Dhir, Ashish and Dhir, Somdutta and Proudfoot, Nick J. and Jopling, Catherine L. (2015) Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. Nature Structural and Molecular Biology, 22 . pp. 319-327. ISSN 1545-9993

Microprocessor mediates transcriptional termination in long noncoding microRNA genes.

Ashish Dhir[1], Somdutta Dhir[1], Nick J. Proudfoot[1], Catherine L. Jopling[2]

[1]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK. [2]School of Pharmacy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

Joint communicating authors: NJP (nicholas.proudfoot@path.ox.ac.uk) and CLJ (catherine.jopling@nottingham.ac.uk).

Abstract

MicroRNA (miRNA) play a major role in the post-transcriptional regulation of gene expression. Mammalian miRNA biogenesis begins with co-transcriptional cleavage of RNA polymerase II (Pol II) transcripts by the Microprocessor complex. While most miRNA are located within introns of protein coding genes, a substantial minority of miRNA originate from long non coding (lnc) RNA where transcript processing is largely uncharacterized. Here, by detailed characterization of liver-specific lnc-pri-miR-122 and genome-wide analysis, we show that most lnc-pri-miRNA do not use the canonical cleavage and polyadenylation (CPA) 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 novel RNase III-mediated, polyadenylation-independent mechanism of Pol II transcription termination in mammalian cells.

MicroRNA (miRNA) are ~22 nucleotide (nt) RNA that play a major role in the regulation of gene expression in most eukaryotic species. In humans, miRNA are thought to post-transcriptionally repress at least 60% of mRNA by binding to targets in the 3' untranslated region (UTR)[1]. miRNA show both tissue- and developmental stage-specific expression patterns. Tight regulation of biogenesis of individual miRNA occurs both at the level of transcription and at downstream processing steps, and is essential for normal development and physiology [2].

miRNA transcription is generally mediated by RNA polymerase II (Pol II), which synthesizes primary (pri-) miRNA transcripts that can extend to several kilobases in length and are typically capped and polyadenylated[3,4]. The mature miRNA is located within a hairpin structure that is recognized by the Microprocessor complex, comprising the dsRNA binding protein DGCR8 and the RNase III endonuclease Drosha[5]. Microprocessor cleavage releases a ~70nt hairpin precursor (pre-) miRNA, which is exported and processed by Dicer to generate a mature miRNA[5]. In a few cases, pre-miRNA are generated independent of Microprocessor, either from small debranched introns (mirtrons)[6], or from unusually short Pol II transcripts[7].

Similar to RNA processing events required for mRNA generation, excision of pre-miRNA is co-transcriptional[8,9]. Most miRNA derive from introns of protein coding genes, where co-transcriptional Microprocessor cleavage does not inhibit splicing, allowing co-expression of miRNA and mRNA from the same host transcript[10,11]. In contrast, Drosha processing of a miRNA located in a protein coding gene exon can inhibit production of the spliced host mRNA[12]. Importantly, 17.5% of miRNA are located in long non coding (lnc)RNAs (Supplementary Fig. 1), which we define as lnc-pri-miRNA. Processing of these transcripts has not been characterized.

Transcriptionalal termination of Pol II transcribed genes is tightly coupled to 3' end processing[13]. mRNA 3' end formation occurs co-transcriptionally by a Pol II associated cleavage and polyadenylation (CPA) mechanism. This involves recognition of the polyadenylation site (PAS), including a canonical AAUAAA sequence and additional, more degenerate, sequence elements. RNA cleavage occurs 10-30 nucleotides downstream of the AAUAAA sequence, followed by the addition of a polyA (pA) tail to the resulting 3' end. Multiple protein complexes are required for this process with endonucleolytic cleavage at the PAS being mediated by cleavage and pA specific factor, CPSF-73[14]. CPA creates an entry site for the 5'-3' exonuclease Xrn2 (Rat1 in yeast) which acts as a 'torpedo', degrading the nascent transcript and contributing to the displacement of Pol II[15,16], while the CPA factor Pcf11 also contributes to transcription termination by associating with the Pol II CTD and dismantling the elongation complex[17,18]. In budding yeast, the RNase III protein Rnt1 can mediate polyadenylation-independent 3' end formation on pre-mRNA as well as Pol I transcripts[19-22], but a similar mammalian pathway has not been identified.

In this study, we characterize the processing of an exonic lnc-pri-miRNA, *hcr*. This gene encodes miR-122, which is highly and specifically expressed in the liver, where it is important for the regulation of cholesterol metabolism and hepatitis C virus (HCV) replication[23-26]. We find that Microprocessor cleavage at the pre-miR-122 hairpin mediates transcription termination, generating non-polyadenylated transcripts that are rapidly degraded in the nucleus. Using genome-wide nascent RNA-sequencing, we find that this mechanism of transcription termination occurs on most lnc-pri-miRNA, but not protein coding pri-miRNA. These lnc-pri-miRNA generally do not use CPA at canonical PAS even when the Microprocessor is depleted, revealing a fundamental difference in RNA processing of lncRNA and mRNA. We also identify a biological role for Microprocessor-mediated termination in preventing transcriptional interference with downstream genes. This identification of an

RNase III-mediated, CPSF-73-independent mechanism of Pol II transcriptional termination provides a mammalian parallel to the yeast Rnt1 pathway.

