Transcription is required for establishment of germline methylation marks at imprinted genes

Mita Chotalia,1,4 Sebastien A. Smallwood,1 Nico Ruf,1 Claire Dawson,1 Diana Lucifero,1 Marga Frontera,1 Katherine James,2 Wendy Dean,1 and Gavin Kelsey1,3,5

1Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB2 2AT, United Kingdom; 2Emmanuel College, University of Cambridge, Cambridge CB2 3AP, United Kingdom; 3Centre for Trophoblast Research, University of Cambridge, Cambridge, United Kingdom

Genomic imprinting requires the differential marking by DNA methylation of genes in male and female gametes. In the female germline, acquisition of methylation imprint marks depends upon the de novo methyltransferase Dnmt3a and its cofactor Dnmt3L, but the reasons why specific sequences are targets for Dnmt3a and Dnmt3L are still poorly understood. Here, we investigate the role of transcription in establishing maternal germline methylation marks. We show that at the Gnas locus, truncating transcripts from the furthest upstream Nesp promoter disrupts oocyte-derived methylation of the differentially methylated regions (DMRs). Transcription through DMRs in oocytes is not restricted to this locus but occurs across the prospective DMRs at many other maternally marked imprinted domains, suggesting a common requirement for transcription events. The transcripts implicated here in gametic methylation are protein-coding, in contrast to the noncoding antisense transcripts involved in the monoallelic silencing of imprinted genes in somatic tissues, although they often initiate from alternative promoters in oocytes. We propose that transcription is a third essential component of the de novo methylation system, which includes optimal CpG spacing and histone modifications, and may be required to create or maintain open chromatin domains to allow the methylation complex access to its preferred targets.

[Keywords: Genomic imprinting; DNA methylation; oocytes; transcription]

Supplemental material is available at http://www.genesdev.org.

Received July 4, 2008; revised version accepted November 12, 2008.

Genomic imprinting in mammals is controlled by discrete elements, termed imprinting control regions (ICRs), that become methylated in male or female gametes and retain this modification as a memory of parental origin in the somatic tissues of the conceptus, where it is acted upon to cause monoallelic silencing of imprinted genes [Reik and Walter 2001a]. All imprinted domains possess differentially methylated regions [DMRs] with germline-derived methylation [although only a subset have been functionally verified as ICRs], and most such marks are laid down in the female germline [Reik and Walter 2001b]. Establishment of methylation imprints in oocytes occurs postnatally, in growing oocytes arrested in meiosis I [Lucifero et al. 2004; Hiura et al. 2006], and requires the cooperative activities of the DNA methyltransferase Dnmt3a and its nonenzymatic cofactor Dnmt3L [Bourc’his et al. 2001; Hata et al. 2002; Kaneda et al. 2004]. Dnmt3L acts in the maternal germline apparently specifically for the establishment of DMRs [Bourc’his et al. 2001; Hata et al. 2002].

DMRs are CpG-rich sequences that fulfill the criteria of CpG islands [CGIs], but the properties that distinguish them from the majority of CGIs and result in their gametic methylation remain to be fully elucidated. At a sequence level, DMRs frequently contain tandem repeats [Neumann et al. 1995; Hutter et al. 2006; Reinhart et al. 2006], whose functional significance is supported by some transgenic studies [Neumann et al. 1995; Reinhart et al. 2006]; however, not all maternal germline DMRs possess this feature [Arnaud et al. 2003], and tandem repeats also occur in the CGIs of apparently nonimprinted loci [Yamada et al. 2004]. Structural studies have revealed that the Dnmt3a/3L complex methylates DNA templates in a periodic pattern, with an optimal spacing of CpGs of 8–10 base pairs [bp] [Jia et al. 2007; Jurkowska et al. 2008]. Furthermore, it has been suggested that this periodicity characterizes maternal germline DMRs [Jia et al. 2007]. This finding could help explain the prevalence
of tandem repeats in DMRs, but the degree to which an 8- to 10-bp CpG periodicity is specific to DMRs remains to be confirmed [Ferguson-Smith and Greally 2007]. Biochemical studies have shown that Dnmt3L interacts with histone H3, but only when H3 is unmethylated at Lys 4 [Ooi et al. 2007], revealing that the chromatin signature of target sequences is also an important factor. Methylated H3K4 is recognized as a mark of active promoters and transcription units [Bernstein et al. 2005; Kim et al. 2005].

