Imprint stability and plasticity during development

Sarah-Jayne Mackin, Avinash Thakur*, and Colum P Walsh

Genomic Medicine Research Group, School of Biomedical Sciences, Ulster University, Northern Ireland, UK

Correspondence should be addressed to C P Walsh; Email: cp.walsh@ulster.ac.uk

*(A Thakur is currently at Terry Fox Laboratory, BC Cancer Research Centre, 675 West 10th Avenue, Room 13-112, Vancouver, BC V5Z 1L3, Canada)

Abstract

There have been a number of recent insights in the area of genomic imprinting, the phenomenon whereby one of two autosomal alleles is selected for expression based on the parent of origin. This is due in part to a proliferation of new techniques for interrogating the genome that are leading researchers working on organisms other than mouse and human, where imprinting has been most studied, to become interested in looking for potential imprinting effects. Here, we recap what is known about the importance of imprints for growth and body size, as well as the main types of locus control. Interestingly, work from a number of labs has now shown that maintenance of the imprint post implantation appears to be a more crucial step than previously appreciated. We ask whether imprints can be reprogrammed somatically, how many loci there are and how conserved imprinted regions are in other species. Finally, we survey some of the methods available for examining DNA methylation genome-wide and look to the future of this burgeoning field.

Introduction

Genomic imprinting is a classic epigenetic phenomenon, and after X-inactivation, one of the best understood. In its simplest form, imprinting refers to genes showing mono-allelic expression depending on their parent of origin (Bartolomei & Ferguson-Smith 2011, Barlow & Bartolomei 2014). There is however considerable variation seen in where in the body and when the imprinted expression is seen. Imprints need to be set up when the alleles are separated during germ line development and then the imprints need to be interpreted by the transcriptional machinery in the target tissues. Recent work has advanced our understanding of (1) the processes behind establishment of imprints as well as (2) how the imprinting mark is maintained and interpreted. For the former, there is an excellent review by Kelsey (Hanna & Kelsey 2014). Here, we wish to focus on the latter process of maintenance, as well as ask what we have learned from species other than mouse, and touch on the technical difficulties, which arise when working in organisms other than mouse and human, where the most resources are available.

Importance of known imprints for body mass and feeding

Effects on growth and body size

One of the most enduring explanations for why imprinting may have arisen is the parental conflict hypothesis (Haig & Graham 1991, Moore & Haig 1991). Many imprinting phenomena involve effects on growth, nutrition or metabolic balance, and this theory was partly based on observations from uniparental conceptuses where the paternal contribution seemed particularly important for the embryo, while the maternal genome was important for placental development (Barton et al. 1984). This turned out to be partly due to the paternal expression of the major growth factor insulin-like growth factor-2 (IGF2), and the maternal expression of an antagonistic binding partner insulin-like growth factor-2 receptor (IGF2R), one of the first observations leading to the idea of parental conflict in imprinting. Several imprinted genes were also shown to be dysregulated in the placentas of these mouse conceptuses (Walsh et al. 1994), including genes specifically imprinted only in placenta such as Mash2 (Guillemot et al. 1995). Indeed, the placenta is generally considered to have a higher number of imprinted genes than in other tissues (Court et al. 2014); though in the mouse and human, many of these are maternally expressed. An alternative to the parental conflict hypothesis, called coadaptation theory, also stresses the importance of interactions between offspring and parents for nutrient provision and acquisition (Keverne & Curley 2008).

Much of the support for these theories has been gleaned in mouse systems. Interestingly however, a single nucleotide substitution in a non-coding region of IGF2 underlies a major QTL in pigs that affects muscle growth, heart size and fat deposition (Jeon et al. 1999,
Neurological effects involving feeding and weight gain

Another general feature of imprinted genes is that many are transcribed primarily in the neurons and in some cases only show imprinted expression there. While initially this posed difficulties for the parental conflict hypothesis, which was originally proposed on the basis of imprinting affecting placentation and embryonic growth, with the discovery that some of the genes controlled feeding behaviour or weight gain later in life, these results could be reconciled with the hypothesis. Thus, Prader–Willi syndrome in humans (OMIM176270), where transcription of the neuronal genes at the SNRPN cluster are affected, involves excessive eating (hyperphagia), while the Peg3 gene in mouse is crucial for feeding behaviour (Curley et al. 2005). Likewise, loss of expression of the Gnas and Gnasxl genes in mice can give hypo- and hyperphagia respectively, while a number of other imprinted genes also have effects on fat/body mass ratios (Peters 2014).

Furthermore, the aforementioned LOS or large calf syndrome can often cause a reluctance to suckle in young cattle. This may be due to the dysregulation of a number of imprinted genes reported in these animals (Zhang et al. 2004, Chen et al. 2015). However apart from these few studies, little work appears to have been done to date investigating imprinted genes and feeding behaviours outside of the mouse and human systems.

