


PRIMARY RESEARCH

Open Access



Identification of long non-coding RNA using single nucleotide epimutation analysis: a novel gene discovery approach

Mohammad Amin Kerachian^{1,2,3,4*}  and Marjan Azghandi^{3,5}

Abstract

Background Long non-coding RNAs (lncRNAs) are involved in a variety of mechanisms related to tumorigenesis by functioning as oncogenes or tumor-suppressors or even harboring oncogenic and tumor-suppressing effects; representing a new class of cancer biomarkers and therapeutic targets. It is predicted that more than 35,000 ncRNA especially lncRNA are positioned at the intergenic regions of the human genome. Emerging research indicates that one of the key pathways controlling lncRNA expression and tissue specificity is epigenetic regulation.

Methods In the current article, a novel approach for lncRNA discovery based on the intergenic position of most lncRNAs and a single CpG site methylation level representing epigenetic characteristics has been suggested.

Results Using this method, a novel antisense lncRNA named LINC02892 presenting three transcripts without the capacity of coding a protein was found exhibiting nuclear, cytoplasmic, and exosome distributions.

Conclusion The current discovery strategy could be applied to identify novel non-coding RNAs influenced by methylation aberrations.

Keywords Long non-coding RNA, Epigenetic, LINC02892, Exosome, Intergenic, Discovery

Introduction

Long non-coding RNAs (lncRNAs) comprise different species of RNA which exceed 200 nucleotides that are not usually translated into proteins (limited protein-coding capacity) [1]. They modulate the gene expression at various levels, including transcriptional, post-transcriptional, and epigenetic processing [2, 3]. Additionally, growing evidence has revealed that lncRNAs could play an important role in various cancers by regulating oncogenes or tumor-suppressors, or even harboring oncogenic and tumor-suppressing effects, representing a new class of cancer biomarkers and therapeutic targets [4–8]. Dysregulation of lncRNAs normally affects cellular functions such as apoptosis resistance, cell proliferation, tumor suppressor evasion, metastasis promotion, and angiogenesis activation in tumorigenesis [9–11], reported

*Correspondence:

Mohammad Amin Kerachian

amin.kerachian@mail.mcgill.ca; amin.kerachian@ryerson.ca

¹Medical Genetics Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

²Department of Medical Genetics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

³Cancer Genetics Research Unit, Reza Radiotherapy and Oncology Center, Mashhad, Iran

⁴Department of Chemistry and Biology, Toronto Metropolitan University, Toronto, ON, Canada

⁵Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

in breast cancer [12], glioblastoma [13], liver cancer [14], leukemia [15], colorectal cancer (CRC) [6] and several other cancers [16]. Their expression and function can be influenced by mutation [17] or epigenetic changes, including DNA methylation [8]. Epigenetic modifications have key roles in cancer biology and cell growth [18–20]. Recent studies of DNA methylation analysis in tumor cells have identified several thousand differential methylated regions (DMRs) [21] with less than 3% mapped to promoters. The majority of DMRs are found in introns or intergenic regions [22]. It is widely known that tumor cells display global demethylation of intergenic regions expressing large hypomethylation across different types of tumors [21, 23–25]. Of note, one potential function of intergenic DMRs is to regulate the non-coding RNA (ncRNA) expression [22]. It is predicted that more than 35,000 ncRNA especially lncRNA are positioned at the intergenic regions [26]. Emerging research indicates that one of the key pathways controlling lncRNA expression and tissue specificity is epigenetic regulation [27, 28]. Similar to germline genetic mutations, constitutive aberrant methylation may serve as the first hit (according to Knudson's model of tumor development) in patients with cancer [29] especially at the intergenic regions. Changes in methylation could be due to single CpG methylation errors at different positions [30].

We have previously suggested an algorithm to identify methylated CpG sites (accessible in GitHub through the following link: https://github.com/Genetics-Research-Laboratory-RROC/Candidate_Primer_Region_Finder) using methylation-sensitive high resolution melting (MS-HRM), on data from methylation next-generation sequencing (mNGS). It is feasible that methylation aberrations in crucial single CpG sites could impact the function of the lncRNA similar to single nucleotide polymorphisms (SNPs) of lncRNAs, leading to different impacts on its expression and function [31–33]. Therefore, in this article based on the intergenic position of lncRNAs and single CpG site methylation, an approach for novel lncRNA discovery linked to tumorigenesis is suggested. The newly discovered lncRNA would be attributed to the analyzed cancer type. Furthermore, we used bioinformatics tools and laboratory experiments to identify and validate the novel lncRNAs.

Materials and methods

Identification and validation of single CpG epimutation

Single CpG epimutations were identified by mNGS [34] and verified by MS-HRM assay. Briefly, a CpG site discovery step was performed based on unbiased methylome sequencing using SureSelectXT Methyl-Seq in CRC and control groups (six individuals each) using an algorithm to identify methylated CpG sites accessible in GitHub through the following link: <https://github.com/>

[Genetics-Research-Laboratory-RROC/Candidate_Primer_Region_Finder](https://github.com/Genetics-Research-Laboratory-RROC/Candidate_Primer_Region_Finder). Then, specific primers for bisulfite-converted sequences were designed (MethPrime 2.0 software package) and synthesized (Metabion, Germany). Prior to use, MS-HRM assays were evaluated on methylated and unmethylated bisulfite converted control DNA and the optimal annealing temperatures were determined empirically.

