



Perspective on long noncoding RNA functionality

Agnik Halder¹, Sanga Mitra², Jayprokas Chakrabarti^{3,4}

¹Department of Bioinformatics, Maulana Abul Kalam Azad University of Technology, West Bengal, India; ²RNA Biology Section, Laboratory of Cell and Molecular Biology, NIDDK, National Institute of Healths, Bethesda, MD 20892, USA; ³Computational Biology Group, Indian Association for the Cultivation of Science, Kolkata-700032, India; ⁴Gyanxet, BF 286 Salt Lake, Kolkata, India

Correspondence to: Jayprokas Chakrabarti, Senior Author. Computational Biology Group, Indian Association for the Cultivation of Science, Kolkata-700032, India. Email: j.chakrabarti@gyanxet.com; Sanga Mitra. RNA Biology Section, Laboratory of Cell and Molecular Biology, NIDDK, National Institute of Healths, Bethesda, MD 20892, USA. Email: sanga.mitra@nih.gov.

Comment on: Liu SJ, Horlbeck MA, Cho SW, *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 2017;355. pii: aah7111.

Submitted Jun 06, 2017. Accepted for publication Jun 19, 2017.

doi: 10.21037/tcr.2017.07.20

View this article at: <http://dx.doi.org/10.21037/tcr.2017.07.20>

Introduction

Long non-coding RNAs or lncRNAs, are non-protein coding transcripts longer than 200 nucleotides, distinguishing them from small regulatory RNAs such as microRNAs (miRNAs), short interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), and other short RNAs. As the number of lncRNAs-genes has gone above and beyond the number of protein coding genes in human, lncRNAs have begun to gain widespread attention in recent years, as a result the need to identify them as well as determine their functions has gained precedence. To understand the functions, it is important to know about regulation of chromatin structures and the roles lncRNAs play in the facilitation of transcription factors. In this article, we discuss the methods for identification of lncRNAs, focusing mainly on the CRISPR/CAS9 and its derivative CRISPRi, and the known functions of the lncRNAs in the transcription regulation of nearby and distant genes. Liu *et al.* developed a CRISPR interference (CRISPRi) platform targeting 16,401 lncRNA loci in seven diverse cell lines, including six transformed cell lines and human induced pluripotent stem cells (iPSCs) (1). They identified the functions of 499 lncRNAs, which were found to be required for robust cell growth, and attempted to build a machine learning model to correlate the features of these lncRNAs to data obtained from ENCODE, FANTOM and Vista and from other sites on enhancer maps, expression levels, chromosomal looping data, conservation and copy number variations.

It is known that lncRNAs regulate complex transcriptional and post transcriptional network, critically involved in many disease pathways, including cancer, and take part in crucial epigenetic controls, yet much of their functionalities have remained beyond experimental methods. Out of the total of 27,692 human lncRNA transcripts in GENCODE, functionalities of just 184 are listed. lncRNAs of all kinds have been implicated in a range of growth and developmental processes, but knowledge of their mechanisms are yet to be explored. As the number of lncRNA transcripts begin to mount rapidly, significantly beyond the present count, the work of Liu *et al.*, dealing with their functions and as important nodes in transcription network, could not have been timelier.

Human protein coding genes number about 20,684 making up less than two percent of the genome. However, transcriptomic-studies over the last decade indicate 70 to 90 percent of the human genome is transcribed. It stands to reason, therefore, that the cellular functionality derives significantly from non-coding genes (2). A large chunk of these noncoding genes transcribes to lncRNAs. The genomic locations of the lncRNA genes, especially if they are very near mRNA-genes, may make it difficult to unambiguously determine the lncRNA-functionality. Considering the genomic locations of lncRNA-genes, the lncRNAs are classed into: (I) sense (i.e., the lncRNA sequence overlaps with the sense strand of a mRNA-gene); (II) antisense (i.e., the sequence overlaps with the antisense strand of a mRNA gene); (III) bidirectional (the sequence

is oriented head to head with a mRNA gene within 1 kilobase); (IV) intronic (the sequence is located inside the intron of a mRNA gene); (V) intergenic (the sequence is found in a region between two mRNA genes).

