Published in final edited form as:

Genomics. 2016 August ; 108(2): 64–77. doi:10.1016/j.ygeno.2016.07.003.

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

Heeyoun Bunch[1,†], **Brian P Lawney**[2], **Adam Burkholder**[3], **Duanduan Ma**[4], **Xiaofeng Zheng**[5], **Shmulik Motola**[4], **David C Fargo**[3], **Stuart S Levine**[4], **Yaoyu E Wang**[2], and **Guang Hu**[5]

[1]Department of Genetics and Complex Diseases, Harvard School of Public Health, Boston, MA 02115, USA

[2]Center for Cancer Computational Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02130, USA.

[3]Integrative Bioinformatics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27705, USA.

[4]BioMicro Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

[5]Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, National Institute of Health, Research Triangle Park, NC 27705, USA.

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, >1,000 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.

[†]**Corresponding Author Information:**Dr. Heeyoun Bunch, 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, Phone: 617-992-6288, heeyounbunch@gmail.com.

Competing interests: The authors declare no conflict of interest.

Accession codes. ChIP-seq genomic described in this study have been deposited in the Gene Expression Omnibus database under accession number GSE48253 and GSE75170.

Keywords

Long Non-Coding RNAs; RNA Polymerase II Promoter-Proximal Pausing; TRIM28

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 proteincoding 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 Immuno-Precipitation 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 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 proteincoding 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 Pol II pausing. It has been shown that stimulus-inducible and developmental lncRNAs tend to utilize Pol II promoter-proximal pausing in spite of tissue-specific, diverse biological roles that lncRNAs play in mammalian cells. Together, our study provide the first direct statistic measurement of the degree of Pol II pausing in lncRNA genes, which presents Pol II pausing to be a transcriptional mechanism that regulates the expression of a number of lncRNA genes in mammals.

Results

RNA polymerase II promoter-proximal pausing at murine and human lncRNA genes

Firstly, we aimed to measure the frequency of Pol II promoter-proximal pausing among annotated lncRNA genes in murine cells. A pausing index parameter used to define paused genes at protein-coding genes [29, 39] was applied to evaluate Pol II pausing at lncRNAs in murine embryonic stem (mES) cells. Briefly, TSS-proximal windows were defined as +/−250 nt, and gene body windows were defined as +500–2,500; if the size of a gene was smaller than 2,500 bp, the gene body window was defined as +500 to the gene end. Pausing index, the ratio of Pol II occupancy between the TSS and gene body [29, 39], was calculated to determine whether a gene is paused or not. A set of 1,535 lncRNAs derived from UCSC KnownGene annotations was included for the statistical analysis through the Pol II ChIP-seq data. Strikingly, Pol II promoter-proximal pausing occurred in approximately 47.6% of the mappable lncRNAs (at least 50% of unique alignment in the TSS or gene body windows, n= 645) (defined by pausing index over 2 [39]) (n= 307; Fig. 1A; Supplementary Data 1). The Pol II occupancy profile of these paused lncRNAs derived from a metagene analysis illustrated the prominent Pol II enrichment at the TSSs of these lncRNAs, in contrast to the profile of total lncRNAs (Fig. 1B). Since a large number of lncRNA genes are transcribed divergently from neighboring protein-coding genes, we investigated whether the Pol II peak in the promoter-proximal site included Pol II counts from protein-coding genes. As expected, it was shown that 50.9% of the lncRNA genes included in the study (782 out of 1,535 lncRNA genes) appeared to divergently transcribed. However, only 1 and 67 lncRNA genes among these lncRNAs were located within 1,000 and 5,000 bp from TSSs of protein-coding genes, respectively (Supplementary Data 2). This indicated that the Pol II peak between approximately −300 and +300, shown in Fig. 1B, represents Pol II promoter-proximal pausing in lncRNA genes, not in protein-coding genes. In addition, we note that the Pol II profile of lncRNA genes is highly comparable with one of protein-coding genes (Fig. S1), visualizing the established Pol II pausing in mES cells [14]. These data suggested that Pol II pausing could be an important transcriptional regulatory step in a number of lncRNAs in mES cells.

While the protein-coding regions that occupy less than 10% of the genome are 85% identical and evolutionally conserved, the non-coding regions show much less conservation in sequences between mouse and human [40]. Therefore, we next asked whether Pol II pausing also occurs at human lncRNA encoding genes in human embryonic kidney 293 (HEK293) cells as seen at ones in mES cells. Although lncRNAs are typically defined as ncRNAs whose sizes are longer than 200 nucleotides (nt), the lncRNAs filtered for our Pol II pausing analysis were equal or longer than 1,000 nucleotides (nt). We reasoned that a sufficient window for gene body could be critical to achieve a more accurate determination of Pol II pausing status, given that Pol II pausing index is calculated through the ratio of Pol II occupancy between TSS and gene body. The pausing index parameters used in mouse lncRNA pausing analysis described above were applied to the human lncRNA. A total of 12,270 ncRNA genes from UCSC hg19 KnownGene were initially downloaded, and 2,482 lncRNAs sizing over 1,000 nt were analyzed for Pol II pausing (Supplementary Data 3). The result showed prevalent Pol II pausing in human lncRNA encoding genes, indicating 46.6 %

of the examined lncRNAs to harbor Pol II paused in the promoter-proximal site (paused genes, Fig. 1C). Metagene analysis of lncRNA genes (all genes; paused genes with pausing index over 2; paused genes with pausing index over 8) showed significant Pol II accumulation in the TSSs (Fig. 1D). Pol II peaks in the TSS were higher (Fig. 1D left and top right panels) while gene body Pol II occupancies were slightly lower (Fig. 1D left and bottom right panels) for the genes with pausing index (PI) over 8 than the ones with PI over 2 or all genes. Compared to the metagene profile of mESC lncRNAs (Fig. 1B, all), human lncRNAs (Fig. 1D, all) displayed more prominent Pol II occupancy in the TSS. This suggested that Pol II pausing might take place more significantly at lncRNAs whose sizes are longer than 1,000 nt in humans than mice. A chromosome view of *RAM26-AS1* represents Pol II promoter-proximal pausing in human lncRNA genes (Fig. 1E). Importantly, we note that the degrees of Pol II pausing at lncRNAs are statistically comparable between mouse (47.6%) and human (46.6%) in spite of reported little conservation in non-coding gene sequences between the two species [40, 41]. These statistic data further suggest that Pol II pausing could be a conserved, important mechanism in lncRNA genes to regulate transcription as it has been established in protein-coding genes in mammals.

Role of TRIM28 in regulating Pol II pausing and transcription in lncRNAs in mES cells

It has been previously shown that TRIM28 regulates Pol II pausing and pause release at a number of protein-coding genes [29]. In addition, another study has reported that TRIM28 regulates erythropoiesis through controlling microRNA expression [42]. Based on these observations and the noticeable Pol II pausing observed in mammalian lncRNA genes (Fig. 1), we hypothesized that the novel pausing factor, TRIM28 may regulate Pol II pausing at noncoding genes. To investigate the potential role of TRIM28 in Pol II pausing, TRIM28 was knocked-down by short interfering RNA (siRNA) in mES cells. Pol II ChIP-seq analysis was performed, comparing global Pol II occupancies between WT and TRIM28 knock-down (KD) mES cells. The analysis indicated that TRIM28 KD led to increased or decreased pausing indices in 30.6 % (n= 93) of mappable lncRNAs (n= 304, Figs. 2A and S2). Among these 93 genes, the pausing index was decreased at 69 lncRNAs (74.2 %) and the index was increased at 24 genes (25.8 %) by TRIM28 KD (Fig. 2A, Table 1). These data suggested that TRIM28 KD attenuated Pol II pausing in a subset of murine lncRNA genes. Consistent with this finding, the metagene analysis of lncRNAs (n= 69) with decreased-pausing indices upon TRIM28 KD showed increased Pol II gene body occupancies (Fig. 2B). Pol II occupancy in the total annotated lncRNAs (n= 1,535), comparing WT and TRIM28 KD, is shown in the heatmaps (Figs. 2C and S2). In Fig. 2D, chromosome views of the *Stamos* gene (*Stam* opposite strand) depict an increase in the Pol II occupancy, in the immediately downstream of the pausing site (pausing site marked with a blue box; the immediately downstream of the pausing site with a red box), in TRIM28 KD mES cells. These results suggested a potential function of TRIM28 in stabilizing Pol II pausing in a subset of murine lncRNAs during the transcriptionally inactive state, similarly to the role of this factor in protein-coding genes [29].

We next examined the function of TRIM28 in ncRNA expression in murine cells. NcRNA gene expression profiles between WT and TRIM28 KD mES cells were compared by the microarray analysis. A total of 39,414 non-coding genes were included for this analysis.

Among them, 308 ncRNAs showed a fold change of at least 1.5 and also satisfied $p < 0.05$. Fifty-seven and 251 ncRNAs decreased and increased RNA expression upon TRIM28 KD, respectively (Fig. S3, Supplementary Data 4). This means that TRIM28 KD activated 81.5% of affected ncRNAs while it repressed 18.5% of these ncRNAs. Notable ncRNAs with fold changes (>+/−1.50; $p < 0.05$) and nomenclatures were listed in Table 2. A hierarchical clustering analysis comparing the relative level of the 308 ncRNAs in WT and TRIM28 KD mES cells depicted the function of TRIM28 to regulate the expression of a subset of ncRNA genes (Fig. 2E). Together, ChIP-seq and microarray data (Figs. 2A–E and S2–4) comparing WT and TRIM28 KD mES cells suggest that TRIM28 controls the transcription of a subset of non-coding genes.

Pol II pausing in inducible lncRNAs in mammalian cells

Engagement of Pol II in the TSS before processive elongation has been considered to be a preparatory stage in transcriptional activation at which transcriptional elements such as participating protein and nucleic acid molecules are primed for synchronized induction upon the receipt of activating signals [43, 44]. In addition, immediate early serum responsive proteincoding genes are known to harbor Pol II paused at the promoter proximal site. This suggests that Pol II pausing may be important to assist prompt and synchronized transcriptional activation in such genes upon induction.

We therefore screened for human lncRNAs that were activated or suppressed during serum induction and the resultant cell cycle transition from G_0 to G_1 in HEK293 cells. Using ChIP-seq analysis, Pol II occupancy at lncRNA genes was then monitored and compared between control (S0) and serum induced (S15) cells. Genes with noticeable pausing index changes (pausing index ratio change > 2) when induced were collected. We analyzed a total of 2,491 lncRNAs whose sizes were longer than 1,000 bp, selecting a total of 1,990 lncRNAs that possess non-zero pausing indices in both control and serum induced cells (Supplementary Data 5). We found that the pausing indices of 37.1% of these lncRNAs were decreased (n= 483) or increased (n= 217) by over 2 fold upon serum induction (Supplementary Data 5). LncRNA genes with pausing index decreased or increased by over 3 fold upon serum induction were listed in Table 3. Among those lncRNAs whose pausing index decreased (n= 483), 109 genes (22.6%) appeared to be non-paused (by the criterion of pausing index < 2) prior to induction. By contrast, 374 genes (77.6%) appeared to have been paused (pausing index > 2) in uninduced, control cells (Fig. 3A; Supplementary Data 5). These data suggest that expression of a number of human lncRNAs become induced during the G_0 to G_1 transition and that many of these early serum-inducible long non-coding genes harbored Pol II pausing.

