Genomic imprinting is an epigenetic process resulting in the monoallelic parent-of-origin-specific expression of a subset of genes in the mammalian genome. The parental alleles are differentially marked by DNA methylation during gametogenesis when the genomes are in separate compartments. How methylation machinery recognizes and differentially modifies these imprinted regions in germ cells remains a key question in the field. While studies have focused on determining a sequence signature that alone could distinguish imprinted regions from the rest of the genome, recent reports do not support such a hypothesis. Rather, it is becoming clear that features such as transcription, histone modifications and higher order chromatin are employed either individually or in combination to set up parental imprints.

Address
Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, 415 Curie Boulevard, Philadelphia, PA 19104, USA

Corresponding author: Bartolomei, Marisa S (bartolom@mail.med.upenn.edu)

Introduction
Genomic imprinting affects a small number of genes in the mammalian genome and results in parent-of-origin-specific monoallelic expression [1–3]. Imprinted genes, which are typically conserved among mammals, play essential roles in the growth and development of the fetus, as well as in post-natal behavior and metabolism. Further, while many imprinted genes are ubiquitously imprinted, some exhibit tissue-specific or temporal-specific imprinting patterns. The best-defined class of genes that display restricted imprinting are those that are imprinted exclusively in the placenta, including Ash2, Phlda2, Slc22a2, and Slc22a3 [4]. Notably, imprinted genes are located in approximately 1Mb clusters throughout the genome, although singletons have been described (http://www.har.mrc.ac.uk/research/genomic_imprinting/). These clusters typically contain genes that are expressed exclusively from the maternally or paternally inherited chromosomes. Additionally, each of these clusters is under the control of a discrete region, termed imprinting control region (ICR, also designated imprinting center or imprinting control element).

One critical attribute of imprinted genes is that they must be marked with their parental origin so that the correct allele-specific expression patterns are observed in somatic tissues. The parental-specific mark must be stable and heritable so that imprinting is maintained throughout development. They must also be erasable so that imprints can be reset from a biparental somatic pattern (germ cells are derived from the soma in mammals) to the germline-specific pattern that reflects the sex of the individual. Moreover, the most logical time for alleles to be marked is in the germline when they are in separate compartments and can be differentially modified as epigenetic reprogramming takes place (Figure 1). DNA methylation is the epigenetic modification that appears to be most integral to the marking of parental alleles, although post-translational histone modifications are clearly involved in imprinted gene expression (see below). Allele-specific DNA methylation has been described at all of the imprinted gene clusters and many other single imprinted genes. These differentially methylated regions (DMRs) are established in the germline by de novo DNA methyltransferases (DNMTs, see below) and are either maintained throughout development (primary DMRs) or arise later in development (secondary DMRs), often as a consequence of imprinted expression. Furthermore germ-line-specific DMRs often expand or contract during development [5*], although the reason for this is unclear. Many primary DMRs are also ICRs as gene targeting experiments have shown that deletion of the region corresponding to the DMR causes loss of imprinting of multiple genes in cis [1]. Additionally, all identified ICRs to date are marked by differential methylation. Curiously, many more maternally methylated than paternally methylated DMRs/ICRs have been identified. Subsequently, all DMRs referred to here are primary DMRs unless otherwise noted.

Once imprints are set in the germline, they must be properly maintained following fertilization when the genome undergoes a period of considerable reprogramming (Figure 1). This reprogramming involves the resetting of DNA methylation patterns, which is now known to involve conversion of methylcytosine to hydroxymethylcytosine [6*,7*], probably followed by replication dependent passive demethylation [8*]. Importantly, DNA methylation imprints must not only survive the
Figure 1

Timeline of epigenetic reprogramming. The first wave of genome-wide DNA demethylation takes place shortly after fertilization, with the maternal genome passively demethylated and the methylcytosine of the paternal genome converted into hydroxymethylcytosine, which is passively eliminated. Imprinted DMRs (ICRs) are maintained despite this demethylation event. De novo DNA methylation occurs around implantation. In both male and female primordial germ cells another wave of DNA demethylation initiates as the cells migrate toward the genital ridge. All DMRs are also erased at this time. In male germ cells methylation imprints are acquired in prospermatogonia around E15.5–17.5. In the female germine methylation imprints are not acquired until after birth, in growing oocytes. Transcripts have been correlated with acquisition of DNA methylation in both the male and female germlines. Activities in male and female germ cells are represented in blue and red, respectively.

