Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs

Ashish Dhir1, Somdutta Dhir1, Nick J Proudfoot1 & Catherine L Jopling2

MicroRNAs (miRNAs) play a major part in the post-transcriptional regulation of gene expression. Mammalian miRNA biogenesis begins with cotranscriptional cleavage of RNA polymerase II (Pol II) transcripts by the Microprocessor complex. Although most miRNAs are located within introns of protein-coding transcripts, a substantial minority of miRNAs originate from long noncoding (lnc) RNAs, for which transcript processing is largely uncharacterized. We show, by detailed characterization of liver-specific lnc-pri-miR-122 and genome-wide analysis in human cell lines, that most lncRNA transcripts containing miRNAs (lnc-pri-miRNAs) do not use the canonical cleavage-and-polyadenylation pathway but instead use Microprocessor cleavage to terminate transcription. Microprocessor inactivation leads to extensive transcriptional readthrough of lnc-pri-miRNA and transcriptional interference with downstream genes. Consequently we define a new RNase III–mediated, polyadenylation-independent mechanism of Pol II transcription termination in mammalian cells.
RESULTS

Inc-pri-miR-122 transcripts are capped but not polyadenylated

First, we characterized the processing of Inc-pri-miR-122 (Fig. 1a). By northern analysis, we identified mature miR-122 and two Inc-pri-miR-122 transcripts of ~4.8 and ~1.9 kb in total RNA from human liver and in the human hepatocellular carcinoma cell line Huh7 but not in HeLa or HepG2 cells (Fig. 1b). Intron- and exon-specific probes showed that the larger Inc-pri-miR-122 transcript was unspliced, whereas the smaller transcript corresponded to inefficiently spliced RNA lacking an internal 3-kb intron (Fig. 1c). The transcript size indicated that the 3′ end lay close to the pre-miR-122 hairpin, ~2.5 kb upstream of a previously identified polyadenylated 3′ end27 (Fig. 1a). We did not detect any longer Inc-pri-miR-122 transcripts. Immunoprecipitation with an antibody directed against the mG cap demonstrated that Inc-pri-miR-122 was capped, similarly to GAPDH mRNA, a result in agreement with existing cap analysis of gene expression (CAGE) data28 (Fig. 1d). However, quantitative reverse-transcription PCR (RT-qPCR) and northern analysis of pa-selected RNA indicated that Inc-pri-miR-122 was nonpolyadenylated, in contrast to GAPDH mRNA but similarly to U6 small nuclear RNA (snRNA) (Fig. 1e).

Microprocessor cleavage generates Inc-pri-miR-122 3′ end

To map Inc-pri-miR-122 3′ ends at nucleotide resolution, we developed a pA tail–independent 3′ rapid amplification of cDNA ends (3′ RACE) technique. We mapped Inc-pri-miR-122 3′ ends to just upstream of the site of Drosha cleavage on the 5′ arm of the pre-miR-122 hairpin. We observed some heterogeneity of 3′-end location, presumably due to exonucleolytic trimming after Drosha cleavage (Fig. 2b and Supplementary Fig. 2). Both unspliced and spliced Inc-pri-miR-122 were retained in the nucleus and rapidly degraded (Fig. 2c,d), as expected for nonpolyadenylated transcripts.

These results suggested that the Inc-pri-miR-122 3′ end is generated by Drosha cleavage and not by CPA. To characterize the role of the Microprocessor in Inc-pri-miR-122 3′-end formation, we compared the effects of short interfering RNA (siRNA)-mediated knockdown of DGCR8.
CPA does not occur on Inc-pri-miR-122 transcripts

The transcriptional readthrough that we observed after DGCR8 knockdown implied that Inc-pri-miR-122 does not switch to the efficient CPA mechanism of transcriptional termination when the Microprocessor mechanism is inhibited. To determine this directly, we pA-selected RNA from Huh7 cells with or without DGCR8 knockdown. Similarly to U6 snRNA, Inc-pri-miR-122 transcripts remained pA negative (pA−) even when their 3′-end formation was compromised by DGCR8 depletion. In contrast, GAPDH mRNA was strongly pA positive (pA+) under both conditions (Fig. 4a). By next-generation sequencing analysis of chromatin-associated RNA from Huh7 cells (chromatin RNA–seq), we found that DGCR8 depletion leads to extensive transcriptional readthrough for over 5 kb downstream of the pre-miR-122 hairpin in Inc-pri-miR-122 (Fig. 4b), despite the presence of several consensus PASs.

These results indicated that Pol II transcribing Inc-pri-miR-122 fails to recognize PASs even when Microprocessor cleavage is inhibited. This surprising finding suggested that Pol II might be recruited to the endogenous Inc-pri-miR-122 promoter in a CPA-refractory form. To investigate this further, we cloned Inc-pri-miR-122 under the control of the HIV long terminal repeat (LTR) promoter with or without pre-miR-122 hairpin deletion (Fig. 5a). We transfected the resulting wild-type (WT) and deleted (Δ) plasmids into HeLa cells, which do not express endogenous Inc-pri-miR-122 (Fig. 1b), together with a plasmid encoding the Tat transcriptional activator. We confirmed that mature miR-122 was expressed from the WT

![Figure 2](https://example.com/figure2.png)