Imprinting disorders in hybrid species: the case of the hinny

A classic example of parent-of-origin effects in livestock is the horse–donkey cross: when the horse is the sire, the resultant offspring is a mule; if the donkey is the sire, then a hinny results, and these are quite distinguishable animals even though their genomic complements are the same. Initially, it was argued that this may be due to differences in uterine environment and the placental–uterine exchange. Early investigation identified that serum concentrations of the placental hormone equine chorionic gonadotrophin were considerably higher in mule than in hinny pregnancies, suggestive of paternal genome-specific expression (Allen 1969). Recent use of RNA sequencing of trophoblast tissue from embryos derived from each reciprocal cross of horse and donkey (Wang et al. 2013) identified 15 ‘core’ imprinted genes that are conserved in equids, mouse and human, including genes such as SNRPN, H19 and PEG3.

Interestingly, paternal expression seems favoured, as 10/15 core imprints – and all of the larger group of 78 genes identified in that study – showed paternal bias, but with many only imprinted in placenta. Additionally, some of the genes in the second, larger group showed incomplete or variable silencing. The authors speculate that this variability may reflect the flexibility of the structure and function of the placenta, which varies widely between mammals. Interestingly, XIST does not seem to be one of the imprinted genes, and X-inactivation in the placenta is random in both the horse parent and the hybrids (Wang et al. 2012).

Establishment of parental-specific methylation at imprints

It was recognised already from early studies that imprints would most likely be established in the germ line, where each allele of the diploid pair is separate and therefore can be separately marked. While studies on germ cell development were long hampered by the
difficulty of access to the early stages of development of this migratory cell population, and the small numbers of cells involved, recent advances in cell labelling and separation, coupled with the refinement of sequencing techniques and single-cell approaches (see below) have meant that we now have a much deeper understanding of the process of imprint establishment. This area has been dealt with very expertly by others in a recent review (Hanna & Kelsey 2014). We will concentrate here on insights gained in some studies published since then, confining ourselves first to a brief recap of the varied imprinting mechanisms.

The main types of imprinted loci and their structures
Current studies suggest that there are three main types of imprinted locus, controlled by different mechanisms. This has implications for where the imprint controllers are located and what they look like, as well as how the locus might respond to different perturbations.

Insulator
The archetype here is the cluster of genes around the H19 locus (Bartolomei & Ferguson-Smith 2011): these are controlled by an intergenically located ICR that is not a CpG island (CGI) (Bird 1987) but is relatively CG rich, making it sensitive to DNA methylation. The ICR contains several repeats of a sequence that binds the insulator protein CTCF. Methylation blocks CTCF binding on the paternal allele at the locus (Bell & Felsenfeld 2000, Hark et al. 2000), preventing H19 transcription but facilitating expression of Igf2 and, in mouse, the Insulin 2 gene expressed in the yolk sac. The mark on the ICR that is set up in the germ cells is called a primary differentially methylated region (1 DMR) or gametic DMR (gDMR). Methylation that occurs post implantation, for example, on the H19 promoter after it has already been silenced, is known as a secondary (2) DMR or somatic DMR (sDMR). There are a number of other imprinted loci with CTCF-binding sites, in some cases lacking CGI, where this protein is thought to play an important role (Prickett et al. 2013).

Long non-coding RNA promoter
Here the index locus is Igf2r that, unlike H19, is an orphan imprinted gene. The DMR at the Igf2r promoter was, against expectation, the secondary DMR, and the primary DMR and functional ICR were found to be located at an intragenic island. This turned out to control transcription of an antisense lncRNA, whose transcription blocked the sense transcript in cis (Sleutels et al. 2002). This type of arrangement is also seen at the Kcnq1, Grb10 and Dlk1 loci among others (Barlow & Bartolomei 2014) and as more lncRNA are uncovered, this is becoming the largest group.

Promoter
For some imprinted loci such as Snrpn, Plagl1 and Grb10, transcription appears to be directly regulated by methylation of the promoter. Here, a promoter CGI is methylated to block transcription rather than any indirect mechanism. Although there is no antisense transcript, the Snrpn promoter does drive the production of a long RNA called Snurf/Snrpn that extends far downstream and contains multiple small RNA species that are processed from it (Buiting 2010), so it could be argued that it falls into the lncRNA category. However, for other genes in this group such as Grb10, no associated lncRNA have so far been found.

Recent insights into imprint establishment
Seminal work from the Bestor lab (Ooi et al. 2007) showed that the methyltransferase cofactor DNMT3L, which is only required in germline, bound to CGI but was blocked by methylation of the fourth lysine on the histone 3 tail (H3K4me3), and further that imprinting was disrupted in the knockout (KO) mice. This link seemed strengthened by the finding that KDM1B, a histone demethylase, was required to establish DNA methylation at imprints (Ciccone et al. 2009). However, the idea that DNMT3L might be specific to imprints was weakened by studies charting the establishment of methylation in oocytes, which showed that not only did many thousands of non-imprinted CGI become hypermethylated during oocyte maturation in a DNMT3L-dependent fashion, but DNMT3L was also required for methylation of many other non-CGI sequences (Kato et al. 2007, Smallwood et al. 2011, Kobayashi et al. 2012), suggesting that DNMT3L was in fact more of a general cofactor for the de novo methyltransferases, required primarily in the germline. What appeared instead to distinguish imprinted loci from other genes was that they could maintain their DMR in the face of the two main waves of remodelling in the early embryo, namely active and passive demethylation in the first few cell divisions and then the widespread de novo methylation seen post implantation. While few imprints are established in the male germ line, indications are that here too it is the maintenance of these DMR in the face of remodelling that separates them from other loci (Wang et al. 2014).