Considering that many of the disease related loci in humans are known to be on lncRNA-gene regions, their functions, presently largely unknown, are of immediate interest. Amongst them are:

Regulation of chromatin structure

Chromatin structures are important in the facilitation of transcription factors and in regulation of the gene promoters.

Genomic DNA is present in the eukaryotic nucleus and is hierarchically packaged in the chromatin by the histones. There are multiple levels of the DNA packaging. Among them the essential repeating unit of chromatin is the nucleosome, consisting of 146 base pairs of DNA wrapped in 1.7 superhelical turns around an octamer of histone proteins (3). The linker histones (H1 or H5) binds in a manner which sorts the nucleosome arrays into a much more condensed 30-nm chromatin fiber (3) which is typically hypothesized as being the second structural level of DNA organization.

Higher-order structure refers to the accumulation of nucleosomes, i.e., conformation in the 3D space. Mitotic/meiotic chromosome has around 15,000–20,000 folds of DNA packaged into them (4). Chromatin structures are experimentally challenging using light and electron microscopy mainly due to the difficulties in resolving nucleosomes and linker DNAs in the compact chromatin in the nucleus (5).

Large scale chromatin loops are crucial to chromatin functions. The various loop phenomena that have been reported, may constitute distinct level(s) of chromatin higher-order structure (6). Chromatin loopings between nuclear receptor response elements and activated promoters depend dynamically on lncRNAs interactions with chromatin remodeling complexes, contributing to remodel chromatin and hence regulate gene expression.

Examples are also provided in a recent review by Li *et al.* (2016) which discussed the detailed mechanisms of triple helix formation by lncRNAs and DNAs *in vivo*. Out of the seven lncRNAs, six were seen to form triple helices as well as enlist chromatin modifiers (see Table 1, Li *et al.* 2016). It is quite noteworthy that these proteins are also involved in *de novo* chromatin modifications.

Transcription regulation

lncRNAs control transcription of neighboring genes, as well as interact with chromatin at a plethora of different locations across numerous chromosomes to regulate distal gene expressions.

Nearby effects

There are different mechanisms by which lncRNAs are seen to be regulating the transcriptional process. They can interfere with the promoters of protein-coding genes by themselves or by mediating demethylation of those promoters. They are also able to intervene as co-activators of certain transcription factors or, on the contrary, act as decoys of transcription factors, blocking their binding to DNA. In addition, it has been observed that the lncRNA MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript 1) interacts with a type of splicing factors, contributing to the regulation of the mRNA splicing process (6).

In a recent survey on the human genome with GENCODE annotation¹³, it was found that a second class of lncRNAs exhibited enhancer-like function. These lncRNAs are crucial for the robust expression of their neighboring protein-coding genes in multiple cell lines as well as transcription activation from the thymidine kinase promoter in luciferase reporter assay. The assembly of transcription factors is supported by the lncRNAs only when these lncRNAs are activated on heterologous promoters. This falls similar in line to that of the human HOTAIR RNA, where the lncRNA-mediated assembly of different histone modifying complexes are necessary for their DNA binding (7) and silencing of HOXD gene expression.

Reduction of an lncRNA from the MyoD1 enhancer led to MyoD1 chromatin accessibility and RNA polymerase II (PolII) occupancy depletion (8) and consequent downregulation of MyoD1 expression. This means that enhancer related transcripts can regulate enhancer activity by changing nearby chromatin accessibility/structure. Similarly, lncRNAs from p53 bound enhancers influenced chromatin conformations to activate enhancer activity; also at estrogen receptor bound enhancers inhibition of lncRNA production by blocking transcriptional elongation had no effect on chromatin looping, still downregulated target gene activation (9). This has led to the conclusion that, enhancer associated lncRNAs have many RNA based pathways for regulation of transcription.