**Figure 2** 3′-end mapping, subcellular distribution and rapid turnover of lnc-pri-miR-122. (a) Gene map of Inc-pri-miR-122 showing positions of primers for RT-qPCR and 3′-end mapping, Fwd, forward primer. (b) Poly(A) polymerase (PAP)-dependent 3′-end mapping of Inc-pri-miR-122 with 3′ RACE. Position of 3′ RACE PCR product shown by gel fractionation and location of mapped 3′-end cleavage products are shown by red arrowheads on the pre-miR-122 hairpin structure (sequencing in Supplementary Fig. 2). M, marker. (c) Pri-miR-122 and GAPDH mRNA distribution determined between cell fractions (WC, whole cell; N, nuclear; C, cytoplasmic) by RT-PCR with indicated primers. Fractionation control, western blot showing purity of nuclear and cytoplasmic fractions by use of cytoplasmic- and nuclear-specific protein antibodies. (d) RNA stability after actinomycin D inhibition of transcription at various time points, measured by RT-qPCR of lnc-pri-miR-122 transcripts versus GAPDH mRNA.

RNA levels are expressed relative to the levels at time 0, which were set to 1. Huh7 cells were used in all experiments. Error bars, s.d. (n = 3 independent experiments). Original images of gels, autoradiographs and blots used in this study can be found in Supplementary Data Set 1.

![Figure 3](https://example.com/figure3.png)

**Figure 3** Microprocessor defines transcriptional termination of Inc-pri-miR-122. (a,b) Mapping nascent transcription across Inc-pri-miR-122 (a) versus GAPDH (b) in Huh7 cells. Gene maps show positions of primers used in RT-qPCR analysis of chromatin RNA levels (left) and bromo-UTP–labeled nuclear run on RNA (BrU-NRO, right). (c) Western blot showing effective depletion of DGCR8 and CPSF-73 by siRNA transfection in Huh7 cells. Error bars, s.d. (n = 3 independent experiments). Original images of gels, autoradiographs and blots used in this study can be found in Supplementary Data Set 1.
but not the Δ plasmid (Fig. 5b). Unspliced and spliced Inc-pri-miR-122 transcripts generated from the WT plasmid were the same size as the endogenous transcripts in Huh7 cells (Fig. 5c), thus indicating that the site of 3′-end formation is independent of promoter or cell type. Deletion of the pre-miR-122 hairpin or depletion of Drosha or DGCR8 led to production of unspliced and spliced Inc-pri-miR-122 transcripts that migrated at a higher molecular weight than WT transcripts (Fig. 5d,e). pA fractionation indicated that Inc-pri-miR-122 RNA generated from the WT plasmid was pA−, similarly to U6 snRNA, but became pA+ when DGCR8 was depleted, similarly to GAPDH mRNA (Fig. 5f). We mapped the 3′ end of Δ RNA by 3′ RACE to a PAS 91 nt downstream of the pre-miR-122 hairpin (pA1, Fig. 5a). Mutagenesis indicated that this PAS was necessary for 3′-end generation in Δ but not WT Inc-pri-miR-122 transcripts (Fig. 5g). This confirmed that in a plasmid context, Inc-pri-miR-122 3′ ends were generated by the Microprocessor, but in the absence of this processing CPA occurred at a site downstream. Importantly, this was in contrast to the endogenous transcripts, for which transcriptional readthrough occurred when Microprocessor cleavage was inhibited, and the PAS was not used (Figs. 3 and 4).

Microprocessor terminates transcription of most Inc-pri-miRNAs
Having demonstrated that chromatin RNA-Seq is an effective method of identifying defects in transcriptional termination after Microprocessor depletion (Fig. 4b), we extended this analysis to HeLa cells on a genome-wide scale. We chose HeLa cells because they express a relatively high number of miRNAs. We found that depletion of either Drosha or DGCR8 by siRNA treatment (Fig. 6e) resulted in transcriptional readthrough in most of the expressed Inc-pri-miRNAs (Supplementary Table 1). MIR181A1HG and MIR17HG are shown as specific examples (Fig. 6a and Supplementary Fig. 4a), and meta-gene analysis showed a general termination defect with transcription extending more than 10 kb downstream of Inc-pri-miRNA after Microprocessor depletion (Fig. 6b). A few Inc-pri-miRNAs did not show transcriptional readthrough after Microprocessor depletion (Supplementary Table 1); MIRLET7BH is shown as an example (Supplementary Fig. 4b). Importantly, Dicer knockdown did not affect Inc-pri-miRNA transcriptional termination, thus confirming that the effects of Microprocessor depletion are direct and not due to loss of mature miRNAs (Supplementary Fig. 5).
In marked contrast, protein-coding genes that contain intronic pre-miRNAs showed no termination defect after Microprocessor depletion. Rather, intronic sequence containing pre-miRNA was stabilized (higher reads), as shown for MCM7 (Fig. 6c), which has three miRNAs (miR-25, miR-93 and miR-106b) in its penultimate intron. Presumably, this selective intron stabilization was caused by the loss of Drosha-mediated cotranscriptional cleavage. Metagene analysis revealed that Microprocessor activity generally had no effect on transcriptional termination for these protein-coding miRNA genes (Fig. 6d). This implies a functional difference between the processing of protein-coding pri-miRNA transcripts, which use CPA-mediated termination irrespective of internal Microprocessor cleavage, and that

