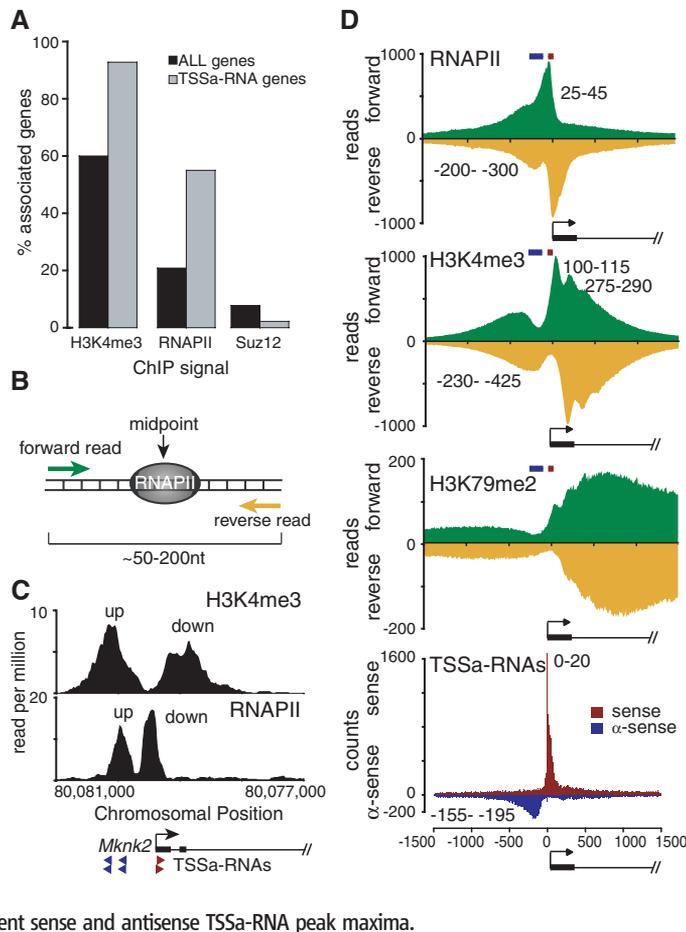


Fig. 4. Relationship between TSSa-RNAs, chromatin structure, and RNAPII. **(A)** Percentage of genes associated with H3K4me3, RNAPII and Suz12. ALL, all mouse genes. *t* test gives $P < 2.2 \times 10^{-16}$ for all marks. **(B)** Schematic of factor binding site mapping using Chip-seq reads. **(C)** Chromosomal position versus enrichment ratio for H3K4me3-modified nucleosomes and RNAPII for representative gene *Mknk2*. TSSa-RNAs are shown as arrowheads. **(D)** Metagene profiles for forward (green) and reverse (yellow) reads for ChIP-seq data (first three panels) and TSSa-RNAs from the sense (red) and antisense (blue) strand (bottom panel). Panels are aligned at the TSS. The TSS is denoted by the arrow. Black numbers on the profiles define the midpoint between forward and reverse peaks. Red and blue bars above the ChIP-seq profiles represent sense and antisense TSSa-RNA peak maxima.



that the sense complex may promote antisense preinitiation complex formation in the upstream region (16). Second, as RNAPII elongates the sense transcript, negative supercoiling of the

will occur upstream, perhaps promoting the antisense initiation process (17). This divergent transcription could structure chromatin and nascent RNA at the TSS for subsequent regulation.

RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters

Pascal Preker,¹ Jesper Nielsen,² Susanne Kammler,^{1*} Søren Lykke-Andersen,¹ Marianne S. Christensen,¹ Christophe K. Mapendano,¹ Mikkel H. Schierup,² Torben Heick Jensen^{1†}

Studies have shown that the bulk of eukaryotic genomes is transcribed. Transcriptome maps are frequently updated, but low-abundant transcripts have probably gone unnoticed. To eliminate RNA degradation, we depleted the exonucleolytic RNA exosome from human cells and then subjected the RNA to tiling microarray analysis. This revealed a class of short, polyadenylated and highly unstable RNAs. These promoter upstream transcripts (PROMPTs) are produced ~0.5 to 2.5 kilobases upstream of active transcription start sites. PROMPT transcription occurs in both sense and antisense directions with respect to the downstream gene. In addition, it requires the presence of the gene promoter and is positively correlated with gene activity. We propose that PROMPT transcription is a common characteristic of RNA polymerase II (RNAPII) transcribed genes with a possible regulatory potential.

