3464–3473 *Nucleic Acids Research, 2009, Vol. 37, No. 10* *Published online 1 April 2009*
doi:10.1093/nar/gkp205

C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts

Marie-Line Bortolin-Cavaillé[1,2], Marie Dance[1,2], Michel Weber[1,2] and Jérôme Cavaillé[1,2,*]

[1]Université de Toulouse, UPS, Laboratoire de Biologie Moléculaire Eucaryote and [2]CNRS, LBME, F-31000 Toulouse, France

Received February 10, 2009; Revised March 11, 2009; Accepted March 12, 2009

ABSTRACT

MicroRNAs are tiny RNA molecules that play important regulatory roles in a broad range of developmental, physiological or pathological processes. Despite recent progress in our understanding of miRNA processing and biological functions, little is known about the regulatory mechanisms that control their expression at the transcriptional level. C19MC is the largest human microRNA gene cluster discovered to date. This 100-kb long cluster consists of 46 tandemly repeated, primate-specific pre-miRNA genes that are flanked by *Alu* elements (*Alus*) and embedded within a ~400- to 700-nt long repeated unit. It has been proposed that C19MC miRNA genes are transcribed by RNA polymerase III (Pol-III) initiating from A and B boxes embedded in upstream *Alu* repeats. Here, we show that C19MC miRNAs are intron-encoded and processed by the DGCR8-Drosha (Microprocessor) complex from a previously unidentified, non-protein-coding Pol-II (and not Pol-III) transcript which is mainly, if not exclusively, expressed in the placenta.

INTRODUCTION

MicroRNAs (miRNAs) are tiny, endogenously expressed non-protein coding RNAs (~19–23 nt in length) that act as antisense RNA to regulate eukaryotic gene expression mainly at the post-transcriptional level. Given their ability to interact with a huge number of target mRNAs, microRNAs are recognized as key regulators of gene expression that play pivotal roles in a broad range of cellular and developmental processes (1).

The biogenesis of microRNAs is a complex process that requires the involvement of two members of the ribonuclease (RNase) III family, Drosha and Dicer, that act sequentially in the nucleus and in the cytoplasm, respectively (2). First, the primary transcripts (pri-miRNAs) are converted by the Drosha-DGCR8 complex (the so-called Microprocessor) into pre-miRNAs that fold into characteristic, ~70-nt long, irregular hairpin-like structures. Through the exportin-5 pathway, pre-miRNAs are then exported to the cytoplasm where they are further processed by Dicer (complexed with TRBP), which excises, from one of the two arms of the pre-miRNA hairpin, the mature single-stranded miRNAs. Concomitantly with the final steps of processing, miRNAs incorporate into specific ribonucleoprotein (RNP) complexes (miRNPs or miRISC). Through their partial base-pairing with the 3′-UTRs of protein-coding mRNAs, they trigger gene silencing by preventing translation of targeted mRNAs and/or by accelerating their degradation (3,4).

In mammals, the vast majority of miRNA genes (~80%) are positioned within introns of longer primary transcripts that can be either protein coding or mRNA-like transcripts that are synthesized by RNA polymerase II (Pol-II) (5–8). Intergenic miRNA genes are independently transcribed from their own promoters whereas intronic microRNAs are transcribed with their host genes and are likely co-transcriptionally processed before the complete removal of the host intron (6). Interestingly, several pre-miRNAs (the so-called mirtrons) correspond precisely to the spliced-out intron and thus do not require the involvement of the Microprocessor (9–11).

The expression of many microRNA genes is tightly regulated in a developmental and/or tissue-specific manner and, as a consequence, mis-regulation of microRNAs is associated with many human diseases, notably in cancers (12). Thus, furthering our knowledge of miRNA transcription is of primary interest to fully describe miRNA-mediated gene regulation. However, despite the recent genome-wide identification of chromatin signatures that predict some miRNA gene transcriptional start sites and promoter regions (13,14) and the characterization of several factors that activate or repress microRNA transcription (15–20), the transcriptional regulatory mechanisms governing the spatiotemporal expression pattern of the vast majority of mammalian miRNA gene loci remains largely unknown.

Interestingly, microRNA genes tend to be organized into clusters comprising different or related microRNA gene copies (21). Such operon-like gene structures are

*To whom correspondence should be addressed. Tel: +33 0 561335927; Fax: +33 0 561335886; Email: cavaille@ibcg.biotoul.fr

thought to allow co-expression of most, if not all, miRNA members belonging to a same cluster (22) and are very likely to have functional significance (23). The largest human miRNA gene cluster (also referred to as chromosome 19 miRNA cluster, C19MC), maps to chromosome 19q13.41 and extends over a ~100-kb long region. It harbours 46 pre-miRNA genes that are mainly, if not exclusively, expressed in the placenta (24). Remarkably, most of C19MC microRNA genes are highly related to each other and are positioned within 400–700 bp repeated sequences bounded by *Alu* repeats (*Alus*). Borchert *et al.* (25) have proposed that upstream *Alus* provide Pol-III promoters that drive the expression of the downstream C19MC pre-miRNA genes. These would be the only examples of Pol-III transcribed miRNA genes, and thus if unique, would have a major impact on our understanding of the evolution and the expression of miRNA genes in mammals. Indeed, C19MC accounts for ~8% of all known human microRNA genes (59 out of 695) and the fact that it is only found in the primate lineage strongly suggests that it is involved in body-plan innovation and/or phenotypic plasticity (24,26,27).

In this work, we re-examine microRNA gene organization and expression at C19MC and we demonstrate that C19MC microRNA genes are unlikely to be significantly expressed as Pol-III-dependent transcription units. Rather we privilege a model whereby C19MC microRNAs are processed from previously unidentified, Pol-II placenta-specific non-protein-coding transcripts.

MATERIALS AND METHODS

Unless otherwise noted, all techniques for cloning and manipulating nucleic acids were performed according to standard protocols.

Oligonucleotides, siRNAs

DNA oligonucleotides and siRNAs were purchased from Sigma and Qiagen, respectively. The 5′-exon(1): ttc tctagaaatccagcccaattctcttgg; 3′-exon(1): gttcacttgagttccag (c/t)acagggacacacagt; 5′-exon(2): catcaggactgtgtgtttctgtg; 3′-exon(2): tatattta accatgagaattgagc; 5′-exon(3): cccccatg aggactgtgcgc; 3′-exon(3): gagaacagcaatggactttgagctg; 3′-intron(1) caaat(g/t)tattaccaagatcagc; 5′-intron(1): tgcaa tctttatttttgtgtccatttt. h2b-fd: gaccaaagcgcagaagaaag; h2b-rev:ggtcgagcgcttgttgtaat; 5.8S rRNA: tcctgcaattcacattaatt ctcgcagctagc; $tRNA^{Tyr}$: atcgaaccagcgacctaaggat; pre-$tRNA^{Tyr}$ (intron): ggatgtctcctgctgaggaagtagct; snaR-A: gacccatgtggaccaggctggcctcgaact; hsa-mir-515-5p: cagaaag tgctttctttggagaa; hsa-mir-519a: gtaacactctaaaaggatgcacttt; hsa-mir-515-3p: acgctccaa aagaaggcactc; hsa-mir517a: aac agtctaaagggatgcacgat; si-control: uucuccgaagcugucacgutt; si-DGCR8: auccguugaucucgaggaatt; si-Drosha: aaggacca aguauucagc aag.

