Small non-coding RNAs and genomic imprinting

H. Royo\textsuperscript{a} M.-L. Bortolin\textsuperscript{a} H. Seitz\textsuperscript{b} J. Cavaillé\textsuperscript{a}

\textsuperscript{a}LBME-CNRS, UMR 5099, IFR 109, Toulouse (France); \textsuperscript{b}UMMS, Department of Biochemistry and Molecular Pharmacology, Lazare Research Building, Worcester, MA (USA)

Manuscript received 1 June 2005; accepted in revised form for publication by F. Ishino, 19 July 2005.

Abstract. Experimental and computer-assisted approaches have led to the identification of hundreds of imprinted small RNA genes, mainly clustered in two chromosomal domains (human 15q11\textendash}q13 and 14q32 loci). The genes are only detected in placental mammals and belong to the C/D RNA and microRNA gene families. These are small non-coding RNAs involved in RNA-guided post-transcriptional RNA modifications and RNA-mediated gene silencing, respectively. Here, we discuss their potential functions and report the identification of novel small RNA genes lying within (or nearby) known imprinted chromosomal domains.

The world of RNA has recently been upset by the discovery that a large proportion of mammalian genomes give rise to transcripts that do not fall into the three historical RNA classes: ribosomal RNA, transfer RNA or messenger RNA. Indeed, systematic bio-informatic approaches and cDNA cloning strategies have led to the identification of thousands of small and large untranslated transcripts or non-coding RNAs (also called non-messenger RNAs, mmRNAs), whose existence was completely unexpected ten years ago (Huttenhofer et al., 2005). Among the non-coding RNAs that have revolutionized our understanding of eukaryotic gene expression are the tiny, 21–23-nucleotide-long small RNAs that play a pivotal role in gene silencing, genome defense and genome stability (Tomari and Zamore, 2005).

From the vagueness that emerged from efforts to understand the biological relevance of this RNA population, the importance of non-coding RNAs in gene expression and more specifically in epigenetic control has been acknowledged. Among the best studied non-coding RNAs involved in epigenetic regulation, \textit{Xist} is a spliced, 17-kb-long non-coding RNA that has the remarkable property to ‘coat’ the future inactive X chromosome. Although the molecular events underlying this process are still largely unknown, \textit{Xist} expression is required to trigger \textit{cis}-transcriptional silencing of genes on one X chromosome to achieve gene dosage compensation in female mammalian cells (Heard, 2004). Epigenetic control also involves members of the widely conserved RNA interference (RNAi) process – small interfering RNAs (siRNAs) and small heterochromatic RNAs (shRNAs) – that guide gene silencing through heterochromatin formation or RNA-directed DNA methylation (Almeida and Allshire, 2005).

Genomic imprinting is the epigenetic marking of homologous alleles via differential cytosine methylation, histone tail modifications and replication timing that results in the preferential expression of one allele according to its parental origin (Reik and Walter, 2001). The presence of non-coding RNA genes, often (but not always) transcribed in the opposite orientation to protein-coding genes, is a recurrent theme in imprinted domains. Although the functions of these imprint-ed non-coding RNAs are still puzzling, they are believed to...
Contribute to genomic imprinting control (O’Neill, 2005). Consistent with this idea, the imprinted non-coding RNA Air appears to play a direct role in the allele-specific repression of several neighboring imprinted genes at the mouse Igf2r locus (Sleutels et al., 2002). Recent work has also made an intriguing feature of some small RNAs emerge: the expression of many C/D RNA and microRNA (miRNA) genes is subjected to genomic imprinting at human 15q11 – q13 and 14q32 loci. Here, we review their potential functions as well as the identification of additional small RNA genes that we have shown – or we strongly suspect – to be imprinted. Due to space limitations, the list of referenced publications is far from exhaustive and readers are invited to consult more comprehensive reviews about C/D RNA and miRNA genes (Bachellerie et al., 2002; Kiss, 2002; Ambros, 2004; Bartel, 2004; Kim, 2005).

C/D RNAs and microRNAs: Big world of small RNAs

Biosynthesis and functions of C/D small RNA methylation guides

Small RNAs of the C/D type (C/D RNAs) are metabolically stable, 60- to 300-nucleotide-long RNAs that have been described in eukaryotes and Archaeabacteria (human C/D RNA sequences are available at http://www-snorna.biotoul.fr/). In eukaryotic cells, they reside in the nucleus: either in the nucleolus (they are called C/D snoRNAs for C/D small nuclear RNAs) or in the Cajal bodies (they are called scaRNAs for small Cajal body-specific RNAs). In vertebrates, the vast majority of C/D RNA genes are located within introns of protein-coding genes whose products are often – but not always – part of the translation machinery. C/D RNA genes can also be intron-encoded within non-coding RNA genes. Whatever their genomic organization is, most of the C/D RNAs appear to be processed from the host-gene introns through exonucleolytic degradations of the debranched lariat.