Consistently, the heatmaps of Pol II for lncRNA encoding genes that underwent decreases in pausing indices upon serum induction showed a prominent Pol II promoter-proximal pausing profile prior to induction, represented by concentrated peaks near TSS (Fig. 3B, PI Decreased Genes). On the other hand, the genes with increased-pausing index displayed broad, spread peaks near the TSS in the un-induced state (Fig. 3B, PI Increased Genes, Serum –). In addition, the lncRNA genes with decreased pausing indices upon serum induction displayed mild increases in Pol II occupancy in the promoters, TSSs, and gene

bodies (Fig. 3B). In contrast, the lncRNA genes that showed increased-pausing indices upon serum induction showed Pol II occupancy increases in the divergent direction, including the promoter and upstream regions (Fig. 3B). Since many lncRNA genes are reportedly transcribed divergently to protein-coding genes [33], this may imply that the pausing index-increased lncRNA genes could be located divergently to the protein-coding genes whose expression becomes increased upon serum induction.

We next carried out a metagene analysis of lncRNA genes comparing the Pol II counts in the TSS and gene body between control and serum induced samples (Fig. 3C). This analysis showed that pausing index-decreased lncRNA genes increased gene body Pol II occupancies upon serum induction (n= 483, Fig. 3C, upper panel). On the other hand, pausing index-increased lncRNA genes showed decreased-gene body Pol II occupancies upon serum induction (n= 217, Fig. 3C, bottom panel). These data suggest the gene body Pol II occupancy to be a determinant of pausing index changes and a negative correlation between pausing index and the level of gene body Pol II occupancy. Comparison of the median Pol II occupancy between control and serum induced samples showed a consistent correlation between the gene body Pol II occupancy and pausing index in individual genes (Fig. 3D). For the pausing index increased and decreased genes, the values of empirical 95% confidence interval, −0.18 to +0.01 and 0.25 to 0.35, indicated the tendency of a mild decrease and increase in gene body Pol II, respectively. Unlike the gene body regions, there was no clear trend in the TSS Pol II occupancies in pausing-decreased and - increased genes, consistent with the results shown in Fig. 3C (Fig. 3D). This further confirmed the change in gene body Pol II occupancy to determine the pausing index increase or decrease upon serum induction. A chromosome view of *SNHG8* illustrating Pol II occupancy increase in the gene body upon serum induction was shown in Fig. 3E.

In our previous study, gene activation leads to phosphorylation of TRIM28 at S824, a modification that reverses its ability to stabilize Pol II pausing at a model paused gene, *HSPA1B* [29]. In addition, TRIM28 becomes phosphorylated at S824 at a number of immediate early protein-coding genes upon transcriptional activation [30]. Since TRIM28 regulates Pol II pausing in a subset of lncRNA encoding genes (Fig. 2), we hypothesized that TRIM28 might be modulated at these genes in a similar manner to stimulus-inducible protein-coding genes. Therefore, we examined levels of TRIM28 phosphorylated S824 (pTRIM28) at serum-activated ncRNA genes in humans. Genome-wide occupancy of pTRIM28 was monitored in control and serum induced HEK293 cells by ChIP-seq. Among 1,634 non-coding genes with non-zero pTRIM28 counts in both control and serum induced cells, 264 non-coding genes contained increased pTRIM28 levels and 284 non-coding genes decreased pTRIM28 counts by over 50% upon serum induced transcriptional activation (Supplementary Data 6). To further examine changes in mean Pol II occupancy associated with serum induction shown in figure 3C, we next investigated the correlation between pTRIM28 levels and Pol II occupancy in the serum-induced samples. The lncRNA genes with increased pTRIM28 levels upon serum induction also increased Pol II gene body occupancies, suggesting a positive correlation between transcriptional activation and pTRIM28 accumulation (Figs. 3F, 3G, and S5). On the other hand, the correlation between pTRIM28 levels and Pol II occupancy was less obvious in the genes in which pTRIM28 was decreased upon serum induction (p= 0.01; Figs. 3G and S5). These findings suggest that

TRIM28 becomes phosphorylated at S824 during gene expression at some human lncRNAs as reported in a number of protein-coding genes [29, 30].

We asked if the lncRNAs that harbor Pol II promoter-proximal pausing could be inducible as in the case of protein-coding genes [16]. To characterize the functions of lncRNA genes, GO analysis was performed for the murine and human lncRNAs separately, and the data were compared (Supplementary Data 7). The common cellular processes assigned to both human and murine lncRNAs were proliferation, protein functional regulation, protein modification, receptor signaling pathways, cellular organization/trafficking, metabolism, and immune response (Fig. 4A and B). In mES cells, a number of pausing-regulated lncRNAs were found to be involved in development and differentiation (Fig. 4A and B). This is in accordance with the fact that Pol II pausing occurs preliminarily in developmental protein-coding genes [45]. Interestingly, our analysis showed that the majority of human embryonic kidney lncRNAs were involved in regulating metabolic processes. Protein modifications such as phosphorylation and ubiquitination also appeared to be regulated by a number of lncRNAs in HEK293 cells. In mES cells, many lncRNAs were also found to regulate lipid metabolism and immune cell activation and maturation (Supplementary Data 7). These data suggest tissue-specific pathways that the murine and human lncRNAs are involved in. Importantly, it is estimated that about 59.1% of total paused lncRNAs gene counts (n= 719, *italic* in Fig. 4A) were stimulus-inducible in mouse and 65.9% (n= 946, *italic* in Fig. 4A) in humans, indicating that a large number of mammalian lncRNA genes, both longer than 1,000 nt and regulated by Pol II pausing, tend to be inducible. We note that such stimulus-inducibility is an important characteristic of paused protein-coding genes [16, 45]. Together, the data suggest tissue-specific, inducible functions of a subset of lncRNA genes utilizing Pol II pausing in mES and HEK293 cells as well as an intrinsic property of Pol II pausing benefiting to prime such lncRNA genes for synchronized transcriptional activation.

Discussion

Mammalian genomes are occupied dominantly by non-coding genes. While it is unclear what proportion of these protein-non-coding transcripts is functional, approximately 98% of the genome is transcribed and protein-coding sequences occupy only about 1.5–2.5% of the genome in humans [46, 47]. This estimation may imply the existence of diverse mechanisms to regulate the non-coding gene expression and to modulate their biological functions. However, compared to protein-coding gene expression, mechanisms to regulate non-coding gene expression and function have only recently begun to be understood. In addition, in spite of the significance and diversity of ncRNA biological functions, it is noted that only a limited number of studies regarding general mechanisms to regulate transcription of ncRNAs have been reported. In this regard, our study presented in this report would contribute to the growing mechanistic understanding of non-coding genes transcription.

For the first time, we statistically measured and identified the statistic prevalence of Pol II pausing in lncRNA genes through genomic analysis in murine and human cells. The results have shown that Pol II promoter proximal pausing occurs at a large number of lncRNA genes in mammalian cells, statistically comparable to the extent of Pol II pausing in mRNA genes. Importantly, the frequencies of Pol II pausing for lncRNA genes (47.6 % in mice and

46.6 % in humans) reported in this study are similar to, or even slightly higher than the frequency in coding genes (about 30%) [16]. This statistic values suggest that Pol II pausing could be a major checkpoint in lncRNA transcription as in protein encoding genes in mammalian cells.

Pol II pausing occurs in protein-coding genes, especially in stimulus-inducible and developmental genes that are expressed instantly for cellular signals and demands [16, 48], Pol II pausing appears as a wise cellular provision in that Pol II and transcriptional microenvironment could be prepared and ready at the pausing site for following processive elongation, immediately upon receipt of activating signals. In this study, we have presented a set of human immediate early lncRNAs whose expression could be controlled by serum induction, and have found that many of these lncRNAs harbor Pol II pausing (Fig. 3 A–C). In addition, lncRNAs that utilize Pol II pausing appear to be involved in stimulus-responsive biological pathways such as the immune response, development, proliferation, and synapsis (Fig. 4A and B). Therefore, Pol II pausing appears to equip these inducible lncRNA genes for efficient expression as widely conceived for paused protein-coding genes, perhaps for similar mechanistic advantages.

Pol II pausing has been shown to be regulated by various transcription factors at protein-coding genes. It remains to be understood whether well-characterized pausing factors established for protein-coding genes such as NELF, DSIF, MYC, and P-TEFb could play such roles also in noncoding gene transcription. Here, we investigated TRIM28 for ncRNA transcription as one of the pausing and elongation factors for a large number of protein-coding genes [29, 30]. In recent studies, TRIM28 has been reported to regulate differentiation and embryonic development [42, 49-52]. Some of these studies have shown that TRIM28 controls the expression of non-coding RNAs in regulation of differentiation and development [42, 52]. Our study indicates a role of TRIM28 in regulating Pol II pausing and expression of a subset of ncRNAs in mammalian cells (Fig. 2). Importantly, some serum induced lncRNAs showed elevated levels of pTRIM28 (Fig. 3F–G), an indicative marker for Pol II pause release and processive transcriptional elongation in serum-responsive protein-coding genes [29, 30]. These results suggest that TRIM28 may regulate Pol II pausing at stimulus-inducible lncRNA genes by a similar mechanism to proteincoding genes in which the activation of many of these genes involves phosphorylation of TRIM28 at S824 by DNA-PK and ATM. Recent studies have provided overlapping general mechanisms of transcriptional regulation involving functions of Mediator complex and topoisomerase I between non-coding RNA and protein-coding gene transcription [53, 54]. It is notable that the mammalian Mediator complex is known to be an important regulator both for activator-dependent transcription and Pol II pausing [27, 55]. This may also support that many lncRNAs could be signal-inducible and utilize Pol II pausing as a transcriptional regulatory mechanism.

In summary, we have shown that RNA polymerase II promoter-proximal pausing occurs prevalently both in murine and human long non-coding RNA genes (Fig. 4C). The statistic significance of RNA polymerase II promoter-proximal pausing in mammalian lncRNA genes is comparable with that of protein-coding genes. In addition, TRIM28 regulates Pol II pausing and expression in a subset of murine lncRNA genes (Fig. 4C). Serum inducible

lncRNA genes in human cells harbor Pol II paused in the promoter-proximal site, and some of these lncRNA genes accumulate phosphorylated TRIM28 at Serine 824, an indicative marker for Pol II pause release and processive elongation in protein-coding genes in humans (Fig. 4C). These data suggest a potential conservation of Pol II pausing regulatory mechanisms between protein-coding and noncoding gene transcription. Mammalian lncRNA genes harboring Pol II pausing are involved in stimulus-inducible biological pathways in spite of tissue-specificity. We suggest that these observations provide an insight into regulatory mechanisms of ncRNA expression and propose Pol II pausing as a general mechanism to regulate the transcription of stimulus-inducible lncRNA genes between early and processive elongation (Fig. 4C).

Materials and Methods

Cell culture and Cell Preparation Conditions.