In vitro studies indicate lncRNAs interact cohesively and introduce looping interactions between their enhancers and the promoters of neighboring mRNA genes.

The human growth hormone (*bGH*) HS1 enhancer is activated by lncRNA transcription downstream of HS1 (10). The molecular mechanism of this activation was investigated by adding a transcriptional terminator at the site, leading interestingly to reduced lncRNA transcription, together with downregulation of *bGH* expression (10).

Regulation of nearby gene expression are mediated by *cis* acting lncRNA actions on the same chromosome and is allele specific. Remarkably, *trans* acting lncRNA pathways also regulate local gene expression. More studies are required to evaluate the relative contributions of *cis* and *trans* acting lncRNA activities in regulating nearby gene expressions.

Let us consider the instance of lncRNA Jpx, which is transcribed from the active X chromosome. It has been noted that Jpx RNA upregulates the *Xist* gene (which lies adjacent to it) by detaching the CTCF protein. This occurs on the inactive X allele, during the process of X chromosome inactivation (XCI) (11). In a quite indistinguishable manner, the enhancer elements present in the locus of *Evf2* are in fact regulated by it. This is brought about by the binding of distal-less homeobox 2, *DLX2*, and methyl CpG binding protein 2, *MECP2*. Methylation of this enhancer is then downregulated and the expression of the nearby *Dlx6* gene is controlled (11).

As evident from *fbp1+* geneactivation in yeast (12), RNA independent functions in gene activation are influenced by lncRNA loci, arising out of transcription affecting local chromatin accessibility. Thus, lncRNAs which are enhancer dependent act locally, through either RNA dependent or independent pathways, to increase the transcriptional activity of chromosomally close mRNA genes.

Distal effects

Long range lncRNA interactions with chromosomes are known to impact mRNA gene regulation. The lncRNAs *TFF1* and *NRIP1*, exists 27 Mb apart on chromosome 21, come spatially close by DNA looping interactions. Induced by E2, the looping depends on the *NRIP1* (9).

The molecular pathways in targeting lncRNAs to distal binding sites are under active investigation, the spatial genomic configuration clearly has a role. The intergenic lncRNAs are turning out to be more involved in gene regulation than earlier anticipated.

Amongst other known lncRNA functions with direct or

indirect effect on mRNA-gene expressions are:

- ❖ Regulation of mRNA translation: some lncRNAs decrease the translation of mRNAs by reducing polysomes, inducing what is known as ribosome 'drop-off'. Interestingly, lncRNAs may also enhance translation by overlapping at the 5' end of target mRNAs regulating the interactions of polysomes-mRNA;
- ❖ The ceRNA effect: interactions with other ncRNAs: lncRNAs behave as sponges for miRNAs (13) affecting their availability and as a result their functions. These lncRNAs are termed as competing endogenous RNAs, or ceRNAs;
- ❖ Regulating mRNA stability: lncRNAs regulate the stability of their target mRNAs by activating the SMD (Staufen 1-Mediated messenger RNA Decay) pathways involved in degrading mRNAs (14). There are indications that lncRNAs increase the stability of mRNAs by binding to them, regulating their degradation;
- ❖ Organization of nuclear structures: long intergenic noncoding RNA Firre (Functional Intergenic Repeating RNA Element) is involved in the maintenance of the nuclear architecture (15) across chromosomes;
- ❖ Regulation of protein activity: the subtype termed sno-lncRNAs, because their sequence is flanked by snoRNA-genes, bind to alternative splicing factors regulating and interfering in the splicing process.

Despite the differences, lncRNAs share important similarities with mRNAs. Many of the lncRNAs are transcribed by RNA polymerase II; the promoters have histone marks (16) controlling gene expression regulation; splicing events are known, and these transcripts are polyadenylated and 5' capped. lncRNAs expression levels are significantly more tissue specific than the mRNAs even though the lncRNAs are expressed at much lower levels than the latter. It is this tissue specificity of lncRNA expression which has been critical in the attraction of widespread attention to their putative therapeutic use because of the possibility of selective targeting.