Two dominating mechanisms have been described for mediating imprinting in clusters [1–3]. Thus far, the most evolutionarily ancient [12], but seemingly least utilized mechanism is the insulator model of imprinting, which is employed by the H19/Igf2 imprinted locus. The maternally expressed H19 gene and paternally expressed Igf2 gene share enhancers and their reciprocal imprinting is governed by a CTCF-dependent insulator that is located between the genes. On the maternal allele in mouse, CTCF binds to 4 binding sites within the ICR, generating an insulator that prevents Igf2 from accessing the shared enhancers that are located on the H19 side of the insulator. Thus, the insulator effectively acts as an enhancer-blocker. On the paternal chromosome, methylation at the ICR not only prevents CTCF from binding, allowing Igf2 to engage the enhancers, but it is also required for methylation at the H19 promoter and silencing of H19 (Figure 2a). A more commonly employed and recently evolved mechanism of imprinting uses a long non-coding RNA (ncRNA). In this case, the ICR includes a differentially methylated promoter that regulates the expression of an ncRNA; when unmethylated the ncRNA is expressed and represses cis-linked genes. By contrast, when the ICR is methylated, the ncRNA is repressed and the cis-linked genes are expressed. The maternally expressed Igf2r gene was first shown to use such a mechanism (Figure 2b). At this locus, the ncRNA Airn represses Igf2r ubiquitously and Slc22a2 and Slc22a3 in the placenta. Just how this repression is mediated is unclear but it has been proposed that Airn interacts with the Slc22a3 promoter and the H3K9 histone methyltransferase G9a in placenta, thereby epigenetically silencing transcription [13]. Alternatively, transcription through the domain has been suggested to silence genes in cis [2]. It is possible that both mechanisms are used, but in a tissue-specific manner.

Here, we will focus on current developments in the field of imprinting. Although the mechanisms underlying genomic imprinting are slowly being elucidated, some of the most significant developments have centered on characterizing the earliest steps in the imprinting process, the recognition and marking of imprinted regions in germ cells.

Methylation machinery and recognition of DMRs in germ cells

Experiments by Jaenisch and colleagues, which deleted the maintenance methyltransferase Dnmt1, first supported a central role for DNA methylation for imprinted gene expression [14]. Studies of conditional knockouts of the de novo methyltransferases Dnmt3a and Dnmt3b in germ cells provided evidence that DNMT3a is required...
for *de novo* methylation of all DMRs (both maternal and paternal) except for the paternally methylated *Rasgrf1* ICR, which requires both DNMT3a and DNMT3b [15,16]. A defect in the establishment of methylation of germline DMRs was also observed in mice deficient for DNMT3-like (DNMT3L) [17,18]. DNMT3L, which lacks methyltransferase activity, has the ability to interact with both DNMT3a and DNMT3b. Oocytes deficient for DNMT3L lacked methylation at maternal DMRs. Additionally, *Dnmt3L* null females exhibited a maternal lethal phenotype, with embryos dying around E9.5. These embryos were hypomethylated at all maternally methylated DMRs but global methylation did not appear reduced [17]. DNMT3L also plays a critical role in establishment of paternal methylation imprints. Though reports vary on the extent that paternal DMRs were affected in DNMT3L deficient male germ cells, ICRs at *H19*, *Rasgrf1* and *Gtl2* (Ig-DMR) require DNMT3L for full methylation [15,16,18,19].

Though the DNA methylation machinery is now well-established, it remains to be determined how this machinery is recruited to specific CpGs. Recently, a link has been found between histone methylation and DNA methylation. DNMT3L has binding affinity for nucleosomes containing unmethylated H3K4, which is abolished with the addition of methyl groups to this residue [20]. These results suggest that patterns of histone methylation could dictate patterns of DNA methylation, which could then be stably inherited. Furthermore, oocytes that lack lysine demethylase 1B (KDM1B), which functions as an H3K4 demethylase, were shown to be hypomethylated at a number of imprinted loci [21], suggesting that an unmethylated H3K4 is necessary for DNA methylation. Consistently, male germ cells assayed at E15.5 (the onset of *de novo* methylation) revealed high levels of H3K4me3 at the maternally methylated *Snrpn* ICR and KvDMR but absence of H3K4me3 at paternally methylated *H19* ICR and Ig-DMR, suggesting that H3K4me3 prevents maternally methylated DMRs from acquiring
DNA methylation in the male germline [22*]. Interestingly, the association of DNA methylation with an unmethylated H3K4 and the protection from DNA methylation with methylation of H3K4 is not limited to imprinted loci, but has been observed genome-wide [23,24].