Results

Lnc-pri-miR-122 transcripts are capped but not polyadenylated.

Initially, to characterize the processing of a biologically important lnc-pri-miRNA, we focused on *hcr*, henceforth known as lnc-pri-miR-122 (Fig. 1a). As well as miR-122, lnc-pri-miR-122 generates two transcripts of ~4.8 and ~1.9 kb identified by northern analysis of both human liver and the human hepatocellular carcinoma cell line Huh7, but not of the non-miR-122-expressing HeLa and HepG2 cells (Fig. 1b). Northern blotting using intron- and exon-specific probes indicates that the larger transcript is unspliced, while the smaller transcript corresponds to inefficiently spliced RNA that lacks an internal 3 kb intron (Fig. 1c). The transcript size indicates that the 3' end lies close to the pre-miR-122 hairpin, ~2.5kb upstream of a previously identified polyadenylated 3' end[27]. We did not detect any longer lnc-pri-miR-122 transcripts. Immunoprecipitation with an antibody directed against the m^7G cap demonstrates that lnc-pri-miR-122 is capped, similar to GAPDH mRNA and in agreement with existing CAGE data[28] (Fig. 1d). However, quantitative RT-PCR (qPCR) and northern analysis of polyA-selected RNA indicates that lnc-pri-miR-122 is non-polyadenylated, in contrast to GAPDH mRNA, but similar to U6 snRNA (Fig. 1e).

The 3' end of lnc-pri-miR-122 is generated by Microprocessor cleavage.

To map lnc-pri-miR-122 3' ends at nucleotide resolution, we developed a pA tail-independent 3'RACE technique based on *in vitro* tailing of the RNA by poly(A) polymerase (PAP). Lnc-pri-miR-122 3' ends were mapped 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 following Drosha cleavage (Fig. 2b, Supplementary

Fig. 2). Both unspliced and spliced lnc-pri-miR-122 are not significantly exported to the cytoplasm (Fig. 2c), and are rapidly degraded (Fig. 2d), as expected for non-polyadenylated transcripts.

These above results suggest that the lnc-pri-miR-122 3' end is generated by Drosha cleavage and not by CPA. To further characterize the role of the Microprocessor in lnc-pri-miR-122 3' end formation, we compared the effects of siRNA-mediated knockdown in Huh7 cells of DGCR8 versus the CPA endonuclease CPSF-73. Depletion of both proteins was effective (Fig. 3c). qPCR analysis of RNA isolated from nuclear chromatin, which is enriched in nascent transcripts[29], was used to compare profiles across lnc-pri-miRNA-122 and the protein coding gene GAPDH. The level of each qPCR product was normalized to the intron 1 product as a measure of basal nascent transcription, and is shown relative to control siRNA-treated cells. These nascent transcript profiles showed a clear increase in level downstream of the pre-miR-122 sequence in DGCR8 but not CPSF-73-depleted cells (Fig. 3a). To further test for this affect, we compared the effect of depleting either Drosha or Dicer on lnc-pri-miRNA-122 termination (Supplementary Fig. 3). Drosha, like DGCR8, depletion elicited readthough transcription while Dicer depletion had no effect. This argues against indirect effects of microRNA depletion on termination and strongly suggests a direct role for the Microprocessor in the termination process of this lncRNA. In contrast, GAPDH nascent transcripts show no effect of DGCR8 depletion, but as expected show increased level downstream of the PAS indicating a strong termination defect following CPSF-73 depletion (Fig. 3b). These results were confirmed by nuclear run on (NRO) analysis, which again shows that DGCR8 knockdown causes a termination defect for lnc-pri-miR-122 but not GAPDH, while CPSF-73 knockdown inhibits termination in GAPDH but not lnc-pri-miR-122 (Fig. 3a,b). In sum, we establish that the Microprocessor dictates lnc-pri-miR-122 3' end formation and transcriptional termination, in contrast to GAPDH which relies on the orthodox CPA complex.

Lnc-pri-miR-122 transcriptional termination does not occur efficiently in the absence of Microprocessor cleavage.

The transcriptional readthrough we observe following DGCR8 knockdown implies that lnc-pri-miR-122 does not switch to the efficient CPA mechansism of transcriptional termination when the Microprocessor mechanism is inhibited. To address this question directly, we used poly(A) fractionation to determine the poly(A) status of lnc-pri-miR-122 in Huh7 cells with or without DGCR8 knockdown. Similar to U6 snRNA, lnc-pri-miR-122 transcripts remain p(A)-, even when their 3' end formation is compromised by DGCR8 depletion. In contrast GAPDH mRNA is strongly p(A)+ under both conditions (Fig. 4a). This indicates that efficient CPA does not occur on lnc-pri-miR-122 when Microprocessor-dependent termination is inhibited, despite the presence of several canonical PAS. To investigate the effects of Microprocessor depletion on lnc-pri-miR-122 transcriptional termination in more detail, we isolated chromatin-associated RNA from Huh7 cells and analyzed it by next generation sequencing (chromatin RNA-seq). We observed extensive transcriptional readthrough for over 5 kilobases downstream of the pre-miR-122 hairpin following DGCR8 depletion (Fig. 4b).