For Pol II pausing study, HEK293 cells were grown in DMEM supplemented with 10% FBS and 1% Penicillin/Streptomycin (P/S) solution and synchronized to G_0 for gene expression. For serum induction experiments, HEK293 cells were grown to about 80% confluence. The cells were incubated in DMEM including 0.1% FBS and 1% P/S solution for 17.5 hours and then induced by serum by incubating in DMEM supplemented with 18% FBS and 1% P/S solution for 15 minutes. Control cells were serumstarved for 17.5 hours. Mouse ESCs, E14Tg2a, were obtained from Mutant Mouse Research Resource Centers and American Type Culture Collection. They were cultured on gelatin-coated plates in M15 medium: DMEM (Invitrogen) supplemented with 10% FBS, 10 μM 2-mercaptoethanol, 0.1 mM nonessential amino acids (Invitrogen), 1× EmbryoMax nucleosides (Millipore), 1,000 U of ESGRO (Millipore). To determine the knockdown efficiency of siRNAs, 50×10^3 ESCs were transfected with siRNAs at 50 nM in one well of a 24-well plate. Cells were harvested 48 hours after transfection and total RNA was extracted from the cells using manufacturer's protocol (Qiagen) for qRT-PCR and microarray experiment.

Chromatin Immunoprecipitation & Sequencing.

For ChIP-seq sample preparation, ESCs (2.5×10^6) were transfected with control or Trim28 siRNAs (SiRNA D-040800-02, si2; D-040800-03, si3, Thermo scientific) at 50 nM using lipofectamine 2000 (Invitrogen) in one 10 cm plate for each replicate. After 48 hours transfection, ESCs were harvested and processed according to the protocol in [56] for ChIP-seq analysis. ChIP experiment was conducted using the IP-star (Diagenode, Denville, NJ) and according to the manufacturer's protocol. Briefly, 5 million cells were immunoprecipitated using an antibody, sc-899 (Santa Cruz Biotechnology, 3 μg/each IP) against whole Pol II. Illumina libraries were constructed using the SPRI-TE Nucleic Acid Extractor (Beckman Coulter, Indianapolis, IN) with a size selection of 200-400bp. The samples were sequenced on a HiSeq sequencing platform and resulting short reads were aligned against the mm9 reference mouse genome. The antibodies listed above have been validated for the relevant species and applications, and the validation is provided on the manufacturers' websites.

For HEK293 cells, ChIP sample preparation was conducted following Abcam X-ChIP protocol with mild modifications. Cell lysis buffer includes 5 mM PIPES (pH 8.0), 85 mM KCl, 0.5% NP-40 and fresh protease inhibitors described above. Nuclei lysis buffer including 50 mM Tris-Cl (pH 8.0), 10 mM EDTA, 1% SDS was added before sonication. Sonication was performed at 12% amplitude for 30 sec with 2 mins intervals on ice, Sonic Dismembrator Model 500, Fisher Scientific and was optimized to produce DNA segments ranging between +100 and +1,000 bp on a DNA gel. Briefly, about 3.3 million cells were immunoprecipitated using an antibody, sc-899 (Santa Cruz Biotechnology, 3 μg/each IP) or A300-767A (Bethyl, 3 μg/each IP) against whole Pol II or phosphorylated TRIM28 at residue S824, respectively. Illumina libraries were constructed using Kapa Biosciences Illumina library prep kits according to the manufacturer's protocol. Samples were quantified by quantitative PCR and pooled at equal ratio before sequencing on the Illumina HiSeq2000 sequencing platform. The resulting short reads were aligned against the hg19 reference Human genome.

Bioinformatics: Pol II ChIP-seq. mESC:

Full details of mESC ChIP-seq analyses are found in [29]. Briefly, input and ChIP samples were filtered to exclude reads with mean quality score less than 20, aligned to mm9 using bowtie 0.12.8 retaining unique alignments with 2 or fewer mismatches, de-duplicated using Picard tools' MarkDuplicates, and center-shifted 75nt based on fragment length estimates obtained via HOMER 4.1. Raw data has been deposited to the Gene Expression Omnibus, under accession GSE48253. A set of 1,535 lncRNAs were used for all analyses, derived from UCSC knownGene annotations downloaded March 18th 2013, also present in NONCODE v4, identified using associated ensembl transcript identifiers. LncRNAs were identified following the exclusion of transcripts annotated to random chromosomes or chrM, as well as those shorter than 1000 nt, and the consolidation of transcripts with identical TSS. Also excluded were transcripts with less than 50% uniquely positioned TSS-proximal (+/− 250 nt) or gene body windows (TSS+500 to 2500 or end). We termed it as mappability, and over 50% mappability was defined as "mappable" in the text. Pausing indices were calculated as the depth-normalized, input-subtracted count of fragment centers per uniquely mappable nt in the TSS-proximal region divided by that of the gene body. Heatmaps were generated using Partek Genomics Suite 6.5, depicting normalized, input-subtracted fragment center counts in 50 nt bins, tiling the region −500 to +999 nt relative to the TSS, and sorted by WT normalized, input-subtracted counts in the TSS+250 to 999 region. Metagene analyses depict the per-lncRNA mean count of normalized, input-subtracted fragment centers in 50 nt bins. Pausing indices calculated based on mm9 annotations were compared to expression data derived from the mm10-based MTA 1.0 array via identical UCSC knownGene identifiers. **HEK293:** For details for ChipSeq analyses please refer to our previous paper [29]. Several replicates were included for each experimental and input control conditions. Fastq files of each sample or control was aligned to hg19 reference genome independently using the BWA (v0.5.9) with default parameters. Following alignment, reads aligning to the same genomic coordinates were removed (Picard MarkDuplicates), and filtered to retain only primary (unique) alignments. Gene annotation was referred to UCSC hg19 knownGene table downloaded in January 19, 2015. The downstream Pol II pausing analyses were performed as described above for mESC. The

combined input counts were subtracted from each window, and the windows with negative counts were set to zero. We further used Model-based Analysis for Chip-Seq software [57] to identify Pol II binding peaks. The windows outside of the significant peak regions detected by MACS were excluded from the final report.

GO analysis.

GO analysis was performed using Bioconductor's GOStats package [58]. We perform an unconditioned hypergeometric test for enrichment of GO terms between a group of interest versus a background set of genes. For the background set, we use all lnRNAs studied that have unique GO annotations (n=338 (mouse), 1,752 (human)) and we test for an over-representation of GO terms among the annotated genes where pausing index is larger or equal to 2.

Microarray.

For each sample, 200 ng of total RNA were amplified and labeled using Affymetrix Genechip WT Plus Reagent Kit (Catalog #902280; Affymetrix) according to the protocol provided by the supplier. Affymetrix Mouse Transcriptome 1.0 arrays were hybridized for 16 hours at 45C and washed and stained according to the protocol described in Affymetrix GeneChip Expression Analysis Manual (Fluidics protocol FS450-0001). Arrays were scanned using an Affymetrix Genechip Scanner 7G. MTA1.0 array quality control, normalization, and gene level expression calculation were performed by Affymetrix Expression Console Software. Differential expression statistics was calculated by Affymetrix Transcriptome Analysis Console software, comparing the normalized expression level between the three replicates in the TRIM28 knock down samples versus the three replicates in the wild type control samples. The genes with fold change greater than or equal 1.5x up regulated or down regulated and p values smaller than or equal to 0.05 are considered as differentially expressed. Differentially expressed non-conding RNA was further selected for downstream study. Hierarchical clustering of relative expression was performed using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/) following gene normalization. Heat map was created using Java Treeview [59].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement:

We appreciate S. K. Calderwood in Beth Israel Deaconess Medical Center & Harvard Medical School for discussions and reading the manuscript and P. J. Park and L. Jung in Brigham Women Hospital & Harvard Medical School for providing initial bioinformatics analysis and discussions. We are grateful to D. J. Taatjes in the University of Colorado, Boulder and to J. Rinn in the Broad Institute for their critical reading and helpful comments of the manuscript. H.B. is grateful to D. Bunch, Joy, R. Baker, M. Seaquist, K. Perkins, D. Verrengia, G. Hugenberger, and P.S.C. for their loving support and encouragement throughout the work. This study was supported by grants from the Harvard Joint Center for Radiation Therapy to H.B.

References:

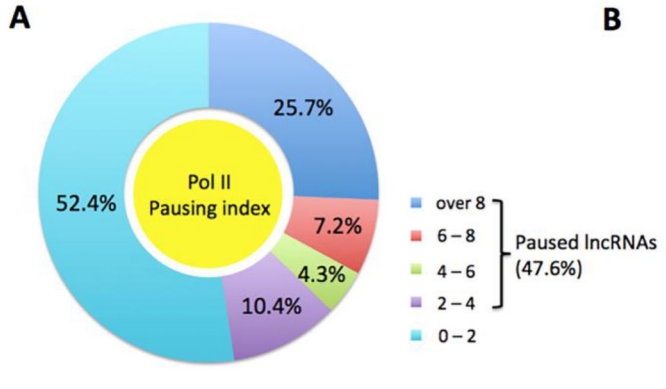
1. Bierhoff H, Dammert MA, Brocks D, Dambacher S, Schotta G, Grummt I: Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. Mol Cell 2014, 54:675–682. [PubMed: 24768537]
2. Lee JT, Bartolomei MS: X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 2013, 152:1308–1323. [PubMed: 23498939]
3. Sarachana T, Dahiya N, Simhadri VL, Pandey GS, Saini S, Guelcher C, Guerrera MF, Kimchi-Sarfaty C, Sauna ZE, Atreya CD: Small ncRNA Expression-Profiling of Blood from Hemophilia A Patients Identifies miR-1246 as a Potential Regulator of Factor 8 Gene. PLoS One 2015, 10:e0132433. [PubMed: 26176629]
4. Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, Chen LL: Long noncoding RNAs with snoRNA ends. Mol Cell 2012, 48:219–230. [PubMed: 22959273]
5. Carpenter S, Ricci EP, Mercier BC, Moore MJ, Fitzgerald KA: Post-transcriptional regulation of gene expression in innate immunity. Nat Rev Immunol 2014, 14:361–376. [PubMed: 24854588]
6. Abdelmohsen K, Panda AC, Kang MJ, Guo R, Kim J, Grammatikakis I, Yoon JH, Dudekula DB, Noh JH, Yang X, et al.: 7SL RNA represses p53 translation by competing with HuR. Nucleic Acids Res 2014.
7. Geisler S, Coller J: RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 2013, 14:699–712. [PubMed: 24105322]
8. Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG: A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science 2005, 309:1570–1573. [PubMed: 16141075]
9. Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, Molnar Z, Ponting CP: Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol 2010, 11:R72. [PubMed: 20624288]
10. Beaulieu YB, Kleinman CL, Landry-Voyer AM, Majewski J, Bachand F: Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. PLoS Genet 2012, 8:e1003078. [PubMed: 23166521]
11. Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R: Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res 2012, 22:1616–1625. [PubMed: 22955974]
12. Soreq L, Guffanti A, Salomonis N, Simchovitz A, Israel Z, Bergman H, Soreq H: Long non-coding RNA and alternative splicing modulations in Parkinson's leukocytes identified by RNA sequencing. PLoS Comput Biol 2014, 10:e1003517. [PubMed: 24651478]
13. Fu Y, Dominissini D, Rechavi G, He C: Gene expression regulation mediated through reversible m(6)A RNA methylation. Nat Rev Genet 2014, 15:293–306. [PubMed: 24662220]
14. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA: c-Myc regulates transcriptional pause release. Cell 2010, 141:432–445. [PubMed: 20434984]
15. Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K: Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science 2010, 327:335–338. [PubMed: 20007866]
16. Adelman K, Lis JT: Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet 2012, 13:720–731. [PubMed: 22986266]
17. Jonkers I, Kwak H, Lis JT: Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife 2014, 3:e02407. [PubMed: 24843027]
18. Core LJ, Lis JT: Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 2008, 319:1791–1792. [PubMed: 18369138]
19. Kwak H, Fuda NJ, Core LJ, Lis JT: Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science 2013, 339:950–953. [PubMed: 23430654]
20. Gaertner B, Johnston J, Chen K, Wallaschek N, Paulson A, Garruss AS, Gaudenz K, De Kumar B, Krumlauf R, Zeitlinger J: Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep 2012, 2:16701683.