**Figure 6** Microprocessor-dependent chromatin RNA–seq profiles across pri-miRNAs from HeLa cells. (a) Reads (RPKM) across MIR181A1HG, showing readthrough profiles after Microprocessor depletion. (b) Metagene analysis of all expressed Inc-pri-miRNAs. (c) Reads across a protein-coding gene (MCM7) containing an intronic miRNA cluster, showing intron accumulation after Microprocessor depletion. (d) Metagene analysis of all expressed protein-coding genes containing miRNAs. Direction of transcription is indicated by green arrows and positions of miRNA by red vertical lines in a and c. Metagene profiles show transcription unit (between TSS and transcription end site (TES)) followed by 10 kb of 3′ flanking region in b and d (Mann Whitney U test, $P < 0.0001$, two-tailed; $n$ denotes number of genes in b and d). (e) Western blot showing effective depletion of Drosha and DGC8R8 by siRNA transfection in HeLa cells. Original images of gels, autoradiographs and blots used in this study can be found in Supplementary Data Set 1.

**Figure 7** Microprocessor-dependent termination prevents transcriptional interference. (a) Chromatin RNA–seq profiles across the MIR17HG-GPCS tandem gene locus, showing readthrough transcription after Microprocessor depletion. Black arrow and red box highlight reduction in GPCS exon 1 peak after Microprocessor depletion. Direction of transcription is indicated by green arrows and miRNA by red vertical lines. (b) RT-qPCR analysis of chimeric MIR17HG-GPCS transcripts versus GPCS exon 1. Gene-specific reverse-transcription primer (black arrowhead) in exon 1 and PCR amplicons are indicated by black bars. (c) Chromatin RNA–seq profiles across convergent OGFRL1-LINC00472 gene locus. (d) Read quantification for protein-coding genes subject to transcriptional interference after Microprocessor depletion. (e) mRNA levels of GPCS and OGFRL1, determined by RT-qPCR with exon-specific primers. RNA levels are expressed relative to control siRNA, which was set to 1. All values are normalized to GAPDH mRNA. (f) Western blot showing reduction in protein levels of GPCS and OGFRL1 by transcriptional interference. All experiments used HeLa cells. Error bars, s.d. (n = 3 independent experiments). Original images of gels, autoradiographs and blots used in this study can be found in Supplementary Data Set 1.
**Figure 8** Inc-pri-miRNA may be CPA incompetent or competent after Microprocessor depletion. (a, b) RNA sequencing of pA-fractionated nuclear RNA from siRNA-treated HeLa cells. (a) Nuclear RNA-sequencing profile of MIR17HG pA⁺ and pA⁻ RNA after DGCR8 depletion. The position of the PAS (pA) is marked by a dashed vertical line. The direction of transcription is indicated by a green arrow and positions of miRNA by red vertical lines in a and b. (c) Model showing CPA in protein-coding pri-miRNA allows generation of spliced mRNA and miRNA, whereas Microprocessor-driven termination in Inc-pri-miRNA generates miRNA and a pA⁻ host transcript that is rapidly degraded. Inc-pri-miRNA may be CPA incompetent, thus leading to transcriptional readthrough and interference when Microprocessor cleavage is inhibited, or CPA competent, thus allowing effective transcription termination even in the absence of Microprocessor.

Red lightning bolt with black fill depicts CPA-mediated cleavage. Blue lightning bolt with red fill depicts Microprocessor-mediated cleavage.

Most Inc-pri-miRNAs remain pA⁺ after Microprocessor depletion

Similarly to endogenous Inc-pri-miR-122, Microprocessor-terminated Inc-pri-miRNAs appear to be insensitive to the presence of cryptic PASs, which are invariably present within the gene and 3’ flanking regions. To further investigate the use of PASs in pri-miRNAs, we performed nuclear pA⁺ and pA⁻ RNA sequencing in HeLa cells with or without DGCR8 knockdown. We found that the majority of Inc-pri-miRNAs existed as predominantly pA⁺ transcripts, and those that showed extensive readthrough upon loss of Microprocessor remained pA⁺ (Supplementary Table 3). MIR17HG is shown as a specific example (Fig. 8a). It is remarkable that for these Pol II transcripts...
PASs remain opaque to RNA processing by the CPA complex. However, a few lnc-pri-miRNAs use PASs to some extent, especially after Microprocessor inactivation. Thus MIRLET7BHG transcripts switched from mainly pA+ to pA− after Microprocessor depletion (Fig. 8b and Supplementary Table 3), and efficient termination occurred at a canonical PAS positioned immediately downstream of pre-miR-let7b. This is similar to the switch to CPA at a downstream PAS that we observed in ectopically expressed lnc-pri-miR-122 (Fig. 5), thus indicating that this distinction is biologically relevant.

**DISCUSSION**

We have identified a CPSF-73–independent, Microprocessor-driven transcription termination mechanism for pri-miR-122 lncRNA. This results in the production of unstable nuclear unspliced and spliced lnc-pri-miR-122 transcripts with 3′ ends defined by Drosha cleavage (Fig. 1 and 2). By genome-wide analysis, we found that transcriptional termination by the Microprocessor is a feature shared with most other lnc-pri-miRNA genes in both HeLa and Huh7 cells (Fig. 6 and Supplementary Tables 1 and 2).