The failure of Pol II transcribing lnc-pri-miR-122 to recognize PAS even when Microprocessor cleavage is inhibited is surprising, and suggests that Pol II might be recruited to the endogenous lnc-pri-miR-122 promoter in a CPA refractory form. To investigate this further, we generated a cis-deletion of the pre-miR-122 hairpin in a plasmid context. We cloned lnc-pri-miR-122 under the control of the human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter with or without pre-miR-122 hairpin deletion (Fig. 5a). The resulting WT and Δ plasmids were transfected into HeLa cells, which do not express endogenous lnc-pri-miR-122 (Fig. 1b), together with a plasmid encoding the Tat transcriptional activator. We confirmed that mature miR-122 is expressed from the WT but not Δ plasmid (Fig. 5b). Unspliced and spliced

lnc-pri-miR-122 transcripts generated from the WT plasmid are the same size as the endogenous transcripts in Huh7 cells (Fig. 5c), indicating that the site of 3' end formation is independent of promoter or cell type. Deletion of the pre-miR-122 hairpin leads to loss of the correct sized lnc-pri-miR-122 transcripts, which are replaced by unspliced and spliced transcripts migrating at a higher molecular weight (Fig. 5d). siRNA-mediated depletion of Drosha or DGCR8 leads to the same shift in size, supporting the idea that an alternative mechanism of 3' end generation occurs downstream when Microprocessor cleavage is inhibited (Fig. 5e). Poly(A) fractionation indicated that lnc-pri-miR-122 RNA generated from the WT plasmid is p(A)-, similar to U6 snRNA, but becomes p(A)+ when DGCR8 is depleted, similar to GAPDH mRNA (Fig. 5f). The 3' end of Δ RNA was mapped by 3'RACE to a PAS 91nt downstream of the pre-miR-122 hairpin (pA1, Fig 5a). Mutagenesis indicated that this PAS is necessary for 3' end generation in Δ, but not WT lnc-pri-miR-122 transcripts (Fig. 5g). This confirms that in a plasmid context, lnc-pri-miR-122 3' ends are generated by the Microprocessor, but in the absence of this processing, CPA occurs at a site downstream. Importantly, this is in contrast to the endogenous transcripts, where transcriptional readthrough occurs when Microprocessor cleavage is inhibited, and the PAS is not used (Fig. 3, 4). This suggests that for the endogenous lnc-pri-miRNA-122 gene, chromatin context or promoter specificity prevents efficient CPA in the absence of Microprocessor cleavage.

Genome-wide analysis shows that most lnc-pri-miRNA use Microprocessor-mediated transcriptional termination.

Having demonstrated that chromatin RNA-seq is an effective method of identifying defects in transcriptional termination following Microprocessor depletion (Fig. 4b), we extended this analysis to HeLa cells on a genome-wide scale. HeLa cells were chosen because they express a relatively high number of miRNA. We found that either Drosha or DGCR8 depletion, by siRNA treatment (Fig. 6e), results in transcriptional readthrough in most of the expressed lnc-pri-miRNAs (Supplementary Table 1). MIR181A1HG is given

as a specific example (Fig. 6a), while metagene analysis shows a general termination defect with transcription extending more than 10 kb downstream of lnc-pri-miRNA genes following Microprocessor depletion (Fig. 6b). MIR17HG is shown as a further example (Supplementary Fig. 4a). A few lnc-pri-miRNA do not shown transcriptional readthrough following Microprocessor depletion (shaded area, Supplementary Table 1); MIRLET7BHG is shown as an example (Supplementary Fig. 4b). Importantly, Dicer knockdown does not affect lnc-pri-miRNA transcriptional termination, again confirming that the effects of Microprocessor depletion are direct and not due to loss of mature miRNA (Supplementary Fig. 5).

In marked contrast, protein coding genes that harbor intronic pre-miRNAs show no termination defect following Microprocessor depletion. Rather, intronic sequence containing pre-mRNA is stabilized (higher reads) as shown for MCM7 (Fig. 6c), which has three miRNA (miR-25, 93 and 106b) in its penultimate intron. Presumably this selective intron stabilization is caused by the loss of Drosha-mediated co-transcriptional cleavage[8]. Metagene analysis reveals that Microprocessor activity generally has no effect on transcriptional termination for these protein coding miRNA genes (Fig. 6d). This implies a functional difference between the processing of miRNA in the introns of protein coding as compared to lnc-pri-miRNA genes. Protein coding genes utilize a CPA-mediated termination process irrespective of internal Microprocessor cleavage. In contrast lnc-pri-miRNA rely on the Microprocessor for efficient termination. Chromatin RNA-seq analysis in Huh7 cells also identified transcriptional readthrough in lnc-pri-miRNA, but not protein coding pri-miRNA, following DGCR8 depletion. Although the pattern of pri-miRNA expression differs between Huh7 and HeLa, many of the same lnc-pri-miRNA were affected in both cell lines (Supplementary Table 2).