21. Basnet H, Su XB, Tan Y, Meisenhelder J, Merkurjev D, Ohgi KA, Hunter T, Pillus L, Rosenfeld MG: Tyrosine phosphorylation of histone H2A by CK2 regulates transcriptional elongation. Nature 2014, 516:267–271. [PubMed: 25252977]
22. Wu CH, Yamaguchi Y, Benjamin LR, Horvat-Gordon M, Washinsky J, Enerly E, Larsson J, Lambertsson A, Handa H, Gilmour D: NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev 2003, 17:1402–1414. [PubMed: 12782658]
23. Gilchrist DA, Nechaev S, Lee C, Ghosh SK, Collins JB, Li L, Gilmour DS, Adelman K: NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev 2008, 22:1921–1933. [PubMed: 18628398]
24. Peterlin BM, Price DH: Controlling the elongation phase of transcription with P-TEFb. Mol Cell 2006, 23:297–305. [PubMed: 16885020]
25. Byun JS, Fufa TD, Wakano C, Fernandez A, Haggerty CM, Sung MH, Gardner K: ELL facilitates RNA polymerase II pause site entry and release. Nat Commun 2012, 3:633. [PubMed: 22252557]
26. Cheng B, Li T, Rahl PB, Adamson TE, Loudas NB, Guo J, Varzavand K, Cooper JJ, Hu X, Gnatt A, et al.: Functional association of Gdown1 with RNA polymerase II poised on human genes. Mol Cell 2012, 45:38–50. [PubMed: 22244331]
27. Allen BL, Taatjes DJ: The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol 2015, 16:155–166. [PubMed: 25693131]
28. Galbraith MD, Allen MA, Bensard CL, Wang X, Schwinn MK, Qin B, Long HW, Daniels DL, Hahn WC, Dowell RD, Espinosa JM: HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell 2013, 153:1327–1339. [PubMed: 23746844]
29. Bunch H, Zheng X, Burkholder A, Dillon ST, Motola S, Birrane G, Ebmeier CC, Levine S, Fargo D, Hu G, et al.: TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. Nat Struct Mol Biol 2014, 21:876–883. [PubMed: 25173174]
30. Bunch H, Lawney BP, Lin YF, Asaithamby A, Murshid A, Wang YE, Chen BP, Calderwood SK: Transcriptional elongation requires DNA break-induced signalling. Nat Commun 2015, 6:10191. [PubMed: 26671524]
31. Sanso M, Levin RS, Lipp JJ, Wang VY, Greifenberg AK, Quezada EM, Ali A, Ghosh A, Larochelle S, Rana TM, et al.: P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. Genes Dev 2016, 30:117–131. [PubMed: 26728557]
32. McNamara RP, Reeder JE, McMillan EA, Bacon CW, McCann JL, D'Orso I: KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Mol Cell 2016, 61:39–53. [PubMed: 26725010]
33. Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, Almada AE, Lin C, Sharp PA, Giallourakis CC, Young RA: Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc Natl Acad Sci U S A 2013, 110:2876–2881. [PubMed: 23382218]
34. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al.: The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012, 22:1775–1789. [PubMed: 22955988]
35. Core LJ, Waterfall JJ, Gilchrist DA, Fargo DC, Kwak H, Adelman K, Lis JT: Defining the status of RNA polymerase at promoters. Cell Rep 2012, 2:1025–1035. [PubMed: 23062713]
36. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA: Divergent transcription from active promoters. Science 2008, 322:1849–1851. [PubMed: 19056940]
37. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT: Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet 2014, 46:1311–1320. [PubMed: 25383968]
38. Maxwell CS, Kruesi WS, Core LJ, Kurhanewicz N, Waters CT, Lewarch CL, Antoshechkin I, Lis JT, Meyer BJ, Baugh LR: Pol II docking and pausing at growth and stress genes in C. elegans. Cell Rep 2014, 6:455–466. [PubMed: 24485661]

39. Gilchrist DA, Fromm G, dos Santos G, Pham LN, McDaniel IE, Burkholder A, Fargo DC, Adelman K: Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. Genes Dev 2012, 26:933–944. [PubMed: 22549956]
40. Washietl S, Kellis M, Garber M: Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Res 2014, 24:616–628. [PubMed: 24429298]
41. Ponting CP, Oliver PL, Reik W: Evolution and functions of long noncoding RNAs. Cell 2009, 136:629–641. [PubMed: 19239885]
42. Barde I, Rauwel B, Marin-Florez RM, Corsinotti A, Laurenti E, Verp S, Offner S, Marquis J, Kapopoulou A, Vanicek J, Trono D: [A KRAB/KAP1-miRNA cascade regulates erythropoiesis through stage-specific control of mitophagy]. Med Sci (Paris) 2014, 30:12–15. [PubMed: 24472449]
43. Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA: RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet 2007, 39:1512–1516. [PubMed: 17994019]
44. Wu JQ, Snyder M: RNA polymerase II stalling: loading at the start prepares genes for a sprint. Genome Biol 2008, 9:220. [PubMed: 18466645]
45. Chen F, Gao X, Shilatifard A: Stably paused genes revealed through inhibition of transcription initiation by the TFTTH inhibitor triptolide. Genes Dev 2015, 29:39–47. [PubMed: 25561494]
46. Fatica A, Bozzoni I: Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet 2014, 15:7–21. [PubMed: 24296535]
47. Palazzo AF, Gregory TR: The case for junk DNA. PLoS Genet 2014, 10:e1004351. [PubMed: 24809441]
48. Gaertner B, Zeitlinger J: RNA polymerase II pausing during development. Development 2014, 141:1179–1183. [PubMed: 24595285]
49. Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB: Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. Science 2012, 335:1499–1502. [PubMed: 22442485]
50. Hosoya T, Clifford M, Losson R, Tanabe O, Engel JD: TRIM28 is essential for erythroblast differentiation in the mouse. Blood 2013, 122:3798–3807. [PubMed: 24092935]
51. Hu G, Kim J, Xu Q, Leng Y, Orkin SH, Elledge SJ: A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev 2009, 23:837–848. [PubMed: 19339689]
52. Fasching L, Kapopoulou A, Sachdeva R, Petri R, Jonsson ME, Manne C, Turelli P, Jern P, Cammas F, Trono D, Jakobsson J: TRIM28 Represses Transcription of Endogenous Retroviruses in Neural Progenitor Cells. Cell Rep 2014.
53. Puc J, Kozbial P, Li W, Tan Y, Liu Z, Suter T, Ohgi KA, Zhang J, Aggarwal AK, Rosenfeld MG: Ligand-dependent enhancer activation regulated by topoisomerase-I activity. Cell 2015, 160:367–380. [PubMed: 25619691]
54. Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R: Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 2013, 494:497–501. [PubMed: 23417068]
55. Malik S, Roeder RG: Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci 2005, 30:256–263. [PubMed: 15896744]
56. Lee TI, Johnstone SE, Young RA: Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc 2006, 1:729–748. [PubMed: 17406303]
57. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008, 9:R137. [PubMed: 18798982]
58. Falcon S, Gentleman R: Using GOstats to test gene lists for GO term association. Bioinformatics 2007, 23:257–258. [PubMed: 17098774]
59. Saldanha AJ: Java Treeview--extensible visualization of microarray data. Bioinformatics 2004, 20:3246–3248. [PubMed: 15180930]

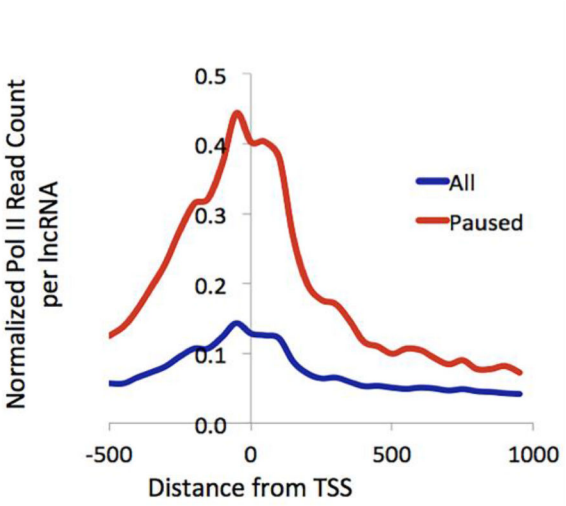
Highlights:

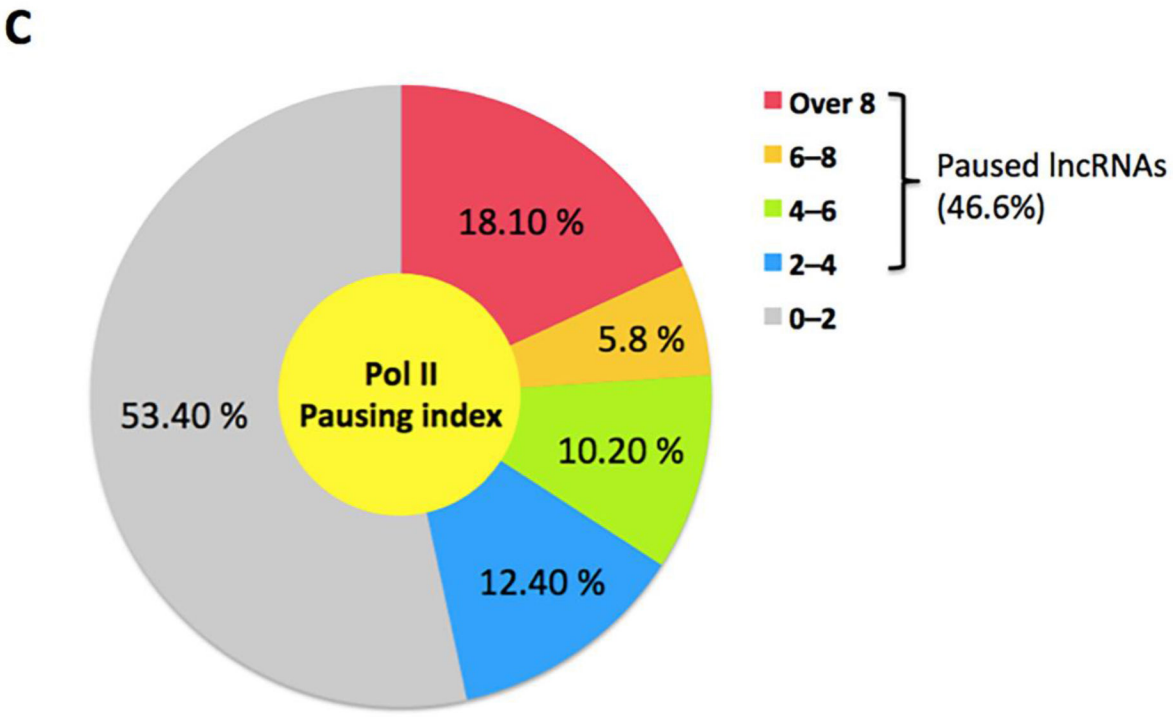
1. RNA polymerase II promoter-proximal pausing occurs prevalently at murine and human long noncoding RNA genes.
2. The statistic significance of RNA polymerase II promoter-proximal pausing in mammalian lncRNA genes is comparable with that of protein-coding genes.
3. A Pol II pausing-stabilizing factor in protein-coding genes, TRIM28 regulates transcription in lncRNAgenes.
4. A subset of serum inducible lncRNA genes in human cells are regulated by Pol II pausing.
5. Mammalian paused lncRNA genes are involved in stimulus-inducible biological pathways in spite of tissue-specific functions that they are involved in.