Previous evidence has indicated that pri-miRNA are typical capped, polyadenylated Pol II transcripts. This has been clearly established for protein-coding genes containing intronic miRNA5 but is also true of the few lnc-pri-miRNA for which the 3′ end has been characterized, such as pri-miR-21 (refs. 4,33) and Caenorhabditis elegans let-7 (ref. 34). Our genome-wide analysis confirmed that CPA does occur in a minority of lnc-pri-miRNAs (Supplementary Table 3) but showed that the Microprocessor-mediated termination mechanism predominates (Supplementary Tables 1 and 2). A previous study showed that Drosha processing of pre-miRNA hairpins can attenuate downstream transcription by providing an entry site for Xrn2. However, those experiments were carried out with plasmid constructs that lacked a PAS, and they did not provide evidence that Microprocessor cleavage could actually replace CPA as a mode of transcriptional termination.

A role for Drosha cleavage at the HIV LTR in preventing productive transcription elongation in the absence of Tat also indicates that the Microprocessor can disrupt the transcription machinery but does not connect miRNA processing with transcriptional termination. In contrast, we have demonstrated that Microprocessor cleavage mediates transcriptional termination on endogenous pri-miRNA transcripts and moreover that this is limited to lnc-pri-miRNA. This departs from the clear current consensus that pri-miRNAs are typical capped and polyadenylated transcripts (recent review in ref. 5), a view derived from analysis of protein-coding pri-miRNAs and confirmed for protein-coding genes by our genome-wide analysis.

The Microprocessor termination pathway adds to a short list of noncanonical mechanisms of Pol II transcriptional termination. Termination of histone miRNA does not involve polyadenylation, similarly to pri-miR-122 termination, but requires cleavage by CPSF-73, in common with CPA36. In yeast, the Nrd1-Nab3-Sen1 pathway terminates Pol II transcription of snRNAs, small nucleolar RNAs and cryptic unstable transcripts37–39, whereas in mammals the Integrator complex mediates transcriptional termination and subsequent processing of snRNAs40,41. Importantly, both the Nrd1 and Integrator pathways are used for termination only on short transcripts. In contrast, the Microprocessor mechanism described here provides an alternative to CPA in terminating transcription several kilobases downstream of initiation.

The closest parallel to this Microprocessor-dependent termination pathway is in budding yeast, in which Rnt1 cleavage leads to a pA−-independent termination of Pol II transcription when CPA fails, thus preventing readthrough transcription and subsequent transcriptional interference20,22. Rnt1-terminated transcripts are rapidly degraded, similarly to lnc-pri-miR-122. Transcriptional termination after either Rnt1 cleavage or CPA occurs as a result of Rat1 or Xrn2 degradation of the nascent transcript downstream of the cleavage site15,16,20. However, we observed no effect of Xrn2 depletion on pri-miR-122 transcriptional termination (data not shown), a result in contrast to the role for Xrn2 in RNA degradation after Drosha cleavage of an intergenic clustered pri-miRNA5. The mechanism of Pol II termination after Microprocessor cleavage of lnc-pri-miR-122 may involve other nucleases or termination factors, such as Pcf11 (refs. 18,42) or the mammalian ortholog of Sen1, senataxin43. Of note, we observed variable levels of chromatin RNA−seq signal downstream of the pre-miRNA hairpin after DGC8 knockdown among different Microprocessor-terminated lnc-pri-miRNAs. For example, the amplitude of readthrough transcription is higher in MIR181A1HG and MIR17HG than lnc-pri-miR-122 (Figs. 4b and 6, Supplementary Fig. 4 and Supplementary Tables 1 and 2). The decrease in lnc-pri-miR-122 readthrough transcription may result from low-level Microprocessor-driven termination mediated by residual DGC8 after siRNA transfection, with gene-specific differences possibly due to some pre-miRNA hairpins competing more effectively for the remaining Microprocessor. Alternatively, it is possible that other non-CPA mechanisms can displace transcribing Pol II from these genes.

Although Microprocessor-driven termination is a common feature of lnc-pri-miRNAs, it is not universal. This raises the question of whether specific genetic features are necessary for Microprocessor-mediated termination to occur. We find that this mechanism can be used by lncRNAs irrespective of whether the pre-miRNAs are located in an exon or intron (Supplementary Table 4c), at a range of distances from the TSS (Supplementary Table 4a,b). It is possible that the efficiency of termination is affected by the Microprocessor cleavage event itself. For example, protein cofactors are known to assist in Microprocessor release of specific pre-miRNA3, and sequence features surrounding the pre-miRNA hairpin can influence the efficiency of processing44. Many Microprocessor-terminated lnc-pri-miRNAs contain clustered pre-miRNA hairpins; this raises the interesting question of whether Microprocessor cleavage at a specific pre-miRNA drives transcription termination. Because some lnc-pri-miRNA transcriptional readthrough occurs after Microprocessor cleavage, similarly to the continued Pol II transcription after cleavage at a PAS in CPA-dependent termination, it is not possible to precisely define which cleavage event drives termination from our chromatin RNA−seq data. A recent study defined a ‘Microprocessing index’ (MPI) that shows variable cleavage efficiency for different pre-miRNAs45. Of the 13 HeLa lnc-pri-miRNAs detected in this study, 11 contain a pre-miRNA hairpin with MPI <-1.0, indicating efficient cotranscriptional processing, and seven of these contain a hairpin with MPI <-3.0, indicating highly efficient processing45. Although the pool of lnc-pri-miRNAs that we detect in HeLa is too small to draw statistically robust conclusions, our results raise the possibility that rapid Microprocessor cleavage may be important for transcriptional termination.