For biological validation of the identified CpG sites, genomic DNA were isolated from formalin-fixed paraffin-embedded (FFPE) (40 cancerous and 40 normal colon tissues) and fresh (28 cancerous and 28 normal colon tissues) samples using QIAamp DNA FFPE Tissue Kit and QIAamp Fast DNA Tissue kit, respectively (Qiagen, Germany). All patients gave written informed permission to retain and analyze their samples for purposes of this study. The procedures and protocols in the present study were approved by the regional ethics committee. Subsequently, DNA was bisulfite-converted using EpiTect Fast Bisulfite Conversion Kit (Qiagen, Germany) according to the manufacturer's instructions and amplified using the LightCycler 96 (Roche, Mannheim, Germany).

Identification of novel long non-coding RNA

RNA-Seq data analysis

RNA-Seq dataset for normal and colon cancer was obtained from the NCBI Sequence Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>), using the accession number SRR2089755 [35]. The raw reads were processed by removing the low-quality sequences (<10% 'N' bases and >85% QA>20 bases) and ribosomal sequences with Tophat [36]. All subsequent analyses were performed using clean reads. Clean reads were aligned to the GRCh38 reference genome using Tophat [36], during which only 2 mismatches and 2 gaps were allowed for each reading. The mapped reads were then assembled using Cufflinks [37] to identify the known and novel transcripts.

In-silico discovery of novel lncRNA

We screened for potential lncRNAs on genome confined to the discovery CpG sites, based on the following filter criteria: (1) length >200 nucleotides (nt); (2) open reading frame (ORF) length <400 nt; (3) no match to PFAM protein families database [38] (E value >1e-5); (4) iSeeRNA [39] non-coding score >0.5; and (5) the Coding Potential Assessment Tool (CPAT) [40] coding probability >0.375; (6) removal of the transcripts mapped within the 1 kb flanking regions of an annotated gene. Gene expression level was measured by the number of uniquely mapped reads per kilobase of exon region in a gene per million mappable reads (RPKM) [41].

For annotation of the novel lncRNA, the ncRNA sequence database (RNACentral) [42] was used to align the lncRNA to screen for any sequence homology.

In-silico evaluating the coding potentiality of lncRNA

Among the tools for evaluating coding potential, CPAT [40], CPC (Cording-Potential Calculator) [43], and RNA-code [44] were used for the evaluation of the coding potentiality of the novel lncRNAs.

In-silico subcellular localization

Subcellular localization of lncRNAs was predicted using iLoc-LncRNA [45] and lncLocator [46].

Experimentally validation of the novel lncRNA

Tissue expression of novel lncRNA

For experimental validation of the RNA-Seq results, a total RNA from 40 to 40 FFPE cases (cancerous) and control (normal) tissues, CRC cell lines (Caco-2, HCT 116, HT-29, SW480, and SW48) purchased from Pasteur Institute of Iran, were isolated using RNeasy FFPE kit (Qiagen, Germany) and Acuzol (Bioneer, South Korea), respectively. cDNA was synthesized using the RocketScript RT premix (Bioneer, Korea). The gene-specific primer targeting the novel lncRNA and *GAPDH* (as a reference gene) were designed (by primer premier 6.0 software) and synthesized (Eurofins, Germany). Reverse Transcription Quantitative PCR (RT-qPCR) reaction was carried out using HOT FIREPol qPCR mix with EvaGreen (Solis BioDyne- Estonia) on the LightCycler 96 (Roche, Mannheim, Germany) and all experiments were conducted in duplicate for each sample and performed according to the digital MIQE guidelines [47].

Sequencing of the novel lncRNA

The full-length lncRNA was obtained using the 5'- and 3'-RACE System for Rapid Amplification of cDNA Ends (RACE) standard method [48]. PCR products were separated on a 3% agarose gel. Gel products were extracted with a Gel Extraction kit (Bioneer, South Korea), cloned into pTZ57R/T vector, and sequenced by directionally using M13 forward and reverse primers.

Protein coding potentiality

The novel lncRNA named "Long intergenic non-protein coding RNA 2892 (LINC02892)" cDNA was synthesized from HT29 cells by RT-PCR. For the test of the protein-coding potentiality of LINC02892, the enhanced green fluorescent protein (EGFP) coding sequence was inserted into the 3' end of the putative LINC02892 open reading frame (ORF), and the fusion gene LINC02892-EGFP was cloned into the restriction sites; Nhe I and Xho I of plasmid pcDNA3.1 (Invitrogen, California, USA). Then, plasmid transfections were performed using Lipofectamine 2000 (Invitrogen, California, USA) and GFP expression was measured by fluorescence microscopy images.

Cellular fractionation and organelle isolation

A total of 1×10^6 cells were washed twice in cold phosphate buffered saline (PBS) and then incubated in hypotonic buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5, 10 mM KCl, 350 mM sucrose, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), and 0.1% Triton X-100) on ice for 10 min. After 5 min of centrifugation at 2,000 g, the supernatant was collected as the cytoplasmic fraction, and after additional washing, the remainder was considered as nuclear pellets, which was resuspended in lysis buffer (10 mM HEPES, pH 7.0, 100 mM KCl, 5 mM MgCl₂, 0.5% NP-40, 10 μM DTT and 1mM phenylmethanesulfonyl fluoride (PMSF)) to prepare the nuclear lysate. Cytoplasmic fraction was then centrifuged in an ultracentrifuge at 100,000 g at 4 °C for 40 min to pellet the exosomes. The supernatant was carefully removed, and the crude exosome-containing pellets were resuspended in 1 mL of ice-cold PBS. The second round of ultracentrifugation (100,000 g at 4 °C for 40 min) was carried out, and the resulting exosome pellet was resuspended in 500 μL of PBS. In addition, transmission electron microscope (TEM) study was performed according to standard techniques [49] to corroborate the presence of exosomes.