Recent high-throughput analyses have revealed that >90% of all human DNA is transcribed (1). The vast majority of these

transcripts are noncoding, thus challenging the classical definition of what constitutes a gene and, by association, a promoter (2–4). Further-

References and Notes

- G. Orphanides, D. Reinberg, *Cell* **108**, 439 (2002).
- J. M. Calabrese, A. C. Seila, G. W. Yeo, P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18097 (2007).
- A. Marson *et al.*, *Cell* **134**, 521 (2008).
- Materials and methods are available as supporting material on Science Online.
- P. Carninci *et al.*, *Science* **309**, 1559 (2005).
- L. A. Boyer *et al.*, *Nature* **441**, 349 (2006).
- T. I. Lee *et al.*, *Cell* **125**, 301 (2006).
- M. Sultan *et al.*, *Science* **321**, 956 (2008).
- L. J. Core, J. J. Waterfall, J. T. Lis, *Science* **322**, 1845 (2008); published online 4 December 2008 (10.1126/science.1162228).
- A. Saunders, L. J. Core, J. T. Lis, *Nat. Rev. Mol. Cell Biol.* **7**, 557 (2006).
- D. A. Gilchrist *et al.*, *Genes Dev.* **22**, 1921 (2008).
- M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, R. A. Young, *Cell* **130**, 77 (2007).
- A. Barski *et al.*, *Cell* **129**, 823 (2007).
- T. S. Mikkelsen *et al.*, *Nature* **448**, 553 (2007).
- A. R. Kays, A. Schepartz, *Chem. Biol.* **7**, 601 (2000).
- H. Xiao, J. T. Lis, H. Xiao, J. Greenblatt, J. D. Friesen, *Nucleic Acids Res.* **22**, 1966 (1994).
- L. F. Liu, J. C. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7024 (1987).
- We thank G. Zheng, C. Whittaker, S. Hoersch, and A. F. Seila. A.C.S. was supported by NIH postdoctoral fellowship 5-F32-HD051190 and G.W.Y. by the Crick-Jacobs Center for Computational Biology. This work was supported by NIH grants RO1-GM34277 and HG002668, NCI grant P01-CA42063, and the NCI Cancer Center Support (core) grant P30-CA14051. The data discussed in this publication have been deposited in National Center for Biotechnology Information's Gene Expression Omnibus under accession numbers GSE13483 and GSE12680.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1162253/DC1
Materials and Methods

Figs. S1 to S7

Tables S1 to S5

Data Files

References

24 June 2008; accepted 12 November 2008

Published online 4 December 2008;

10.1126/science.1162253

Include this information when citing this paper.

more, additional short-lived RNAs might have escaped detection. With the aim of identifying such transcripts, we used RNA interference in HeLa cells to deplete hRrp40, a core component of the human 3' to 5' exoribonucleolytic exosome, one of the major RNA degradation complexes (fig. S1A) (5). This resulted in a severe processing defect of the known exosome substrate 5.8S ribosomal RNA (fig. S1B), demonstrating diminished exosome function. Oligo dT-primed, double-stranded cDNA from cells that had been treated with either a control [enhanced green fluorescent protein (eGFP)] or hRrp40 small interfering RNA (siRNA) was hybridized to an encyclopedia of DNA elements (ENCODE) tiling array, which covers a representative ~1% of the human genome (1). Comparison of array data to public gene annotations revealed overall stabilization of mRNAs (exons in Fig. 1A), as expected. RNA from intronic and intergenic regions were largely unaffected, with the exception of a 1.5-kb region immediately upstream of transcription start sites (TSSs) that was stabilized ~1.5-fold on average (Fig. 1A). The relative stabilization of

RNA expressed from a 500-kb region exemplifies this: Four of the five genes in this region display peaks of stabilized RNA upstream of their annotated promoters (Fig. 1B).