We found that Microprocessor-driven transcriptional termination is used by most lnc-pri-miRNAs (73%) but not protein-coding pri-miRNAs (Fig. 6 and Supplementary Tables 1 and 2), thus demonstrating a fundamental difference in RNA processing between lncRNAs and protein-coding transcripts. The Microprocessor mechanism has parallels to another CPA-independent mechanism of transcriptional termination used by the lncRNA MALAT1, in which 3′-end formation and concomitant release of a small RNA is mediated by the tRNA-biogenesis endonucleases RNase P and RNase Z46. However, lncRNAs derived from bidirectional firing at protein-coding gene promoters,
known as upstream antisense RNA transcripts, terminate transcription by CPA and tend to use promoter-proximal PASs. The difference between a PAS and Inc-pri-miRNA transcripts as described in this study and CPA-competent IncRNAs derived from antisense promoter activity may relate to promoter specificity. Pol II elongation complexes set up on a protein-coding-gene promoter may be CPA responsive irrespective of promoter directionality. In contrast, Inc-pri-miRNA promoters may form a different type of Pol II elongation complex that is CPA nonresponsive but Microprocessor active. A role for the promoter is supported by our observation that most Inc-pri-miRNAs, including Inc-pri-miR-122, do not use CPA even when the Microprocessor is depleted, instead showing extensive transcriptional readthrough and remaining pA− (Figure 8a and Supplementary Table 3). In contrast, ectopically expressed Inc-pri-miR-122 uses Microprocessor cleavage to mediate transcriptional termination but switches to CPA when this mechanism is inhibited (Figure 5). Therefore, the PAS downstream of pre-miR-122 can function, but not in the context of the endogenous gene. The biological relevance of this distinction is clear from our observation of a switch to CPA in MIR17HG transcripts after Microprocessor depletion (Figure 8b).

IncRNA genes are generally thought to be similar to mRNA genes, with similar chromatin profiles, transcriptional regulation and splice signals. However, IncRNA genes tend to have fewer and longer exons than mRNA genes, and there is a trend for less efficient splicing of IncRNAs than of mRNAs. We find that CPA is inefficient on Inc-pri-miRNA (Supplementary Table 3). Because splicing and CPA function to generate stable cytoplasmic coding transcripts, there would be little need for these processes to occur on most IncRNAs. Microprocessor-driven transcriptional termination occurs on IncRNA genes with either intronic or exonic miRNAs (Supplementary Table 4c), thus suggesting that splicing in IncRNA transcripts is not functionally important. It is possible that spliced transcripts generated from IncRNA genes, such as pri-miR-122, are simply a byproduct of transcription; because the splicing machinery is recruited cotranscriptionally, some default recognition and processing of splice sites occurs. IncRNAs remain relatively uncharacterized, and it is possible that multiple classes of IncRNA may exist that use different mechanisms of RNA processing.

The use of Microprocessor-driven 3′-end formation by a subset of pri-miRNA genes raises the question of why these genes use this mechanism. A major consequence is the generation of pA− Inc-pri-miR-122 transcripts that are rapidly degraded (Figures 1 and 2). This suggests that this mode of termination might have evolved to limit the production of spliced cytoplasmic pri-miRNA transcripts. The concurrent generation of a spliced miRNA that occurs from intronic coding miRNA genes may not be desirable for all miRNA genes. miR-122 expression is very high in liver cells, with an average of 66,000 copies per cell, and the accumulation of a similar number of copies of the spliced host transcript might be problematic. The miRNA genes that we identify as using the Microprocessor-mediated transcriptional termination pathway include others that can be highly expressed and are biologically important, such as the miR-17–92a (MIR17HG) cluster, which is highly expressed in embryonic cells and cancer.

Our genome-wide analysis identified a few examples of Microprocessor depletion leading to transcriptional interference with downstream genes in either tandem or convergent orientation (Figures 7 and Supplementary Figure 8). For two such genes, we observed a strong reduction in mRNA and protein levels after DGCGR8 knockdown, confirming that gene expression is affected. It is unclear how widespread this transcriptional interference is. It is likely that it will be influenced by the relative transcriptional activity of the upstream and downstream genes, the distance between them and their chromatin context, and that tissue-specific examples will exist. Because Drosha expression changes in different tissues and during differentiation, transcriptional interference may also differ in these situations.

In conclusion, we have identified a new RNase III–mediated transcriptional termination pathway in mammalian cells. This Microprocessor mechanism provides an alternative to CPA, generating unspliced and spliced transcripts of multiple kilobases, and is specific to IncRNA. We propose a model in which Microprocessor-driven transcriptional termination of Inc-pri-miRNAs prevents readthrough and interference with downstream genes. At the same time, a miRNA is produced, and the host transcript is not polyadenylated and so is rapidly turned over in the nucleus (Figure 8c). This may be valuable for highly expressed miRNAs such as miR-122, allowing the cell to achieve high levels of miRNAs without concomitant generation of high levels of unwanted host transcripts.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. RNA-sequencing data have been deposited in the Gene Expression Omnibus database under accession number GSE58838.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of the Proudfoot laboratory for advice and encouragement. This work was supported by a Programme grant (991805/Z/10/Z) from the Wellcome Trust, a European Research Council Advanced Award (339270-polyloop) to N.J.P. and a Biotechnology and Biological Sciences Research Council David Phillips Fellowship (BB/F02360X/1) to C.I.J. High-throughput sequencing was performed by the Genomics group at The Oxford Wellcome Centre for Human Genetics.