Cell culture, transient transfection and drug treatment

JEG3 and HeLa cells were grown in DMEM (Gibco) supplemented with 10% fetal bovine serum and antibiotics. For RNA interference experiments, 10×10^6 exponentially growing JEG3 cells were trypsinized, washed in PBS and suspended in 200 µl of Optimem (Gibco) before being electroporated at 270V/950 µFaraday (Biorad Genepulser Xcell) with 10 µl of siRNA (100 µM. Electroporated cells were then incubated 72 h before further analysis. RNA pol-II was inhibited by adding α-amanitine (#A2263, Sigma-Aldrich) to the cell culture medium to a final concentration of 20 µg/ml. The steady-state level of the pre-miRNAs was evaluated by adding Actinomycine D (#A1410, Sigma-Aldrich) to the cell culture medium to a final concentration of 5 µg/ml.

RNA isolation, RT–PCR, northern blot analysis and RNAse protection assay

Total RNA was prepared from cells using Trizol (Invitrogen) and treated with RQ1 DNase (Promega) and proteinase K (Sigma) before storage at −20°C in RNAse-free water. Human total RNAs were purchased from Ambion (The FirstChoice human total RNA survey panel, #AM6000). For RT–PCR analysis, 10 µg of total RNA was reverse-transcribed (Superscript II RTase, Invitrogen) at 42°C for 2 h, using random hexamer primers or specific primers, and ~1/10–1/20 of cDNA products were amplified by PCR (30 cycles with GoTaq polymerase, Promega). PCR products were then cloned into pGEM-Teasy (Promega) and sequenced. For northern blot analysis, 10 µg of total RNA was fractionated by electrophoresis on a 15% acrylamide/7 M urea denaturating gel. Electrotransfer was performed onto nylon membranes (BrightStar Plus membrane, Ambion), followed by UV light irradiation. Northern blot hybridization was carried out with 5′-32P-labeled-DNA oligonucleotide probes, with an overnight incubation at 50°C in 5× SSPE, 1%SDS, 5× Denhardt's, 150 µg/ml yeast tRNA. Membranes were washed twice with 0.1%SSPE, 0.1%SDS at room temperature before autoradiography. For RNAse protection assay (RPA), antisense pri-miRNA probes were prepared by *in vitro* transcription of linearized plasmids (template) with T7 RNA polymerase in the presence of α-[^{32}P] CTP. 10 µg of total RNA was then incubated with excess antisense RNA probe (overnight, 50°C) and digested with a mixture of RNAse A and RNAse T1. The RNAse-protected fragments were resolved on a 6% acrylamide/7 M urea denaturating gel.

Chromatin immunoprecipitation

HeLa or JEG3 cells were grown to 100% confluence and then processed as previously described (28). Briefly, cells were cross-linked 10 min in 1% formaldehyde before the reaction was stopped with 0.125 M of glycine. Cells were scrapped, centrifuged and the pellet was resuspended in cellular lysis buffer, then Dounce-homogenized, centrifuged again and resuspended in nuclear lysis buffer. Samples were sonicated and 200 µg of DNA were used for each immunoprecipitation. After saturation, cellular extract was incubated overnight at 4°C with 2 µg of either mouse IgM directed against RNA polymerase II phosphorylated on serine 5 (H14, Covance) or IgG anti-fibrillarin. Then, saturated protein A/G sepharose and IgG anti-mouse IgM (Sigma) were added and the incubation was continued for 2 h. The immunoprecipitate was

subsequently centrifuged and washed. Beads and input were resuspended in TE buffer and treated with RNase A (0.05 μg/μl). Cross-link was reversed with 0.25% SDS overnight at 70°C. Samples were then treated with Proteinase K (0.25 μg/μl) and DNA fragments were purified by consecutive phenol/chloroform extractions followed by ethanol precipitation. Pellets were resuspended in 100 μl of ultra-pure water. For PCR, 5 μl of ChIP, GoTaq polymerase (Promega) and 28 cycles of amplification were used. PCR products were cloned using the pGEM T-easy kit (Promega) and individual clones were sequenced.

RESULTS

C19MC miRNA genes are intron-encoded within a novel, primate-specific non-protein-coding RNA gene

To further characterize how C19MC microRNAs are generated, we searched for ESTs covering the 100-kb long C19MC cluster both in human and in monkey. We identified 58 ESTs in the human locus, 74% of them cloned from placenta and 7% from germ cells (not shown). Based on sequence analysis of BF773110, one of the very few spliced ESTs, we identified a 123-nt long, spliced exon that displays strong sequence similarity to 37 genomic segments of C19MC, most of which are regularly spaced upstream and downstream of pre-miRNA genes (Figure 1A and B). The overall intron–exon genomic organization of many of these C19MC pre-miRNA genes appears to be well conserved among several primates including Rhesus, Chimpanzee, Marmoset, but apparently not in Bushbaby, nor in any other nonprimate genomes that we have analysed (Figure 2A and data not shown), supporting the notion that C19MC has appeared recently (24), most likely in the common ancestor that leads to Haplorhini or Anthropoidea lineages (27).

Examining orthologous loci from other primates with a dbEST BLAST search using the sequence of the human consensus repeated exon, we identified two ESTs isolated from a *Papio anubis* placenta cDNA library (FC114510 and FC115727). The 3′-end of FC114510 and the 5′-end of FC115727 overlap over 84/85 nt with a single mismatch and thus likely correspond to the same exon. Altogether, these two ESTs define 7 exons and overlap a 14 617-nt long genomic segment containing eight pre-miRNA sequences (Figure 2B). These seven exons can be classified into three categories: (i) Exons III, IV and V- and the 3′-end of exon I- are highly similar to the human C19MC consensus exon we defined from BF773110. Furthermore, four highly similar sequences (IVb, Vb, Vc and VIb), not included in these two ESTs, were localized in this genomic segment, showing that a consensus exon, or exon-like sequence, precedes each pre-miRNA sequence; (ii) Exons II and VI correspond to inverted *Alu* sequences; (iii) Exon VII shows high sequence identity with 5 ESTs from the human C19MC locus, three of which are polyadenylated, leaving open the possibility that it represents the last exon of these transcripts (Figure 2B).

Detection of a placenta-specific mRNA-like transcript generated at C19MC

In order to further confirm that one or several previously undetected transcript(s) is/are generated at C19MC, we performed strand-specific RT–PCR using primers matching the consensus of the repeated exons (Figure 1B). As shown in Figure 2C, in agreement with the existence of spliced-, repeated exon-containing mRNA-like transcripts overlapping C19MC, a characteristic ladder-like pattern was revealed with 4 discrete bands differing by about 120 nt increments (i.e.the length of one exon). The specificity of these RT–PCR products was further confirmed by cloning and sequencing. None of them displayed any obvious open reading frame (data not shown). Surprisingly, unexpected nucleotide differences with the available human genomic sequence did not allow us to position precisely most of these spliced exons along C19MC, suggesting that either the C19MC is a highly polymorphic locus and/or that its sequence is not entirely correct. Nevertheless, as illustrated in Figure 2D, the use of additional primers designed to specifically recognize selected exons unambiguously revealed alternatively spliced transcripts overlapping the miR-517a and miR-515-1 gene studied by Borchert *et al*. As observed for *P. anubis* ESTs, antisense *Alu* sequences are also included in mature transcripts, again suggesting that C19MC *Alus* are prone to exonization (29). Although their complex repeated structures are still poorly characterized (i.e. we cannot formally distinguish between a single or several transcription unit(s) and we do not know if all the repeated exons are efficiently spliced), these non-protein-coding transcripts are collectively termed C19MC-HG for C19MC Host-Gene.

