nearly twice that of *H. hampei*, even if they belong to the same family. Moreover, the number of exon 6s of all analyzed species was much smaller than the *D. melanogaster*, which has 48 exon 6 variants (**Figure 1B**). Although the number of variable exons varies between different species, *Dscam1* of most Coleoptera species can potentially generate tens of thousands of protein isoforms (the potential protein isoforms of *R. ferrugineus* and *C. maculatus Dscam1* are uncertain due to the lack of genomic sequence in exon 6 clusters).

Downstream RNA Pairing Mediates Mutually Exclusive Splicing of Exon 4 Cluster

Phylogenetic analyses revealed that most variable exon 4s were orthologous in Coleoptera species (**Supplementary Figures 1, 2**), indicating that the variable exon 4s derived from the common ancestor and less exon duplication or loss occur during the evolutionary process. This was consistent with the previous studies, which suggested that most exon 4s might be orthologous in the insects (Lee et al., 2010). To decipher the mechanism for *Dscam1* exon 4 mutually exclusive splicing, we first searched the conserved intron element. Docking site-selector sequence base pairing mediating mutually exclusive splicing in *Dscam1* exon 4 has been identified in *Drosophila* and Hymenoptera species (Yang et al., 2011; Yue et al., 2016). However, the primary sequences of the docking sites between *Drosophila* and Hymenoptera species were different. Therefore, we speculated that the primary sequences of the docking site in the exon 4 cluster were evolutionarily specific in the Coleoptera species. Through sequence alignment, we found a conserved intron element (docking site) downstream of the last variable exon 4 (**Figure 2A**). Indeed, the docking site sequences in Coleoptera were different from *Drosophila* and Hymenoptera species, indicating a clade-specific docking site in Coleoptera species. Moreover, only one apparent docking site has been found, similar to the exon 4 cluster of *Drosophila*, while there was a docking site on both sides of the exon 4 cluster in Hymenoptera species (Yue et al., 2016). Through RNA secondary structure prediction, evolutionarily conserved selector sequences complementary to the docking site were identified, and all the selector sequences were located downstream of the variable exons (**Figures 2B,C** and **Supplementary Figures 3–5**). Moreover, clear evidence of compensatory structural covariations and evolutionary intermediates exist within the core region of the RNA secondary structure formed by docking site-selector base pairing (**Figures 2A,B**). Due to the distant

[2]https://www.ebi.ac.uk/Tools/msa/clustalo

[3]http://unafold.rna.albany.edu/?q=mfold

[4]http://weblogo.berkeley.edu/logo.cgi

[5]https://www.megasoftware.net/

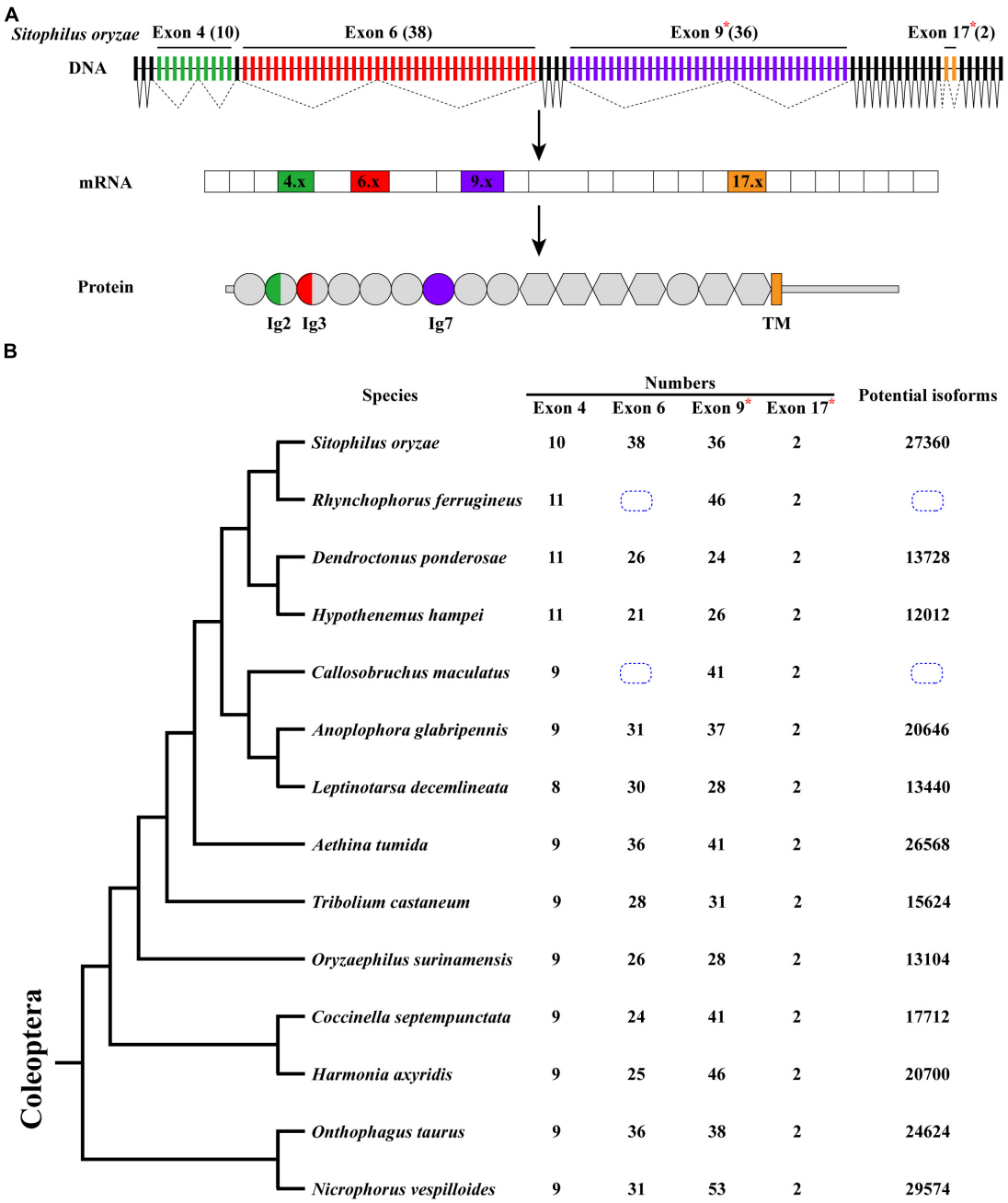


FIGURE 1 | *Dscam1* gene structure and molecular diversity of Coleoptera species. **(A)** Schematic diagram of the *Dscam1* gene structures of *S. oryzae*. Variable exons are marked by colored boxes, constitutive exons as black boxes. Dscam1 protein includes 10 immunoglobulin (Ig) domains (circles), six fibronectin type III domains (hexagons), one TM domain, and cytoplasmic tails. The variable exons 4 and 6 encode half Ig2 and Ig3 domains, respectively, while exons 9 and 17 encode the whole Ig7 and transmembrane domains, respectively. Variable exon 11 and 24 clusters of *S. oryzae* are evolutionarily homologous to exon 9 and 17 clusters of *D. melanogaster* and are marked with an "*" and named exons 9 and 17. **(B)** A phylogenetic tree of Coleoptera species is shown on the left. Evolutionary relationships of taxa were drawn with MEGA X. The number of variable exons in each cluster is shown in the middle, and the total potential isoforms are shown on the right. The blue dotted line box indicates that the number of exon 6 cannot be defined.

evolutionary relationship, the docking sites in *O. taurus* and *N. vespilloides* were less conserved compared to that in other species. However, conserved RNA secondary structures within these species were found (**Supplementary Figure 6**). Taken together, these results suggested that the downstream RNA base pairing could mediate the mutually exclusive splicing of variable exon 4 cluster, and the docking site showed to be clade specific.

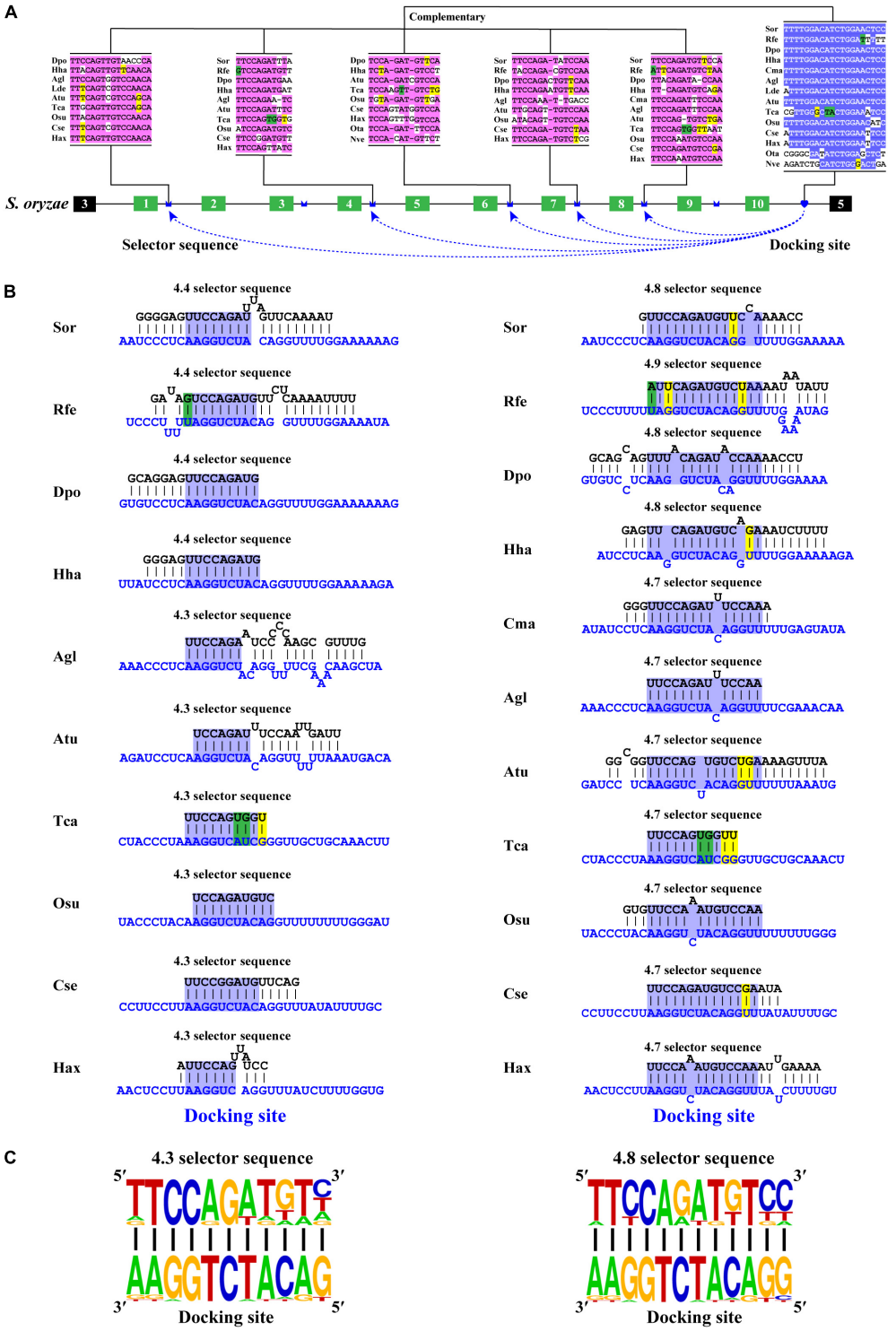


FIGURE 2 | Conserved downstream RNA pairings mediate mutually exclusive splicing of *Dsacm1* exon 4. **(A)** Schematic diagram of the *Dscam1* exon 4 of *S. oryzae*. The docking site (marked by blue heart) and each selector sequence (marked by blue crowns) are complementary. The conserved nucleotide sequences of the docking site and selector are highlighted in different colors. The base sequences are shown from 5′ to 3′. Abbreviations of the species name are shown in **Supplementary Table 1**. **(B)** The RNA secondary structures between the docking site and 4.3 and 4.8 selector sequences are shown among Coleoptera species. The sequences that make up the core region of the RNA secondary structure are highlighted in blue. The selector sequences are shown in black font, and the docking sites are shown in blue font. Nucleotides of compensatory structural covariations that maintain the base pairing are shaded in green, and their evolutionary intermediates (U-G, G-U) are shaded in yellow. **(C)** The most frequent nucleotides at each position of the 4.3 and 4.8 selector sequences among species are complementary to the docking site.

Most Selector Sequences of Exon 6 Cluster Are Partially Located in Variable Exon Region

After annotating the exon 6 cluster of *Dscam1* in Coleoptera, we calculated the length of introns between two variable exons. Surprisingly, up to 82% of intron lengths were <150 bp. More interestingly, more than 45% of intron lengths were <50 bp (**Figure 3A**). Due to the small intron (<50 bp), maybe nearly half of the selector sequences will be located in the exon region to avoid the steric hindrance. To test our hypothesis, we identified the evolutionarily conserved docking site of exon 6 cluster through sequence alignment and marked the selector sequences located upstream of each variable exon 6 by long-range competitive RNA secondary prediction. Notably, almost all exon 6s could find the corresponding selector sequence (**Supplementary Figures 7, 8**). These results indeed illustrated and consolidated that the mechanism of mutually exclusive splicing of the exon 6 cluster was regulated by the competitive RNA secondary structure. Moreover, it suggested that we may have found the correct selector sequences.

To explore the distribution of the selector sequences, we divided the location of the selector sequences into three types: completely located in the exon region, located in the intron–exon boundary region, and completely located in the intron region (**Figure 3B**). In the exon 6 cluster of *D. melanogaster Dscam1*, 85% (41 out of 48) of the selector sequences were completely located in intron regions, while the remaining seven selector sequences were located in the intron–exon boundary region (**Figure 3C**; Graveley, 2005). On the contrary, after analyzing the distribution of exon 6 selector sequences of 12 Coleoptera species, we found that 56% (14 out of 25) of the selector sequences of *D. ponderosae* exon 6 were completely located in intron regions, 12% (three out of 25) of the selector sequences were completely located in the exon region, and 32% (eight out of 25) of the selector sequences were located in the intron–exon boundary region. More obviously, only 7% (two out of 28) of the selector sequences of *T. castaneum* exon 6 completely located in intron regions, while 26 out of 28 selectors included the exon sequences (**Figure 3C**). In conclusion, our discovery in the exon 6 cluster of Coleoptera *Dscam1* expanded our understanding that the selector sequences can be located in or included the variable exon sequence, not just in the intron region.

Dual RNA Pairing Mediates Mutually Exclusive Splicing of Exon 9 Cluster

Next, we decoded the mutually exclusive splicing mechanism of the *Dscam1* exon 9 cluster. Previous studies have reported that the unidirectional-competitive RNA secondary structure regulates splicing of *Dscam1* exon 9 in *Drosophila* (Yang et al., 2011), bidirectional RNA base pairing in Lepidoptera, and Hymenoptera *Dscam1* exon 9 (Yue et al., 2016). What is more, the primary sequences of the docking site showed to be clade specific between *Drosophila*, Lepidoptera, and Hymenoptera. Likewise, through genome sequence alignment and RNA secondary structure prediction, two intron elements (upstream docking site and downstream docking site) in the exon 9 cluster were found. However, 10 out of 14 chosen species shared a conserved upstream docking site (**Figure 4A**); the primary sequence of upstream docking sites in *D. ponderosae*, *H. hampei*, *O. taurus*, and *N. vespilloides* was specific (shown later). By contrast, the downstream docking sites were conserved in 14 species. Moreover, both the primary sequences of upstream or downstream docking sites were clade specific compared to that of Lepidoptera and Hymenoptera.

Through further RNA secondary structure prediction, many downstream selector sequences complementary to the upstream docking site and many upstream selector sequences complementary to the downstream docking site were identified (**Supplementary Figures 9–14**). However, due to the poor homology between the variable exon 9s of *Dscam1* in Coleoptera species, it was difficult to confirm the conservativeness of evolutionarily corresponding selector sequences. Alternatively, we selected two selector sequences paired with upstream or downstream docking sites in each species. Through the alignment of so many downstream and upstream selector sequences, respectively, the core area of downstream and upstream selector sequences can form base pairing to upstream and downstream docking sites, respectively (**Figures 4B,C**). Moreover, compensatory structural covariations and evolutionary intermediates were shown to be formed by docking site-selector base pairing (**Figures 4A,B**). Upstream and downstream base pairings can form a relatively strong remote competitive RNA secondary structure (**Figures 4D,E**). Therefore, we concluded that clade-specific upstream and downstream docking sites regulated the mutually exclusive splicing of the *Dscam1* exon 9 cluster in Coleoptera species.

The Primary Structure of the Docking Site Is Specific, but the RNA Secondary Structure Is Conserved

Bidirectional competitive RNA secondary structure has been identified in the exon 9 cluster. However, in the process of intron sequence alignment, the upstream docking site of *D. ponderosae* and *H. hampei* showed specificity compared to the other 10 species, but they were evolutionarily conserved (**Figure 5A**). Recently, a midge-specific docking site in the exon 6 cluster has been identified (Hong et al., 2020). Therefore, we suspected that species-specific upstream docking sites existed in the exon 9 cluster of *D. ponderosae* and *H. hampei*. Through RNA secondary structure prediction, many downstream and upstream selector sequences were complementary to the upstream and downstream docking sites, respectively (**Figures 5B,C** and **Supplementary Figure 15**). Similarly, due to the poor evolutionary correspondence between variable exon 9s in *D. ponderosae* and *H. hampei*, we selected four selector sequences of each species for further analysis. Through selector sequences alignment, the upstream or downstream selector sequences shared a core conserved region, and the core region could interact with the upstream or downstream docking site *via* base pairing (**Figures 5D,E**). In addition, compensatory structural covariations and evolutionary intermediates exist within the core

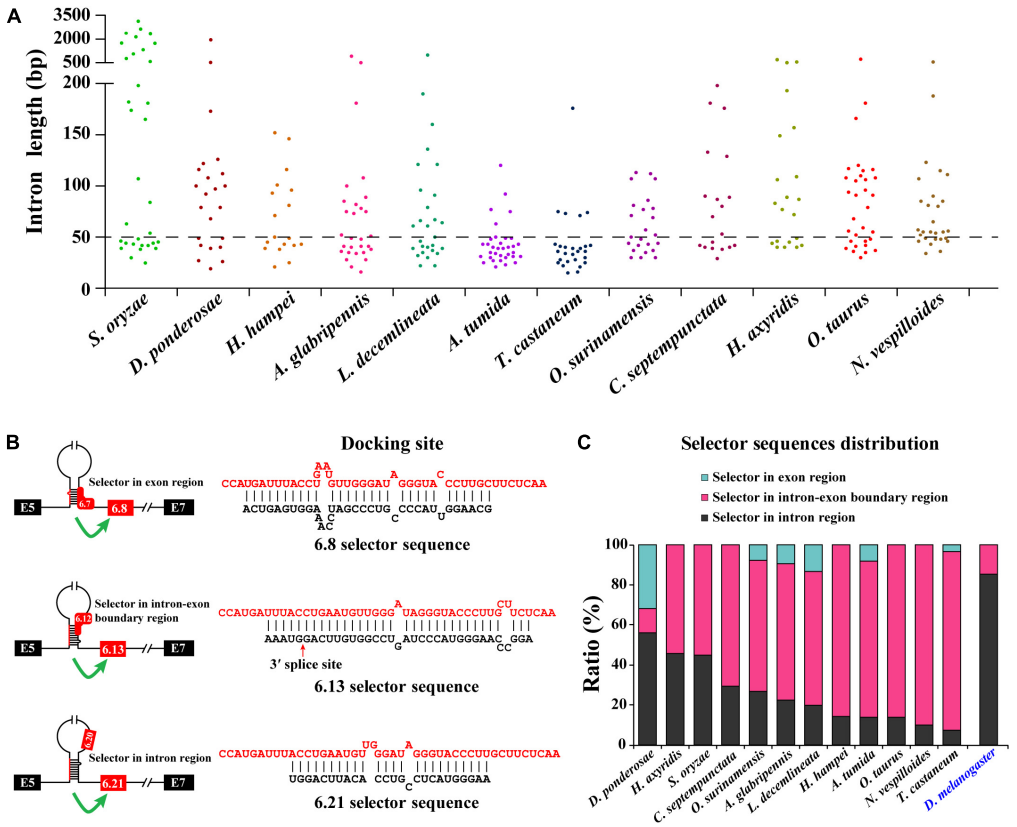


FIGURE 3 | Most exon 6 selector sequences are partially located in the exon sequence of Coleoptera *Dscam1* exon 6. **(A)** The intron length between two variable exon 6s of *Dscam1* in Coleoptera species is shown. **(B)** Three types of the location of selector sequences and the corresponding secondary structures are shown in *A. tumida*. The selector sequences are shown in black font, and the docking sites are shown in red font. **(C)** Comparison of the distribution of exon 6 selector sequences between Coleoptera species and *D. melanogaster*.

region of the RNA secondary structure formed by docking site-selector base pairing (**Figures 5B–E**). Likewise, the upstream docking sites were species specific, but the base pairings were conserved at the secondary structure level within species in *O. taurus* and *N. vespilloides* (**Supplementary Figures 16, 17**). However, all species shared a common region of downstream docking sites to form the downstream RNA base pairings. Hence, a species-specific docking site but with conserved RNA secondary structure could mediate alternative splicing of *Dscam1* exon 9.

Summary of Bidirectional Competitive RNA Secondary Structure in Exon 9

In this study, we identified bidirectional RNA base pairing in *Dscam1* exon 9 in Coleoptera species. Overall, 10 out of 14 chosen species shared a conserved upstream docking site, while the upstream docking site in *D. ponderosae*, *H. hampei*, *O. taurus*, and *N. vespilloides* was species specific. Besides, upstream docking sites between *D. ponderosae* and *H. hampei* were evolutionarily conserved. For the downstream base pairing, all chosen species shared a conserved downstream docking site (**Figure 6**). Taken together, we considered that during the evolution process, the primary sequences of the docking site would be mutated, but the base pairings in the secondary structure level were still conserved. Moreover, the dual docking sites may make up the splicing abnormality caused by the mutation of the docking site during evolution. Therefore, the bidirectional RNA secondary structure may be an adaptation of the organism to the evolution process.

DISCUSSION AND CONCLUSION

Through the comparative analyses of 14 species in Coleoptera, We propose a potential mechanism that competing RNA secondary structure could mediate mutually exclusive splicing in Coleoptera *Dscam1*. Downstream base pairings directed the splicing of variable exon 4s. In the exon 6 cluster, we expanded the location of the selector sequence that may be located in the exon region. Moreover, species- or clade-specific docking sites could mediate the splicing of exon 9 by forming a bidirectional competitive RNA secondary structure. These studies have provided more evidence for the view that competitive RNA secondary structures mediate *Dscam1* alternative splicing from an evolutionary perspective.

The mutually exclusive alternative splicing model of *Dscam1* exon 6 cluster guided by competitive secondary structure was proposed as early as 2005. Even if it has undergone evolution

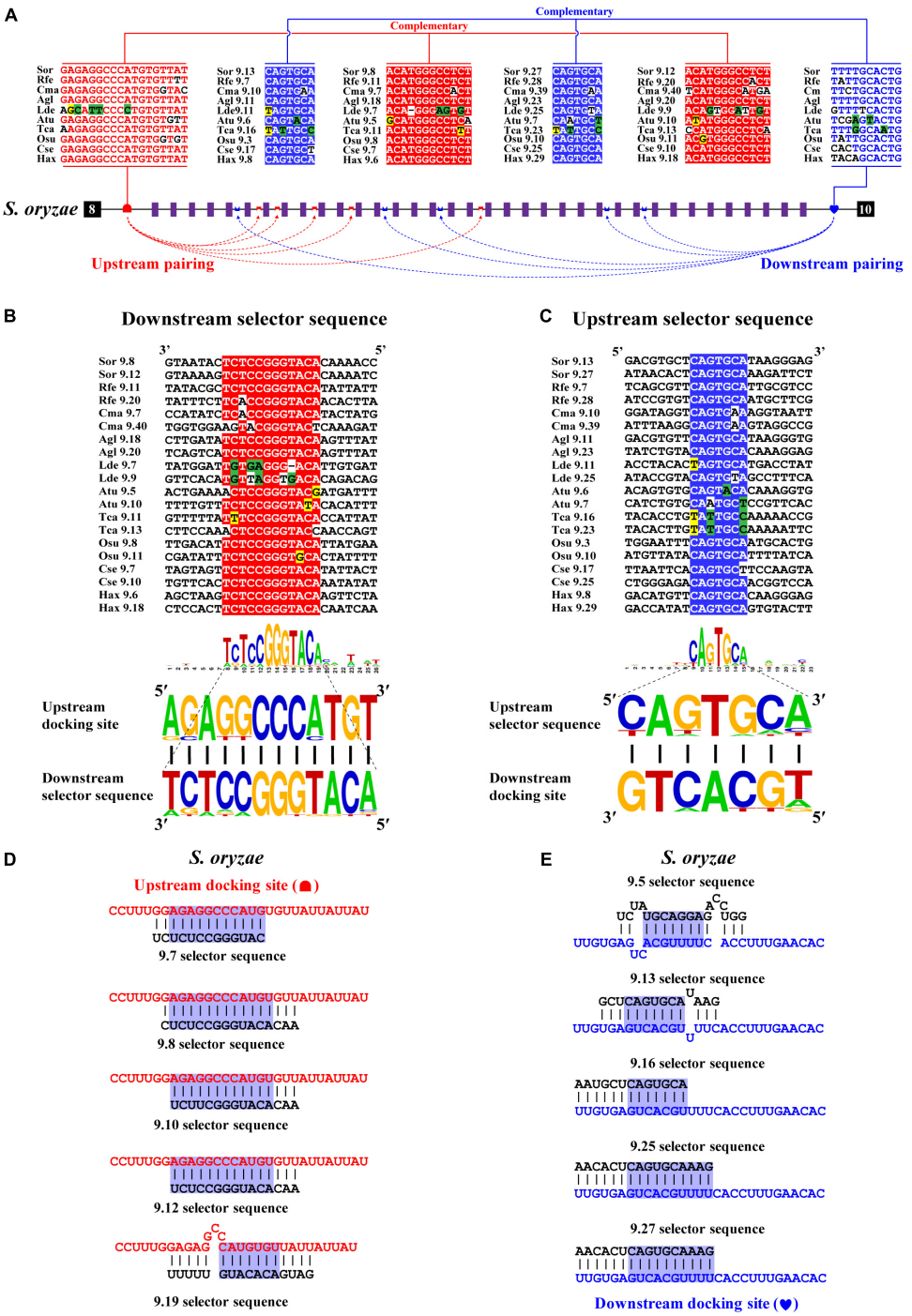


FIGURE 4 | Conserved dual docking site and selector sequences base pairing of Coleoptera *Dscam1* exon 9. **(A)** Schematic diagram of the *Dscam1* exon 9 of *S. oryzae*. Constitutive exons are depicted as black boxes and variable exon 9 as purple boxes. Upstream docking site (marked by red semicircles) and downstream docking site (marked by blue heart) complementary to the downstream selector sequences (marked by red saddle shapes) and upstream selector sequences (marked by blue crowns), respectively. The dashed arrow represents the RNA–RNA interaction of upstream or downstream base pairings. The most frequent nucleotides at upstream and downstream docking sites are depicted in red and blue, respectively, while the most frequent nucleotides at the selectors are depicted in red and blue, respectively. The base sequences are shown from 5′ to 3′. **(B,C)** Upstream and downstream selector sequences alignment. The core regions of the downstream or upstream selector sequences are highlighted red or blue, respectively. The most frequent nucleotides at each position of the downstream or upstream selector sequences are complementary to the upstream or downstream docking sites, respectively. Nucleotides of compensatory structural covariations that maintain the base pairing are shaded in green, and their evolutionary intermediates (U-G, G-U) are shaded in yellow. **(D,E)** The secondary structures between upstream or downstream base pairing are shown in *S. oryaze*. The sequences that make up the core region of the stem are highlighted in blue. The upstream and downstream selector sequences are shown in black font; upstream and downstream docking sites are shown in red and blue fonts, respectively.

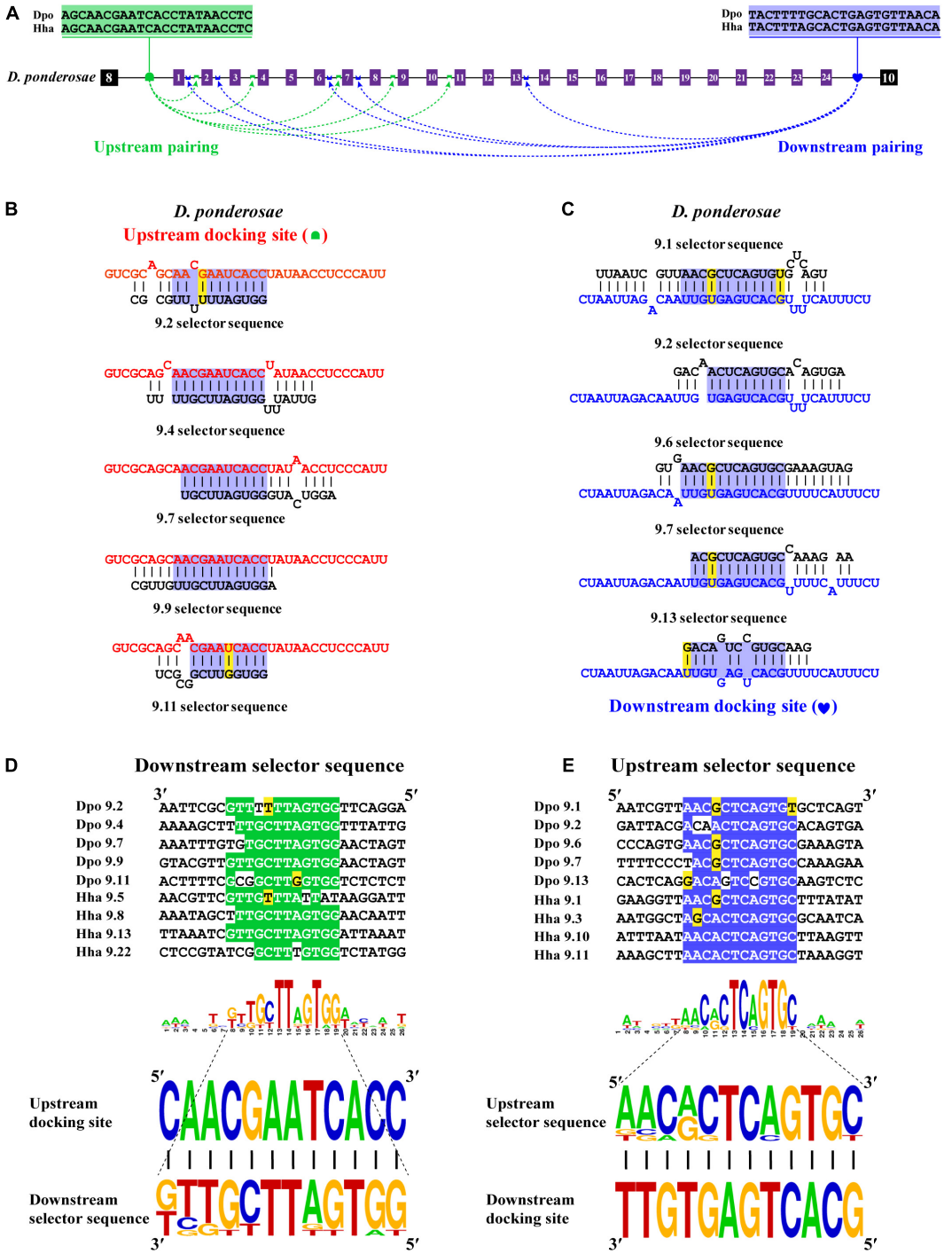


FIGURE 5 | Species-specific upstream docking site in *D. ponderosae* and *H. hampei Dscam1* exon 9. **(A)** Schematic diagram of the *Dscam1* exon 9 of *D. ponderosae*. Upstream docking site (marked by green semicircles) and downstream docking site (marked by blue heart) complementary to the downstream selector sequences (marked by green saddle shapes) and upstream selector sequences (marked by blue crowns), respectively. The dashed arrow represents the RNA–RNA interaction of upstream or downstream pairing. The most frequent nucleotides at upstream and downstream docking sites are depicted in green and blue, respectively. The base sequences are shown from 5′ to 3′. **(B,C)** The secondary structures between upstream or downstream base pairing are shown in *D. ponderosae*. The sequences that make up the core region of the stem are highlighted in blue. The upstream and downstream selector sequences are shown in black font; upstream and downstream docking sites are shown in red and blue fonts, respectively. **(D,E)** Upstream and downstream selector sequences alignment. The core regions of the downstream or upstream selector sequences are highlighted green or blue, respectively. The most frequent nucleotides at each position of the downstream or upstream selector sequences are complementary to the upstream or downstream docking sites, respectively. Nucleotides of compensatory structural covariations that maintain the base pairing are shaded in green, and their evolutionary intermediates (U-G, G-U) are shaded in yellow.

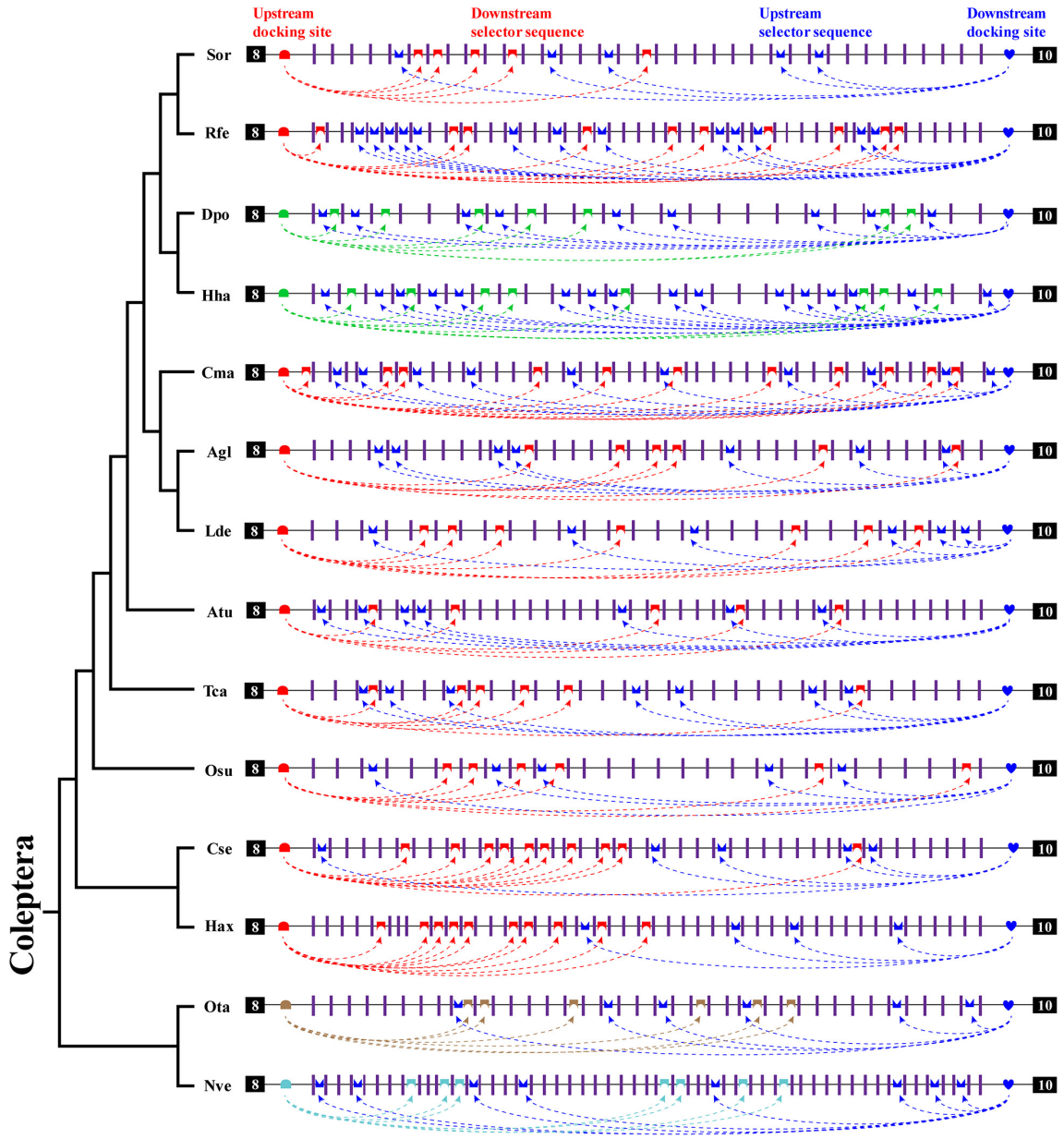


FIGURE 6 | A summary of bidirectional RNA pairing of *Dscam1* exon 9 in Coleoptera species. Overview of the arrangement of the docking site and selector sequence of exon 9 cluster of Coleoptera *Dscam1*. Symbols used are the same as in **Figure 4**, and the exons, introns, docking sites, and selectors are not drawn to scale. Specific upstream or downstream docking sites are shown in different colors. The dashed arrow represents the RNA–RNA interaction of upstream or downstream pairing. The phylogenetic tree of Coleoptera species is shown on the left.

for 300 million years, the docking site of the exon 6 cluster is conserved through the entire Insecta (Graveley, 2005). Recently, a midge-specific docking site but base-pairing conserved in secondary structure level in the exon 6 cluster has been found (Hong et al., 2020), indicating a species-specific docking site in the exon 6 cluster. Our study also predicted the secondary structure in the Coleoptera exon 6 cluster, and most selector sequences were partly located in the exons. This was different from the previous view and had a new inspiration for the identification of the selector sequence. Overall, the docking site of *Dscam1* exons 4 and 9 is clade or species specific and less conserved to exon 6. Therefore, less apparent docking sites make some researchers question the mechanism model of competitive RNA secondary structure regulating the alternative splicing of exons 4 and 9 clusters (Haussmann et al., 2019; Ustaoglu et al., 2019). In this study, through sequence alignment, we identified the clade- or species-specific docking sites of Coleoptera *Dscam1* exon 4 and exon 9 clusters, but the docking site-selector base pairings are conserved in the secondary structure level, which provided more evidence for *Dscam1* exon 4 and 9 clusters

of competitive RNA secondary structure to regulate mutually exclusive alternative splicing.