We also investigated the effect of Microprocessor depletion on levels of transcription at gene 5’ ends (TSS). While no change in transcript level was

observed for lnc-pri-miRNA genes, protein coding genes hosting pre-miRNA showed transcript reduction following DGCR8 and especially Drosha knockdown (Supplementary Fig. 6a, b). This suggests that the Microprocessor may have a positive influence on gene transcription, as previously noted[30]. It also suggests that transcriptional initiation differs at lnc-pri-miRNA and protein coding pri-miRNA gene promoters. As a further control for the quality of our chromatin RNA-seq data we show that duplicate libraries display high correlation (Supplementary Fig. 7).

Microprocessor cleavage of lnc-pri-miRNA prevents transcriptional interference with downstream genes.

We identified specific examples of lnc-pri-miRNA genes in which the transcriptional readthrough induced by Microprocessor depletion extended into a downstream protein coding gene, either in convergent or tandem orientation. We reasoned that such readthrough transcription might downregulate the invaded gene by a transcriptional interference mechanism[31]. For the tandem MIR17HG-GPC5 locus, Microprocessor depletion caused the -MIR17HG transcript to extend over 20 kb, reading into GPC5 (Fig. 7a and Supplementary Fig. 8). Chimeric transcripts were readily detected, as was a significant reduction in RNA levels from GPC5 exon 1 (Fig. 7b,d and Supplementary Fig. 8). We also tested the consequence of this interference effect on GPC5 gene expression. Significantly both GPC5 mRNA and protein levels were reduced 5 fold (Fig, 7e,f). These data demonstrate a clear transcriptional interference effect caused by loss of Microprocessor-mediated termination. For the convergent OGFRL1-LINC00472 locus, loss of Microprocessor caused - LINC00472 to read through into OGFRL1, again causing transcriptional down-regulation (Fig. 7c,d). OGFRL1 mRNA levels dropped 5 fold (Fig. 7e) while protein levels were reduced 2 fold (Fig. 7f). Possibly this protein has a higher stability than GPC5. These data imply that convergent transcription can also induce gene inactivation, possibly by Pol II collision effects[32]. Notably Dicer depletion had no effect on either mRNA level

(Fig. 7e), indicating that the effects of DGCR8 knockdown are due to transcriptional interference and not miRNA-mediated mRNA destabilization.

Most lnc-pri-miRNA remain poly(A)- following Microprocessor depletion. Similar to endogenous lnc-pri-miR-122, Microprocessor-terminated lnc-pri-miRNA appear to be insensitive to the presence of cryptic PAS, invariably present within their gene and 3' flanking regions. To further investigate the use of PAS in pri-miRNA, we also performed nuclear poly(A)+ and poly(A)- RNA-seq in HeLa cells both with or without DGCR8 knockdown. We found that the majority of lnc-pri-miRNA exist as predominantly p(A)- transcripts, and those that show extensive readthrough upon loss of Microprocessor continue to remain p(A)- (Supplementary Table 3). MIR17HG is shown as a specific example (Fig. 8a). It is remarkable that for these Pol II transcripts PAS remain opaque to RNA processing by the CPA complex. However, a few lnc-pri-miRNA do utilize PAS to some extent, especially following Microprocessor inactivation. Thus MIRLET7BHG shows a switch from mainly p(A)- RNA to p(A)+ following Microprocessor depletion (Fig. 8b, Supplementary Table 3). Interestingly for this gene, a canonical PAS positioned immediately downstream of pre-miR-let7b promotes efficient termination if Microprocessor termination is inhibited. This is similar to the switch to CPA at a downstream PAS that we observe in ectopically expressed lnc-pri-miR-122 (Fig. 5), and indicates 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, 2). Genome-wide analysis indicates that transcriptional termination by the Microprocessor is a feature shared with most other lnc-pri-miRNA genes (Fig. 6, Supplementary Table 1). While we focus on HeLa cells in this study, we have also carried out

chromatin RNA-seq in Huh7 cells and find Microprocessor-driven transcriptional termination is also used by most lnc-pri-miRNA expressed in this cell line (Supplementary Table 2).

Previous evidence indicated that pri-miRNA are typical Pol II transcripts, with a 5' cap and a 3' pA tail. This is clearly established for pre-mRNA in which miRNA are located in introns[8], but is also true of the few lnc-pri-miRNA for which the 3' end has been characterized, such as pri-miR-21[4,33] and *C. elegans let-7*[34]. Our genome-wide analysis confirms that CPA does occur in a minority of lnc-pri-miRNA (Supplementary Tables 3), but we find that the Microprocessor-mediated termination mechanism predominates (Supplementary Tables 1-2). A previous indication that such a mechanism was possible came from a study showing that Drosha processing of pre-miRNA hairpins can attenuate downstream transcription by providing an entry site for Xrn2. However, these experiments were carried out using plasmid constructs that lacked a poly(A) signal, and did not provide evidence that Microprocessor cleavage could actually replace CPA as a mode of transcriptional termination[9]. A role for Microprocessor cleavage at the human immunodeficiency virus (HIV) long terminal repeat (LTR) in allowing Xrn2 entry and preventing productive transcription elongation also indicates that Drosha cleavage can disrupt the transcription machinery[35], 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 it is limited to lnc-pri-miRNA. This departs from the clear current consensus that pri-miRNA are typical capped and polyadenylated transcripts (see recent review[5]), a view derived from analysis of protein coding pri-miRNA and confirmed for protein coding genes by our genome-wide analysis.