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

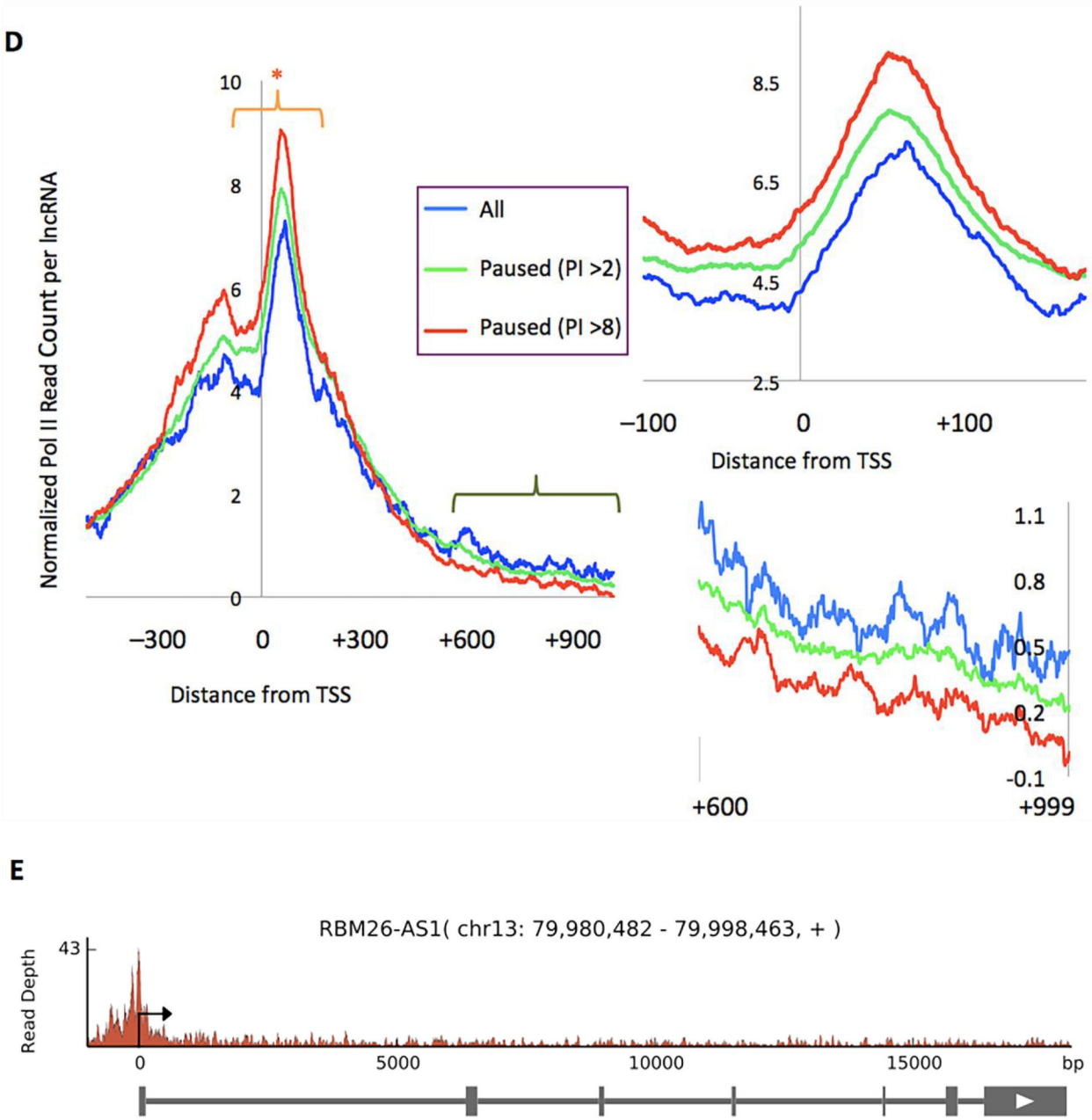
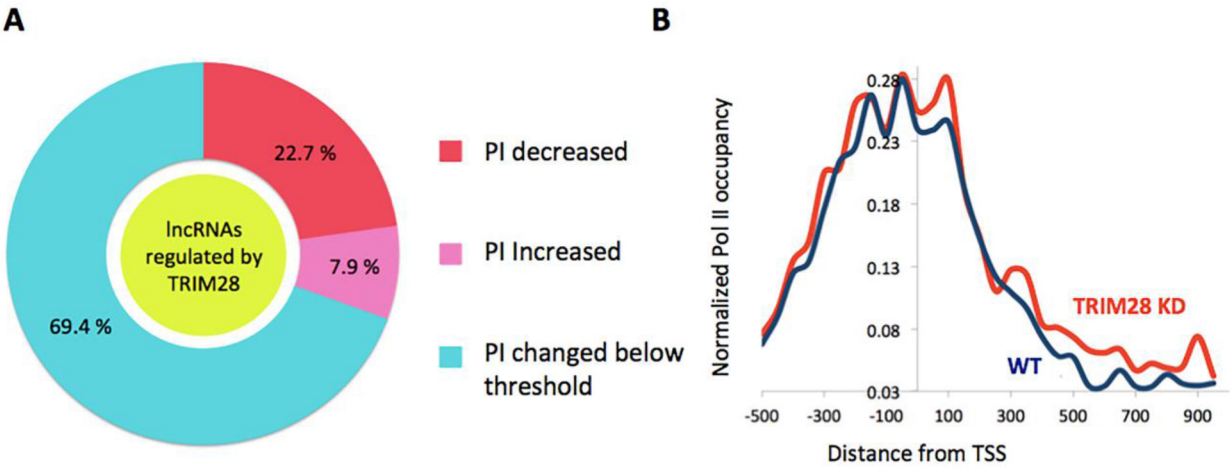
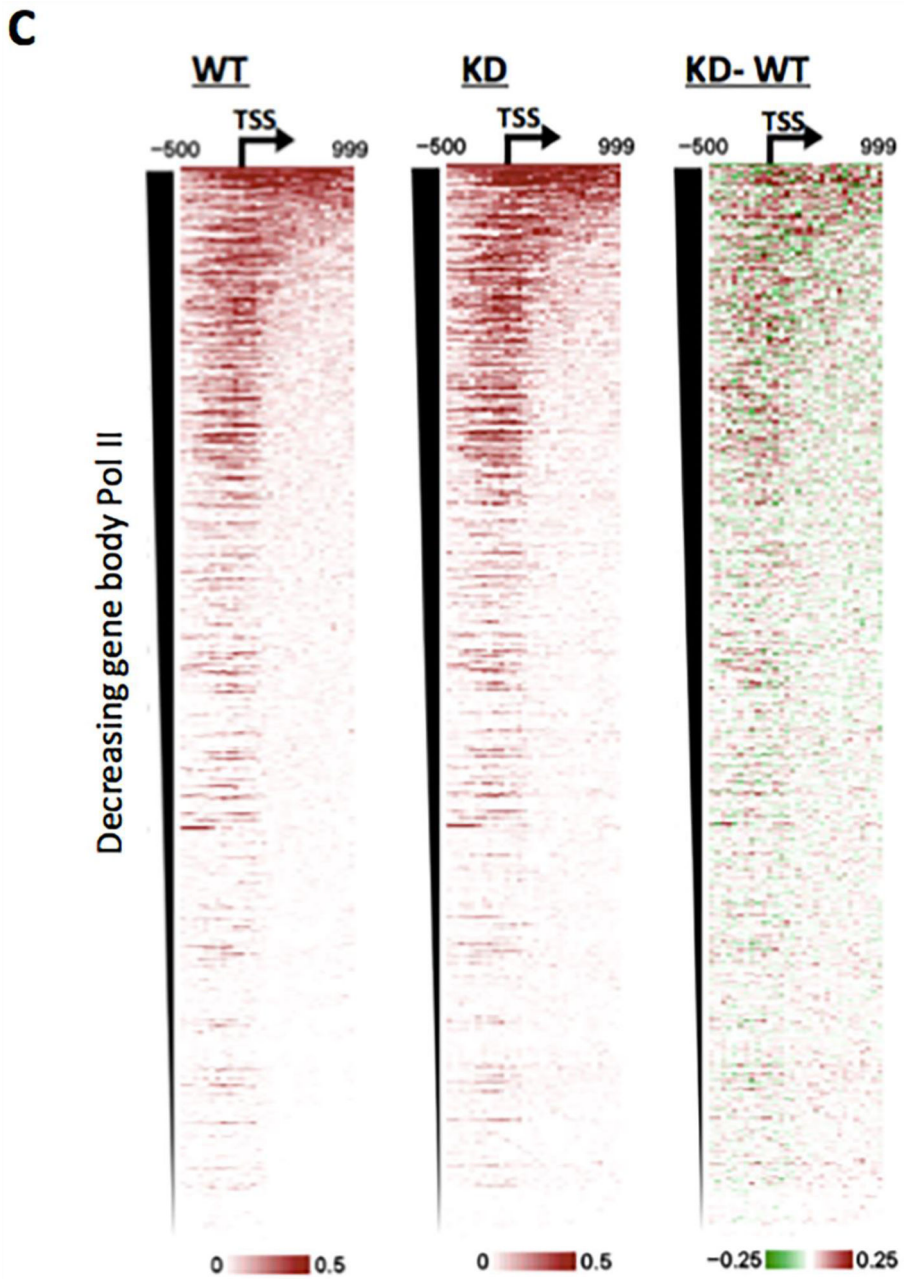


Fig. 1. Mammalian lncRNAs are regulated by Pol II promoter-proximal pausing.

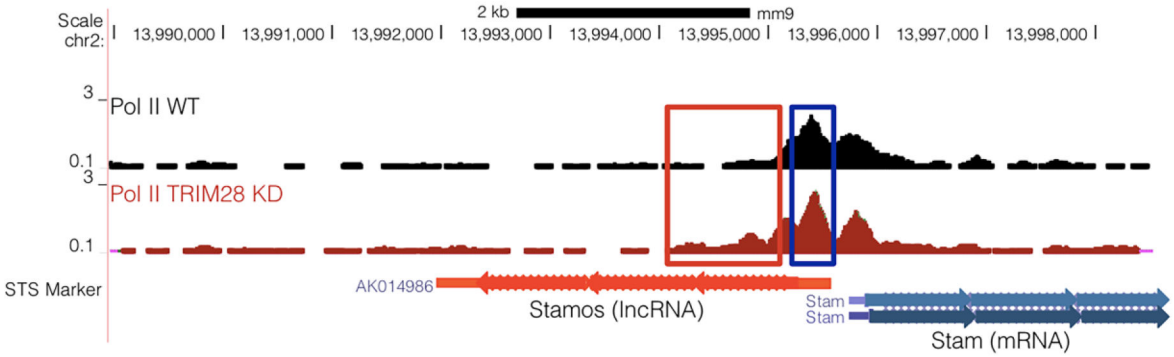
(A) Pausing indices of annotated lncRNAs (n= 1,535) in mouse embryonic stem (mES) cells. The graph was generated with pausing index calculable genes (non-zero gene body Pol II occupancy, n= 645). LncRNAs with pausing indices over 2 are 307 including 166 lncRNAs with pausing indices over 8. **(B)** Pol II metagene analysis comparing total (n= 1,535) 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= 2,482) 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= 1,752; 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.