We have used the Mfold program, which uses a minimum free energy algorithm, to perform RNA secondary structure prediction (Zuker, 2003). The prediction results were similar to some other programs, for example, RNAstructure, a program that calculates the base-pairing probabilities for RNA or DNA sequences by predicting the lowest free energy structures (Mathews et al., 2004), and RNAfold, a program that also uses the minimum free energy algorithm and has a partition function for computing base-pairing probabilities (Bompfunewerer et al., 2008). Although the competitive RNA secondary structures of Coleoptera *Dscam1* were shown in this paper, experimental verification of these predicted secondary structures is difficult due to the limitation of technical means. The main reasons are as follows: First, it is difficult to construct an expression vector due to the large size of the variable exon cluster (30,000 bp in *S. oryzae Dscam1* exon 9). Second, even if the minigene of the variable exon cluster was constructed, the variable exons may not be spliced normally (Graveley, 2005). Third, using the CRISPR-Cas9 system to directly perform mutation *in vivo* seems hard to carry out in practice due to the lack of model organisms in Coleoptera. However, it will be interesting if there are useful systems to solve the experimental verification problems in the future.

Coleopteran insects have not been thoroughly studied, and there is no established genetic manipulation system as mature as the model organism *D. melanogaster*. The experimental operation is difficult. Therefore, all the secondary structures and their effects described in this article are predicted. In the future, it is necessary to conduct systematic research on Coleoptera, explore its genetic research tools, and further experimentally verify the regulatory effect of our proposed RNA secondary structure on alternative splicing.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YJ conceived the project. HD, LL, JS, and YF found the sequence. LL, YF, SZ, and JS predicted the secondary structure. YS, YF, LL, and JS made the evolutionary tree. HD, LL, and JS drew the pictures. XZ, BX, JZ, and FS analyzed the data. HD and LL wrote the manuscript. All the authors discussed the results and commented on the manuscript.

FUNDING

This work was supported by research grants from the National Natural Science Foundation of China (91940303, 31630089, and 91740104) and the Natural Science Foundation of Zhejiang Province (LD21C050002).

ACKNOWLEDGMENTS

We thank the members of Jin Lab for their hard work and for their suggestions and discussion on this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.644238/full#supplementary-material

REFERENCES

Anastassiou, D., Liu, H., and Varadan, V. (2006). Variable window binding for mutually exclusive alternative splicing. *Genome Biol.* 7:R2.

Armitage, S. A., Peuss, R., and Kurtz, J. (2015). Dscam and pancrustacean immune memory - a review of the evidence. *Dev. Comp. Immunol.* 48, 315–323. doi: 10.1016/j.dci.2014.03.004

Bhadra, M., Howell, P., Dutta, S., Heintz, C., and Mair, W. B. (2020). Alternative splicing in aging and longevity. *Hum. Genet.* 139, 357–369. doi: 10.1007/s00439-019-02094-6

Bocak, L., Barton, C., Crampton-Platt, A., Chester, D., Ahrens, D., and Vogler, A. P. (2014). Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Syst. Entomol.* 39, 97–110. doi: 10.1111/syen.12037

Bompfunewerer, A. F., Backofen, R., Bernhart, S. H., Hertel, J., Hofacker, I. L., Stadler, P. F., et al. (2008). Variations on RNA folding and alignment: lessons from Benasque. *J. Math. Biol.* 56, 129–144. doi: 10.1007/s00285-007-0107-5

Bonnal, S. C., Lopez-Oreja, I., and Valcarcel, J. (2020). Roles and mechanisms of alternative splicing in cancer - implications for care. *Nat. Rev. Clin. Oncol.* 17, 457–474. doi: 10.1038/s41571-020-0350-x

Bouchard, P., Smith, A. B. T., Douglas, H., Gimmel, M. L., Brunke, A. J., and Kanda, K. (2017). "Biodiversity of coleoptera," in *Insect Biodiversity*, eds R. G. Foottit and P. H. Adler (Hoboken, NJ: John Wiley & Sons), 337–417.

Brites, D., Brena, C., Ebert, D., and Du Pasquier, L. (2013). More than one way to produce protein diversity: duplication and limited alternative splicing of an adhesion molecule gene in basal arthropods. *Evolution* 67, 2999–3011.

Chen, T. W., Wu, T. H., Ng, W. V., and Lin, W. C. (2011). Interrogation of alternative splicing events in duplicated genes during evolution. *BMC Genomics* 12(Suppl. 3):S16. doi: 10.1186/1471-2164-12-S3-S16

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

Cvetkovska, V., Hibbert, A. D., Emran, F., and Chen, B. E. (2013). Overexpression of Down syndrome cell adhesion molecule impairs precise synaptic targeting. *Nat. Neurosci.* 16, 677–682. doi: 10.1038/nn.3396

Dong, Y., Taylor, H. E., and Dimopoulos, G. (2006). AgDscam, a hypervariable immunoglobulin domain-containing receptor of the Anopheles gambiae innate immune system. *PLoS Biol.* 4:e229. doi: 10.1371/journal.pbio.0040229

Gallego-Paez, L. M., Bordone, M. C., Leote, A. C., Saraiva-Agostinho, N., Ascensão-Ferreira, M., and Barbosa-Morais, N. L. (2017). Alternative splicing: the pledge, the turn, and the prestige : the key role of alternative splicing in human biological systems. *Hum. Genet.* 136, 1015–1042. doi: 10.1007/s00439-017-1790-y

Graveley, B. R. (2005). Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123, 65–73. doi: 10.1016/j.cell.2005.07.028

Graveley, B. R., Kaur, A., Gunning, D., Zipursky, S. L., Rowen, L., and Clemens, J. C. (2004). The organization and evolution of the dipteran and hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. *RNA* 10, 1499–1506. doi: 10.1261/rna.7105504

Hatje, K., and Kollmar, M. (2013). Expansion of the mutually exclusive spliced exome in *Drosophila*. *Nat. Commun.* 4:2460.

Hatje, K., Rahman, R. U., Vidal, R. O., Simm, D., Hammesfahr, B., Bansal, V., et al. (2017). The landscape of human mutually exclusive splicing. *Mol. Syst. Biol.* 13:959.

Hattori, D., Demir, E., Kim, H. W., Viragh, E., Zipursky, S. L., and Dickson, B. J. (2007). Dscam diversity is essential for neuronal wiring and self-recognition. *Nature* 449, 223–227. doi: 10.1038/nature06099

Haussmann, I. U., Ustaoglu, P., Brauer, U., Hemani, Y., Dix, T. C., and Soller, M. (2019). Plasmid-based gap-repair recombineered transgenes reveal a central role for introns in mutually exclusive alternative splicing in Down Syndrome Cell Adhesion Molecule exon 4. *Nucleic Acids Res.* 47, 1389–1403. doi: 10.1093/nar/gky1254

Hong, W., Shi, Y., Xu, B., and Jin, Y. (2020). RNA secondary structures in Dscam1 mutually exclusive splicing: unique evolutionary signature from the midge. *RNA*. 26, 1086–1093. doi: 10.1261/rna.075259.120

Hughes, M. E., Bortnick, R., Tsubouchi, A., Baumer, P., Kondo, M., Uemura, T., et al. (2007). Homophilic Dscam interactions control complex dendrite morphogenesis. *Neuron* 54, 417–427. doi: 10.1016/j.neuron.2007.04.013

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, O. S., Wild, R., et al. (2007). A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 318, 1913–1916. doi: 10.1126/science.1146954

Jin, Y., Dong, H., Shi, Y., and Bian, L. (2018). Mutually exclusive alternative splicing of pre-mRNAs. *Wiley Interdiscip Rev. RNA* 9:e1468. doi: 10.1002/wrna.1468

Kim, H. K., Pham, M. H. C., Ko, K. S., Rhee, B. D., and Han, J. (2018). Alternative splicing isoforms in health and disease. *Pflugers Arch.* 470, 995–1016.

Kise, Y., and Schmucker, D. (2013). Role of self-avoidance in neuronal wiring. *Curr. Opin. Neurobiol.* 23, 983–989. doi: 10.1016/j.conb.2013.09.011

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lee, C., Kim, N., Roy, M., and Graveley, B. R. (2010). Massive expansions of Dscam splicing diversity via staggered homologous recombination during arthropod evolution. *RNA* 16, 91–105. doi: 10.1261/rna.1812710

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7287–7292. doi: 10.1073/pnas.0401799101

Matthews, B. J., Kim, M. E., Flanagan, J. J., Hattori, D., Clemens, J. C., Zipursky, S. L., et al. (2007). Dendrite self-avoidance is controlled by Dscam. *Cell* 129, 593–604. doi: 10.1016/j.cell.2007.04.013

May, G. E., Olson, S., McManus, C. J., and Graveley, B. R. (2011). Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. *RNA* 17, 222–229. doi: 10.1261/rna.2521311

Montes, M., Sanford, B. L., Comiskey, D. F., and Chandler, D. S. (2019). RNA splicing and disease: animal models to therapies. *Trends Genet.* 35, 68–87. doi: 10.1016/j.tig.2018.10.002

Ng, T. H., and Kurtz, J. (2020). Dscam in immunity: a question of diversity in insects and crustaceans. *Dev. Comp. Immunol.* 105:103539. doi: 10.1016/j.dci.2019.103539

Nilsen, T. W., and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi: 10.1038/nature08909

Olson, S., Blanchette, M., Park, J., Savva, Y., Yeo, G. W., Yeakley, J. M., et al. (2007). A regulator of Dscam mutually exclusive splicing fidelity. *Nat. Struct. Mol. Biol.* 14, 1134–1140. doi: 10.1038/nsmb1339

Pandey, M., Stormo, G. D., and Dutcher, S. K. (2020). Alternative splicing during the *Chlamydomonas reinhardtii* cell cycle. *G3 (Bethesda)* 10, 3797–3810.

Pervouchine, D. D., Khrameeva, E. E., Pichugina, M. Y., Nikolaienko, O. V., Gelfand, M. S., Rubtsov, P. M., et al. (2012). Evidence for widespread association of mammalian splicing and conserved long-range RNA structures. *RNA* 18, 1–15. doi: 10.1261/rna.029249.111

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., et al. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671–684. doi: 10.1016/s0092-8674(00)80878-8

Smith, C. W. (2005). Alternative splicing–when two's a crowd. *Cell* 123, 1–3. doi: 10.1016/j.cell.2005.09.010

Soba, P., Zhu, S., Emoto, K., Younger, S., Yang, S. J., Yu, H. H., et al. (2007). *Drosophila* sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization. *Neuron* 54, 403–416. doi: 10.1016/j.neuron.2007.03.029

Suresh, S., Crease, T. J., Cristescu, M. E., and Chain, F. J. J. (2020). Alternative splicing is highly variable among Daphnia pulex lineages in response to acute copper exposure. *BMC Genomics* 21:433. doi: 10.1186/s12864-020-06831-4

Suyama, M. (2013). Mechanistic insights into mutually exclusive splicing in dynamin 1. *Bioinformatics* 29, 2084–2087. doi: 10.1093/bioinformatics/btt368

Ustaoglu, P., Haussmann, I. U., Liao, H., Torres-Mendez, A., Arnold, R., Irimia, M., et al. (2019). Srrm234, but not canonical SR and hnRNP proteins, drive inclusion of Dscam exon 9 variable exons. *RNA* 25, 1353–1365. doi: 10.1261/rna.071316.119

Wang, X., Li, G., Yang, Y., Wang, W., Zhang, W., Pan, H., et al. (2012). An RNA architectural locus control region involved in Dscam mutually exclusive splicing. *Nat. Commun.* 3:1255.

Wang, Y., Bao, Y., Zhang, S., and Wang, Z. (2020). Splicing dysregulation in cancer: from mechanistic understanding to a new class of therapeutic targets. *Sci. China Life Sci.* 63, 469–484. doi: 10.1007/s11427-019-1605-0

Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L., and Clemens, J. C. (2004). Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* 118, 619–633. doi: 10.1016/j.cell.2004.08.021

Woodcock, T. S., Boyle, E. E., Roughley, R. E., Kevan, P. G., Labbee, R. N., Smith, A. B., et al. (2013). The diversity and biogeography of the Coleoptera of Churchill: insights from DNA barcoding. *BMC Ecol.* 13:40. doi: 10.1186/1472-6785-13-40

Xu, B., Meng, Y., and Jin, Y. (2020). RNA structures in alternative splicing and back-splicing. *Wiley Interdiscip Rev. RNA* 12:e1626.

Xu, B., Shi, Y., Wu, Y., Meng, Y., and Jin, Y. (2019). Role of RNA secondary structures in regulating Dscam alternative splicing. *Biochim. Biophys. Acta Gene Regul. Mech.* 1862:194381. doi: 10.1016/j.bbagrm.2019.04.008

Yang, Y., Zhan, L., Zhang, W., Sun, F., Wang, W., Tian, N., et al. (2011). RNA secondary structure in mutually exclusive splicing. *Nat. Struct. Mol. Biol.* 18, 159–168.

Yue, Y., Hou, S., Wang, X., Zhan, L., Cao, G., Li, G., et al. (2017). Role and convergent evolution of competing RNA secondary structures in mutually exclusive splicing. *RNA Biol.* 14, 1399–1410. doi: 10.1080/15476286.2017.1294308

Yue, Y., Yang, Y., Dai, L., Cao, G., Chen, R., Hong, W., et al. (2016). Long-range RNA pairings contribute to mutually exclusive splicing. *RNA* 22, 96–110. doi: 10.1261/rna.053314.115

Zhang, S. Q., Che, L. H., Li, Y., Dan, L., Pang, H., Slipinski, A., et al. (2018). Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nat. Commun.* 9:205.

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

Zipursky, S. L., and Grueber, W. B. (2013). The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* 36, 547–568. doi: 10.1146/annurev-neuro-062111-150414

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

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

frontiers in Genetics

REVIEW
published: 31 March 2021
doi: 10.3389/fgene.2021.626234

LncRNAs: Architectural Scaffolds or More Potential Roles in Phase Separation

Jie Luo[1], Lei Qu[2], Feiran Gao[3,4], Jun Lin[1], Jian Liu[3,4]* and Aifu Lin[2,5]**

[1] Department of Obstetrics and Gynecology, Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [2] College of Life Sciences, Zhejiang University, Hangzhou, China, [3] Department of Respiratory and Critical Care Medicine, The Second Affiliated Hospital, Zhejiang University School of Medicine, Zhejiang University, Hangzhou, China, [4] Zhejiang University-University of Edinburgh Institute (ZJU-UoE Institute), Zhejiang University School of Medicine, Zhejiang University, Haining, China, [5] Breast Center of The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China*

Biomolecules specifically aggregate in the cytoplasm and nucleus, driving liquid-liquid phase separation (LLPS) formation and diverse biological processes. Extensive studies have focused on revealing multiple functional membraneless organelles in both the nucleus and cytoplasm. Condensation compositions of LLPS, such as proteins and RNAs affecting the formation of phase separation, have been gradually unveiled. LncRNAs possessing abundant second structures usually promote phase separation formation by providing architectural scaffolds for diverse RNAs and proteins interaction in both the nucleus and cytoplasm. Beyond scaffolds, lncRNAs may possess more diverse functions, such as functioning as enhancer RNAs or buffers. In this review, we summarized current studies on the function of phase separation and its related lncRNAs, mainly in the nucleus. This review will facilitate our understanding of the formation and function of phase separation and the role of lncRNAs in these processes and related biological activities. A deeper understanding of the formation and maintaining of phase separation will be beneficial for disease diagnosis and treatment.

Keywords: phase separation, lncRNAs, nuclear bodies, signaling transduction, therapeutics treatments

OPEN ACCESS

Edited by:
Xiao Li,
Texas Heart Institute, United States

Reviewed by:
Bo Wen,
Fudan University, China
Xianghui Fu,
Sichuan University, China

****Correspondence:***
Jun Lin
linjun@zju.edu.cn
Jian Liu
jianl@intl.zju.edu.cn
Aifu Lin
linaifu@zju.edu.cn

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *05 November 2020*
Accepted: *25 February 2021*
Published: *31 March 2021*

Citation:
Luo J, Qu L, Gao F, Lin J, Liu J and Lin A (2021) LncRNAs: Architectural Scaffolds or More Potential Roles in Phase Separation. Front. Genet. 12:626234. doi: 10.3389/fgene.2021.626234

INTRODUCTION

The assembly of liquid-liquid phase separation (LLPS) in cells mediates numerous membraneless compartments' formation, such as stress granules (Wheeler et al., 2016; Wang M. et al., 2018; Gui et al., 2019), RNA-protein complexes, termed ribonucleoprotein (RNP) granules (Murakami et al., 2015; Pitchiaya et al., 2019), PGL-1/3 granules (Zhang et al., 2018), nuclear paraspeckles (Fox et al., 2018; Hupalowska et al., 2018; Yamazaki et al., 2018), and receptor clusters (Su et al., 2016). These compartments are involved in various physiological processes and pathological conditions. These two and three-dimensional membraneless organelles have well-defined boundaries, allowing specific biomolecules, such as proteins and nucleic acids, to be concentrated within liquid droplets and exchanged with the surrounding microenvironment (Banani et al., 2017). By creating distinct physical and unique biochemical compartments, phase separation facilitates temporal and spatial control of signaling transduction and biochemical reactions (Nott et al., 2015; Chong and Forman-Kay, 2016; Su et al., 2016). Phase separation transitioning from liquid to gel/solid implicates various

central nervous diseases caused by aberrant aggregation of proteins common in amyotrophic lateral sclerosis (ALS) (Kim et al., 2013; Gasset-Rosa et al., 2019) and frontotemporal dementia (FTD) (Murakami et al., 2015). Dynamic liquid droplets formed by LLPS are believed to be driven by multivalent interactions between biomacromolecules containing intrinsically disordered regions (IDRs)/prion-like domains (PrLDs) or RGG/RG sequence (Kim et al., 2013; Banani et al., 2017; Chong et al., 2018). Those interactions always include charge-charge, pi-pi, and cation-pi interactions (Alberti et al., 2019). Those PrLDs and RGG sequences of RNA binding proteins (RBPs) possess small polar residues and aromatic, positively charged amino acids, which are critical elements for intermolecular interactions (Maharana et al., 2018; Alberti et al., 2019). Those RBPs contribute to the formation of RNP granules and nucleus paraspeckles through interaction with diverse RNAs in the manner of LLPS (Patel et al., 2015; Fox et al., 2018). In addition, LLPS is sensitive to its surrounding environment. Biophysical features of LLPS components (usually specific proteins and nucleic acids) and environmental factors (such as temperature, concentration of salt solution, pH, co-solute, the concentration of other macromolecules, and the modification of phase-separation-related components) have an enormous influence on intermolecular interactions between RBPs and RNAs (Brangwynne et al., 2015; Nott et al., 2015; Reichheld et al., 2017; Franzmann and Alberti, 2019). Post-translational modifications (PTMs), such as phosphorylation (Larson et al., 2017; Zhang et al., 2018), methylation (Qamar et al., 2018; Ryan et al., 2018), ubiquitination (Dao et al., 2018), and SUMOylation (Jin, 2019; Qu et al., 2020) of proteins and m6A modification of RNA (Ries et al., 2019), modulate LLPS formation through regulating protein-protein or protein-RNA interaction, which are affected by the net charge distribution of those molecules.

As a member of phase separation, RNA cooperates with protein partners to drive LLPS formation and modulates the properties of droplets (Huo et al., 2020). Emerging pieces of evidence have reported that RNA not only serves as a scaffold in phase separation due to their abundant secondary structures (Jain and Vale, 2017; Fay and Anderson, 2018; Maharana et al., 2018), but also for their ability to decrease the viscosity of protein components and promote the diffusion of protein components (Elbaum-Garfinkle et al., 2015). Long non-coding RNAs (lncRNAs) are longer than 200 nt in length and unable to code proteins, but they play critical roles in cell metabolism and tumor development, largely depending on their subcellular localization (Zhang et al., 2014). Nuclear lncRNAs regulate transcription, epigenetic modification, and splicing processes of mRNAs (Schmitt and Chang, 2016; Tang et al., 2017). Evidence reveals that lncRNAs regulate mRNA translation and degradation by complementary base pairing and serve as an RNA sponge by interacting with the miRNA in cytosol (Yoon et al., 2013). Our previous studies have revealed that lncRNAs coordinate diverse signal transduction pathways, such as PIP3, HIF1-α, Hippo, Hedgehog, and NF-κB, to promote tumor development (Lin et al., 2016, 2017; Zheng et al., 2017; Sang et al., 2018). Compared to small RNAs, lncRNAs are more capable of providing binding sites for RBPs involved in phase separation (Chujo et al., 2016; Chujo and Hirose, 2017; Fox et al., 2018; Yamazaki et al., 2018). The classical paraspeckles, which are mainly constituted by lncRNA *NEAT1* and numerous RBPs, sequester component proteins and RNAs in the nucleus to mediate gene expression by extensive polymerization and multivalent interaction of LLPS components (Fox et al., 2018; Yamazaki et al., 2018). However, further investigation is needed to understand how lncRNAs coordinate phase separation in different subcellular localization to contribute to diseases (such as degenerative diseases) and tumor development.

This review summarized current advances about phase separation and related lncRNAs in nucleus and cytosol during numerous biological processes (**Figure 1**). We have also summarized the lncRNAs referred to in this review (**Table 1**). Finally, potential therapeutic targets in phase-separation-related lncRNAs and phase separation components during disease development are also summarized.

PHASE SEPARATION

The membrane organelles in eukaryotic cells are well-defined by their membrane-boundaries which provide relatively independent compartments for their specific function (Hyman et al., 2014; Nott et al., 2015; Alberti et al., 2019). For example, endoplasmic reticulum (ER) is involved in the processing of protein and the synthesis of lipids; Golgi apparatus participates in the processing, sorting, and transporting of proteins. Lysosomes function as the cleaning machines for misfolding and pathological proteins; mitochondria provide cellular fuel. However, how do membraneless organelles assemble proteins, nucleic acids, and other molecular components into phase separation? What are the roles of these membraneless organelles in biomolecules metabolic processes, stress sensing, signaling pathways transduction, and gene expression regulation remain largely unknown. Since Hyman and Brangwynne first reported the formation of germline P granules by phase separation in worm embryo cells in 2009 (Brangwynne et al., 2009), the number of studies on phase separation touching myriad cellular functions have increased significantly.

The regulation of gene expression is a prominent event in healthy and diseased states and involves many factors (such as enhancers and coactivators). Recent studies suggested that gene regulation is always accompanied by phase separation assembled by numerous IDR proteins (Guo et al., 2019). Using live-cell super-resolution light-sheet imaging, a previous study found that mediator coactivator coordinates RNA polymerase II (RNA pol II) to regulate the assembly of mediator cluster at enhancer, thus activating gene expression (Cho et al., 2018). Typically, enhancers can activate promoters within the locus (Palstra et al., 2003). Those phase separation-mediated enhancers cause gene bursting expression. Transcriptional factors (TFs) MED1 and BRD4 condensate at super enhancers' (SEs) foci to coactivate gene transcription. This phase separation formed by SEs and TFs confers robust gene expression, which could explain why cancer cells acquire large SEs at driver oncogenes and results in bursting gene expression from a new perspective

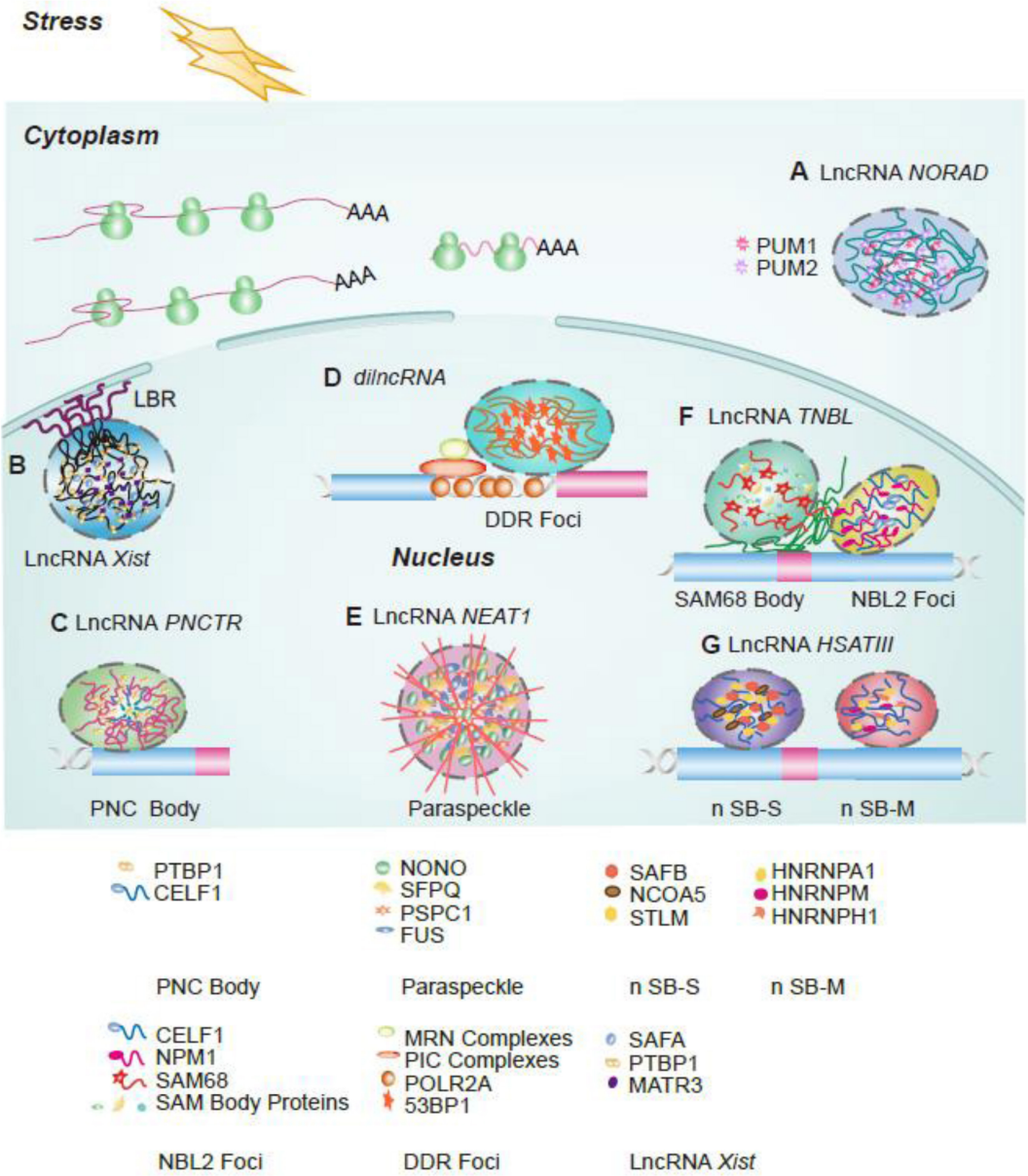


FIGURE 1 | The graphical abstract of phase-separation related LncRNAs involved in cellular function. **(A)** LncRNA *NORAD* functions as a multivalent binding platform for PUM1/2 proteins in cytoplasm; **(B)** LncRNA *Xist* mediates X chromosome silence and subsequently drives interaction between inactivated X chromosome and Lamin-B receptor (LBR); **(C)** LncRNA *PNCTR* sequesters PTBP1 in the perinucleolar compartment (PNC) and modulates splicing regulation function of PTBP1 protein; **(D)** *DilncRNA* synthesized at DSB foci and coordinates DDR proteins to promote the formation of DDR foci to response to DBS; **(E)** LncRNA *NEAT1* functions as scaffolds to recruit CARM1, PSPC1, and p54nrb proteins to regulate cell differentiation and embryo development in paraspeckle; **(F)** LncRNA *TNBL* is accumulated as a perinucleolar aggregate at NBL2 loci and close to SAM68 body and is involved in genome organization, splicing regulation, and mRNA stability, respectively; **(G)** LncRNA *HSATIII* is involved in two nuclear bodies, n SB-M and n SB-S, formation to respond to thermal stress.

(Sabari et al., 2018). The composition of amino acids of TFs' activation domain in mammalian OCT4 and yeast GCN4 is vital for forming phase separation. Phase separation also coordinates multiple signaling pathways (such as estrogen receptor (ER) and Yes-associated protein (YAP) signaling axis) to respond to stress (Boija et al., 2018; Cai et al., 2019). Changes in the components of phase separation often have an impact on their function. Phase separation formed by the histidine-rich domain (HRD) of cyclin T1 and DYRK1A contributes a lot to phosphorylated C-terminal domain (CTD). Disruption of HRD interaction downregulated gene expressions (Lu et al., 2018). Phase separation accumulated at chromatin foci is significantly dependent on the conformation of nucleosomes. A loose conformation of nucleosomes means the activation of chromosomes, while tight condensation suggests the

TABLE 1 | The LncRNAs referenced in this review.

LncRNAs	Subcellular localization	Biological function	References
LncRNA *LINKA*	Cytoplasm	Hyperactivate AKT, HIF1-α signaling pathway, and downregulate antigen presentation related genes to promote drug resistance and immune escaping and remodel glycolysis reprogram of cancer cells.	Lin et al., 2016, 2017; Hu et al., 2019
LncRNA *BRCA4*	Nuclear	Coordinate hippo and hedgehog signaling pathways to aberrantly regulate glycolysis and advance breast cancer development.	Xing et al., 2014; Zheng et al., 2017
LncRNA *CamK-A*	Cytoplasm	Assist the $Ca2^{+}$ signaling pathway to aberrantly regulate glycolysis and remodel tumor microenvironment.	Sang et al., 2018
LncRNA *NEAT1*	Nuclear	Function as a scaffold for paraspeckle components and sequester specific proteins (such as CARM1) promotes cell differentiation and embryo development. Attenuate activation of p-53 and confer cancer cell drug resistance (LLPS).	Chen and Carmichael, 2009; Adriaens et al., 2016; Fox et al., 2018; Hupalowska et al., 2018; Yamazaki et al., 2018
LncRNA *MAYA*	Cytoplasm	Mediate heterodimerization of ROR1 and HER3 and promote activation of YAP, thus facilitating breast cancer bone metastasis.	Li et al., 2017
LncRNA *HOTAIR*	Nuclear	Assist PRC2 complex to recruit to histone and be responsible for the silence transcription of HOXD gene.	Rinn et al., 2007
LincRNA	Nuclear	Bind to a series of chromatin-modifying proteins to maintain the pluripotent state of ESCs.	Guttman et al., 2011
LncRNA *NORAD*	Both nuclear and cytoplasm	Assemble a topoisomerase complex at targeted chromatin foci to stabilize genome (Nuclear). Function as a multivalent binding platform for PUM1/2 proteins, and thus maintaining genomic stability (LLPS).	Lee et al., 2016; Tichon et al., 2016; Munschauer et al., 2018
LncRNA *Xist*	Nuclear	Recruit epigenetic factors to chromosome loci and mediate X chromosome silence (LLPS).	Heard and Disteche, 2006; Moindrot et al., 2015; Cerase et al., 2019
LncRNA *TCF7*	Nuclear	Recruit epigenetic factors SWI/SNF promoting TCF expression, thus activating Wnt pathway to promote self-renewal of liver cancer stem cells.	Wang et al., 2015
eRNA	Nuclear	Bind to multiple TFs and coactivator to alter the chromosomal architecture and thus regulating gene expression.	Li et al., 2013; Liu et al., 2014; Pnueli et al., 2015
LncRNA *GATA6-AS*	Nuclear	Recruit and inactive epigenetic factor LOXL2 and regulate endothelial gene expression and angiogenic activity in responding to hypoxia.	Neumann et al., 2018
LncRNA *AGPG*	Both nuclear and cytoplasm	Stabilize PFKFB3 by blocking its ubiquitination and degradation thus promoting glycolysis in cancer cells.	Liu et al., 2020
lncRNA *HOXB-AS3*	Both nuclear and cytoplasm (according to genecard)	Encode peptide HOXB-AS3 regulating splicing of pyruvate kinase M (PKM) and thus reprogramming glucose metabolism.	Huang et al., 2017
LOC100507537/LINC00948	Sarcoplasmic reticulum membrane (according to genecard)	Encode peptide activating the SERCA pump to promote $Ca2^{+}$ up-taking into sarcoplasmic reticulum (SR) and enhancing muscle contractility.	Anderson et al., 2015; Nelson et al., 2016
LncRNA *meiRNA*	Nuclear	Function as an architectural scaffold promoting the formation of sme2 chromosomal loci (phase droplet) and mediate pairing of homologous chromosomes (LLPS).	Shichino et al., 2014; Ding et al., 2019

(Continued)

TABLE 1 | Continued

LncRNAs	Subcellular localization	Biological function	References
LncRNA *HSATIII*	Nuclear	Function as an architectural scaffold interacting with two hnRNPs to promote nuclear stress bodies formation upon thermal stress exposure (LLPS).	Aly et al., 2019
LncRNA *PNCTR*	Nuclear	Function as an architectural scaffold sequestering PTBP1 in the perinucleolar compartment, thus modulating splicing of PTBP1 protein and promoting cancer cell survival (LLPS).	Yap et al., 2018
LncRNA *TNBL*	Nuclear	Accumulate as a perinucleolar aggregate at NBL2 loci and close to SAM68 body and thus responding to nuclear functions and RNA metabolism (LLPS).	Dumbovic et al., 2018
DilncRNA	Nuclear	Is synthesized at DSB foci and coordinates DDR proteins to promote the formation of DDR foci to respond to DBS (LLPS).	Pessina et al., 2019
LncRNA *MajSAT*	Nuclear	Functions as a scaffold promoting HP1α and SAFB to form PCH foci in Pericentromeric heterochromatin (LLPS).	Huo et al., 2020
LncRNA *TERRA*	Nuclear	Considering the enrichment of LncRNA *TERRA* in APB and interaction between LncRNA *TERRA* and epigenetic modification factors and RBPs, LncRNA *TERRA* may also play a functional role in the telomere foci by providing a platform for multiple proteins interaction (LLPS).	Min et al., 2019

formation of heterochromatin. Heterochromatin protein 1 (HP1) is known to finely tune heterochromosome phase separation by participating in weak multivalent interaction of nucleosomes (Larson et al., 2017; Sanulli et al., 2019). H1 histone and the 10n + 5 inter-nucleosome spacing promotes the phase separation of chromatin and decreases dynamics in droplets (Gibson et al., 2019). Those models of heterochromatin formation provide a new perspective to understand phase separation in regulating the conformation of chromatin. Regulation of gene expression by phase separation broadens our understanding of the mechanism of aberrant expression at the transcriptional level in numerous diseases, facilitating the development of new strategies to identify key components involving the formation and maintenance of phase transition. A novel CRISPR-Cas9-based optogenetic technology was used to explore the formation of droplets impacted by the chromatin microenvironment. This study suggested that phase separation is preferentially formed at low-density genomic regions and promotes genomic rearrangements, thus contributing to the activation of gene expression. On the contrary, at high-density genomic regions, small droplets ultimately dissolve, contributing to the disappearance of phase separation (Shin et al., 2018). These pieces of evidence indicated that the structure of genome and phase separation affected each other, both of which have an enormous impact on gene expression. The existence of phase separation could explain the aberrant patterns of gene expression well.