Results

The current study was inspired and extended by our previous work, in which SureSelectXT assay and methylation array observations revealed two-track methylation shifts for 'potentially functioning' sites like CpG islands (CGIs), CPG shores, promoters, 5'- from other 'relatively non-functioning intergenic sites [34]. As results, the algorithm found 194 regions and the two best locations with the highest differential methylation rates between case and control groups were subjected for lncRNA discovery.

In this study, we discovered a novel lncRNA termed "LINC02892". In order to characterize and verify the newly discovered lncRNAs, we used bioinformatics instruments and laboratory experiments to offer a path to discover lncRNA based on a single epimutation. Our path would be different with the general RNA-Seq searching publishes every day for lncRNA discovery (Fig. 1, Roadmap to detect lncRNA).

Validation of single CpG epimutation

In our previous study, single CpG epimutations were identified by mNGS assay [50]. In order to biologically validate the mNGS results, primer sets were used to target the different regions on the bisulfite-modified DNA. Methylation-sensitive high-resolution melting assay results were in accordance with the mNGS. The real-time PCR was conducted with the LightCycler® 96 and their results were demonstrated in Supplementary Fig. 1.

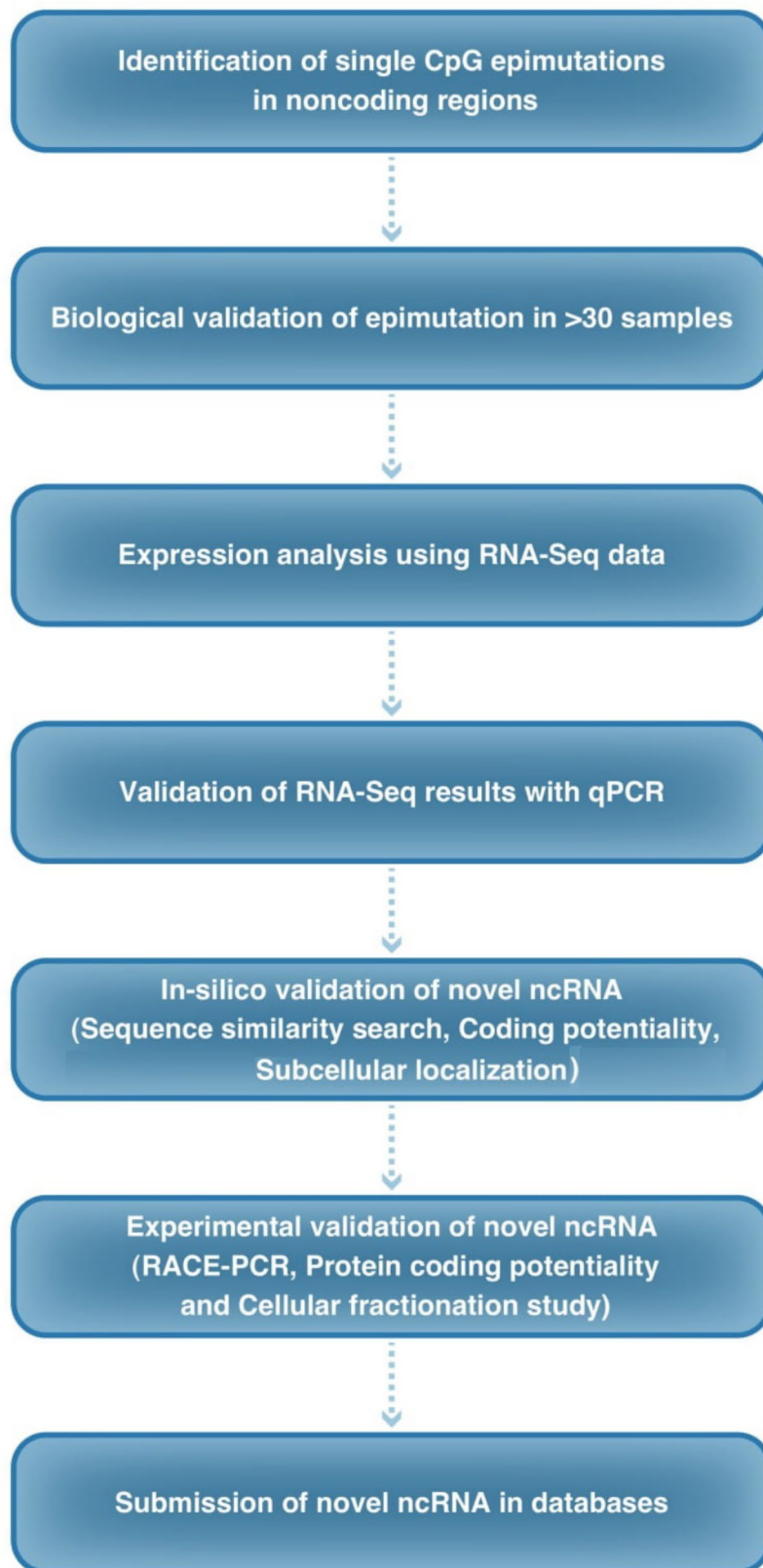


Fig. 1 Roadmap for discovering novel lncRNA based on single epimutation

RNA-Seq data analysis and annotation of novel lncRNA

Based on a single CpG epimutation position, high-throughput RNA sequence analysis was used to identify the novel lncRNAs on genome in colon tissues (cancerous and normal). The RNA-Seq dataset for normal and colon cancer was obtained from the NCBI Sequence Read Archive database. The RNA-Seq reads were successfully mapped onto one of the CpG epimutation positions and there was no expression statement for the second CpG site.