The Microprocessor termination pathway adds to a short list of non-canonical mechanisms of Pol II transcriptional termination. Termination of histone mRNA does not involve polyadenylation, similar to pri-miR-122, but requires cleavage by CPSF-73, in common with CPA[36]. In yeast, the Nrd1-Nab3-Sen1 pathway terminates Pol II transcription of small nuclear (sn)RNA, small nucleolar (sno)RNA and cryptic unstable transcripts (CUTs)[37-39]. There is currently no evidence for the Nrd1 pathway in mammals, which use the multi-subunit Integrator complex for transcriptional termination and subsequent processing of snRNAs[40,41]. Importantly, both the Nrd1 and Integrator pathways are only used for termination 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 Rnt1-mediated termination in budding yeast. In addition to functions in rRNA biogenesis, Rnt1 cleavage leads to pA-independent termination of Pol II transcription when CPA fails, acting as a fail-safe mechanism to prevent read-through transcription and subsequent transcriptional interference[20,22]. Rnt1-terminated transcripts are rapidly degraded, similar to lnc-pri-miR-122. Transcriptional termination following either Rnt1 cleavage or CPA occurs as a result of Rat1/Xrn2 degradation of the nascent transcript downstream of the cleavage site[15,16,22]. However we did not observe any effect of Xrn2 depletion on pri-miR-122 transcriptional termination (data not shown), in contrast to the role for Xrn2 in RNA degradation following Drosha cleavage of an intergenic clustered pri-miRNA[9]. It is possible that other nucleases or termination factors such as Pcf11[18,42] or the mammalian ortholog of Sen1, Senataxin[43], may be involved. Further research will be required to elucidate the mechanism of Pol II termination following Microprocessor cleavage on pri-miR-122. Of note, we observe a variation in levels of chromatin RNA-seq signal downstream of the pre-miRNA hairpin following DGCR8 knockdown among different Microprocessor-terminated lnc-pri-miRNA. For example, the amplitude of

readthrough transcription is higher in MIR181A1HG and MIR17HG (Fig. 6, Supplementary Fig. 4,Supplementary Table 1-2) than on lnc-pri-miR-122 (Fig. 4b, Supplementary Table 2). This loss of transcription may be the result of low-level Microprocessor-driven termination mediated by residual DGCR8 following 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-miRNA, 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-miRNA is 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 may be affected by the Microprocessor cleavage event itself. For example, protein cofactors are known to assist in Microprocessor release of specific pre-miRNA[5], while sequence features surrounding the pre-miRNA hairpin can influence the efficiency of processing[44]. Many of the pri-miRNA that use this mechanism contain clustered pre-miRNA hairpins, which raises the interesting question of whether Microprocessor cleavage at a specific pre-miRNA drives transcription termination. As some transcriptional readthrough occurs following Microprocessor cleavage, similar to the continued Pol II transcription following cleavage at a PAS in CPA-dependent termination, it is not possible to precisely define which cleavage event drives termination based on our chromatin RNA-seq data. A recent study used a similar chromatin RNA-seq approach to define a ‘Microprocessing index’ (MPI) that shows variable cleavage efficiency for different pre-miRNA[45]. Of the 13 HeLa lnc-pri-miRNAs detected in this study, 11 contain a pre-miRNA hairpin with MPI <-1.0 indicating efficient co-transcriptional processing, and 7 of these contain a

hairpin with MPI <-3.0, indicating highly efficient processing[45]. Although the pool of lnc-pri-miRNA that we detect in HeLa is too small to draw statistically robust conclusions, this raises the possibility that rapid Microprocessor cleavage may be important for transcriptional termination.

By genome-wide analysis, we find that Microprocessor-driven transcriptional termination is used by many lnc-pri-miRNA (73%), but not protein coding genes containing pre-miRNA (Fig. 6 and Supplementary Tables 1 and 2). Overall our results demonstrate a fundamental difference in RNA processing between lncRNA and protein coding transcripts. The Microprocessor mechanism has parallels to another CPA-independent mechanism of transcriptional termination used by the lncRNA MALAT1, where 3' end formation and concomitant release of a small RNA is mediated by the tRNA biogenesis endonucleases RNase P and RNase Z[46]. However, lncRNA derived from bi-directional firing at protein coding gene promoters, known as upstream antisense (ua)RNA transcripts, terminate transcription by CPA and tend to use promoter proximal PAS[47,48]. This is related to enrichment in PAS and depletion in U1 splice sites in upstream antisense regions; it has been demonstrated that U1 snRNA recruitment acts to block nearby cryptic PAS[49]. The difference between p(A)- lnc-pri-miRNA transcripts as described in this study and CPA-competent lncRNA 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 lnc-pri-miRNA promoters may form a different type of Pol II elongation complex that is CPA-non-responsive but Microprocessor active. A role for the promoter is supported by our observation that most lnc-pri-miRNA, including lnc-pri-miR-122, do not use CPA even when the Microprocessor is depleted, instead showing extensive transcriptional readthrough and remaining p(A)- (Fig. 8a, Supplementary Table 3). In contrast, ectopically expressed lnc-pri-miR-122 uses Microprocessor cleavage to mediate transcriptional termination but switches

to CPA when this mechanism is inhibited (Fig. 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 that MIRLET7BHG uses Microprocessor cleavage to terminate transcription but switches to CPA following Microprocessor depletion (Fig. 8b).