Aside from the determination of the functionalities of a large number of lncRNAs, the work of Liu *et al.* sheds important semi-quantitative insight into the cell-specificity of their expressions. They worked on seven fairly diverse cell lines, namely, (I) chronic myeloid leukemia cell line, K562; (II) cervical cancer cell line, HeLa; (III) glioblastoma line, U87; (IV) mammaryadenocarcinoma line, MCF7; (V) mammary adenocarcinoma line MDA-MB-231; (VI) human embryonic kidney line, HEK293T; and (VII) a human iPSC

line. These cell lines were studied in ENCODE.

The identification of lncRNAs and determination of their functions have proven to be an integral problem in need of a solution, and it relies on the detection of transcription from genomic regions that has not been annotated as protein coding. This can be achieved by the direct detection of the transcribed RNA. However, usual gene expression microarrays are only able to detect the expression of protein-coding mRNAs. Thus the need for unbiased RNA detection methods becomes paramount. Some methods which are being considered as a probable solution to the said problem: tiling arrays, serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE) and high-throughput RNA sequencing (RNA-seq). Alternatively, detection of some specific histone marks (such as H3K4–H3K36 domains) usually signifies transcriptionally active chromatin which as a result can detect transcription from genomic DNA analyses.

While there are other methods to study the functions of lncRNAs, many of them are not suitable for large-scale screening studies. CRISPR/CAS9 method is generally appropriate for mRNA genes as the introduced indels alter the reading frames, reducing the expression of the targeted mRNA-gene. Liu *et al.* used the CRISPR interference, CRISPRi, technology to repress the expression of lncRNA genes to determine their functions. Transcriptional repression with CRISPRi employs transcription inhibitors complexed with endonuclease-dead CAS9, dCAS9, to suppress transcription. Single Guide RNA, sgRNA, targets specific DNA sites. The method involves blocking transcription by chromatin modifications within the window of around 1 kb encompassing the transcription start site TSS. It is a highly efficient way for manyfold reduction of targeted lnc-gene expression. The inhibitive chromatin modification H3K9me3 induced significantly reduced off-target effects. To optimize the on-target activity, Liu *et al.* developed an improved sgRNA library, towards genome wide screening of lncRNA gene functions, that also has the potential to suppress any residual off-target effects. The important transcriptome annotations led to the identification of the important lncRNA gene sets, a third of these lncRNAs were prioritized based on a panel of normal and cancer cell lines. Per TSS, ten sgRNAs were designed, using hCRISPRi-v2.1, to target all the chosen lncRNA genes. This invaluable library, CRiNCL, accompanies the Liu *et al.* script.

As of January 2016, 294 lncRNAs have been functionally annotated in LncRNADB, the majority of these, 184, are in humans. A list of all the functional lncRNAs from LncRNADB is stored in a spreadsheet. Liu *et al.*, found the functions of 499 lncRNAs that do not have coding genes

in their neighborhood from their CRISPRi screening experiment. These lncRNAs were found to modify cell growth. Of these 499, the functions of eight were listed in LncRNADB. Interestingly, a large number out of these 499 of cell-growth modifying lncRNAs were found in the iPSC line leading to the surmise that these cells were differentiating towards cells with reduced growth rates and towards diminished pluripotency.

To comprehend the consequences, RNA-seq was performed following the CRISPRi in a subset of the 499 cases in three cell lines. When different sgRNAs were used to target the same lncRNA gene TSS, highly correlated transcriptomes emerged. By contrast, sgRNAs targeting different lncRNA-gene TSS with similar phenotypes resulted in dissimilar transcriptomes, implying distinct pathways. The result of this systematic analysis through the various cell lines implied that the lncRNA-genes are nodal elements in fairly intricate transcriptional network, disturbing any of these lnc-nodes leads to its unique characteristic disruption of the transcription downstream. The authors have tried to correlate the transcription disruption caused by suppressing lncRNA expression by CRISPRi with the chromosomal location of the disrupted transcription nodes and reported of evidence of correlation in just 14 cases where the lncRNA gene expression repression affected the transcription scenario in the immediate neighborhood, i.e., within 20 genes.