To validate these results, we subjected RNA from exosome-depleted versus control cells to oligo dT-primed reverse transcription followed by quantitative polymerase chain reaction (RT-qPCR) analyses of a region upstream of 20 TSSs, all of which confirmed a statistically significant stabilization under hRrp40 knockdown conditions (Fig. 1C and fig. S2A). Depletion of an additional exosome component (hRrp46) resulted in similar levels of stabilization, whereas depletion of other factors involved in RNA turnover (hUpf1, hXrm1, hXrm2, hDcp2, PARN) had no effect (fig. S2B), indicating that promoter upstream transcripts (PROMPTs) are exosome-specific targets. Individual depletion of hRrp6 or hRrp44, the catalytically active exosome subunits, resulted in no or only modest stabilization. Depletion of both, however, caused levels of stabilization comparable to that observed upon depletion of hRrp40 (Fig. 1C and fig. S2A), suggesting that hRrp6 and hRrp44 act redundantly to degrade PROMPTs. This stabilization of PROMPTs in exosome-depleted cells is reminiscent of that of

Saccharomyces cerevisiae cryptic unstable transcripts that, like PROMPTs, are also transcribed from nongenic regions (6).

To overview the average RNA stabilization profile around all 1594 annotated ENCODE TSSs, we aligned array data from the hRrp40 and control knockdown experiments, as well as the ratio of the two, relative to each other (Fig. 2A, top). Because of the different levels of stabilization of exonic and intronic RNA (Fig. 1A), we only considered data derived from exonic sequences downstream of the TSSs (fig. S3). Moreover, because many genes have multiple TSS clusters (i.e., promoters) that may confound analyses, we also aligned array data from 64 selected genes with only one major TSS cluster (low-complexity genes) (Fig. 2A, bottom, and table S1). Both alignments revealed an average RNA stabilization profile over a ~2-kb region upstream of the TSS with a peak around -1 kb (Fig. 2A). In control cells, RNA levels are near background, whereas they are greatly elevated upon hRrp40 depletion. RNA levels in the hRrp40-depleted cells drop to background levels nearing the TSS, indicating that stabilized transcripts are distinct from their neighboring mRNAs. Thus, PROMPTs constitute a class of unstable transcripts, and we refer to the PROMPT-encoding DNA as the "PROMPT region." Short RNAs produced around TSSs have previously been reported, most notably promoter-associated short RNAs, which were on average 0.5 kb on either side of the TSS (4). These are, however, physically separate from PROMPTs by several hundred base pairs (fig. S4). In contrast, a few verified

PROMPT regions show weak signs of transcriptional activity in other data sets, such as scattered cap analysis of gene expression tags (markers of transcription initiation events) (7) and expressed sequence tags unassigned to known genomic features (fig. S5).

We next examined whether PROMPTs were sense or antisense relative to the mRNA produced from the downstream positioned genes. Orientation-specific RT-qPCR performed on RNA from either hRrp40 depleted- or control cells demonstrated that, regardless of directional preference, both sense and antisense transcripts were detectable in PROMPT regions (Fig. 2B). In the presence of actinomycin D, which inhibits spurious synthesis of potential second-strand cDNA artifacts (8), this bidirectionality of PROMPTs was still observed (fig. S6). Moreover, both sense and antisense RNAs were stabilized to a similar extent by hRrp40 depletion (Fig. 2B), demonstrating that both species are exosome substrates. When aligning array data to the TSSs of PROMPT regions where either sense or antisense RNA production predominates, they displayed patterns similar to the average PROMPT profile (fig. S7). Taken together, these data suggest a complex pattern of RNA polymerase II (RNAPII) activity in either orientation upstream of individual gene promoters. This observation was supported by nonexhaustive rapid amplification of cDNA ends (RACE) analyses of eight PROMPT regions, which often reveals multiple 5' and 3' ends (fig. S8).

To investigate the requirements for transcription upstream of promoters, we transiently transfected HeLa cells with a plasmid containing the

¹Centre for mRNP Biogenesis and Metabolism, Department of Molecular Biology, C. F. Møllers Alle, Building 1130, Aarhus University, Denmark. ²Bioinformatics Research Center, C. F. Møllers Alle, Building 1110, Aarhus University, Denmark. *Present address: Exiqon A/S, Byggestubben 9, 2850 Vedbæk, Denmark. †To whom correspondence should be addressed. E-mail: thj@mb.au.dk