Our analysis with short-read mapping along with approximately 250 reads were successfully mapped onto a single CpG epimutation position on chromosome 21. The novel lncRNA, identified on chromosome 21 was further classified by comparison with the known gene annotations using RNAcentral sequence search tool. The similarity searches against a comprehensive set of ncRNAs showed that the LINC02892 sequence is similar to a long ncRNA in Pan troglodytes (Orangutan) with identity and query coverage of 70% and 79.9%, respectively (Fig. 2 A and 2B).

5'- and 3'-rapid amplification of cDNA ends (RACE) assay

Based on the sequence of LINC02892, the experiments of 5'- and 3'-RACE assay were initiated with total RNA from HT29 cells and resulted in three 888, 603, and 382-nucleotide (nt) antisense transcripts (Fig. 2 C), which the transcript #1 is the same as the transcript annotated with RNA-Seq data. In the current study, the three novel transcripts were identified with seven, five, and three exons, respectively (Fig. 2D). LINC02892 transcripts were submitted to NCBI under the accession numbers: Banklt2400105, LINC02892, MW248922; Banklt2400122, LINC02892, MW248923; Banklt2400131, LINC02892, MW248924; Banklt2400132, LINC02892, MW248925.

Subcellular localization

In-silico subcellular localization revealed cytoplasmic, dual nuclear/cytoplasmic, and exosomal distributions for transcript #1, #2, and #3, respectively (Fig. 3 A).

Moreover, to determine the cellular localization of the LINC02892 transcripts, the nuclear, cytoplasm, and exosome RNAs from the HT29 and SW48 cell lines were isolated, and the expression of lncRNA- LINC02892 transcripts in all subcellular locations were measured. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), small nuclear RNA U1 (*U1*), and *BCAR4* lncRNA were utilized as controls for cytoplasm, nucleus, and exosome, respectively. The RT-qPCR data of cellular fractionation assay in both cell lines demonstrated that the distribution of LINC02892 transcripts were clearly similar to that of the nuclear-localized *U1* snRNA, the exosomal retained *BCAR4* mRNA, and the protein-coding *GAPDH* mRNA (Fig. 3B).

To characterize the lncRNA that is enriched in the exosomes from the cell line, the extracted exosomes were examined and confirmed by TEM (data not shown).

Protein coding potentiality

The coding potential calculator tools predicted that LINC02892 displayed no protein-coding potentiality. A protein's potential score of transcripts was less than zero, which meant that the transcript has no capacity for coding a protein. Furthermore, the coding potential analysis revealed that LINC02892 sequence could not code any proteins. Although UniProt showed a putative peptide prediction of 28 amino acids for LINC02892 transcript #1, the putative ORF of LINC02892 transcript #1 was not expressed as an N-terminal enhanced green fluorescent protein fusion protein (Fig. 3 C and 3D).

LINC02892 is upregulated in colorectal cancer tissue and cell lines

RNA-Seq data analysis indicated that the LINC02892 expression level was significantly high in tumorous tissues compared with adjacent normal tissues. To further confirm this observation, we obtained 40 FFPE CRC tumors and their adjacent normal FFPE tissues from CRC patients. LINC02892 expression was examined by RT-qPCR and its upregulation was observed in tumoral samples. The RT-qPCR results demonstrated that in FFPE samples, CRC tissues indicated a significant 5.11-fold overexpression of the LINC02892 as compared to the corresponding normal tissues (p -value < 0.005) (Supplementary Fig. 2). Moreover, we profiled LINC02892 expression in CRC cell lines (Caco-2, HCT 116, HT-29, SW480, and SW48) and found that the recent lncRNA ubiquitously was overexpressed in all tested CRC cell lines with higher levels compared to the normal cell line. These findings confirmed the RNA-Seq results derived from the NCBI SRA database.

Discussion

Over the past decade, lncRNAs have been identified as significant players in gene regulation. They are often differentially expressed and widely associated with a majority of cancer types [51]. In a wide number of biological functions such as apoptosis, lncRNAs have been involved, and their roles are strongly associated with the cellular compartments where they are located [52]. Previous studies have shown that by acting as tumor suppressors or oncogenes, lncRNAs have significant roles in cancer [53]. Emerging research has indicated that DNA methylation is a significant epigenetic regulator of lncRNA expression, and the expression pattern of lncRNAs can be affected by epigenetic changes in DNA methylation which could lead to carcinogenesis [54–58].