We then examined the expression profile of C19MC-HG transcripts in a panel of 20 human tissues. As shown in Figure 2E, bottom, C19MC-HG exhibits a strict tissue-specific expression pattern with strongest expression in the placenta, as well as weaker expression in the testes, a finding consistent with the expression pattern of ESTs matching C19MC. It is worth noting that C19MC miRNAs are also mostly, if not exclusively, detected in the placenta (Figure 2E, top), strongly supporting the notion that these miRNAs derive from these placenta-specific C19MC-HG transcripts.

Active RNA polymerase II is associated with C19MC in JEG-3 cells

Because some of these observations are at odds with the notion that C19MC miRNAs are derived from Pol-III transcripts (25), we examined next Pol-II recruitment at C19MC in the choriocarcinoma (placenta) JEG3 cell line that endogenously expresses C19MC microRNAs (Figure 3A, left), as well as C19MC HG (not shown). However, the steady-state level and/or the (alternative) RNA splicing of spliced C19MC-HG in JEG3 cells is/are different from that we described in the placenta, because only the shorter and most abundant PCR products are revealed by using exon-specific primers and not a characteristic ladder-like pattern as seen in placenta tissue (Figure 2C). Given the complex repeated

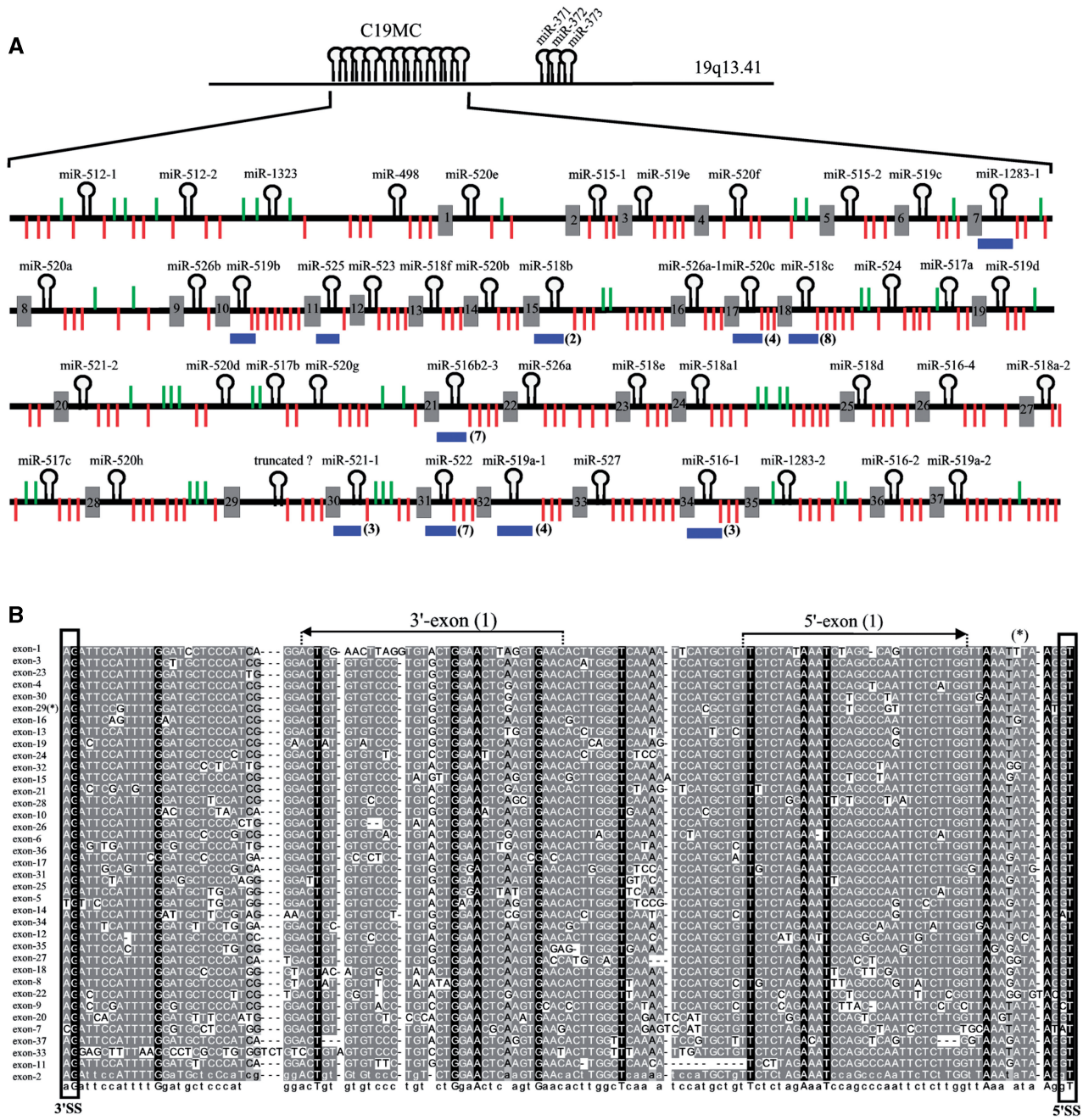

Figure 1. Revisiting microRNA gene organization at C19MC, the largest human microRNA gene cluster. (**A**) Schematic representation of the ~100-kb long C19MC (HG18: 58 860 000–58 962 300) mapping at human chromosome 19q13.41. Pre-miRNA genes are symbolized as stem–loop structures. Repeated exons of C19MC-HG are indicated as grey boxes and *Alus* (as annotated at http://genome.ucsc.edu) are indicated by vertical bars, with green and red bars corresponding to the sense and antisense orientations relative to the pre-miRNA genes, respectively. Horizontal blue bars show miRNA gene loci that have been identified by RNA polymerase-II ChIP experiments described in Figure 3A (the number of sequenced clones is indicated in brackets). (**B**) Sequence alignment of human exons 1–37. The multiple sequence alignment was generated with Multalin (http://bioinfo.genopole-toulouse.prd.fr/multalin/cgi-bin/multalin.pl) and conserved nucleotides were colored with GeneDoc (http://www.Cris.com/~ketchup/genedoc.shtml). Donor (5′SS) and acceptor (3′SS) splice sites are indicated. Asterisks indicate an alternative 5′SS found in a subset of spliced exons found in cDNA clones we sequenced (not shown). The position of primers 3′-exon(1) and 5′-exon(1) used in Figure 2C are indicated on the top. An LTR sequence inserted in exon 29 has been deleted.

organization at C19MC, we voluntarily designed 'consensus primers' matching repeated sequences for ChIP experiments, in order to simultaneously amplify multiple miRNA gene copies in a single PCR reaction. As expected, C19MC pri-miRNA sequences were enriched for Pol-II in JEG3 but not in HeLa cells, whereas the H2B gene, used as positive control for Pol-II, was enriched in both cell lines (Figure 3A, right). The specificity of these PCR products was confirmed by sequencing 41 randomly picked clones assigned to 11 different miRNA gene

