Genomics 108 (2016) 64–77

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

RNA polymerase II promoter-proximal pausing in mammalian long non-coding genes

Heeyoun Bunch [a,*], Brian P. Lawney [b], Adam Burkholder [c], Duanduan Ma [d], Xiaofeng Zheng [e], Shmulik Motola [d], David C. Fargo [c], Stuart S. Levine [d], Yaoyu E. Wang [b], Guang Hu [e]

[a] Department of Genetics and Complex Diseases, Harvard School of Public Health, Boston, MA 02115, USA
[b] Center for Cancer Computational Biology, Dana Farber Cancer Institute, Boston, MA 02130, USA
[c] Integrative Bioinformatics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27705, USA
[d] BioMicro Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[e] Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, National Institute of Health, Research Triangle Park, NC 27705, USA

ARTICLE INFO

Article history:
Received 19 March 2016
Received in revised form 11 July 2016
Accepted 14 July 2016
Available online 16 July 2016

Keywords:
Long non-coding RNAs
RNA polymerase II promoter-proximal pausing
TRIM28

ABSTRACT

Mammalian genomes encode a large number of non-coding RNAs (ncRNAs) that greatly exceed mRNA genes. While the physiological and pathological roles of ncRNAs have been increasingly understood, the mechanisms of regulation of ncRNA expression are less clear. Here, our genomic study has shown that a significant number of long non-coding RNAs (lncRNAs, >1000 nucleotides) harbor RNA polymerase II (Pol II) engaged with the transcriptional start site. A pausing and transcriptional elongation factor for protein-coding genes, tripartite motif-containing 28 (TRIM28) regulates the transcription of a subset of lncRNAs in mammalian cells. In addition, the majority of lncRNAs in human and murine cells regulated by Pol II promoter-proximal pausing appear to function in stimulus-inducible biological pathways. Our findings suggest an important role of Pol II pausing for the transcription of mammalian lncRNA genes.

1. Introduction

Mammalian genomes are composed of a large number of non-coding RNAs (ncRNAs) that greatly exceed protein-coding genes, and significant physiological roles for these molecules have been discovered. Diverse mechanisms by which ncRNAs regulate gene expression have been reported such as silencing transcription [1–3], altering splicing [4,5], modulating translation [6], and regulating RNA transport [7,8]. Although lncRNA transcripts are not translated into proteins, they share similar features with protein-coding genes. For example, lncRNA genes are transcribed by RNA polymerase II (Pol II), the same and sole enzyme to transcribe protein-coding genes, undergo 5′ capping and polyadenylation, typically contain multiple exons, and are post-transcriptionally modified [9–13]. In spite of these characteristics and functions of an increasing number of lncRNA genes that have recently been reported, how transcription of lncRNA genes is regulated in mammalian cells is incompletely understood.

In mRNA gene transcription, it was originally thought that transcriptional initiation in which general transcription factors and Pol II are recruited to the promoter site and promoter escape was the major rate-limiting step for transcriptional activation in metazoan cells. However, advances in genome-wide analyses such as Chromatin Immunoprecipitation and Sequencing (ChIP-seq) and RNA Sequencing (RNA-Seq) have shown that the early elongating complex of Pol II becomes paused, about 20–100 bp downstream from the Transcriptional Start Site (TSS) in a substantial number of genes, identifying an additional, major regulatory step between early and processive elongation [14–16]. This step, characterized by TSS-engaged, stalled Pol II, is termed Pol II promoter proximal pausing. Pol II promoter proximal pausing occurs in a broad range of metazoan genes [16,17]. Approximately 30% of protein-coding genes and 70% of signal-induced or developmental genes are known to harbor Pol II paused at the promoter-proximal site [15,16,18]. While apparently repressive regarding gene expression, Pol II pausing is considered not to simply stall transcription. Pausing may be a proactive step to prepare the early elongating Pol II complex and the microenvironment for processive elongation upon gene activation [16,19–21].