Other evidence indicates that sequences outside of DMRs can make an essential contribution to methylation. DMRs on transgenes often fail to adopt imprinted methylation, unless in the context of large constructs or in the presence of additional sequence elements (e.g., Sleutels and Barlow 2001). In two human imprinted gene syndromes, deletions many kilobases from DMRs have been shown to disrupt methylation. Maternally transmitted microdeletions upstream of the SNRPN DMR cause Angelman syndrome (AS) as a result of loss of methylation (LoM) of this DMR, the overlap between deletions defines a <1-kb region 35kb upstream of the DMR that appears to be required for methylation [Buiting et al. 1999]. In Pseudohypoparathyroidism type 1b [PHP1b], an endocrine disorder associated with imprinting mutations in the GNAS locus, maternal allele methylation of the GNAS DMRs is disrupted by deletions in the neighboring NESP55 or STX16 genes, the latter being ~220 kb from the affected DMR [Bastepe et al. 2003, 2005]. How such remote elements are involved in methylation of the associated DMRs is unclear, but their existence strongly suggests that DMR sequence alone is insufficient.

A potential factor that has not been explored in any detail is transcription, although there are lines of evidence suggestive of a role for transcription in methylation of germline DMRs, and a transcription-based mechanism could provide a unifying explanation for the observations above. Many maternal germline DMRs are contained within transcription units, including those at four recently evolved imprinted retrotransposons [Wood et al. 2007], and in general, intragenic CGIs are more likely to be methylated than those at promoters [Yamada et al. 2004; Illingworth et al. 2008]. In a transgene study, the Air DMR was appropriately methylated in some mouse lines when located in an intron of an Aprt transgene [Sleutels and Barlow 2001]. Finally, in somatic cells, noncoding antisense transcripts have been implicated in methylation of the “somatic” DMRs through which they transcribe [Sleutels et al. 2002; Williamson et al. 2006; Shin et al. 2008], although the mechanistic aspects of this remain to be established [Pauler et al. 2007]. Here, we directly test the hypothesis that transcription is required for establishment of methylation of DMRs in female gametes. At several maternally marked imprinted domains, we detect transcripts traversing the DMRs in growing oocytes prior to and at the time of de novo methylation. Using the mouse Gnas locus as a model, we show that disruption of the Nesp transcript, which initiates furthest upstream in this imprinted domain, prevents normal establishment of methylation of the germline DMRs. In striking contrast to monoallelic silencing of imprinted genes in somatic tissues by antisense transcription, the transcripts we believe to be involved in gametic methylation are protein-coding. Our results thus reveal a novel, transcriptionally mediated component in imprint establishment. Transcription across DMRs in germ cells may be essential for remodeling histone modifications or to create chromatin domains permissive for de novo methylation. These findings provide a basis for understanding how new imprinted loci evolve and a novel molecular explanation for imprinting errors.

Results

Transcription across the Gnas locus DMRs in growing oocytes

We are using the mouse Gnas locus to investigate mechanisms of imprint establishment. The locus comprises a series of overlapping, imprinted, protein-coding transcripts Gnas, Gnasxl, and Nesp and the imprinted noncoding transcripts Nespas and 1A (Fig. 1A). An extensive DMR covering the Gnasxl and Nesp promoters and a second DMR covering the 1A promoter govern monoallelic expression of these transcripts [Williamson et al. 2004, 2006; Liu et al. 2005]. Both DMRs are methylated on the maternal allele and acquire methylation in oocytes by a Dnmt3L dependent mechanism [Liu et al. 2005; Arnaud et al. 2006]. An additional DMR with paternal allele methylation covers the Nesp exons but is methylated after fertilization [i.e., a somatic DMR], while the Gnas promoter is located within a constitutively unmethylated CGI [see Fig. 1B; Liu et al. 2000]. This pattern of methylation makes Gnas a particularly informative imprinted cluster in which to investigate how imprints are established. Any model must explain why adjacent elements like the 1A DMR and Gnas CGI are marked differently. We are especially interested in the Nesp transcript (which in somatic tissues is expressed only from the maternal allele), because its start site is furthest upstream in the locus, such that transcription events initiating at the Nesp promoter traverse the whole cluster, including the two germline DMRs. In addition, maternally transmitted deletions of the NESP region in humans are associated with LoM of the NESPAS/GNASXL and 1A DMRs in some PHP1b families [Bastepe et al. 2005]. To establish a possible role for Nesp transcription in methylation acquisition, we first showed by RT–PCR, using primer sets that detect transcription events that traverse the maternal germline DMRs at Nespas/Gnasxl and 1A, that Nesp transcripts could be detected in growing oocytes obtained at 5, 10, and 15 d post-partum [dpp] [Fig. 1C]. These germline DMRs start to become methylated from day 15 [Fig. 1B]. In contrast, Nespas, Gnasxl, and 1A transcripts were not detected by RT–PCR during oocyte growth under the same conditions in which Nesp was readily detected [Fig. 1C], suggesting that the Nespas, Gnasxl, and 1A promoters are inactive even before the onset of methylation of their DMRs, whereas Gnas transcripts were present throughout. Interestingly, by 5’RACE analysis, we found that the start site utilized by Nesp transcripts in oocytes was distinct.
from the major start site used in somatic tissues (Supplementary Fig. 1). In conclusion, transcription from Nesp occurs across the DMRs in oocytes before and possibly during methylation establishment, while the DMRs themselves may correspond to inactive promoters.