An important issue in lncRNA gene expression has been the tissue specificity. Here the authors observed that 89.4% of the lncRNA hits were unique to just one out of the seven cell lines. Total number of lncRNA genes expressed in the seven cell lines numbered 1329 of which 82.6% of the lncRNA hits modified growth in just one cell line. Clearly, the lncRNA gene expressions are significantly more cell specific than the mRNA genes. A detailed analysis of the lncRNA LINC expressions in U87, K562, MCF7 and HeLa led to the conclusion that the cell specific differences in expression were not due to varied activity of CRISPRi across cell lines, but connected to the divergences in their transcriptional networks.

The authors attempted to build a machine learning model to correlate the features of these 499 lncRNAs to other observed parameters, such as enhancer maps, expression levels, chromosomal looping data, conservation and copy number variation obtained from ENCODE, FANTOM and Vista and from a few other sites. It turned out that the expression levels in cell lines, lncRNA genes within 1kb of FANTOM enhancers, lncRNAs within 5kb of cancer associated SNPs and the number of exons in

the lncRNA genes were significantly correlated for the expression data in the 499 cases, but with a few exceptions. The correlations of cell-specificity of lncRNA expressions to proximity of enhancers, and other chromosomal contacts led to the surmise that the higher order chromatin structure might be the underlying reason. In particular the differences between chromosomal looping of the lncRNA gene promoters and target genes could be one crucial factor.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned and reviewed by the Section Editor Long Chen (Department of PET-CT center at the Yunnan Tumor Hospital, The Third Affiliated Hospital of Kunming Medical University, Kunming, China; Department of Biochemistry and Molecular Biology of Kunming Medical University, Kunming, China).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tcr.2017.07.20>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Liu SJ, Horlbeck MA, Cho SW, et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 2017;355. pii: aah7111.
- Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics* 2013;193:651-69.
- Li G, Reinberg D. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* 2011;21:175-86.
- Woodcock CL, Ghosh RP. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol* 2010;2:a000596.
- Wang Y, Maharana S, Wang MD, et al. Super-resolution microscopy reveals decondensed chromatin structure at transcription sites. *Sci Rep* 2014;4:4477.
- Tripathi V, Ellis JD, Shen Z, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 2010;39:925-38.
- Tsai MC, Manor O, Wan Y, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 2010;329:689-93.
- Mousavi K, Zare H, Dell'orso S, et al. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* 2013;51:606-17.
- Vance KW, Ponting CP. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet* 2014;30:348-55.
- Yoo EJ, Cooke NE, Liebhaber SA. An RNA-independent linkage of noncoding transcription to long-range enhancer function. *Mol Cell Biol* 2012;32:2020-9.
- Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* 2010;143:390-403.
- Hirota K, Miyoshi T, Kugou K, et al. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* 2008;456:130-4.
- Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;146:353-8.
- Park E, Maquat LE. Staufen-mediated mRNA decay. *Wiley Interdiscip Rev RNA*. 2013;4:423-35.
- Cheng L, Ming H, Zhu M, et al. Long noncoding RNAs as Organizers of Nuclear Architecture. *Sci China Life Sci* 2016;59:236-44.
- Kaikkonen MU, Lam MT, Glass CK. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res* 2011;90:430-40.

Cite this article as: Haldar A, Mitra S, Chakrabarti J. Perspective on long noncoding RNA functionality. *Transl Cancer Res* 2017;6(Suppl 6):S1049-S1053. doi: 10.21037/tcr.2017.07.20