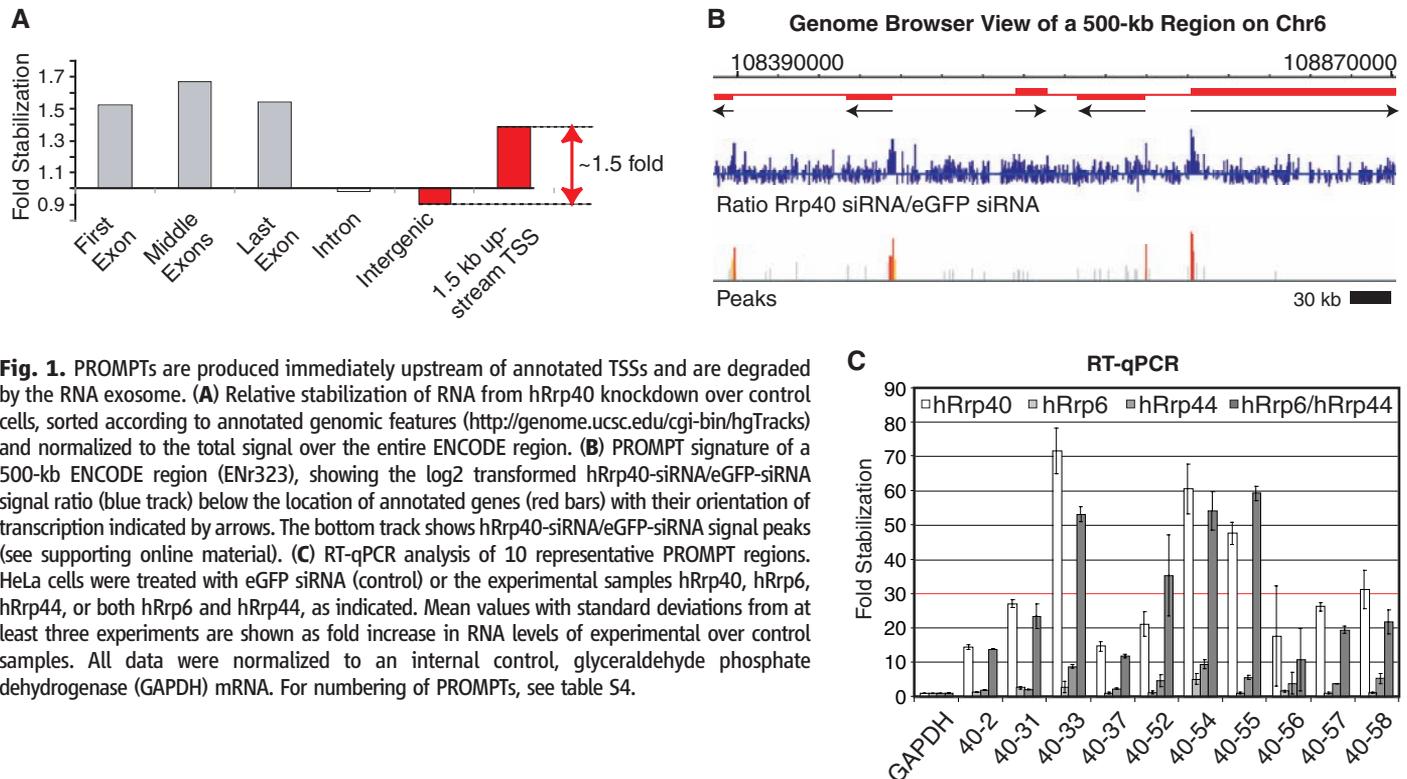


Fig. 1. PROMPTs are produced immediately upstream of annotated TSSs and are degraded by the RNA exosome. (A) Relative stabilization of RNA from hRrp40 knockdown over control cells, sorted according to annotated genomic features (<http://genome.ucsc.edu/cgi-bin/hgTracks>) and normalized to the total signal over the entire ENCODE region. (B) PROMPT signature of a 500-kb ENCODE region (ENr323), showing the log₂ transformed hRrp40-siRNA/eGFP-siRNA signal ratio (blue track) below the location of annotated genes (red bars) with their orientation of transcription indicated by arrows. The bottom track shows hRrp40-siRNA/eGFP-siRNA signal peaks (see supporting online material). (C) RT-qPCR analysis of 10 representative PROMPT regions. HeLa cells were treated with eGFP siRNA (control) or the experimental samples hRrp40, hRrp6, hRrp44, or both hRrp6 and hRrp44, as indicated. Mean values with standard deviations from at least three experiments are shown as fold increase in RNA levels of experimental over control samples. All data were normalized to an internal control, glyceraldehyde phosphate dehydrogenase (GAPDH) mRNA. For numbering of PROMPTs, see table S4.