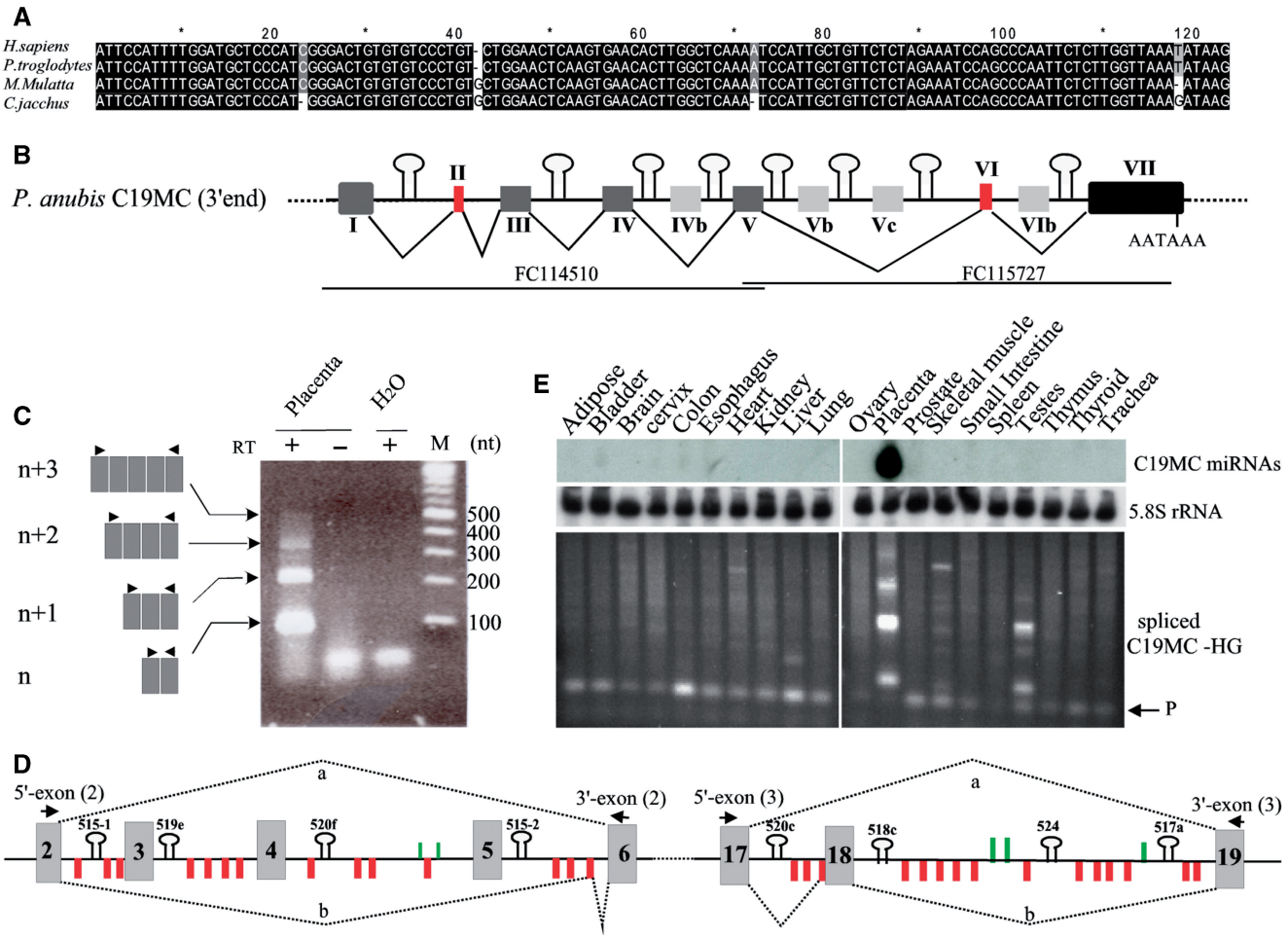

Figure 2. Identification of a novel, placenta-specific and repeated exon-containing noncoding RNA gene. (**A**) Sequence alignment of the consensus, repeated non-protein-coding exons identified in several primate species. (**B**) Exon–intron organization at *P. anubis* C19MC identified by ESTs analysis. The genomic structure of the *P. anubis* locus (AC184009.2 *Papio* BAC clone) encoding ESTs FC114510 and FC115727 is shown with sequences similar to the human C19MC exon indicated in dark grey, while similar exons not spliced in these two ESTs are indicated in light grey. The two exons exapted from inverted *Alu* elements are in red and pre-miRNA sequences are symbolised as stem–loop structures. Note that the 3′ exon VII (black rectangle) is similar to five ESTs at the 3′-end of the human C19MC, three of them (BF508730, BQ024418, BQ024579) are polyadenylated. For clarity, only the two exapted *Alus* are shown. (**C**) Experimental detection of (a) novel, placenta-specific mRNA-like transcript(s). Detection of spliced C19MC-HG mRNA-like transcript(s) by RT–PCR using primers as indicated in Figure 1B. (**D**) Schematic representation of the genomic region spanning miR-515-1 and miR-517a miRNA gene loci proposed to be transcribed by Pol-III (25). Dotted lines represent two alternative splicing events (a, b) identified by sequencing RT-PCR products using primers as indicated above exons. Other symbols are as in Figure 1. (**E**) Tissue-specific expression pattern of C19MC miRNAs. Total RNA from human tissues was fractionated on a denaturing 15% acrylamide/7 M urea gel and analysed by northern blot with a mixture of ^{32}P-labelled oligonucleotides antisense to mature miR-517a, miR-515-1 and miR-519a1 that have been analysed extensively by Borchert *et al*. The same membrane was probed with a 5.8S rRNA probe to control gel loading. The tissue-specific expression pattern of spliced C19MC-HG transcripts was also assayed by RT–PCR using the same set of human tissues; P, primers.

loci throughout C19MC (Figure 1A). Although it is not expected to be exhaustive (e.g. bias during PCR due to preferential amplification of some miRNA genes is quite likely), this approach provides strong support for the involvement of Pol-II in the transcription of microRNA genes at C19MC.

miRNA gene expression at C19MC is sensitive to alpha-amanitin treatment in choriocarcinoma cell lines

To further confirm the involvement of Pol-II in C19MC miRNA transcription, JEG3 cells were treated with α-amanitin at a concentration (20 μg/ml) that specifically inhibits Pol-II transcription. As shown in Figure 3B, left, the level of the pre-miRNAs, but not that of the fully processed and metabolically stable microRNAs, was dramatically reduced in α-amanitin-treated cells, while the level of pre-tRNA^{Tyr} or that of snaR-A (30), used as Pol-III transcript controls, remained unchanged. As expected, the use of actinomycin D at high concentration (5 μg/ml) that rapidly block the three RNA polymerases similarly affected the steady-state level of both pre-miRNAs and unspliced pre-tRNA^{Tyr} (Figure 3B, right). From these experiments, we conclude that Pol-II transcription significantly contributes to *in vivo* C19MC gene expression.