In effect, this marks a real shift in thinking with regards to imprinting, moving the emphasis away from the mechanisms that might target regions to become imprinted and towards identifying factors that are responsible for maintaining methylation differences at specific loci in the face of extensive epigenomic remodelling. Since this is a crucial new insight, we wish to expand on these findings below and begin by looking in more detail at the protein factors that may be involved prior to implantation, and then look at events post implantation.
Maintenance of imprints pre implantation

Methylation marks are applied in a sex-specific manner during gametogenesis at the gDMRs and several proteins play important roles at various stages in this process. In particular, Pgc7 (also referred to as Stella) is a critical component required in early maternal development and encodes a protein that has a SAP-like domain and a splicing factor-like domain (Aravind & Koonin 2000). It has an important role in protecting the methylation status of imprinted genes by limiting demethylation in early embryogenesis (Nakamura et al. 2007) and as such is crucial for normal development. PGC7 has been experimentally shown to be actively expressed in primordial germ cells (PGCs) in mouse from E7.25 to E15.5 (Sato et al. 2002) but more importantly in the immature ovaries of the neonate females (Payer et al. 2003). Through depletion of the protein in the oocytes, Payer et al. (2003) found there was a reduction in the number of blastocysts, successful implantations and a reduction in the number of viable offspring. Nakamura et al. (2007) initially showed that PGC7 was required to maintain methylation on most imprinted loci in the preimplantation embryo, and later went on to show that PGC7 bound to H3K9me2 at both paternally and maternally marked loci protects against active demethylation by TET3, a finding confirmed by others (Nakamura et al. 2012, Szabó & Pfeifer 2012).

Another factor necessary for DNA methylation maintenance at ICRs in both mouse and human is ZFP57 (Kruppel-like zinc finger protein). Removal of the protein from the mouse zygote by use of both maternal and zygotic deletions caused embryonic death mid-gestation with a loss of all maternally methylated imprints (Li et al. 2008). In humans too, mutations in ZFP57 causes hypomethylation at multiple imprinted loci (Mackay et al. 2008). ZFP57 binds to a target hexanucleotide found at imprinted gDMRs, but only when methylated, and recruits DNMT1 and its cofactor UHRF1 to maintain methylation on the marked alleles (Quenneville et al. 2011). Additionally, through KAP1, it recruits SETDB1 that modifies the chromatin by adding H3K9me3. Indeed, loss of KAP1 also causes a failure to maintain imprints in the preimplantation embryo (Messerschmidt et al. 2012). Following this protective step in early development, DNA methylation is thought to be maintained along with this chromatin mark throughout the offspring’s life at the respective ICRs (Proudhon et al. 2012).

Post-implantation changes at transient and stable imprints

The small number of known imprinted loci characteristically show (1) methylation differences between gametes (gDMRs), (2) dependence on DNMT3L (Kato et al. 2007) and (3) an ability to maintain differential methylation during preimplantation development (Hanna & Kelsey 2014). However, the genome-wide studies of the methylation landscape in germ cells mentioned earlier found that a relatively large number of loci (>1000) were methylated in a DNMT3L-dependent fashion in oocytes, but not sperm, and that many of them maintained high levels of methylation on the maternal allele until at least the blastocyst stage (Borgel et al. 2010, Smallwood et al. 2011, Shirane et al. 2013). Initial work by Michael Weber and colleagues (Borgel et al. 2010), suggested that not only imprinted loci, but also genes that were specifically expressed in the germline showed these features. Work from our own lab confirmed this (Rutledge et al. 2014) and further uncovered a class of genes expressed in brain showing these characteristics, suggesting three classes of genes with DNMT3L-programmed gDMRs: imprinted, germline and brain-specific genes (Fig. 1A). However, in stark contrast to the canonical imprinted loci, most of the brain and germline genes examined lost their differential methylation in adult tissues tested due to gains in methylation on the paternal, unmethylated allele. However, it was possible that among these latter two classes there were some true imprinted genes, which would show allele-specific expression in a tissue- or stage-specific fashion.