A targeted allele to truncate Nesp transcription

As a functional test of the role of Nesp transcription, we designed a targeted allele to prematurely terminate transcription. We replaced the splice donor site of the second Nesp exon with a cassette comprising the final exon, polyadenylation, and downstream sequences of the rabbit β-globin gene [Fig. 2A,B]; this cassette has been used effectively elsewhere to terminate noncoding RNAs to examine their effect in somatic tissues on imprinted monoallelic expression (Sleutels et al. 2002; Shin et al. 2008). After germline transmission of the targeted allele and excision of the selection markers (this allele is termed Nesptrun), we checked for the effect on the Nesp transcript. RT–PCR assays of newborn brain tissue (where Nesp is highly expressed) failed to detect the normal Nesp transcript after maternal transmission of the Nesptrun allele (Fig. 2C). Complete ablation of the normal Nesp transcript was verified by real-time PCR (data not shown), and 3’RACE analysis confirmed that the expected premature polyadenylation event had

Figure 1. The Nesp transcript is detected in growing oocytes coincident with establishment of methylation of germline DMRs in the Gnas locus. (A) Scheme of the mouse Gnas locus (Plagge and Kelsey 2006), showing the organization of the overlapping, protein-coding transcripts Nesp, Gnasxl, and Gnas and the noncoding Nespas and 1A transcripts. Transcripts expressed from the maternal allele are indicated above the line, from the paternal allele below the line; Gnas exhibits tissue-specific imprinting, with repression of the paternal allele in a subset of tissues. The location of the DMRs is shown by the rows of filled circles on the methylated allele; the Gnas promoter resides in a biallelically unmethylated CGI (open circles). The positions of the PCR products for bisulphite analysis in B and Figures 3 and 4 are indicated by the hatched bars, and the primers for the RT–PCR assays in C and Figure 2 by labeled arrows. (B) Bisulphite sequences of the Nespas/Gnasxl and 1A DMRs in oocytes isolated at postnatal days 5, 10, and 15 and mature metaphase II (MII) oocytes from adult females. Each row represents the CpG sites of an individual sequenced clone, with filled circles depicting methylated CpG sites (missing circles represent CpGs for which sequence was ambiguous). Sequences obtained from two independent bisulphite treatments are indicated by bracketed sets of methylation profiles. Methylation of these DMRs in MII oocytes has been described previously (Liu et al. 2000; Coombes et al. 2003). (C) RT–PCRs for Nesp, Nespas, Gnasxl, 1A, and Gnas transcripts in day 5, 10, 15 oocytes. Co indicates amplification control with cDNA from an E13.5 embryo; H indicates no template control. The left lanes on each gel show a 100-bp marker ladder, with the 100- or 200-bp markers indicated.
occurred (Fig. 2C). Finally, analysis of oocytes from Nesptrun/trun homozygous females revealed that the Nesp transcript that normally traverses the DMRs was undetectable (Fig. 2D) and the expected truncation product was detected by 3'9RACE (Supplemental Fig. 2), suggesting that transcription had indeed been terminated upstream of the germline DMRs.

Nesp truncation causes loss of germline methylation at the Gnas locus

To examine whether truncation of Nesp altered the methylation of the germline DMRs, we first analyzed DNA of tissues from newborn pups (denoted Nesp\textsuperscript{trun/+}) inheriting the Nesp\textsuperscript{trun} allele maternally. Nesp\textsuperscript{trun} carrier females were crossed with C57BL/6j males or with males carrying the Gnas region of chromosome 2 from Mus spretus to enable discrimination of parental alleles using sequence polymorphisms. In the carrier females, the Nesptrun allele was of paternal origin, so that their progeny represented the first maternal transmission of the truncation allele. All samples were tested by combined bisulphite and restriction analysis (COBRA) (representative COBRA results are shown in Supplemental Fig. 3), and PCR products from a subset were cloned and
sequenced to determine complete methylation profiles. As shown in Figures 3A and 4A, pups inheriting the \( Nesp^{\text{tran}} \) allele from females showed LoM of the maternal germline DMRs at \( Nespas/Gnasxl \) and/or 1A. Where informative polymorphisms were present [as in amplicons F3 and F5 in Fig. 3A], it was possible to confirm in the bisulphite sequence profiles that the normally methylated maternal allele had indeed lost methylation. Unexpectedly, the \( Nesp^{\text{tran}/+} \) mutants differed in their degree of LoM [Fig. 3A,B], and this was the case even among littermates. All pups inheriting the \( Nesp^{\text{tran}} \) allele maternally fully lacked methylation of the 1A DMR [amplicon F6 in Fig. 3]. In 68% of \( Nesp^{\text{tran}/+} \) pups, there was also LoM of the \( Gnasxl \) exon and promoter [amplicons F4, F5], and in 24%, LoM extended further upstream to the \( Nespas \) first exon [amplicon F2]. This last category therefore had complete LoM of all the maternal germline DMRs (e.g., mutants #A25 and #A62 in Fig. 3A). In this class of \( Nesp^{\text{tran}/+} \) mutant, we would expect expression of the antisense \( Nespas \) on the maternal allele to be derepressed, which should result in acquisition of methylation on the maternal allele of the somatic DMR at \( Nesp \) [Williamson et al. 2006], and this was found to be the case [amplicon F1 in Fig. 3A]. No LoM at any DMR was observed in seven wild-type littermate controls; paternal transmission of the \( Nesp \) truncation also had no effect on