AUTHOR CONTRIBUTIONS

A.D. and C.I.J. performed molecular biology experiments; S.D. performed bioinformatics analysis; A.D., N.J.P. and C.I.J. designed the experiments and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
13. Proudfoot, N.J. Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782 (2011).
14. Mandel, C.R. *et al.* Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* **444**, 953–956 (2006).
15. West, S., Gromak, N. & Proudfoot, N.J. Human 5′ → 3′ exoribonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522–525 (2004).
16. Kim, M. *et al.* The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517–522 (2004).
17. Zhang, Z., Fu, J. & Gilmour, D.S. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. *Genes Dev.* **19**, 1572–1580 (2005).
18. Zhang, Z. & Gilmour, D.S. Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Nat. Cell Biol.* **21**, 65–74 (2006).
19. Kawasumi, J., Miscoho, H., Bragila, P., Rondon, A. & Proudfoot, N.J. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev.* **22**, 1082–1092 (2008).
20. Rondon, A.G., Miscoho, H.E., Kawasumi, J. & Proudfoot, N.J. Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Mol. Cell* **36**, 88–98 (2009).
21. El Haj, A., Koper, M., Kufel, J. & Tollervey, D. Efficient termination of transcription by RNA polymerase I requires the 5′ exoribonuclease Rat1 in yeast. *Genes Dev.* **22**, 1069–1081 (2008).
22. Ghazal, G. *et al.* Yeast RnaIII triggers polyadenylation-independent transcription termination. *Mol. Cell* **36**, 99–109 (2009).
23. Chang, J. *et al.* miR-122, a mammalian liver-specific microRNA, is processed from *hcr* mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA Biol.* **1**, 106–113 (2004).
24. Esau, C. *et al.* miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab.* **3**, 87–98 (2006).
25. Elmén, J. *et al.* LNA-mediated microRNA silencing in non-human primates. *Nature* **452**, 896–899 (2008).
26. Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M. & Sarnow, P. Modulation of RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. *Genes Dev.* **19**, 1572–1580 (2005).
27. Li, Z.Y. *et al.* Positive regulation of hepatic miR-122 expression by HNF4α. *J. Hepatol.* **55**, 602–611 (2011).
28. Chien, C.H. *et al.* Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Res.* **39**, 9345–9356 (2011).
29. Wang, L.X. *et al.* DNA-RNA differences are generated in human cells within seconds after RNA exits polymerase II. *Cell Reports* **6**, 906–915 (2014).
30. Gromak, N. *et al.* Drosophila regulates gene expression independently of RNA cleavage function. *Cell Reports* **5**, 1499–1510 (2013).
31. Greger, I.H., Aranda, A. & Proudfoot, N. Balancing transcriptional interference and initiation on the GAL7 promoter of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **97**, 8415–8420 (2000).
32. Prescott, E.M. & Proudfoot, N.J. Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci. USA* **99**, 8796–8801 (2002).
33. Ribas, J. *et al.* A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Res.* **40**, 6821–6833 (2012).
34. Bracht, J., Hunter, S., Euchus, R., Weeks, P. & Pasquinelli, A.E. Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. *RNA* **10**, 1586–1594 (2004).
35. Wagschal, A. *et al.* Microprocessor, Setx, Xm2, and Rrp6 co-operate to induce premature termination of transcription by RNApol II. *Cell* **150**, 1147–1157 (2012).
36. Kolev, N.G., Yario, T.A., Benson, E. & Steitz, J.A. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3′-end maturation. *EMBO Rep.* **9**, 1013–1018 (2008).
37. Steinmetz, E.J., Conrad, N.K., Brew, D.A. & Corden, J.L. RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. *Nature* **413**, 327–331 (2001).
38. Arigo, J.T., Eyles, D.E., Carroll, K.L. & Corden, J.L. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* **23**, 841–851 (2006).
39. Vasilijeva, L., Kim, M., Mutschler, H., Buratowski, S. & Meinhard, A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **15**, 795–804 (2008).
40. Bailait, D. *et al.* Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005).
41. O’Reilly, D. *et al.* Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res.* **42**, 264–275 (2014).
42. Kim, M. *et al.* Distinct pathways for snoRNA and mRNA termination. *Mol. Cell* **24**, 723–734 (2006).
43. Skourti-Stathaki, K., Proudfoot, N.J. & Gromak, N. Human senataxin resolves RNA-DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
44. Auyeung, V.C., Ulltzy, I., McGearry, S.E. & Bartel, D.P. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* **152**, 844–858 (2013).
45. Conrad, T., Marsico, A., Gehre, M. & Orom, U.A. Microprocessor activity controls differential miRNA biogenesis in vivo. *Cell Reports* **9**, 542–554 (2014).
46. Wilusz, J.E., Freier, S.M. & Spector, D.L. 3′-end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919–932 (2008).
47. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. & Sharp, P.A. Promoter directionality is controlled by U1 snRNA and polyadenylation signals. *Nature* **499**, 360–363 (2013).
48. Ntini, E. *et al.* Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928 (2013).
49. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
50. Tligner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).
51. Mogilyansky, E. & Rigoutsos, I. The mir-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ.* **20**, 1603–1614 (2013).
52. Sperber, H. *et al.* miRNA sensitivity to Drosha levels correlates with pre-miRNA secondary structure. *RNA* **20**, 621–631 (2014).
ONLINE METHODS

PCR primers and siRNA sequences. Sequences are shown in Supplementary Table 5.