D



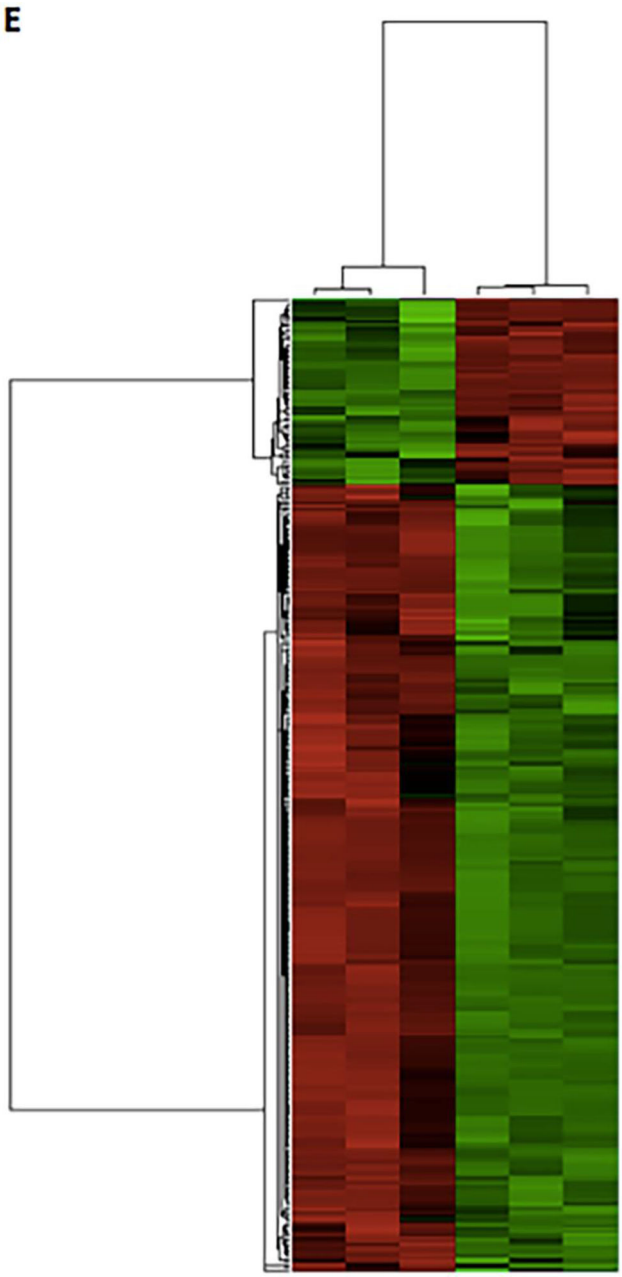
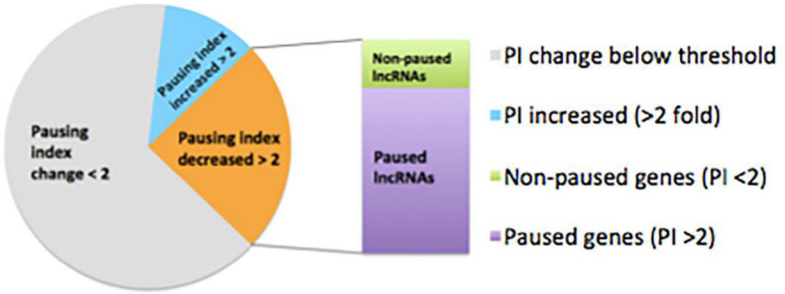


Fig. 2. TRIM28 KD regulates the expression of a number of lncRNAs in mES cells.

(A) LncRNAs categorized by pausing index change between WT and TRIM28 KD mES cells. Genes with pausing indices increased (PI increased, n= 24) or decreased (PI decreased, n= 69) over 2 fold, or changed less than 2 fold (PI change below threshold, n= 211) upon TRIM28 KD. **(B)** A metagene analysis of Pol II occupancy at lncRNAs with decreased pausing indices upon TRIM28 KD (n= 69, p = 6.009e-6). TRIM28 KD increased Pol II occupancy in the gene body (> +300 from TSS) at these genes. The p-value shown was calculated by a paired Wilcoxon signedrank test, using the sums of normalized, input-adjusted reads in the window between +300 and +999. **(C)** Heatmaps of Pol II occupancy at 1,535 lncRNAs in WT, TRIM28 KD (KD), and KD-WT, illustrating the overall change of

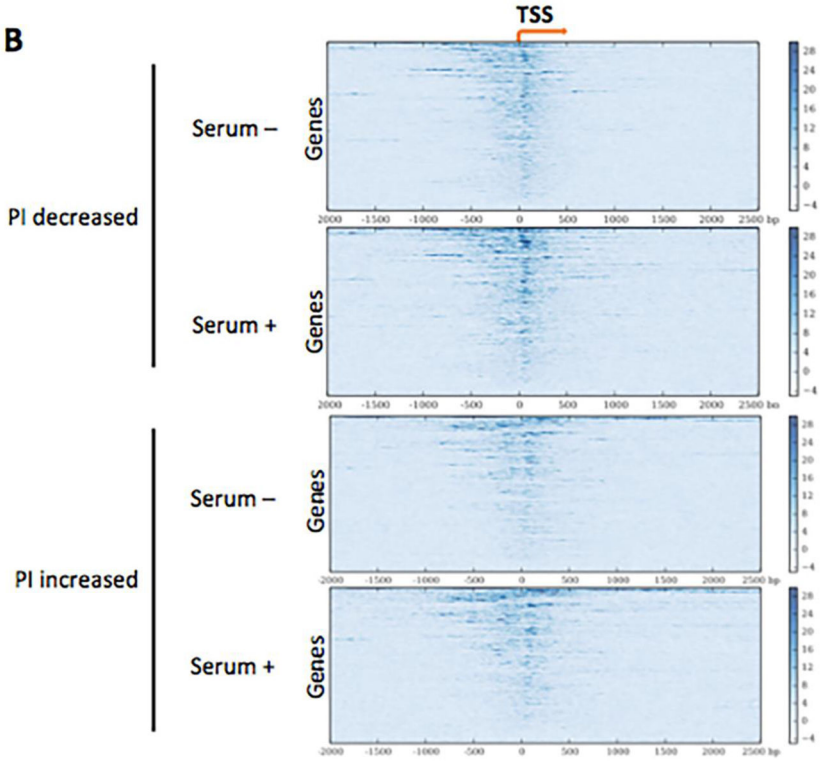
Pol II occupancy at these genes upon TRIM28 KD (p-value for TSS to +999, 1.413e–05). **(D)** A chromosome view of *Stamos* (*AK014986*) depicting increased Pol II occupancy in the downstream of Pol II pausing site (marked with red box) upon TRIM28 KD. **(E)** Hierarchical clustering comparing gene expression profiles (n= 308; fold change > 1.5; $P < 0.05$) between TRIM28 KD (the first set of triplicates from the left) and WT (the second set of triplicates) mES cells. Red, increased expression; Green, decreased expression; Black, no change.

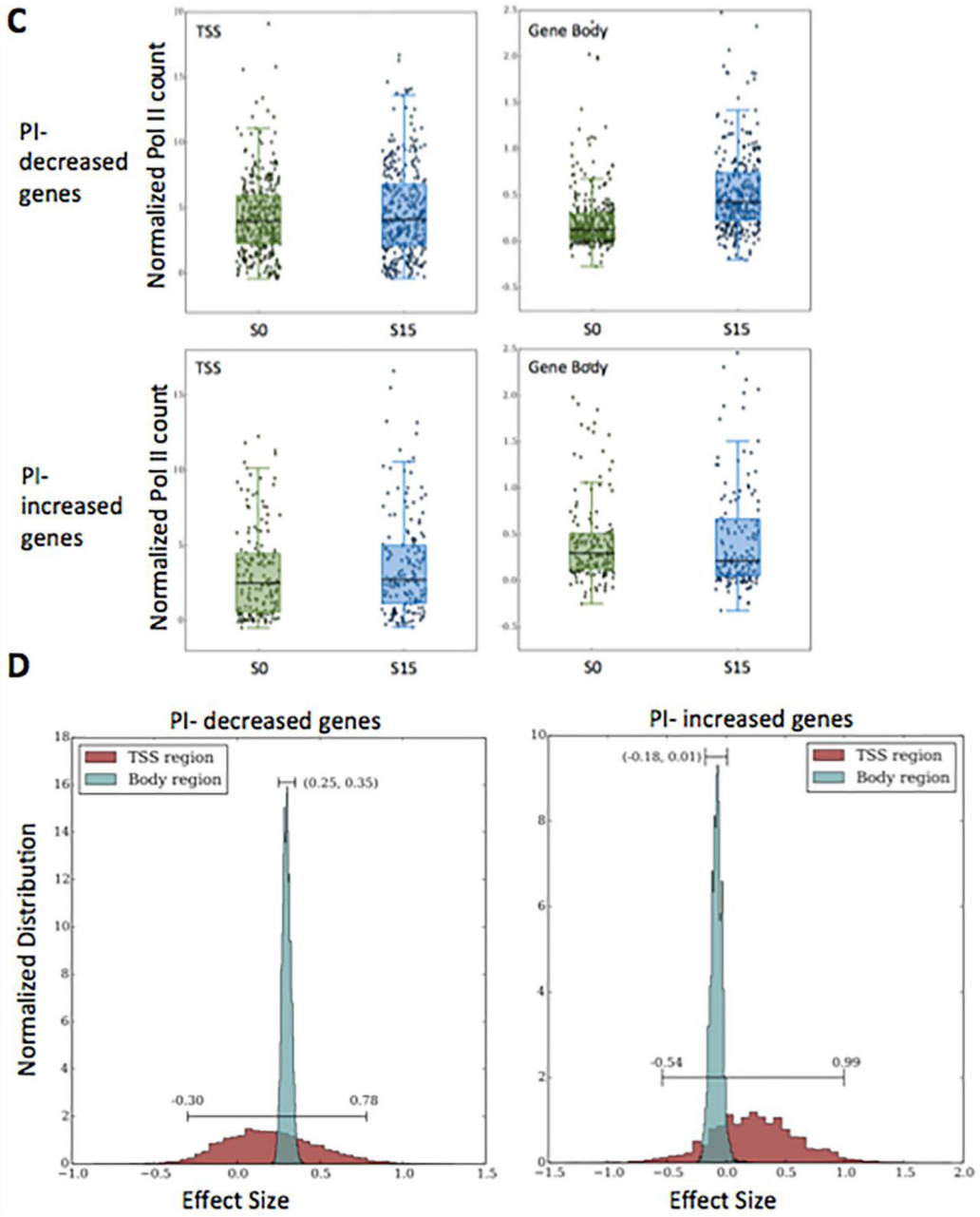
A



	Total lncRNA	PI available (non zero in both S0 and S15)	PI change below threshold	PI decreased in S15 (< 2 fold)	PI increased in S15 (> 2 fold)
Gene (n)	2491	1990	1290	483	217