C/D RNAs owe their names to canonical structural motifs, the C-box (consensus 5’-PuUGAUGA-3’) and the D-box (consensus 5’-CUGA-3’) found close to their 5’ and 3’ termini respectively. They also contain C’ and D’ boxes that are more degenerate and that occupy an internal position within the RNA sequence. Many C/D RNAs contain conserved antisense elements (8 to 21 nucleotides long) positioned upstream from the D- and/or the D’-motifs. While several vertebrate C/D RNAs participate in pre-ribosomal RNA (pre-rRNA) cleavages (i.e. U3, U8 …), the vast majority of them pair with RNA targets on which they guide ribose methylations at specific ribonucleotides (the modified nucleotide is always paired to the fifth nucleotide upstream from the D or D’ box, Fig. 1A). C/D snoRNAs modify the Pol I-transcribed pre-rRNAs or the Pol III-transcribed U6 spliceosomal snoRNA while scaRNAs modify the Pol II-transcribed U1, U2, U4 and U5 spliceosomal snRNAs (Fig. 1A). Finally, a broad proportion of C/D RNAs – including most of the imprinted C/D RNAs described in this review (see below) – seem to be devoid of any cellular RNA target, leaving open the possibility that additional cellular transcripts (including miRNAs?) might be targeted by C/D RNAs as well. Alternatively, these ‘orphan C/D RNAs’ might play a different role from RNA modification guiding in which no pairing with a target RNA is needed.

Biosynthesis and functions of microRNA genes

miRNAs belong to the largest eukaryotic small RNA family discovered so far, with more than 200 different RNA species described in mouse and human (miRNA sequences are available at http://www.sanger.ac.uk/Software/Rfam/mirna). They are 21- to 23-nucleotide-long single-stranded RNA molecules that are processed from one arm of an irregular 60- to 70-nucleotide-long hairpin structure (the so-called premiRNA). Pre-miRNA genes harbor different genomic organizations. They can be transcribed from their own promoters either as independent entities or as polycistrons, or they can be included in larger transcription units of coding or non-coding genes (Kim, 2005). A pre-miRNA gene is first transcribed by RNA polymerase II as a several kb long pri-miRNA (the miRNA primary transcript). Then, the RNase-III-type enzyme Drosha – assisted by DGC8 – cleaves the large pri-miRNA precursor to give rise to the pre-miRNA. The pair of cuts made by Drosha establishes one end of the miRNA. The pre-miRNA is then translocated to the cytoplasm by exportin 5-mediated export, where the other extremity of the miRNA is generated by Dicer, another multidomain RNase-III, to yield a short RNA duplex. This duplex is unwound upon loading on the RISC complex (the mature miRNA-containing ribonucleoparticle), giving rise to the single-stranded mature miRNA.

miRNAs can trigger gene silencing at the post-transcriptional level, by base pairing with a target RNA. Two pathways can be adopted (Fig. 1B): (i) If perfect (or almost perfect) complementarity is shared, the target RNA is directed for cleavage. (ii) If the miRNA presents a partial complementarity with an mRNA (generally in its 3’ UTR part), then the translation of the targeted mRNA is made non-productive by still poorly characterized mechanisms. In this context, the most 5’ part of the miRNA (2nd to 8th nucleotides, also called the ‘seed’) plays a critical role for target recognition (Doench and Sharp, 2004). Recent studies have challenged these two simple rules by demonstrating that imperfect RNA duplexes can also destabilize the targeted miRNAs to some extent (Jing et al., 2005; Lim et al., 2005). Moreover, it has been shown that a plant miRNA with full complementarity to its target can also act at the translation level (Aukenan and Sakai, 2003). Hence, the mode of action of miRNAs might be more complicated than expected. A further layer of complexity occurs in plant systems in which a miRNA has been shown to direct asymmetric DNA methylation within the gene it targets (Bao et al., 2004). Whether such a mechanism operating in the nucleus at the DNA level can also take place in mammals is unknown.