To identify such genes, Ferguson-Smith and colleagues used ZFP57 binding as an extra criterion in addition to the three indicated above (Strogantsev et al. 2015). Using this approach, they could confirm recently identified imprints such as Cdh15 and Gpr1, and also uncovered evidence for imprinting at the Fkbp6 gene. Fkbp6 showed methylation in oocytes but not sperm in their study (Strogantsev et al. 2015), and maternal methylation was maintained in placenta at e16.5. In keeping with this, there was also predominant expression from the paternal allele as assayed using a transcribed SNP in interspecific mouse crosses, and this allele also carried H3K4me3 marks. Notably, in brain, both transcription and H3K4me3 continued to be associated with the paternal allele, even though the DNA methylation was assayed at 80%, rather than the expected 50%. This suggests that (1) either a subset of cells in brain continues to maintain the DNA methylation difference (in keeping with the low levels of transcription in this tissue); (2) some of the methylation being assayed may in fact be 5-hydroxymethylation (5hmC- see below), which is more prominent in brain and may cloud the picture, or most intriguingly (3) the DNA methylation difference is not as important as the histone marks. In addition to this complexity, Fkbp6 may show differences in imprinting due to genetic variation, since sperm methylation varies between mouse strains (Rutledge et al. 2014), and there have been reports of polymorphic imprinting in humans (Hanna et al. 2016). ZFP57 is also capable of mediating such strain-specific effects (Strogantsev et al. 2015).
Using one of the other three criteria above, namely dependence on DNMT3L, the Bourc’his lab also identified novel imprinted gDMRs including Cdh15, AK008011, Zfp777 and Zfp787 (Proudhon et al. 2012). The Cdh15 gDMR was found to control transcription of an allele-specific RNA in the hypothalamus, but there was no evidence for imprinted expression at the other three loci. All four gDMRs again lost their parental mark through gains in methylation in most adult tissues, highlighting the crucial role of protection from de novo methylation for imprint stability. They proposed that such genes, which would include Fkbp6, Cdh15, Gpr1 and others, should be referred to as transient imprints. The subtlety of regulation of these loci led to the question of how functionally important this was. In a subsequent paper, they showed however that for the Gpr1 locus, DNA methylation did indeed play an important role in fine-tuning transcription levels (Duffie et al. 2014). At this locus, interestingly, the histone marks seem to play a secondary role.

In contrast to these transient imprints, the canonical imprinted loci maintain methylation differences at the gDMR even in tissues that do not express the associated gene (Woodfine et al. 2011, Court et al. 2014, Wang et al. 2014), highlighting the existence of mechanisms to (1) prevent demethylation of the methylated (usually maternal) allele and (2) protect the unmethylated (usually paternal) allele from gaining methylation. Studies comparing the sizes of DMR concluded that a certain degree of shrinkage occurs from the gametic state as the embryo matures (Court et al. 2014) but that the boundaries do not markedly shift upstream or downstream, consistent with a mechanism centred on the sequence determinants in the gDMRs.

Possible role of other cytosine modifications at imprinted loci

The dynamic changes in DNA methylation seen during early development are now known to be driven both by passive dilution through replication in the absence of maintenance activity, as well as active demethylation via the action of the three ten-eleven translocation (Tet) enzymes (Guo et al. 2014, Wang et al. 2014). The Tets have been shown to catalyse the oxidation of normal 5-methylcytosine (5mC), first to 5-hydroxymethylcytosine (5hmC), then to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC), at which stage the modified nucleotide is excised from DNA by the base excision repair enzyme thymidine DNA glycosylase (TDG) or possibly removed through an as-yet uncharacterised
decarboxylase activity (reviewed in Hahn et al. 2014). Active demethylation explains the rapid loss of 5mC in the pronuclei of the fertilised egg in the absence of replication, particularly notable for the male pronucleus, and lies behind much of the swift demethylation seen at some other time points such as germ cell specification and preimplantation development (reviewed in Hill et al. 2014). Blocking the action of the Tets thus prevents the demethylation at ICR in the germ cells required to reset the imprints, and it is only through the protective action of PGC7 (Nakamura et al. 2012, Szabó & Piefer 2012) that imprints can be maintained pre implantation (Fig. 1B), found that most imprinted gDMRs lost DNA methylation and could not recover it even when DNMT1 was restored using a cDNA (1KO + 1), unlike the rest of the genome. However, when the rescued ES cells were used to make mice, the imprints were restored following passage through the germline (Tucker et al. 1996). This is thought to be due to the imperviousness of imprints to the normal reprogramming events seen in the soma. Indeed, the inability to reprogram in somatic tissues is a key difference between the imprinted gDMRs and the other two classes of gDMR, namely those found in germline and brain tissues (Borgel et al. 2010, Rutledge et al. 2014), whose methylation can be restored in 1KO + 1 cells (Fig. 1B). We and others have confirmed these results both in ES cells (Chen et al. 2003, Thakur et al. 2016) and in adult cell lines (Chen et al. 2003, Wernig et al. 2007) (Fig. 1B and C). Interestingly, imprints appear to be resistant to demethylation as well as remethylation: treatment of cells with 5′Aza-2′ deoxycytidine (Aza) results in robust demethylation of most targets, but not imprints (O’Neill et al. 2018).

Can somatic reprogramming of imprints occur?

One of the key features of imprinted regions is their programming in the germ line, when the parental alleles are separated. Early work showed that in Dnmt1−/− ES cells (1KO – Fig. 1B), imprints lost DNA methylation and could not recover it even when DNMT1 was restored using a cDNA (1KO + 1), unlike the rest of the genome. However, when the rescued ES cells were used to make mice, the imprints were restored following passage through the germline (Tucker et al. 1996). This is thought to be due to the imperviousness of imprints to the normal reprogramming events seen in the soma. Indeed, the inability to reprogram in somatic tissues is a key difference between the imprinted gDMRs and the other two classes of gDMR, namely those found in germline and brain genes (Borgel et al. 2010, Rutledge et al. 2014), whose methylation can be restored in 1KO + 1 cells (Fig. 1B). We and others have confirmed these results both in ES cells (Chen et al. 2003, Thakur et al. 2016) and in adult cell lines (Chen et al. 2003, Wernig et al. 2007) (Fig. 1B and C). Interestingly, imprints appear to be resistant to demethylation as well as remethylation: treatment of cells with 5′Aza-2′ deoxycytidine (Aza) results in robust demethylation of most targets, but not imprints (O’Neill et al. 2018).