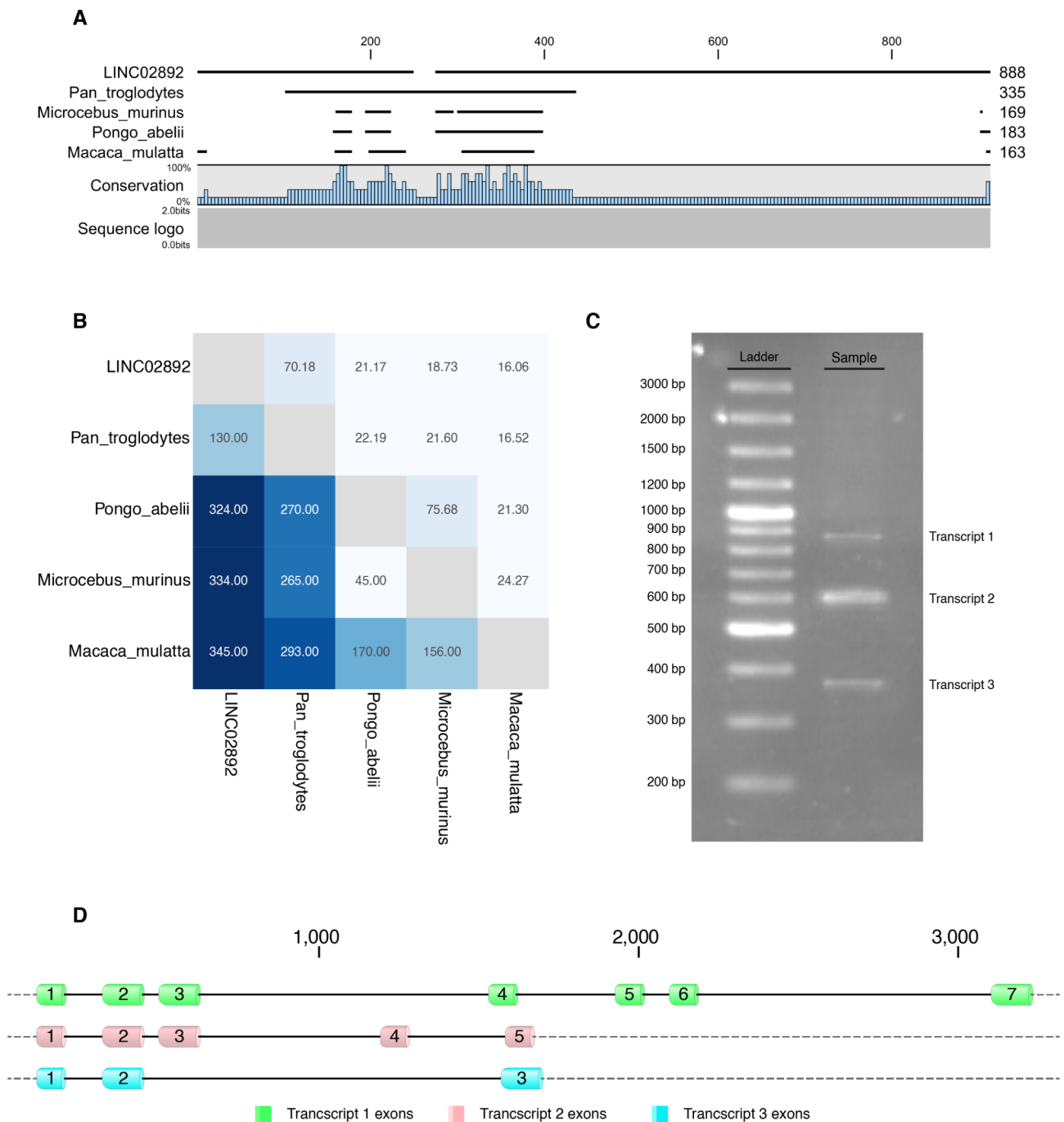


Fig. 2 (A) Alignment of the LINC02892 sequence from humans and other organisms. (B) Pairwise comparison among complete sequences of LINC02892. The upper comparison gradient indicated the percentage identity between two sequences, and the lower comparison gradient indicated the distance between two sequences. (C) The length of LINC02892 transcripts determined by RACE PCR assays. (D) Schematic intron-exon diagram of the LINC02892 transcripts. The exons and introns are marked as boxes and lines, respectively

The most abundant RNA modification in eukaryotic cells is N6-methyladenosine (m6A) [59]. RNA methylation usually occurs at the RRm6ACH consensus motif ([G/A/U][G/A]m6A[U/A/C]) [60, 61] and is abundant in 3' untranslated regions (3'UTRs), between stop codons and within long internal exons [62, 63]. In addition, in

precursor mRNAs (pre-RNAs) and lncRNAs, m6A modification occurs [64, 65]. Proteins that can add, remove, or recognize m6A-modified sites and change substantial biological processes are m6A “writers,” “erasers” and “readers,” respectively [61]. Moreover, DNA methylation depends upon DNA methyltransferases (DNMTs) [66].