Genetic manipulations and computational work have enabled several miRNA targets to be found in Caenorhabditis elegans, Drosophila melanogaster and Arabidopsis thaliana (in plants, target identification is relatively easy because most functional miRNAs are fully complementary to their targets).
They turn out to be involved in diverse biological functions including development timing, cell proliferation, cell death and fat metabolism. In vertebrates, addressing the function of miRNAs is challenged by the lack of genetic tools and much recent progress has relied on bio-informatic predictions. Several thousands of endogenous vertebrate miRNA targets have been isolated through computer-assisted approaches (John et al., 2004; Kiriakidou et al., 2004; Krek et al., 2005; Lewis et al., 2005; Xie et al., 2005). Although only little in vivo evidence of predicted targets exists so far (Poy et al., 2004; Yekta et al., 2004; Davis et al., 2005), these computational predictions of miRNA targets indicate that a broad range of mRNAs (estimated to be 10–30% of the population of mRNAs) are potentially targeted by miRNAs, thus leading to the concept that miRNAs are 'micromanagers' that may mediate combinatorial gene expression control (Bartel and Chen, 2004). But concretely, few genetic data about the biological role of vertebrate miRNAs are available. In zebraﬁ sh, the brain defects resulting from Dicer knock-out are rescued by miR-430, which constitutes a convincing argument that the miR-430 family is involved in brain morphogenesis (Giraldez et al., 2005). Deeper studies are absolutely required to learn more about the biological roles of vertebrate miRNAs during embryogenesis and/or in adults.

Imprinted small RNA genes

The use of computational methods and systematic cDNA cloning strategies have given rise to the identiﬁcation of a large collection of more than 100 C/D RNA and miRNA genes in two imprinted loci conserved in mammals: human 14q32 (the so-called Callipyge domain) and 15q11–q13 (the so-called Prader-Willi domain). They seem to share several structural and functional characteristics: (i) They are grouped into clusters of homologous repeated gene copies, with most – if not all – embedded within introns of large non-coding genes subjected to a complex spectrum of alternative splicing events. (ii) They are poorly conserved in evolution (i.e. they are not found in non-eutherian mammals). (iii) They display a tissue-speciﬁc expression pattern with preponderant expression in the adult brain. (iv) With few exceptions, they lack an obvious functional antisense element against a cellular transcript and their molecular and biological functions remain highly elusive.

Imprinted small RNA genes at the Prader-Willi locus

The human 15q11–q13 region is subjected to genomic imprinting. Two clinical disorders – the Prader-Willi and Angelman syndromes (PWS and AS, respectively) – are associated with defects in this imprinted locus. PWS is a rare developmental and complex behavioral disorder notably characterized by hypotonia, hyperphagia leading to obesity and moderate mental retardation. It results from the lack of expression of one or some paternally-expressed gene(s). AS patients manifest distinct neurological features and this syndrome is due to the absence of expression of UBE3A, a maternally-expressed gene (Nicholls and Knepper, 2001).

The PWS/AS locus includes paternally-expressed protein-coding genes: the makorin ring ﬁnger protein 3 gene (MKRN3), the MAGE-like 2 protein gene (MAGEL2), the neuronal growth suppressor gene NDN and the SNURF-SNRPN bicistronic gene (SNRPN upstream reading frame-small nuclear ribonucleoprotein N). Numerous paternally-expressed C/D snoRNA genes map downstream from SNURF-SNRPN. They are mainly organized into two large clusters: HBII-85 and HBII-52 clusters that contain 27 and 47 highly homologous snoRNA gene sequences, respectively (Cavaille et al., 2000; de los Santos et al., 2000; Meguro et al., 2001). HBII or MBII nomenclature is used for human or mouse brain snoRNAs of class II, respectively (Huttenhofer et al., 2001). Other C/D RNAs, not encoded by multiple repeated-
gene copies (HBII-436, HBII-437, HBII-438a/HBII-438b), are also found within this region (Fig. 2; Runte et al., 2001).

Most – if not all – C/D snoRNA genes are contained within the introns of a large paternally-expressed polycistronic transcript (over 460 kb) believed to initiate at the SNURF-SNRPN promoter region, that contains a cis-acting imprinting center, and to overlap the maternally-expressed UBE3A gene in an antisense orientation (Runte et al., 2001). Recent-ly, a complex array of tissue-specific, alternatively spliced RNA isoforms generated from this large transcription unit has been described. This may account for the different gene expression patterns of SNURF/SNRPN and the C/D RNA genes (Landers et al., 2004).

Fig. 2. Human imprinted domains and small RNA genes. The position of several imprinted genes at human imprinted domains is indicated by squares (protein-coding or large non-coding RNA genes), vertical bars (C/D snoRNA genes), or triangles (pre-miRNA genes). Maternally, paternally and bi-allelically expressed genes are represented in pink, blue and black respectively; genes whose imprinted status is not determined are filled in grey. Sense of transcription is indicated by arrows. Note that imprinted status of small RNA genes at the Dlk1-Gtl2 domain has only been demonstrated in the mouse. Names of putative or experimentally detected imprinted small RNA genes are written in red below their location. ‘n’ indicates the number of gene copies of snoRNA or pre-miRNA genes. Dotted lines show different splicing events that occur in the NESP/anti-NESP/GNAS region. The picture is not drawn to scale.