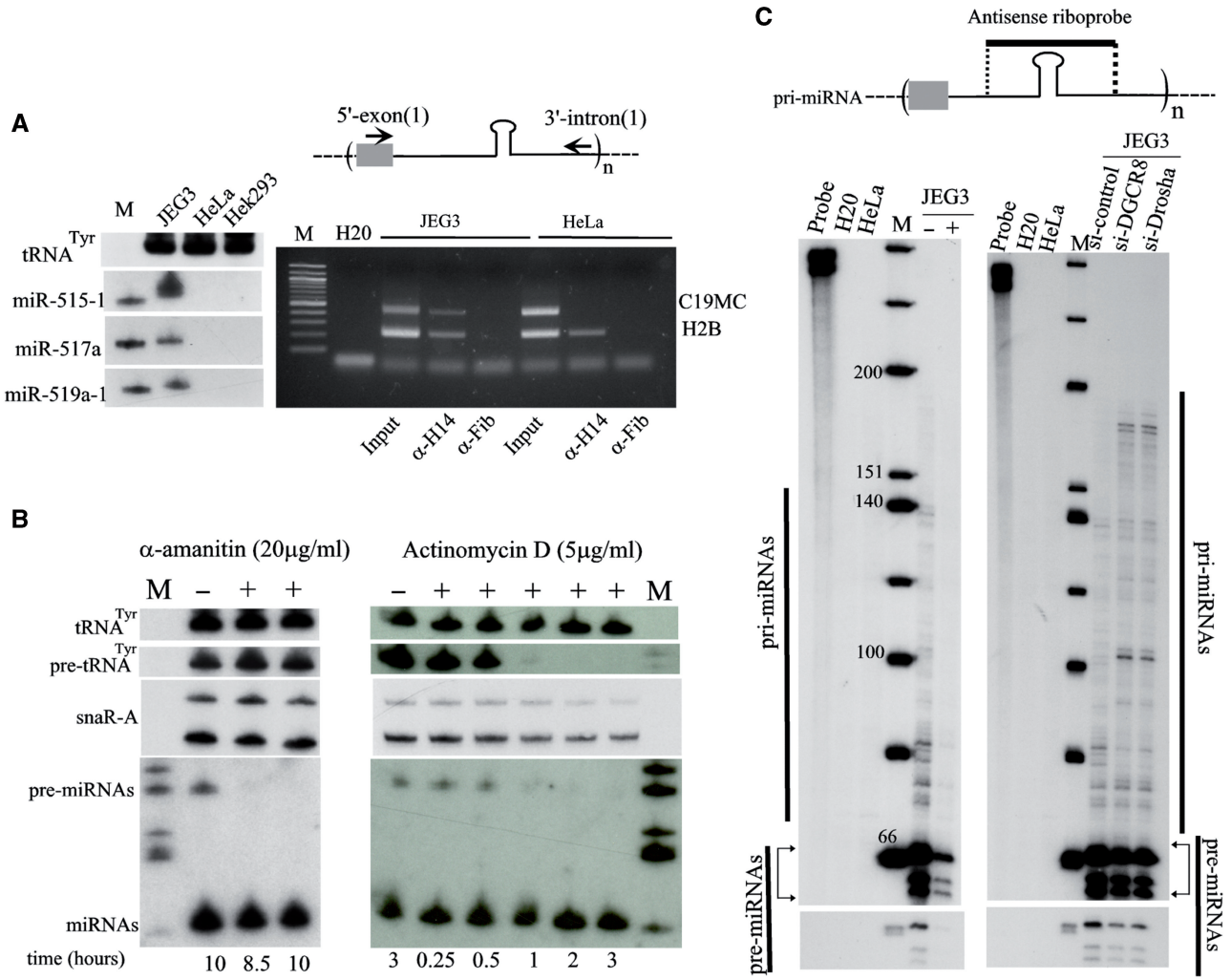

Figure 3. C19MC-HG as a Pol-II, pri-miRNA transcript. (**A**) Pol-II is recruited at the endogenously-expressed C19MC locus. Left: miR-515-1, miR-517a and miR-519a-1 genes are endogenously-expressed in JEG3, but not in HeLa or HEK293 cells. miRNA expression was monitored by northern blot with specific oligonucleotide probes. Note that since microRNA genes mapping to C19MC are highly related to each other, causing potential cross-hybridization, it is extremely difficult to formally demonstrate their specific expression by northern blot although stringent hybridization conditions were used. Right: ChIPs were performed in JEG3 and HeLa cells with anti-Pol-II antibodies (H14, Covance) that recognize the phosphorylated CTD serine 5. Antibodies against fibrillarin (Fib) were used as a negative control. C19MC and H2B genes (used as positive control for Pol-II) were simultaneously detected by multiplexed PCR. A representative ethidium bromide-stained agarose gel of three independent ChIPs is shown and miRNA gene loci enriched for RNA Pol-II are shown in Figure 1A. (**B**) C19MC gene expression is sensitive to α-amanitin. JEG-3 cells were either treated by α-aminitin (left) or actinomycin D (right) as indicated and C19MC microRNA gene expression was assayed by northern blot using a mixture of ^{32}P-labelled oligonucleotides antisense to mature miR-517a, miR-515-1 and miR-519a-1. By probing the membrane with an antisense probe to the mature miR-517 family, the same α-amanitin sensitivity was also specifically demonstrated for expression of mir-517a, one of the two C19MC microRNA genes not preceded by any (TTTT) sequence (not shown). The same membrane was also probed with a tRNA^{Tyr} and pre-tRNA^{Tyr}-specific probe that recognizes introns, as well as with a snaR-A specific probe. (**C**) C19MC-HG transcripts are processed by Microprocessor. Unspliced C19MC-HG expression was monitored by RNAse A/T1 mapping with a mixture of three related ^{32}P-labelled antisense riboprobes to ~180- to 200-nt long segments of C19MC spanning the pre-miRNA and the surrounding intronic sequences of the three randomly chosen miR-518b, miR-518e and miR-523 gene loci. Left: JEG-3 cells were either treated (+) or not (−) by α-aminitin (20 μg/ml, 9 h). Right: JEG-3 cells were transiently transfected with siRNA-negative control or siRNAs directed against Drosha and DGCR8 mRNAs. HeLa cells were used as a negative control. Probe: ~2000 c.p.m. of undigested riboprobes. H20: RNAse A/T1 mapping carried out without any RNA. At the bottom of the gel, a shorter exposure is shown for pre-miRNA signals.

C19MC-HG transcripts are processed by the Microprocessor

Given that the Microprocessor (Drosha-DGCR8) complex converts pri-miRNAs to pre-miRNAs, we reasoned that if C19MC-HG transcripts represent pri-miRNA transcripts, then their steady-state levels should increase in Microprocessor-depleted cells, as previously observed for other miRNA gene loci (31–33). We therefore knocked-down the Microprocessor by transiently transfecting JEG3 cells with short interfering RNA (siRNAs), previously shown to silence Drosha and DGCR8 mRNAs (32,33). We then simultaneously monitored the steady-state levels of pri-miRNAs and of pre-miRNAs by RNAse A/T1 mapping. As shown in Figure 3C, left,

RNAse A/T1 mapping performed with total RNA extracted from JEG-3, but not from HeLa cells, revealed pre-miRNAs species migrating as the expected ~70-nt species, as well as additional, weaker pri-miRNA signals with slower electrophoretic migration. These various RNA species, slightly shorter than the predicted fully protected RNA duplexes, might reflect the fact that C19MC-HG transcripts are short-lived RNA molecules undergoing rapid RNA decay. In addition, given the repeated organization of C19MC-HG, RNA species migrating at intermediate positions are also expected, since they might represent internally cleaved RNA products derived from imperfect RNA duplexes formed between the riboprobe and the repeated segments of C19MC with partial sequence similarity. In line with our hypothesis that C19MC-HG transcripts are processed by the Microprocessor, the steady-state level of pri-miRNA was increased in Drosha and DGCR8 depleted-cells with a pattern of protected RNA fragments notably different from that seen in control cells, while the level of pre-miRNAs was decreased as expected (Figure 3C, right). The increase of the steady-state level of pri-miRNA species was further confirmed by real-time PCR (data not shown). Consistent with the fact that these protected pri-miRNA species are synthesized by Pol-II, they were not detected in α-amanitin-treated JEG3 cells (Figure 3C, left). These data confirm the notion that C19MC-HG corresponds to Pol-II transcripts processed by the Microprocessor, and can therefore be considered as *bona fide* pri-miRNA transcripts from which miRNAs are derived.