LncRNA genes have generally been thought to be similar to mRNA genes, with similar chromatin profiles, transcriptional regulation and splice signals. However, lncRNA genes tend to have fewer and longer exons than mRNA genes, and there is a trend for less efficient splicing of lncRNA than of mRNA[50,51]. The lack of CPA we observe on most lnc-pri-miRNA implies that this process is similarly inefficient on this class of lncRNA. As splicing and CPA function to generate a stable cytoplasmic coding transcript, there would be little need for these processes to occur on most lncRNA. We find that Microprocessor-driven transcriptional termination occurs on lncRNA genes with either intronic or exonic miRNA, suggesting that splicing in lncRNA genes is not functionally important. Possibly spliced transcripts generated from lncRNA genes such as pri-miR-122 are simply a by-product of transcription; as the splicing machinery is recruited co-transcriptionally, some default recognition and processing of splice sites occurs. LncRNA remain relatively uncharacterized, and it is possible that multiple classes of lncRNA 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 p(A)- pri-miR-122 transcripts that are rapidly degraded (Fig. 1, 2). This suggests that the evolutionary purpose of this mode of termination might be to limit the production of a spliced cytoplasmic pri-miRNA transcript. The concurrent generation of a spliced mRNA 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[23], and the accumulation of a similar number of copies of the spliced host transcript might be problematic. The miRNA genes 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[52].

Our genome-wide analysis identifies a few examples of Microprocessor depletion leading to transcriptional interference with downstream genes (Fig. 7, Supplementary Fig. 8). This can occur for genes that have tandem or convergent locations relative to the lncRNA that uses Microprocessor for transcriptional termination. For two such genes, we observed a strong reduction in mRNA and protein levels following DGCR8 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. As Drosha expression changes in different tissues and during differentiation[53], transcriptional interference may also differ in these situations.

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

Genomic data accession codes

Expression Omnibus accession number [GEO:GSE58838] at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=kzaxguooffwnvsn&acc=GSE58838.

Acknowledgements

We thank members of the NJP lab for advice and encouragement. This work was supported by a Programme grant from the Wellcome Trust and an ERC Advanced Award to NJP and a BBSRC David Phillips Fellowship to CLJ. High throughput sequencing was performed by the Genomics group at The Oxford Wellcome Centre for Human Genetics.

Author Contributions

AD and CLJ performed molecular biology experiments; SD performed bioinformatics analysis; AD, NJP and CLJ designed the experiments and wrote the manuscript.

Figure Legends

Figure 1: Lnc-pri-miR-122 transcripts are capped but not polyadenylated.

a. Gene map showing exon-intron structure and 3' positioned pre-miR-122 hairpin. Position of primers and probe are indicated. **b.** Northern blot showing spliced and unspliced lnc-pri-miR-122 from human liver and three cell lines as indicated. Levels of γ-actin mRNA and mature miR-122 were also measured. Note reduced levels of liver RNA were employed. **c.** Northern blot detecting unspliced and spliced lnc-pri-miR-122 in human liver RNA using intron and exon-specific probes. **d.** Levels of m^7G-cap immunoselected RNA measured by qPCR as indicated. **e.** Proportion of polyA+ RNA relative to polyA- for transcripts indicated measured by qPCR, and for lnc-pri-miR-122 and γ-actin measured by northern blotting. Error bars are SD.

Figure 2: 3' end mapping, subcellular distribution and rapid turnover of lnc-pri-miR-122.

a. Position of primers for qPCR and 3' end mapping in lnc-pri-miR-122. **b.** PolyA polymerase (PAP)-dependent 3'end mapping using 3' RACE. Position of 3' RACE PCR product shown by gel fractionation and location of predominant 3' end cleavage products marked by arrows (see Supplementary Fig. 2). **c.** Pri-miR-122 and GAPDH mRNA distribution were determined between cell fractions (WC denotes whole cell, N nuclear and C cytoplasmic) by RT-PCR using indicated primers. Western blot shows purity of nuclear and cytoplasmic fractions by use of cytoplasmic and nuclear specific protein

antibodies. **d.** RNA stability following actinomycin D (ActD) inhibition of transcription at various time points measured by qPCR of lnc-pri-miR-122 transcripts versus GAPDH mRNA. All experiments use Huh7 cells. Error bars are SD.