Figure 3. Maternal transmission of the \( Nesp \) truncation results in loss of methylation of maternal germline DMRs. [A] Bisulphite sequence profiles from neonatal brain DNA from \( Nesp^{\text{tran}/+} \) and wild-type [+/-] pups. A total of 34 \( Nesp^{\text{tran}/+} \) pups was analyzed by COBRA [see B], and sequences for one or more amplicon were obtained from seven mutants. Representative sequences from \( Nesp^{\text{tran}/+} \) mutants with complete LoM [#A25, #A62] or partial LoM [#A43, #A47] of the maternal germline DMRs are shown. Sequence polymorphisms allowed maternal and paternal alleles to be discriminated for products F1, F3, and F5. [B] Summary showing the frequency of LoM across the DMRs in neonatal brain in \( Nesp^{\text{tran}/+} \) mutants collected from 11 litters. The number of \( Nesp^{\text{tran}/+} \) pups typed by COBRA for amplicons F2, F3, and F6 was 34; for F4 was 22, and for F5 was 26; seven wild-type pups were similarly scored and found to have the normal pattern of DMRs. These data are a combination of COBRA results from crosses of \( Nesp^{\text{tran}/+} \) females with C57BL/6j males and with males carrying a \( M. \) spretus-derived Gnas locus; the variation in methylation pattern among \( Nesp^{\text{tran}/+} \) pups was similar in the two crosses. See Figure 1A for the location of the amplicons within the DMRs.
residual methylation of the Nespas/Gnasxl DMR [Fig. 4B]. This confirmed a failure of methylation in the maternal germline, and also that the variable degree of LoM observed among Nesp\textsuperscript{trans}/+ pups most likely reflects differences among oocytes in the degree of LoM across the locus. (We cannot exclude that some modulation of methylation patterns may also occur after fertilization.) We do not know the reason for the partial establishment of methylation at the DMR closest to the β-globin cassette. The directionality of the partial methylation suggested that the cassette may not be fully effective at terminating transcription downstream from Nesp, so that if some run-through transcription occurs across a prospective DMR during oocyte growth, it might be sufficient to allow methylation, particularly those most proximal to the cassette. Although there was no evidence for run-through transcription in our 3' RACE analysis [Fig. 2B], this technique may not be sufficiently sensitive to pick up longer, less abundant transcription products that might occur if transcription termination is not fully effective. Real-time PCR assays for primary transcripts at the Nespas/Gnasxl DMR in Nesp\textsuperscript{trans}/+ brain RNA did reveal the expected substantial down-regulation of transcription but were not sensitive enough to exclude the possibility of residual transcription downstream from the termination cassette (data not shown). Alternatively, there may be an effect on local chromatin organization downstream from transcription termination that permits some acquisition of methylation of the DMR sequences in closest proximity to the cassette.