Phase separation transition from a liquid to a gel or solid leads to degenerative neurological diseases (Wang and Zhang, 2019). Heterogeneous nuclear ribonucleoproteins (hnRNPs) containing IDRs or PrLDs, such as FUS, hnRNPA1, or TAR DNA-binding protein 43 (TDP-43), are found rich in many aging-associated diseases (Kim et al., 2013; Patel et al., 2015; Gui et al., 2019; Mann et al., 2019). Tau droplets formed by phosphorylated or mutant Tau with IDRs undergoing LLPS contributes to Alzheimer's disease (Wegmann et al., 2018). Fused in sarcoma (FUS) is an RNA-binding protein involved in RNA transcription, splicing, transporting, and translation. With classical IDR and low-complexity domain (LCD), FUS protein transitions from a liquid to aggregated state, promoting LLPS formation at the sites of DNA damage, which is associated with ALS (Patel et al., 2015). As membraneless organelles, phase separation can sequester specific components to accelerate or inhibit unique cellular function, and thus advance disease development. Mislocalization and aberrant aggregation of misfolded TDP-43 sequester importin-α and Nup62 in the cytoplasm. Depletion of importin-α and Nup62 in the nucleus induces RnaGap1, Ran, and Nup107 mislocalization, thus promoting cell death and causing advanced ALS and FTD (Gasset-Rosa et al., 2019). Phase separation can also contribute to the development of degenerative neurological diseases. Degenerative neurological-disease-related mutations can also affect the formation of phase separation. Recent studies reported that ALS/FTD related mutation-induced FUS phase transition from liquid droplets to irreversible hydrogels, which impairs RNP function and advances disease (Murakami et al., 2015; Patel et al., 2015). Similarly, the ALS-related mutations in the TDP-34 C-terminal domain (CTD) disrupt phase separation and impair interaction within the phase droplets, which promotes

LLPS transition into solid aggregation, thus aggravating the ALS condition (Conicella et al., 2016). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 also contribute to ALS (Kim et al., 2013). Elucidation of the exact mechanism involved in the molecular properties, formation, regulation, and function of membraneless organelles can help us explore novel therapeutic approaches to treating aging-related disorders. Optogenetic approaches used in controlling phase separation formation of TDP-43 reveals that LCD of TDP-43 are competitively bound by RNA. And oligonucleotides composed of the TDP-43 target sequence can moderate the neurotoxicity caused by aggregation of TDP-43 (Mann et al., 2019). Dysregulation of phase separation in aging-related protein accelerates the malignant transition, but is not a one-way process. Extensive exploration of those processes helps us better understand the development of aging-related diseases.

LncRNAs IN CELL BEHAVIOR

Nearly 98% of human genome encodes as non-coding RNAs (ENCODE Project Consortium et al., 2007; Schmitt and Chang, 2016). For such a large amount of non-coding RNAs (ncRNAs), their cellular function has intrigued many researchers. According to the size, ncRNAs are divided into small ncRNAs and long ncRNAs (Brosnan and Voinnet, 2009; Liu et al., 2019). LncRNAs are poorly conserved in terms of their nucleotide sequences, even though they can be found in many species (Johnsson et al., 2014; Beermann et al., 2016). Secondary structures of lncRNAs enable them to interact with DNAs, proteins, and RNAs, allowing them to participate in multiple cellular processes (Fernandes et al., 2019).

Emerging evidence has revealed that different subcellular localizations of lncRNAs engage in numerous biological processes, including regulation of gene transcription, chromatin remodeling, cancer-related signaling pathways, and organism development (Li et al., 2017; Zheng et al., 2017; Sun et al., 2018; Sarropoulos et al., 2019). Nuclear-localized lncRNAs are involved in transcriptional and post-transcriptional modification and chromatin organization (Sun et al., 2018). The intended transcriptional regulation function of lncRNAs largely relies on multiple interactions between lncRNAs and other molecules (such as DNA, proteins, and RNAs). LncRNA *HOTAIR* assists the PRC2 complex to accumulate at histone and is responsible for the silence of HOXD gene (Rinn et al., 2007). Researchers identified dozens of lncRNAs involved in binding to a series of chromatin-modifying proteins to maintain the pluripotent state of stem cells (Guttman et al., 2011). LncRNA *NORAD* assembles a topoisomerase complex at targeted chromatin foci to stabilize the genome (Munschauer et al., 2018). Classical lncRNA *Xist* mediates X-chromosome inactivation by recruiting protein complexes to repress epigenetic marks and encompass the X-chromosome (Heard and Disteche, 2006). LncRNA *TCF7* recruits SWI/SNF5 complexes to TCF7 promoter to mobilize nucleosomes and remodel chromatin conformation, promoting liver cancer stem cells self-renewal (Wang et al., 2015). All studies mentioned above suggested that lncRNAs are vital for gene expression at the epigenetics level. LncRNAs also act as local regulators to influence the expression of nearby genes by *cis* regulation (Guil and Esteller, 2012; Engreitz et al., 2016). Enhancer RNAs (eRNAs) transcribed from bidirectional ncRNA can bind to multiple TFs and coactivators to alter the chromosomal architecture and regulate gene expression (Li et al., 2013; Liu et al., 2014; Pnueli et al., 2015). The emerging roles of eRNAs significantly extend our understanding of the function of gene transcription regulated by lncRNAs.

Dysregulation of lncRNAs in cells and tissues is associated with malignant transformation and various pathological processes (Xing et al., 2014; Neumann et al., 2018), which is always coordinated by multiple classical signaling pathways. Our previous studies have suggested that lncRNA *LINK-A* is involved in breast cancer drug resistance (Lin et al., 2017), hypoxia (Lin et al., 2016), and immunosuppressive microenvironment (Hu et al., 2019). LncRNAs *CamK-A*, *BRCA4*, and *AGPG* wires up NF-kB (Engreitz et al., 2016), Hippo and Hedgehog (Zheng et al., 2017), and PFKFB3 glycolytic enzyme complexes (Liu et al., 2020), respectively, to remodel glucose metabolism and tumor microenvironment, promoting tumor development. LncRNAs are characterized to have specific tissue distribution, which implies their functional role in development and differentiation. Through genome-wide analysis, Luo et al. (2016) found that divergent lncRNAs regulate about 168 genes coding transcription factors and developmental regulators in embryonic stem cells (ESCs), which implies lncRNAs may be developmentally regulated (Schmitz et al., 2016). The developmental lncRNAs atlas constructed by Sarropoulos et al. (2019) revealed that lncRNAs show species specificity and dynamic expression pattern from early organogenesis to adulthood suggesting that the time, lineage-, and organ-specific lncRNAs are responsible for specific functions during organogenesis and organism development. The functions of lncRNAs are much more than what has been mentioned above. Recent advances in deep-sequencing technologies have identified that some lncRNAs have the ability to encode functional peptides (Anderson et al., 2015; Nelson et al., 2016; Huang et al., 2017). And more interestingly, coordination of the phase separation formation by lncRNAs has been reported in many recent studies.

LncRNAs AND PHASE SEPARATION

The essential features of phase separation are mainly determined by their components. Phase separation related to the regulation of gene expression always takes place in the nucleus. Nuclear bodies, clustering factors, super-enhancers, and chromatin foci are often related to phase separation and transcriptional regulation. Those phase separation membraneless organelles are storage compartments for many RNAs and RNA binding proteins. RNAs involved in the formation of phase separation function as scaffolds and eRNA, whereas phase separation, in turn, impacts the behavior of RNAs, such as synthesis (Pefanis et al., 2015; Nair et al., 2019). The nucleoplasm is a natural pool abundant with diverse membraneless nuclear bodies regulating gene expression (Chujo and Hirose, 2017; Ninomiya and Hirose, 2020), especially

paraspeckle and chromosome loci. Those nuclear bodies are usually composed of multiple RNAs and RBPs containing PrLD and RGG sequence (Van Treeck and Parker, 2018; Alberti et al., 2019). In phase separation nuclear bodies, RNAs critically regulate the phase behavior of RBPs with PrLD. Different RNA/protein ratios exert different influences on phase separation transition. To some extent, RNAs act as a buffer in the nucleus where high RNA concentrations keep RBPs soluble (Van Treeck and Parker, 2018). Changes at RNA levels or RNA binding abilities of RBPs cause aberrant phase transitions (Maharana et al., 2018). This makes us consider that RNAs in phase separation can competitively bind with proteins containing IDR, which attenuates protein self-aggregation. This implies that RNAs with a bigger size, especially lncRNAs, may be more efficient in buffering phase separation.

In addition to the buffering function in the nucleus, many lncRNAs often serve as scaffolds for nuclear bodies' formation. LncRNAs act as seeds to recruit specific component proteins by RNA-proteins interactions (Fox et al., 2018). Those RBPs always recruit additional proteins to induce the formation of LLPS and control gene expression under certain stimulations. Among those nuclear bodies, paraspeckle is a sound model system with well-defined RNAs and protein components for the study of phase separation (Fox et al., 2018). LncRNA *NEAT1* has good architectural functions to provide a scaffold for multiple RNA-binding proteins (RBP) in paraspeckles construction (Adriaens et al., 2016). Gene expression is affected by the size and number of paraspeckles, which can sequester specific RBPs and/or RNA away from nucleoplasm to achieve the regulation (Chen and Carmichael, 2009). The paraspeckles formation is similar to cytoplasmic stress granules, which are another membraneless organelle (Fox et al., 2018). Both paraspeckles and stress granules can respond to cellular stress and function by sequestering specific components to regulate stress response-related gene expression. Those membraneless organelles seem to be more flexible than compartmentalized organelles during stress response due to their dynamic disassembling and assembling. The aberrant gene expression in paraspeckle is often associated with cancer progression (Adriaens et al., 2016). Another typical transcriptional element enhancer is also responsible for gene bursting transcription. Phase separation model suggested that super-enhancers (SEs) consisted of cluster enhancers involved in high transcriptional activity of related genes (Pefanis et al., 2015; Hnisz et al., 2017). In the SEs foci, phase separation regulates the degradation and accumulation of eRNAs, which finally affects the stability of the genome (Pefanis et al., 2015). The function of eRNAs and SEs in phase separation provides us new insights into the regulation of gene expression. In addition to the regulation of nuclear body formation by lncRNAs, delineating the phase behavior mediated by lncRNAs beyond the nucleus can shed light on the impact of cytoplasm condensations on signaling transduction and cellular metabolism. LncRNA *NORAD* retains PUM1/PUM2 protein in the cytoplasm to form RNP granule, leading to chromatin instability in response to DNA damage. In this study, LncRNA *NORAD* functions as a platform to sequester PUM1/PUM2, negatively regulating PUMILIO activity in the cytoplasm. This leads to elevated key mitotic, DNA repair, and recruitment of DNA replication factors. In this RNP granule, lncRNA *NORAD* may coordinate interferon response pathway proteins IFIT1/2/3/5 to regulate this process (Lee et al., 2016; Tichon et al., 2016). Although little is known about the role of lncRNAs in cytoplasm phase separation, one can speculate multiple potential functions of lncRNAs in forming and maintaining phase separation and many other biological processes.

LncRNAs Modulate Phase Separation in Nucleus

In mammalian cells, there are various nuclear bodies. They are mainly involved in the regulation of gene expression by transcriptional epigenetic modification. Chromosome homologous pairing and separation, chromatin remodeling, and RNA splicing are common events in the nucleus, often mediated by phase separation. Many nuclear bodies are well-defined by the enrichment of specific proteins and RNAs (Ninomiya and Hirose, 2020). Beyond function as architectural RNAs, lncRNAs also serve as eRNAs that exist in phase separation droplets (Chujo et al., 2016). The mechanism of how those nuclear bodies exert their functions remains poorly understood. There may be three reasons. First, those nuclear bodies function as a reaction tank sequestering specified molecules, such as enzymes and their substrates. Second, they act as a sequestering compartment, which can condensate specific molecules and protect them from degradation, or sequester from nucleoplasm, to impair their function. Third, they can form an organizational hub that anchors chromatin loci to remodel chromatin and regulate gene expression.

Paraspeckle

Paraspeckle was first reported in 2002 as a marker of paraspeckle component proteins 1(PSPC1) and subsequently found to be mainly localized in mammal cell nuclei (Fox et al., 2002). In addition to PSPC1, paraspeckle consists of over 40 different proteins and the structure lncRNAs *NEAT1* (Mao et al., 2011; Fox et al., 2018). *NEAT1* depletion completely abolished the formation of paraspeckle (Sasaki et al., 2009; Shevtsov and Dundr, 2011). PSPC1 was first reported to be enriched in paraspeckle, but later it was reported that PSPC1 together with NONO and SFPQ were dispensable for paraspeckle formation (Sasaki et al., 2009; Naganuma et al., 2012). EM and super-resolution microscopy have revealed that paraspeckle is a spherical shape with a shell and core. 3′ and 5′ ends of lncRNA *NEAT1* are extended out of paraspeckle in the form of bundles (West et al., 2016). Once formed, paraspeckle sequestered specific RNAs and proteins to alter the levels of those components, changing the cellular processes (Prasanth et al., 2005; Chen and Carmichael, 2009). Paraspeckle in the nucleus participates in many cellular processes, usually related to stress response and cancer. P53 regulates the transcription of *NEAT1*, which promotes the formation of paraspeckle and confers breast cancer cell drug resistance (Adriaens et al., 2016). High level PSPC1 expression in cancer cells activates the TGF-β pathway and promotes metastasis (Salvador and Gomis, 2018). Recent studies link paraspeckle to mitochondrial homeostasis against

the stress response. Classical paraspeckle-mitochondria crosstalk provides a nice model for understanding the role of *NEAT1* and paraspeckle in cancer and neurodegeneration (Nishimoto et al., 2013; Adriaens et al., 2016; Fox, 2018; Wang Y. et al., 2018).

As a classical nuclear body, paraspeckles are involved in gene expression regulation and retention of mRNAs and proteins (Hirose et al., 2014; Wang Y. et al., 2018). Recent studies wired paraspeckles with phase separation and found paraspeckle are more likely to form droplets (Fox et al., 2018). Previous studies revealed that many paraspeckle proteins (such as RBM14 and FUS) containing IDR are responsible for phase separation formation (Hennig et al., 2015; Patel et al., 2015; West et al., 2016; Shin et al., 2017). During preimplantation development of mouse embryo, activation of histone by the histone coactivator associated arginine methyltransferase 1 (CARM1) is necessary for the upregulated expression of a subset of pluripotency genes. The function of CARM1 is maintained by paraspeckle integrity and dependent on lncRNA *NEAT1* and NONO (Hupalowska et al., 2018). A specific sequence in 3′-UTR of RNA makes it prone to be bound with paraspeckle components. The latest study reported that paraspeckle lncRNA *NEAT1* and four major proteins are responsible for retaining circadian mRNA to regulate gene expression at post-transcriptional level (Torres et al., 2016). The size and number of paraspeckles significantly affect gene expression. In contrast, the assembly of paraspeckle is mainly determined by the level of *NEAT1* and components proteins such as SFPQ and FUS. An earlier study revealed that the bigger a paraspeckle becomes, the more SFPQ is needed. The decreasing level of SFPQ in the nucleus altered the targeted gene expression, which also occurred in other nuclear bodies (Chen and Carmichael, 2009; Imamura et al., 2014; Wu et al., 2016). As the paraspeckle component proteins, such as SFPQ and NONO, are involved in pri-miRNA processing, sequestering both of those proteins can affect miRNA processing (Jiang et al., 2017). Studies indicate that expression and mutation of core paraspeckle structure *NEAT1-2* are often related to multiple cancers (Fujimoto et al., 2016; Rheinbay et al., 2017). All of this evidence indicates that phase separation in paraspeckle can regulate gene expression and RNA-related processes, promoting disease and cancer development.

Paraspeckles are involved in various cellular processes. The core structure of lncRNA *NEAT1* is responsible for building paraspeckles. The protein components of paraspeckles usually contain IDR, which promotes the formation of phase separation. Therefore, investigating the structural *NEAT1* RNA or phase separation proteins in paraspeckles will help develop new strategies for targeted therapies.

Chromatin Foci

In addition to phase separation related to paraspeckles involved in RNAs and proteins, phase separation formed at chromatin is common to regulate gene transcription and chromosome segregation. Phase separation formed at chromatin is usually affected by the surrounding microenvironment, such as nucleosome state and modification of histone. Exposing histone tails of nucleosome makes the interaction of inter-nucleosome tighter and thus promotes the phase separation formation by HP1 (Gibson et al., 2019; Sanulli et al., 2019). Phase separation prefers to form at low-density chromatin compared to high-density regions, referred to as heterochromatin. This preference caused by phase separation usually results in reorganization of chromatin and thus alters gene expression (Gibson et al., 2019). Many studies suggest that lncRNAs in specific chromatin loci also function as scaffolds to recruit chromatin-modifying complexes, promoting the epigenetic regulation of gene expression. X-chromosome inactivation (XIC) is a critical epigenetic mechanism for balancing gene dosage between XY and XX in eutherian mammals. Recent studies suggest that the process of X-chromosome inactivation is involved in phase separation mediated by *Xist* (Cerase et al., 2019). *Xist* drives phase separation by enriching chromatin remodeling factors, such as Spen, Ptbp1, HnrnpK, and PRC1/2 IDR-protein (Moindrot et al., 2015). This recruitment leads to deacetylation of histone and chromatin condensation. After inactivation, the X-chromosome is sequestered by specific interactions between *Xist* and Lamin-B receptor (Cerase et al., 2019). Pericentromeric heterochromatin (PCH) formation is also a phase separation process, mainly mediated by HP1α and lncRNA *MajSAT*. In these PCH foci, the R/G-rich domain of RNP protein SAFB is responsible for recognizing lncRNA *MajSAT*. SAFB-*MajSAT* interaction functions as a scaffold for the 3D organization of heterochromatin (Huo et al., 2020). What is impressive in this study is that, although the SAF family proteins SAFA/B have a similar functional domain, only SAFB confers the formation of PCH foci. The factor contributing to this difference is interesting for future studies. Telomeres, a special part of the chromosome, consist of DNA-protein complexes involved in chromosome end protection. It has been reported that many cancer cells can escape senescence by altering the length of telomeres, which is also termed alternative lengthening of telomeres (ALT). LncRNA TElomeric Repeat-containing RNA (LncRNA *TERRA*), transcribed at telomeres, is a main hallmark of ALT (Roake and Artandi, 2017; Bettin et al., 2019). Evidence has indicated that lncRNA *TERRA* acts as a scaffold to promote the recruitment of epigenetic modification factors (such as PRC2 and HP1) and diverse RBPs (such TLS/FUS and TRF2) (Deng et al., 2009; Takahama et al., 2013; Montero et al., 2018), which always appear in numerous phase separations. Simultaneously, lncRNA *TERRA* was reported to be enriched in ALT-associated PML body (APB), one of the promyelocytic leukemia (PML) bodies, which are nuclear membraneless organelles formed by LLPS and are involved in mitosis by recruiting multivalent proteins with small ubiquitin-like modification (SUMO) sites and SUMO-interacting motifs (SIMs) (Arora et al., 2014; Banani et al., 2016). A recent study reported an artificial model system where APB could form telomere cluster condensates by LLPS *in vivo*. During this process, BLM helicase and RAD52 are responsible for the formation of telomeres' foci (Min et al., 2019). Considering the enrichment of lncRNA *TERRA* in APB and the interactions between lncRNA *TERRA* and epigenetic modification factors and RBPs, we speculate that lncRNA *TERRA* may also play a functional role in the telomere foci. However, the detailed mechanism needs to be further investigated. Of note, why both lncRNA *MajSAT* and

LncRNA *TERRA*, repetitive RNAs, are preferred to be selected to participate in the formation of phase separation needs further exploration.

Corrective pairing and segregation of homologous chromosomes in meiosis are critical to producing haploids. LncRNA *sem2 RNA* helps Smp (sme2RNA-associated protein) protein form three chromosome loci and determine the specificity of chromosomal loci for fusion. It indicates the importance of Smp proteins in the accumulation of lncRNA and the critical role of lncRNA-mediated chromosome homologous pairing in *Schizosaccharomyces pombe* (Ding et al., 2019). In the fission yeast, the *meiRNA* plays a crucial role in recognizing and pairing homologous chromosomes during meiotic prophase. LncRNA *meiRNA* recruits Mmi1 protein to sem2 dot to promote meiosis, which is pivotal for selective elimination of meiosis-specific transcripts (Shichino et al., 2014). Enhancers and SEs are good partners to explain the bursting expression of genes. Recent studies reveal that enhancers, SEs, and eRNA may be involved in phase separation. Transcribed from bidirectional ncRNA, eRNAs act as enhancers and alter the chromosomal architecture during the transcription process (Li et al., 2013; Liu et al., 2014; Pnueli et al., 2015). Under the acute stimulation of 17β-estradiol (E2), eRNA and several TFs provide a conductive microenvironment for the assembly of enhancer RNA–dependent ribonucleoprotein (eRNP), regulating signal-inducible transcription (Nair et al., 2019). eRNAs wire DNA and TFs together and thus promote gene expression. RNA-exosome regulates the degradation and terminates transcription of eRNA lncRNA *CSR*, which coordinates SEs to promote the stability of chromatin in long range (Pefanis et al., 2015). eRNAs highly expressed in many cancers may be responsible for drug resistance by promoting related gene expression, which indicates that certain eRNAs can be diagnostic markers and targets for cancer treatment (Zhang et al., 2019).

The diverse functions of lncRNA combined with phase separation in chromatin loci show a spectacular panoramic view for understanding the regulation of gene expression at the transcriptional level. The inactivation and segregation of chromatin and gene bursting expression can be well-interpreted by the phase separation model. The chromatin loci formed by phase separation through specific proteins and lncRNA provide us with new strategies to explore the abnormal cellular processes and develop novel therapy.

Nuclear Stress Bodies

The nucleoplasm is a natural pool for diverse nuclear bodies to regulate gene expression (Chujo and Hirose, 2017; Ninomiya and Hirose, 2020). Nuclear bodies accumulated at specific nucleus sites affect the biogenesis, maturation, storage, and sequestration of specific proteins and RNAs, thus altering cellular events to respond to stress stimuli. Under thermal stress, lncRNA *HSATIII* acts as the structural scaffold for the HNRNPM and SAFB foci formation and retains numerical RBPs to regulate gene expression (Aly et al., 2019). To respond to DNA double-strand breaks (DSB), damage-induced long non-coding RNA (*dilncRNA*) is synthesized at DSB foci, also called DNA-damage-response (DDR) foci. *DilncRNA*, together with DDR proteins, such as 53BP1, promotes the formation of DDR foci to regulate the transcriptional activity of genes mediating the DSB signal pathway (Pessina et al., 2019). In a wide range of cancers, lncRNAs and related RBPs are often aberrantly transcribed. LncRNA *PNCTR* recruits RBP PTBP1 to form a nuclear body called peri-nucleolar compartment (PNC), where lncRNA *PNCTR* modulates cellular localization of PTBP1 by changing the splicing of PTBP1, an activator of the intrinsic branch of apoptosis. The alteration of PTPB1 cellular localization results in its inhibition and thus promotes cell survival (Yap et al., 2018). In colon cancer, upregulated lncRNA *TNBL* is accumulated at the subset of NBL2 loci and forms dense aggregates, which sequesters SAM68 RBPs and nucleic acids. This SAM68 nuclear body may disrupt nuclear organization (Dumbovic et al., 2018). LncRNAs involved in many events in nuclear bodies can enrich our insights to better understand the function of lncRNAs in phase separation.

PROSPECTIVE

This review mainly summarized the current findings on phase separation and the potential roles of phase-separation related lncRNAs. The formation of phase separation involves multiple molecules, such as RNAs, proteins, and related chromatin. Maintenance of phase separation relies on its surrounding environments, such as pH, temperature, and the concentration of salt solution. Sometimes phase transition is largely determined by the sequence of RNAs, proteins, and the PTM of proteins. Phase separation has expanded our understanding of biochemical reactions and biological processes in membraneless organelles. LncRNAs mainly function as architectural scaffolds for diverse RNA and protein interaction in this process. Phase separation coordinating lncRNAs in multiple nuclear bodies are mainly involved in regulating gene expression, chromatin remodeling, RNA splicing, and homologous chromosome separation in the nucleus. However, lncRNAs involved in cytosolic phase separation are less reported. Several studies have revealed that phase-separation related lncRNAs in cytoplasm participate in signaling transduction (Lee et al., 2016; Tichon et al., 2016). This evidence inspires us to explore more about cytosolic lncRNAs-mediated phase separation. Combining the function of lncRNAs and phase separation together, current studies on both are only the tip of the iceberg. Major questions have yet to be answered in these emerging fields about phase separation and lncRNAs. The most concerning problem is identification of the factors that confer the special components in phase droplets. Proteins contained with LDR, PrLD, or RNA with a repetitive sequence are more likely to form phase separation. Maybe the distribution of net charge and the advanced structure of RNAs and proteins are major factors, which have a great influence on multivalent interactions. Other environmental factors, such as pH, temperature, and the concentration of salt solution, are also important for phase separation formation. The second problem is how the subcellular localization of lncRNAs and phase separation-related proteins affects the phase separation formation. It seems that more functional phase separated droplets

tend to form in the nucleus, which is mostly related to the formation of heterochromatin. What factors contribute to this preference needs to be further elucidated. Numerous nuclear bodies exert different roles in gene expression and epigenetic regulation. Why different lncRNA are selected in different functional phase separation droplets needs to be further explored. Moreover, which factors and signaling pathways are involved in the dynamically assembled and disassembled phase separation droplets upon different environmental stress is of significance. It is also of great importance to precisely identify the role of lncRNAs in sensing stress stimulations, signal transduction, and maintenance of phase separation. Such discoveries will help better understand and develop better therapeutic treatments for phase-separation related diseases.

AUTHOR CONTRIBUTIONS

AL contributed to the study design and data analysis. JLuo wrote the manuscript. LQ and FG contributed to the figure and table design. AL, JLiu, and JLin edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (81872300 and 81672791) and the Zhejiang Provincial Natural Science Fund for Distinguished Young Scholars of China (LR18C060002) to AL.

ACKNOWLEDGMENTS

We acknowledge the funding support by the National Natural Science Foundation of China (81872300 and 81672791), the Zhejiang Provincial Natural Science Fund for Distinguished Young Scholars of China (LR18C060002) to AL, and we also acknowledge the authors of the publications that will be cited in the scope of this review.

REFERENCES

Adriaens, C., Standaert, L., Barra, J., Latil, M., Verfaillie, A., Kalev, P., et al. (2016). p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat. Med.* 22, 861–868. doi: 10.1038/nm.4135

Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* 176, 419–434. doi: 10.1016/j.cell.2018.12.035

Aly, M. K., Ninomiya, K., Adachi, S., Natsume, T., and Hirose, T. (2019). Two distinct nuclear stress bodies containing different sets of RNA-binding proteins are formed with HSATIII architectural noncoding RNAs upon thermal stress exposure. *Biochem. Biophys. Res. Commun.* 516, 419–423. doi: 10.1016/j.bbrc.2019.06.061

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., and McAnally, J. R. (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

Arora, R., Lee, Y., Wischnewski, H., Brun, C. M., Schwarz, T., and Azzalin, C. M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* 5:5220.

Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298. doi: 10.1038/nrm.2017.7

Banani, S. F., Rice, A. M., Peeples, W. B., Lin, Y., Jain, S., Parker, R., et al. (2016). Compositional control of phase-separated cellular bodies. *Cell* 166, 651–663. doi: 10.1016/j.cell.2016.06.010

Beermann, J., Piccoli, M. T., Viereck, J., and Thum, T. (2016). Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. *Physiol. Rev.* 96, 1297–1325. doi: 10.1152/physrev.00041.2015

Bettin, N., Oss Pegorar, C., and Cusanelli, E. (2019). The emerging roles of TERRA in telomere maintenance and genome stability. *Cells* 8:246. doi: 10.3390/cells8030246

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., et al. (2018). Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175, 1842–1855.e16.

Brangwynne, C., Tompa, P., and Pappu, R. (2015). Polymer physics of intracellular phase transitions. *Nat. Phys.* 11, 899–904. doi: 10.1038/nphys3532

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., et al. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324, 1729–1732. doi: 10.1126/science.1172046

Brosnan, C. A., and Voinnet, O. (2009). The long and the short of noncoding RNAs. *Curr. Opin. Cell Biol.* 21, 416–425.

Cai, D., Feliciano, D., Dong, P., Flores, E., Gruebele, M., Porat-Shliom, N., et al. (2019). Phase separation of YAP reorganizes genome topology for long-term YAP target gene expression. *Nat. Cell Biol.* 21, 1578–1589. doi: 10.1038/s41556-019-0433-z

Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., and Tartaglia, G. G. (2019). Phase separation drives X-chromosome inactivation: a hypothesis. *Nat. Struct. Mol. Biol.* 26, 331–334. doi: 10.1038/s41594-019-0223-0

Chen, L. L., and Carmichael, G. G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* 35, 467–478. doi: 10.1016/j.molcel.2009.06.027

Cho, W. K., Spille, J. H., Hecht, M., Lee, C., Li, C., Grube, V., et al. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415. doi: 10.1126/science.aar4199

Chong, P. A., and Forman-Kay, J. D. (2016). Liquid-liquid phase separation in cellular signaling systems. *Curr. Opin. Struct. Biol.* 41, 180–186. doi: 10.1016/j.sbi.2016.08.001

Chong, P. A., Vernon, R. M., and Forman-Kay, J. D. (2018). RGG/RG Motif regions in RNA binding and phase separation. *J. Mol. Biol.* 430, 4650–4665. doi: 10.1016/j.jmb.2018.06.014

Chujo, T., and Hirose, T. (2017). Nuclear bodies built on architectural long noncoding RNAs: unifying principles of their construction and function. *Mol. Cells* 40, 889–896.

Chujo, T., Yamazaki, T., and Hirose, T. (2016). Architectural RNAs (arcRNAs): a class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim. Biophys. Acta* 1859, 139–146. doi: 10.1016/j.bbagrm.2015.05.007

Conicella, A. E., Zerze, G. H., Mittal, J., and Fawzi, N. L. (2016). ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-Complexity C-Terminal domain. *Structure* 24, 1537–1549. doi: 10.1016/j.str.2016.07.007

Dao, T. P., Kolaitis, R. M., Kim, H. J., O'Donovan, K., Martyniak, B., Colicino, E., et al. (2018). Ubiquitin modulates liquid-liquid phase separation of UBQLN2 via disruption of multivalent interactions. *Mol. Cell* 69, 965–978.e6.

Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., and Lieberman, P. M. (2009). TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* 35, 403–413. doi: 10.1016/j.molcel.2009.06.025

Ding, D. Q., Okamasa, K., Katou, Y., Oya, E., Nakayama, J. I., Chikashige, Y., et al. (2019). Chromosome-associated RNA-protein complexes promote pairing

of homologous chromosomes during meiosis in Schizosaccharomyces pombe. *Nat. Commun.* 10:5598.

Dumbovic, G., Biayna, J., Banús, J., Samuelsson, J., Roth, A., Diederichs, S., et al. (2018). A novel long non-coding RNA from NBL2 pericentromeric macrosatellite forms a perinucleolar aggregate structure in colon cancer. *Nucleic Acids Res.* 46, 5504–5524. doi: 10.1093/nar/gky263

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C. C., Eckmann, C. R., Myong, S., et al. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U S A.* 112, 7189–7194. doi: 10.1073/pnas.1504822112

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., et al. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539, 452–455. doi: 10.1038/nature20149

Fay, M. M., and Anderson, P. J. (2018). The role of RNA in biological phase separations. *J. Mol. Biol.* 430, 4685–4701. doi: 10.1016/j.jmb.2018.05.003

Fernandes, J. C. R., Acuña, S. M., Aoki, J. I., Floeter-Winter, L. M., and Muxel, S. M. (2019). Long non-Coding RNAs in the regulation of gene expression: physiology and disease. *Noncoding RNA.* 5:17. doi: 10.3390/ncrna5010017

Fox, A. H. (2018). A mitochondria-paraspeckle crosstalk. *Nat. Cell Biol.* 20, 1108–1109. doi: 10.1038/s41556-018-0207-z

Fox, A. H., Lam, Y. W., Leung, A. K., Lyon, C. E., Andersen, J., Mann, M., et al. (2002). Paraspeckles: a novel nuclear domain. *Curr. Biol.* 12, 13–25.

Fox, A. H., Nakagawa, S., Hirose, T., and Bond, C. S. (2018). Paraspeckles: where long noncoding RNA meets phase separation. *Trends Biochem. Sci.* 43, 124–135. doi: 10.1016/j.tibs.2017.12.001

Franzmann, T. M., and Alberti, S. (2019). Prion-like low-complexity sequences: key regulators of protein solubility and phase behavior. *J. Biol. Chem.* 294, 7128–7136. doi: 10.1074/jbc.tm118.001190

Fujimoto, A., Furuta, M., Totoki, Y., Tsunoda, T., Kato, M., Shiraishi, Y., et al. (2016). Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* 48, 500–509.

Gasset-Rosa, F., Lu, S., Yu, H., Chen, C., Melamed, Z., Guo, L., et al. (2019). Cytoplasmic TDP-43 De-mixing independent of stress granules drives inhibition of nuclear import, loss of nuclear TDP-43, and cell death. *Neuron* 102, 339–357.e7.

Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., et al. (2019). Organization of chromatin by intrinsic and regulated phase separation. *Cell* 179, 470–484.e21.

Gui, X., Luo, F., Li, Y., Zhou, H., Qin, Z., Liu, Z., et al. (2019). Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* 10:2006.

Guil, S., and Esteller, M. (2012). Cis-acting noncoding RNAs: friends and foes. *Nat. Struct. Mol. Biol.* 19, 1068–1075. doi: 10.1038/nsmb.2428

Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnese, A., Hannett, N. M., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548. doi: 10.1038/s41586-019-1464-0

Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, 295–300. doi: 10.1038/nature10398

Heard, E., and Disteche, C. M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* 20, 1848–1867. doi: 10.1101/gad.1422906

Hennig, S., Kong, G., Mannen, T., Sadowska, A., Kobelke, S., Blythe, A., et al. (2015). Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* 210, 529–539. doi: 10.1083/jcb.201504117

Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., et al. (2014). NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell* 25, 169–183. doi: 10.1091/mbc.e13-09-0558

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., and Sharp, P. A. (2017). A phase separation model for transcriptional control. *Cell* 169, 13–23. doi: 10.1016/j.cell.2017.02.007

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

Huang, J. Z., Chen, M., Chen, D., Gao, X. C., Zhu, S., Huang, H., et al. (2017). A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol. Cell* 68, 171–184.e6.

Huo, X., Ji, L., Zhang, Y., Lv, P., Cao, X., Wang, Q., et al. (2020). The nuclear matrix protein SAFB cooperates with major satellite RNAs to stabilize heterochromatin architecture partially through phase separation. *Mol. Cell* 77, 368–383.e7.

Hupalowska, A., Jedrusik, A., Zhu, M., Bedford, M. T., Glover, D. M., and Zernicka-Goetz, M. (2018). CARM1 and paraspeckles regulate pre-implantation mouse embryo development. *Cell* 175, 1902–1916.e13.

Hyman, A. A., Weber, C. A., and Jülicher, F. (2014). Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* 30, 39–58.

Imamura, K., Imamachi, N., Akizuki, G., Kumakura, M., Kawaguchi, A., Nagata, K., et al. (2014). Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol. Cell* 53, 393–406. doi: 10.1016/j.molcel.2014.01.009

Jain, A., and Vale, R. D. (2017). RNA phase transitions in repeat expansion disorders. *Nature* 546, 243–247. doi: 10.1038/nature22386

Jiang, L., Shao, C., Wu, Q. J., Chen, G., Zhou, J., Yang, B., et al. (2017). NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. *Nat. Struct. Mol. Biol.* 24, 816–824. doi: 10.1038/nsmb.3455

Jin, J. (2019). Interplay between ubiquitylation and SUMOylation: empowered by phase separation. *J. Biol. Chem.* 294, 15235–15236. doi: 10.1074/jbc.h119.011037

Johnsson, P., Lipovich, L., Grandér, D., and Morris, K. V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta* 1840, 1063–1071.