Figure 3: Microprocessor defines transcriptional termination of lnc-pri-miR-122.

a. and b. Mapping nascent transcription across lnc-pri-miR-122 versus GAPDH in Huh7 cells. Gene maps show positions of amplicons used in analysis of chromatin RNA levels (left) and bromo UTP-labelled nuclear run on RNA (right). **c.** Western blot showing effective depletion of DGCR8 and CPSF-73 by siRNA transfection in Huh7 cells. Error bars are SD.

Figure 4: Microprocessor depletion leads to generation of poly(A)-transcriptional readthrough products on lnc-pri-miR-122

a. qPCR analysis of lnc-pri-miR-122 versus U6 or GAPDH mRNA controls using poly(A)+ or – fractionated RNA from siRNA-treated Huh7 cells. Error bars are SD. **b.** Chromatin RNA-seq analysis of lnc-pri-miR-122 in control or DGCR8 siRNA-treated Huh7 cells. Boxed 3' region is zoomed in and shown on right hand side. Position of miRNA (miR-122) is shown by red vertical line.

Figure 5: Ectopically expressed lnc-pri-miR-122 switches to CPA when Microprocessor activity is inhibited.

a. Schematic of lnc-pri-miR-122 expression construct, with locations of northern probe, qPCR product and PAS. WT and Δ plasmids were generated with and without the pre-miR-122 hairpin. **b.** Northern blot showing that mature miR-122 is expressed from the WT, but not Δ, plasmid following transfection into HeLa cells. **c.** Northern blot showing spliced and unspliced lnc-pri-miR-122 transcripts generated from the WT plasmid are the same size as endogenous transcripts in Huh7 cells. **d.** Northern blot showing spliced and unspliced lnc-pri-miR-122 transcripts generated from the Δ plasmid are larger

than the WT transcripts. **e.** Northern blot showing transcripts generated from the WT plasmid increase in size following Drosha or DGCR8 depletion. **f.** qPCR analysis of poly(A)+ or - fractionated RNA extracted from siRNA-treated HeLa cells transfected with the WT plasmid. Lnc-pri-miR-122 is predominantly pA- in control siRNA-treated cells but pA+ following DGCR8 depletion. **g.** Northern blot showing that mutation of pA1 does not affect migration of WT lnc-pri-miR-122, but leads to loss of unspliced and spliced transcripts generated from the Δ plasmid. Error bars are SD.

Figure 6: Microprocessor dependent chromatin RNA-seq profiles across pri-miRNA from HeLa cells.

a. Reads (in RPKM) across (MIR181A1HG) showing readthrough profiles following Microprocessor depletion. **b.** Metagene analysis of all expressed lnc-pri-miRNA. **c.** Reads across a protein coding gene (MCM7) harboring an intronic miRNA cluster showing intron accumulation following Microprocessor depletion **d.** Metagene analysis of all expressed protein coding genes harboring intronic miRNA. Direction of transcription indicated by green arrow and positions of miRNA by red vertical lines in **a** and **c**. Metagene profiles show transcription unit (between transcription start site, TSS and transcription end site, TES) followed by 10 kb of 3' flanking region in **b** and **d** ($P < 0.0001$, Mann–Whitney U-test, for both cases). **e.** Western blot showing effective depletion of Drosha and DGCR8 by siRNA transfection in HeLa cells.

Figure 7: Microprocessor dependent termination prevents transcriptional interference.

a. Chromatin RNA-seq profiles across MIR17HG-GPC5 tandem gene locus showing readthrough transcription following Microprocessor depletion. Black arrow denotes reduction in GPC5 exon 1 peak following Microprocessor depletion. **b.** qPCR analysis of chimeric transcripts versus GPC5 exon 1. RT primer (arrowhead) in exon 1 and PCR amplicons indicated by black bars.

Error bars are SD. **c.** Chromatin RNA-seq profiles across convergent OGFRL1- LINC00472 gene locus. **d.** Read quantification for protein coding genes subject to transcriptional interference following Microprocessor depletion. **e.** mRNA levels of GPC5 and OGFRL1 were determined by qPCR using exon specific primers. All values are normalized to GAPDH mRNA. DGCR8 depletion reduced levels of both mRNA by 5 fold while Dicer depletion had no effect. **f.** Protein levels of GPC5 and OGFRL1 are reduced by transcriptional interference. Direction of transcription indicated by green arrow and miRNA by red vertical lines. All experiments used HeLa cells.

Figure 8: Lnc-pri-miRNA may be CPA-incompetent or competent following Microprocessor depletion.

RNA-seq of poly(A)+ or – fractionated Nuclear RNA from siRNA-treated HeLa cells. **a.** Nuclear RNA-seq profile showing that MIR17HG remains pA- and shows extensive transcriptional readthrough when DGCR8 is depleted. **b.** Nuclear RNA-seq profile showing that MIRLET7BHG transcripts completely switch to pA+ and terminate at a downstream PAS when DGCR8 is depleted. Note that pA- transcripts in siCntrl fraction terminate at the let-7b pre-miRNA just upstream of the PAS. Position of the PAS (pA) is marked by a dashed black vertical line. Direction of transcription indicated by 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, while Microprocessor-driven termination in lnc-pri-miRNA generates miRNA and a polyA- host transcript that is rapidly degraded. Lnc-pri-miRNA may be CPA-incompetent, leading to transcriptional readthrough and interference when Microprocessor cleavage is inhibited, or CPA-competent allowing effective transcription termination even in the absence of Microprocessor. Red thunderbolt with black fill depicts CPA mediated cleavage. Blue thunderbolt with red fill depicts Microprocessor mediated cleavage.