Antibodies. Anti-glycans (ab124886; Abcam), anti-OGFRL1 (SC-137654; Santa Cruz), anti-actin (A2103; Sigma), anti-Drosha (ab12286; Abcam), anti-DGCRR (10996-1-AP; Proteintech), anti-CPSP73 (A301-090A; Bethyl), anti-α-tubulin (T5168; Sigma), anti–histone H3 (ab1791; Abcam), anti–β-tubulin (ab6046; Abcam), and anti–Dicer (1D6; Abcam) were used. Validation of primary antibodies is provided on the manufacturers’ websites.

Plasmid constructs. Construction of βwt (formerly labeled HIVβ) has been described previously53. The pri-miR-122 WT construct was made by insertion of a genomic PCR fragment generated with the primers Pri122qF/Pri122qR on Huh7 genomic DNA. The resulting βwt PCR fragment was ligated into a cloning vector prepared by long-range PCR amplification of βwt with primers Open_BetaqR/B10 qF with Prime STAR HS DNA polymerase (Takara). The QuikChange II XL site-directed mutagenesis kit (Stratagene) was used to generate the various mutants with the following primer sets: Δ (DELTARTAqF/DELTARTAqR) and pA1mt (pA1lMqF/pA1lMqR).

Cell culture and transfection of siRNA and plasmids. HeLa, Huh7 and HepG2 cells were maintained in DMEM supplemented with 10% FBS. For Huh7 and HepG2 cell culture, 1% nonessential amino acids (Invitrogen) was also included in the culture medium. RNA was performed with Lipofectamine RNAiMax (Invitrogen), with siRNA delivered at 30 nM final concentration. A second siRNA treatment was performed at 48 h, and cells were harvested at 72 h after the first hit. Lipofectamine 2000 (Invitrogen) was used to deliver 0.1 µg pri-miR-122 plasmid and 0.025 µg pTAT per well of a 6-well plate.

RNA isolation and northern blot. RNA was isolated with TRIzol reagent (Ambion) according to the manufacturer’s instructions. Total RNA from human liver was purchased from Agilent Technologies (cat. no. 540017). Northern blotting was carried out with standard procedures on equal molar quantities of RNA. Membranes were probed with a random-primed 32P-labeled DNA fragment corresponding to nt 3077–3707 (exon probe) and nt 1563–2005 (intron probe) of pri-miR-122 in Ultrahyb (Ambion). A fragment corresponding to nt 685–1117 of 18S actin was used as a loading control. For miRNA northern blots, the small-RNA fraction was isolated by dissolving total RNA in 30 µL of TE buffer with addition of equal amounts of PEG solution (20% PEG 8000 and 2M NaCl). Samples were mixed and incubated for at least 30 min on ice; this was followed by centrifugation at 14,000g for 15 min and isopropanol precipitation of the supernatant. Small RNAs were run on an 18% polyacrylamide (19:1) urea gel and analyzed as described before54 with a 32P-end-labeled oligonucleotide complementary to miR-122.

Reverse transcription and real-time qPCR analysis. Total RNA was treated with DNase I (Roche) and reverse-transcribed with SuperScript Reverse Transcriptase III (Invitrogen) and random primers (Invitrogen). Real-time quantitative PCR (qPCR) was performed with 2× Sensimix SYBR mastermix (Bioline) and analyzed with the v3 chemistry. The primers used are listed in Supplementary Table 5.

Chromatin RNA isolation. The procedures for separating nuclear RNA into chromatin-associated and released fractions have been described before56. Chromatin-associated RNA from Huh7 cells was analyzed with RT-qPCR as detailed above. The primers used are listed in Supplementary Table 5.

ChIP analysis. ChIP analysis was carried out as described before57. 5 µg antibody (rabbit total RNAIP2, N-20; sc-899; Santa Cruz Biotechnology) was used per ChIP. Immunoprecipitated DNAs were used as templates for qPCR.

RNA sequencing. Chromatin-associated RNA was isolated as described above with the omission of tRNA in solution preparation. Nuclear poly(A)+ or poly(A)− RNAs were prepared as described above. RNA-seq was performed by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics, University of Oxford. RNA samples were ribodepleted with Ribo-Zero rRNA removal kit (Human/Mouse/Rat, Epitrend RHZH10142). Libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina, v1.0 (cat. no. E7420) according to the manufacturer’s guidelines, with the exception of use of our own 8-bp tags for indexing according to ref. 57. Libraries were sequenced on an Illumina HiSeq 2000 with 100-bp and 50-bp paired-end reads, v3 chemistry.

Bioinformatic analysis. Data sets. All human miRNAs and their genomic coordinates were obtained from miRBase release 20 (ref. 58). Annotation for protein-coding genes (GRCh37) was obtained from Ensembl release 74 (ref. 59). Annotation for lncRNAs was obtained from the GENCODE v7 catalog of human long noncoding RNAs49.