DISCUSSION

In this work, we show that many, if not all, C19MC microRNA genes are positioned within introns of one or several non-protein-coding transcript(s) that, after completion of (alternative) RNA splicing, consist of repeated exons as well as inverted *Alu* sequences (Figures 1 and 2B, 2D). These novel transcripts, collectively called C19MC-HG, represent Pol-II, pri-miRNA transcripts because: (i) the overall intron–exon genomic organization is conserved among primates (Figure 2A); (ii) they display the same tissue-specific expression pattern as the embedded miRNAs (Figure 2E); (iii) the steady-state level of intron-containing RNA species surrounding miRNA gene loci is increased in Microprocessor-depleted cells (Figure 3C); (iv) Pol-II is associated throughout C19MC in microRNA-expressing cells (Figure 3A); (v) the expression of C19MC-HG is sensitive to α-amanitin (Figure 3B).

microRNA gene organization at C19MC is therefore similar to that we previously described at C14MC (also referred to as the Mirg cluster), another large tandemly repeated array of microRNA genes residing at the imprinted Dlk1-Dio3 domain (34–36). However, despite an obvious overall similarity, i.e. repeated microRNAs processed from introns of large, non-protein-coding transcripts, these two evolutionarily distinct miRNA gene clusters differ to some extent. First, C19MC is only found in primates while C14MC appears to be conserved among eutherian species (37). Second, C19MC is particularly enriched with *Alus* (25) while C14MC lies within a genomic region depleted of *Alus* (34). Intriguingly, C19MC-encoded *Alus* are massively inserted in the antisense (minus) strand with respect to transcription of C19MC miRNA genes (Figure 4A). The structural and/or functional significance of this unexpected bias, if any, is currently not understood. Third, C19MC miRNA genes, in contrast to C14MC, are found within longer sequences that are repeated. Altogether, these observations suggest a different evolutionary history for these miRNA gene cluster: *Alu*-mediated rearrangements events at C19MC might have facilitated the emergence of novel repeated miRNAs (27) while C14MC miRNAs might have arisen by tandem duplication of one or several ancestral pre-miRNA gene(s) (37). We are aware, however, that given their more ancient origin, the surrounding pre-miRNA gene sequences at C14MC might have accumulated mutations over time and therefore may no longer be detectable as repeated units.

Our findings challenge a previous report that promotes the idea that C19MC microRNAs are generated by Pol-III initiating at *Alu*-derived promoters (25). Upon stress conditions, a transient and weak reactivation of some *Alus* has been described (38–40). Thus, one could argue that Pol-II mediated transcription generate the bulk of C19MC microRNA genes under normal condition while under certain conditions that remain to be identified, Pol-III could occasionally transcribe C19MC genes. This explanation is, however, extremely unlikely. Indeed, while *Alu* sequences might facilitate post-transcriptional regulation of C19MC-HG transcripts by promoting (alternative) RNA splicing (e.g. exonization of *Alus* as illustrated in Figure 2B and D), or might have played a role in the diversification of C19MC miRNA genes in primates [e.g. *Alu*-mediated recombination as previously discussed (27)], an involvement of *Alu*-embedded Pol-III in transcribing miRNA genes at the C19MC locus is not supported by the miRNA- *Alu* gene organization. First, the vast majority of these putative Pol-III promoters (81.7%; 211 out of 258 *Alus* annotated at UCSC version HG18, bp 58 860 000–58 962 300) are inserted in the reverse orientation with respect to the direction of transcription of the pre-miRNA genes (Figures 1A and 4A and 4B). Second, even *Alus* in the correct orientation are unlikely to drive expression of miRNAs since in the vast majority of cases several Pol-III transcriptional stop signals, e.g. stretches of T_4 ot T_5, are found upstream of the pre-miRNA sequences. TTTT sequences are also present in 22 out of 46 pre-miRNA genes (~48%) including within the mature sequences of miR-519a-1, miR-515-1 and miR-517c studied by Borchert *et al.* (Figure 4B). In fact, only 2 out of 46 pre-miRNA genes, namely miR-517a and miR-517b, do not have any (TTTT) stretches. We are aware that, in some rare cases, Pol-III can travel through short polyT stretches. However, it is largely accepted that four or five consecutive T residues, or the related TTATT sequence, constitute a very efficient RNA Pol-III terminator (41). Accordingly, a single run of T_4 and T_5 terminates Pol-III transcription in 65% and 20% of the 464 human and mouse tRNA genes, respectively (42). It is also