B





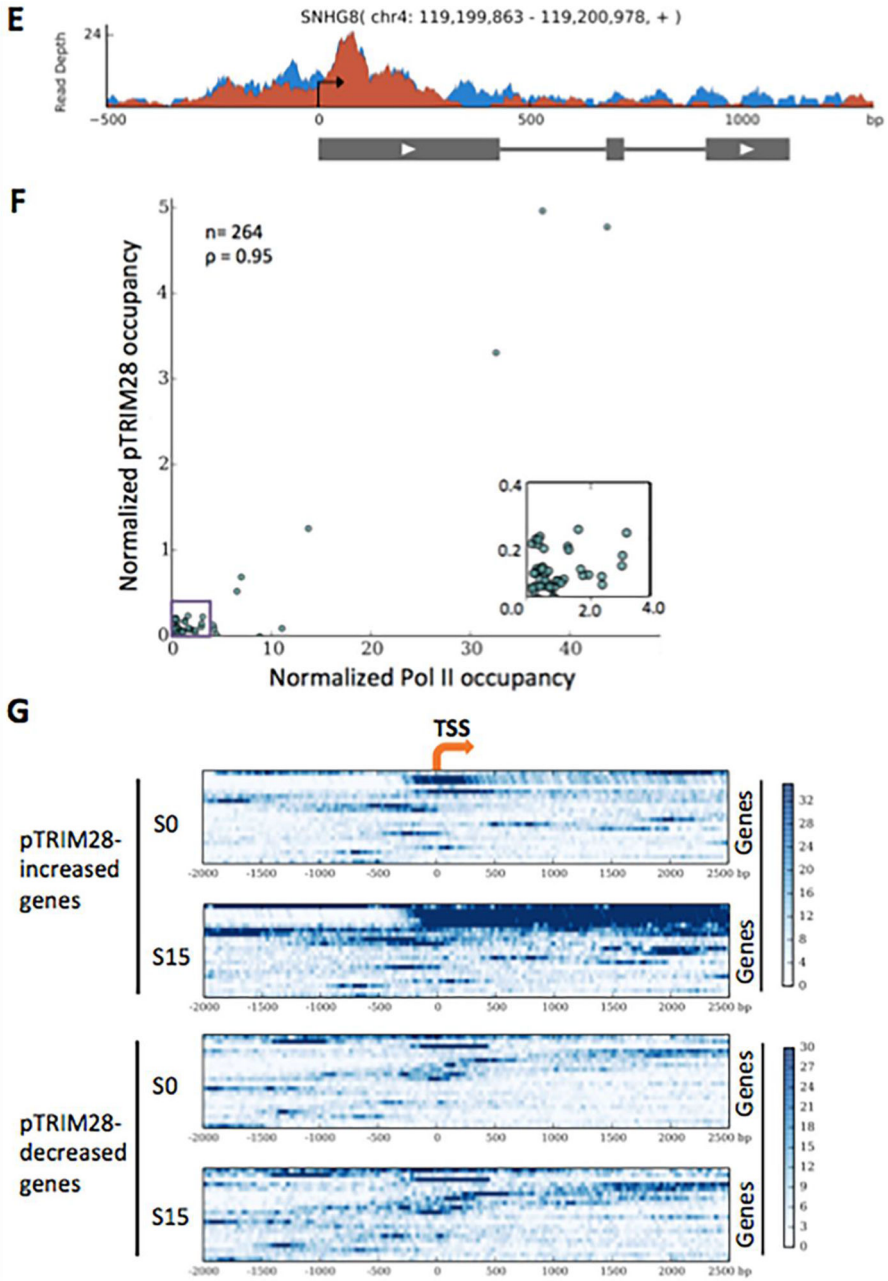


Fig. 3. Stimulus-inducible long non-coding genes are regulated by Pol II promoter-proximal pausing in mammalian cells.

(A) Pausing index changes of human lncRNAs between control and serum induced HEK293 cells. Pausing index decreased genes tend to be more paused, suggesting that serum-induced lncRNAs are paused as in coding genes. PI, pausing index; S0, control, non-serum induced HEK293 cells; S15, serum-induced HEK293 cells. **(B)** Pol II heatmaps of pausing index (PI) decreased and increased long non-coding genes. Pausing index-decreased genes displayed a Pol II peak in a defined locus near TSS (Transcription Start Site) in control cells (Serum −). Pausing index increased genes showed weak Pol II TSS peaks in a broad area, implying less Pol II pausing. Upon serum induction (Serum +), Pol II occupancy was increased in TSS and

downstream (toward gene body) in pausing index-decreased genes while the Pol II occupancy increase was observed in TSS and upstream (toward promoters) in pausing index-increased genes. **(C)** Box plots showing Pol II occupancy changes in TSSs and gene bodies in pausing index (PI) decreased and increased genes. TSS Pol II becomes slightly increased in both groups of genes while gene body Pol II is increased and decreased in pausing index decreased and increased genes, respectively. **(D)** A bootstrap resampling procedure. For each bootstrap sample (1e5 total), the effect size was computed as the difference between the median value of the samples in control and serum induced states (for both TSS and body regions). A positive value indicates an increase in pol-II occupancy. The normalized empirical distribution was plotted with horizontal bars indicating the 2.5% and 97.5% percentiles. **(E)** A chromosome view of *SNHG8*, representing increased Pol II occupancy in the gene body upon serum induction. Red and blue peaks are in control cells and serum-induced cells, respectively. **(F)** Correlation analysis between TRIM28 phosphorylation at S824 (pTRIM28) and Pol II occupancy, showing that a subset of long non-coding genes (n= 264) accumulated Pol II and pTRIM28 simultaneously upon serum induction. The long non-coding genes with increased pTRIM28 in serum-induced cells were collected and correlated with Pol II occupancy. The inset on right is a zoomed view for the boxed region in the main plot. Pearson's correlation coefficient = 0.95. **(G)** Pol II heatmaps of pTRIM28 increased or decreased genes upon serum induction (S15), showing a simultaneous increase of Pol II occupancy and pTRIM28 at some long noncoding genes (n= 20). For statistic values, normalized, mean Pol II occupancy was calculated for each gene in the window of interest (−2,000, +2,500 bp). Both pTRIM28 increased and decreased genes demonstrated a difference in Pol II occupancy ($p < 0.001$ by Wilcoxon signedrank test) and pTRIM28 increased genes showed greater changes in Pol II occupancy following induction ($p = 0.01$ by Mann-Whitney U test).

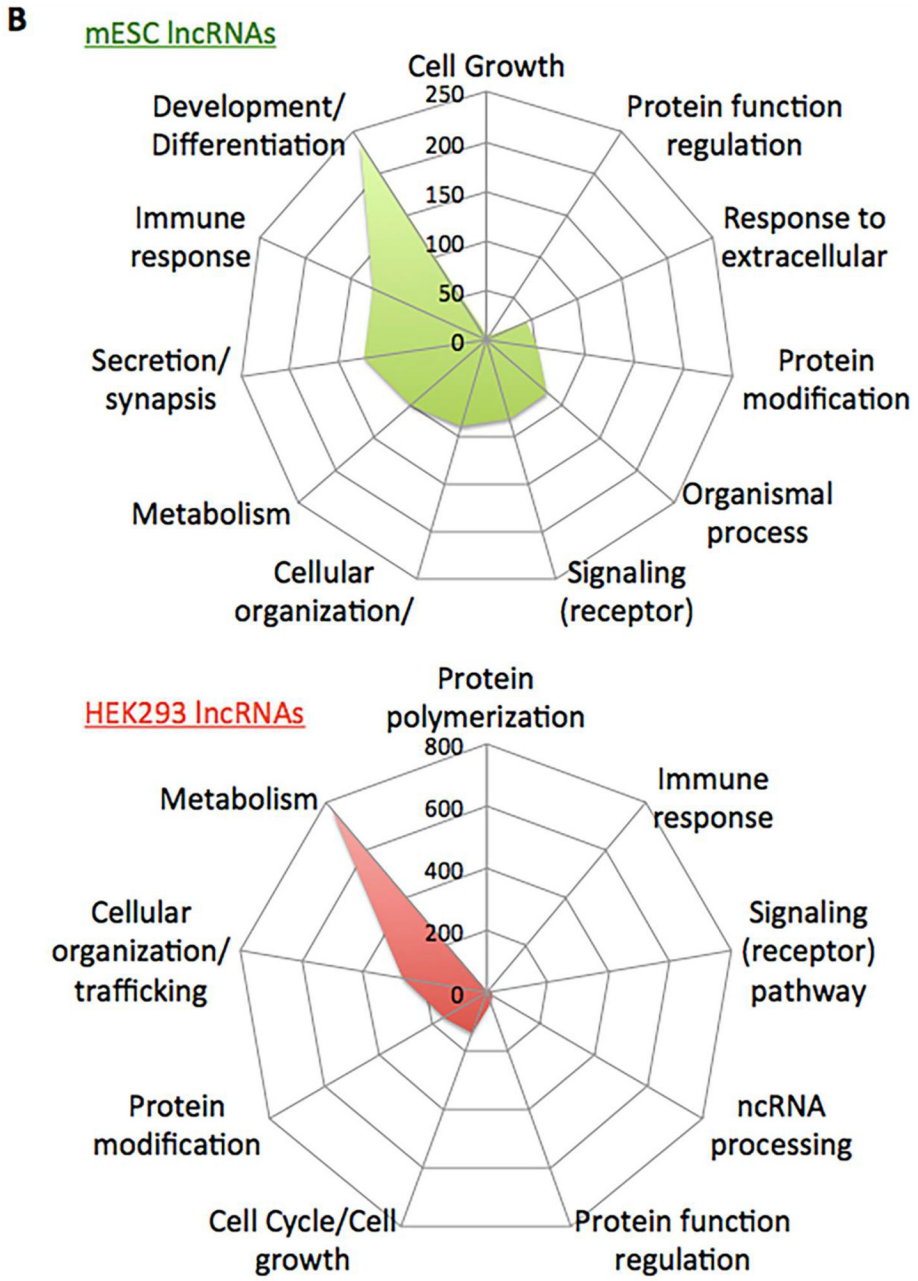
A

Mouse lncRNAs

Biological Function	Gene counts
Cell Growth	6
Protein function regulation	8
Response to extracellular chemical substances	45
Protein modification	52
Organismal process	81
Signaling (receptor) pathway	82
Cellular organization/ trafficking	90
Metabolism	100
Secretion/synapsis	124
Immune response	126
Development/ Differentiation	236
Others	266

Human lncRNAs

Biological Function	Gene counts
Protein polymerization	8
Immune response	17
Signaling (receptor) pathway	22
ncRNA processing	23
Protein function regulation	34
Cell Cycle/Cell growth	136
Protein modification	156
Cellular organization/ trafficking	269
Metabolism	771