Kim, H. J., Kim, N. C., Wang, Y. D., Scarborough, E. A., Moore, J., Diaz, Z., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 495, 467–473. doi: 10.1038/nature11922

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., et al. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* 547, 236–240. doi: 10.1038/nature22822

Lee, S., Kopp, F., Chang, T. C., Sataluri, A., Chen, B., Sivakumar, S., et al. (2016). Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell* 164, 69–80. doi: 10.1016/j.cell.2015.12.017

Li, C., Wang, S., Xing, Z., Lin, A., Liang, K., Song, J., et al. (2017). A ROR1-HER3-lncRNA signalling axis modulates the Hippo-YAP pathway to regulate bone metastasis. *Nat. Cell Biol.* 19, 106–119. doi: 10.1038/ncb3464

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516–520. doi: 10.1038/nature12210

Lin, A., Hu, Q., Li, C., Xing, Z., Ma, G., Wang, C., et al. (2017). The LINK-A lncRNA interacts with PtdIns (3,4,5)P3 to hyperactivate AKT and confer resistance to AKT inhibitors. *Nat. Cell Biol.* 19, 238–251. doi: 10.1038/ncb3473

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

Liu, H., Wan, J., and Chu, J. (2019). Long non-coding RNAs and endometrial cancer. *Biomed. Pharmacother.* 119:109396. doi: 10.1016/j.biopha.2019.109396

Liu, J., Liu, Z. X., Wu, Q. N., Lu, Y. X., Wong, C. W., Miao, L., et al. (2020). Long noncoding RNA AGPG regulates PFKFB3-mediated tumor glycolytic reprogramming. *Nat. Commun.* 11:1507.

Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M. J., et al. (2014). Enhancer activation requires trans-recruitment of a mega transcription factor complex. *Cell* 159, 358–373. doi: 10.1016/j.cell.2014.08.027

Lu, H., Yu, D., Hansen, A. S., Ganguly, S., Liu, R., Heckert, A., et al. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* 558, 318–323. doi: 10.1038/s41586-018-0174-3

Luo, S., Lu, J. Y., Liu, L., Yin, Y., Chen, C., Han, X., et al. (2016). Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell. Stem. Cell* 18, 637–652. doi: 10.1016/j.stem.2016.01.024

Maharana, S., Wang, J., Papadopoulos, D. K., Richter, D., Pozniakovsky, A., Poser, I., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* 360, 918–921. doi: 10.1126/science.aar7366

Mann, J. R., Gleixner, A. M., Mauna, J. C., Gomes, E., DeChellis-Marks, M. R., Needham, P. G., et al. (2019). RNA binding antagonizes neurotoxic phase transitions of TDP-43. *Neuron* 102, 321.e–338.e.

Mao, Y. S., Sunwoo, H., Zhang, B., and Spector, D. L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* 13, 95–101. doi: 10.1038/ncb2140

Min, J., Wright, W. E., and Shay, J. W. (2019). Clustered telomeres in phase-separated nuclear condensates engage mitotic DNA synthesis through BLM and RAD52. *Genes Dev.* 33, 814–827. doi: 10.1101/gad.324905.119

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., et al. (2015). A pooled shRNA screen identifies Rbm15, spen, and Wtap as factors required for xist RNA-Mediated silencing. *Cell Rep.* 12, 562–572. doi: 10.1016/j.celrep.2015.06.053

Montero, J. J., López-Silanes, I., Megías, D., Fraga, F. M., Castells-García, Á, and Blasco, M. A. (2018). TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin. *Nat. Commun.* 9:1548.

Munschauer, M., Nguyen, C. T., Sirokman, K., Hartigan, C. R., Hogstrom, L., Engreitz, J. M., et al. (2018). The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* 561, 132–136. doi: 10.1038/s41586-018-0453-z

Murakami, T., Qamar, S., Lin, J. Q., Schierle, G. S., Rees, E., Miyashita, A., et al. (2015). ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* 88, 678–690. doi: 10.1016/j.neuron.2015.10.030

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y. F., Goshima, N., and Hirose, T. (2012). Alternat'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* 31, 4020–4034. doi: 10.1038/emboj.2012.251

Nair, S. J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M. J., et al. (2019). Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat. Struct. Mol. Biol.* 26, 193–203. doi: 10.1038/s41594-019-0190-5

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

Neumann, P., Jaé, N., Knau, A., Glaser, S. F., Fouani, Y., Rossbach, O., et al. (2018). The lncRNA GATA6-AS epigenetically regulates endothelial gene expression via interaction with LOXL2. *Nat. Commun.* 9:237.

Ninomiya, K., and Hirose, T. (2020). Short tandem repeat-enriched architectural RNAs in nuclear bodies: functions and associated diseases. *Noncoding RNA.* 6:6. doi: 10.3390/ncrna6010006

Nishimoto, Y., Nakagawa, S., Hirose, T., Okano, H. J., Takao, M., Shibata, S., et al. (2013). The long non-coding RNA nuclear-enriched abundant transcr_2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis. *Mol. Brain.* 6:31. doi: 10.1186/1756-6606-6-31

Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., et al. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* 57, 936–947. doi: 10.1016/j.molcel.2015.01.013

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35, 190–194. doi: 10.1038/ng1244

Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., et al. (2015). A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* 162, 1066–1077. doi: 10.1016/j.cell.2015.07.047

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., et al. (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* 161, 774–789. doi: 10.1016/j.cell.2015.04.034

Pessina, F., Giavazzi, F., Yin, Y., Gioia, U., Vitelli, V., Galbiati, A., et al. (2019). Functional transcription promoters at DNA double-strand breaks mediate RNA-driven phase separation of damage-response factors. *Nat. Cell Biol.* 21, 1286–1299. doi: 10.1038/s41556-019-0392-4

Pitchiaya, S., Mourao, M. D. A., Jalihal, A. P., Xiao, L., Jiang, X., Chinnaiyan, A. M., et al. (2019). Dynamic recruitment of single RNAs to processing bodies depends on RNA functionality. *Mol. Cell* 74, 521–533.e6.

Pnueli, L., Rudnizky, S., Yosefzon, Y., and Melamed, P. (2015). RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α-subunit gene. *Proc. Natl. Acad. Sci. U S A.* 112, 4369–4374. doi: 10.1073/pnas.1414841112

Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., et al. (2005). Regulating gene expression through RNA nuclear retention. *Cell* 123, 249–263. doi: 10.1016/j.cell.2005.08.033

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

Qamar, S., Wang, G., Randle, S. J., Ruggeri, F. S., Varela, J. A., Lin, J. Q., et al. (2018). FUS phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions. *Cell* 173, 720–734.e15.

Qu, W., Wang, Z., and Zhang, H. (2020). Phase separation of the *C. elegans* polycomb protein SOP-2 is modulated by RNA and sumoylation. *Protein Cell* 11, 202–207. doi: 10.1007/s13238-019-00680-y

Reichheld, S. E., Muiznieks, L. D., Keeley, F. W., and Sharpe, S. (2017). Direct observation of structure and dynamics during phase separation of an elastomeric protein. *Proc. Natl. Acad. Sci. U S A.* 114, E4408–E4415.

Rheinbay, E., Parasuraman, P., Grimsby, J., Tiao, G., Engreitz, J. M., Kim, J., et al. (2017). Recurrent and functional regulatory mutations in breast cancer. *Nature* 547, 55–60.

Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B. F., et al. (2019). m6A enhances the phase separation potential of mRNA. *Nature* 571, 424–428. doi: 10.1038/s41586-019-1374-1

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323. doi: 10.1016/j.cell.2007.05.022

Roake, C. M., and Artandi, S. E. (2017). Approaching TERRA firma: genomic functions of telomeric noncoding RNA. *Cell* 170, 8–9. doi: 10.1016/j.cell.2017.06.020

Ryan, V. H., Dignon, G. L., Zerze, G. H., Chabata, C. V., Silva, R., Conicella, A. E., et al. (2018). Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol. Cell* 69, 465–479.e7.

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361:eaar3958. doi: 10.1126/science.aar3958

Salvador, F., and Gomis, R. R. (2018). Paraspeckle factor turns TGF-β1 pro-metastatic. *Nat. Cell Biol.* 20, 367–369. doi: 10.1038/s41556-018-0078-3

Sang, L. J., Ju, H. Q., Liu, G. P., Tian, T., Ma, G. L., Lu, Y. X., et al. (2018). LncRNA CamK-A regulates Ca2+-Signaling-Mediated tumor microenvironment remodeling. *Mol. Cell* 72, 71–83.e7

Sanulli, S., Trnka, M. J., Dharmarajan, V., Tibble, R. W., Pascal, B. D., Burlingame, A. L., et al. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature* 575, 390–394. doi: 10.1038/s41586-019-1669-2

Sarropoulos, I., Marin, R., Cardoso-Moreira, M., and Kaessmann, H. (2019). Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* 571, 510–514. doi: 10.1038/s41586-019-1341-x

Sasaki, Y. T., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. U S A.* 106, 2525–2530. doi: 10.1073/pnas.0807899106

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

Schmitz, S. U., Grote, P., and Herrmann, B. G. (2016). Mechanisms of long noncoding RNA function in development and disease. *Cell Mol. Life Sci.* 73, 2491–2509. doi: 10.1007/s00018-016-2174-5

Shevtsov, S. P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. *Nat. Cell Biol.* 13, 167–173. doi: 10.1038/ncb2157

Shichino, Y., Yamashita, A., and Yamamoto, M. (2014). Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. *Open Biol.* 4:140022. doi: 10.1098/rsob.140022

Shin, Y., Berry, J., Pannucci, N., Haataja, M. P., Toettcher, J. E., and Brangwynne, C. P. (2017). Spatiotemporal control of intracellular phase transitions using light-activated optodroplets. *Cell* 168, 159–171.e14.

Shin, Y., Chang, Y. C., Lee, D. S. W., Berry, J., Sanders, D. W., Ronceray, P., et al. (2018). Liquid nuclear condensates mechanically sense and restructure the genome. *Cell* 175, 1481–1491.e13.

Su, X., Ditlev, J. A., Hui, E., Xing, W., Banjade, S., Okrut, J., et al. (2016). Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science.* 352, 595–599. doi: 10.1126/science.aad9964

Sun, Q., Hao, Q., and Prasanth, K. V. (2018). Nuclear long noncoding RNAs: key regulators of gene expression. *Trends Genet.* 34, 142–157. doi: 10.1016/j.tig.2017.11.005

Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., et al. (2013). Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem. Biol.* 20, 341–350. doi: 10.1016/j.chembiol.2013.02.013

Tang, Y., Cheung, B. B., Atmadibrata, B., Marshall, G. M., Dinger, M. E., Liu, P. Y., et al. (2017). The regulatory role of long noncoding RNAs in cancer. *Cancer Lett.* 391, 12–19. doi: 10.1016/j.canlet.2017.01.010

Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., et al. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by pumilio proteins in human cells. *Nat. Commun.* 7:12209.

Torres, M., Becquet, D., Blanchard, M. P., Guillen, S., Boyer, B., Moreno, M., et al. (2016). Circadian RNA expression elicited by 3'-UTR IRAlu-paraspeckle associated elements. *eLife* 5:e14837.

Van Treeck, B., and Parker, R. (2018). Emerging roles for intermolecular RNA-RNA interactions in RNP assemblies. *Cell* 174, 791–802. doi: 10.1016/j.cell.2018.07.023

Wang, M., Tao, X., Jacob, M. D., Bennett, C. A., Ho, J. J. D., Gonzalgo, M. L., et al. (2018). Stress-Induced low complexity RNA activates physiological amyloidogenesis. *Cell Rep.* 24, 1713–1721.e4.

Wang, Y., He, L., Du, Y., Zhu, P., Huang, G., Luo, J., et al. (2015). The long noncoding RNA lncTCF7 promotes self-renewal of human liver cancer stem cells through activation of Wnt signaling. *Cell Stem Cell.* 16, 413–425. doi: 10.1016/j.stem.2015.03.003

Wang, Y., Hu, S. B., Wang, M. R., Yao, R. W., Wu, D., Yang, L., et al. (2018). Genome-wide screening of NEAT1 regulators reveals cross-regulation between paraspeckles and mitochondria. *Nat. Cell Biol.* 20, 1145–1158. doi: 10.1038/s41556-018-0204-2

Wang, Z., and Zhang, H. (2019). Phase separation, transition, and autophagic degradation of proteins in development and pathogenesis. *Trends Cell Biol.* 29, 417–427. doi: 10.1016/j.tcb.2019.01.008

Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K. M., Bennett, R. E., Dujardin, S., et al. (2018). Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J.* 37:e98049.

West, J. A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., et al. (2016). Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.* 214, 817–830. doi: 10.1083/jcb.201601071

Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R., and Parker, R. (2016). Distinct stages in stress granule assembly and disassembly. *eLife* 5:e18413.

Wu, H., Yin, Q. F., Luo, Z., Yao, R. W., Zheng, C. C., Zhang, J., et al. (2016). Unusual processing generates SPA LncRNAs that sequester multiple RNA binding proteins. *Mol. Cell.* 64, 534–548. doi: 10.1016/j.molcel.2016.10.007

Xing, Z., Lin, A., Li, C., Liang, K., Wang, S., Liu, Y., et al. (2014). lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* 159, 1110–1125. doi: 10.1016/j.cell.2014.10.013

Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y. S., Fox, A. H., et al. (2018). Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. *Mol. Cell* 70, 1038–1053.e7.

Yap, K., Mukhina, S., Zhang, G., Tan, J. S. C., Ong, H. S., and Makeyev, E. V. (2018). A short tandem repeat-enriched RNA assembles a nuclear compartment to control alternative splicing and promote cell survival. *Mol. Cell* 72, 525–540.e13.

Yoon, J. H., Abdelmohsen, K., and Gorospe, M. (2013). Posttranscriptional gene regulation by long noncoding RNA. *J. Mol. Biol.* 425, 3723–3730. doi: 10.1016/j.jmb.2012.11.024

Zhang, G., Wang, Z., Du, Z., and Zhang, H. (2018). mTOR regulates phase separation of PGL granules to modulate their autophagic degradation. *Cell* 174, 1492–1506.e22.

Zhang, K., Shi, Z. M., Chang, Y. N., Hu, Z. M., Qi, H. X., and Hong, W. (2014). The ways of action of long non-coding RNAs in cytoplasm and nucleus. *Gene* 547, 1–9. doi: 10.1016/j.gene.2014.06.043

Zhang, Z., Lee, J. H., Ruan, H., Ye, Y., Krakowiak, J., Hu, Q., et al. (2019). Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat. Commun.* 10:4562.

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

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

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 01 April 2021
doi: 10.3389/fgene.2021.642969

7SK Acts as an Anti-tumor Factor in Tongue Squamous Cell Carcinoma

*Bowen Zhang[1,2], Sainan Min[3], Qi Guo[1,2], Yan Huang[1,2], Yuzhu Guo[1], Xiaolin Liang[1,2], Li-ling Wu[4], Guang-yan Yu[3] and Xiangting Wang[1,2]**

[1] Department of Geriatrics, The First Affiliated Hospital of University of Science and Technology of China, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [2] Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, China, [3] Department of Oral and Maxillofacial Surgery, Peking University School and Hospital of Stomatology, Beijing, China, [4] Department of Physiology and Pathophysiology, Peking University School of Basic Medical Sciences, Key Laboratory of Molecular Cardiovascular Sciences, Ministry of Education, and Beijing Key Laboratory of Cardiovascular Receptors Research, Beijing, China

Increasing evidence has shown the mechanistic insights about non-coding RNA 7SK in controlling the transcription. However, the biological function and mechanism of 7SK in cancer are largely unclear. Here, we show that 7SK is down-regulated in human tongue squamous carcinoma (TSCC) and acts as a TSCC suppressor through multiple cell-based assays including a migration assay and a xenograft mouse model. The expression level of 7SK was negatively correlated with the size of tumors in the 73 in-house collected TSCC patients. Through combined analysis of 7SK knockdown of RNA-Seq and available published 7SK ChIRP-seq data, we identified 27 of 7SK-regulated genes that were involved in tumor regulation and whose upstream regulatory regions were bound by 7SK. Motif analysis showed that the regulatory sequences of these genes were enriched for transcription factors *FOXJ3* and *THRA*, suggesting a potential involvement of *FOXJ3* and *THRA* in 7SK-regulated genes. Interestingly, the augmented level of *FOXJ3* in TSCC patients and previous reports on *THRA* in other cancers have suggested that these two factors may promote TSCC progression. In support of this idea, we found that 21 out of 27 aforementioned 7SK-associated genes were regulated by *FOXJ3* and *THRA*, and 12 of them were oppositely regulated by 7SK and *FOXJ3/THRA*. We also found that *FOXJ3* and *THRA* dramatically promoted migration in SCC15 cells. Collectively, we identified 7SK as an antitumor factor and suggested a potential involvement of *FOXJ3* and *THRA* in 7SK-mediated TSCC progression.

Keywords: 7SK, tongue squamous cell carcinoma, tumor suppressor, *FOXJ3*, *THRA*

OPEN ACCESS

Edited by:
Liang Chen,
Wuhan University, China

Reviewed by:
Xiong Ji,
Peking University, China
Francesco Acquati,
University of Insubria, Italy

***Correspondence:**
Xiangting Wang
wangxt11@ustc.edu.cn

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *17 December 2020*
Accepted: *17 February 2021*
Published: *01 April 2021*

Citation:
Zhang B, Min S, Guo Q, Huang Y, Guo Y, Liang X, Wu L-l, Yu G-y and Wang X (2021) 7SK Acts as an Anti-tumor Factor in Tongue Squamous Cell Carcinoma. Front. Genet. 12:642969. doi: 10.3389/fgene.2021.642969

INTRODUCTION

7SK is highly abundant and evolutionarily conserved non-coding RNA (ncRNA) in vertebrates. It is transcribed by RNA polymerase III (Pol III) and predominantly localized in the nucleus (Wassarman and Steitz, 1991). 7SK is best known for acting as a scaffold with multiple proteins to form small nuclear ribonucleoprotein (snRNP) with *LARP7, HEXIM1*, and *MEPCE*. 7SK snRNP

has the ability to inhibit gene transcriptional elongation by sequestering the positive transcription elongation factor b (P-TEFb) (He et al., 2008; Abasi et al., 2016; Eichhorn et al., 2018). P-TEFb consists of CDK8 and cyclin T1 or T2 and regulates gene transcriptional elongation. Release of 7SK from snRNP complex leads to P-TEFb activation and transcriptional elongation.

Despite its well-investigated molecular mechanism in snRNP activity control, the biological function of 7SK is largely unclear. It has been shown that the expression level of 7SK was down-regulated in breast cancer, colon cancer, and Chronic Myelogenous Leukemia (CML) (Abasi et al., 2016). These reports suggested that 7SK may possess a tumor-suppressive role. Recently, an inhibitory role of LARP7 in tumor progression has been reported. In this report, knocking downLARP7 resulted in decreasing 7SK snRNP and released P-TEFb to positively regulate EMT associated genes, such as Slug, FOXC2, and ZEB2, which ultimately caused increased EMT, invasion, and metastasis in breast cancer (Ji et al., 2014). In view of the critical role of 7SK as a general regulator in gene transcription control and the suggestive dysregulation of 7SK in the previously reported cancer types, it would be of great interest to explore the biological function of 7SK in cancer and the underlying regulatory mechanism.

Tongue squamous cell carcinoma (TSCC) is a malignant tumor characterized by high incidence, mortality, and risk of metastasis to lymph node and other organs in the early stages (Annertz et al., 2012; Michikawa et al., 2012). Due to the limited understanding of the molecular mechanism, the targeted therapy is less involved for TSCC (Omura, 2014; Guo et al., 2018; Chen et al., 2020). It is urgent to explore the novel regulatory molecules to pursue potential targeted therapy development and drug discovery in TSCC. NcRNAs have been shown to participate in the regulation of TSCC progression. For example, long non-coding RNA (lncRNA) *KCNQ1OT1* is highly expressed in TSCC tissues, and knockdown of *KCNQ1OT1* repressed TSCC proliferation via miR-211-5p mediated Ezrin/Fak/Src signaling (Zhang et al., 2018). And miR-22 has been shown to be able to promote TSCC apoptosis upon *cis*-platin treatment through PI3K/Akt/NF-κB signaling (Gu et al., 2018).

Here, we showed that 7SK was down-regulated in 63% of the TSCC patients, and the expression level of 7SK was negatively correlated with the size of the tumor. We also found that knockdown of 7SK promoted cell growth and migration in cultured SCC15 cells and xenograft tumor formation in nude mice. Through RNA-seq and data mining of the reported ChIRP-seq (Flynn et al., 2016), we discovered 27 7SK-regulated genes that showed occupancy by 7SK RNA. Furthermore, we built a potential correlation between 7SK and FOXJ3/THRA through identifying FOXJ3 and THRA binding elements on the 5′ regulatory regions of these genes. Among these 27 7SK-regulated genes, 21 of them were also regulated by FOXJ3 and THRA, including 12 genes that are showed to be oppositely controlled by 7SK and FOXJ3/THRA. Interestingly, the expression level of FOXJ3 was augmented in TSCC patients and previous reports of other cancer have suggested an oncogenic role of THRA. In support of this, we also found that FOXJ3 and THRA dramatically promoted migration in SCC15 cells. All in all, we have reported a tumor-suppressive role of 7SK in TSCC and suggested a putative functional involvement of FOXJ3 and THRA in 7SK-mediated TSCC progression.

MATERIALS AND METHODS

Cell Culture

Human tongue squamous cell carcinoma cell line SCC15 was purchased from the American Type Culture Collection (ATCC). SCC15 cells were cultured in a DMEM/F-12 medium (Sigma, United States) supplemented with 10% fetal bovine serum (FBS, Gibco, United States), 1% penicillin, and streptomycin (WISENT Inc., CA) and incubated under humidified atmosphere of 5% CO_2 at 37°C and maintained at the confluency of approximately 80%.

Plasmid Construction

The small hairpin RNAs (shRNAs) targeting human 7SK were subcloned into the expression vector pLKO.1 using *Age*I and *Eco*RI restriction enzymes (Thermo Fisher Scientific, United States) under the control of the U6 promoter. The resultant constructs were designated as sh7SK-1 and sh7SK-2. Sequences used for sh7SK are listed in **Supplementary Table 5**.

Transfection

Cells were pre-seeded on 60 mm dishes at a density of 6×10^5/dish the day before transfection. ShRNAs target 7SK or the negative control in the PLKO.1 vector and siRNAs target FOXJ3 and THRA were transfected into cells using Lipofectamine 3,000 (Invitrogen, United States) following the manufacturer's instructions. Transfectants were collected 48 h after transfection. SiRNAs were purchased from RiboBio, and sequences used for siFOXJ3 and siTHRA are listed in **Supplementary Table 5**.

RNA Extraction, Reverse Transcription (RT), and Real-Time Quantitative PCR (RT-qPCR)

Cultured cells were washed twice in ice-cold phosphate-buffered saline (PBS) and lysed with TRIzol reagent (Ambion, United States) to isolate total RNA, according to the manufacturer's instructions. After treatment with DNase I (Thermo Fisher Scientific, United States), the first-strand cDNA was synthesized using HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, CHINA). The cDNA was analyzed with real-time PCR using a SYBR Green qPCR kit (Vazyme, CHINA). All primers for each gene are listed in **Supplementary Table 6**. Quantitative PCR was conducted at 95°C for 5 min followed by 40 cycles of 95°C for 10 s and 60°C for 30 s in LightCycler® 96 of Roche. The $2^{-\Delta\Delta Ct}$ method was used to quantify the relative expression using GAPDH as the endogenous control. Primers used for RT-qPCR are listed in **Supplementary Table 6**.

Apoptosis Analysis

SCC15 cells were transfected with sh7SK or shRNA control, respectively. SCC15 cells were harvested by trypsinization and washed with PBS. For apoptosis analysis, 1×10^5 cells were

harvested and an Annexin V Staining Kit (BD Biosciences, United States) was used according to the manufacturer's instructions 48 h after transfection. The stained cells were analyzed with Cellometer Vision Image Cytometer (Nexcelom Bioscience, United States). Data were analyzed in FCS4 Express Cytometry (*De Novo* Software, United States).

RNA-Seq

The control and two 7SK knockdown SCC15 cell lines were collected for RNA-Seq. RNA library construction and sequencing were performed with BGISEQ-500. About ∼6 GB raw RNA-seq data were obtained for each sample. For each experiment, three biological replicates were sequenced. The differential expressed genes (DEGs) were determined by using Fold change ≥ 2 and diverge probability ≥ 0.8 as criteria. The raw data is uploaded to NCBI SRY, the accession No. is PRJNA686697.

Patient Sample Collection

Tumor and matched non-malignant tissues of 73 patients were recruited from Peking University School and Hospital of Stomatology. All of the fresh tissue specimens were delivered under liquid nitrogen after being dissected carefully at least 1.5 cm from the tumor margins and stored at −80°C before use. The surgical samples contained at least 70% tumor or normal cells, which was confirmed by a pathologist using Hematoxylin-Eosin staining. The clinicopathological factors of patients were calculated in **Supplementary Table 1**. Tumor sizes and clinical stages were informed from clinical examination and radiogram and classified patients into clinical stage I–IV in terms of TNM staging system. With respect to lymph node metastasis, the status was judged according to histopathologic examination of the regional lymph nodes. The research was approved by the Ethics Committees for Human Experiments of Peking University School and Hospital of Stomatology (Approval number PKUSSIRB-2013009). All patients signed an informed consent document before tissue collection.

Cell Migration Assay

To begin, 8 μm-pore polycarbonate membrane divided transwell chambers into upper and lower chambers. And 3.0×10^4 cells resuspended with serum-free DMEM/F12 medium were added to the upper chamber and medium containing 20% FBS, which was added to the lower chamber before the chamber was incubated at 37°C in 5% CO_2 for 24 h. After the removal of non-migrated cells on the upper side of the membrane by cotton swab, the membrane was fixed with 95% ethanol and stained with hematoxylin-eosin (ZSGB, CHINA). The migrated cells were counted using light microscopy. The experiment was performed three times independently.

Scratch Wound Healing Assay

Cells were plated into 6-well plates and incubated at 37°C in 5% CO_2 until reached to 100% confluence. Straight scratches were made after treating cells with 1 μg/ml mitomycin C (Roche, United States) for 2 h. After we washed it with 1 × PBS, serum-free medium was added and the cells were further cultured for 20 h. The cells from six views in each well were photographed at 0 and 20 h. The migration distance was assessed using Image-pro plus 6.0. The average distance of the six counted views was calculated and used for evaluating the wound healing effect. The wound healing experiment was performed three times independently.

MTS Cell Proliferation Assay

Cell proliferation was assessed by CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega, Madison, WI, United States), a 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) assay. Squamous cell carcinoma cells were seeded in 96-well plates (2,000 cells per well). Then 20 μl MTS reagent was mixed in each well after 24 or 48 h and incubated in 37°C for 4 h before detection. The plate was measured absorbance at 490 nm.

Nude Mice Xenograft Model

All experimental procedures were approved by the Ethics Committee of Animal Research, Peking University Health Science Center. Four female BALB/c nude mice were obtained from Peking University Health Science Center. And 6×10^6 of 7SK knockdown and control cells were subcutaneously injected at the armpits. Nude mice xenograft tumor models were assessed after 3 weeks and hematoxylin-eosin (HE) staining was performed to analyze tumor pathological feature under the microscope.

ChIRP-Seq Analysis

The 7SK ChIRP-Seq data included hg19_7SK_ChIRP-seq_HeLa.bw and hg19_7SK_ChIRP-seq_HeLa_Input.bw were obtained from Gene Expression Omnibus database under accession code GSE69141 (Flynn et al., 2016). The obtained data were uploaded to UCSC Custom Tracks and analyzed for the occupancy of 7SK on the 59 DEGs.

Statistical Analysis

Data were presented as the mean ± SEM. Statistical analysis of the results were performed using unpaired Student's *t*-tests with GraphPad Prism, GraphPad Software, La Jolla California, United States[1]. $P < 0.05$ was considered significant.

RESULTS

Downregulation of 7SK Correlates With Tumor Size in TSCC Patients

To investigate whether 7SK is involved in human TSCC progression, we detected the expression level of 7SK in clinical specimens. Our results showed that the expression level of 7SK in 46 out of 73 TSCC samples (63%) were significantly decreased when compared to the matched non-malignant counterpart (Tumor/Non-malignant, T/N) (**Figures 1A,B**). Statistical analysis showed that the downregulation of 7SK in TSCC was

[1]www.graphpad.com

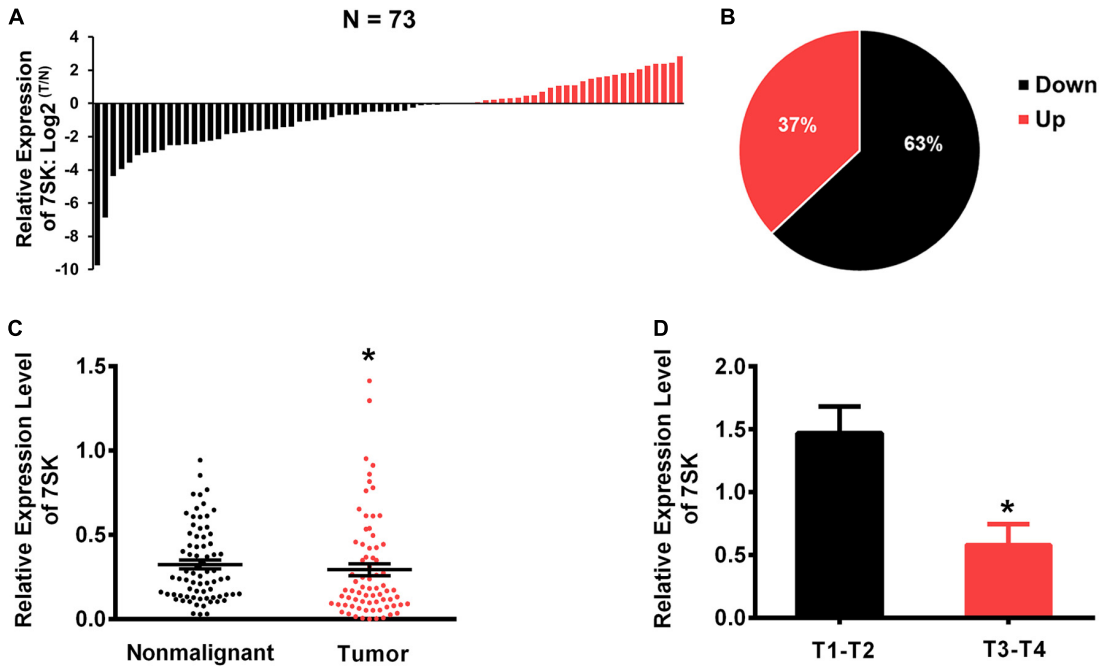


FIGURE 1 | The expression levels of 7SK in TSCC patients. **(A)** The Log_2 fold changes of 7SK expression (in tumors and non-malignant samples) in each of 73 TSCC patients. **(B)** Pie chart showing ratios of different 7SK expression in tumors and non-malignant areas of 73 patients. **(C)** Scatterplot showing the expression levels of 7SK in 73 TSCC tumors and the corresponding non-malignant samples. **(D)** The expression levels of 7SK in different T stages of the 73 patients. *$P < 0.05$.

significant (**Figure 1C**). To further establish the correlation between 7SK expression and TSCC risk, corresponding 7SK expression levels of clinicopathological features were calculated and summarized in **Supplementary Table 1**. Our results showed that patients with low 7SK expression had poorer T stage than those patients with higher 7SK expression (**Figure 1D**), indicating a negative correlation between 7SK expression and the tumor size of TSCC. Note that there is no significant difference between 7SK expression and clinicopathological factors such as age, gender, or lymph nodes metastatic (**Supplementary Table 1**) or prognosis of TSCC patients (**Supplementary Figure 1**). The association between tumors and reduced 7SK expression was further examined by using cultured tumor cell lines. Decreased expression level of 7SK was also observed in human tongue squamous carcinoma (TSCC) SCC15 cells, compared with the normal oral keratinocyte HOK cells (**Supplementary Figure 2**). The down-regulated 7SK in TSCC patients and SCC15 cells indicates a tumor suppressor role of 7SK. SCC15 cells were chosen as a cell model for further investigations.

Knockdown of 7SK Leads to Increased Cell Proliferation and Migration, but Decreased Apoptosis in SCC15 Cell–

To pursue the biological role of 7SK in SCC15 cell, we generated two 7SK shRNAs. Both 7SK shRNAs showed more than 50% knockdown efficiency (**Figure 2A**). To test the effects of 7SK on SCC15 proliferation, cell death or migration, we performed MTS assay, apoptotic assay detected by Annexin V/PI staining, wound healing assay, and transwell assay (**Figures 2B–E** and **Supplementary Figure 3**). Apoptotic, wound healing, and transwell assays were performed in the presence of transiently transfected 7SK shRNAs. For MTS assay, we used the most efficient single stable clones among the tested sh7SK-1 and sh7SK-2 transfectants (**Supplementary Figure 4**). Compared with the control shRNA transfected SCC15 cells, 7SK shRNAs significantly enhanced SCC15 cell proliferation detected by MTS assay (**Figure 2B**) but markedly reduced apoptotic rate (**Figure 2C**). In addition, 7SK shRNAs markedly fastened wound closure rate in wound healing assay (**Figure 2D**), and possessed higher migration in transwell assay (**Figure 2E**). Consistently with the enhanced effect of 7SK knockdown in transwell assay, our data showed that overexpression of 7SK significantly decreased the cell migration rate in SCC15 cells (**Supplementary Figures 3A,B**).