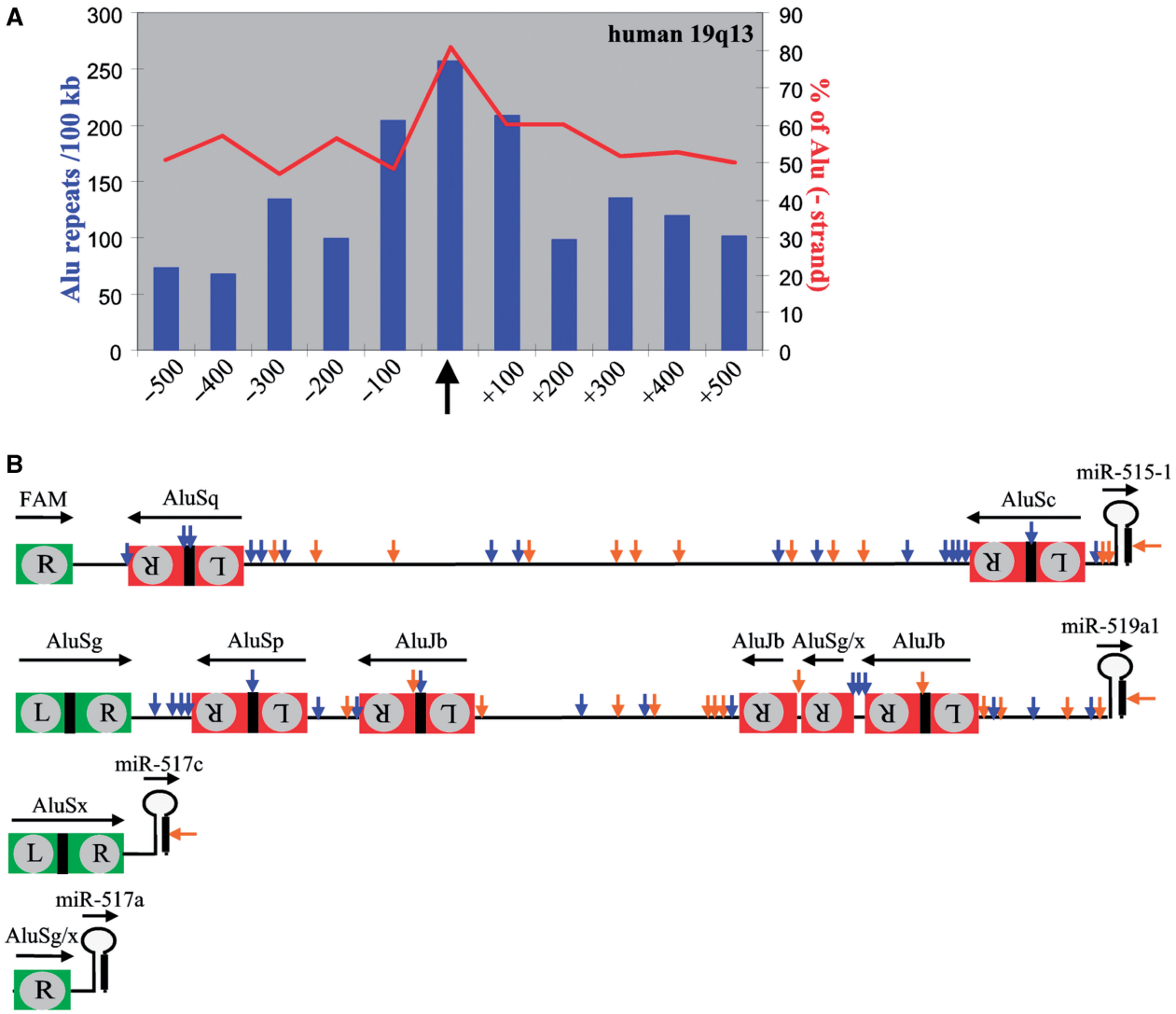

Figure 4. The C19MC microRNA-*Alu* arrangements are not consistent with robust Pol-III microRNA expression *in vivo*. (**A**) High density and strand bias of *Alu* repeat insertion at C19MC. Distribution of *Alu* repeats (histograms) and percentage of *Alus* in the minus strand (red curves) along C19MC. A 500 kb flanking either side of C19MC (the position of which is indicated as black arrow) was analysed. (**B**) Schematic representation of miRNA gene loci studied by Borchert *et al.* (25). Upstream *Alus* are indicated by boxes, with green and red boxes oriented in the sense and antisense orientations with respect to transcription of the downstream pre-miRNA genes (stem–loop structures). *Alus* have a dimeric structure with two related, but not equivalent, monomers: left (L) and right (R) arms. The A and B boxes of the internal Pol-III promoter are located in the left arm. Note that for miR-515-1 and miR-517a genes, the first correctly oriented *Alu* consists only of the right arm not expected to contain a functional Pol-III promoter. Blue and orange vertical arrows correspond to T_5 (or $T_{>5}$) and T4 stretches, respectively, predicted to act as Pol-III transcriptional stop signals.

noteworthy that transcription of some *Alu* repeats can terminate at a run of only three T residues or even at atypical A-rich pol-III terminators (43). Therefore, the putative *Alu*-pre-miR-515-1 and -pre-miR-519a-1 transcripts predicted by Borchert *et al.* are unlikely to be generated *in vivo* because of the presence of 4 and 8 stretches of TTTT(T), respectively (Figure 4B). In addition, their two upstream *Alus* are positioned in a reverse orientation. Assuming that the nearest correctly oriented *Alu*-embedded Pol-III promoters (located at ~3-kb away) are functional, which has not been demonstrated, then there are ~30 stretches of TTTT(T) between the pre-miRNA genes and the putative *Alu*-resident promoter. Third, even in the case of the miR-517a pre-miRNA gene, the presence of a functional two box (A and B) RNA Pol III promoter is questionable since the upstream annotated *Alu* (AluSg/x) only consists of the right arm of the monomer (Figure 4B). Altogether, and taking into account that endogenous *Alu*-derived Pol-III promoters in their native chromatin context are extremely weak (e.g. *Alus* are usually highly methylated) and transcribed *Alu* elements rely on upstream sequences for transcription (44,45), we conclude that *Alu*-pre-miRNA gene arrangements at C19MC are unlikely to confer robust and significant RNA Pol-III transcription *in vivo*. It is worth noting, however, that it does not exclude the possibility that some

pre-miRNA genes located elsewhere in the genome are transcribed by Pol-III (13), or even that some *Alus* within C19MC generate short and unstable Pol-III transcripts.

Our study opens up new avenues to further investigate the evolution, genomic organization and expression at C19MC. We are aware, however, that the high density of repeated sequences at C19MC (over two-third of the 100-kb sequences) and the fact that C19MC-HG transcripts are inherently short-lived RNA species (i.e rapidly degraded by the Drosha–DGCR8 complex), renders technically challenging proof of the formal existence of a single, large pri-miRNA transcript overlapping the entire 100 kb miRNA gene cluster. As a case in point, we failed to detect any significant and reproducible RNA signals by northern blot (data not shown). Thus, we cannot formally rule out the possibility that multiple miRNA promoters dispersed within C19MC contribute to miRNA gene expression. In this regard, the two closely spaced miR-127 and miR-433 genes were recently shown to be processed from two overlapping, independent pri-miRNA transcripts (46). More sophisticated experiments are now required to further identify miRNA promoter region(s) and other *cis*-acting regulatory elements that confer the placenta-specific gene expression pattern of C19MC.

ACKNOWLEDGEMENTS

We thank E. Käs and C. Monod for careful reading of the article and S. Labialle, G. Canal and P. Vitali for their continuous and helpful discussions. We are very indebted to many colleagues and anonymous reviewers who encouraged us to publish this work. We are also grateful to G. Legube and D. Trouche for their technical advice with Chromatin immunoprecipitations.

FUNDING

CNRS and Université Paul Sabatier; EMBO (EMBO Young Investigator Programme); the European Union (CallimiR); l'Agence Nationale de la Recherche (ANR blanche snosca et ARN jeunes chercheurs); l'Association pour la Recherche contre le Cancer (ARC). Funding for open access charge: CNRS.

Conflict of interest statement. None declared.