Multiple transcription factors are known to regulate Pol II promoter-proximal pausing. These include factors such as DSIF [5,6-dichloro-1-b-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor] [22], NELF [Negative Elongation Factor] [22,23], Myc [14], P-TEFb [Positive Transcription Elongation Factor b] [24,25], TFIIS [General Transcription Factor IIS] [1], Gdown1 [26], and the Mediator complex [27] for

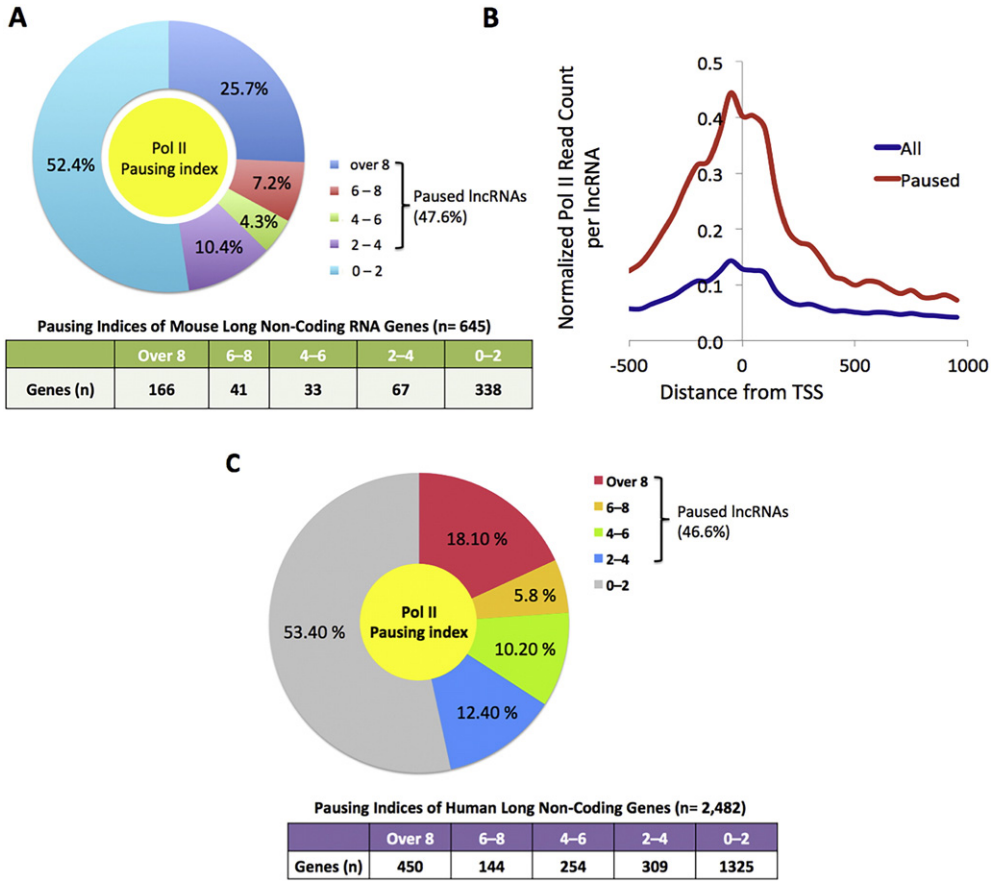
* Corresponding author at: Harvard School of Public Health, Department of Genetics and Complex Diseases, Harvard Medical School, Goldenson Bldg. Room 559, 220 Longwood Avenue, Boston, MA 02115, USA.
E-mail address: heeyounbunch@gmail.com (H. Bunch).

http://dx.doi.org/10.1016/j.ygeno.2016.07.003
0888-7543/

protein-coding genes. DSIF, NELF, and Gdown1 are known to mediate and stabilize Pol II pausing while P-TEFb, CDK8-Mediator [28], TFIIS, and MYC help release the paused Pol II. In our previous studies, another novel pausing factor TRIM28 has been identified to stabilize Pol II pausing in a broad range of protein-coding genes [29]. We have also found recently that TRIM28 is phosphorylated at residue S824 by phosphatidylinositol 3-kinase-related kinases (PIKK) such as DNA-PK and ATM (Ataxia Telangiectasia Mutated) during transcriptional activation and Pol II pause release at human *HSPA1B* [29]. Furthermore, our recent study has indicated that phosphorylated TRIM28 (S824) is increased in the transcribed regions of a number of serum activated protein-coding genes in humans [30]. These data suggest that TRIM28 has a biphasic characteristic: it represses transcription during the un-induced state while it functions to positively regulate transcriptional elongation during transcriptional activation. Consistently, other recent studies reporting the interaction of TRIM28 with P-TEFb suggest the important function of TRIM28 in transcriptional elongation [31,32].