β -globin gene under control of the strong cytomegalovirus promoter (pCMV) that is preceded by 2.2 kb of bacteriophage λ DNA (Fig. 2C). This resulted in transcript production from the λ DNA, demonstrating that PROMPT-like transcription can be initiated independent of the underlying DNA sequence. Transcripts arising from the λ DNA region cannot be read-through products from transcription around the plasmid because β -globin transcript levels reach background immediately downstream of the transcription termination site. Again, 5'- and 3'-RACE analyses were employed to map some transcription start- and end points, which substantiated the observation of dynamic and complex RNAPII activity in the region (fig. S9). Deletion of the CMV promoter resulted in the concomitant elimination of PROMPT and β -globin gene transcription (Fig. 2C and fig. S9). Thus, the generation of transcripts

upstream of an active gene appears to depend on the gene promoter.

To further characterize the transcriptional activity and its origin in PROMPT regions, we compared PROMPT patterns to RNAPII occupancy, transcription factor binding, and chromatin modifications using public data sets generated by the ENCODE project (table S2). In two representative examples, the PROMPT region is covered by markers of active transcription, RNAPII and acetylated histone 3 (H3K9ac), whereas the transcription initiation factor TAF1 peaks at the TSS (Fig. 3A). The generality of this observation was examined by creating composite profiles of the 64 low-complexity regions encompassing PROMPT and TSS sequences. PROMPTs generally overlap with RNAPII, marks of active chromatin, and DNase hypersensitive sites (9, 10), but not with peaks of transcription initiation factors;

e.g., TAF1 or E2F1 (10, 11) (Fig. 3B and fig. S10). Although this reinforces the concept of substantial transcription activity upstream of bona fide genes, the TSS-restricted localization of transcription initiation factors supports our conclusion using CMV/ Δ CMV plasmids and argues against the presence of an independent PROMPT promoter.

A link between transcriptional activity in PROMPT and gene regions is further supported by scatter plots showing a strong positive correlation between total average RNAPII chromatin immunoprecipitation (ChIP) signal within the first 1.5 kb up- and downstream of all 1594 ENCODE TSSs (Fig. 4A). This relation is also evident from raw RNA expression data from the hRrp40 depletion experiment (Fig. 4B). With slopes of up to 0.7, these plots indicate that transcription activity in the PROMPT region is comparable to that in the beginning of the gene.