7SK Deficiency of SCC15 Cells Strongly Induces Tumor Formation in Nude Mice

To determine the *in vivo* role of 7SK, we conducted nude mice xenograft assay by using the same sh7SK-1 stable clone for the aforementioned MTS assay. The control shRNA or sh7SK-1 cells were injected subcutaneously into the armpits of four nude mice and were examined 3 weeks later. Consistently with our prediction, in the right armpits transplanted with sh7SK cells we observed obvious lumps formation while the left armpits transplanted with control shRNA did not form any lump during the observed time window (**Figure 2F**). All transplanted nude mice developed progressive cachexia, and the

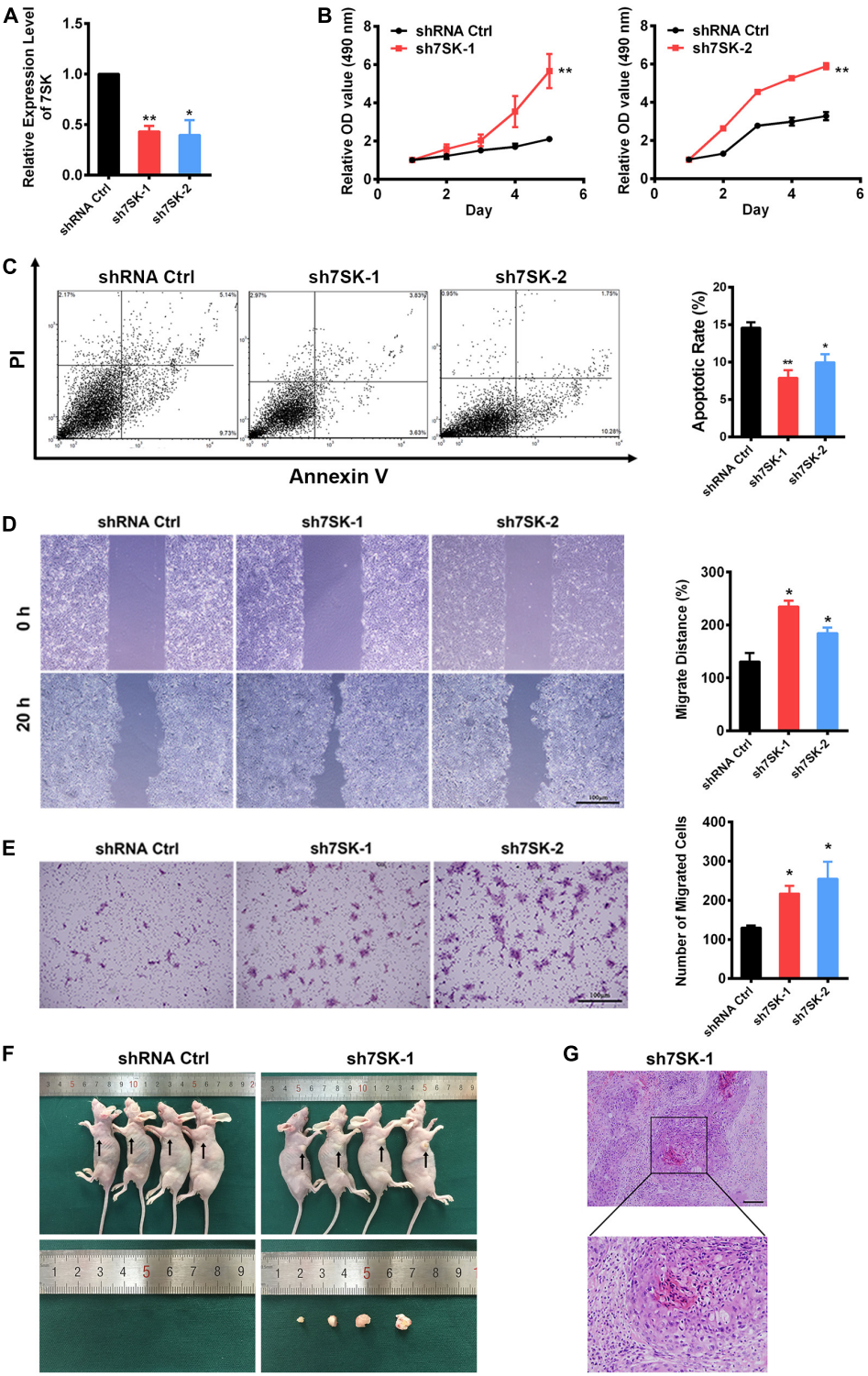


FIGURE 2 | 7SK inhibits tumor progression *in vitro* and *in vivo*. **(A)** Efficiency of 7SK knockdown in SCC15 cells, $n = 3$. **(B–E)** Apoptosis assay detected by Annexin V/PI staining **(B)**, MTS assay **(C)**, wound healing assay **(D)**, and transwell assay **(E)** in 7SK knockdown SCC15 cells, $n = 3$. For **(C–E)**, the representative images and corresponding statistical analysis are shown. The scale bar is 100 μm. **(F)** Images of nude mice injected with control or sh7SK-1 SCC15 cells, as well as the lumps from the injection of sh7SK-1 SCC15 cells. **(G)** HE staining of lumps from sh7SK-1 SCC15 cells injection, $n = 4$. The scale bar is 50 μm. Data represent the mean ± SEM. $^*P < 0.05$. $^{**}P < 0.01$.

skin area of the sh7SK lumps festered. HE staining of the sh7SK lump tissues showed abnormal mitoses, high atypia, and nests composed of polygonal cells, which are typical characteristics of squamous cell carcinoma (**Figure 2G**). All in all, our results suggest that knockdown of 7SK promoted cell growth and migration in cultured SCC15 cells, and xenograft tumor formation in nude mice.

RNA-Seq Reveals 7SK Regulated Genes That Are Enriched in Cell Proliferation, Migration, and Cell Death in SCC15 Cells

To systematically identify downstream genes regulated by 7SK, we performed RNA sequencing (RNA-Seq) for 7SK knockdown SCC15 cells with two shRNAs. Analyses on the size of clean reads, mapping rates, and quality control have confirmed the suitable quality for the obtained RNA-Seq data (**Supplementary Tables 2–4**). Correlation analysis shown in **Supplementary Figure 5** indicated that all triplicates in each treatment are comparable and reproducible. A total of 127 up-regulated genes and 138 down-regulated genes were identified (**Figure 3A** and **Supplementary Table 7**). In concert with the enhanced tumor progression in 7SK knockdown SCC15 cells, 40 of 127 up-regulated genes were found to be enriched in gene ontology (GO) terms that were involved in proliferation, migration, and angiogenesis; while 19 of the 138 down-regulated genes were enriched for cell death and cornification by GO analysis (**Figure 3B**). A heatmap of these 59 genes (40 up-regulated and 19 down-regulated) is shown in **Figure 3C**. Among these, 27 genes (star-marked in **Figure 3C**) have been further identified to have occupancy with 7SK RNA (detailed explanation is provided under the headline "ChIRP-Seq Data Mining Reveals PAX5, THRA, FOXJ3, and Two Novel Motifs on 7SK-Associated Genes"), and validated by RT-qPCR (**Supplementary Figure 6**).

ChIRP-Seq Data Mining Reveals PAX5, THRA, FOXJ3, and Two Novel Motifs on 7SK-Associated Genes

To investigate if 7SK has physical interactions with the 5′ regulatory regions of the above identified 59 7SK-regulated genes to regulate the expression, we took advantage of the 7SK ChIRP-Seq data published by Chang lab (Flynn et al., 2016). Our data mining results indicated that 27 of the 59 genes showed occupancy of 7SK (**Supplementary Figure 7**, and star-marked in **Figure 3C**). To further investigate the potential regulatory mechanism of these 27 7SK-associated genes, we conducted a bioinformatic analysis to explore the common motifs within the 5′ regulatory regions of these 27 genes up to 3 Kb upstream from their transcription start sites and obtained 5 motifs. Then, we performed motif analysis on the five motifs, and we found that motifs 1 and 2 do not match any known binding elements

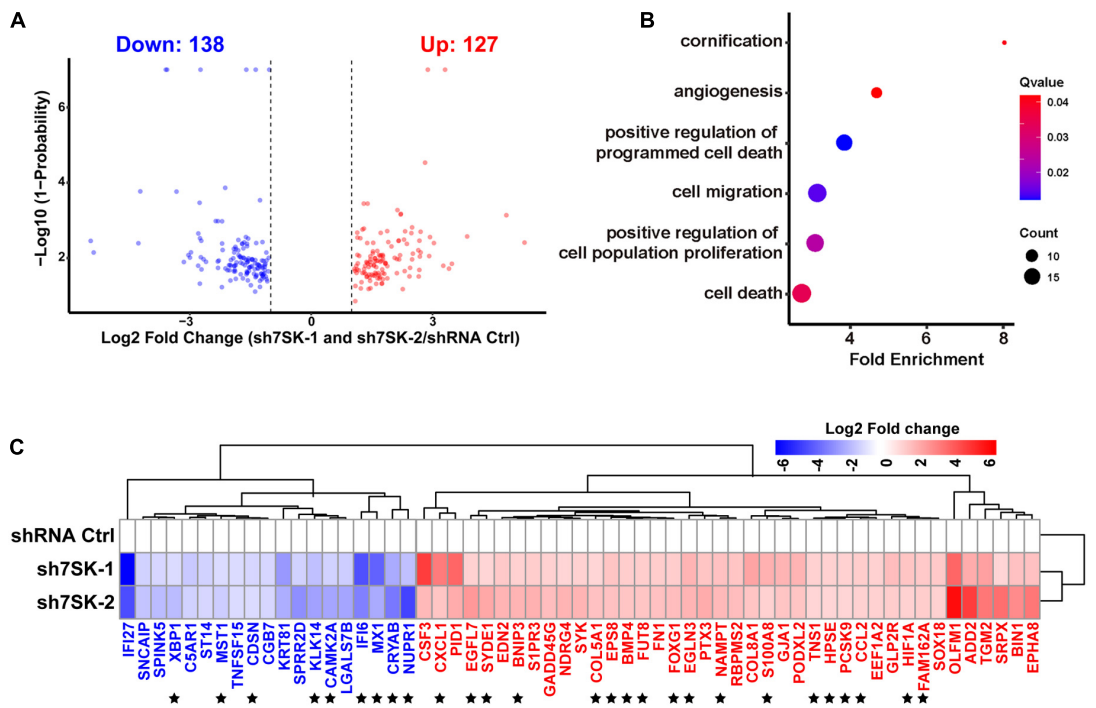


FIGURE 3 | 7SK regulates genes that are involved in tumor progression. **(A)** Volcano plot of commonly regulated genes identified from sh7SK-1 and sh7SK-2 RNA-seq compared with shRNA Ctrl. **(B)** Biological processes related to tumor progression and significantly enriched by Gene Ontology (GO) analysis using 7SK regulated genes from **(A)**. **(C)** The Log_2 fold changes of the genes that are identified from **(B)** are illustrated on a heat map. ★, 27 genes that are further identified as 7SK directly associated genes shown in **Figure 4**. Blue: down-regulated genes in the presence of 7SK shRNAs; red: up-regulated genes in the presence of 7SK shRNAs, $n = 3$.

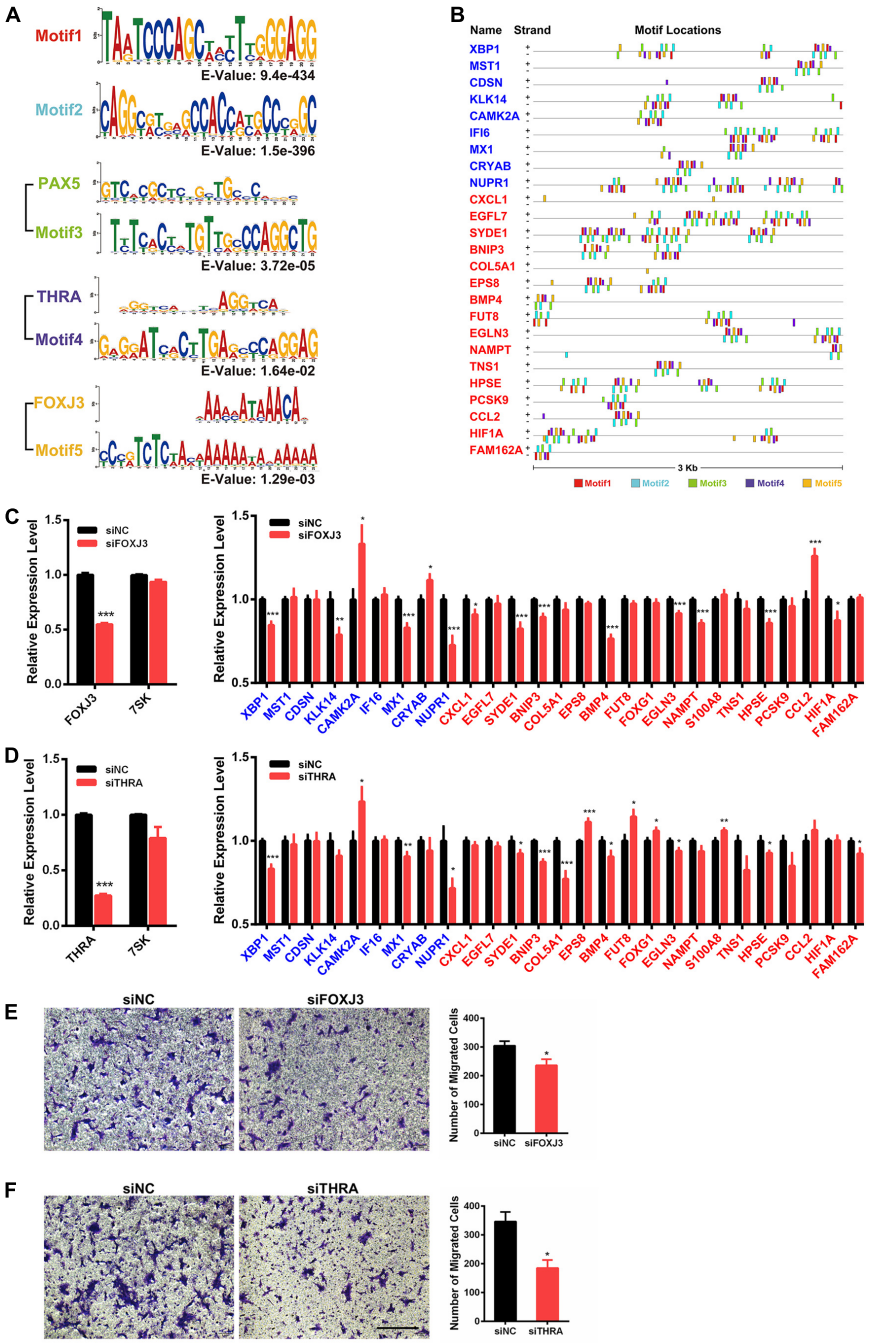


FIGURE 4 | Identification of *FOXJ3* and *THRA* as two transcription factors that oppositely regulate 7SK associated genes and SCC15 migration. **(A)** The top five motifs (including *FOXJ3* and *THRA*) identified on the upstream regulatory regions of the 27 7SK associated genes. **(B)** Motif distributions on the upstream regulatory regions of the 27 7SK-associated genes. **(C,D)** RT-qPCR of *FOXJ3*, *THRA*, 7SK, and the 27 7SK-associated genes in the presence of siFOXJ3 **(C)** and siTHRA **(D)** treated SCC15 cells. **(E,F)** The representative images (left) and the according statistical analysis (right) of transwell assay after *FOXJ3* **(E)** and *THRA* **(F)** knockdown. The scale bar is 100 μm, $n = 3$. $^*P < 0.05$.

to the reported transcription factors, while motifs 3, 4, and 5 possess similar DNA binding motifs to *PAX5*, *THRA*, and *FOXJ3*, respectively (**Figure 4A**). The distribution maps of the 5 motifs on the upstream regulatory regions of the 27 7SK-associated genes are shown in **Figure 4B**.

7SK-Associated Genes Are Regulated by FOXJ3 and THRA

Next, we attempted to identify the factors that may participate in the regulation of the 7SK-associated genes in TSCC. Our RT-qPCR results showed that PAX5 is barely detectable

in SCC15 cells; in contrast, both FOXJ3 and THRA are relatively abundantly expressed in SCC15 cells (**Supplementary Figure 8A**). Therefore, we focused on FOXJ3 and THRA for further investigation. We found that FOXJ3 was significantly upregulated in TSCC patients, and THRA was constantly expressed in normal and TSCC tissues from TCGA database (**Supplementary Figure 8B**). Our RT-qPCR results showed that 21 out of these 27 genes exhibited significant changes in the presence of either knockdown of *FOXJ3 (XBP1, KLK14, CAMK2A, MX1, CRYAB, NUPR1, CXCL1, SYDE1, BNIP3, BMP4, EGLN3, NAMPT, HPSE, CCL2,* and *HIF1A)* or *THRA (XBP1, CAMK2A, MX1, NUPR1, SYDE1, BNIP3, COL5A1, EPS8, BMP4, FUT8, FOXG1, EGLN3, S100A8, HPSE,* and *FAM162A*). Among these 21 *FOXJ3*- and *THRA*-regulated genes, 9 of them (*XBP1, CAMK2A, MX1, NUPR1, SYDE1, BNIP3, BMP4, EGLN3,* and *HPSE*) are common target genes shared by FOXJ3 and THRA (**Figures 4C,D**). Interestingly, among these 21 genes, 12 genes (57%) are oppositely regulated by 7SK and *FOXJ3*/*THRA*. Particularly, four genes that positively regulate tumor migration are identified to be negatively regulated by 7SK but positively regulated by *FOXJ3* and *THRA*. These four genes are *CXCL1*, *SYDE1*, *COL5A1*, and *HIF1A* (Lo et al., 2017; Byun et al., 2019; Feng et al., 2019; Yang et al., 2019). Consistently with these findings, knockdown of either *FOXJ3* or *THRA* significantly decreased the cell migration rate in SCC15 cells (**Figures 4E,F**), which is opposite to 7SK.

Altogether, our results suggest a putative functional involvement of the transcriptional factors *FOXJ3* and *THRA* in the 7SK-associated gene expression control and tumor migration in TSCC.

DISCUSSION

Currently, 7SK is considered a paradigm in RNA-regulated transcription (Diribarne and Bensaude, 2009). However, the biological function of 7SK is largely unclear. Here, we identified 7SK acting as a novel tumor suppressor in TSCC, and suggested a putative functional involvement of FOXJ3 and THRA in 7SK-mediated TSCC progression.

In TSCC, we found that 7SK was down-regulated in 63% of the 73 TSCC patients, and the down-regulated 7SK is correlated with the size of tumor. In functional assays, knockdown of 7SK in cell and xenograft experiments showed a tumor suppressor role of 7SK in TSCC. Numerous reports have linked non-coding RNAs with cancer (Wapinski and Chang, 2011; Wang et al., 2019; Huang et al., 2020). For example, *HOTAIR* is found to be associated with tumor metastasis in breast cancer through binding with polycomb repressive complex 2 (PRC2) and the lysine-specific demethylase1 (LSD1) to regulate gene expression (Burd et al., 2010; Tsai et al., 2010). Our previous work showed that ncRNA$_{CCND1}$s bind to FUS/TLS, thus causing an allosteric effect and inhibiting the histone acetyltransferase activity of CBP/P300, leading to gene transcription suppression of *CCND1* in Hela cells (Wang et al., 2008). Interestingly, ncRNA$_{CCND1}$s can crosstalk with another lncRNA, namely, LINC00473, to fine-tune the expression level of *CCND1* in breast cancer cells (Shi and Wang, 2019).

FOXJ3 (Forkhead box protein J3), *THRA* (thyroid hormone receptor alpha), and *PAX5* (paired box protein Pax-5) were identified through motif comparison analysis with 7SK binding genes. Due to the barely detectable expression level of *PAX5* in SCC15 cells, we chose not to focus on *PAX5* in the current study. However, we do not exclude the potential regulatory role of *PAX5* in 7SK-associated genes in other types of cancer. *FOXJ3* is a transcription factor that is down-regulated in colorectal cancer and lung cancer (Ma et al., 2016; Chen et al., 2018). In TSCC, the expression level of *FOXJ3* is up-regulated. *THRA* is a nuclear receptor that binds to T3 (Thyroxine 3) (Markossian et al., 2018). Interestingly, although the expression level of *THRA* has not been found to be changed in TSCC patients, our analysis, using TCGA database derived head and neck squamous cancer (HNSC) patients, found certain mutations located in the ligand-binding domain and zinc finger domain of *THRA*. These findings suggest that possible dysregulations of *THRA* in cancer may involve the changes of its DNA binding or ligand-receptor interaction abilities. *THRA* has been shown to be an adverse prognostic signature for breast cancer (Wu et al., 2020). The up-regulation of *FOXJ3* in multiple types of cancers and the link between *THRA* and the breast cancer prognostic prediction suggest oncogenic functions of these two transcription factors, which is opposite to that of 7SK. Supporting this, our results from transwell assay have demonstrated opposite roles of *FOXJ3* and *THRA* in TSCC to 7SK. In addition, we have also identified a set of genes that are regulated in the opposite way by 7SK and *FOXJ3*/*THRA*, including four positive regulators in tumor migration (*CXCL1*, *SYDE1*, *COL5A1*, and *HIF1A*). It is likely that 7SK and *FOXJ3*/*THRA* possess competitive bindings on their commonly regulated genes to control the targeted genes' activation and repression at the transcription level. It would be of great interest to demonstrate whether a directly functional link between *FOXJ3*/*THRA* and 7SK is involved in 7SK-mediated genes and 7SK-mediated tumor progression events in the near future.

In summary, the identification of 7SK in controlling TSCC progression expands our understanding of ncRNAs in cancer biology. Our work provides a potential area of developing novel diagnosis and therapeutic markers for TSCC.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI Bioproject: https://www.ncbi.nlm.nih.gov/bioproject/686697, accession PRJNA686697.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committees for Human Experiments of

Peking University School and Hospital of Stomatology (Approval number PKUSSIRB-2013009). The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Ethics Committee of Animal Research, Peking University Health Science Center.

AUTHOR CONTRIBUTIONS

XW initiated the project and supervised the whole project. SM, LW, and GY contributed to the xenograft experiments and the experiments using TSCC patient samples. BZ, YG, and XL performed the cell-based assays. QG, YH, and BZ performed the RNA-seq and data analysis. BZ conducted all the other experiments and data analysis. BZ and XW wrote the manuscript. All authors read and accepted the final version.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (31970598 and 31471226 to XW) and the Fundamental Research Funds for the Central Universities (YD2070002010 to XW).

ACKNOWLEDGMENTS

We thank TS and YH for supporting this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.642969/full#supplementary-material

REFERENCES

Abasi, M., Bazi, Z., Mohammadi-Yeganeh, S., Soleimani, M., Haghpanah, V., Zargami, N., et al. (2016). 7SK small nuclear RNA transcription level down-regulates in human tumors and stem cells. *Med. Oncol.* 33:128. doi: 10.1007/s12032-016-0841-x

Annertz, K., Anderson, H., Palmer, K., and Wennerberg, J. (2012). The increase in incidence of cancer of the tongue in the Nordic countries continues into the twenty-first century. *Acta Otolaryngol.* 132, 552–557. doi: 10.3109/00016489.2011.649146

Burd, C. E., Jeck, W. R., Liu, Y., Sanoff, H. K., Wang, Z., and Sharpless, N. E. (2010). Expression of linear and novel circular forms of an INK4/ARF-associated non-coding rna correlates with atherosclerosis risk. *PLoS Genet.* 6:e1001233. doi: 10.1371/journal.pgen.1001233

Byun, Y., Choi, Y. C., Jeong, Y., Lee, G., Yoon, S., Jeong, Y., et al. (2019). MiR-200c downregulates HIF-1 and inhibits migration of lung cancer cells. *Cell. Mol. Biol. Lett.* 24:28. doi: 10.1186/s11658-019-0152-2

Chen, J. L., Liu, L., Cai, X. J., Yao, Z. G., and Huang, J. H. (2020). Progress in the study of long noncoding RNA in tongue squamous cell carcinoma. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 129, 51–58. doi: 10.1016/j.oooo.2019.08.011

Chen, L. Z., Ding, Z., Zhang, Y., He, S. T., and Wang, X. H. (2018). MiR-203 over-expression promotes prostate cancer cell apoptosis and reduces ADM resistance. *Eur. Rev. Med. Pharmacol. Sci.* 22, 3734–3741.

Diribarne, G., and Bensaude, O. (2009). 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. *RNA Biol.* 6, 122–128. doi: 10.4161/rna.6.2.8115

Eichhorn, C. D., Yang, Y., Repeta, L., and Feigon, J. (2018). Structural basis for recognition of human 7SK long noncoding RNA by the La- related protein Larp7. *Proc. Natl. Acad. Sci. U.S.A.* 115, E6457–E6466. doi: 10.1073/pnas.1806276115

Feng, G., Ma, H. M., Huang, H. B., Li, Y. W., Zhang, P., Huang, J. J., et al. (2019). Overexpression of COL5A1 promotes tumor progression and metastasis and correlates with poor survival of patients with clear cell renal cell carcinoma. *Cancer Manag. Res.* 11, 1263–1274. doi: 10.2147/Cmar.S188216

Flynn, R. A., Do, B. T., Rubin, A. J., Calo, E., Lee, B., Kuchelmeister, H., et al. (2016). 7SK-BAF axis controls pervasive transcription at enhancers. *Nat. Struct. Mol. Biol.* 23, 231–238. doi: 10.1038/nsmb.3176

Gu, Y. X., Liu, H., Kong, F. R., Ye, J. H., Jia, X. T., Zhang, Z. J., et al. (2018). miR-22/KAT6B axis is a chemotherapeutic determiner via regulation of PI3k-Akt-NF-kB pathway in tongue squamous cell carcinoma. *J. Exp. Clin. Cancer Res.* 37:164. doi: 10.1186/s13046-018-0834-z

Guo, Y. Z., Sun, H. H., Wang, X. T., and Wang, M. T. (2018). Transcriptomic analysis reveals key lncRNAs associated with ribosomal biogenesis and epidermis differentiation in head and neck squamous cell carcinoma. *J. Zhejiang Univ. Sci. B* 19, 674–688. doi: 10.1631/jzus.B1700319

He, N., Jahchan, N. S., Hong, E., Li, Q., Bayfield, M. A., Maraia, R. J., et al. (2008). A la-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. *Mol. Cell* 29, 588–599. doi: 10.1016/j.molcel.2008.01.003

Huang, Y., Guo, Q., Ding, X. P., and Wang, X. T. (2020). Mechanism of long noncoding RNAs as transcriptional regulators in cancer. *RNA Biology* 17, 1680–1692. doi: 10.1080/15476286.2019.1710405

Ji, X. D., Lu, H. S., Zhou, Q., and Luo, K. X. (2014). LARP7 suppresses P-TEFb activity to inhibit breast cancer progression and metastasis. *Elife* 3:e02907. doi: 10.7554/eLife.02907

Lo, H. F., Tsai, C. Y., Chen, C. P., Wang, L. J., Lee, Y. S., Chen, C. Y., et al. (2017). Association of dysfunctional synapse defective 1 (SYDE1) with restricted fetal growth-SYDE1 regulates placental cell migration and invasion. *J. Pathol.* 241, 324–336. doi: 10.1002/path.4835

Ma, W. Q., Yu, Q., Jiang, J., Du, X. P., Huang, L. L., Zhao, L. L., et al. (2016). miR-517a is an independent prognostic marker and contributes to cell migration and invasion in human colorectal cancer. *Oncol. Lett.* 11, 2583–2589. doi: 10.3892/ol.2016.4269

Markossian, S., Guyot, R., Richard, S., Teixeira, M., Aguilera, N., Bouchet, M., et al. (2018). CRISPR/Cas9 editing of the mouse thra gene produces models with variable resistance to thyroid hormone. *Thyroid* 28, 139–150. doi: 10.1089/thy.2017.0389

Michikawa, C., Uzawa, N., Kayamori, K., Sonoda, I., Ohyama, Y., Okada, N., et al. (2012). Clinical significance of lymphatic and blood vessel invasion in oral tongue squamous cell carcinomas. *Oral Oncol.* 48, 320–324. doi: 10.1016/j.oraloncology.2011.11.014

Omura, K. (2014). Current status of oral cancer treatment strategies: surgical treatments for oral squamous cell carcinoma. *Int. J. Clin. Oncol.* 19, 423–430. doi: 10.1007/s10147-014-0689-z

Shi, X. M., and Wang, X. T. (2019). LINC00473 mediates cyclin D1 expression through a balance between activation and repression signals in breast cancer cells. *FEBS Lett.* 593, 751–759. doi: 10.1002/1873-3468.13353

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* 329, 689–693. doi: 10.1126/science.1192002

Wang, F., Ren, D. L., Liang, X. L., Ke, S. W., Zhang, B. W., Hu, B., et al. (2019). A long noncoding RNA cluster-based genomic locus maintains proper development and visual function. *Nucleic Acids Res.* 47, 6315–6329. doi: 10.1093/nar/gkz444

Wang, X. T., Arai, S., Song, X. Y., Reichart, D., Du, K., Pascual, G., et al. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454, 126–130. doi: 10.1038/nature06992

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

Wassarman, D. A., and Steitz, J. A. (1991). Structural-analyses of the 7sk Ribonucleoprotein (Rnp), the most abundant human small Rnp of unknown function. *Mol. Cell. Biol.* 11, 3432–3445. doi: 10.1128/Mcb.11.7.3432

Wu, F., Chen, W., Kang, X., Jin, L., Bai, J., Zhang, H., et al. (2020). A seven-nuclear receptor-based prognostic signature in breast cancer. *Clin. Transl. Oncol.* doi: 10.1007/s12094-020-02517-1

Yang, C. C., Yu, H. C., Chen, R., Tao, K., Jian, L., Peng, M. X., et al. (2019). CXCL1 stimulates migration and invasion in ER-negative breast cancer cells via activation of the ERK/MMP2/9 signaling axis. *Int. J. Oncol.* 55, 684–696.

Zhang, S. Y., Ma, H. Y., Zhang, D. M., Xie, S. L., Wang, W. W., Li, Q. X., et al. (2018). LncRNA KCNQ1OT1 regulates proliferation and cisplatin resistance in tongue cancer via miR-211-5p mediated Ezrin/Fak/Src signaling. *Cell Death Dis.* 9:742. doi: 10.1038/s41419-018-0793-5

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

The reviewer XJ declared a shared affiliation, with no collaboration, with several of the authors SM, LW, GY to the handling editor at the time of the review.

frontiers in Genetics

REVIEW
published: 26 August 2021
doi: 10.3389/fgene.2021.668313

Epigenetic Regulation of the Vascular Endothelium by Angiogenic LncRNAs

Noeline Subramaniam[1,2], *Ranju Nair*[2,3] *and Philip A. Marsden*[1,2,3,4]*

[1] Marsden Lab, Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada, [2] Marsden Lab, Keenan Research Centre in the Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada, [3] Marsden Lab, Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada, [4] Department of Medicine, University of Toronto, Toronto, ON, Canada

The functional properties of the vascular endothelium are diverse and heterogeneous between vascular beds. This is especially evident when new blood vessels develop from a pre-existing closed cardiovascular system, a process termed angiogenesis. Endothelial cells are key drivers of angiogenesis as they undergo a highly choreographed cascade of events that has both exogenous (e.g., hypoxia and VEGF) and endogenous regulatory inputs. Not surprisingly, angiogenesis is critical in health and disease. Diverse therapeutics target proteins involved in coordinating angiogenesis with varying degrees of efficacy. It is of great interest that recent work on non-coding RNAs, especially long non-coding RNAs (lncRNAs), indicates that they are also important regulators of the gene expression paradigms that underpin this cellular cascade. The protean effects of lncRNAs are dependent, in part, on their subcellular localization. For instance, lncRNAs enriched in the nucleus can act as epigenetic modifiers of gene expression in the vascular endothelium. Of great interest to genetic disease, they are undergoing rapid evolution and show extensive inter- and intra-species heterogeneity. In this review, we describe endothelial-enriched lncRNAs that have robust effects in angiogenesis.

Keywords: endothelial cell, lncRNA, angiogenesis, epigenetics, gene regulation, vascular

OPEN ACCESS

Edited by:
Xiao Li,
Texas Heart Institute, United States

Reviewed by:
Tijana Mitic,
University of Edinburgh, United Kingdom
Chien-Ling Huang,
Hong Kong Polytechnic University, China

***Correspondence:**
Philip A. Marsden
p.marsden@utoronto.ca

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *16 February 2021*
Accepted: *17 May 2021*
Published: *26 August 2021*

Citation:
Subramaniam N, Nair R and Marsden PA (2021) Epigenetic Regulation of the Vascular Endothelium by Angiogenic LncRNAs. Front. Genet. 12:668313. doi: 10.3389/fgene.2021.668313

INTRODUCTION

The cardiovascular system is a complex and dynamic network of blood vessels pumping blood from the heart to the rest of the body. Through the blood vessels, nutrients and oxygen are delivered to the cells, and carbon dioxide and waste products are removed. This occurs at a rapid rate in advanced species, and to maintain this, there is a tight interplay of multiple hemodynamic forces including circumferential stretch, hydrostatic pressure, shear stress and rates of blood flow. At the interface between the circulating blood and the vascular wall is the vascular endothelium, acting as a dynamic barrier. The endothelium is a monolayer of endothelial cells (ECs) that lines the entire closed cardiovascular system. We have previously argued that ECs are professional sensors of hemodynamic forces (Ku et al., 2019). ECs sense and respond to these forces, which in turn affect EC phenotype.

A finite number of cis-DNA elements and associated trans-factors mediate the nuclear-based response to the interplay of these varied hemodynamic factors. One such cis-DNA element that is activated by atheroprotective, laminar flow is the shear stress response element (SSRE), which was first identified in the promoter region of platelet-derived growth factor-B (PDGF-B) (Resnick et al., 1993). We now know that the SSRE is detected in many other flow-regulated EC

genes such as intercellular adhesion molecule 1 (ICAM1), endothelin-1 (ET-1/EDN1), monocyte chemoattractant protein 1 (MCP1)/chemokine (C-C motif) ligand 2 (CCL2), and the prototypic EC gene responsible for nitric oxide production, endothelial nitric oxide synthase (eNOS) (Ku et al., 2021). Another important flow-regulated cis-DNA element is Krüppel-like factor (KLF). The KLFs are zinc-finger transcription factors (Anderson et al., 1995). In particular, it is known that KLF2 and KLF4 are important flow-regulated transcription factors that signal through the MEF5/ERK5/MEF2 pathway to mediate transcription of many flow-responsive genes (Dekker et al., 2002b; Parmar et al., 2005; Ohnesorge et al., 2010; Le et al., 2013; Sangwung et al., 2017). KLF2 regulates vascular tone by inducing expression of eNOS (Dekker et al., 2005; Parmar et al., 2005). Models of atherosclerosis confirm the critical role for these KLF transcription factors in vascular homeostasis. In apolipoprotein E (ApoE) deficient mice with hemizygous KLF2 deficiency, there is a notable increase in atherosclerosis (Atkins et al., 2008). Similar results are observed with EC-specific loss of KLF4 in ApoE deficient mice (Zhou et al., 2012). Together, these studies demonstrate an important atheroprotective role for KLF2 and KLF4. Finally, there is the myocyte enhancer factor-2 (MEF2) family, members of which bind to the promoter region of KLF2 and regulate its expression under shear stress (Kato et al., 1997; Parmar et al., 2005; Wang et al., 2010). EC-specific deletions in mice of MEF2 factors, Mef2a, -c, and -d, disrupt vascular homeostasis (Lu et al., 2021). Combined deletion of these MEF2 factors significantly decreased KLF2/KLF4 expression. In summary, these independent cis-DNA elements, namely the SSRE, KLF and MEF2 elements mediate transcriptional responses to changes in shear stress.

Adding to this classic cis-trans paradigm, EC gene expression is also regulated by epigenetic mechanisms. Broadly defined, epigenetics refers to chromatin-based mechanisms important in the regulation of gene expression that do not involve changes to the DNA sequence *per se* (Matouk and Marsden, 2008; Yan et al., 2010; Webster et al., 2013). Epigenetic mechanisms include DNA methylation, histone modifications and RNA-based mechanisms, including long non-coding RNAs (lncRNAs). They are highly responsive to changes in the environment, making them quite dynamic. Epigenetic mechanisms have profound effects on many biological processes in which ECs participate in, especially hemodynamic regulation and angiogenesis. Short non-coding RNAs, such as microRNAs (miRNAs), have been gaining scientists' attention for their post-transcriptional effects since the early 2000s (Battistella and Marsden, 2015). In recent years, a newer class of non-coding RNAs called long non-coding RNAs have emerged, and they have been shown to be important regulators of gene expression in health and disease. LncRNAs tend to be enriched in the nucleus, where they can act as epigenetic modifiers of gene expression (Man and Marsden, 2019). Some of the best studied lncRNAs are X-inactive specific transcript (XIST) which is known to inactivate one of the X chromosomes in females; HOX transcript antisense RNA (HOTAIR), which is involved in limb development; and antisense non-coding RNA in the Inhibitors of CDK4 (INK4) locus (ANRIL), which is strongly correlated with cardiovascular disease risk. The identification and characterization of angiogenic lncRNAs has introduced the idea that lncRNAs may serve as biomarkers and/or therapeutic targets for diseases in which angiogenesis is disrupted, such as in cancers or cardiovascular disease. In this review, we will discuss how nuclear endothelial-enriched lncRNAs, affect EC angiogenesis. We will especially highlight the STEEL, GATA6-AS and MANTIS lncRNAs.

LONG NON-CODING RNAs

Historically believed to be "transcriptional noise" or "dark matter," lncRNAs have emerged as key modulators of many biological processes. Scientists have identified thousands of lncRNAs, with the online database "LncBook" citing > 270,000 lncRNAs in humans (Ma et al., 2019). However, the number of lncRNAs that have been functionally characterized is ∼<1% of those identified (Quek et al., 2015). LncRNAs are primarily characterized by their length as >200 nucleotides long, mainly to distinguish this class of non-coding molecules from shorter transcripts (Mercer et al., 2009; Mattick and Rinn, 2015). LncRNAs can be 5′-capped, spliced, polyadenylated, and often have low expression levels relative to protein-coding genes (Guttman et al., 2009; Derrien et al., 2012). As their name suggests, they are not translated into proteins and thus, often have trivial or non-functional open reading frames (ORFs). This can be assessed through bioinformatic analysis of coding domain sequence, secondary structure, di/tri-nucleotide sequence frequencies and cross-species conservation (Ulitsky and Bartel, 2013; Ventola et al., 2017). LncRNAs can be classified based on several criteria, but broadly are often grouped by their organization relative to other genes, due to a lack of clarity on their sequence-structure-function relationship. They can be described as intronic, intergenic, antisense, bidirectional, enhancer, or promoter-associated lncRNAs.