However, recently two ES cell systems have been described in which imprints can apparently be restored, breaking one of the main rules of imprinting. Wong and colleagues, working with UHRF1-deficient ES cells (Qi et al. 2015), found that most imprinted gDMRs lost methylation, as expected, since UHRF1 appears to be a vital cofactor for DNMT1 in somatic cells (Bostick et al. 2007, Sharif et al. 2007). Surprisingly however, they found that when they rescued these cells using a UHRF1 cDNA that not only was methylation restored on bulk DNA, but also is a subset of the imprinted loci showed recovery. In particular, the H19, Nnat and Dlk1 gDMRs showed some recovery of methylation, though not to WT levels, and ZFP57 binding was restored at 2/3 loci (Qi et al. 2015). While it would be reasonable to assume that an underlying chromatin mark might be retained, allowing restoration of methylation, there was no clear correlation between recovery and chromatin marks at the loci (Qi et al. 2015).

Using a different ES cell system, we too found somatic recovery of imprints could occur under some conditions. In cells lacking DNMT3A and DNMT3B (3abKO), imprinted gDMRs lost their parental allele-specific methylation to almost the same extent as in DNMT1-deficient cells. However, in contrast to rescue experiments in 1KO cells, when DNMT3A2 was added back to the 3abKO cells, we found that DNA methylation was restored at most gDMRs examined (11/14), with many achieving levels similar to the WT cells (Fig. 1B). Results were confirmed using up to three different techniques (Thakur et al. 2016).
Taken together, these two sets of results independently confirm that in some ES cell types, methylation can be restored on imprinted gDMRs outside of the germline opening the way for further exploration of the factors and signals that may be involved.

Implications for reproductive biology

How many imprinted loci are there?

In the index species mouse, where most data are available and there have been some dedicated searches, the number of verified imprinted loci remains relatively steady at around 150. Despite a few high-profile studies suggesting more may exist, the number of genes with confirmed uniparental expression has not increased greatly. Many of the known imprinted genes in mouse ([http://www.mousebook.org/imprinting-gene-list](http://www.mousebook.org/imprinting-gene-list) (accessed 24/1/2018)) are also imprinted in human ([http://igc.otago.ac.nz/1601summarytable.pdf](http://igc.otago.ac.nz/1601summarytable.pdf) (accessed 24/1/2018)). A recent study by Wang et al. (2014) identified a small number of new germline DMR, for which uniparental methylation has been verified by ourselves or others (Thakur et al. 2016), but the effects on transcription are yet to be confirmed. Likewise a study in human by Court et al. (2014) using a combination of arrays and whole-genome bisulfite sequencing (WGBS – see below) added some new DMR. These were largely confined to known imprinted loci, although some were novel placenta-specific DMR. One caveat with the latter is that the differential methylation appears to have been established postnatally, which would require a radical rethink of mechanisms. However, some evidence exists to suggest that this may also occur in mouse (Wagschal & Feil 2006).

In addition to the ‘true’ imprints, there are a number of ‘transient’ imprinted genes that show widespread imprinting in the early embryo that later becomes confined to one tissue (see ‘Post-implantation changes’ section above). This larger group of neuronal and germline genes have gDMRs in the early embryo that become erased post-implantation, largely through the de novo methylation of the unmethylated paternal allele (Borgel et al. 2010, Proudhon et al. 2012, Rutledge et al. 2014). These loci have not all been tested for uniparental expression in early development, but may contribute to parent-of-origin effects in early embryogenesis.

How conserved are imprints in other species?

Initial reports of imprinting in mouse were quickly followed by the investigation of corresponding human loci, and the subsequent discovery that at least one of the index loci IGF2R was not imprinted in all humans, but may instead show polymorphic imprinting (Xu et al. 1993). As the number of well-characterised imprinted loci in mouse grew, the general trend was for the homologous locus in human to be imprinted, but with a number of notable exceptions (Hanna & Kelsey 2014, Peters 2014). In general, though, the emerging complexity of imprint locus control has tended to discourage attempts to verify in detail in the second species, with an assumption of similarity being adopted on the whole unless forced to be re-evaluated due to clinical or experimental evidence.

Some early studies also tried to map the index imprinted loci in marsupials in order to test parental conflict theories as well as to investigate evolutionary mechanisms. This met with some success, despite the difficulty of the exercise. There is now support for an origin for imprinting at the time of divergence of marsupials (metatheria) and egg-laying mammals (montremes), which lack imprinting (Rentfree et al. 2013). The oldest imprinted loci with a conserved DMR are H19 and PEG10, though a number of other imprinted genes such as IGF2R and PEG1 are also imprinted, but lack a conserved DMR. Many other loci appear to have acquired imprinting in eutheria (Rentfree et al. 2013). PEG10 is a retroposon-derived gene, and it has long been suggested that there is a link between methylation at newly acquired retrotansposons and imprinting (Yoder et al. 1997). It is interesting in this context that rodentia are notable for having the highest number of imprinted loci and also a unique methyltransferase locus Dnmt3c, initially thought to be a pseudogene based on an early draft of the rat genome (Lees-Murdock et al. 2004) but now shown to produce a functional enzyme confined to the male rodent germline that appears dedicated to retroposon methylation (Barau et al. 2016).