The paternally-expressed candidate gene(s) involved in PWS is (are) still largely unknown. Studies of knock-out mice models suggest that the candidate PWS gene(s) lie(s) between Snrpn and the MBII-52 gene cluster (reviewed in Nicholls and Knepper, 2001). Rare balanced translocations in human PWS patients have also enabled the identification of a minimal critical region that starts at intron 17 of SNRPN gene and extends to the end of the HBII-85 snoRNA gene cluster (Wirth et al., 2001; Gallagher et al., 2002; Schule et al., 2005). This minimal critical region includes 47 copies of HBII-85 and HBII-438a. Since the mouse 7C imprinted locus lacks HBII-438, this latter is unlikely to play a significant role in PWS. On the contrary, HBII-85 snoRNA has been proposed to be, wholly or partly, at the origin of PWS-associated pathogenesis.
(Cavaille et al., 2000; Wirth et al., 2001; Gallagher et al., 2002; Schule et al., 2005). More sophisticated studies are now required to determine how the lack of expression of HBII-85 might be linked to PWS etiology.

Any major role of HBII-52 in contributing to PWS has recently received support. Indeed, individuals carrying a deletion covering the entire HBII-52 cluster do not show any obvious clinical phenotype indicating that lack of expression of HBII-52 genes alone is not sufficient to promote PWS (Runte et al., 2005). Interestingly, HBII-52 has a conserved, 18-nucleotide-long antisense element against a brain-specific pre-mRNA, the serotonin receptor 5-HT$_2C$ (Cavaille et al., 2000). The 5-HT$_2C$ receptor is a trans-membrane signaling receptor coupled to heterotrimeric G proteins. Its mRNA possesses five A-to-I (adenosine to inosine) editing sites (A-E edited sites, Fig. 3A). A-to-I RNA editing is the site-specific modification of adenosine to inosine within precursor mRNAs that is catalyzed by members of the ADAR (adenosine deaminase acting on RNA) family. As translation occurs, an inosine is read as a guanine and consequently, a combination of 5-HT$_2C$ mRNA isoforms with different sequences exists and gives rise to sensitively different proteins, whose binding properties to G proteins differ. Importantly, the nucleotide predicted to be targeted by HBII-52 – the edited C-site – is known to play a pivotal role in regulating serotonergic signal transduction (Burns et al., 1997; Niswender et al., 1999), suggesting that MBII-52 might specifically modulate the efficiency of RNA editing at this site (Fig. 3A). Although direct proof of the in vivo role of HBII-52 is still lacking, we have recently shown that targeting the C-site for 2’O-methylation by MBII-52 significantly decreases the efficiency of RNA editing at that site specifically (Vitali et al., 2005). An alternative splice donor site in 5-HT$_2C$ pre-mRNA leading to a truncated form of the receptor is located 13 nucleotides from its segment of complementarity with HBII-52 (Fig. 3A). Thus, 2’O-methylation at the C site and/or the base-pairing between HBII-52 and the pre-mRNA could also regulate the choice of splice site by the splicing machinery (Cavaille et al., 2000).

Imprinted small RNA genes at the Callipyge locus

The 14q32 domain spans over roughly 1 Mb and contains three protein-coding genes expressed from the paternal allele: DLK1 (Delta-like 1), RTL1 (Retrotransposon-like 1), also called PEG11 (Paternally-expressed gene 11) in sheep, and DIO3 (type 3 deiodinase) genes. Interestingly, RTL1 sequence analysis reveals that it has homology to Ty3/gypsy retrotransposon (Lynch and Tristem, 2003; Youngson et al., 2005). It harbors an intronless open reading frame (ORF) potentially encoding a protein with the gag and pol domains of gypsy-like retroelements. An intriguing feature of this imprinted chromosomal domain is that it contains a complex population of large as well as small non-coding RNA genes. It includes Gil2, expressed as a spliced and polyadenylated RNA with multiple alternatively spliced isoforms, numerous C/D RNA and miRNA genes and two antisense transcripts to Rtl1 and Dio3 genes (note that seven and six miRNAs are processed from five pre-miRNA genes encoded in the murine and human anti-Rtl1 gene, respectively). Apart from anti-Dio3 whose imprinted expression remains to be demonstrated, all of these non-coding RNA genes are only expressed from the maternally inherited allele and their imprinted status is controlled by an intergenic, differentially methylated region (called IG-DMR) located between genes Gil2 and Dlk1 (Lin et al., 2003). It has been argued that all these non-coding RNA genes might belong to a single transcription unit starting from the Gil2 promoter region (Seitz et al., 2004a).