REFERENCES

1. Bushati,N. and Cohen,S.M. (2007) microRNA functions. *Annu. Rev. Cell Dev. Biol.*, **23**, 175–205.
2. Kim,V.N. (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.*, **6**, 376–385.
3. Kim,V.N., Han,J. and Siomi,M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10**, 126–139.
4. Filipowicz,W., Bhattacharyya,S.N. and Sonenberg,N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, **9**, 102–114.
5. Rodriguez,A., Griffiths-Jones,S., Ashurst,J.L. and Bradley,A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, **14**, 1902–1910.
6. Kim,Y.K. and Kim,V.N. (2007) Processing of intronic microRNAs. *EMBO J.*, **26**, 775–783.
7. Cai,X., Hagedorn,C.H. and Cullen,B.R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, **10**, 1957–1966.
8. Lee,Y., Kim,M., Han,J., Yeom,K.H., Lee,S., Baek,S.H. and Kim,V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.*, **23**, 4051–4060.
9. Berezikov,E., Chung,W.J., Willis,J., Cuppen,E. and Lai,E.C. (2007) Mammalian mirtron genes. *Mol. Cell*, **28**, 328–336.
10. Okamura,K., Hagen,J.W., Duan,H., Tyler,D.M. and Lai,E.C. (2007) The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell*, **130**, 89–100.
11. Ruby,J.G., Jan,C.H. and Bartel,D.P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature*, **448**, 83–86.
12. Lee,Y.S. and Dutta,A. (2008) MicroRNAs in cancer. *Annu. Rev. Pathol.*, **4**, 199–227.
13. Ozsolak,F., Poling,L.L., Wang,Z., Liu,H., Liu,X.S., Roeder,R.G., Zhang,X., Song,J.S. and Fisher,D.E. (2008) Chromatin structure analyses identify miRNA promoters. *Genes Dev.*, **22**, 3172–3183.
14. Marson,A., Levine,S.S., Cole,M.F., Frampton,G.M., Brambrink,T., Johnstone,S., Guenther,M.G., Johnston,W.K., Wernig,M., Newman,J. *et al.* (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, **134**, 521–533.
15. He,L., Thomson,J.M., Hemann,M.T., Hernando-Monge,E., Mu,D., Goodson,S., Powers,S., Cordon-Cardo,C., Lowe,S.W., Hannon,G.J. *et al.* (2005) A microRNA polycistron as a potential human oncogene. *Nature*, **435**, 828–833.
16. He,L., He,X., Lim,L.P., de Stanchina,E., Xuan,Z., Liang,Y., Xue,W., Zender,L., Magnus,J., Ridzon,D. *et al.* (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**, 1130–1134.
17. Rao,P.K., Kumar,R.M., Farkhondeh,M., Baskerville,S. and Lodish,H.F. (2006) Myogenic factors that regulate expression of muscle-specific microRNAs. *Proc. Natl Acad. Sci. USA*, **103**, 8721–8726.
18. Rosenberg,M.I., Georges,S.A., Asawachaicharn,A., Analau,E. and Tapscott,S.J. (2006) MyoD inhibits Fstl1 and Utrn expression by inducing transcription of miR-206. *J. Cell Biol.*, **175**, 77–85.
19. Chang,T.C., Yu,D., Lee,Y.S., Wentzel,E.A., Arking,D.E., West,K.M., Dang,C.V., Thomas-Tikhonenko,A. and Mendell,J.T. (2008) Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat. Genet.*, **40**, 43–50.
20. Barroso-delJesus,A., Romero-Lopez,C., Lucena-Aguilar,G., Melen,G.J., Sanchez,L., Ligero,G., Berzal-Herranz,A. and Menendez,P. (2008) Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter. *Mol. Cell Biol.*, **28**, 6609–6619.
21. Megraw,M., Sethupathy,P., Corda,B. and Hatzigeorgiou,A.G. (2007) miRGen: a database for the study of animal microRNA genomic organization and function. *Nucleic Acids Res.*, **35**, D149–D155.
22. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
23. Kim,Y.K., Yu,J., Han,T.S., Park,S.Y., Namkoong,B., Kim,D.H., Hur,K., Yoo,M.W., Lee,H.J., Yang,H.K. *et al.* (2009) Functional links between clustered microRNAs: suppression of cell-cycle inhibitors by microRNA clusters in gastric cancer. *Nucleic Acids Res.*, **37**, 1672–1681.
24. Bentwich,I., Avniel,A., Karov,Y., Aharonov,R., Gilad,S., Barad,O., Barzilai,A., Einat,P., Einav,U., Meiri,E. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.*, **37**, 766–770.
25. Borchert,G.M., Lanier,W. and Davidson,B.L. (2006) RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.*, **13**, 1097–1101.
26. Wang,X., Zhang,X. and Li,Y. (2008) Complicated evolutionary patterns of microRNAs in vertebrates. *Sci. China C Life Sci.*, **51**, 552–559.
27. Zhang,R., Wang,Y.Q. and Su,B. (2008) Molecular evolution of a primate-specific microRNA family. *Mol. Biol. Evol.*, **25**, 1493–1502.
28. Tyteca,S., Vandromme,M., Legube,G., Chevillard-Briet,M. and Trouche,D. (2006) Tip60 and p400 are both required for

UV-induced apoptosis but play antagonistic roles in cell cycle progression. *EMBO J.*, **25**, 1680–1689.
29. Sorek,R., Ast,G. and Graur,D. (2002) Alu-containing exons are alternatively spliced. *Genome Res.*, **12**, 1060–1067.
30. Parrott,A.M. and Mathews,M.B. (2007) Novel rapidly evolving hominid RNAs bind nuclear factor 90 and display tissue-restricted distribution. *Nucleic Acids Res.*, **35**, 6249–6258.
31. Han,J., Lee,Y., Yeom,K.H., Kim,Y.K., Jin,H. and Kim,V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, **18**, 3016–3027.
32. Lee,Y., Ahn,C., Han,J., Choi,H., Kim,J., Yim,J., Lee,J., Provost,P., Radmark,O., Kim,S. *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.
33. Gregory,R.I., Yan,K.P., Amuthan,G., Chendrimada,T., Doratotaj,B., Cooch,N. and Shiekhattar,R. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–240.
34. Cavaille,J., Seitz,H., Paulsen,M., Ferguson-Smith,A.C. and Bachellerie,J.P. (2002) Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.*, **11**, 1527–1538.
35. Seitz,H., Royo,H., Bortolin,M.L., Lin,S.P., Ferguson-Smith,A.C. and Cavaille,J. (2004) A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.*, **14**, 1741–1748.
36. Royo,H. and Cavaille,J. (2008) Non-coding RNAs in imprinted gene clusters. *Biol. Cell*, **100**, 149–166.
37. Glazov,E.A., McWilliam,S., Barris,W.C. and Dalrymple,B.P. (2008) Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Mol. Biol. Evol.*, **25**, 939–948.
38. Li,T.H., Kim,C., Rubin,C.M. and Schmid,C.W. (2000) K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. *Nucleic Acids Res.*, **28**, 3031–3039.
39. Li,T.H. and Schmid,C.W. (2001) Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. *Gene*, **276**, 135–141.
40. Liu,W.M., Chu,W.M., Choudary,P.V. and Schmid,C.W. (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.*, **23**, 1758–1765.
41. Bogenhagen,D.F. and Brown,D.D. (1981) Nucleotide sequences in Xenopus 5S DNA required for transcription termination. *Cell*, **24**, 261–270.
42. Braglia,P., Percudani,R. and Dieci,G. (2005) Sequence context effects on oligo(dT) termination signal recognition by Saccharomyces cerevisiae RNA polymerase III. *J. Biol. Chem.*, **280**, 19551–19562.
43. Hess,J., Perez-Stable,C., Wu,G.J., Weir,B., Tinoco,I. Jr and Shen,C.K. (1985) End-to-end transcription of an Alu family repeat. A new type of polymerase-III-dependent terminator and its evolutionary implication. *J. Mol. Biol.*, **184**, 7–21.
44. Khanam,T., Rozhdestvensky,T.S., Bundman,M., Galiveti,C.R., Handel,S., Sukonina,V., Jordan,U., Brosius,J. and Skryabin,B.V. (2007) Two primate-specific small non-protein-coding RNAs in transgenic mice: neuronal expression, subcellular localization and binding partners. *Nucleic Acids Res.*, **35**, 529–539.
45. Ludwig,A., Rozhdestvensky,T.S., Kuryshev,V.Y., Schmitz,J. and Brosius,J. (2005) An unusual primate locus that attracted two independent Alu insertions and facilitates their transcription. *J. Mol. Biol.*, **350**, 200–214.
46. Song,G. and Wang,L. (2008) Transcriptional mechanism for the paired miR-433 and miR-127 genes by nuclear receptors SHP and ERRgamma. *Nucleic Acids Res.*, **36**, 5727–5735.