**Disrupted imprinted expression of Gnas locus transcripts as a result of loss of germline methylation**

The LoM observed in Nesp\textsuperscript{trans}/+ pups is predicted to result in altered imprinting and expression of the Gnas locus transcripts, but the allelic expression profile should now depend on the precise methylation pattern in a given mutant. To investigate this, we typed mutants for methylation pattern by COBRA and separated them into three epigenotype classes based on the extent of LoM (Supplemental Fig. 5A). RNA was extracted from brain and brown adipose tissue, real-time RT–PCR used to quantify expression levels, and the Sequenom MassARRAY system used to determine allelic usage, based on a sequence polymorphism between C57BL/6 and spretus derived alleles in Gnas exon 11 that is common to all sense transcripts of the locus. Expression of the Nesp, Gnasxl, IA, and Gnas transcripts was found to be altered largely as expected from the methylation profile in the three mutant epigenotype classes (Supplemental Fig. 5B,C). For example, LoM of the IA DMR was associated with derepression of the maternal allele of the IA transcript, and this was accompanied by down-regulation of the maternal allele of Gnas in brown adipose tissue but not in brain, consistent with the role of the IA DMR or transcript in tissue-specific imprinting of Gnas [Williamson et al. 2004; Liu et al. 2005]. Nesp\textsuperscript{trans}/+ pups had anomalous phenotypes consistent with deregulated imprinting of the Gnas locus. A proportion of pups was noticeably heavier and edematous and had the broad-necked appearance.
Transcription and imprint establishment

(Date of publication) [Plagge and Kelsey 2006]. On typing by COBRA, these pups were found to have LoM confined to the 1A DMR (Supplemental Fig. 6B), which causes reduced expression of maternal Gnas in tissues with imprinted expression [as in Supplemental Fig. 5]. The remaining Nespm+/− pups were lighter at birth and were found to have more extensive LoM, giving up-regulation of maternal Gnas transcripts as well as reduction in maternal Gnas. All epigenotype classes had very poor postnatal survival, with only a small proportion of the pups with LoM restricted to the 1A DMR surviving to weaning, pups with more extensive LoM did not survive more than a few days after birth (Supplemental Fig. 6C).

Table 1. Locations of germline DMRs in imprinted loci

<table>
<thead>
<tr>
<th>Locus [DMR]</th>
<th>Chromosome</th>
<th>DMR methylation</th>
<th>DMR location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gnas (Nesp/Gnasx1)</td>
<td>2</td>
<td>Maternal</td>
<td>Intron</td>
<td>Coombes et al. 2003</td>
</tr>
<tr>
<td>Gnas (1A)</td>
<td>2</td>
<td>Maternal</td>
<td>Intron</td>
<td>Li et al. 2000</td>
</tr>
<tr>
<td>Mcts2</td>
<td>2</td>
<td>Maternal</td>
<td>Intron</td>
<td>Wood et al. 2007</td>
</tr>
<tr>
<td>Peg10</td>
<td>6</td>
<td>Maternal</td>
<td>Promoter</td>
<td>Ono et al. 2003</td>
</tr>
<tr>
<td>Peg1</td>
<td>6</td>
<td>Maternal</td>
<td>Intron</td>
<td>Lucifero et al. 2002</td>
</tr>
<tr>
<td>Nap11s</td>
<td>6</td>
<td>Maternal</td>
<td>Intron</td>
<td>Wood et al. 2007</td>
</tr>
<tr>
<td>Peg3</td>
<td>7</td>
<td>Maternal</td>
<td>Intron</td>
<td>Kim et al. 2003</td>
</tr>
<tr>
<td>Sarpn</td>
<td>7</td>
<td>Maternal</td>
<td>Intron</td>
<td>Mapendano et al. 2006</td>
</tr>
<tr>
<td>Kcnq1 (KvDMR)</td>
<td>7</td>
<td>Maternal</td>
<td>Intron</td>
<td>Yatsuki et al. 2002</td>
</tr>
<tr>
<td>Inpp5f</td>
<td>7</td>
<td>Maternal</td>
<td>Intron</td>
<td>Wood et al. 2007</td>
</tr>
<tr>
<td>Zac1</td>
<td>10</td>
<td>Maternal</td>
<td>Intron</td>
<td>This study</td>
</tr>
<tr>
<td>Grb10</td>
<td>11</td>
<td>Maternal</td>
<td>Intron</td>
<td>Arnaud et al. 2003</td>
</tr>
<tr>
<td>U2af1-rs1</td>
<td>11</td>
<td>Maternal</td>
<td>Intron</td>
<td>Wood et al. 2007</td>
</tr>
<tr>
<td>Peg13</td>
<td>15</td>
<td>Maternal</td>
<td>Intron</td>
<td>Ruf et al. 2007</td>
</tr>
<tr>
<td>Slc38a4</td>
<td>15</td>
<td>Maternal</td>
<td>Promoter</td>
<td>G. Kelsey [unpubl.]</td>
</tr>
<tr>
<td>Igf2r (Air)</td>
<td>17</td>
<td>Maternal</td>
<td>Intron</td>
<td>Stoger et al. 1993</td>
</tr>
<tr>
<td>Impact</td>
<td>18</td>
<td>Maternal</td>
<td>Intron</td>
<td>Okamura et al. 2000</td>
</tr>
<tr>
<td>H19</td>
<td>7</td>
<td>Paternal</td>
<td>Intergenic</td>
<td>Tremblay et al. 1997</td>
</tr>
<tr>
<td>Rasgfr1</td>
<td>9</td>
<td>Paternal</td>
<td>Intergenic</td>
<td>Shibata et al. 1998</td>
</tr>
<tr>
<td>Dlk1-Gtl2 (IG-DMR)</td>
<td>12</td>
<td>Paternal</td>
<td>Intergenic</td>
<td>Takada et al. 2002</td>
</tr>
</tbody>
</table>