Even with the observation that DNA methylation requires a favorable histone environment, it is still unclear what (if any) sequence signature distinguishes DMRs from other CpG rich regions in the genome. One such signature has been proposed to be the spacing of CpGs at DMRs. Molecular modeling of the DNMT3a/DNMT3L complex indicated an optimal periodicity of CpGs for methylation at about 8–10 base pairs [25]. Interestingly, Jia et al. [25] reported this periodicity in 12 maternal DMRs. However, recent studies were unable to detect this trend at DMRs in both oocytes and sperm [5*] or were unable to detect a difference in CpG spacing between methylated and unmethylated CpG islands (all CpG rich regions in the genome, of which DMRs are only a small subset) in oocytes and sperm [26*]. Overall, the role of CpG spacing as a signature for germline DMRs remains unclear.

Another proposed sequence feature that could distinguish DMRs from the rest of the genome is the composition of repetitive elements at imprinted loci [27]. Sequence analysis reveals a depletion of SINE elements [28,29] but an enrichment of tandem repeats [30] at imprinted regions. Thus, it has been hypothesized that the presence of certain repetitive elements or the absence of others could be the key distinguishing trait recognized by methylation machinery [27]. Recently, it has been confirmed that DMRs have a higher overall concentration of tandem repeats than CpG islands [5*,30]. However, this comparative analysis could be skewed by the Rasgrf1 ICR, which is exceptionally repeat rich. Moreover, there are still a substantial number of CpG islands that contain more tandem repeats than many DMRs [5*]. Genome-wide analysis of CpG islands found no correlation between frequency of tandem repeats and methylation status [26*]. Thus, although DMRs are located in regions within or adjacent to tandem repeats, it does not seem that this is their key distinguishing feature, or necessarily a feature that would attract methylation machinery. Rather than being a single sequence trait of all DMRs, each DMR may contain a locus-specific signature that differentiates it from the rest of the genome.

**Active transcription as a signal for de novo methylation**

Many reports have recently uncovered a link between transcription and methylation, which at first glance is counter to expectations given the general observation that DNA methylation is associated with gene repression. For example, as described above, KDM1B is critical for establishment of methylation of many maternal DMRs [21*]. Interestingly, the human ortholog of KDM1B, LSD2, is associated with gene bodies of actively transcribed genes [31]. Additionally, the PWWP domain of DNMT3a binds H3K36me3, a mark of transcriptional elongation, which increases the methyltransferase activity of DNMT3a in vitro [32]. These observations suggest that de novo methylation is targeted to sites of active transcription. Analysis of CpG islands in day 10 oocytes, the time at which de novo methylation is first apparent, has provided further evidence for the requirement of transcription for methylation. Intragenic methylated CpG islands are more likely to be within active transcription units than unmethylated intragenic CpG islands. Moreover, methylated CpG islands within annotated promoters more frequently overlap transcripts than unmethylated CpG islands within promoters [26*]. One possible explanation for the link between active transcription and methylation is that transcription may help maintain open chromatin, which in turn allows methylation machinery to access DNA.

Multiple imprinted loci have been shown to require transcription for establishment of DNA methylation; the first of these being the Gnas locus [33*]. This locus encodes numerous transcripts; the protein coding transcripts Gnas, Gnas1, and Nesp and the non-coding transcripts Nespas and IA. This region contains two maternally methylated DMRs, one that encompasses the Gnas1 and Nesp promoters, which acts as the ICR, and another that covers the IA promoter. Chotalia et al. made a targeted mutation in the mouse that truncated the Nesp transcript, the furthest upstream transcript, and observed hypomethylation (in varying degrees) at all DMRs at the Gnas locus in newborn pups [33*]. Furthermore, analysis of mutant oocytes revealed a loss of methylation at the DMRs, suggesting a defect in methylation establishment. Interestingly, transcripts were also detected in growing oocytes at the maternally methylated DMRs of the Grib10, Igf2r, Impact, Kingl, Zac1, and Surpr imprinted loci [33*,34] (Figure 1).

