FEBS Letters xxx (2015) xxx–xxx

FEBS Letters

journal homepage: www.FEBSLetters.org

Computational identification of epigenetically regulated lncRNAs and their associated genes based on integrating genomic data

Tingting Zhao [a], Jinyuan Xu [b], Ling Liu [b], Jing Bai [b], Lihua Wang [c], Yun Xiao [b,*], Xia Li [b,*], Liming Zhang [a,*]

[a] Department of Neurology, The First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150001, China
[b] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150086, China
[c] Department of Neurology, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150086, China

ARTICLE INFO

Article history:
Received 31 August 2014
Revised 19 November 2014
Accepted 13 January 2015
Available online xxxx

Edited by Alfonso Valencia

Keywords:
Computational method
Integration
Long non-coding RNA
Transcriptome
Epigenome
Disease

ABSTRACT

Long non-coding RNAs (lncRNAs) are new players in various biological processes. However, understanding of lncRNAs is still in its infancy. Here, we proposed an integrative method to identify epigenetically regulated lncRNAs and their associated genes. By combining RNA-seq data and ChIP-seq data for histone H3 trimethylated at lysine 4 (H3K4me3) and H3K27me3, we identified 699 H3K4me3-regulated and 235 H3K27me3-regulated lncRNAs, each with an average of 238 and 307 associated genes, respectively. By analyzing Polycomb repressive complex 2 (PRC2) binding maps, we observed that the negatively related genes of most epigenetically regulated lncRNAs were enriched for PRC2-binding genes. In addition, through enrichment analysis, we inferred some lncRNAs with aberrant epigenetic modifications in glioblastoma and Alzheimer's disease. Together, we describe a method for the analysis of lncRNAs and demonstrate how integration of multi-omics data can improve understanding of lncRNAs.

1. Introduction

With the emergence of next-generation sequencing technologies, the characterization of mammal transcriptome is allowed at an unprecedented resolution [1]. Genome-wide transcriptome analyses have revealed that almost the entire genome is transcribed to some degree [2], however, only a minority of transcripts can be translated into proteins [3]. Among the human genomic regions without protein-coding potential, thousands of long non-coding transcripts (lncRNAs) are identified [4].

LncRNAs are emerging as a class of important regulatory molecules showing developmental and tissue-specific expression patterns [5]. Accumulating evidence has shown that lncRNAs participate in many essential biological processes, such as genomic imprinting, maintenance of pluripotency, immune response and development [6]. Moreover, recent studies have linked lncRNAs to human diseases, such as neurodegenerative disorders, cardiovascular diseases and cancers [7]. They demonstrate that aberrant lncRNA expression may be a major contributor to the pathogenesis of disease [8,9].

However, the regulation of lncRNA expression has so far been poorly addressed relative to protein-coding genes (PCGs) [6]. A large number of studies have been carried out to analyze the regulation of PCGs, and demonstrate that epigenetic modifications of their promoters can substantially influence chromatin structure and, in turn, control their expression [10–12]. Recent studies performed genome-wide analysis of the epigenetic modifications of lncRNAs in different cell and tissue types and revealed that lncRNAs showed similar epigenetic modification patterns as PCGs [13–15]. They observed that active histone H3 trimethylated at lysine 4 (H3K4me3) and histone H3 trimethylated at lysine 36 (H3K36me3) were both enriched in highly expressed lncRNAs, whereas Polycomb-mediated histone H3 trimethylated at lysine 27 (H3K27me3) was enriched in lowly expressed lncRNAs. In addition, genome-wide epigenetic modification was also used to identify novel lncRNAs [16]. These findings highlighted that epigenetic modifications may critically contribute to the regulation of lncRNA expression.

Emerging evidence indicates that lncRNAs play various important roles in global gene regulation, such as sensory, guiding and scaffolding capacities [17]. LncRNAs can recruit chromatin-modifying complexes to specific genomic loci for modulating chromatin

* Corresponding authors at: Department of Neurology, The First Clinical College of Harbin Medical University, Harbin 150081, China. Fax: +86 451 86615922 (L. Zhang).

E-mail addresses: xiaoyun@ems.hrbmu.edu.cn (Y. Xiao), lixia@hrbmu.edu.cn (X. Li), zhanglimingjack@163.com (L. Zhang).

http://dx.doi.org/10.1016/j.febslet.2015.01.013
0014-5793/

states and transcriptional activity of their targets in *cis* or in *trans* [18]. Recently, genomic maps of a few lncRNAs demonstrated that lncRNAs could occupy many genomic sites [19]. Another part of lncRNAs, such as HOX transcript antisense RNA (HOTAIR), can serve as a scaffold for histone modifying complexes and guide them to their targets for silencing expression through H3K4 demethylation and H3K27 methylation [19,20]. Some lncRNAs, such as GAS5, can act as decoys for gene repression [21]. Besides, lncRNAs can also form RNA–RNA interaction for changing RNA structures [22]. Such complex regulatory mechanisms [23] impede the comprehensive characterization of regulatory targets of lncRNAs.