*Evidence for upstream start sites from ESTs.
DMR promoter were only weakly detected in growing oocytes. Perhaps surprisingly, even for many of the genes for which the DMR is in an intron of the characterized transcription unit (e.g., Grb10, Igf2r, Kcnq1, Zac1, and Impact), showing the locations of the primers used for RT–PCRs in relation to the germline DMRs (filled circles). The schemes are not to scale, and for simplicity, not all exons of the genes are shown. Characterized start sites or start sites determined by 5’RACE analysis are indicated by arrows, with those above the line representing start sites detected in somatic cells and those below the line in oocytes. Novel exons identified by 5’RACE are shown as gray boxes. On the right, RT–PCRs for these loci for transcripts traversing the DMRs in day 5, day 10, day 15 and MII oocytes are shown. For both Igf2r and Grb10, RT–PCRs labeled with Ex1-F/Ex3-R1 assay transcripts from the canonical somatic promoter and RT–PCRs labeled with Un1-F/Ex3-R2 assay the novel start sites identified in oocytes by 5’RACE. The multiple bands for Zac1 reflect alternative splicing of the various 5’ untranslated exons, exon 8 being the first coding exon. (B) RT–PCR analysis of transcripts for the H19 paternal germline DMR. The RT–PCR represents amplicon 4 from Schoenfelder et al. [2007]. (C) RT–PCR analysis of transcripts crossing the intragenic CGIs of four nonimprinted genes. Gene Sp6 represents locus PvuII 44, Chst8 is PstI 53, Sema6c is PstI 58, and Adra1b is PvuII 66 from Song et al. (2005). In the schemes to the left, the locations of the CGIs are represented by the open circles. In each panel, the lanes marked Co represent control amplification from E13.5 embryo RNA, and H indicates no template control. The left lanes on each gel show a 100-bp marker ladder, with size marker relevant to the RT–PCR product indicated by arrows.

In conclusion, the frequent location of maternal germline DMRs within transcription units and the fact that transcripts traversing these DMRs are detected in growing oocytes suggest that transcription is a common requirement for de novo methylation of maternal imprint marks.

Discussion
In this study, we provide evidence that transcriptional events are required for the establishment of DNA methylation imprints in the female germline. Functional
evidence in support of this concept was provided by truncating the Nesp transcript at the Gnas locus, which resulted in LoM of the maternal germline DMRs. LoM of the 1A DMR was fully penetrant on maternal transmission of the truncation allele, as well as in oocytes, but there was variable LoM of the Nesp/Gnasxl DMR. In humans, deletions of the NESP55 region also result in LoM of the GNAS DMRs on the maternal allele [Bastete et al. 2005]. We also suggest that transcription is a common requirement for maternal germline DMRs, because most such DMRs are located within transcription units, and where this is apparently not the case, further examination provided evidence for alternative upstream start sites used in oocytes. Furthermore, for all eight maternal germline DMRs we tested, transcripts traversing the prospective DMRs were detected in growing oocytes. In contrast, this was not the case for the H19 DMR, which is methylated in the male germline, and intragenic CGIs of nonimprinted loci were also not consistently transcribed in oocytes and serve here as biologically appropriate controls.

Possible mechanistic links between transcription and DNA methylation establishment

What is the role of transcription and how can it be incorporated into a model that accounts for the normal methylation pattern of Gnas and other imprinted loci? We expect that the logic that explains imprinted methylation at Gnas will apply to all maternally marked imprinted domains. Based on the interaction of Dnmt3L with histone H3 [Ooi et al. 2007], we anticipate that the prospective DMRs at Nesp/Gnasxl and 1A have low levels of methylated H3K4 in growing oocytes, which could reflect relative inactivity of the associated promoters in growing oocytes. Consistent with this possibility, we found that Nesp, Gnasxl, and 1A transcripts were not readily detected during oocyte growth. The same model might apply to the Zac1 promoter DMR. We suggest that if prospective DMRs are inactive promoters with unmethylated H3K4, then they are likely to be packaged into relatively condensed chromatin, so transcription from upstream promoters may be necessary to create or maintain an open chromatin environment that allows Dnmt3a/Dnmt3L access to its preferred targets. Transcription might also play a more active role. There could be a direct interaction between nascent transcripts or the transcription complex and the de novo methylation machinery. Alternatively, transcription could recruit factors, such as LSD1 or SMCX [Shi et al. 2004; Iwase et al. 2007], necessary for demethylating H3K4 at the DMRs. Future experiments will seek to address the temporal relationship between ongoing transcription and de novo methylation.