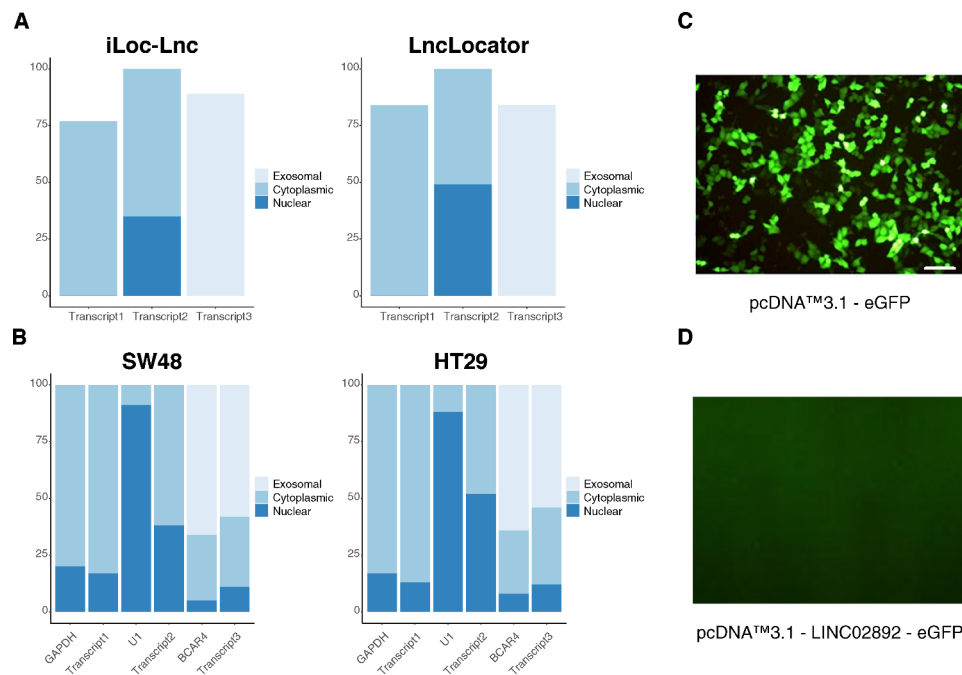


Fig. 3 (A) In silico subcellular localization of LINC02892 transcripts. (B) qRT-PCR assay following nuclear, cytoplasmic and exosome fractionation detecting the distribution of the indicated LINC02892 transcripts in HT29 and SW48 cell lines. The qRT-PCR data, represented as a percentage of the total amount of detected transcripts, are presented as means \pm SD from three independent experiments performed in triplicate. (C and D) Fluorescence microscopy of HT29 cells that had been transfected with the indicated plasmid (scale bars, 100 μ m)

For DNA methylated in CpG islands, there are proteins called “Methyl-CpG-binding domains (MBDs)” which are required for binding to methylated DNA [67]. MBD can also bind up with RNA and influence the methylation of DNA [68]. Hence, some RNAs could direct DNA methylation. MiRNA could also influence the methylation of mRNA [69] and thus, RNA directing RNA methylation also exists. However, DNA causing RNA methylation has not been explored yet.

In the current study, an integrated methylation and transcriptome analysis was conducted to identify the crosstalk between DNA methylation and lncRNA. We identified an intergenic lncRNA based on methylation characteristics. During the past decade, due to the development of relevant biotechnology and computational methods, a growing number of newly detected lncRNAs have been reported [70]. To discover lncRNAs, there are two common methods: (1) RNA sequencing (RNA-Seq) using next-generation sequencers and (2) microarrays [71]. Owing to the development of NGS technology, lncRNA identification is now more easily achievable and several assay-based sequencing protocols have been developed to predict lncRNAs [72]. However, the identification of lncRNA relying only on RNA-Seq or microarray has some limitations. Firstly, their data are predictive and secondly, since the expression of lncRNAs are mostly low, they could be lost during normalization and trimming of the data or become absent in RNA sequencing of

numerous samples. Furthermore, more complementary techniques are needed to identify the potential lncRNAs.

Since intergenic hypomethylation is crucial in tumorigenesis, aberration methylation of single nucleotide CpG sites could act as a landmark to discover long intergenic non-protein coding RNAs. It has been reported that lncRNAs are often located at crucial sites including regions of SNPs, amplifications, or common breakpoints [73], and intergenic regions [74]. Several studies have indicated that lncRNAs SNPs can prone the patients to CRC via deregulation of downstream pathways, proposing polymorphisms as CRC risk factors [8].

The DMR of DNA in intergenic regions could be related to the expression of intergenic ncRNAs [75]. Once the methylation statuses of single nucleotide CpG sites throughout the DNA genome are determined, they could be easily validated by MS-HRM. Then, the existence of a potential ncRNA could be investigated in RNA-Seq datasets as well as in-silico studies. Unlike other ncRNAs, lncRNAs are not quite conserved between species [76], causing annotation less informative in lncRNA discovery. To further confirm, gene expression should be conducted on cancer and normal tissues.

Conclusion

In summary, based on our discovery platform, we found a novel antisense lncRNA named “LINC02892”, which has three transcripts with no capacity of coding a protein that exhibits nuclear, cytoplasmic, or exosome distributions.

Our study characterized the crosstalk between DNA methylation and lncRNA, providing a novel pipeline to identify intergenic lncRNAs like LINC02892 which could be important in tumorigenesis of CRC. Further studies are necessary to validate the efficiency of this new method.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12935-022-02752-2>.

Supplementary Figure 1. Methyl Specific High Resolution Melting peaks of CpG epimutation in chromosome21 analysis, normal samples (blue) and CRC patients (red).

Supplementary Figure 2. Real time-PCR analysis of LINC02892 gene expression in patients with CRC and normal (control) FFPE tissues (p-value <0.005). The error bars represent standard deviation (SD).

Acknowledgements

Our special thanks go to Mr. Reza Khayami for designing the figures.

Authors' contributions

Conceived and designed the study: MAK. Performed the experimental procedures: MA. Analyzed the data and drafted the manuscript: MAK and MA. All authors read and approved the final manuscript.

Funding

This study was supported financially by Mashhad University of Medical Sciences (Grant number: 991511).

Data availability

The authors declare that the datasets on which the conclusions of this manuscript rely on are deposited in publicly available repositories.

Declarations

Ethics approval and consent to participate

The current study was approved by Mashhad University of Medical Sciences (MUMS) ethics committee.

Consent for publication

All authors are consent for the publication of this work.

Competing interests

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants, or patents received or pending, or royalties.