ART and imprints

In livestock species, there has also been a lively interest in imprinted loci (O’Doherty et al. 2012), partly due to the widespread use of artificial reproduction techniques (ART), which are thought to be particularly associated with perturbed methylation and development including syndromes such as the aforementioned LOS (see above), enlarged placenta and perinatal death. These have been tied to changes in methylation at imprinted loci in a number of studies in cattle (O’Doherty et al. 2012, 2014). While studies in mouse have found substantial support for ART-induced alterations in imprints (Rivera et al. 2008, Denomme & Mann 2012), in humans, any link between methylation changes at imprinted loci and standard IVF has been controversial, with most reviews of the area not able to fully exclude an effect on imprints or other epigenetically regulated loci (Grafodatskaya et al. 2013). However, the addition of large enough concentrations of inhibitors such as aphidicolin have conclusively been shown to alter the programmed demethylation of the genome (Guo et al. 2014) at the pronuclear stage, so
Table 1 Technologies for genome-wide methylation analysis in animals.

<table>
<thead>
<tr>
<th>Type</th>
<th>Acronym</th>
<th>Technology</th>
<th>Spp</th>
<th>Cost</th>
<th>Coverage</th>
<th>meC</th>
<th>Input</th>
<th>Res</th>
<th>Notes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Array</td>
<td>450K/EPIC</td>
<td>450K/850K probe array (Illumina)</td>
<td>Hu</td>
<td>L</td>
<td>Exome, CGI, regulatory</td>
<td>M</td>
<td>L</td>
<td>No TEs</td>
<td>Bibikova et al. (2011)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EDMA</td>
<td>EmbryoGENE DNA methylation analysis</td>
<td>Bo</td>
<td>L</td>
<td>Exome, CGI, promoter</td>
<td>M</td>
<td>L</td>
<td>Qualitative, no TEs</td>
<td>Shojaei Saadi et al. (2014)</td>
<td></td>
</tr>
<tr>
<td>Reseq</td>
<td>WGBS</td>
<td>Whole-genome bisulfite sequencing</td>
<td>Any</td>
<td>H</td>
<td>Total</td>
<td>H</td>
<td>H</td>
<td>Bioinformatic analysis-heavy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RRBS</td>
<td>Reduced representation bisulfite sequencing</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>M</td>
<td>Partial CGI cover</td>
<td>Cokus et al. (2008), Lister et al. (2008), Meissner et al. (2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PBAT</td>
<td>Post-bisulfite adaptor tagging</td>
<td>Any</td>
<td>M</td>
<td>Total</td>
<td>H</td>
<td>L</td>
<td>Amplification-free</td>
<td>Miura et al. (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MeDIP-seq</td>
<td>Methylated DNA immunoprecipitation</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>M</td>
<td>Favour sparse CpG</td>
<td>Weber et al. (2005)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MBD-CAP</td>
<td>Methyl binding domain CAPture</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>M</td>
<td>Favour CGI</td>
<td>Brinkman et al. (2010), Serre et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>uCGI</td>
<td>MRE-seq</td>
<td>Methylation-sensitive restriction enzyme</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>M</td>
<td>uCGI-specific</td>
<td>Maunakea et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>CFP1-seq</td>
<td>CFP1 protein IP</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>M</td>
<td>uCGI-specific</td>
<td>Thomson et al. (2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>scWGBS</td>
<td>Single-cell WGS</td>
<td>Any</td>
<td>H</td>
<td>Total</td>
<td>H</td>
<td>L</td>
<td>Bioinfo-heavy</td>
<td>Farlik et al. (2015)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>scRRBS</td>
<td>Single-cell RRBS</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>L</td>
<td>Bioinfo-heavy</td>
<td>Guo et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>scPBAT</td>
<td>Single-cell PBAT</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>L</td>
<td>Bioinfo-heavy</td>
<td>Smallwood et al. (2014)</td>
<td></td>
</tr>
<tr>
<td>modC</td>
<td>SMRT</td>
<td>Single molecule real time (PacBio)</td>
<td>Any</td>
<td>H</td>
<td>Total</td>
<td>H</td>
<td>H</td>
<td>All modifications</td>
<td>Flusberg et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>hMeDIP-Seq</td>
<td>5hmC antibody IP</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>M</td>
<td>5hmC-specific</td>
<td>Jin et al. (2011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oxBSeq</td>
<td>Oxidative bisulfite sequencing (CEGX)</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>M</td>
<td>5hmC and 5mC</td>
<td>Booth et al. (2012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB-Seq</td>
<td>TET-assisted bisulfite sequencing</td>
<td>Any</td>
<td>M</td>
<td>Total</td>
<td>H</td>
<td>H</td>
<td>5hmC and 5mC</td>
<td>Yu et al. (2012, 2018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fCAB-seq</td>
<td>JBP1 5hmC-binding protein IP</td>
<td>Any</td>
<td>M</td>
<td>Total</td>
<td>H</td>
<td>H</td>
<td>5fC-specific</td>
<td>Song et al. (2013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JBP1-seq</td>
<td>JBP1 5hmC-binding protein IP</td>
<td>Any</td>
<td>M</td>
<td>Fraction</td>
<td>M</td>
<td>M</td>
<td>5hmC-specific</td>
<td>Cui et al. (2014)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Type: Array, microarray; Reseq, resequencing; IP, immunoprecipitation; uCGI, targets unmethylated CGI; SC, single-cell; modC, targets modified 5mC. Spp: species; Hu, human-specific; Bov, bovine-specific; any, can in theory be applied to any. Cost: L, Low; M, Medium; H, High. Coverage: genome-wide coverage: Exome, exons; CGI, CpG islands; Regulatory, promoters and enhancers; Total, sequences whole genome; Fraction, subsets the genome but includes all types of sequence. meC: fraction of all meC covered: Low, Medium, High. Input: amount and purity of DNA: Low, lower quality and amount needed; Medium, better quality and more needed; High, highly pure DNA at higher amounts. Res: resolution at target, i.e. are all 5mC covered in target areas (e.g., RRBS) or only a fraction at each location (arrays): Low, Medium, High. Notes: TE, transposable elements; 5hmC, 5-hydroxymethylcytosine detection only; 5fC, 5-formylcytosine only.
environmental perturbations of imprinting remain at least theoretically plausible.