It should be noted that both epigenetic modification and targeting regulation can be involved in the regulation of many important biological processes [24]. Epigenetic alterations can transcriptionally activate or silence some key regulators (such as transcription factors and lncRNAs). In such case, increased (or decreased) expression of these regulators mediated by epigenetic alterations can subsequently affect their downstream targets by various targeting mechanisms [25]. For example, ZEB1, an important transcription factor, has been associated with the transition during development and the regulation of the epithelial-mesenchymal transition (EMT). When the ZEB1 promoter converts from a bivalent to active chromatin state, its increased expression leads to expression changes of many differentiation-related genes by direct and indirect means, which in turn drives the generation of cancer stem cells [26]. A recent study reported that a lncRNA DLX1AS can function as a key determinate of neurogenesis by converting from a bivalent to active chromatin state during neuronal differentiation [27]. The epigenetic alteration results in the increased expression of DLX1AS, which then induce an increase in the expression of its protein-coding gene neighbors, DLX1 and DLX2, two important transcription factors in neuron differentiation [28].

Based on the above evidence, it is reasonable to assume that epigenetic alterations at the promoters of lncRNAs can change their expression levels, and in turn influences the expression of their downstream target genes by direct and indirect means (Fig. 1). Such cascade relationships allow us to identify epigenetically regulated lncRNAs and their affected targets whose expression is dynamically dependent on epigenetic states at lncRNA promoters. Notably, both direct and indirect targets of lncRNAs can be influenced by lncRNA expression changes through information flow in complex regulatory network. Such direct and indirect effects have been reported in transcription factors and miRNAs [29–31]. Thus, these direct and indirect targets should show similar dependence of expression on epigenetic states of lncRNA promoters. We referred to the direct and indirect targets as the relevant genes of lncRNAs. Based on the above consideration, we combined transcriptome from RNA-seq data and epigenome from ChIP-seq data across numerous human cell lines to identify a particular type of cascade relationship in which epigenetic alterations at lncRNA promoters influence the expression of lncRNAs and in turn change the expression of their relevant genes. Subsequently, we used the relevant genes of epigenetically regulated lncRNAs to characterize the functions of these lncRNAs. In addition, using expression profiles of PCGs from glioblastoma (GBM) and Alzheimer's disease (AD), we inferred GBM- and AD-associated lncRNAs with aberrant epigenetic modifications, whose relevant genes were significantly enriched in differentially expressed genes.

2. Materials and methods

2.1. Data resource

The RNA-seq data for 27 cell lines (22 were paired-end sequenced and 5 were single-end sequenced) were downloaded from Gene Expression Omnibus (GEO) (Table S1). Of these RNA-seq data, 20 were obtained from the ENCODE Consortium (GSE 30567), and the other 7 were collected from different studies (GSE52657, GSE30786, GSE52447, GSE37061, GSE43070, GSE 45428, SPX209063 and SPX209064).

A total of 74 ChIP-seq data referring to two epigenetic marks (H3K4me3 and H3K27me3) and their corresponding input controls in the 27 cell lines were used (Table S2). Among these cell lines, ChIP-seq data for H3K4me3 in 27 cell lines and ChIP-seq data for

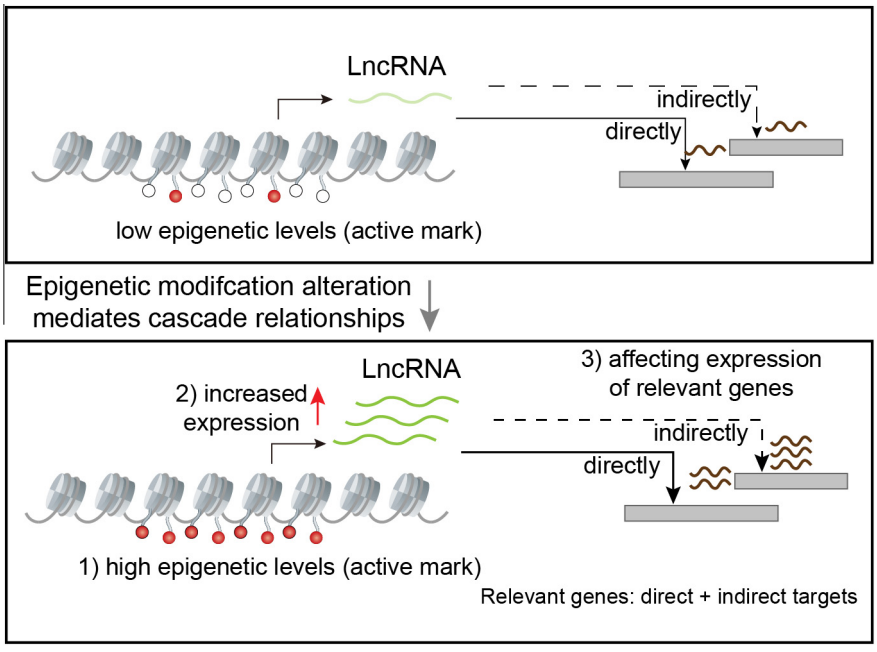

Fig. 1. Epigenetic-mediated cascade relationship. Epigenetic alterations at lncRNA promoters can influence their expression and in turn affect their downstream target genes by direct and indirect means.

Please cite this article in press as: Zhao, T., et al. Computational identification of epigenetically regulated lncRNAs and their associated genes based on integrating genomic data. FEBS Lett. (2015), http://dx.doi.org/10.1016/j.febslet.2015.01.013