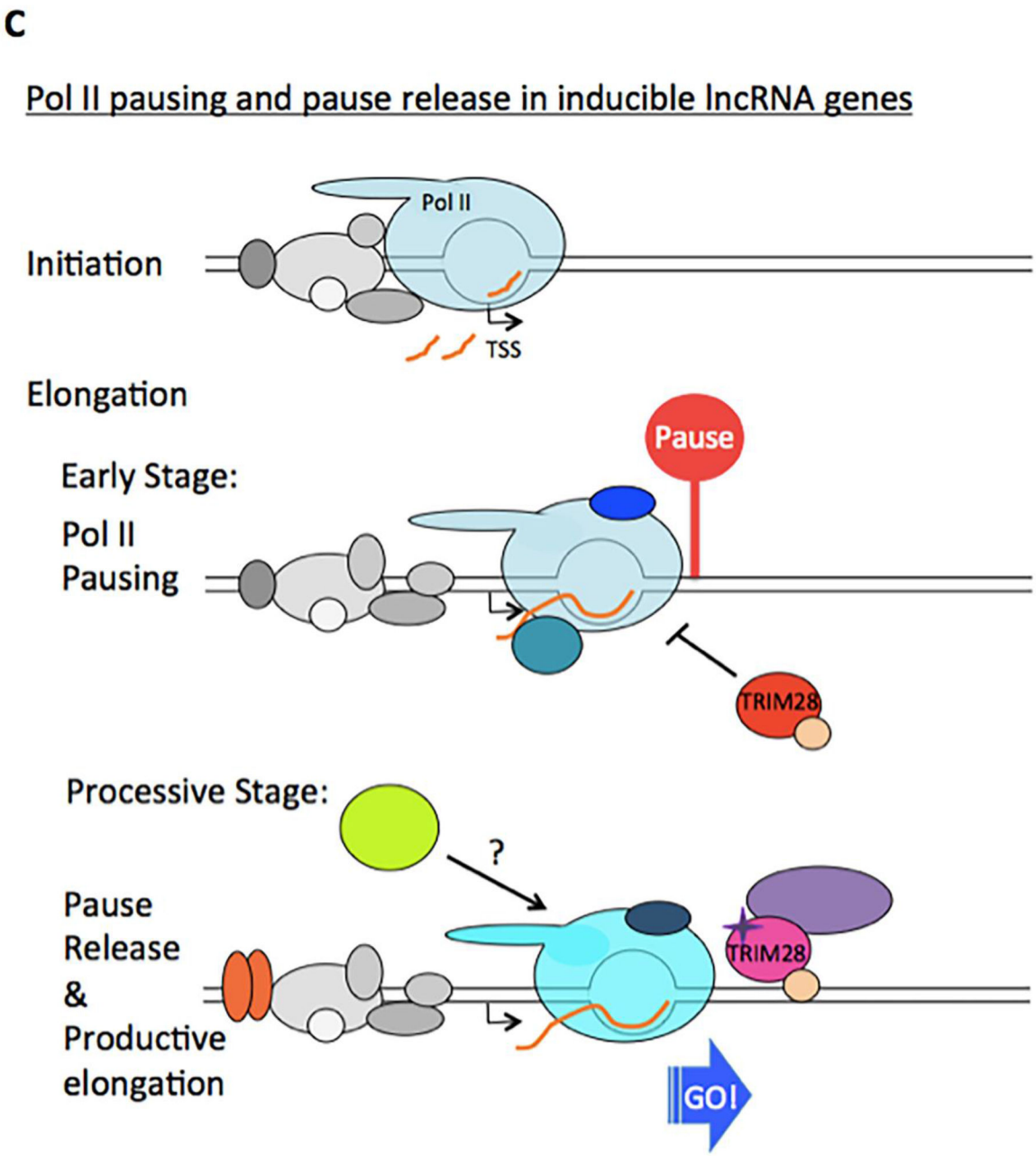


Fig. 4. Paused lncRNA genes are involved in stimulus-inducible biological pathways.

(A) GO analysis of murine and human long non-coding genes, showing common and unique biological pathways. Stimulus-inducible pathways are in *italic*. **(B)** Radar graphs showing the major biological pathways for lncRNAs in human HEK293 cells (upper) and in mES cells (bottom), implying tissue-specific functions of lncRNAs. In spite of tissue-specific functions, paused long non-coding genes mainly regulate stimulus-inducible pathways. Metabolism appears to be regulated by many lncRNAs in both, especially in human HEK293 cells and a number of mES lncRNAs are involved in development/differentiation. **(C)** A model of RNA polymerase II promoter-proximal pausing in inducible lncRNA genes. In the uninduced state of transcription, Pol II is paused in the promoter-proximal site of

lncRNA genes. TRIM28 stabilizes Pol II pausing in a subset of these lncRNA genes. Upon transcriptional activation, Pol II is released from the pausing site, which involves phosphorylation of TRIM28 at S824 (marked with a purple star).

Table 1.

Long non-coding RNAs with increased or decreased pausing indices upon TRIM28 KD in mES cells

Gene	Known function	Pausing index, WT	Pausing index, KD	Pausing index, % of WT
1700012B15Rik	RAB10 opposite strand	604.61	27.45	4.54
4930581F22Rik		98.66	6.50	6.59
AK015092		124.33	8.80	7.08
AK015714		876.61	62.57	7.14
AK040705		19.89	1.65	8.31
B230206F22Rik	Ftx, a positive regulator for XIST	83.89	8.65	10.31
AK131739		53.29	5.79	10.86
1700003G18Rik		389.80	51.08	13.10
AK042797		17.25	2.35	13.61
4930465K10Rik		176.44	24.21	13.72
4933407K13Rik		98.86	15.31	15.48
AK079957		86.87	16.15	18.60
AK135254		44.78	8.55	19.08
AK005608		28.19	5.45	19.32
AK014986	Signal transduction adaptor molecule 1 opposite strand	54.05	12.06	22.32
Nphs1as	Nephrosis 1 homolog opposite strand	3.57	1.71	47.96
4632419I22Rik	Brip1 opposite strand	14.23	35.53	249.68
F730043M19Rik		20.28	128.24	632.40
Trmt61b	Trmt61b tRNA methyltransferase 61B	5.19	60.05	1156.04
1700020I14Rik		12.34	241.35	1956.12
AK162965		15.69	356.57	2272.69
4930483J18Rik		74.36	2420.54	3255.31

Table 2.

Non-coding RNAs with increased or decreased expression upon TRIM28 KD in mES cells

Gene Symbol	Fold Change	Gene Description
C330046G13Rik	−2.77	RIKEN cDNA
Gm14074	−2.74	Predicted gene 14074
Mir290	−2.62	MicroRNA 290
Pramel7	−2.5	Preferentially expressed antigen in melanoma like 7
Mir291a; Mir292; Mir291b; Mir293; Mir294; Mir295	−2.35	MicroRNA 291a; microRNA 292; microRNA 291b; microRNA 293; microRNA 294; microRNA 295
Traj1	−2.04	T cell receptor alpha joining 1
FTX_2	1.53	FTX transcript, XIST regulator conserved region 2
Rmst	1.56	Rhabdomyosarcoma 2 associated transcript
Mir302a, Mir302c	1.57	MicroRNA 302a and 302c
Mir205	1.73	MicroRNA 205
LOC100041014	2.32	Sycp3 like Y-linked pseudogene
LOC100041346	2.34	Sycp3 like Y-linked pseudogene
LOC380994	2.35	Sycp3 like Y-linked related
Gm21627	2.36	Predicted gene, 21627
Gm21294	2.37	Predicted gene, 21294
Gm5773	2.41	Predicted pseudogene 5773
Gm9141	2.42	Predicted gene 9141
Gm12865	2.44	Predicted gene 12865
Gm21815	2.47	Predicted gene, 21815
Gm11544	2.47	Predicted gene 11544
Gm11546	2.48	Predicted gene 11546
Gm20871	2.5	Sycp3 like Y-linked pseudogene
Gm13354	2.53	Predicted gene 13354
Trgj1	2.56	T cell receptor gamma joining 1
Gm13988	2.57	Predicted gene 13988
Gm11331	2.58	Predicted gene 11331
Gm14219	2.58	Predicted gene 14219
Gm10807	2.59	Predicted gene 10807
Gm26040	2.63	Predicted gene, 26040
Gm6507	2.63	Predicted gene 6507
Fam171b	2.71	Family with sequence similarity 171, member B
Gm14944	2.73	Predicted gene 14944
Gm14742	2.95	Predicted gene 14742
Gm13106	3.07	Predicted gene 13106
Olfr1372-ps1	3.17	Olfactory receptor 1372
Gm25163	3.44	Predicted gene 25163

Gene Symbol	Fold Change	Gene Description
Gm11543	3.48	Predicted gene 11543
2410024N13Rik	3.86	RIKEN cDNA 2410024N13
Gm9125	4.67	Predicted gene 9125
Gm4982	5.2	Predicted gene 4982

Table 3.

Long non-coding RNAs with pausing index changes over 3 fold upon serum induction in human cells

Gene Symbol	Pausing index Uninduced	Pausing index Induced	Pausing index ratio	Gene Description
BTN2A3P	7.05	0.09	82.35	Butyrophilin, subfamily 2
LINC00265	41.87	1.08	38.69	lncRNA 265
MIR7-3HG	44.89	1.52	29.55	MIR7-3 host gene
MAFG-AS1	46.57	1.81	25.76	MAFG antisense RNA 1
LOC100506207	91.45	3.88	23.54	Uncharacterized
LOC100506804	183.51	8.73	21.01	Uncharacterized
C4orf29	111.51	5.40	20.66	Chromosome 4 ORF 29
OSER1-AS1	112.41	6.05	18.58	OSER1 antisense RNA 1
STAG3L4	99.55	7.18	13.87	Stromal antigen 3-like 4
NBR2	18.59	1.69	11.02	Neighbor of BRCA1 gene 2
LOC389765	16.31	2.01	8.10	Kinesin family member 27
GUSBP1	121.40	15.31	7.93	Glucuronidase, beta pseudogene 1
INO80B-WBP1	2.02	0.26	7.88	INO80B-WBP1 readthrough
FLJ31104	33.65	5.00	6.74	LOC441072, uncharacterized
CASC2	293.94	63.79	4.61	Cancer susceptibility candidate 2
FOXN3-AS1	3.22	0.73	4.41	FOXN3 antisense RNA 1
SNHG8	11.44	2.73	4.19	snRNA host gene 8
CECR5-AS1	17.12	4.37	3.92	CECR5 antisense RNA 1
PITPNA-AS1	8.59	2.33	3.69	PITPNA antisense RNA 1
TRAM2-AS1	13.74	3.85	3.57	TRAM2 antisense RNA 1
LOC100131060	11.26	3.27	3.44	lncRNA 1135
LOC643837	27.77	8.17	3.40	lncRNA 1128
LOC100506190	2.68	0.80	3.37	lncRNA 963
AL133568	7.93	2.38	3.33	CYP51A1 antisense RNA 1
LOC150776	6.09	1.91	3.18	Sphingomyelin phosphodiesterase 4
TUG1	4.52	1.43	3.17	Taurine up-regulated 1
LOC645513	16.33	5.23	3.12	Uncharacterized
DPH6-AS1	5.64	1.84	3.07	DPH6 antisense RNA 1
UBL7-AS1	155.84	50.89	3.06	UBL7 antisense RNA 1
LOC441204	2.79	465.19	0.01	Uncharacterized
LINC00116	3.91	70.41	0.06	lncRNA 116
LOC399829	2.03	18.23	0.11	lncRNA 1168
AK056368	4.70	39.85	0.12	Uncharacterized
AK055623	8.89	73.22	0.12	BST2 interferon stimulated positive regulator
LOC256880	36.29	208.31	0.17	Uncharacterized
LINC00240	11.68	64.57	0.18	lncRNA 240

Gene Symbol	Pausing index Uninduced	Pausing index Induced	Pausing index ratio	Gene Description
ZNF503-AS2	1.10	4.12	0.27	ZNF503 antisense RNA 2
LOC100289509	3.39	11.93	0.28	KCNIP2 antisense RNA 1
LOC645166	0.34	1.19	0.29	Lymphocyte-specific protein 1 pseudogene