To address the mechanisms that regulate transcription of lncRNA genes, we noted interesting characteristics that many lncRNA genes display. Studies have indicated that the majority of mammalian lncRNA genes are transcribed divergently from protein-coding genes [33,34]. It has been also shown that in more complex organisms such as human and mouse, protein-coding and enhancer genes are more bidirectional than in simpler organisms such as *Drosophila* [35,36]. In addition, divergent transcriptional initiation at a number of promoters and enhancers has been reported in human cells [37]. This divergent transcription might be beneficial in respect of coordinating the expression of coding and non-coding gene pairs, utilizing transcription factors, and compacting the genome. Interestingly, we have noticed that there are two distinctive Pol II peaks in the upstream and downstream close to the TSSs of protein-coding genes in mouse [29]. While the upstream peak could be the result of poised or docked Pol II for protein-coding genes that has been reported in *C. elegans* [38], whether it may indicate Pol II pausing in lncRNA genes is less clear and has not been statistically evaluated.

In this study, we thus aimed to determine to what extent Pol II promoter-proximal pausing occurs in lncRNA gene transcription. Pol II occupancy was analyzed to determine the significance of Pol II pausing in murine and human lncRNA genes. In addition, we attempted to characterize serum-inducible lncRNAs during the cell-cycle progression from G_0 to G_1 regarding Pol II pausing, because stimulus-inducibility is one of signature characteristics of protein-coding genes that utilize Pol II pausing. Lastly, we examined the function of TRIM28 in Pol II pausing and expression at lncRNA genes. Our genomic data have shown that Pol II pausing is a prevalent phenomenon in a subset of lncRNAs both in murine and human cells. The statistical occurrence of Pol II pausing in both of these species is comparable to what has been reported for protein-coding genes. Our data have suggested that TRIM28 regulates the expression of a subset of lncRNAs. In addition, we have examined the biological functions of murine and human lncRNAs that are regulated by



Pausing Indices of Mouse Long Non-Coding RNA Genes (n= 645)

	Over 8	6–8	4–6	2–4	0–2
Genes (n)	166	41	33	67	338

Pausing Indices of Human Long Non-Coding Genes (n= 2,482)

	Over 8	6–8	4–6	2–4	0–2
Genes (n)	450	144	254	309	1325

Fig. 1. Mammalian lncRNAs are regulated by Pol II promoter-proximal pausing. (A) Pausing indices of annotated lncRNAs ($n = 1535$) in mouse embryonic stem (mES) cells. The graph was generated with pausing index calculable genes (non-zero gene body Pol II occupancy, $n = 645$). LncRNAs with pausing indices over 2 are 307 including 166 lncRNAs with pausing indices over 8. (B) Pol II metagene analysis comparing total ($n = 1535$) versus paused lncRNAs ($n = 307$), displaying a significant accumulation of Pol II paused at TSSs in many lncRNAs in mES cells. (C) Pausing indices of lncRNAs ($n = 2482$) in human embryonic kidney 293 cells, indicating a large number of human lncRNAs to harbor Pol II paused in the promoter-proximal region. (D) A metagene analysis of Pol II depicting prevalent Pol II pausing in lncRNAs (all, blue, $n = 1752$; paused (PI > 2), green, $n = 839$; paused (PI > 8), red, $n = 316$) in humans. An orange bracket with a star marks a region (−100 to +150) for a zoom-in view shown in the top right panel. A green bracket marks the gene body region (>+600 from the TSS) that displays diverse Pol II occupancies in the 3 gene groups described above, a zoom-in view shown in the bottom right panel. PI, pausing index. (E) A chromosome view of *RBM26-AS1* showing accumulated Pol II occupancy in the promoter-proximal site.