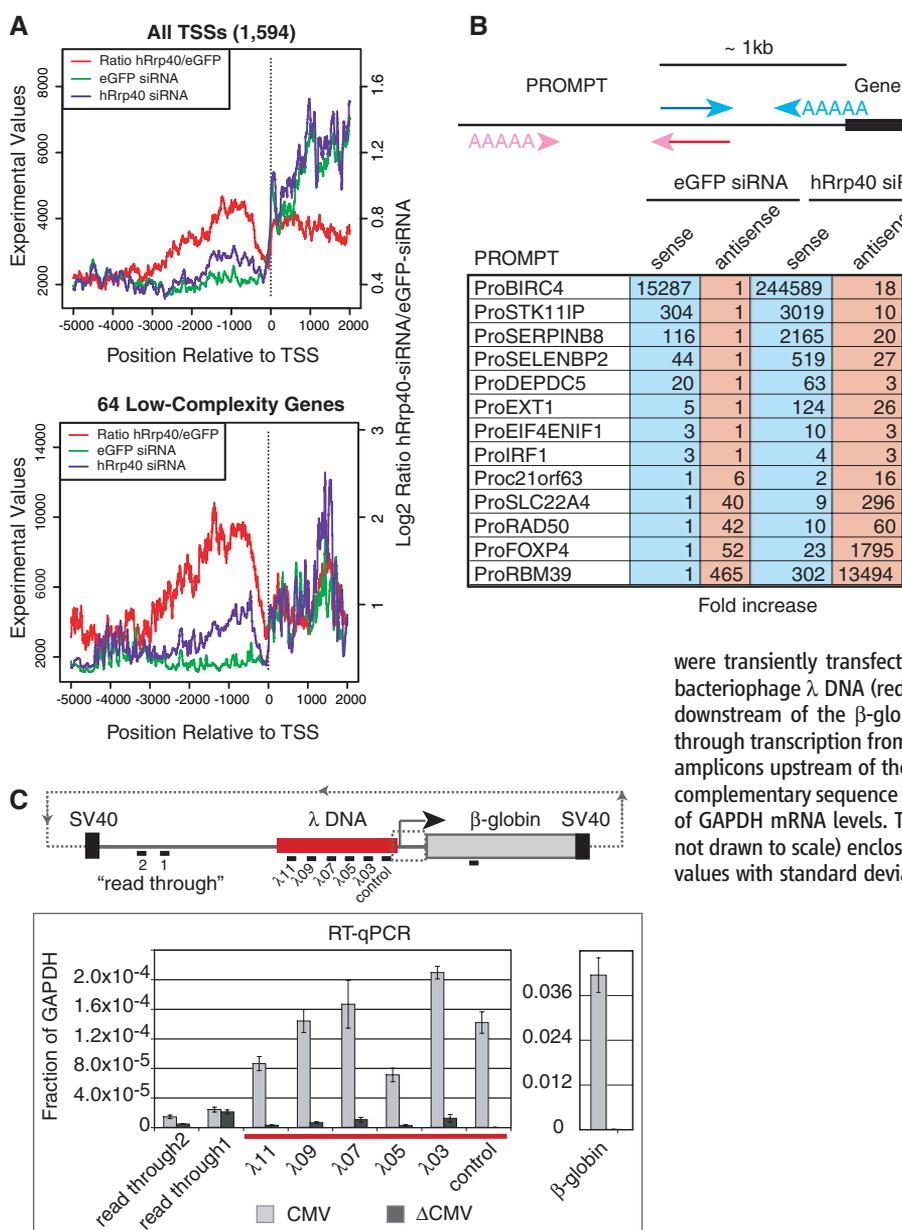


Fig. 2. PROMPT expression maps to 0.5 to 2.5 kb (i) upstream of TSSs, (ii) can occur in both orientations, and (iii) requires the gene promoter. **(A)** Composite RNA profiles upstream of all 1594 (top) or 64 low-complexity (bottom) TSSs. Raw (single-channel) data (smoothed over a 10-bp window) from hRrp40-siRNA treated cells, control (eGFP) siRNA-treated cells, and their ratio are shown as indicated. The left y axis denotes values for raw data, and the right y axis denotes the log₂-transformed ratio of the raw data, scaled to center at zero. Positions in base pairs of RNA signals relative to TSSs are shown on the x axes. **(B)** The sense (blue)/antisense (red) directionality of selected PROMPTs was determined by RT-qPCR with gene-specific primers (~1 kb upstream the TSS) in either orientation in combination with a T₂₀VN primer that hybridizes to the 3' poly(A) tail. Fold increases relative to the lowest value in control cells (set to 1) are plotted. PROMPTs are ordered such that the one with the highest preference for sense transcription is at the top. **(C)** Generation of promoter-upstream transcription in nonhuman DNA. Plasmids containing the β -globin gene under control of a viral promoter (CMV) or its Δ CMV control were transiently transfected into HeLa cells. Both constructs have an insertion of bacteriophage λ DNA (red bar) upstream and a strong SV40 poly(A) site (black box) downstream of the β -globin gene. RNA levels were analyzed by RT-qPCR. Read-through transcription from the β -globin promoter was measured with the use of two amplicons upstream of the λ DNA ("read through"). The "control" amplicon has no complementary sequence in the Δ CMV plasmid. Values on the y axis are percentages of GAPDH mRNA levels. The dashed box in the linear plasmid representation (top, not drawn to scale) encloses the region that is deleted in the Δ CMV construct. Mean values with standard deviations ($n = 3$) are shown.