LncRNAs can be present in the nucleus, cytoplasm, or mitochondria, and they may also be secreted (Rackham et al., 2011; van Heesch et al., 2014; see **Figure 1**). A lncRNA can also be expressed in multiple compartments, such as GAS5, which is expressed in both the nucleus and the cytoplasm (Kino et al., 2010). Since subcellular localization often confers function, the mechanism of action of a lncRNA can be inferred, in part, by defining where they are targeted (Cai and Cullen, 2007; Quinodoz and Guttman, 2014; Romero-Barrios et al., 2018; Hou et al., 2019; Mishra and Kanduri, 2019; Rom et al., 2019). Nuclear lncRNAs like XIST or HOTAIR are often important mediators of regulating epigenetic mechanisms (Mercer et al., 2009). They can act in *cis* or in *trans* by interacting with neighboring or non-neighboring genes to exert their effects (Rinn et al., 2007; Wang et al., 2011; Engreitz et al., 2013; Novikova et al., 2013). Cytoplasmic lncRNAs like Tie1-AS can interact with protein-coding genes, and others still can act as a scaffold for protein-protein interactions (Li et al., 2010). Many lncRNAs, including ones we will highlight in this review regulate gene expression through chromatin-based mechanisms.

Finally, non-coding RNAs can be enriched in particular cell- or tissue-types. For example, spliced-transcript endothelial-enriched lncRNA (STEEL) is an endothelial-enriched

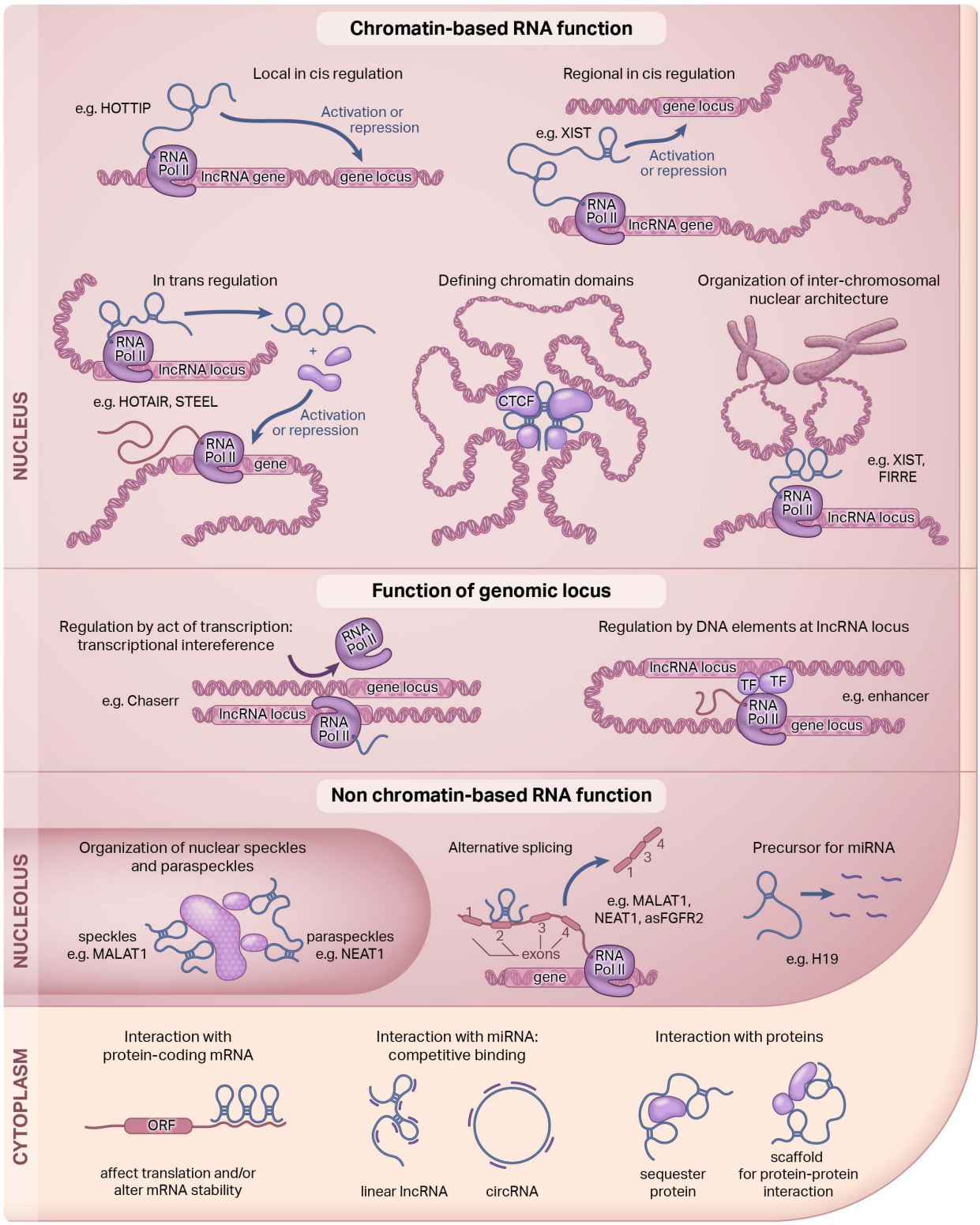


FIGURE 1 | Functions of long non-coding RNAs. Long non-coding RNAs (lncRNAs) are a diverse class of molecules that are distributed throughout the cell. The function of lncRNAs are dependent, in part, on their subcellular localization. Here, we illustrate known lncRNA functions in the nucleus and the cytoplasm. Our review focuses on the nuclear function of lncRNAs. This figure was adapted; it was originally published in Current Opinion in Pharmacology, Volume 45, Hon-Sum Jeffrey Man and Philip A Marsden, LncRNAs and epigenetic regulation of vascular endothelium: genome positioning system and regulators of chromatin modifiers, pp. 72–80, Copyright Elsevier, 2019.

lncRNA that has EC-specific functions (Man et al., 2018). In contrast, there are lncRNAs like metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) that is expressed at very high levels and is found widely across almost all cell types. Some lncRNAs are highly upregulated by environmental stimuli that are especially relevant to ECs, such as the lncRNA that enhances eNOS expression (LEENE), which is induced under physiological blood flow and pulsatile shear stress (Miao et al., 2018). Another example is GATA6-AS, which is hypoxia-responsive (Neumann et al., 2018). Disease-associated lncRNAs have also been identified, with the best example being ANRIL (McPherson et al., 2007; Holdt et al., 2010; Congrains et al., 2012). Through genome-wide association studies (GWAS), exons 13-19 of ANRIL comprise a disease-associated haplotype noted for a marked increased risk of coronary artery disease (CAD). Now we know that it is also associated with other cardiovascular diseases, such as ischemic peripheral vascular disease and ischemic stroke (Zeggini et al., 2007; Foroud et al., 2012; Kremer et al., 2015; Kong et al., 2016; Tan et al., 2019). Evidently, lncRNAs are biologically important and functionally diverse.

How to Detect lncRNAs

There are many methods to identify lncRNAs, but microarrays were by far the method of choice for a long time—until the advent of deep RNA-sequencing (RNA-seq; **Table 1**; Mockler et al., 2005). While microarrays are high throughput, cost-effective, and computationally manageable, they also limit novel lncRNA discovery due to pre-determined probe sets; optimal probe coverage and density; and background noise from cross-hybridizations or weak binding (Uchida, 2017). Deep RNA-seq has become the current method of choice because it enables researchers to discover non-annotated transcripts, single nucleotide variations (SNVs), splice variants, novel splice junctions and gene fusion events (Sultan et al., 2008; Trapnell et al., 2009; Edgren et al., 2011; Djebali et al., 2012; Quinn et al., 2013). Moreover, deep RNA-seq has greater specificity and sensitivity enabling detection of low expression and rare transcripts as well as cell- and tissue-specific lncRNAs (Wang et al., 2014; Liu et al., 2015; Li et al., 2016). Thus, an important parameter of RNA-seq for accurate quantification is read depth. For RNAs of moderate abundance, ∼30–40 million reads are needed whereas for higher coverage (e.g., detecting rare and lowly expressed transcripts), reads of up to 500 million are recommended (Fu et al., 2014).

RNA-seq and next generation sequencing (NGS) has expanded significantly and over 400 methods have been established over the last decade (Hadfield and Retief, 2018). Emerging approaches include single-cell RNA-seq (scRNA-seq) which examines gene expression at a single-cell resolution, or assay for transposase-accessible chromatin sequencing (ATAC-seq) which locates regions of open chromatin, in genomic regions devoid of protein-coding genes. However, RNA-seq is not without its challenges. The analyses are more difficult and require more computational power; it is more costly; and multiple cycles of polymerase chain reaction (PCR) may introduce some amplification bias (Uchida, 2017). Importantly, lncRNA discovery is limited by annotations and genome build accuracy.

LncRNAs are difficult to annotate because of their low expression levels, our limited understanding of their sequence-function relationship and their lack of evolutionary conservation. Thus, lncRNAs are currently annotated primarily based on transcriptomic evidence (Uszczynska-Ratajczak et al., 2018). The 2 main annotation approaches are automated or manual. With manual annotation, humans strategically put together transcriptomic and genomic data to build models that can create relatively accurate annotations. Automated annotation uses transcriptome assembly approaches that are quick and not costly, but typically result in incomplete and inaccurate annotations. To date, manual annotations are more accurate. Moreover, manual annotations have a higher quality assessment of lncRNA coding potential (via mass spectrometry, PhyloCSF, UniProt, and Pfam) (Sonnhammer et al., 1998; Apweiler et al., 2004; Lin et al., 2011). The most widely used manual annotation is GENCODE, followed by Reference Sequence (Refseq) (Harrow et al., 2012; Pruitt et al., 2014; Fang et al., 2018). There has been a historical bias toward using cell lines, adult tissues and tumor samples to build these reference databases. LncRNAs specifically expressed in rare cell populations, in response to various environmental stimuli, and in development may be excluded from these annotations. As technologies continue to advance, we predict that these databases will become more comprehensive with higher confidence lncRNA annotations.

How to Study lncRNAs

In order to study lncRNAs and their epigenetic and non-epigenetic functions, an arsenal of molecular biology techniques are employed by scientists. In **Table 2**, we outline the most commonly used methodologies to study lncRNA-chromatin and lncRNA-protein interactions. To identify chromatin associated lncRNAs, approaches have been broadly classified into either "one-to-many" or "all-to-all." One-to-many approaches include

TABLE 1 | Methods to detect a lncRNA.

Method	Function	Advantages	Disadvantages	Example lncRNA
Microarrays	To detect RNAs	High throughput; computationally manageable	Background noise from cross-hybridizations or weak binding	HOTAIR, STEEL (Rinn et al., 2007; Man et al., 2018)
RNA sequencing	To detect RNAs	Able to detect new transcripts; high sensitivity with increased read depth	LncRNA discovery limited by annotations and genome build accuracy	GATA6-AS (Neumann et al., 2018)

TABLE 2 | Methods to study lncRNA function.

Method	Function	Advantages	Disadvantages	LncRNA example	Method Reference
Chromatin Isolation by RNA purification (ChIRP)	To identify lncRNA-chromatin interactions	Probes can be designed without knowledge of structure or functional domains of target lncRNA	Background noise from non-specific binding	MEG3, HOTAIR (Chu et al., 2011; Iyer et al., 2017)	Chu et al., 2011
Capture hybridization analysis of RNA targets (CHART)	To identify lncRNA-chromatin interactions	Probes can be designed without knowledge of structure or functional domains of target lncRNA	Background noise from non-specific binding	NEAT1, MALAT1 (West et al., 2014)	Simon et al., 2011
RNA antisense purification (RAP)	To identify lncRNA-chromatin interactions	Longer probes mitigate background noise	Need probes to overlap entire length of lncRNA for capture	XIST, FIRRE (Engreitz et al., 2013; Hacisuleyman et al., 2014)	Engreitz et al., 2013
Mapping RNA-genome interactions (MARGI)	To identify lncRNA-chromatin interactions	Identifies native RNA-chromatin interactions *in vivo* and *in vitro*	Moderate sensitivity might decrease detection of low abundance chromatin-associated RNAs	XIST, SNHG1, NEAT1, MALAT1 (Sridhar et al., 2017)	Sridhar et al., 2017
Global RNA interaction with DNA sequencing (GRID-seq)	To identify lncRNA-chromatin interactions	Identifies genome-wide lncRNA-chromatin interactions *in situ*	Moderate sensitivity might decrease detection of low abundance chromatin-associated RNAs	MALAT1, NEAT1 (Li X. et al., 2017)	Li X. et al., 2017
Chromatin-associated RNA sequencing (ChAR-seq)	To identify lncRNA-chromatin interactions	Identifies genome-wide lncRNA-chromatin interactions *in situ*	Moderate sensitivity might decrease detection of low abundance chromatin-associated RNAs	*roX1* and *roX2* in *Drosophila* (Bell et al., 2018)	Bell et al., 2018
RNA and DNA interacting complexes ligated and sequenced (RADICL-seq)	To identify lncRNA-chromatin interactions	RNase H and actinomycin D decrease bias for nascent transcripts; improved genomic coverage and unique mapping efficiency	Moderate sensitivity might decrease detection of low abundance chromatin-associated RNAs	MALAT1 (Bonetti et al., 2020)	Bonetti et al., 2020
Chromosome conformation capture (3C)	Characterizing spatial topology of long-range DNA interactions	High throughput; many variations have emerged	Risk of artifacts during data analysis	ANRIL (Nakaoka et al., 2016)	Dekker et al., 2002a; Lieberman-Aiden et al., 2009; Han et al., 2018
RNA Immunoprecipitation (RIP)	Identifies lncRNA-protein interactions	Sensitive and specific for RNA detection	Need specific antibodies for protein targets	HOTTIP (Hu et al., 2019)	Lerner and Steitz, 1979; Ule et al., 2018
Cross-linking and immunoprecipitation (CLIP)	Identifies lncRNA-protein interactions	No nucleases used; do not need special reagents/equipment	UV light can cause mutations; low sensitivity	NEAT1 (Wen et al., 2020)	Ule et al., 2003
RNA pull-down	Identifies lncRNA-protein interactions	Improved discovery of weak/transient binding	Artificially increasing lncRNA of interest may generate false positives	STEEL (Man et al., 2018)	Marín-Béjar and Huarte, 2015
Chromatin immunoprecipitation (ChIP)	Identifies proteins associated with specific genomic regions	Can identify histone proteins and histones with modifications (e.g., methylation, acetylation)	Need specific antibodies	HOTTIP and ANRIL (Nakaoka et al., 2016; Hu et al., 2019)	Massie and Mills, 2012; Xie et al., 2016
RNA fluorescence *in situ* hybridization (RNA FISH)	Visualization of subcellular localization of lncRNA	Branched chain approaches can detect low abundance lncRNAs with single-cell resolution	Hybridization artifacts	MEG3 (Cabili et al., 2015)	Gall and Pardue, 1969
RNA interference (e.g., siRNA, shRNA)	Guilt by association defined by lncRNA knockdown	High knockdown efficiency	Potential off-target effects could lead to decreased specificity; transfection method artifacts	STEEL (Man et al., 2018)	Dorsett and Tuschl, 2004; Taxman et al., 2006
Antisense oligonucleotides (ASOs)	To silence lncRNA in order to assess function	Ideal for targeting non-coding nuclear RNAs	Potential off-target effects could lead to decreased specificity	MALAT1 (Gong et al., 2019)	Crooke, 2017
Clustered regularly interspersed short palindromic Repeats (CRISPR)	Ablate native lncRNA locus	Can edit any regulatory element (e.g., enhancer, promoter, etc.); no mediator machinery involved	Need to have "CRISPRable" genomic locus; challenge to study primary human cell types	MANTIS, PRANCR (Leisegang et al., 2017; Cai et al., 2020)	Jinek et al., 2012

chromatin isolation by RNA purification (ChIRP), capture hybridization analysis of RNA targets (CHART) and RNA antisense purification (RAP) (Chu et al., 2011; Simon et al., 2011; Engreitz et al., 2013). These methods are based on hybridizations of biotin-labeled probes targeted to lncRNAs of interest, followed by pull-down of the associated chromatin fraction. These techniques are limited by background noise from non-specific binding. RAP has relatively less background noise due to the use of longer probes, but these capture approaches could all be improved with further background corrections (Li and Fu, 2019). ChIRP, CHART, and RAP can be combined with other techniques to discover more interactions within a single experiment. One example is domain-specific ChIRP (dsChIRP), a method by which lncRNAs domain-by-domain are assessed to identify functional elements (Quinn et al., 2014). Deep sequencing and/or mass spectrometry can be also be combined to obtain more high-resolution data on lncRNA interactions.

All-to-all approaches enable global detection of RNA-chromatin interactions across all RNAs. These methods include mapping RNA-genome interactions (MARGI), global RNA interaction with DNA sequencing (GRID-seq), and chromatin-associated RNA sequencing (ChAR-seq) (Li X. et al., 2017; Sridhar et al., 2017; Bell et al., 2018). More recently, RADICL-seq (RNA and DNA interacting complexes ligated and sequenced) was developed (Bonetti et al., 2020). These techniques are based on a bivalent linker in which one end ligates to an RNA and the other end ligates to a restriction-digested DNA. Specifically, MARGI maps chromatin-RNA interactions through ligation of a lncRNA to its target genomic sequences, generating RNA-DNA chimeric sequences prior to sequencing. GRID-seq and ChAR-seq are very similar but GRID-seq employs a linker with 2 restriction sites for a type IIS restriction enzyme, MmeI, followed by digestion to produce size-specific products. On the other hand, ChAR-seq employs sonication of ligated products to generate smaller fragments prior to library preparation. RADICL-seq is similar to GRID-seq in that its linker also has 2 restriction sites (for *Eco*P15I), but it also uniquely employs RNase H and actinomycin D to reduce bias toward abundant nascent transcripts. These all-to-all methods are limited by their moderate sensitivity, which could decrease the detection of low abundance chromatin-associated RNAs.

To visualize 3D chromatin interactions, chromosome conformation capture (3C) has historically been used to characterize long-range DNA contacts (Dekker et al., 2002a; Han et al., 2018). Many adaptations of this technique exist, including chromosome conformation capture-on-chip (4C), or chromosome conformation capture carbon copy (5C). These variations require knowledge of the target loci, but another adaptation referred to as Hi-C, provides unbiased locations of chromatin interactions across the genome (Lieberman-Aiden et al., 2009).

To study lncRNA-protein interactions, scientists utilize RNA immunoprecipitation (RIP), cross-linking and immunoprecipitation (CLIP), and RNA pull-down (Lerner and Steitz, 1979; Ule et al., 2003, 2018; Marín-Béjar and Huarte, 2015). Another immunoprecipitation-based approach is chromatin immunoprecipitation (ChIP) which identifies protein-DNA interactions (Massie and Mills, 2012; Xie et al., 2016). These techniques have further advanced in recent years. For example, RIP can be combined with APEX (engineered ascorbate peroxidase)-catalyzed proximity biotinylation of endogenous proteins (APEX-RIP) to improve the spatial resolution of RNA mapping (Kaewsapsak et al., 2017). A recent study used APEX-RIP to generate a transcriptome-wide RNA atlas (Fazal et al., 2019).

Another powerful technique is RNA fluorescence *in situ* hybridization (FISH) which allows researchers to visualize RNA and DNA molecules while retaining cell morphology (Gall and Pardue, 1969; Collins et al., 1997; Cabili et al., 2015). The quantitative strength of this assay has increased with improvements of the branched DNA signal amplification technology to amplify the signal 1,000–10,000-fold. FISH improves the detection of lowly expressed lncRNAs. Recently, fluorescence in situ RNA sequencing (FISSEQ) was developed (Lee et al., 2015). FISSEQ provides high throughput information on tissue-specific gene expression while maintaining spatial context.

To silence lncRNAs and study their functions, there are many well-established techniques. There is RNA interference (RNAi), which is widely used, and is very efficient at knocking down RNAs (Watts and Corey, 2012; Chery, 2016). RNAi works best at targeting cytoplasmic lncRNAs. To silence nuclear lncRNAs, antisense oligonucleotides (ASOs) can be used in which RNase H is recruited to hydrolyze RNA in a DNA: RNA complex, causing transcriptional silencing of the target lncRNA. Finally, there is clustered regularly interspaced short palindromic repeats (CRISPR) which is similar to RNAi except that it is not reliant on mediator machinery (Jinek et al., 2012; Awwad, 2019). CRISPR can directly target genomic regions, allowing scientists to target regulatory regions like promoters or enhancers. CRISPR interference (CRISPRi) has been used for large-scale, systematic lncRNA screens in cell lines, demonstrating how this tool can be used to identify lncRNAs and further study their functions (Koch, 2017; Liu et al., 2017, 2020; Cai et al., 2020). Importantly, not all lncRNAs can be studied using CRISPR because CRISPR efficiency is affected by internal or bidirectional promoters (Goyal et al., 2017).

LncRNAs are a highly heterogeneous and functionally diverse class of molecules. As we illustrate in this section, there is a growing number of methods that enable functional characterization of lncRNAs. Many of these methods can be combined to efficiently map the lncRNA interactome. The data generated are increasingly being archived on a multitude of public databases. We direct readers to reviews that outline existing lncRNA databases (Peng et al., 2020; Pinkney et al., 2020). There is no doubt that high-throughput technologies will continue to advance to produce higher quality lncRNA data and improve our overall understanding of their molecular functions.

ANGIOGENESIS

The *de novo* formation of blood vessels from angioblasts and circulating hematopoietic stem cells is called vasculogenesis (Cines et al., 1998). In the absence of healthy vascular

development, embryonic lethality results. Post-natally, new blood vessels form from pre-existing blood vessels in a closed cardiovascular system in a process called angiogenesis. Angiogenesis plays an important role in development, wound healing, and many other physiological processes (Schmidt and Carmeliet, 2010). ECs are instrumental in the orchestration of angiogenesis. It is a highly choreographed cascade of events that involves both exogeneous (e.g., hypoxia and VEGF) and endogenous regulatory inputs. There are two main types of angiogenesis: sprouting angiogenesis and intussusceptive angiogenesis. In sprouting angiogenesis, as the name suggests, new blood vessels grow via "sprout" formation from existing vessels through EC proliferation and migration (Ackermann et al., 2014). In contrast, intussusceptive angiogenesis, also referred to as splitting angiogenesis, there is little dependence on EC proliferation and migration. Instead, the ECs reorganize, and the cells invade the lumen forcing the vessel to split (Konerding et al., 2012). Both forms of angiogenesis, which are believed to occur in nearly all organs and tissues, produce new vasculature, most often in capillary beds (Adair and Montani, 2010). Sprouting angiogenesis, hereafter angiogenesis, is better understood and the most well-studied. As such, it will be the focus of this review.

A multitude of factors work together in a dynamic network to maintain tight regulation of angiogenesis. This is to prevent insufficient or over-vascularization from occurring. A major environmental regulator is hypoxia, which can be defined as an imbalance between oxygen supply and demand. The main intracellular signaling molecule is hypoxia inducible factor (HIF), a transcription factor comprised of HIF-alpha and HIF-beta subunits (Shweiki et al., 1992). The HIF-alpha subunit is functionally regulated by oxygen-dependent post-translational modifications of prolyl residues. In normoxia, the prolyls are hydroxylated, and Von Hippel-Lindau (VHL) recruits the E3 ubiquitin ligase complex to ubiquitinate the HIF-alpha subunits and make them a target for proteasomal degradation (Chen et al., 2009). In contrast, in hypoxic conditions, there is no hydroxylation and therefore no subsequent ubiquitination. Thus, HIF is not degraded and instead accumulates and translocates to the nucleus. Together with the constitutively expressed HIF-beta, it can bind to the hypoxia-responsive cis-DNA element (HRE) and modify transcription of a number of genes, many of which are involved in angiogenesis. This includes genes like matrix metalloproteinase-2 (MMP-2), angiopoietin-2 (Ang-2), Tie-2, PDGF, Delta-like 4 (DLL4) and many more factors. However, the most important target is arguably vascular endothelial growth factor (VEGF). The development of normal vasculature is heavily dependent on a VEGF gradient (see **Figure 2**). When the gene dose of VEGF is reduced by 50%, it causes embryonic lethality due to vascular deficiencies (Carmeliet et al., 1996; Ferrara et al., 1996). Conversely, overexpression of VEGF, as seen in tumors, causes exuberant EC activation, leading to a disorganized

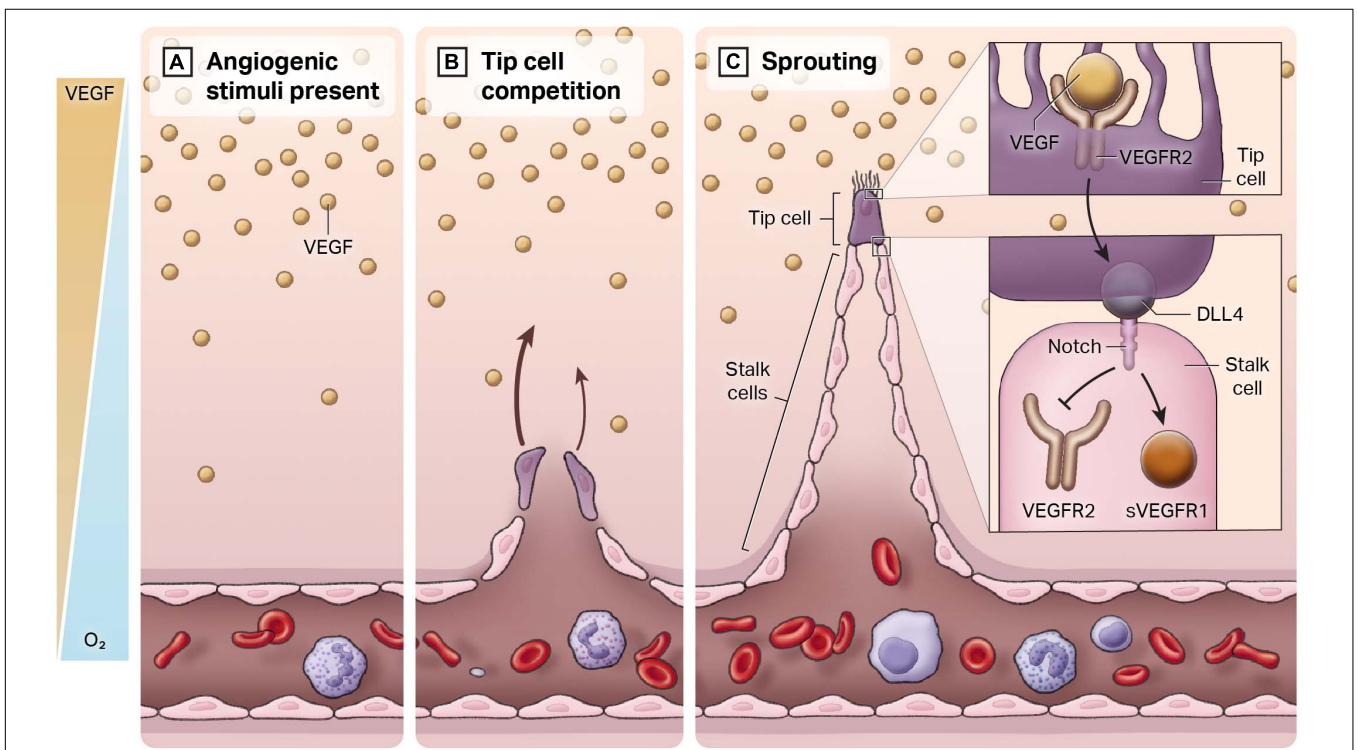


FIGURE 2 | Endothelial heterogeneity in the angiogenic response. New blood vessels develop from a pre-existing closed cardiovascular system, a process termed angiogenesis. **(A)** In regions of hypoxia, there is an increase in the release of the pro-angiogenic ligand VEGF. An inverse gradient is established between free oxygen concentration and VEGF. **(B)** Quiescent endothelial cells respond to VEGF and compete to be the leader or the "tip" of the newly forming blood vessel. **(C)** The cell that becomes the tip cell represses expression of tip markers in adjacent cells, which are known as stalk cells. Stalk cells proliferate to form the body of the vessel.

vasculature (Jain, 2005). Since VEGF is a potent regulator of angiogenesis, VEGF is often targeted therapeutically to treat diseases in which this process is dysregulated.

In healthy adult blood vessels, vascular ECs exhibit a low rate of cell number turnover (Sender and Milo, 2021). Adult humans exhibit approximately 1 trillion ECs, with a mean life span typically of 3–10 years. This long-life span can be contrasted with red blood cells or circulating white blood cells, which approximate 120 and 7 days, respectively. Importantly, the basal low rate of turnover of ECs can be markedly augmented. Our current understanding of angiogenesis can be encapsulated by the tip-stalk paradigm (see **Figure 2**). Once hypoxia has helped establish the VEGF gradient from non-endothelial cell types (e.g., macrophages), ECs respond and begin to migrate toward this angiogenic stimulus. These ECs are called "tip cells," and they are morphologically characterized by filopodia and stress fibers, which facilitate invasion into the surrounding tissue and creates a clear path for sprouting to commence (Eilken and Adams, 2010). VEGF binds to vascular endothelial growth factor receptor 2 (VEGFR2) receptors in tip cells, activating DLL4, which then binds to Notch receptors in adjacent ECs (Adams and Alitalo, 2007). Notch signaling is activated in adjacent ECs and suppresses tip genes, such as VEGFR2 and DLL4, to prevent these ECs from also becoming tip cells. These cells are referred to as "stalk cells." In addition to positional identity, stalk cells are highly proliferative, lack filopodia and contribute to lumen formation. They express soluble VEGFR1 (sVEGFR1), which sequesters VEGF-A in a regulatory manner to prevent VEGF-induced signaling (del Toro et al., 2010).

Once multiple sprouts have formed, tip cells of different sprouts will anastomose with each other, which is believed to occur through filopodial interactions (Bentley et al., 2009). This creates vessel networks. Wnt signaling is also important in angiogenesis. Specifically, the interplay between Notch signaling and Wnt signaling causes an upregulation of β-catenin expression. β-catenin is important in stabilizing tight junctions and activating PDGF-B expression. PDGF-B promotes recruitment of mural cells/support cells, which is an indicator of healthy and mature vessels and prevents vessel "leakiness" (Gavard and Gutkind, 2008; Reis et al., 2012). Lumenization and blood flow will further stabilize these new vessel connections (Chappell et al., 2011; Potente et al., 2011). Once the vessels have been perfused, the ECs will move toward a quiescent phenotype. It is important to note that the underlying mechanisms, in particular of these final phases of sprouting angiogenesis, are not well understood.

As inferred from above, hemodynamic forces must regulate angiogenesis. In regions of the body in which angiogenesis occurs such as in muscle that undergoes remodeling following exercise, there are markedly increased levels of shear stress. In contrast, there is low shear stress in a tumor. Of note, angiogenesis occurs at both these levels of shear stress, though in opposite directions (Kaunas et al., 2011). In normal physiological conditions, shear stress varies depending on the vessel type. In arteries, shear stress can range between 10 and 70 dynes/cm^2, whereas in veins, shear stress can range between 1 and 6 dynes/cm^2 (Lipowsky, 1995; Malek, 1999). Thus, it is not the absolute value of shear stress that induces angiogenesis, but instead, the deviation from normal levels detected by the ECs that stimulates this physiological response.

NUCLEAR ENDOTHELIAL-ENRICHED ANGIOGENIC lncRNAs

In the last several years, lncRNAs that regulate biological processes like angiogenesis have emerged. In this section, we will highlight angiogenic lncRNAs that act through chromatin-based mechanisms to effect angiogenesis, including STEEL, GATA6-AS, and MANTIS (**Table 3**). These lncRNAs are also of particular interest because they are regulated by environmental stimuli that regulate angiogenesis: hemodynamic forces and hypoxia.

STEEL

STEEL was the first flow-regulated, endothelial-enriched lncRNA identified. STEEL is a pro-angiogenic lncRNA that links decreased laminar shear stress with pro-angiogenic programming. It is downregulated by laminar flow (10 dynes/cm^2) and enriched in microvascular ECs (Man et al., 2018). In static conditions, STEEL maintains eNOS and KLF2 expression at basal levels required for angiogenesis. Consistent with its nuclear localization, we found evidence that STEEL regulates eNOS and KLF2 by transcriptional mechanisms. First, STEEL knockdown decreased heterogeneous nuclear RNA (hnRNA) levels and decreased RNA polymerase II (Pol II) loading at the proximal promoters of eNOS and KLF2. Moreover, STEEL regulates chromatin accessibility, nucleosome occupancy and histone 3 lysine 4 trimethylation (H3K4me3) at both eNOS and KLF2 proximal promoters. Interestingly, a feedback loop exists in which laminar shear stress (10 dynes/cm^2) induces high levels of eNOS and KLF2, which in turn represses STEEL expression. In this way, STEEL functions as a rheostat of angiogenesis that responds to shear stress conditions. Mechanistically, they identified a lncRNA-protein interaction that presents a new mechanism for genomic targeting of the poly-ADP ribosylase 1 (PARP1), which contributes to transcriptional regulation, DNA damage repair, and cardiovascular disease (Chaudhuri and Nussenzweig, 2017). Using RNA pulldown followed by mass spectrometry and RIP to identify and confirm this interaction, respectively, and ChIP to demonstrate an effect of STEEL knockdown on PARP1 occupancy at the eNOS and KLF2 promoters. Together, these mechanistic studies of the STEEL lncRNA provide evidence for epigenetic regulation of gene expression and novel lncRNA-protein interactions. The underlying mechanisms of these interactions require further investigation.

Regarding its angiogenic functions, STEEL was shown to affect blood vessel formation both *in vitro* and *in vivo*. Using a Matrigel network assay in HUVEC, STEEL knockdown decreased network formation while STEEL overexpression increased network formation (Man et al., 2018). Cell migration is characteristic of tip cells and cell proliferation is characteristic of stalk cells. Both assessments are key measures of sprouting angiogenesis and were therefore examined. As expected, through scratch wound assay, carboxyfluorescein

TABLE 3 | Summary of endothelial-enriched angiogenic lncRNA function.

LncRNA	Epigenetic mechanism of action	Function in angiogenesis	References
STEEL	In static conditions, STEEL is recruited to EC proximal promoter regions, eNOS and KLF2, where STEEL's association will enhance transcription.	STEEL increases angiogenesis *in vitro* and *in vivo.*	Man et al., 2018
GATA6-AS	Hypoxia-responsive GATA6-AS interacts with LOXL2 and deactivates H3K4me3 to repress transcription of COX-2 and POSTN.	GATA6-AS increases sprouting *in vitro* but decreases blood vessel formation *in vivo.*	Neumann et al., 2018
MANTIS	When JARID1B is repressed, MANTIS increases and associates with BRG1, which is associated with the SWI/SNF remodeling complex and is stabilized by BAF155.	MANTIS increases transcription of pro-angiogenic factors (i.e., COUP-TFII, SMAD6, SOX18) *in vitro*.	Leisegang et al., 2017, 2019

succinimidyl ester (CFSE) labeling and bromodeoxyuridine (BrdU) incorporation, STEEL overexpression promoted EC proliferation and migration. These results were confirmed *in vivo* using a mouse model. Collagen modules with stromal cells were coated with either control or STEEL overexpressing ECs transduced and implanted into immunocompromised mice. Using micro-computed tomography (micro-CT) imaging, it was found that STEEL-transduced implants had more vessels that were perfused compared to control implants, which not only had fewer vessels, but also demonstrated extravasation and pooling. Further examination of the STEEL-transduced implants' vascular networks revealed that the blood vessels showed mural cell support, as indicated by smooth muscle actin staining, demonstrating vessel maturity. Of note, STEEL-induced angiogenesis did not display evidence of morphologically abnormal vessels (e.g., contrast leakage, lack of pericyte coverage).

GATA6-AS

Another endothelial-enriched lncRNA is the antisense transcript of GATA6 (GATA6-AS). GATA6-AS is upregulated approximately 2.5-fold under chronic hypoxia (24 h) and localized primarily to the nucleus (Neumann et al., 2018). Using mass spectrometry, it was found that GATA6-AS interacts with lysyl oxidase-like 2 (LOXL2), a known hypoxia regulator (Bignon et al., 2011). Nuclear LOXL2 is a known co-repressor of transcriptional activity and deactivates H3K4me3. With GATA6-AS repression, there was a 30% decrease in H3K4me3, which also occurred under hypoxia. Curiously, when LOXL2 was repressed, H3K4me3 was increased, and the majority of GATA6-AS regulated genes were inversely expressed when compared to GATA6-AS repression. Cyclooxygenase-2 (COX-2) and periostin (POSTN) were more closely examined with ChIP-PCR. COX-2 catalyzes the production of prostaglandins in ECs, which contributes to flow-mediated vasodilation whereas POSTN acts through Erk/VEGF signaling to stimulate angiogenesis (Koller et al., 1993; Duffy et al., 1998). GATA6-AS silencing markedly decreased H3K4me3 at the promoter regions of both these genes, pointing toward an epigenetic role for GATA6-AS on EC gene expression.