It is important to highlight in our model that none of the factors implicated in DMR methylation—CpG periodicity, promoter activity/H3K4 methylation status, and transcription—is sufficient by itself. Rather, we expect that all have to apply in order for a DMR to become methylated. This is illustrated, for example, by the Gnas CGI. It has a CpG periodicity apparently similar to the Nespas/Gnasxl and 1A DMRs [data not shown], and transcripts initiating at Nesp in oocytes traverse the CGI just as they do these DMRs. In this case, it might be that the methylation state of H3K4 at the Gnas CGI is inappropriate for Dnmt3L engagement, and the fact that Gnas transcripts are readily detected in oocytes suggests that the Gnas CGI houses an active promoter marked by H3K4me3 and protected from DNA methylation. Similarly, although intragenic CGIs of some nonimprinted loci may be transcribed in oocytes, without the appropriate CpG periodicity or H3K4 methylation state they would not be targets for Dnmt3a/Dnmt3L.

Oocyte transcription units and the evolution of imprinted loci

It is notable that in all the cases we studied, the transcripts traversing the DMRs and that we propose are involved in methylation establishment are protein-coding transcripts, and this sets them apart from the noncoding antisense transcripts, such as Air, Rcnqiot1, and Nespas, which are involved in the silencing and methylation of imprinted promoters in somatic tissues [Pauel et al. 2007]. This suggests that they encode proteins with functions in the oocyte, or the preimplantation embryo, and the activity of these transcription units therefore may have provided the opportunity for imprinting to arise. It has been proposed that imprinted domains could evolve through insertion of CpG-rich elements, such as retroposons [Wood et al. 2007], into transcription units active in oocytes, and this is fully consistent with our model. Such events could be the origin of DMRs such as Air and KvDMR1, which became the promoters for long, noncoding RNAs that subsequently came to orchestrate domain-wide monoallelic expression. In many cases, we detected alternative promoters for transcripts in oocytes distinct from the major promoters used in somatic tissues (which, generally, did not alter their coding potential). This suggests a new possibility for the evolution of imprinting, whereby recruitment of oocyte-expressed promoters upstream of susceptible CGIs could give rise to new DMRs and establish novel imprinted domains. This might explain loci such as Grb10, Snrpn, and Zac1 that are characterized by multiple promoters for transcripts with the same coding potential. Consistent with this possibility, while the Gnas locus is highly conserved, we did not find evidence for a Nesp homolog in noneutherian mammals. In marsupials, the Gnas locus does contain a Gnasxl homolog, but neither the Gnasxl CGI nor any other CGI in the locus is a DMR [G. Kelsey and E. Ivanova, unpubl.]. It is possible that the oocyte promoters of these “imprinting” transcripts represent a specific subset of promoters with shared, specific transcription factors and that subsequently have been selected especially for their ability to imprint genes.

Implications for imprinted gene disorders

Our model also has important implications for the nature of genetic defects giving rise to imprinting disorders in
humans. It offers an explanation for how microdeletions disrupt imprinting of GNAS in PHP1b [Bastepe et al. 2003, 2005] or cause LoM of the SNRPN DMR in AS [Buiting et al. 1995; Dittrich et al. 1996], and this should encourage a search for similar lesions in other imprinted disorders. SNRPN is characterized as having multiple, alternative upstream start sites [Buiting et al. 1995; Dittrich et al. 1996], so the AS microdeletions could involve essential oocyte promoters or enhancers or otherwise prevent germ-line transcription across the SNRPN DMR. In mouse oocytes deficient in the KRAB finger protein Zfp57, the Snrpn DMR is unmethylated [Li et al. 2008], and it will be interesting to examine whether this factor is involved in regulating transcription events across the Snrpn DMR. It is interesting to note that an alternative upstream promoter has been described for human ZAC1 [Valleley et al. 2007], similar to that we describe here for mouse Zac1, and such elements become candidate regions to analyze in other imprinted disorders. It is also possible that defects affecting multiple imprinting loci, such as cases of maternal hypomethylation syndrome [Mackay et al. 2006], could arise because of mutations within transcription factors required for the expression of subsets of the oocyte transcripts implicated here in imprint establishment. Finally, differences in the timing of the transcription events through DMRs could explain why imprinting of the various domains appears to be established at different times during oocyte growth [Obata and Kono 2002; Lucifero et al. 2004; Hiura et al. 2006].