References

1. Friedman, R.C., Farh, K.K., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. **19**, 92-105 (2009).
2. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597-610 (2010).
3. Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051-60 (2004).
4. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957-66 (2004).
5. Ha, M. & Kim, V.N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509-24 (2014).
6. Ruby, J.G., Jan, C.H. & Bartel, D.P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83-6 (2007).
7. Xie, M. et al. Mammalian 5'-Capped MicroRNA Precursors that Generate a Single MicroRNA. *Cell* **155**, 1568-80 (2013).
8. Morlando, M. et al. Primary microRNA transcripts are processed co-transcriptionally. *Nat. Struct. Mol. Biol.* **15**, 902-9 (2008).
9. Ballarino, M. et al. Coupled RNA processing and transcription of intergenic primary microRNAs. *Mol. Cell. Biol.* **29**, 5632-8 (2009).
10. Kim, Y.K. & Kim, V.N. Processing of intronic microRNAs. *EMBO J.* **26**, 775-83 (2007).

11. Baskerville, S. & Bartel, D.P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**, 241-7 (2005).

12. Sundaram, G.M. et al. 'See-saw' expression of microRNA-198 and FSTL1 from a single transcript in wound healing. *Nature* **495**, 103-6 (2013).
13. Proudfoot, N.J. Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770-82 (2011).
14. Mandel, C.R. et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953-6 (2006).
15. West, S., Gromak, N. & Proudfoot, N.J. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522-5 (2004).
16. Kim, M. et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517-22 (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**, 1572-80 (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. *Mol. Cell* **21**, 65-74 (2006).
19. Kawauchi, J., Mischo, H., Braglia, P., Rondon, A. & Proudfoot, N.J. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev.* **22**, 1082-92 (2008).
20. Rondon, A.G., Mischo, H.E., Kawauchi, J. & Proudfoot, N.J. Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. *Mol. Cell* **36**, 88-98 (2009).

21. El Hage, A., Koper, M., Kufel, J. & Tollervey, D. Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes Dev* **22**, 1069-81 (2008).
22. Ghazal, G. et al. Yeast RNase III triggers polyadenylation-independent transcription termination. *Mol. Cell* **36**, 99-109 (2009).
23. Chang, J., E. Nicolas, 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 Biology* **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. Elmen, J. et al. LNA-mediated microRNA silencing in non-human primates. *Nature* **452**, 896-9 (2008).
26. Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M. & Sarnow, P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309**, 1577-81 (2005).
27. Li, Z.Y. et al. Positive regulation of hepatic miR-122 expression by HNF4alpha. *J. Hepatol.* **55**, 602-11 (2011).
28. Chien, C.H. et al. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Res.* **39**, 9345-56 (2011).
29. Wang, I.X. et al. RNA-DNA Differences Are Generated in Human Cells within Seconds after RNA Exits Polymerase II. *Cell Rep.* **6**, 906-15 (2014).
30. Gromak, N. et al. Drosha regulates gene expression independently of RNA cleavage function. *Cell Rep.* **5**, 1-12 (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**, 8415-20 (2000).

32. Prescott, E.M. & Proudfoot, N.J. Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci. USA* **99**, 8796-801 (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**, 6821-33 (2012).

34. Bracht, J., Hunter, S., Eachus, R., Weeks, P. & Pasquinelli, A.E. Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. *RNA* **10**, 1586-94 (2004).

35. Wagschal, A. et al. Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* **150**, 1147-57 (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**, 1013-8 (2008).

37. Steinmetz, E.J., Conrad, N.K., Brow, D.A. & Corden, J.L. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413**, 327-31 (2001).

38. Arigo, J.T., Eyler, 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**, 841-51 (2006).

39. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S. & Meinhart, 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. Baillat, D. et al. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265-76 (2005).
41. O'Reilly, D. et al. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res.* **42**, 264-75 (2014).
42. Kim, M. et al. Distinct pathways for snoRNA and mRNA termination. *Mol. Cell* **24**, 723-34 (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., Ulitsky, I., McGeary, S.E. & Bartel, D.P. Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. *Cell* **152**, 844-58 (2013).
45. Conrad, T., Marsico, A., Gehre, M. & Orom, U.A. Microprocessor Activity Controls Differential miRNA Biogenesis In Vivo. *Cell Rep.* **9**, 542-54 (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**, 919-32 (2008).
47. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. & Sharp, P.A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360-3 (2013).
48. Ntini, E. et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923-8 (2013).
49. Kaida, D. et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664-8 (2010).

50. Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775-89 (2012).
51. Tilgner, 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**, 1616-25 (2012).
52. 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**, 1603-14 (2013).
53. Sperber, H. et al. miRNA sensitivity to Drosha levels correlates with pre-miRNA secondary structure. *RNA* **20**, 621-31 (2014).