The snoRNA genes are arranged into two main clusters that contain nine and 31 copies of relatively degenerate 14q(I) and 14q(II) C/D RNA genes, respectively (Cavaille et al., 2002). In contrast to the PWS locus, the C/D RNA sequences at 14q32 are relatively poorly conserved in rodents, although the whole repeated gene organization is similar to that of the PWS domain, with most – if not all – C/D snoRNA genes embedded within introns of larger non-coding RNA genes (Rian and MEG8 in mouse and human, respectively). Surprisingly, the rat Dlk1-Gil2 domain is char-

---

**Fig. 3.** Potential functions of imprinted small RNA genes. (A) Predicted base pairing between MBII-52 and 5-HT2C pre-mRNA. The edited C-site (red arrow) predicted to be targeted by MBII-52 is located at the fifth nucleotide upstream from the D-box. An alternative donor splice site is boxed. The editing A to E sites are underlined and the corresponding amino acid changes in the receptor protein encoded by the unedited and edited mRNAs are indicated. Changes in amino acids are indicated.

(B) ‘The battle of sexes’: Maternally-expressed anti-Rtl1 miRNAs direct cleavages within Rtl1 transcripts that derive from the paternally-inherited chromosome. Hypothetical models involving a (some) miRNA(s) in Dlk1 non-productive translation as well as a role of repeated miRNA genes in allele-specific gene silencing are also proposed (see also text).
characterized by a unique ~100-kb-long C/D small RNA gene cluster that contains ~86 nearly identical C/D RNA gene copies (RBII-36), all of them being intron-encoded within a large non-coding RNA gene (Bsr) (Komine et al., 1999; Cavaille et al., 2001). Surprisingly, this small RNA gene cluster is not detected in other mammalian genomes available, including that of mouse. This might reflect a recent appearance only in the Rattus genus of rodents and to some extent recalls other imprinted genes that come from recent lineage-specific insertions in imprinted domains (Chai et al., 2001).

Most of the maternally-expressed miRNA genes locate into two clusters. In human, five pre-miRNA genes map within anti-RTL1 gene and the remaining ones (−40 in silico-predicted genes) are grouped downstream from the snoRNA gene cluster. In the mouse, most – but not all – pre-miRNAs are conserved with several of them being intron-encoded within Mirg, a non-coding RNA gene (Seitz et al., 2003, 2004a; Davis et al., 2005).

Although their spatio-temporal gene expression pattern in the developing embryo is still poorly characterized, C/D RNA and miRNA genes seem to share a similar expression pattern as all of them have been detected in brain, muscle, placenta as well as in pancreatic beta-cells (Cavaille et al., 2002; Seitz et al., 2003, 2004a; Poy et al., 2004; Davis et al., 2005). Lack of expression of the maternally-expressed non-coding RNA genes in mouse embryos with a paternal uniparental disomy for chromosome 12 (patUPD12) causes placental-megaly, skeletal muscle maturation defects and embryonic lethality, highlighting the presumably important functions of these maternally expressed non-coding RNAs in development (Georges et al., 2000).

The seven murine maternally-expressed miRNAs encoded within anti-RTL1 RNA (Fig. 3B) are obviously perfectly complementary to Rtl1 RNA, which led to the hypothesis that they might function as siRNAs to silence Rtl1/PEG11 (Seitz et al., 2003; Davis et al., 2005). Several lines of evidence support this assumption: (i) In knock-out mice deleted for their maternally-inherited IG-DMR, both chromosomes acquire a paternal identity. As a consequence, neither Gtl2 nor the small RNA genes are expressed while Dlk1, Rtl1 and Dio3 are expressed from both alleles. Consistent with a bi-allelic expression, Dlk1 and Dio3 mRNA levels were twice those observed in wild-type littersmates. Rtl1 gene expression, by contrast, is higher (~450%) than expected upon reactivation from the silent maternal allele, suggesting that miRNAs normally silence Rtl1 gene expression (Lin et al., 2003). (ii) The predicted cleavage sites within Rtl1 transcripts have been experimentally mapped in mouse placenta as well as in ovine skeletal muscle (Davis et al., 2005). This is the first evidence showing that RNAi-related mechanisms regulate imprinted genes.