Use of sequencing technologies to investigate known and potential imprinted regions

There have been a number of approaches taken to identify potentially imprinted genes and to more closely define the size of the imprint, partly dictated by the tools and approaches available in the differing species. In humans for example, microarrays have made assessing methylation easier, but low heterozygosity levels and incomplete data make assignment of parental allele and gametic imprints – even with resequencing approaches – problematic, whereas in mice, no methylation arrays are available but interspecific crosses can be generated to maximise heterozygosity. Ruminants on the other hand have a long reproductive cycle and inter-strain hybrids are not commonly used, and with some exceptions, there are no arrays here either. For these reasons, resequencing approaches offer the most widely applicable tool (Table 1). While sequencing can be restrictive in cost, it represents a powerful discovery platform.

Wang and colleagues for example used WGBS to investigate the methylation status of the 54 imprinted DMR in the mouse genome (Wang et al. 2014), and in particular to sort unclassified DMR into gametic or somatic categories. They compared oocyte, sperm and early embryo, and began by confirming they could correctly sort all the DMR whose status was known (29) into gametic vs somatic types. They then used the same criteria to classify 25 imprinted DMRs whose status was unknown into gametic or somatic DMR, which is essential for determining where primary control of each locus is being exerted. At least four of these newly classified DMRs could subsequently be independently confirmed using a pyrosequencing-based assay (Thakur et al. 2016).

In humans, microarray technologies and in particular the Illumina EPIC array, have become so reliable in terms of assessing methylation levels quantitatively that confirmation by a second technology has become almost redundant. The low relative cost of the array makes it an attractive approach for initial screening and a comparison of uniparental disomies could correctly identify all 30 DMRs for which the array had probes (one DMR at MEG3/DLK1 is not covered). They additionally identified 21 novel putative imprints, 15 specific to placenta, albeit a number of them were at known imprinted loci already (Court et al. 2014). However, the low relative resolution of the array meant that they needed to use WGBS to accurately delineate the DMR.

Reliable standardised methylation microarrays are largely confined to humans, though the success of the 450K and EPIC platforms has meant that Illumina appears to be considering a similar BeadChip array for mouse. An exception is cattle, where the EDMA array has been developed for assessing genome-wide methylation (Shojaei Saadi et al. 2014), though this is based on cutting with methylation-sensitive restriction enzymes rather than oligos. While the array gives good internal consistency and technical reproducibility and could correctly identify methylation status at a number of imprinted regions in this species, it gives relative rather than absolute levels of methylation and is therefore difficult to correlate with outputs from many other assays (Desmet et al. 2016), though not all (Ispada et al. 2018). Whole-genome sequencing represents the gold standard for exploratory work and can definitively assess methylation across all sites. Confirming parental origin of the methylation mark requires heterozygous SNPs

Figure 2 Schematic comparison of reduced representation bisulfite sequencing (RRBS) vs post-bisulfite adaptor tagging (PBAT). In RRBS (left), the DNA is initially fragmented by a methylation-insensitive restriction endonuclease such as MspI to an optimal fragment size of between 400 and 600bp, and this fraction isolated following gel electrophoresis or similar. Adaptors are ligated to these DNA fragments that are subsequently treated through bisulfite conversion to convert unmethylated cytosines to thymines for the differential readout. Amplification is carried out using primers specific to the adaptor sequences, but there is substantial loss of sample at this point due to strand breakage (flashes) between adaptors caused by bisulfite. PBAT (right) skips the initial enzyme digestion and instead takes advantage of the bisulfite-induced breakage to fragment the DNA here. Instead of PCR, the adaptors are used for two rounds of random primer extension, which serves to produce numerous reads from a small amount of input genomic DNA.
however. While Liu and colleagues could find sufficient variation in an inter-strain cross for mice (Wang et al. 2014), generating cross-bred cattle is not trivial given cost implications and slow reproductive times and most reported work in the area does not feature such animals.