It was argued that GATA6-AS regulates angiogenesis through endothelial-to-mesenchymal transition (EndMT), a process that can be induced by hypoxia (Neumann et al., 2018). Using an EndMT-assay, repressing GATA6-AS in HUVECs largely inhibited EndMT. Further *in vitro* analysis of angiogenesis using a spheroid assay showed that GATA6-AS silencing significantly decreased sprouting. GATA6-AS repression decreased cell migration, but it did not affect proliferation or apoptosis. The effect GATA6-AS has on the EndMT process may in turn, be affecting angiogenic potential. Using an *in vivo* immune deficient mouse model, HUVEC transfected with control GapmeRs or GapmeRs against GATA6-AS were transplanted. Through histological visualization, GATA6-AS repressed cells had a marked increase in the number of perfused, mature blood vessels compared to controls. These *in vivo* findings are contradictory to the *in vitro* results. The decrease in sprouting *in vitro* may be compensated for through other mechanisms *in vivo*. LOXL2 in the extracellular matrix is also involved in angiogenesis. When LOXL2 was repressed, there was a decrease in sprouting. Interestingly, LOXL2 did not decrease with GATA6-AS silencing. It is clear GATA6-AS is regulating angiogenesis, but the mechanisms require further study.

MANTIS

MANTIS is a flow-regulated lncRNA expressed by ECs. This nuclear lncRNA was identified through inhibition of an EC-enriched H3K4 lysine-specific demethylase 5B (JARID1B) (Leisegang et al., 2017). MANTIS is not specific to ECs; it is also expressed by smooth muscle cells. Steady, laminar flow upregulates MANTIS and is mediated through KLF2 and KLF4 (Leisegang et al., 2019). Moreover, when MANTIS was repressed, HUVEC were unable to align in the direction of flow. Using ChIP, JARID1B was bound to an H3K4me3 region near the MANTIS transcription start site (TSS), which was further increased with JARID1B silencing. Using mass spectrometry, MANTIS was found to be highly associated with Brahma Related Gene 1 (BRG1), an ATPase involved in the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeling complex and important for EC function. BRG1 is stabilized by BAF155. With MANTIS silencing, there was marked reduction in BRG1 and BAF155 binding and the ability of BRG1 to bind to target promoters.

Using CRISPR/Cas9, MANTIS was functionally inactivated in HUVEC, resulting in significantly less tube formation and sprouting. Silencing MANTIS yielded similar results and also decreased cell migration. This also resulted in decreased mRNA and protein expression of factors important in angiogenesis including chicken ovalbumin upstream promoter – transcription

factor 2 (COUP-TFII), SMAD6 and sex determining region Y-box 18 (SOX18). It should be noted that there are many other tip and stalk genes that are also relevant that were not assessed in this study. When these 3 factors are reduced, there is decreased sprouting and yet, overexpression of these factors does not restore sprouting to normal. However, the regulation of these proteins through MANTIS may be important in maintaining healthy sprouting in ECs. MANTIS was knocked down and ATAC-Seq was conducted. BRG1 protein levels were unchanged, but at the TSS of COUP-TFII, SMAD6 and SOX18, there was a decrease in open chromatin. Using micrococcal nuclease (MN) digestion, there was an increase in nucleosomal formation at the TSS of these 3 genes when MANTIS was decreased. MANTIS repression increased H3K27me3, and decreased RNA Pol II at the TSS of all 3 genes. In addition, silencing MANTIS reduced BRG1 binding at the TSS of COUP-TFII, SMAD6 and SOX18. Since BRG1 is known to play a role in nucleosome remodeling, this may suggest that BRG1's interactions with these proteins may be mediated through MANTIS. Further study of MANTIS is needed to better understand these interactions.

ANGIOGENIC lncRNAs IN DISEASE

The majority of lncRNAs that have been identified and characterized have been in diseases. This include cancers (Jin et al., 2020; Teppan et al., 2020; Zhou et al., 2020; Katsushima et al., 2021), diabetes (Taheri et al., 2020; Xu E. et al., 2020; Ismail et al., 2021), cardiovascular diseases (Fang et al., 2020; Meng et al., 2020; Yeh et al., 2020) and ischemic stroke (Gan et al., 2021; Wolska et al., 2021). In this section, we will focus on angiogenic lncRNAs enriched in disease and their epigenetic functions. We will highlight some of the most extensively studied lncRNAs: MALAT1, MEG3 and ANRIL. A summary of these as well as other more recently published disease-associated angiogenic lncRNAs can be seen in **Table 4** (Zhou et al., 2016; Li Y. et al., 2017; Ruan et al., 2018; Niu et al., 2020; Xu X. et al., 2020; Zhang H. et al., 2020; Biswas et al., 2021).

MALAT1

MALAT1, as the name suggests, is implicated in protean cancer cell types. It is extremely abundant in multiple cell types, including vascular ECs (Gutschner et al., 2013; Michalik et al., 2014; Yan et al., 2016). Primarily localized in the nucleus as part of nuclear speckles, MALAT1 associates with the serine/arginine (SR) family of pre-mRNA splicing factors such as SRSF1/2/3; it plays an important role in alternative splicing. When MALAT1 is silenced, it results in reduced nuclear speckle association of many pre-mRNA splicing factors including SF1, U2AF65, SF3a60, and U2snRNP *in vitro* (Tripathi et al., 2010). MALAT1 may have species-specific function. Unexpectedly, Malat1 knockout mice evidenced no change in nuclear speckle markers compared to wildtype mice (Nakagawa et al., 2012). This finding in mice was confirmed by other studies (Eißmann et al., 2012; Zhang et al., 2012).

MALAT1 also plays a critical role in transcriptional regulation, through direct binding to the 3′ end of actively transcribing gene bodies, and mediating localization of unmethylated proteins in nuclear speckles (Engreitz et al., 2014). MALAT1 functions as a molecular scaffold for unmethylated polycomb 2 proteins (PC2), E2F transcription factor, and histones involved in active transcription and the transcriptional coactivator complex (Yang et al., 2011). MALAT1 has a role in regulating expression of

TABLE 4 | Summary of disease-associated angiogenic lncRNA function.

LncRNA	Disease	Epigenetic mechanism of action	Function in angiogenesis	References
HOX transcript antisense RNA (HOTAIR)	Diabetic retinopathy	Histone methylation, histone acetylation, DNA methylation	HOTAIR regulates glucose-mediated increases of angiogenesis in diabetic retinopathy	Biswas et al., 2021
Small nucleolar RNA host gene 14 (SNHG14)	Hepatocellular carcinoma	SNHG14 upregulates PABPC1 expression via H3K27 acetylation	SNHG14 promotes proliferation and tube formation in endothelial cells	Zhang H. et al., 2020
LINC00337	Colorectal cancer	LINC00337 recruits DNMT1 to CNN1 promoter, which inhibits its transcription and increases VEGF-mediated angiogenesis	LINC00337 increases tumor growth and microvascular density	Xu X. et al., 2020
RAB11B Antisense RNA 1 (RAB11B-AS1)	Breast cancer, osteosarcoma	RAB11B-AS1 increases RNA Pol II in hypoxia to upregulates VEGFA and ANGPLT4	HIF2 induces RAB11B-AS1 which increases angiogenic factors	Niu et al., 2020
Metastasis associated lung adenocarcinoma transcript 1 (MALAT1)	Multiple cancers	Formation of molecular scaffolds, splicing and regulating histones and transcription factors	MALAT1 increased proliferation, sprouting and migration in ECs	Li Y. et al., 2017
Maternally expressed 3 (MEG3)	Idiopathic pulmonary fibrosis, cholestatic liver injury	MEG3 interacts with JARID2 which recruits PRC2	MEG3 regulates NOTCH and VEGF pathways	Ruan et al., 2018
Antisense non-coding RNA in the INK4 locus (ANRIL, CDKN2B, CDKN2B-AS1)	Coronary heart disease, ischemic stroke, type 2 diabetes, atherosclerosis	Promoter methylation, chromatin modifications, alternative splicing and post-transcriptional modifications	High glucose upregulates ANRIL in retinal ECs and is involved in VEGF regulation	Zhou et al., 2016

cyclins and cell cycle kinases. Specifically, it regulates S-phase cyclins, p21 and p27Kip1 in mouse (Michalik et al., 2014). Overall, MALAT1's abundance in the cell, varied half-life and structural stability conferred by its 3′ end triple-helix structure, contributes to its functional stability and diversity.

In gastric cancer, MALAT1 promotes vascular mimicry and angiogenesis to establish tumorigenicity and metastasis (Li Y. et al., 2017). When MALAT1 is repressed in HUVEC, ECs were no longer able to form vessels via the tube formation assay. Knockdown was also able to increase EC permeability. With hypoxia, MALAT1 is upregulated and enhances proliferation of ECs *in vitro*. In another study by Michalik et al., MALAT1 knockdown in ECs increased sprouting and migration, but decreased stalk cell proliferation via cell cycle inhibition. Examining a mouse knockout of Malat1, scientists found no affect in adults, but it reduced vascular proliferation and network formation in embryonic retina. In the hind limb ischemia model, Malat1 deficiency decreased neovascularization, capillary density and recovery of blood flow. In thyroid tumors, MALAT1 promotes Fibroblast Growth Factor 2 (FGF2) secretion from tumor-associated macrophages into the tumor microenvironment to mediate angiogenesis (Huang et al., 2017). Together, the role of MALAT1 in angiogenesis is conferred by its role in alternative splicing, molecular scaffold formation and binding to actively transcribed gene loci, in particular the cell cycle genes. Clearly, MALAT1's functions are diverse.

MEG3

Maternally expressed gene 3 (MEG3) is a nuclear and EC-enriched lncRNA that exhibits multiple mRNA transcript variants (Zhang et al., 2010). It is also an imprinted gene (Michalik et al., 2014). Imprinting is an epigenetic phenomenon in which monoallelic silencing of some genes occurs in a parent-of-origin specific manner (Autuoro et al., 2014). This process is thought to be regulated by lncRNAs, though the mechanisms have yet to be fully elucidated. MEG3 is encoded by the imprinted *DLK1-DIO3* locus, and it was found that it interacts with Jumonji And AT-Rich Interaction Domain Containing 2 (JARID2), an important component of PRC2 in pluripotent stem cells (Kaneko et al., 2014). This interaction is needed in order to recruit and assemble PRC2 at a subset of pluripotent stem cell genes. This suggests that the interplay of these RNA-based interactions may participate in the epigenetic regulation of genes involved in the process of transitioning stem cell pluripotency to differentiation. MEG3's binding sites also have GA rich regions critical to guiding MEG3 to chromatin through the formation of RNA-DNA triplex structures (Mondal et al., 2015).

MEG3 was shown to be among the top 10 most abundant lncRNAs in HUVEC, strongly suggesting a clear biological role in ECs (Michalik et al., 2014). It also inhibits VEGF and Notch pathways, which we know are important signaling pathways in angiogenesis (Gordon et al., 2010). Adding to this, MEG3 expression is also upregulated by hypoxia. Ruan et al. overexpressed constitutive HIF-1alpha and found increased activity in the MEG3 promoter. Next, they examined chronic treatment (24 h) of pro-angiogenic growth factors like VEGF, bFGF and Transforming Growth Factor β (TGFβ). There was no effect on MEG3 expression, but there were still notable angiogenic effects. MEG3 knockdown markedly decreased VEGFR2 mRNA and protein expression in HUVEC, which then inhibited cell migration. Moreover, it impaired the ability of ECs to form tube-like structures and significantly decreased sprouting from spheroids in both normoxic and hypoxic conditions (Ruan et al., 2018). It also was found that genes of the TGFβ signaling pathway are direct targets of MEG3 and that MEG3 binds to distal regulatory sites of these genes. Thus, even though there was not a direct effect with pro-angiogenic factors, the downstream factors of these angiogenic pathways are still indirectly regulated by MEG3.

MEG3 can be characterized as a tumor suppressor important in cell cycle regulation and apoptosis (Li et al., 2015). Long-range interaction between distal loops of MEG3 secondary structure forms a pseudoknot which allows MEG3 to upregulate p53 expression (Li et al., 2015; Uroda et al., 2019). MEG3 allelic loss of locus is associated with meningioma pathogenesis and progression (Zhang et al., 2010). Expression of MEG3 in human meningioma cell lines clearly shows marked suppression of tumor cell growth and activation of p53. MEG3 also regulates age-associated decline in endothelial function; MEG3 was significantly upregulated in senescent HUVEC (passages 16–18) compared to earlier HUVEC passages (3–4) (Boon et al., 2016; He et al., 2017; Wu et al., 2017). Scientists found that when Meg3 was repressed in HUVEC, age-mediated inhibition of sprouting was stopped, implying that Meg3 silencing could be a potential way to rescue age-associated impairments in angiogenic potential. In the brain, Meg3 null mice exhibit enhanced vascular density (Gordon et al., 2010). Examining this closer, Meg3 null mice, showed increased VEGFA, VEGFR1, DLL4, among other angiogenic genes. It was previously shown that p53 could bind to Sp1 sites in the VEGFA promoter to negatively regulate VEGFA transcription (Pal et al., 2001). Thus, loss of MEG3 may decrease p53 binding, thereby causing an increase in transcription of genes involved in VEGF signaling. Clearly, MEG3 has an important role as an angiogenic regulator.

ANRIL

ANRIL (also known as CDKN2B or CDKN2B-AS1) is located on chromosome 9p21. GWAS identified this disease-associated locus as a "protein gene desert" (Cheng et al., 2005; Iaconetti et al., 2013; Wahlestedt, 2013; Carninci et al., 2021). ANRIL is transcribed antisense to the INK4b-ARF-INK4a gene cluster (Derrien et al., 2012). Exons 13-19 of ANRIL overlapped with a high-risk haplotype associated with genetic predisposition to coronary artery disease (CAD) (Broadbent et al., 2008). A genetic association is also evident with ischemic stroke, aneurysms, and peripheral vascular diseases (Zeggini et al., 2007; Foroud et al., 2012; Kremer et al., 2015; Kong et al., 2016; Tan et al., 2019). ANRIL is especially enriched in vascular smooth muscle cells (VSMC) and mononuclear phagocytes within atherosclerotic plaques (The Encode Project Consortium, 2012; Zollbrecht et al., 2013; Bai et al., 2014; Nanda et al., 2016). ANRIL is a better genetic predictor of cardiovascular diseases than classical clinical measures such as blood pressure and dyslipidemia (Holdt et al., 2010).

ANRIL has at least 20 linear or circular isoforms associated with atherosclerosis (Burd et al., 2010; Hubberten et al., 2019).

Though the mechanism(s) by which minor frequency alleles of ANRIL still predispose to disease remain to be fully elucidated, it is argued that ANRIL regulates this genomic region *in cis* whereby the risk allele leads to an increase in linear ANRIL, but reduced levels of circular ANRIL (Holdt and Teupser, 2018). Linear ANRIL may function as a scaffold for epigenetic protein complexes that stimulate pro-atherosclerotic cellular functions. ANRIL is highly enriched in the nucleus, playing an active role in chromatin modification (Zhou et al., 2016). It is regulated by promoter methylation, transcription factors, alternative splicing and post-transcriptional modifications. ANRIL interacts with PRC1 & PRC2 to epigenetically repress neighboring genes such as CDKN2A and CDKN2B *in cis*. ANRIL and CDKN2A form a scaffold with H3K27me3 with polycomb Chromobox 7″ (CBX7); ANRIL with CDKN2B interact with PRC2 subunit SUZ12 (Yap et al., 2010). *Trans* activity of ANRIL through PRC1/2 represses distant genes that are dependent on the Alu elements found in ANRIL and in target gene promoters. ANRIL creates a scaffold for WD repeat-containing protein 5 (WDR5), a histone H3K4 presenter and histone deacetylase 3 (HDAC3) coordinating histone modification on target genes of vascular smooth muscle cell phenotypes (Zhang C. et al., 2020).

ANRIL has been shown to be upregulated in human retinal ECs stimulated by high glucose and diabetes. In diabetic retinopathy, ANRIL regulates VEGF through interactions with PRC2 components p300, miR200b, and enhancer of zeste homolog 2 (EZH2) (Thomas et al., 2017). Since VEGF is involved in stimulating vascular permeability, migration and proliferation of ECs, ANRIL upregulating VEGF contributes to promoting endothelial injury, which occurs via tumor necrosis factor-alpha (TNFα)- nuclear factor kappa light-chain-enhancer of activated B cells (NFkB)-ANRIL/YY-IL6 signaling pathways (Zhou et al., 2016). Similarly, in a rat model with diabetes and cerebral infarction, overexpression of ANRIL increased VEGF expression, resulting in increased angiogenesis via NFkB signaling (Zhang et al., 2017). ANRIL also regulates Akt phosphorylation in ECs and scientists recently showed that in mice, ANRIL improves cardiac function and post-ischemic angiogenesis following myocardial infarction by upregulating angiogenesis through Akt activation (Huang et al., 2020). Finally, Zeng et al. recently showed that ANRIL levels were elevated in the serum of thrombosis patients relative to healthy patients (Zeng et al., 2019). To assess the effect of ANRIL on angiogenesis, they took Sprague Dawley rats and injected si-ANRIL and examined lumen formation. They found that there were fewer lumens and smaller lumens in the rats with repressed ANRIL relative to the control group, confirming a role for ANRIL in angiogenesis. These disease associations position ANRIL as a key target for treatment of cardiovascular disease.

CHALLENGES ASSOCIATED WITH STUDYING lncRNAs

There are many challenges associated with studying lncRNAs. The first question when studying a lncRNA is verifying whether it is truly a bona fide lncRNA. RNA-seq data is mapped to the most recent build of the human genome. These reference databases are not comprehensive with respect to lncRNA annotation, thus limiting discovery. Moreover, lncRNAs are typically rare transcripts. Many are only expressed in specific contexts (e.g., development, disease, specific environmental stimuli, etc.), which can make discovery difficult. This review does not address the emerging concepts on the role of lncRNAs as post-transcriptional modifiers of gene expression and function. We and others have recently addressed the cytoplasmic function of lncRNAs (Rashid et al., 2016; Noh et al., 2018; Aillaud and Schulte, 2020; Ho et al., 2021). We also acknowledge that a key concept, and one that warrants deeper study, is the shuttling of lncRNAs in and out of the nucleus. Moreover, the varied RNA transcripts derived from specific lncRNA genes may have distinct subcellular locations. It is paramount that the candidate lncRNA structure and diversity be assessed before proceeding with detailed mechanistic and functional studies.

The paradigm that a gene must have either protein-coding or a non-coding function, but not both, is too simplistic. Some RNAs have both coding and non-coding functions (Robb et al., 2004; Fish et al., 2007). The sONE RNA, is a lncRNA antisense to eNOS that exhibits exon/exon sense/antisense interactions. The sONE locus also has a minor mRNA variant that encodes a protein involved in the autophagy pathway. Human ECs have high levels of eNOS mRNA, but low levels of sONE RNA. When sONE RNA is overexpressed, there is decreased eNOS mRNA and protein expression. Adding to this, sONE RNA is upregulated by hypoxia in ECs and VSMCs. Notably, sONE is primarily localized to the nucleus in normoxia, but with hypoxia, the sONE RNA is shuttled into the cytoplasm.

LncRNAs exhibit protean intra-species allelic diversity. Furthermore, it follows and has been noted that lncRNAs have low inter-species sequence conservation, likely due to rapid evolutionary turnover (Hezroni et al., 2015; Quinn et al., 2016). For lncRNA homologs, generally the length of an alignable sequence is about 5 times shorter than that of a protein-coding gene. A normal lncRNA that is conserved between humans and mice will have about 20% interspecies homology, which decreases to about 5% in fish. As a result, lncRNAs may be absent in model organisms, making it hard for scientists to not only discover lncRNAs, but also to assess their *in vivo* function. There is some sequence conservation in lncRNAs, typically in short sequence islands, and perhaps this is because these are regions that are required for specific interactions with other RNAs, proteins or DNA (Kapusta and Feschotte, 2014; Quinn et al., 2016; Ulitsky, 2016). However, there are other factors to consider in the discussion of conservation besides sequence similarity alone. In fact, scientists have identified many orthologous RNAs with highly divergent sequences, that they would no longer be identifiable as orthologs by sequence similarity alone, but their function is preserved (Ponjavic et al., 2007; Ulitsky et al., 2011; Ulitsky, 2016). Another factor is positional conservation in which lncRNAs can be detected from syntenic loci even in the absence of most, if not all sequence similarity. It is clear that we need to define lncRNA conservation traits/signals. Another debated factor is structural conservation. Scientists have shown that there is limited association between secondary structure and sequence conservation (Managadze et al., 2011; Yang and Zhang, 2015).

Moreover, there is evidence that specific lncRNAs act through specific tertiary or quaternary structural features, such as the triplex elements at the 3′-termini of MALAT1 or Nuclear Enriched Abundant Transcript 1 (NEAT1) (Wilusz et al., 2012). Further study on lncRNA structural conservation is evidently needed to improve our understanding on inter- and intra-species lncRNA conservation.

NEW FRONTIERS: LncRNAs AS DIAGNOSTIC AND THERAPEUTIC TARGETS IN MEDICINE

As we look toward the next decade of lncRNA research, it will be interesting to see more clinical studies evaluating whether lncRNAs have the potential to be used as biomarkers or therapeutic targets for clinical interventions to improve disease outcomes. LncRNAs are lowly expressed, so we know quantification in biological fluids will be challenging. Moreover, they are poorly conserved across species making them difficult to study using *in vivo* models of disease. However, lncRNAs can be highly tissue-specific, which sets up these molecules to be very specific biomarkers. To date, prostate cancer antigen 3 (PCA3) is the only lncRNA approved as a clinical diagnostic biomarker for early detection of prostate cancer (Groskopf et al., 2006).

As for targeting lncRNAs, many scientists agree that the key will be through identifying the optimal delivery system. There has been growing interest in recent years in extracellular vesicles. Though we did not discuss it in this review, many lncRNAs, especially those expressed in cancer, have also been shown to be secreted by extracellular vesicles (Wu et al., 2017; Tellez-Gabriel and Heymann, 2019; Zhao et al., 2019; Fan et al., 2020). Extracellular vesicles are of interest because they are encapsulated by a lipid bilayer, which overcomes concerns with stability. In addition, extracellular vesicles have less immunogenicity and higher *in vivo* stability compared to widely used viral and non-viral vectors (Chen et al., 2021). Exosomes, a subtype of extracellular vesicles, are currently being examined. They have poor efficiency with respect to packaging large nucleic acids, but this is overcome through integration with liposomes or nanoparticles, which improves both specificity and control of delivery. Recent work has found that exosome-liposome hybrids were able to successfully deliver CRISPR-Cas9 systems *in vitro* and *in vivo* (Lin et al., 2018; Tao et al., 2018). This is particularly exciting for the future of precision medicine. Importantly, exosomal studies are not without challenge. There is a high degree of heterogeneity in vesicles, variability between *in vitro* and *in vivo* findings, and difficulty in determining vesicle origin or destination. Advances in isolating and characterizing extracellular vesicle-associated lncRNAs will significantly help move the field forward and has the potential to revolutionize clinical medicine.

Notably, even once we are able to identify "druggable" lncRNAs, it is still unclear what the downstream or off-target effects would be and if they would be adverse. Until clinical trials are conducted, the safety and efficacy of lncRNAs as therapeutic targets remains unknown. Evidently, the emerging study of lncRNAs has many challenges, but recent work underscores the importance of the contribution of lncRNAs to the regulation of angiogenesis in health and disease.

CONCLUSION

Our review is an overview of angiogenic long non-coding RNAs, and their epigenetic regulation of the vascular endothelium. The functional properties of the vascular endothelium are diverse and heterogeneous between vascular beds. Our understanding of angiogenesis to date has largely focused on protein signaling, but recent work by scientists has revealed that long non-coding RNAs, which are a functionally diverse class of molecules, are involved in regulating this process. As the function of lncRNAs is often dependent on their subcellular localization, nuclear lncRNAs act as epigenetic modifiers. Scientists have begun to identify and characterize a sub-class of lncRNAs: angiogenic lncRNAs. This includes: STEEL, GATA6-AS, and MANTIS, and the disease-associated angiogenic lncRNAs: MEG3, MALAT1, or ANRIL. Taken together, these emerging concepts may provide a novel avenue for therapeutic targets or biomarkers for disease in the next decade.

AUTHOR CONTRIBUTIONS

NS, RN, and PAM drafted and edited the manuscript. All authors read and approved the manuscript.

FUNDING

NS and RN were recipients of the Queen Elizabeth II Graduate Scholarships in Science and Technology. This work was supported by a grant to PAM from the Heart and Stroke Foundation of Canada (G-19-0026562).

REFERENCES

Ackermann, M., Houdek, J. P., Gibney, B. C., Ysasi, A., Wagner, W., Belle, J., et al. (2014). Sprouting and intussusceptive angiogenesis in postpneumonectomy lung growth: mechanisms of alveolar neovascularization. *Angiogenesis* 17, 541–551. doi: 10.1007/s10456-013-9399-9

Adair, T. H., and Montani, J.-P. (2010). Angiogenesis. *Colloquium Ser. Integr. Syst. Physiol. Mol. Funct.* 2, 1–84. doi: 10.4199/C00017ED1V01Y201009ISP010

Adams, R. H., and Alitalo, K. (2007). Molecular regulation of angiogenesis and lymphangiogenesis. *Nat. Rev. Mol. Cell Biol.* 8, 464–478. doi: 10.1038/nrm2183

Aillaud, M., and Schulte, L. N. (2020). Emerging roles of long noncoding RNAs in the Cytoplasmic Milieu. *ncRNA* 6:44. doi: 10.3390/ncrna6040044

Anderson, K. P., Kern, C. B., Crable, S. C., and Lingrel, J. B. (1995). Isolation of a gene encoding a functional zinc finger protein homologous to erythroid Krüppel-like factor: identification of a new multigene family. *Mol. Cell. Biol.* 15, 5957–5965. doi: 10.1128/MCB.15.11.5957

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 32, D115–D119. doi: 10.1093/nar/gkh131

Atkins, G. B., Wang, Y., Mahabeleshwar, G. H., Shi, H., Gao, H., Kawanami, D., et al. (2008). Hemizygous deficiency of Krüppel-like factor 2 augments experimental atherosclerosis. *Circ. Res.* 103, 690–693. doi: 10.1161/CIRCRESAHA.108.184663

Autuoro, J., Pirnie, S., and Carmichael, G. (2014). Long noncoding RNAs in imprinting and X Chromosome inactivation. *Biomolecules* 4, 76–100. doi: 10.3390/biom4010076

Awwad, D. A. (2019). Beyond classic editing: innovative CRISPR approaches for functional studies of long non-coding RNA. *Biol. Methods Protoc.* 4:bz017. doi: 10.1093/biomethods/bpz017

Bai, Y., Nie, S., Jiang, G., Zhou, Y., Zhou, M., Zhao, Y., et al. (2014). Regulation of CARD8 expression by ANRIL and Association of CARD8 single nucleotide polymorphism rs2043211 (p.C10X) With Ischemic stroke. *Stroke* 45, 383–388. doi: 10.1161/STROKEAHA.113.003393

Battistella, M., and Marsden, P. (2015). Advances, nuances, and potential pitfalls when exploiting the therapeutic potential of RNA interference. *Clin. Pharmacol. Ther.* 97, 79–87. doi: 10.1002/cpt.8

Bell, J. C., Jukam, D., Teran, N. A., Risca, V. I., Smith, O. K., Johnson, W. L., et al. (2018). Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife* 7:e27024. doi: 10.7554/eLife.27024

Bentley, K., Mariggi, G., Gerhardt, H., and Bates, P. A. (2009). Tipping the balance: robustness of tip cell selection, migration and fusion in angiogenesis. *PLoS Comput. Biol.* 5:e1000549. doi: 10.1371/journal.pcbi.1000549

Bignon, M., Pichol-Thievend, C., Hardouin, J., Malbouyres, M., Bréchot, N., Nasciutti, L., et al. (2011). Lysyl oxidase-like protein-2 regulates sprouting angiogenesis and type IV collagen assembly in the endothelial basement membrane. *Blood* 118, 3979–3989. doi: 10.1182/blood-2010-10-313296

Biswas, S., Feng, B., Chen, S., Liu, J., Aref-Eshghi, E., Gonder, J., et al. (2021). The long Non-Coding RNA HOTAIR is a critical epigenetic mediator of angiogenesis in diabetic retinopathy. *Invest. Ophthalmol. Vis. Sci.* 62:20. doi: 10.1167/iovs.62.3.20

Bonetti, A., Agostini, F., Suzuki, A. M., Hashimoto, K., Pascarella, G., Gimenez, J., et al. (2020). RADICL-seq identifies general and cell type–specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.* 11:1018. doi: 10.1038/s41467-020-14337-6

Boon, R. A., Hofmann, P., Michalik, K. M., Lozano-Vidal, N., Berghäuser, D., Fischer, A., et al. (2016). Long Noncoding RNA Meg3 controls endothelial cell aging and function. *J. Am. Col. Cardiol.* 68, 2589–2591. doi: 10.1016/j.jacc.2016.09.949

Broadbent, H. M., Peden, J. F., Lorkowski, S., Goel, A., Ongen, H., Green, F., et al. (2008). Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum. Mol. Genet.* 17, 806–814. doi: 10.1093/hmg/ddm352

Burd, C. E., Jeck, W. R., Liu, Y., Sanoff, H. K., Wang, Z., and Sharpless, N. E. (2010). Expression of linear and novel circular forms of an INK4/ARF-associated Non-Coding RNA correlates with atherosclerosis risk. *PLoS Genet* 6:e1001233. doi: 10.1371/journal.pgen.1001233

Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Biaesch, A., Padovan-Merhar, O., Regev, A., et al. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* 16:20. doi: 10.1186/s13059-015-0586-4

Cai, P., Otten, A. B. C., Cheng, B., Ishii, M. A., Zhang, W., Huang, B., et al. (2020). A genome-wide long noncoding RNA CRISPRi screen identifies PRANCR as a novel regulator of epidermal homeostasis. *Genome Res.* 30, 22–34. doi: 10.1101/gr.251561.119

Cai, X., and Cullen, B. R. (2007). The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* 13, 313–316. doi: 10.1261/rna.351707

Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., et al. (1996). Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* 380, 435–439. doi: 10.1038/380435a0

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2021). The transcriptional landscape of the mammalian genome. *Science* 309:6.

Chappell, J. C., Wiley, D. M., and Bautch, V. L. (2011). Regulation of blood vessel sprouting. *Seminars Cell Dev. Biol.* 22, 1005–1011. doi: 10.1016/j.semcdb.2011.10.006

Chaudhuri, A. R., and Nussenzweig, A. (2017). The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat. Rev. Mol. Cell. Biol.* 18, 610–621. doi: 10.1038/nrm.2017.53

Chen, L., Endler, A., and Shibasaki, F. (2009). Hypoxia and angiogenesis: regulation of hypoxia-inducible factors via novel binding factors. *Exp. Mol. Med.* 41:849. doi: 10.3858/emm.2009.41.12.103

Chen, Y., Li, Z., Chen, X., and Zhang, S. (2021). Long non-coding RNAs: From disease code to drug role. *Acta Pharmaceutica Sinica B* 11, 340–354. doi: 10.1016/j.apsb.2020.10.001

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-Nucleotide resolution. *Science* 308, 1149–1154. doi: 10.1126/science.1108625

Chery, J. (2016). RNA therapeutics: RNAi and antisense mechanisms and clinical applications. *Postdoc. J.* 4, 35–50.

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., and Chang, H. Y. (2011). Genomic maps of long Noncoding RNA occupancy reveal principles of RNA-Chromatin interactions. *Mol. Cell* 44, 667–678. doi: 10.1016/j.molcel.2011.08.027

Cines, D. B., Pollak, E. S., Buck, C. A., Loscalzo, J., Zimmerman, G. A., McEver, R. P., et al. (1998). Endothelial cells in physiology and in the pathophysiology of vascular disorders. *Blood* 91, 3527–3561.