The sheep Dlk1-Gtl2 imprinted domain has long been a centre of interest for geneticists because it is the site of a real genetic curiosity: the ‘Callipyge phenomenon’. The Callipyge (CLPG) phenotype in sheep is an inherited skeletal muscular hypertrophy in the hind quarters (‘Callipyge’ means ‘beautiful buttocks’ in greek), due to an increased proportion and diameter of fast twitch muscle fibers (Georges et al., 2003). It is due to a single base substitution from A to G in an intergenic region of the sheep telomeric end (Freking et al., 2002; Smit et al., 2003). The mode of inheritance of the CLPG phenotype is governed by an unusual parent-of-origin effect referred to as ‘polar overdominance’. The concept of overdominance refers to the fact that only heterozygous animals with the mutation inherited from the sire exhibit the phenotype. In other words, the following genotypes wtmat/wtmat, CLPGpat/wtmat, wtpat/CLPGmat and CLPGpat/CLPGmat are phenotypically: normal, Callipyge, normal and normal, respectively. In order to identify putative molecular effectors responsible for the phenotype, the expression of genes close to the mutation in hypertrophied muscles of CLPG sheep were profiled (Charlier et al., 2001). The authors report that the Clpg mutation has a cis enhancer effect on genes expressed from the chromosome holding it, without altering their imprinted status: when the mutation is maternally-inherited, levels of expression of Gtl2, the snoRNAs (Carole Charlier, personal communication) and miRNAs (Davis et al., 2005) are increased from the maternal chromosome; when the paternal allele is mutated, paternally expressed gene mRNA (RTL1 and Dlk1) levels are up-regulated. Hence, this suggests that RTL1 and/or Dlk1 might contribute to the muscular hypertrophy of CLPG sheep (Charlier et al., 2001). Indeed, ectopic expression of Dlk1 protein in skeletal muscles of transgenic mice induces a generalized muscular hypertrophy comparable to the sheep CLPG phenotype (Davis et al., 2004). Whether RTL1 over-expression also contributes to some extent to the induction of the phenotype is unknown.

In CLPGpat/CLPGmat individuals, the absence of Dlk1 protein despite relatively abundant Dlk1 mRNA has also been reported (Georges et al., 2003). Concomitantly, the expression of the maternally-expressed genes that have been tested is increased, shedding light on a situation in which miRNA genes from the locus are over-expressed. Considering the general role of miRNAs in translation repression, it has been hypothesized that maternally-expressed miRNAs repress Dlk1 translation through imperfect base-pairing (Fig. 3B) (Davis et al., 2004, 2005; Seitz et al., 2004a). Candidate complementarities between Dlk1 mRNA and miRNAs from the orthologous mouse locus are currently being tested. The case of 14q32 miRNAs repressing genes that are encoded into the same locus recalls the ‘intra-locus regulations’ occurring at the paralogous Hox clusters, in which genes for targets (Hox transcription factor genes) and triggers (miR-196 genes) are physically linked (Yekta et al., 2004).

The existence of maternally expressed negative regulators (miRNAs) of paternally imprinted genes (Rtl1 and also possibly Dlk1) adds a forceful argument to the conflict theory (Moore and Haig, 1991). Briefly, this theory explains the emergence of genomic imprinting by a conflict of interest between maternally versus paternally inherited alleles (e.g. the paternally expressed genes tend to act as growth-factors while maternally expressed genes mediate growth suppressor functions). In the case of intrinsic regulations within the locus at 14q32, we can observe a typical situation in which two paternally expressed genes, Rtl1 and Dlk1, are repressed
(or assumed to be repressed) by maternally expressed genes (miRNAs), indicating that the RNAi mechanism might have been recruited as a weapon by the maternal forces to annihilate paternal genes (Davis et al., 2005). In this regard, it is noticeable that Rtl1, Dlk1 and the miRNAs are expressed during development in the embryo and the placenta, which are the very sites of predilection for the parental conflict.

**Additional microRNA genes mapping near mammalian imprinted domains**

By systematic screening of all experimentally detected miRNAs deposited at the miRNA registry (http://www.sanger.ac.uk/cgi-bin/Rfam/miRNA/browse.pl) and of some in silico predicted, conserved pre-miRNA-like genes (Berezikov et al., 2005), we identified several additional candidate miRNA genes mapping within human or mouse imprinted chromosomal domains, strongly suggesting an imprinted expression (Fig. 2).

**miR-344 and the mouse 7C imprinted domain**

miR-344 was first isolated from rat embryonic primary cortical neurons (Kim et al., 2004). In mouse, miR-344 is encoded by four highly related gene copies predicted to map between Ndn and Snrpn genes (Fig. 2). Homologous miR-344 genes are not detected in the homologous human Prader-Willi domain at 15q11–q13 nor in the non-rodent genomes examined, which indicates that miR-344 might be specific to rodents.

**miR-335 and the human 7q32 imprinted domain**

miR-335 is intron-encoded within PEG1/MEST (intron 2), a paternally expressed gene encoded at the human 7q32 region (corresponding to mouse chromosome 6 sub-proximal region). Northern blot analysis using total RNA extracted from tissues isolated from embryos having uniparental disomy for chromosome 6 (a generous gift of Dr. J. Peter and C. Beechey) shows that miR-335 exhibits the same imprinted status as its host gene (M.L. Bortolin, unpublished). Two additional miRNA genes, miR-29a and miR-29b1, are also encoded at this locus but Northern blot analysis of their imprinted status has been hampered by the presence of two related miRNA species, miR-29c and miR-29b2, encoded on chromosome 1.