Received: 5 October 2021 / Accepted: 12 October 2022

Published online: 04 November 2022

References

- Sun Z, Yang S, Zhou Q, Wang G, Song J, Li Z, et al. Emerging role of exosome-derived long non-coding RNAs in tumor microenvironment. *Mol Cancer*. 2018;17(1):1–9.
- Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015;21(11):1253–61.
- Schmitt AM, Chang HY. Long RNAs wire up cancer growth. *Nature*. 2013;500(7464):536–7.
- Arun G, Diermeier SD, Spector DL. Therapeutic targeting of long non-coding RNAs in cancer. *Trends Mol Med*. 2018;24(3):257–77.
- Yang F, Zhang L, Huo Xs Y, Jh, Xu D, Yuan, Sx, et al. Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology*. 2011;54(5):1679–89.
- Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long non-coding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res*. 2011;71(20):6320–6.
- Hu Y, Wang J, Qian J, Kong X, Tang J, Wang Y, et al. Long noncoding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer. *Cancer Res*. 2014;74(23):6890–902.
- Poursheikhani A, Abbaszadegan MR, Kerachian MA. Mechanisms of long non-coding RNA function in colorectal cancer tumorigenesis. *Asia - Pacific J Clin Oncol*. 2020;17(1):7–23.
- Fang Y, Fullwood MJ. Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genom Proteom Bioinform*. 2016;14(1):42–54.
- Brunner AL, Beck AH, Edris B, Sweeney RT, Zhu SX, Li R, et al. Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol*. 2012;13(8):1–13.
- Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol*. 2012;9(6):703–19.
- Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464(7291):1071–6.
- Yao Y, Ma J, Xue Y, Wang P, Li Z, Liu J, et al. Knockdown of long non-coding RNA XIST exerts tumor-suppressive functions in human glioblastoma stem cells by up-regulating miR-152. *Cancer Lett*. 2015;359(1):75–86.
- Quagliata L, Matter M, Piscuoglio S, Makowska Z, Heim M, Tornillo L, et al. 90 HOXA13 AND Hottip expression levels predict patients' survival and metastasis formation in hepatocellular carcinoma. *Journal of Hepatology*. 2013;(58):S39–S40.
- Yildirim E, Kirby JE, Brown DE, Mercier FE, Sadreyev RI, Scadden DT, et al. Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell*. 2013;152(4):727–42.
- Jiang M-C, Ni J-J, Cui W-Y, Wang B-Y, Zhuo W. Emerging roles of lncRNA in cancer and therapeutic opportunities. *Am J cancer Res*. 2019;9(7):1354.
- Minotti L, Agnoletto C, Baldassari F, Corrà F, Volinia S. SNPs and somatic mutation on long non-coding RNA: new frontier in the cancer studies? *High-throughput*. 2018;7(4):34.
- Hsiao SJ, Nikiforov YE. Molecular approaches to thyroid cancer diagnosis. *Endocrine-related Cancer*. 2014;21(5):T301–T13.
- Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 2008;453(7197):948–51.
- Baylin SB, Jones PA. A decade of exploring the cancer epigenome—biological and translational implications. *Nat Rev Cancer*. 2011;11(10):726–34.
- Kerachian MA, Kerachian M. Long interspersed nucleotide element-1 (LINE-1) methylation in colorectal cancer. *Clin Chim Acta*. 2019;488:209–14.
- Cheung HH, Lee TL, Rennett OM, Chan WY. DNA methylation of cancer genome. *Birth Defects Research Part C: Embryo Today: Reviews*. 2009;87(4):335–50.
- Baylin PAJ. SB The epigenomics of cancer. *Celt*. 2007;128(4):683–92.
- Shen H, Laird PW. Interplay between the cancer genome and epigenome. *Cell*. 2013;153(1):38–55.
- Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, Zeiger M, et al. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med*. 2014;6(8):1–11.
- Bhat SA, Ahmad SM, Mumtaz PT, Malik AA, Dar MA, Urwat U, et al. Long non-coding RNAs: Mechanism of action and functional utility. *Non-coding RNA research*. 2016;1(1):43–50.
- Amin V, Harris RA, Onuchic V, Jackson AR, Charnecki T, Paithankar S, et al. Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs. *Nat Commun*. 2015;6(1):1–10.
- Wang Z, Yang B, Zhang M, Guo W, Wu Z, Wang Y, et al. lncRNA epigenetic landscape analysis identifies EPIC1 as an oncogenic lncRNA that interacts with MYC and promotes cell-cycle progression in cancer. *Cancer Cell*. 2018;33(4):706–20. e9.