Materials and methods

Isolation of growing and mature oocytes

Growing oocytes were collected from C57BL/6J mice [Charles River Laboratories] at 5, 10, and 15 dpp. Ovaries from two to three mice were dissected in PBS [pH 7.2] and digested for 10–15 min at 37°C in PBS containing 2 mg/mL collagenase, 0.025% trypsin, and 0.02 mg/mL DNase. The digestion mixture was diluted by an equal volume of HEPES buffered M2 medium and oocyte-cumulus cell aggregates dissociated by mechanical disruption in a finely drawn pipette. Oocytes were identified as the larger, refractile cells and washed free of somatic cells by transfer to 20–40 μL PBS containing 20 μg/mL proteinase K and 40 mM SDS, and 20 μL of these lysates treated using the Imprint DNA Modification kit (Sigma). Converted DNA from MII oocytes was prepared previously [Ruf et al. 2007]. PCRs from bisulfite- treated oocyte DNA contained the equivalent of 20–35 oocytes. For growing oocyte DNAs, the results shown were obtained from two independent bisulfite treatments performed on a batch of oocytes collected at the same time; for GV oocytes, the results are from single treatments of 80–110 oocytes. Amplification, COBRA (using MseI or TaqI), and cloning and sequencing of PCR products were performed as previously described [Ruf et al. 2007]. Bisulfite sequences were analyzed using BiQ Analyzer [Bock et al. 2005], with a conversion frequency cut-off of 88% and additional sequences with identical patterns of unconverted cytosines [outside of CpGs] discarded as possible clonal amplification products. Primer information is presented in Supplemental Table 2.

Generation of the Nesptrun targeted allele

The targeting construct was prepared using homologous recombination in yeast, as described previously [Plagge et al. 2005], and inserted a 1.2-kb fragment from the rabbit β-globin gene [kindly provided by Dr. Denise Barlow] [Sleutels et al. 2002]; nucleotides 31,392–32,590 in M18818 downstream from Nesp exon 2 [position 141,627 in AL593857.10] in the correct orientation to truncate Nesp transcripts. The construct also contained a selection cassette with URA3 [for selection in yeast] and neo+ [for selection in embryonic stem [ES] cells] flanked by loxP sites. The targeting construct was linearized with NotI and electroporated into E14 129/Ola ES cells. Colonies surviving G418 selection were screened by Southern blot analysis of Affil digested DNA with a 1.5-kb fragment [nucleotides 134,606–136,087], which resides 5’ to the insertion of the targeting construct, and Xhol digested DNA with a 601-bp fragment [nucleotides 147,539–148,139] at the 3’ end. Chimeric mice were generated by injecting targeted ES cells into C57BL/6J blastocysts; germine transmission of the targeted allele was confirmed by PCR of tail lysate DNA, and transmitting male chimaeras were crossed with homozygous female Cre-deletor mice [Schwenk et al. 1995] for ubiquitous excision of the selection cassette. The amplification product specific for the excised allele was cloned and sequenced to confirm the correct Cre recombination event. Nesptun heterozygotes were maintained on a C57BL/6J background by paternal transmission of the Nesptun allele, owing to the high rate of
lacrimal of Nesprin/pups after maternal transmission. Experimental mice were obtained from maternal transmission of the Nesprin allele crossed with C57BL/6J males or C57BL/6JxD2B males. DS2 is a strain that carries the Gnas region of chromosome 2 from Mus spretus on an F1(C57BL/6)xCBA/Ca background [Plagge et al. 2005]. All analysis of Nesprin/pups was done after the first maternal transmission of the Nesprin allele.

GenBank accession numbers

Nucleotide numbers for the Gnas locus referred to in the text correspond to sequence accession number AL593857.10. Accession numbers for the novel exons and start sites identified in oocyte and embryo RNA by 5′RACE and RT–PCR for the imprinted loci shown in Figure 5 are listed in Supplemental Table 3.

Acknowledgments

We are grateful to Denise Barlow for providing the rabbit β-globin transcription termination cassette. We thank the Gene Targeting Facility at the Babraham Institute for generating the knockout mice, members of the Small Animal Facilities for their conscientious mouse husbandry, Anne Segonds-Pichon for assistance with statistical analysis, and Simon Andrews for bioinformatics support. We are very grateful to Jo Peters, Wolf Reik, and Rebecca Oakley for thoughtful discussions and comments on the manuscript. This work was supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom and the European Union (STREP Savebeta contract 036903 in the framework programme 6 [FP6] of the European Community). N.R. was sponsored by a post-doctoral fellowship from the German Academic Exchange Service (DAAD) and a Marie-Curie Intra European Fellowship from the European Community (FP7/2007-2013) under grant agreement number PIEF-GA-2008-220912. K.J. was supported by a Genes & Development Summer Studentship from the Genetics Society.

References