**miR-296 and the human 20q13 imprinted domain**

miR-296 gene maps at the GNAS locus (at human 20q13), an extremely complex imprinted domain, with alternatively spliced transcripts generated from multiple imprinted promoters (Fig. 2). In man, the miR-296 gene is positioned only 1.0 kb downstream from the 3' end of the GNAS antisense transcript RNA, leaving open the possibility that this latter might be the miRNA host gene. In mouse (distal chromosome 2), an additional miRNA gene without any clear human homologous counterpart, miR-298, is located nearby miR-296.

**In silico-predicted hairpin structures within imprinted domains**

Three conserved hairpin structures exhibiting pre-miRNA gene features (Berezikov et al., 2005) are found at two other human imprinted domains: 11p15.5 (can736 and can693) and 7q21 (can714). Can693 is positioned within an intergenic region between CDKN1C and KCNQ1 genes while can736 is located within an intron of the 5' untranslated part of the paternally expressed gene IGF2. However, it is transcribed in the opposite orientation to that of IGF2 (Fig. 2). In mouse, it maps only 0.3 kb from the 5' part of anti-IGF2 transcript. Can736 and can693 genes are embedded within CpG islands, the significance of which, if any, remains obscure. At human 7q21, can714 is homologous to four repeated sequences arranged in a head-to-tail manner within an intron of the untranslated region of PEG10, a paternally-expressed gene. Remarkably, these repeats have the potential to fold into a large and nearly perfect, double-stranded RNA structure (not shown). Further studies are now needed to formally demonstrate the expression of these in silico-predicted miRNA-like genes.

**Towards the understanding of imprinted small RNA genes**

Important issues persist after the characterization of about 200 bona fide or predicted imprinted small RNA genes. What are the molecular signaling pathways controlling their tissue-specific and/or developmentally regulated gene expression patterns? What are their physiological function(s) in the developing embryo, in placenta or in adult tissues? Are they involved in epigenetic regulatory pathways? Identifying their targets is the first step towards understanding their biological relevance.

Given that the vast majority of imprinted small RNA genes are devoid of full and conserved complementarities to any cellular transcripts, addressing the question of their target is a difficult task. First, imprinted small RNAs are specific to eutherian species (i.e. none of the ones tested were detected in chicken or even opossum brain; H. Seitz, unpublished results), which considerably limits the use of comparative sequence analysis for target identification (computer searches often rely on cross genome comparisons to predict conserved mRNA targets). Second, the functional parts of the small RNAs involved in target specificity (the C/D snoRNA sequences upstream from the D/D boxes or the 5' seed for miRNAs) can vary notably from one copy to the other, suggesting that they may target different RNA species or related members of a same gene family. Alternatively, due to the expected functional redundancy between repeated small RNA gene copies, some of them might be simply dispensable (i.e. mutations within functional RNA segments might accumulate without detrimental consequences). Alternatively, a role of imprinted C/D RNAs and miRNAs unrelated to conventional post-transcriptional regulations may also be envisioned (see below).
Recent advances strongly suggest that miRNAs and RNAi-related processes play a key role in neuronal functions, brain morphogenesis as well as in regulation of stem cell self-renewal and cell fate decisions (Cheng et al., 2005; Giraldez et al., 2005). Remarkably, the listing of new potential imprinted miRNA genes has shed light on a possible link between small RNA genes and imprinted loci known to be associated with behavioral phenotypes. HBII-85 C/D RNA genes at human 15q11–q13 are the very example because they represent credible gene candidates for causing Prader-Willi syndrome. Gene dysregulation at the imprinted locus hosting miR-296 (mouse distal chromosome 2) and miR-335 (mouse proximal chromosome 6) genes are also known to give rise to behavioral disturbances (Lefebvre et al., 1998; Plagge et al., 2005). In this regard, the targeted disruption of the Peg1/Mest gene on mouse proximal chromosome 6 induces abnormal maternal behavior including impaired placentophagia (Lefebvre et al., 1998). Given that expression of miR-335, which is intron-encoded in the Peg1/Mest gene, has not been tested, these phenotypes (or some part of them) might also be attributed, to some extent, to a lack of expression of miR-335. In the case of human 14q32/mouse distal 12 domains, an extensive analysis of the phenotypes in the adults is still lacking but given that imprinted small RNA genes at this locus are strongly expressed in the adult brain, brain-associated functions are probable. All in all, it is tempting to speculate that imprinted small RNA genes might participate in the etiology of disorders affecting brain functions.