Downloaded from <http://science.sciencemag.org/> on September 23, 2019

Given their ubiquitous nature, do PROMPTs have a function? A few noncoding RNAs that have been reported to exert regulatory functions

are located in potential PROMPT regions (12, 13). Likewise, a noncoding RNA directly upstream of the sphingosine-kinase1 (SPHK1) gene, which

affects the methylation status of CpG dinucleotides within its promoter (13), is also stabilized in hRrp40 knockdown cells (fig. S11A). It is therefore interesting to note that the methylation level of some CpG dinucleotides within the SPHK1 promoter region is increased upon hRrp40 depletion (fig. S11B). That PROMPTs more generally may affect promoter methylation is further indicated by the finding that for genes with similar expression levels, PROMPT levels are generally higher around promoters with a high CpG score (fig. S11C).

PROMPTs may arise wherever open chromatin presents itself, possibly as the byproduct of an as yet unexplored aspect of the mechanism of gene transcription. Evolution, being an opportunistic force, may then have co-opted at least some of these PROMPTs as part of regulatory mechanisms (fig. S11). One such molecular system could involve the control of CpG (de)methylation, an as of now poorly understood process (14). An alternative, but not mutually exclusive, possibility is that PROMPT transcription may have a more general function by providing reservoirs of RNAPII molecules, which can facilitate rapid activation of the downstream gene, and/or by serving to alter chromatin structure. Clearly, the generality of the PROMPT phenomenon hints at a more complex regulatory chromatin structure around the TSS than was previously anticipated.

Downloaded from <http://science.sciencemag.org/> on September 23, 2019

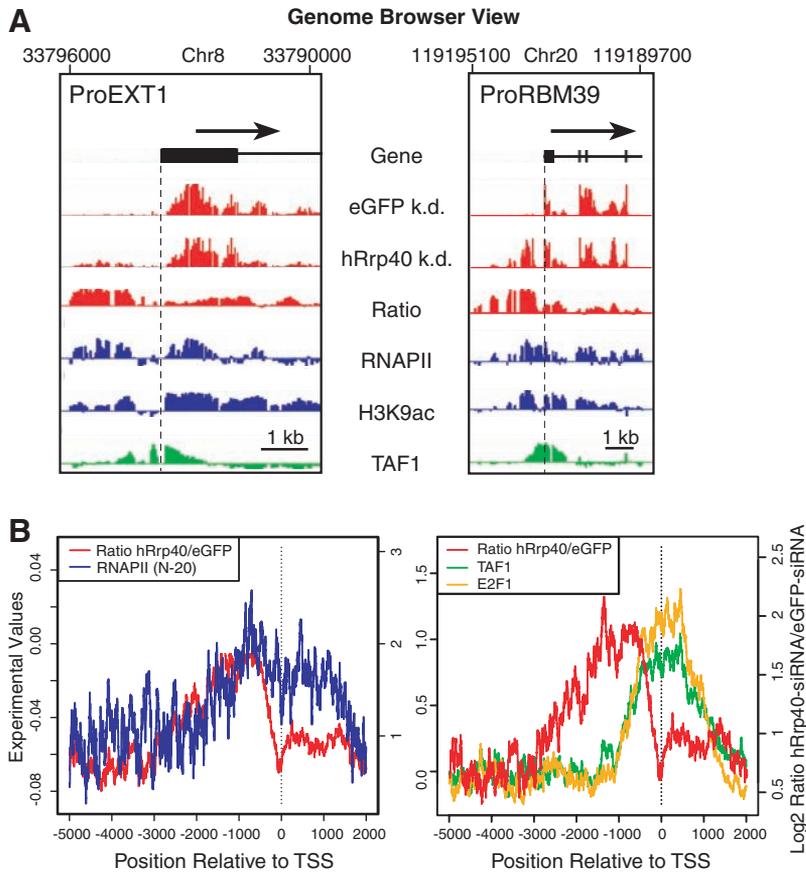


Fig. 3. PROMPT regions are actively transcribed. **(A)** Details of transcript levels from this study compared with previously published ChIP-chip data for PROMPT and 5' regions of two representative genes. Genomic coordinates are shown on top in numbers of base pairs. **(B)** Composite profiles of RNA stabilization in the PROMPT regions of 64 low-complexity TSSs displayed as in Fig. 2A and compared with the indicated data sets.