An alternative, more widely applicable and cost-effective approach for assessing methylation is to subset the genome and do targeted resequencing. Reduced representation bisulfite sequencing has been the most extensively used method in this category so far (Meissner et al. 2005), but is being superseded by post-bisulfite adaptor tagging (PBAT) (Miura et al. 2012), which appears to offer improvements in yield and efficiency, particularly when starting with small amounts of material such as oocytes (Smallwood et al. 2014) (Fig. 2). Apart from cost considerations, a significant bottleneck for resequencing-based analyses is the extensive bioinformatic analysis required, with concomitant need for higher-specification computing hardware, local network clusters and data warehousing, all adding substantial capital and human resource costs to the equation (Lewitter et al. 2009). While this has resulted in the majority of such work being undertaken at large central institutes that have developed the required infrastructure, often with national support, there remains a niche for smaller players out-sourcing the sequencing to larger centres with surplus capacity and who may in future also be able to avail of cloud-based storage and analysis capabilities (Liu et al. 2014).

Another point to consider here is that some species that may be of interest for biologists wishing to understand the origins or phylogenetic reach of imprinting may not be sequenced yet. In cases where a reference genome framework is unavailable, assembling sequence reads and defining variation become considerably more challenging (Cabais et al. 2012).

Finally, the impact of third-generation sequencing technologies is yet to be fully felt, given that this market is currently in flux and no clear front-runner appears yet to have emerged. Current technologies require a separate sequence library for each type of cytosine modification for example (meC, 5hmC, 5iC, 5caC) (Yuet al. 2012), whereas certain new approaches (Table 1) such as PacBio’s technology have the capability to also call methylated bases while sequencing, obviating the need for multiple sequence sets (Lister & Ecker 2009).

Conclusions and future prospects

Imprinting clearly plays an important role in body size determination and nutrient transfer to the young, making assessment of methylation at key loci a useful tool for reproductive biologists and breeders alike. Despite the improvement in technologies for assessing and delineating imprinting, the number of verified imprinted loci has not increased very significantly. However, our understanding of the mechanisms by which these regions are established and maintained in early life, as well as the contexts in which some plasticity in imprinting may occur, have greatly increased. We can look forward to further insights from comparative genomic approaches based on the increasing application of sequencing technologies to an ever-expanding range of organisms, including the so far rather surprisingly neglected livestock species, which play such an important role in human food production and welfare.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

The authors’ work was supported in part by the Medical Research Council (grant number MR/J007773/1) and a grant jointly funded by the ESRC/BBSRC (grant number ES/N000323/1).

Acknowledgements

The authors thank members of the Walsh lab for helpful comments and apologize to colleagues whose work could not be cited owing to length limitations.

References

Allen WR 1969 Factors influencing pregnant mare serum gonadotrophin production. Nature 223 64–65. (https://doi.org/10.1038/223064a0)
Barlow DP & Bartolomei MS 2014 Genomic imprinting in mammals. Cold Spring Harbor Perspectives in Biology 6 (https://doi.org/10.1101/cshperspect.a018382)
Bartolomei MS & Ferguson-Smith AC 2011 Mammalian genomic imprinting. Cold Spring Harbor Perspectives in Biology 3 a018382. (https://doi.org/10.1101/cshperspect.a002592)
Barton SC, Surani MA & Norris ML 1984 Role of paternal and maternal genomes in mouse development. Nature 311 374–376. (https://doi.org/10.1038/311374a0)

Downloaded from Bioscientifica.com at 12/28/2018 09:49:29AM via free access
Imprint stability


Chen Z, Robbins KM, Wells KD & Rivera RM 2013 Large offspring syndrome induced by assisted reproduction. Nature Genetics 45 212–215. (https://doi.org/10.1038/nat.1422088112)


Cui L, Chung TH, Tan D, Sun X & Jia XY 2014 JBP1-seq: a fast and efficient method for genomic DNA methylation profiling of ShmC. Genomics 104 368–375. (https://doi.org/10.1016/j.s感悟.2014.08.023)


Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, Leymaster KA & Jirtle RL & Smith TP 2002b Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. Genome Research 12 1496–1506. (https://doi.org/10.1101/gr.571002)


O’Doherty A, O’Shea LC & Fair T 2012 Bovine DNA methylation imprints are established in an oocyte-size-specific manner, which are coordinated with the expression of the DNMT3 family proteins. Biology of Reproduction 86 1–10. (https://doi.org/10.1095/biolreprod.111.094946)
Mechanisms of Development 113 91–94. ⟨https://doi.org/10.1016/S0925-4773(02)00002-3⟩
Younis S, Schonne M, Massart J, Hjortebjerg R, Sundstrom E, Gustafson U, Bjornholm M, Krook A, Frystyk J, Zierath JR et al. 2018 The ZBED6-IGF2 axis has a major axis has a major effect on growth of skeletal muscle and internal organs in placental mammals. PNAS 115 2048–2057. ⟨https://doi.org/10.1073/pnas.1719278115⟩
Yu M, Han D, Hon GC & He C 2018 Tet-assisted bisulphite sequencing (TAB-seq). Methods in Molecular Biology 1788 645–663. ⟨https://doi.org/10.1007/978-1-4939-7481-8_33⟩