The notion that miRNAs might be involved in viral infections of mammalian cells is emerging, supported by the first evidence that an miRNA (miR-32) is at the basis of an antiviral defense mechanism in human cells (Lecellier et al., 2005). It is worth noting that Rtl1 is a retrotransposon-like gene and its regulation by imprinted miRNAs brings the processes of RNAi, genomic imprinting and genome defense together. These potential links have been previously reviewed by Seitz et al. (2004b).

In plants, numerous cases of RNA-mediated transcriptional gene silencing are associated with specific DNA methylation. In *Schizosaccharomyces pombe*, Dicer-produced siRNAs derived from centromeric repeats are incorporated into RITS (RNA-induced transcriptional silencing complex), a RISC-related complex, to direct heterochromatin formation and gene silencing at the transcriptional level (Matzke and Bircher, 2005). Several lines of evidence in mammalian systems support the existence of nuclear RNA-mediated gene silencing mechanisms: (i) Small RNA-mediated RNA degradation is active in the nucleus (Robb et al., 2005). (ii) DNA methylation and histone H3K9 methylation can be triggered by small interfering RNAs (Kawasaki and Taira, 2004; Morris et al., 2004) although this may not occur in all experimental systems (Park et al., 2004; Svoboda et al., 2004). (iii) Dicer-deficient mouse ES cells show reduced levels of DNA and histone H3K9 methylation of centromeric DNA, high levels of centromeric repeat RNAs and transcriptional derepression of centromeric repeats (Kanellopoulou et al., 2005). A conditional loss-of-function mutant of Dicer in hybrid human-chicken cells also causes accumulation of transcripts from alpha-satellite sequences and leads to abnormal mitotic cells, presumably due to a defect in the formation of centromeric heterochromatin (Fukagawa et al., 2004; Kanellopoulou et al., 2005). Thus, Dicer-related RNAi machinery is required for the silencing of centromeric heterochromatin in vertebrates.

Given that repeated sequences are known to attract silencing and considering the involvement of small RNAs and the RNAi machinery in the construction of repressive chromatin in mammals, it is legitimate to propose that imprinted small RNAs – and more especially the large miRNA gene cluster at 14q32 – could account for certain epigenetic regulations at the basis of genomic imprinting. In many RNA-mediated gene silencing mechanisms, like X chromosome inactivation (mammals), RNA-directed DNA methylation (plants) or heterochromatization at the sexual Mat locus (*S. pombe*), the RNA and/or the RNAi machinery is mainly required to initiate the epigenetic state. Hence, miRNAs might guide epigenetic marking early during development (in the germ line), but once the imprints are installed on gamete haplogenomes, they might be dispensable for the subsequent maintenance and reading of the imprints. DNA methylation plays a central role in genomic imprinting. Interestingly, at mouse distal 12 locus four miRNA genes encoded in *Mirg* as well as the anti-Rtl1 miRNAs are embedded within (or are located next to) CpG islands. The same is true for the miRNA-like genes can693 and can736 at human 11p15.5. In the case of Rtl1 gene, because CpG islands are methylated in patUPD12 embryos (in which miRNA genes are no longer expressed), involvement of the RNAi machinery is unlikely (N. Youngson and A. Ferguson-Smith, personal communication). Concerning the remainder miRNAs, it is now of first interest to investigate their potential involvement in DNA methylation, including in non-canonical contexts as observed in plants (i.e. on cytosines in CpNpG and asymmetric sequence contexts). Alternatively or additionally, miRNAs could help to install an allele-specific chromatin state through recruitment of histone modification complexes. ‘Loss-of-function’ approaches are now indispensable to unravel the biological roles of these small RNAs and to decipher the biological and/or evolutionary significance of this potential link between imprinted loci and small RNA genes.

**Acknowledgements**

We sincerely apologize to all those colleagues whose work, due to space limitations, is not cited. We wish to thank Peter Winterton for careful reading of the manuscript, as well as Anne Ferguson-Smith, Michel Georges, Carole Charlier and all our colleagues from the CallimiR network. We also thank Michel Weber for scientific helpful discussions about the use of the UCSC Genome Browser.
Notes added in proof

Bentwich et al. (2005) recently reported the characterization of many primate-specific microRNAs which are only detected in the placenta and whose genes are organized into a cluster at the human 19q13.4 domain, very close to three previously described miRNA genes (miR-371, miR-372 and miR-373) specifically expressed in human embryonic stem cells. This large cluster is reminiscent to some extent to that described at the human imprint-  
ed 14q32 locus (Cf. Fig. 2). Although these microRNA genes are located relatively far away from the paternally expressed PEG3/ZIM2 gene, it is now important to test their imprinted status.

Many in silico-predicted microRNA genes at the imprinted Dlk1-Gtl2 domain described by Seitz et al. (2004) have now been experimentally detected in human by Bentwich et al. (2005) and Altuvia et al. (2005).