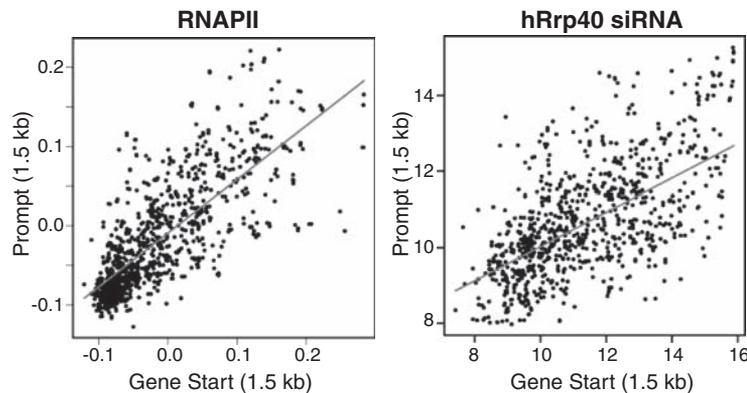


Fig. 4. Overall correlation of PROMPT- and gene-expression levels. **(Left)** Scatter plot of RNAPII distribution as measured by ChIP-chip over all 1594 TSSs in the ENCODE region (data taken from GEO, accession number GSE6391). Data were integrated over 1.5 kb before (y axis, "PROMPT") and after (x axis, "Gene Start") each TSS and plotted against each other. The slope of the linear regression is 0.68 with a P value of $\leq 10^{-300}$ (t test, product-moment correlation) and an r^2 value of 0.61 [degrees of freedom (df) = 1511]. **(Right)** Scatter plot of single-channel RNA microarray signals from hRrp40 siRNA-treated cells created as above with the exception that, in the gene, only data corresponding to exonic DNA were used to remove exon/intron biases (fig. S3). Statistical values are slope = 0.45, P value $< 10^{-137}$, and $r^2 = 0.39$ (df = 1420).

References and Notes

1. E. Birney *et al.*, *Nature* **447**, 799 (2007).
2. M. B. Gerstein *et al.*, *Genome Res.* **17**, 669 (2007).
3. A. Sandelin *et al.*, *Nat. Rev. Genet.* **8**, 424 (2007).
4. P. Kapranov *et al.*, *Science* **316**, 1484 (2007); published online 16 May 2007 (10.1126/science.1138341).
5. Materials and methods are available as supporting material on *Science* Online.
6. F. Wyers *et al.*, *Cell* **121**, 725 (2005).
7. P. Carninci *et al.*, *Nat. Genet.* **38**, 626 (2006).
8. F. Perocchi, Z. Xu, S. Clauder-Munster, L. M. Steinmetz, *Nucleic Acids Res.* **35**, e128 (2007).
9. A. P. Boyle *et al.*, *Cell* **132**, 311 (2008).
10. N. D. Heintzman *et al.*, *Nat. Genet.* **39**, 311 (2007).
11. M. Bieda, X. Xu, M. A. Singer, R. Green, P. J. Farnham, *Genome Res.* **16**, 595 (2006).
12. X. Wang *et al.*, *Nature* **454**, 126 (2008).
13. T. Imamura *et al.*, *Biochem. Biophys. Res. Commun.* **322**, 593 (2004).
14. S. K. Ooi, T. H. Bestor, *Cell* **133**, 1145 (2008).
15. We thank G. Pruijn for antibodies, D. Libri, A. Sandelin, and D. Schubeler for comments on the manuscript, and D. Riishøj and K. Jürgensen for technical assistance. This work was supported by the Danish National Research Foundation and the Danish Natural Science Research Council. Microarray data have been submitted to gene expression omnibus (GEO) at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) under accession number GSE12431.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1164096/DC1
Materials and Methods
Figs. S1 to S11
Table S1 to S4
References

1 August 2008; accepted 12 November 2008
Published online 4 December 2008;
10.1126/science.1164096
Include this information when citing this paper.

RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters

Pascal Preker, Jesper Nielsen, Susanne Kammler, Søren Lykke-Andersen, Marianne S. Christensen, Christophe K. Mapendano, Mikkel H. Schierup and Torben Heick Jensen

Science **322** (5909), 1851-1854.

DOI: 10.1126/science.1164096 originally published online December 4, 2008

ARTICLE TOOLS

<http://science.sciencemag.org/content/322/5909/1851>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2008/12/04/1164096.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/322/5909/1804.full>

REFERENCES

This article cites 13 articles, 3 of which you can access for free
<http://science.sciencemag.org/content/322/5909/1851#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

American Association for the Advancement of Science