We separated the miRNAs into two groups: lnc-pri-miRNA (if the miRNA overlapped with an lncRNA) and protein-coding pri-miRNAs (if the miRNA overlapped with an Ensembl protein-coding gene) on the basis of genomic coordinates, taking the strand orientation into account. miRNA-containing genes having length ≥2000 bp were considered for this study. This resulted in 112 lnc-pri-miRNAs and 967 protein-coding pri-miRNAs. An additional 18 lnc-pri-miRNAs were added to this list, which mapped to genes having lincRNA as Ensembl gene biotype and gene length ≥2000 bp.

Determination of the genomic distribution of miRNA as belonging to protein-coding genes and long noncoding genes was based on Ensembl gene biotype annotation (Supplementary Fig. 1).
Mapping of sequencing reads. Paired-end reads for each sample were mapped to the human genome reference assembly GRh37/hg19 (build 37.2, February 2009) with Bowtie2 alignment software. Prior to alignment, the first 12 nt were trimmed from all the reads, owing to the low quality of the bases. Uniquely mapped reads with no more than two mismatches were retained for further analysis. For nuclear poly(A)+ data, we filtered out reads that had stretches of eight or more genomically encoded As at their 3′ ends. A statistical summary of read alignments can be found in Supplementary Table 6 for HeLa and Huh7 chromatin RNA–seq and HeLa nuclear poly(A)+ and poly(A)− RNA–seq.

Calculation of metagene profiles. We used the Ensembl gene annotation to define transcription start and end sites. In-house Perl and Python scripts were used to compute metagene profiles. To get a list of miRNAs expressed in HeLa cells, miRNA FPKM was calculated with Cufflinks on small RNA–seq data for HeLa cells, downloaded from the ENCODE Experiment Matrix available through the ENCODE Project at UCSC. This resulted in 15 IncRNAs expressing 34 miRNAs in HeLa cells. For the protein-coding data set, miRNA-containing genes with length ≥2 kb and gene-body RPKM ≥1 were considered. 545 genes were identified in this category.

Metagene profiles (Fig. 6) for control siRNA, Drosha siRNA and DGCR8 siRNA in HeLa cells were calculated for a region from the start of the miRNA host gene to 1 kb upstream of the start of the next downstream gene. For this, read counts were normalized to total sequencing depth. The region extending from the TSS to the TES was scaled to 4 kb, and the region from downstream of TES to 1 kb upstream of the start of the next gene was scaled to 10 kb. Normalized read counts were plotted for each 10-bp bin.

To investigate the profile surrounding TSSs upon Microprocessor depletion, we plotted normalized read counts across a region of 1 kb upstream and downstream of annotated TSSs for control siRNA, Drosha siRNA and DGCR8 siRNA in HeLa cells.

For Supplementary Tables 1 and 2, normalized read count (RPKM) was calculated over a region from the TES (3′ end) of the Inc-pri-miRNA to 1 kb upstream of the TSS of the next downstream gene for miRNA-containing IncRNA genes that are expressed (RPKM ≥1) in HeLa and Huh7 cell lines.

Classification of poly(A)+, poly(A)− and bimorphic transcripts. Classification of transcripts into poly(A)+, poly(A)− and bimorphic was done as previously described. Briefly, all expressed transcripts were classified as poly(A)+, poly(A)− and bimorphic predominant transcripts on the basis of their relative abundance, calculated with bases per kilobase of gene model per million mapped bases (BPKM) in the poly(A)+ and poly(A)− sample for each condition (control siRNA and DGCR8 siRNA) (Supplementary Table 3).

53. Dye, M.J. & Proudfoot, N.J. Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Mol. Cell 3, 371–378 (1999).
54. Pall, G.S. & Hamilton, A.J. Improved northern blot method for enhanced detection of small RNA. Nat. Protoc. 3, 1077–1084 (2008).
55. Dye, M.J., Gromak, N. & Proudfoot, N.J. Exon tethering in transcription by RNA polymerase II. Mol. Cell 21, 849–859 (2006).
56. West, S., Proudfoot, N.J. & Dye, M.J. Molecular dissection of mammalian RNA polymerase II transcriptional termination. Mol. Cell 29, 600–610 (2008).
57. Lambie, S. et al. Improved workflows for high throughput library preparation using the transposome-based nextera system. BMC Biotechnol. 13, 104 (2013).
58. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39, D152–D157 (2011).
59. Flicek, P. et al. Ensembl 2014. Nucleic Acids Res. 42, D749–D755 (2014).
60. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
61. Kent, W.J. et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).
62. Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G. & Chen, L.L. Genomewide characterization of non-polyadenylated RNAs. Genome Biol. 12, R16 (2011).
Corrigendum: Microprocessor mediates transcriptional termination in genes encoding long noncoding microRNAs

Ashish Dhir, Somdutta Dhir, Nick J Proudfoot & Catherine L Jopling
Nat Struct. Mol. Biol.; doi:10.1038/nsmb.2982; corrected online 12 March 2015

In the version of this article initially published online, the title 'Microprocessor mediates transcriptional termination in genes encoding long noncoding microRNAs' should have read 'Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs'. The error has been corrected for the print, PDF and HTML versions of this article.