The imprinted Rasgrf1 locus, which harbors a paternally methylated ICR, also requires transcription for methylation establishment [35*]. In spermatogonia of mice mutant for various proteins in the piRNA pathway, including MILI, MIWI2, and MITOPLD, the Rasgrf1 ICR, but not other paternally methylated DMRs, exhibited reduced methylation. Further analysis revealed that a non-coding RNA (named pit-RNA) transcribed from the Rasgrf1 ICR was targeted by piRNAs, causing cleavage of this RNA. The authors propose a model in which targeting of piRNAs to pit-RNA is an important step in sequence specific methylation at the Rasgrf1 locus [35*].

Though paternal-specific methylation of the *H19* ICR and Ig-DMR does not require the piRNA pathway [35*], transcription has been detected at both of these ICRs
In contrast to the absence of transcription at these ICRs in somatic cells or E13.5 male germ cells (before de novo methylation), high levels of RNA can be detected starting at E15.5 (the onset of de novo methylation), through E17.5 in male germ cells (Figure 1). Transcription was detected throughout both ICRs and in the region between H19 and its ICR, as well as downstream of the H19 promoter, and between Gld2 and its ICR [22**]. Transgenic experiments that incorporate H19 sequences at an ectopic locus are consistent with the idea that sequences outside of the ICR are required for DNA methylation establishment in the germline. For example, when the H19 ICR is targeted to an ectopic locus, it fails to establish methylation in sperm [36,37]. By contrast, when the H19 ICR plus upstream sequences (ranging from −4.4 kb to +2.9 kb from the H19 transcriptional start site) is inserted, a subset of spermatooza were fully methylated at this exogenous site [37]. Further, a BAC transgene, including sequences from −7 kb to +140 kb to the H19 transcriptional start site, properly established methylation at the H19 ICR [37]. Consistently, numerous cis-acting ICR mutations that disrupt imprinted gene expression or maintenance of methylation during embryogenesis, still properly establish methylation in sperm [38–40]. Thus, it is possible that these additional sequences contain transcripts necessary for de novo methylation, though this hypothesis is untested. Together, these studies provide evidence for the requirement of transcription at both maternally methylated and paternally methylated ICRs.

**Locus-specific features at the H19 ICR are necessary for proper epigenetic landscape and genomic organization during spermatogenesis**

Accumulating evidence argues against a single signature for all DMRs, but rather, there are likely to be locus-specific features. As described above, the H19/Igf2 locus utilizes a CTCF-dependent insulator to repress Igf2 maternally and methylation of the ICR to repress H19 paternally [41–44] (Figure 2a). Interestingly, asymmetric acquisition of DNA methylation in the germline has been observed at the H19 ICR, with the paternal allele being methylated first [22**,45,46], indicating that the two alleles retain somatic cell memory. Recent studies in male germ cells suggest that CTCF binding may coordinate this somatic memory mark, as functional CTCF binding sites are required for allele-specific histone modifications and inter-chromosomal interactions. Lee et al. observed a slight enrichment of H3K4me2 at the maternal H19 ICR and reciprocally, a slight enrichment of H3K9me3 at the paternal H19 ICR in E13.5–14.5 spermatogonia [46]. However, when the maternal H19 ICR was mutated in such a way that CTCF could not bind, the biased enrichment was no longer detected. Furthermore, delayed de novo methylation at the H19 ICR of the maternal allele was no longer detectable when the maternal allele was mutated at the CTCF sites [46*], suggesting that maternal H3K4me2 (which requires CTCF binding) protects this allele from DNA methylation and needs to be demethylated before de novo DNA methylation can take place. These data indicate that intact CTCF sites coordinate allele specific histone modifications that facilitate marking of the parental origin of each allele.