29. Sloane MA, Ward RL, Hesson LB. Defining the criteria for identifying constitutional epimutations. *Clin Epigenetics*. 2016;8(1):1–2.
30. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 2007;39(4):457–66.
31. Tao R, Hu S, Wang S, Zhou X, Zhang Q, Wang C, et al. Association between indel polymorphism in the promoter region of lncRNA GAS5 and the risk of hepatocellular carcinoma. *Carcinogenesis*. 2015;36(10):1136–43.
32. Li L, Sun R, Liang Y, Pan X, Li Z, Bai P, et al. Association between polymorphisms in long non-coding RNA PRNCR1 in 8q24 and risk of colorectal cancer. *J Experimental Clin Cancer Res*. 2013;32(1):1–7.
33. Xue Y, Gu D, Ma G, Zhu L, Hua Q, Chu H, et al. Genetic variants in lncRNA HOTAIR are associated with risk of colorectal cancer. *Mutagenesis*. 2015;30(2):303–10.
34. Kerachian MA, Javadmanesh A, Azghandi M, Shariatpanahi AM, Yassi M, Davodly ES, et al. Crosstalk between DNA methylation and gene expression in colorectal cancer, a potential plasma biomarker for tracing this tumor. *Sci Rep*. 2020;10(1):1–13.
35. Lee J-R, Kwon CH, Choi Y, Park HJ, Kim HS, Jo H-J, et al. Transcriptome analysis of paired primary colorectal carcinoma and liver metastases reveals fusion transcripts and similar gene expression profiles in primary carcinoma and liver metastases. *BMC Cancer*. 2016;16(1):1–11.
36. Kim D, Perteau G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):1–13.
37. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
38. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427–D32.
39. Sun K, Chen X, Jiang P, Song X, Wang H, Sun H. iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data. *BMC Genomics*. 2013;14(2):1–10.
40. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res*. 2013;41(6):e74–e.
41. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621–8.
42. Williams KP, Lau BY. RNAcentral: A comprehensive database of non-coding RNA sequences. *Nucleic Acids Research*. 2016;45(SAND-2017-0752J).
43. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35(suppl_2):W345–W9.
44. Washietl S, Findeiß S, Müller SA, Kalkhof S, Von Bergen M, Hofacker IL, et al. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA*. 2011;17(4):578–94.
45. Su Z-D, Huang Y, Zhang Z-Y, Zhao Y-W, Wang D, Chen W, et al. iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics*. 2018;34(24):4196–204.
46. Cao Z, Pan X, Yang Y, Huang Y, Shen H-B. The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics*. 2018;34(13):2185–94.
47. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Oxford University Press*; 2009. 1;59(6):pp. 892–902.
48. Yeku O, Frohman MA. Rapid amplification of cDNA ends (RACE). *RNA: Springer*; 2011. pp. 107–22.
49. Jung MK, Mun JY. Sample preparation and imaging of exosomes by transmission electron microscopy. *Journal of visualized experiments: JoVE*. 2018;(131).
50. Kerachian M, Javadmanesh A, Shariatpanahi A, Davodly E, Azghandi M, Yassi M, et al. A simple and cost-effective approach for technical validation of next generation methylation sequencing data. preprint 2019 (10.12103/rs.2.14216/v1).
51. Balas MM, Johnson AM. Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-coding RNA research*. 2018;3(3):108–17.
52. Cabili MN, Dunagin MC, McClanahan PD, Bialesch A, Padovan-Merhar O, Regev A, et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol*. 2015;16(1):1–16.
53. Arab K, Park YJ, Lindroth AM, Schäfer A, Oakes C, Weichenhan D, et al. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol Cell*. 2014;55(4):604–14.
54. Li Q, Wang P, Sun C, Wang C, Sun Y. Integrative analysis of methylation and transcriptome identified epigenetically regulated lncRNAs with prognostic relevance for thyroid cancer. *Front Bioeng Biotechnol*. 2020;7:439.
55. Heilmann K, Toth R, Bossmann C, Klimo K, Plass C, Gerhauser C. Genome-wide screen for differentially methylated long noncoding RNAs identifies *Esrp2* and lncRNA *Esrp2-as* regulated by enhancer DNA methylation with prognostic relevance for human breast cancer. *Oncogene*. 2017;36(46):6446–61.
56. Tang B. Inference of crosstalk effects between DNA methylation and lncRNA regulation in NSCLC. *BioMed research international*. 2018;2018.
57. Zhou Z, Lin Z, Pang X, Tariq MA, Ao X, Li P, et al. Epigenetic regulation of long non-coding RNAs in gastric cancer. *Oncotarget*. 2018;9(27):19443.
58. Bao S, Zhao H, Yuan J, Fan D, Zhang Z, Su J, et al. Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. *Brief Bioinform*. 2020;21(5):1742–55.
59. Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev*. 2015;29(13):1343–55.
60. Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods*. 2015;12(8):767–72.
61. Sun T, Wu R, Ming L. The role of m6A RNA methylation in cancer. *Biomed Pharmacother*. 2019;112:108613.
62. Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, et al. A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev*. 2015;29(19):2037–53.
63. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*. 2012;149(7):1635–46.
64. Warda AS, Kretschmer J, Hackert P, Lenz C, Urlaub H, Höbartner C, et al. Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep*. 2017;18(11):2004–14.
65. Zhang S, Zhao BS, Zhou A, Lin K, Zheng S, Lu Z, et al. m6A demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by sustaining FOXM1 expression and cell proliferation program. *Cancer Cell*. 2017;31(4):591–606. e6.
66. Robertson KD, Jones A. P. DNA methylation: past, present and future directions. *Carcinogenesis*. 2000;21(3):461–7.
67. Clouaire T, Stancheva I. Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell Mol Life Sci*. 2008;65(10):1509–22.
68. Jeffery L, Nakielnny S. Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J Biol Chem*. 2004;279(47):49479–87.
69. Glaich O, Parikh S, Bell RE, Mekahel K, Donyo M, Leader Y, et al. DNA methylation directs microRNA biogenesis in mammalian cells. *Nat Commun*. 2019;10(1):1–11.
70. Sun L, Zhang Z, Bailey TL, Perkins AC, Tallack MR, Xu Z, et al. Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse Klf1 knock-out study. *BMC Bioinformatics*. 2012;13(1):1–12.
71. Uchida S. High-Throughput Methods to Detect Long Non-Coding RNAs. *High-throughput*. 2017;6(3):12.
72. lloft NE, Ponting CP. Predicting long non-coding RNAs using RNA sequencing. *Methods*. 2013;63(1):50–9.
73. Xu M-d, Qi P, Du X. Long non-coding RNAs in colorectal cancer: implications for pathogenesis and clinical application. *Mod Pathol*. 2014;27(10):1310–20.
74. Tsagakis I, Douka K, Birds I, Aspden JL. Long non-coding RNAs in development and disease: conservation to mechanisms. *J Pathol*. 2020;250(5):480–95.
75. Amicone L, Citarella F, Cicchini C. Epigenetic regulation in hepatocellular carcinoma requires long noncoding RNAs. *BioMed research international*. 2015; 10;2015.
76. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*. 2009;136(4):629–41.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.