Collins, M. L., Irvine, B., Tyner, D., Fine, E., Zayati, C., Chang, C., et al. (1997). A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res.* 25, 2979–2984. doi: 10.1093/nar/25.15.2979

Congrains, A., Kamide, K., Oguro, R., Yasuda, O., Miyata, K., Yamamoto, E., et al. (2012). Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. *Atherosclerosis* 220, 449–455. doi: 10.1016/j.atherosclerosis.2011.11.017

Crooke, S. T. (2017). Molecular mechanisms of antisense oligonucleotides. *Nucleic Acid Ther.* 27, 70–77. doi: 10.1089/nat.2016.0656

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002a). Capturing chromosome conformation. *Science* 295, 1306–1311. doi: 10.1126/science.1067799

Dekker, R. J., van Soest, S., Fontijn, R. D., Salamanca, S., de Groot, P. G., VanBavel, E., et al. (2002b). Prolonged fluid shear stress induces a distinct set of endothelial cell genes, most specifically lung Krüppel-like factor (KLF2). *Blood* 100, 1689–1698. doi: 10.1182/blood-2002-01-0046

Dekker, R. J., van Thienen, J. V., Rohlena, J., de Jager, S. C., Elderkamp, Y. W., Seppen, J., et al. (2005). Endothelial KLF2 links local arterial shear stress levels to the expression of vascular tone-regulating genes. *Am. J. Pathol.* 167, 609–618. doi: 10.1016/S0002-9440(10)63002-7

del Toro, R., Prahst, C., Mathivet, T., Siegfried, G., Kaminker, J. S., Larrivee, B., et al. (2010). Identification and functional analysis of endothelial tip cell–enriched genes. *Blood* 116, 4025–4033. doi: 10.1182/blood-2010-02-270819

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

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. doi: 10.1038/nature11233

Dorsett, Y., and Tuschl, T. (2004). siRNAs: applications in functional genomics and potential as therapeutics. *Nat. Rev. Drug Discov.* 3, 318–329. doi: 10.1038/nrd1345

Duffy, S. J., Tran, B. T., New, G., Tudball, R. N., Esler, M. D., Harper, R. W., et al. (1998). Continuous release of vasodilator prostanoids contributes to regulation of resting forearm blood flow in humans. *Am. J. Physiol. Heart Circ. Physiol.* 274, H1174–H1183. doi: 10.1152/ajpheart.1998.274.4.H1174

Edgren, H., Murumagi, A., Kangaspeska, S., Nicorici, D., Hongisto, V., Kleivi, K., et al. (2011). Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol.* 12:R6. doi: 10.1186/gb-2011-12-1-r6

Eilken, H. M., and Adams, R. H. (2010). Dynamics of endothelial cell behavior in sprouting angiogenesis. *Curr. Opin. Cell Biol.* 22, 617–625. doi: 10.1016/j.ceb.2010.08.010

Eißmann, M., Gutschner, T., Hämmerle, M., Günther, S., Caudron-Herger, M., Groß, M., et al. (2012). Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol.* 9, 1076–1087. doi: 10.4161/rna.21089

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X Chromosome. *Science* 341:1237973. doi: 10.1126/science.1237973

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

Fan, T., Sun, N., and He, J. (2020). Exosome-derived LncRNAs in lung cancer. *Front. Oncol.* 10:1728. doi: 10.3389/fonc.2020.01728

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

Fang, Y., Xu, Y., Wang, R., Hu, L., Guo, D., Xue, F., et al. (2020). Recent advances on the roles of LncRNAs in cardiovascular disease. *J. Cell. Mol. Med.* 24, 12246–12257. doi: 10.1111/jcmm.15880

Fazal, F. M., Han, S., Parker, K. R., Kaewsapsak, P., Xu, J., Boettiger, A. N., et al. (2019). Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell* 178, 473–490. doi: 10.1016/j.cell.2019.05.027

Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K. S., et al. (1996). Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* 380, 439–442. doi: 10.1038/380439a0

Fish, J. E., Matouk, C. C., Yeboah, E., Bevan, S. C., Khan, M., Patil, K., et al. (2007). Hypoxia-inducible expression of a natural cis-antisense transcript inhibits endothelial Nitric-oxide Synthase. *J. Biol. Chem.* 282, 15652–15666. doi: 10.1074/jbc.M608318200

Foroud, T., Koller, D. L., Lai, D., Sauerbeck, L., Anderson, C., Ko, N., et al. (2012). Genome-wide association study of intracranial aneurysms confirms role of Anril and SOX17 in Disease Risk. *Stroke* 43, 2846–2852. doi: 10.1161/STROKEAHA.112.656397

Fu, G. K., Xu, W., Wilhelmy, J., Mindrinos, M. N., Davis, R. W., Xiao, W., et al. (2014). Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. *Proc. Natl. Acad. Sci. U S A* 111, 1891–1896. doi: 10.1073/pnas.1323732111

Gall, J. G., and Pardue, M. L. (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparatiONS. *PNAS* 63, 378–383. doi: 10.1073/pnas.63.2.378

Gan, L., Liao, S., Xing, Y., and Deng, S. (2021). The regulatory functions of lncRNAs on Angiogenesis following ischemic stroke. *Front. Mol. Neurosci.* 13:613976. doi: 10.3389/fnmol.2020.613976

Gavard, J., and Gutkind, J. S. (2008). VE-cadherin and claudin-5: it takes two to tango. *Nat. Cell Biol.* 10, 883–885. doi: 10.1038/ncb0808-883

Gong, N., Teng, X., Li, J., and Liang, X.-J. (2019). Antisense oligonucleotide-conjugated nanostructure-targeting lncRNA MALAT1 inhibits cancer metastasis. *ACS Appl. Mater. Interfaces* 11, 37–42. doi: 10.1021/acsami.8b18288

Gordon, F. E., Nutt, C. L., Cheunsuchon, P., Nakayama, Y., Provencher, K. A., Rice, K. A., et al. (2010). Increased expression of angiogenic genes in the brains of mouse Meg3-Null Embryos. *Endocrinology* 151, 2443–2452. doi: 10.1210/en.2009-1151

Goyal, A., Myacheva, K., Groß, M., Klingenberg, M., Duran Arqué, 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

Groskopf, J., Aubin, S. M. J., Deras, I. L., Blase, A., Bodrug, S., Clark, C., et al. (2006). APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. *Clin. Chem.* 52, 1089–1095. doi: 10.1373/clinchem.2005.063289

Gutschner, T., Hämmerle, M., and Diederichs, S. (2013). MALAT1 — a paradigm for long noncoding RNA function in cancer. *J. Mol. Med.* 91, 791–801. doi: 10.1007/s00109-013-1028-y

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

Hacisuleyman, E., Goff, L. A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., et al. (2014). Topological organization of Multi-chromosomal regions by firre. *Nat. Struct. Mol. Biol.* 21, 198–206. doi: 10.1038/nsmb.2764

Hadfield, J., and Retief, J. (2018). A profusion of confusion in NGS methods naming. *Nat. Methods* 15, 7–8. doi: 10.1038/nmeth.4558

Han, J., Zhang, Z., and Wang, K. (2018). 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. *Mol. Cytogenet.* 11:21. doi: 10.1186/s13039-018-0368-2

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

He, C., Yang, W., Yang, J., Ding, J., Li, S., Wu, H., et al. (2017). Long Noncoding RNA MEG3 negatively regulates proliferation and angiogenesis in vascular endothelial cells. *DNA Cell Biol.* 36, 475–481. doi: 10.1089/dna.2017.3682

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., and Ulitsky, I. (2015). Principles of long Noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* 11, 1110–1122. doi: 10.1016/j.celrep.2015.04.023

Ho, J. J. D., Man, J. H. S., Schatz, J. H., and Marsden, P. A. (2021). Translational remodeling by RNA -binding proteins and noncoding RNAs. *Wiley interdiscip. Rev. RNA* 2021:e1647. doi: 10.1002/wrna.1647

Holdt, l. M., beutner, f., scholz, m., gielen, s., gäbel, g., bergert, h., et al. (2010). ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler. Thromb. Vasc. Bio.* 30, 620–627. doi: 10.1161/ATVBAHA.109.196832

Holdt, L. M., and Teupser, D. (2018). Long Noncoding RNA ANRIL: Lnc-ing genetic variation at the chromosome 9p21 locus to molecular mechanisms of atherosclerosis. *Front. Cardiovasc. Med.* 5:145. doi: 10.3389/fcvm.2018.00145

Hou, Y., Zhang, R., and Sun, X. (2019). Enhancer LncRNAs influence chromatin interactions in different ways. *Front Genet.* 10:936. doi: 10.3389/fgene.2019.00936

Hu, X., Tang, J., Hu, X., Bao, P., Deng, W., Wu, J., et al. (2019). Silencing of long Non-coding RNA HOTTIP reduces inflammation in rheumatoid arthritis by demethylation of SFRP1. *Mol. Ther. Nucleic Acids* 19, 468–481. doi: 10.1016/j.omtn.2019.11.015

Huang, J., Ma, L., Song, W., Lu, B., Huang, Y., Dong, H., et al. (2017). LncRNA-MALAT1 promotes angiogenesis of thyroid cancer by modulating tumor-associated macrophage FGF2 protein secretion: the mechanisms of malat1 in TAMs. *J. Cell. Biochem.* 118, 4821–4830. doi: 10.1002/jcb.26153

Huang, Q., Pan, M., Zhou, J., and Yin, F. (2020). Overexpression of long non-coding RNA ANRIL promotes post-ischaemic angiogenesis and improves cardiac functions by targeting Akt. *J. Cell. Mol. Med.* 24, 6860–6868. doi: 10.1111/jcmm.15343

Hubberten, M., Bochenek, G., Chen, H., Häsler, R., Wiehe, R., Rosenstiel, P., et al. (2019). Linear isoforms of the long noncoding RNA CDKN2B-AS1 regulate the c-myc-enhancer binding factor RBMS1. *Eur. J. Hum. Genet.* 27, 80–89. doi: 10.1038/s41431-018-0210-7

Iaconetti, C., Gareri, C., Polimeni, A., and Indolfi, C. (2013). Non-coding RNAs: The "dark Matter" of cardiovascular pathophysiology. *Int. J. Mol. Sci.* 14, 19987–20018. doi: 10.3390/ijms141019987

Ismail, N., Abdullah, N., Abdul Murad, N. A., Jamal, R., and Sulaiman, S. A. (2021). Long Non-coding RNAs (lncRNAs) in cardiovascular disease complication of Type 2 diabetes. *Diagnostics* 11:145. doi: 10.3390/diagnostics11010145

Iyer, S., Modali, S. D., and Agarwal, S. K. (2017). Long Noncoding RNA MEG3 Is an epigenetic determinant of oncogenic signaling in functional pancreatic neuroendocrine tumor cells. *Mol. Cell. Biol.* 37:e00278–17. doi: 10.1128/MCB.00278-17

Jain, R. K. (2005). Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* 307, 58–62.

Jin, K.-T., Yao, J.-Y., Fang, X.-L., Di, H., and Ma, Y.-Y. (2020). Roles of lncRNAs in cancer: focusing on angiogenesis. *Life Sci.* 252:117647. doi: 10.1016/j.lfs.2020.117647

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Kaewsapsak, P., Shechner, D. M., Mallard, W., Rinn, J. L., and Ting, A. Y. (2017). Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife* 6:e29224. doi: 10.7554/eLife.29224

Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., et al. (2014). Interactions between JARID2 and Noncoding RNAs Regulate PRC2

recruitment to chromatin. *Mol. Cell* 53, 290–300. doi: 10.1016/j.molcel.2013.11.012

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

Kato, Y., Kravchenko, V. V., Tapping, R. I., Han, J., Ulevitch, R. J., and Lee, J.-D. (1997). BMK1/ERK5 regulates serum-induced early gene expression through transcription factor MEF2C. *EMBO J.* 16, 7054–7066. doi: 10.1093/emboj/16.23.7054

Katsushima, K., Jallo, G., Eberhart, C. G., and Perera, R. J. (2021). Long non-coding RNAs in brain tumors. *NAR Cancer* 3:41. doi: 10.1093/narcan/zcaa041

Kaunas, R., Kang, H., and Bayless, K. J. (2011). Synergistic regulation of angiogenic sprouting by biochemical factors and wall shear stress. *Cell. Mol. Bioeng.* 4, 547–559. doi: 10.1007/s12195-011-0208-5

Kino, T., Hurt, D., Ichijo, T., Nader, N., and Chrousos, G. (2010). Noncoding RNA Gas5 Is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* 3:ra8. doi: 10.1126/scisignal.2000568

Koch, L. (2017). Screening for lncRNA function. *Nat. Rev. Genet.* 18:70. doi: 10.1038/nrg.2016.168

Koller, A., Sun, D., and Kaley, G. (1993). Role of shear stress and endothelial prostaglandins in flow- and viscosity-induced dilation of arterioles in vitro. *Circ. Res.* 72, 1276–1284. doi: 10.1161/01.RES.72.6.1276

Konerding, M. A., Gibney, B. C., Houdek, J. P., Chamoto, K., Ackermann, M., Lee, G. S., et al. (2012). Spatial dependence of alveolar angiogenesis in post-pneumonectomy lung growth. *Angiogenesis* 15, 23–32. doi: 10.1007/s10456-011-9236-y

Kong, Y., Sharma, R. B., Nwosu, B. U., and Alonso, L. C. (2016). Islet biology, the CDKN2A/B locus and type 2 diabetes risk. *Diabetologia* 59, 1579–1593. doi: 10.1007/s00125-016-3967-7

Kremer, P. H. C., Koeleman, B. P. C., Pawlikowska, L., Weinsheimer, S., Bendjilali, N., Sidney, S., et al. (2015). Evaluation of genetic risk loci for intracranial aneurysms in sporadic arteriovenous malformations of the brain. *J. Neurol. Neurosurg. Psychiatry* 86, 524–529. doi: 10.1136/jnnp-2013-307276

Ku, K. H., Dubinsky Michelle, K., Sukumar Aravin, N., Subramaniam, N., Feasson, M. Y. M., Nair, R., et al. (2021). In vivo function of flow-responsive Cis-DNA elements of the endothelial nitric oxide synthase gene: A role for chromatin-based mechanisms. *Circulation* 2021:1078. doi: 10.1161/CIRCULATIONAHA.120.051078

Ku, K. H., Subramaniam, N., and Marsden, P. A. (2019). Epigenetic determinants of flow-mediated vascular endothelial gene expression. *Hypertension* 74, 467–476. doi: 10.1161/HYPERTENSIONAHA.119.13342

Le, N.-T., Heo, K.-S., Takei, Y., Lee, H., Woo, C.-H., Chang, E., et al. (2013). A crucial role for p90RSK-mediated reduction of ERK5 transcriptional activity in endothelial dysfunction and atherosclerosis. *Circulation* 127, 486–499. doi: 10.1161/CIRCULATIONAHA.112.116988

Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Ferrante, T. C., Terry, R., et al. (2015). Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* 10, 442–458. doi: 10.1038/nprot.2014.191

Leisegang, M. S., Bibli, S.-I., Günther, S., Pflüger-Müller, B., Oo, J. A., Höper, C., et al. (2019). Pleiotropic effects of laminar flow and statins depend on the Krüppel-like factor-induced lncRNA MANTIS. *Eur. Heart J.* 40, 2523–2533. doi: 10.1093/eurheartj/ehz393

Leisegang, M. S., Fork, C., Josipovic, I., Richter, F. M., Preussner, J., Hu, J., et al. (2017). Long Noncoding RNA MANTIS facilitates endothelial angiogenic function. *Circulation* 136, 65–79. doi: 10.1161/CIRCULATIONAHA.116.026991

Lerner, M. R., and Steitz, J. A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *PNAS* 76, 5495–5499. doi: 10.1073/pnas.76.11.5495

Li, J., Hou, R., Niu, X., Liu, R., Wang, Q., Wang, C., et al. (2016). Comparison of microarray and RNA-Seq analysis of mRNA expression in dermal mesenchymal stem cells. *Biotechnol. Lett.* 38, 33–41. doi: 10.1007/s10529-015-1963-5

Li, K., Blum, Y., Verma, A., Liu, Z., Pramanik, K., Leigh, N. R., et al. (2010). A noncoding antisense RNA in tie-1 locus regulates tie-1 function in vivo. *Blood* 115, 133–139. doi: 10.1182/blood-2009-09-242180

Li, X., and Fu, X.-D. (2019). Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat. Rev. Genet.* 20, 503–519. doi: 10.1038/s41576-019-0135-1

Li, X., Zhou, B., Chen, L., Gou, L.-T., Li, H., and Fu, X.-D. (2017). GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* 35, 940–950. doi: 10.1038/nbt.3968

Li, X.-L., Zhou, J., Chen, Z.-R., and Chng, W.-J. (2015). p53 mutations in colorectal cancer- molecular pathogenesis and pharmacological reactivation. *World J. Gastroenterol.* 21, 84–93. doi: 10.3748/wjg.v21.i1.84

Li, Y., Wu, Z., Yuan, J., Sun, L., Lin, L., Huang, N., et al. (2017). Long non-coding RNA MALAT1 promotes gastric cancer tumorigenicity and metastasis by regulating vasculogenic mimicry and angiogenesis. *Cancer Lett.* 395, 31–44. doi: 10.1016/j.canlet.2017.02.035

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Lin, M. F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* 27, i275–i282. doi: 10.1093/bioinformatics/btr209

Lin, Y., Wu, J., Gu, W., Huang, Y., Tong, Z., Huang, L., et al. (2018). Exosome–liposome hybrid nanoparticles deliver CRISPR/Cas9 system in MSCs. *Adv. Sci.* 5:1700611. doi: 10.1002/advs.201700611

Lipowsky, H. H. (1995). "Shear Stress in the Circulation," in *Flow-Dependent Regulation of Vascular Function*, eds J. A. Bevan, G. Kaley, and G. M. Rubanyi (New York: Springer), 28–45.

Liu, S. J., Horlbeck, M. A., Cho, S. W., Birk, H. S., Malatesta, M., He, D., et al. (2017). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 355:7111. doi: 10.1126/science.aah7111

Liu, S. J., Malatesta, M., Lien, B. V., Saha, P., Thombare, S. S., Hong, S. J., et al. (2020). CRISPRi-based radiation modifier screen identifies long non-coding RNA therapeutic targets in glioma. *Genome Biol.* 21:83. doi: 10.1186/s13059-020-01995-4

Liu, Y., Morley, M., Brandimarto, J., Hannenhalli, S., Hu, Y., Ashley, E. A., et al. (2015). RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics* 105, 83–89. doi: 10.1016/j.ygeno.2014.12.002

Lu, Y. W., Martino, N., Gerlach, B. D., Lamar, J. M., Vincent, P. A., Adam, A. P., et al. (2021). MEF2 (myocyte enhancer factor 2) is essential for endothelial homeostasis and the atheroprotective gene expression program. *Arterioscler. Thromb. Vasc. Biol.* 41, 1105–1123. doi: 10.1161/ATVBAHA.120.314978

Ma, L., Cao, J., Liu, L., Du, Q., Li, Z., Zou, D., et al. (2019). LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res.* 47, D128–D134. doi: 10.1093/nar/gky960

Malek, A. M. (1999). Hemodynamic shear stress and its role in atherosclerosis. *JAMA* 282:2035. doi: 10.1001/jama.282.21.2035

Man, H. S. J., Sukumar, A. N., Lam, G. C., Turgeon, P. J., Yan, M. S., Ku, K. H., et al. (2018). Angiogenic patterning by STEEL, an endothelial-enriched long noncoding RNA. *Proc. Natl. Acad. Sci. USA* 115, 2401–2406. doi: 10.1073/pnas.1715182115

Man, H.-S. J., and Marsden, P. A. (2019). LncRNAs and epigenetic regulation of vascular endothelium: genome positioning system and regulators of chromatin modifiers. *Curr. Opin. Pharmacol.* 45, 72–80. doi: 10.1016/j.coph.2019.04.012

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

Marín-Béjar, O., and Huarte, M. (2015). RNA pulldown protocol for in vitro detection and identification of RNA-associated proteins. *Methods Mol. Biol.* 1206, 87–95. doi: 10.1007/978-1-4939-1369-5_8

Massie, C. E., and Mills, I. G. (2012). Mapping protein-DNA interactions using ChIP-sequencing. *Methods Mol. Biol.* 809, 157–173. doi: 10.1007/978-1-61779-376-9_11

Matouk, C. C., and Marsden, P. A. (2008). Epigenetic regulation of vascular endothelial gene expression. *Circ. Res.* 102, 873–887. doi: 10.1161/CIRCRESAHA.107.171025

Mattick, J. S., and Rinn, J. L. (2015). Discovery and annotation of long noncoding RNAs. *Nat. Struct. Mol. Biol.* 22, 5–7. doi: 10.1038/nsmb.2942

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science* 316, 1488–1491. doi: 10.1126/science.1142447

Meng, Q., Pu, L., Luo, X., Wang, B., Li, F., and Liu, B. (2020). Regulatory roles of related long Non-coding RNAs in the process of atherosclerosis. *Front. Physiol.* 11:564604. doi: 10.3389/fphys.2020.564604

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

Miao, Y., Ajami, N. E., Huang, T.-S., Lin, F.-M., Lou, C.-H., Wang, Y.-T., et al. (2018). Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. *Nat. Commun.* 9:292. doi: 10.1038/s41467-017-02113-y

Michalik, K. M., You, X., Manavski, Y., Doddaballapur, A., Zörnig, M., Braun, T., et al. (2014). Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ. Res.* 114, 1389–1397. doi: 10.1161/CIRCRESAHA.114.303265

Mishra, K., and Kanduri, C. (2019). Understanding long noncoding RNA and chromatin interactions: what we know so far. *Noncoding RNA* 5:54. doi: 10.3390/ncrna5040054

Mockler, T. C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S. E., and Ecker, J. R. (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 85, 1–15. doi: 10.1016/j.ygeno.2004.10.005

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., et al. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA–DNA triplex structures. *Nat. Commun.* 6:7743. doi: 10.1038/ncomms8743

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

Nakaoka, H., Gurumurthy, A., Hayano, T., Ahmadloo, S., Omer, W. H., Yoshihara, K., et al. (2016). Allelic imbalance in regulation of ANRIL through chromatin interaction at 9p21 endometriosis risk locus. *PLoS Genet.* 12:e1005893. doi: 10.1371/journal.pgen.1005893

Nanda, V., Downing, K. P., Ye, J., Xiao, S., Kojima, Y., Spin, J. M., et al. (2016). CDKN2B regulates TGFβ signaling and smooth muscle cell investment of hypoxic neovessels. *Circ. Res.* 118, 230–240. doi: 10.1161/CIRCRESAHA.115.307906

Neumann, P., Jaé, N., Knau, A., Glaser, S. F., Fouani, Y., Rossbach, O., et al. (2018). The lncRNA GATA6-AS epigenetically regulates endothelial gene expression via interaction with LOXL2. *Nat. Commun.* 9:237. doi: 10.1038/s41467-017-02431-1

Niu, Y., Bao, L., Chen, Y., Wang, C., Luo, M., Zhang, B., et al. (2020). HIF2-induced long noncoding RNA RAB11B-AS1 promotes hypoxia-mediated angiogenesis and breast cancer metastasis. *Cancer Res.* 80, 964–975. doi: 10.1158/0008-5472.CAN-19-1532

Noh, J. H., Kim, K. M., McClusky, W. G., Abdelmohsen, K., and Gorospe, M. (2018). Cytoplasmic functions of long noncoding RNAs. *WIREs RNA* 9:e1471. doi: 10.1002/wrna.1471

Novikova, I. V., Hennelly, S. P., Tung, C.-S., and Sanbonmatsu, K. Y. (2013). Rise of the RNA machines: exploring the structure of long Non-coding RNAs. *J. Mol. Biol.* 425, 3731–3746. doi: 10.1016/j.jmb.2013.02.030

Ohnesorge, N., Viemann, D., Schmidt, N., Czymai, T., Spiering, D., Schmolke, M., et al. (2010). Erk5 activation elicits a vasoprotective endothelial phenotype via induction of Krüppel-like factor 4 (KLF4)*. *J. Biol. Chem.* 285, 26199–26210. doi: 10.1074/jbc.M110.103127

Pal, S., Datta, K., and Mukhopadhyay, D. (2001). Central role of p53 on regulation of vascular permeability factor/ vascular endothelial growth factor (VPF/VEGF) expression in mammary carcinoma. *Cancer Res.* 61, 6952–6957.

Parmar, K. M., Larman, B., Dai, G., Zhang, Y., Wang, E. T., Moorthy, S. N., et al. (2005). Integration of flow-dependent endothelial phenotypes by Kruppel-like factor 2. *J. Clin. Invest.* 116, 49–58. doi: 10.1172/JCI24787

Peng, L., Liu, F., Yang, J., Liu, X., Meng, Y., Deng, X., et al. (2020). Probing lncRNA–protein interactions: data repositories, models, and algorithms. *Front. Genet.* 10:1346. doi: 10.3389/fgene.2019.01346

Pinkney, H. R., Wright, B. M., and Diermeier, S. D. (2020). The lncRNA toolkit: databases and in silico tools for lncRNA analysis. *ncRNA* 6:49. doi: 10.3390/ncrna6040049

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

Potente, M., Gerhardt, H., and Carmeliet, P. (2011). Basic and therapeutic aspects of angiogenesis. *Cell* 146, 873–887. doi: 10.1016/j.cell.2011.08.039

Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., et al. (2014). RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* 42, D756–D763. doi: 10.1093/nar/gkt1114

Quek, X. C., Thomson, D. W., Maag, J. L. V., Bartonicek, N., Signal, B., Clark, M. B., et al. (2015). lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* 43, D168–D173. doi: 10.1093/nar/gku988

Quinn, E. M., Cormican, P., Kenny, E. M., Hill, M., Anney, R., Gill, M., et al. (2013). Development of strategies for SNP detection in RNA-Seq Data: Application to lymphoblastoid cell lines and evaluation using 1000 genomes data. *PLoS One* 8:e58815. doi: 10.1371/journal.pone.0058815

Quinn, J. J., Ilik, I. A., Qu, K., Georgiev, P., Chu, C., Akhtar, A., et al. (2014). Domain ChIRP reveals the modularity of long noncoding RNA architecture, chromatin interactions, and function. *Nat. Biotechnol.* 32, 933–940. doi: 10.1038/nbt.2943

Quinn, J. J., Zhang, Q. C., Georgiev, P., Ilik, I. A., Akhtar, A., and Chang, H. Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes Dev.* 30, 191–207. doi: 10.1101/gad.272187.115

Quinodoz, S., and Guttman, M. (2014). Long non-coding RNAs: An emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* 24, 651–663. doi: 10.1016/j.tcb.2014.08.009

Rackham, O., Shearwood, A.-M. J., Mercer, T. R., Davies, S. M. K., Mattick, J. S., and Filipovska, A. (2011). Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. *RNA* 17, 2085–2093. doi: 10.1261/rna.029405.111

Rashid, F., Shah, A., and Shan, G. (2016). Long Non-coding RNAs in the cytoplasm. *Genom. Proteom. Bioinform.* 14, 73–80. doi: 10.1016/j.gpb.2016.03.005

Reis, M., Czupalla, C. J., Ziegler, N., Devraj, K., Zinke, J., Seidel, S., et al. (2012). Endothelial Wnt/β-catenin signaling inhibits glioma angiogenesis and normalizes tumor blood vessels by inducing PDGF-B expression. *J. Exper. Med.* 209, 1611–1627. doi: 10.1084/jem.20111580

Resnick, N., Collins, T., Atkinson, W., Bonthront, D. T., Dewey, C. F., and Gimbrone, M. A. (1993). Platelet-derived growth factor B chain promoter contains a cis-acting fluid shear-stress-responsive element. *Med. Sci.* 90, 4591–4595.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX Loci by Non-coding RNAs. *Cell* 129, 1311–1323. doi: 10.1016/j.cell.2007.05.022

Robb, G. B., Carson, A. R., Tai, S. C., Fish, J. E., Singh, S., Yamada, T., et al. (2004). Post-transcriptional regulation of endothelial nitric-oxide synthase by an overlapping antisense mRNA Transcript. *J. Biol. Chem.* 279, 37982–37996. doi: 10.1074/jbc.M400271200

Rom, A., Melamed, L., Gil, N., Goldrich, M. J., Kadir, R., Golan, M., et al. (2019). Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. *Nat. Commun.* 10:5092. doi: 10.1038/s41467-019-13075-8

Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F., and Crespi, M. (2018). Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* 46, 2169–2184. doi: 10.1093/nar/gky095

Ruan, W., Zhao, F., Zhao, S., Zhang, L., Shi, L., and Pang, T. (2018). Knockdown of long noncoding RNA MEG3 impairs VEGF-stimulated endothelial sprouting angiogenesis via modulating VEGFR2 expression in human umbilical vein endothelial cells. *Gene* 649, 32–39. doi: 10.1016/j.gene.2018.01.072

Sangwung, P., Zhou, G., Nayak, L., Chan, E. R., Kumar, S., Kang, D.-W., et al. (2017). KLF2 and KLF4 control endothelial identity and vascular integrity. *JCI Insight* 2:91700. doi: 10.1172/jci.insight.91700

Schmidt, T., and Carmeliet, P. (2010). Bridges that guide and unite. *Nature* 465, 697–699. doi: 10.1038/465697a

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

Shweiki, D., Itin, A., Soffer, D., and Keshet, E. (1992). Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. *Nature* 359, 843–845. doi: 10.1038/359843a0

Simon, M. D., Wang, C. I., Kharchenko, P. V., West, J. A., Chapman, B. A., Alekseyenko, A. A., et al. (2011). The genomic binding sites of a noncoding RNA. *PNAS* 108, 20497–20502. doi: 10.1073/pnas.1113536108

Sonnhammer, E. L., Eddy, S. R., Birney, E., Bateman, A., and Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res.* 26, 320–322. doi: 10.1093/nar/26.1.320

Sridhar, B., Rivas-Astroza, M., Nguyen, T. C., Chen, W., Yan, Z., Cao, X., et al. (2017). Systematic mapping of RNA-chromatin interactions in vivo. *Curr. Biol.* 27, 602–609. doi: 10.1016/j.cub.2017.01.011

Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956–960. doi: 10.1126/science.1160342

Taheri, M., Eghtedarian, R., Dinger, M. E., and Ghafouri-Fard, S. (2020). Emerging roles of non-coding RNAs in the pathogenesis of type 1 diabetes mellitus. *Biomed. Pharmacother.* 129:110509. doi: 10.1016/j.biopha.2020.110509

Tan, C., Liu, J., Wei, J., and Yang, S. (2019). Effects of ANRIL variants on the risk of ischemic stroke: a meta-analysis. *Biosci. Rep.* 39:BSR20182127. doi: 10.1042/BSR20182127

Tao, S.-C., Guo, S.-C., and Zhang, C.-Q. (2018). Modularized extracellular vesicles: the dawn of prospective personalized and precision medicine. *Adv. Sci.* 5:1700449. doi: 10.1002/advs.201700449

Taxman, D. J., Livingstone, L. R., Zhang, J., Conti, B. J., Iocca, H. A., Williams, K. L., et al. (2006). Criteria for effective design, construction, and gene knockdown by shRNA vectors. *BMC Biotechnol.* 6:7. doi: 10.1186/1472-6750-6-7

Tellez-Gabriel, M., and Heymann, D. (2019). Exosomal lncRNAs: the newest promising liquid biopsy. *Cancer Drug Resist.* 2, 1002–1017. doi: 10.20517/cdr.2019.69

Teppan, J., Barth, D. A., Prinz, F., Jonas, K., Pichler, M., and Klec, C. (2020). Involvement of long Non-coding RNAs (lncRNAs) in tumor angiogenesis. *Noncoding RNA* 6:42. doi: 10.3390/ncrna6040042

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

Thomas, A. A., Feng, B., and Chakrabarti, S. (2017). ANRIL: A Regulator of VEGF in Diabetic Retinopathy. *Invest. Ophthalmol. Vis. Sci.* 58:470. doi: 10.1167/iovs.16-20569

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938. doi: 10.1016/j.molcel.2010.08.011

Uchida, S. (2017). High-throughput methods to detect long Non-coding RNAs. *High Throughput.* 6:12. doi: 10.3390/ht6030012

Ule, J., Hwang, H.-W., and Darnell, R. B. (2018). The future of cross-linking and immunoprecipitation (CLIP). *Cold Spring Harb. Perspect. Biol.* 10:a032243. doi: 10.1101/cshperspect.a032243

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R. B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212–1215. doi: 10.1126/science.1090095

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., and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46. doi: 10.1016/j.cell.2013.06.020

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., and Bartel, D. P. (2011). Conserved Function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147, 1537–1550. doi: 10.1016/j.cell.2011.11.055

Uroda, T., Anastasakou, E., Rossi, A., Teulon, J.-M., Pellequer, J.-L., Annibale, P., et al. (2019). Conserved Pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol. Cell* 75, 982–995. doi: 10.1016/j.molcel.2019.07.025

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

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

Ventola, G. M. M., Noviello, T. M. R., D'Aniello, S., Spagnuolo, A., Ceccarelli, M., and Cerulo, L. (2017). Identification of long non-coding transcripts with feature selection: a comparative study. *BMC Bioinform.* 18:187. doi: 10.1186/s12859-017-1594-z

Wahlestedt, C. (2013). Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nat. Rev. Drug Discov.* 12, 433–446. doi: 10.1038/nrd4018

Wang, C., Gong, B., Bushel, P. R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., et al. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* 32, 926–932. doi: 10.1038/nbt.3001

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation. *Nature* 472, 120–124. doi: 10.1038/nature09819

Wang, W., Ha, C. H., Jhun, B. S., Wong, C., Jain, M. K., and Jin, Z.-G. (2010). Fluid shear stress stimulates phosphorylation-dependent nuclear export of HDAC5 and mediates expression of KLF2 and eNOS. *Blood* 115, 2971–2979. doi: 10.1182/blood-2009-05-224824

Watts, J. K., and Corey, D. R. (2012). Gene silencing by siRNAs and antisense oligonucleotides in the laboratory and the clinic. *J. Pathol.* 226, 365–379. doi: 10.1002/path.2993

Webster, A. L. H., Yan, M. S.-C., and Marsden, P. A. (2013). Epigenetics and cardiovascular disease. *Can. J. Cardiol.* 29:12.

Wen, S., Wei, Y., Zen, C., Xiong, W., Niu, Y., and Zhao, Y. (2020). Long non-coding RNA NEAT1 promotes bone metastasis of prostate cancer through N6-methyladenosine. *Mol. Cancer* 19:171. doi: 10.1186/s12943-020-01293-4

West, J. A., Davis, C. P., Sunwoo, H., Simon, M. D., Sadreyev, R. I., Wang, P. I., et al. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* 55, 791–802. doi: 10.1016/j.molcel.2014.07.012

Wilusz, J. E., JnBaptiste, C. K., Lu, L. Y., Kuhn, C.-D., Joshua-Tor, L., and Sharp, P. A. (2012). A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* 26, 2392–2407. doi: 10.1101/gad.204438.112

Wolska, M., Jarosz-Popek, J., Junger, E., Wicik, Z., Porshoor, T., Sharif, L., et al. (2021). Long Non-coding RNAs as promising therapeutic approach in ischemic stroke: a comprehensive review. *Mol. Neurobiol.* 58, 1664–1682. doi: 10.1007/s12035-020-02206-8

Wu, Z., He, Y., Li, D., Fang, X., Shang, T., Zhang, H., et al. (2017). Long noncoding RNA MEG3 suppressed endothelial cell proliferation and migration through regulating miR-2. *Am. J. Transl. Res.* 9, 3326–3335.

Xie, Y., Liu, Y.-G., and Chen, L. (2016). Assessing protein-DNA interactions: Pros and cons of classic and emerging techniques. *Sci. China Life Sci.* 59, 425–427. doi: 10.1007/s11427-016-5046-1

Xu, E., Hu, X., Li, X., Jin, G., Zhuang, L., Wang, Q., et al. (2020). Analysis of long non-coding RNA expression profiles in high-glucose treated vascular endothelial cells. *BMC Endocr. Disord.* 20:107. doi: 10.1186/s12902-020-00593-6

Xu, X., Nie, J., Lu, L., Du, C., Meng, F., and Song, D. (2020). LINC00337 promotes tumor angiogenesis in colorectal cancer by recruiting DNMT1, which suppresses the expression of CNN1. *Cancer Gene Ther.* 2020, 1–13. doi: 10.1038/s41417-020-00277-2

Yan, C., Chen, J., and Chen, N. (2016). Long noncoding RNA MALAT1 promotes hepatic steatosis and insulin resistance by increasing nuclear SREBP-1c protein stability. *Sci. Rep.* 6:22640. doi: 10.1038/srep22640

Yan, M. S.-C., Matouk, C. C., and Marsden, P. A. (2010). Epigenetics of the vascular endothelium. *J. Appl. Physiol.* 109, 916–926. doi: 10.1152/japplphysiol.00131.2010

Yang, J.-R., and Zhang, J. (2015). Human long noncoding RNAs are substantially less folded than messenger RNAs. *Mol. Biol. Evol.* 32, 970–977. doi: 10.1093/molbev/msu402

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

Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., et al. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 Lysine 27 by Polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* 38, 662–674. doi: 10.1016/j.molcel.2010.03.021

Yeh, C.-F., Chang, Y.-C. E., Lu, C.-Y., Hsuan, C.-F., Chang, W.-T., and Yang, K.-C. (2020). Expedition to the missing link: Long noncoding RNAs in cardiovascular diseases. *J. Biomed. Sci.* 27:48. doi: 10.1186/s12929-020-00647-w

Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., et al. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316, 1336–1341. doi: 10.1126/science.1142364

Zeng, R., Song, X.-J., Liu, C.-W., and Ye, W. (2019). LncRNA ANRIL promotes angiogenesis and thrombosis by modulating microRNA-99a and microRNA-449a in the autophagy pathway. *Am. J. Transl. Res.* 11, 7441–7448.

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, B., Wang, D., Ji, T.-F., Shi, L., and Yu, J.-L. (2017). cerebral infarction by activating NF-κB signaling pathway in a. *Oncotarget* 8, 17347–17359.

Zhang, C., Ge, S., Gong, W., Xu, J., Guo, Z., Liu, Z., et al. (2020). LncRNA ANRIL acts as a modular scaffold of WDR5 and HDAC3 complexes and promotes alteration of the vascular smooth muscle cell phenotype. *Cell Death Dis.* 11:435. doi: 10.1038/s41419-020-2645-3

Zhang, H., Xu, H.-B., Kurban, E., and Luo, H.-W. (2020). LncRNA SNHG14 promotes hepatocellular carcinoma progression via H3K27 acetylation activated PABPC1 by PTEN signaling. *Cell Death Dis.* 11, 1–13. doi: 10.1038/s41419-020-02808-z

Zhang, X., Gejman, R., Mahta, A., Zhong, Y., Rice, K. A., Zhou, Y., et al. (2010). Maternally expressed gene 3, an imprinted Noncoding RNA gene, is associated with meningioma pathogenesis and progression. *Cancer Res.* 70, 2350–2358. doi: 10.1158/0008-5472.CAN-09-3885

Zhao, W., Liu, Y., Zhang, C., and Duan, C. (2019). Multiple roles of exosomal long noncoding RNAs in cancers. *Biomed. Res. Int.* 2019:1460572. doi: 10.1155/2019/1460572

Zhou, G., Hamik, A., Nayak, L., Tian, H., Shi, H., Lu, Y., et al. (2012). Endothelial Kruppel-like factor 4 protects against atherothrombosis in mice. *J. Clin. Invest.* 122, 4727–4731. doi: 10.1172/JCI66056

Zhou, L., Zhu, Y., Sun, D., and Zhang, Q. (2020). Emerging roles of long non-coding RNAs in the tumor microenvironment. *Int. J. Biol. Sci.* 16, 2094–2103. doi: 10.7150/ijbs.44420

Zhou, X., Han, X., Wittfeldt, A., Sun, J., Liu, C., Wang, X., et al. (2016). Long non-coding RNA ANRIL regulates inflammatory responses as a novel component of NF-κB pathway. *RNA Biol.* 13, 98–108. doi: 10.1080/15476286.2015.1122164

Zollbrecht, C., Grassl, M., Fenk, S., Höcherl, R., Hubauer, U., Reinhard, W., et al. (2013). Expression pattern in human macrophages dependent on 9p21.3 coronary artery disease risk locus. *Atherosclerosis* 227, 244–249. doi: 10.1016/j.atherosclerosis.2012.12.030

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

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

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

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

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