Inter-chromosomal interactions also appear to be dependent on CTCF binding sites at the H19 ICR in male germ cells. Physical interactions of the H19/Igf2 region with other imprinted regions such as Cog2, Htr2a and Dlk1 have been observed in spermatogonia. However, these interactions were not detectable when a mutated H19 ICR that does not bind CTCF was maternally inherited [47]. Recent reports have indicated that CTCF interacts with cofactors such as cohesins [48–52], the DEAD-box RNA-binding protein p68 (p68), and the steroid receptor RNA activator (SRA) [53] at the H19 ICR. Though analysis of cofactor interaction with CTCF has not been performed in germ cells, it is possible that these various cofactors facilitate higher order chromatin at the H19 locus. The loss of CTCF binding suggests that these factors may no longer mediate inter-chromosomal interactions or allele specific histone modifications. Nevertheless, it still remains to be determined if and how inter-chromosomal interactions are involved in germline reprogramming.

**Conclusions**

Significant questions remain in the study of how imprinted loci are marked with their parental origin, escape post-fertilization reprogramming and are appropriately regulated in the soma. Accumulating evidence suggests that locus-specific sequences, transcription, histone modifications and higher order chromatin structure are employed either individually or in combination for these processes. Here we focused on the process of marking the parental origin of imprinted genes. Interestingly, there is a correlation between transcription at DMRs and methylation establishment. Whereas the Gnas and Rasgrf1 loci clearly require transcription for DNA methylation establishment, cause and effect is less clear at other imprinted loci. The presence of RNA is not sufficient to prove that DNA methylation machinery requires active transcription. For example, Henckel et al. detected transcriptional activity at a number of maternally methylated DMRs in male germ cells [22**], which is contrary to the idea that transcripts are necessary for methylation (as these regions remain unmethylated in male germ cells). Experiments that truncate or delete individual transcripts will be required to determine the role of these transcripts in methylation establishment.

In the future it will be important to interrogate chromatin structure and dynamics during methylation establishment.
in germ cells. Currently, such experiments are not feasible owing to the large number of cells required and the difficulty in obtaining sufficient numbers of germ cells. Nevertheless, as technology advances and becomes more sensitive, elucidation of the mechanism and dynamics of reprogramming in the germline will become possible.

Acknowledgements

We apologize to those whose work could not be mentioned owing to space limitations. We thank members of the laboratory for critical reading of the manuscript. Work in the Bartolomei lab is supported by the NIH (GM051279, HD042026, and HD068157).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


7. Wossidlo M, Nakamura T, Lepikhov K, Marques CJ.


This work demonstrated that the paternal genome in zygotes is not actively demethylated, but rather, methylcytosine is converted to hydroxymethylcytosine.


This work demonstrated that the loss of hydroxymethylcytosine of the paternal genome during early embryogenesis is replication dependent.


This study demonstrated the requirement for KDM1B, a histone H3K4 demethylase, for de novo methylation at a number of imprinted DMRs in oocytes.


This work investigated histone methylation and DNA methylation patterns in primordial germ cells. Additionally, transcripts were detected at paternal DMRs at the time of methylation establishment.


This study investigated over 1000 methylated CpG islands in oocytes to determine features that dictate methylation status and found correlation between transcription and methylation.


33. Chotailia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, • Frontera M, James K, Dean W, Kelsey G: Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev 2009, 23:105-117. First report showing the requirement of transcription for de novo methylation at an imprinted DMR. This work demonstrated the requirement of the Nespr transcript for establishment of methylation at DMRs in the Gnas locus. Additionally, transcripts were detected at numerous other imprinted DMRs in oocytes.


35. Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, • Kuramochi-Miyagawa S, Iida N, Hoki Y, Murphy PJ, Toyoda A et al.: Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science 2011, 332:848-852. This study reports that piRNAs are required for establishment of methylation at the Rasgrf1 ICR. The authors propose that a non-coding RNA at the Rasgrf1 locus is targeted by these piRNAs and that cleavage of this RNA is necessary for de novo methylation.


46. Lee DH, Singh P, Tsai SY, Oates N, Spalla A, Spalla C, Brown L, • Rivas G, Larson G, Rauch TA et al.: CTCF-dependent chromatin bias constitutes transient epigenetic memory of the mother at the H19-Igf2 imprinting control region in prospermatogonia. PLoS Genet 2010, 6:e1001224. Enrichment of H3K4me2 at the maternal H19 ICR and H3K3me3 at the paternal H19 ICR in prospermatogonia was reported. The timing correlates with asymmetric acquisition of DNA methylation on the two parental alleles (the paternal allele is methylated first), all of which required intact CTCF sites on the maternal allele.


