

# Unique patterns of evolutionary conservation of imprinted genes

Martina Paulsen

Received: 30 March 2011 / Accepted: 27 June 2011 / Published online: 7 July 2011  
© Springer-Verlag 2011

**Abstract** During mammalian evolution, complex systems of epigenetic gene regulation have been established: Epigenetic mechanisms control tissue-specific gene expression, X chromosome inactivation in females and genomic imprinting. Studying DNA sequence conservation in imprinted genes, it becomes evident that evolution of gene function and evolution of epigenetic gene regulation are tightly connected. Furthermore, comparative studies allow the identification of DNA sequence features that distinguish imprinted genes from biallelically expressed genes. Among these features are CpG islands, tandem repeats and retrotransposed elements that are known to play major roles in epigenetic gene regulation. Currently, more and more genetic and epigenetic data sets become available. In future, such data sets will provide the basis for more complex investigations on epigenetic variation in human populations. Therein, an exciting topic will be the genetic and epigenetic variability of imprinted genes and its input on human disease.

**Keywords** Genomic imprinting · Mammalian evolution · Repetitive elements · DNA methylation · CpG island

## Epigenetic aspects of mammalian evolution

The genetic diversity of human individuals has a major influence on human health. Hence, the understanding of evolutionary processes in mammalian genomes, especially

in the human, can help to understand individual- or population-specific differences in the genetic components of human diseases. Investigations on sequence conservation may support the identification of conserved functional genomic elements such as regulatory elements, whereas sequence divergence might highlight differences in gene regulation or function between related species, for example, between the human and the mouse that is frequently used as model organism in medical research.

Whereas it is clear that factors such as the mutability of DNA sequences and natural selection play a role in shaping the mammalian genome, the role of epigenetics in the evolution of DNA sequences is more complex and therefore not very easy to access. Epigenetic modifications are linked to the functionality of regulatory elements. Thus, both, the regulatory elements and their epigenetic modifications, are subjected to natural selection. Furthermore, it is known that epigenetic modifications are important for the stability of chromosomes and may directly influence mutability of DNA sequences and subsequent DNA repair.

An interesting group of genes whose evolution is influenced by a complex network of genetic and epigenetic factors and environment are imprinted genes. In contrast to most other genes in mammals, imprinted genes are only expressed from one of the two parental alleles, and the parental origin determines which allele is silenced, and which is activated. During germ cell development, imprinted genes acquire different epigenetic marks in the male and female germ lines. After fertilization, these marks are maintained. The epigenetic difference between the two parental alleles is reflected in the silencing of one gene copy whereas the other copy remains active.

During the past decade, approximately 100 genes have been shown to be affected by genomic imprinting, and there is evidence that more than 1,000 genes exhibit an

---

M. Paulsen (✉)  
FR 8.3 Biowissenschaften, Lehrstuhl für Genetik/Epigenetik,  
Universität des Saarlandes,  
Postfach 151150, 66041 Saarbrücken, Germany  
e-mail: m.paulsen@mx.uni-saarland.de

expression bias towards one of the two parental alleles (Catalogue of Parent of Origin Effects, <http://igc.otago.ac.nz/home.html>; Gregg et al. 2010). The increasing number of genes affected by parental origin-dependent epigenetic regulation emphasizes the relevance of germ line-specific epigenetic modifications. This is particularly important in the light of advances made in artificial reproduction technologies. It is still controversially discussed if these technologies are associated with an elevated risk of epigenetic aberrations (Owen and Segars 2009). Furthermore, there are concerns that an increased environmental exposition to hormone analogues results in fertility problems and possibly also in aberrant imprinting (Kobayashi et al. 2009; Anway et al. 2005).

Investigating the evolution of imprinted genes might highlight the relationships between DNA sequence and epigenetic gene regulation. Imprinted genes should possess specific features in their DNA sequence that are responsible for establishment or maintenance of germ line-specific epigenetic modifications. Hence, imprinted genes may represent ideal models for the question on how interaction of functional and epigenetic factors has shaped the mammalian genome during evolution.

### Special roles of imprinted genes in human evolution?

Human evolution is marked by evolution of cognitive skills. In addition, increased energy consumption of the brain, and changing climates, resulted probably in pronounced changes in energy metabolism. Many imprinted genes act as growth regulators, thereby influencing energy metabolism; others are predominantly expressed in the brain and are functionally linked to postnatal behaviour (Davies et al. 2005). Hence, the contribution of imprinted genes to human evolution might have been quite pronounced. That this might be indeed the case is indicated by the observation that the imprinted region on human chromosome 14 is among the top 20 genomic regions with evidence for strong positive selection during human evolution (Green et al. 2010). In terms of genome evolution, this is particularly interesting as this region contains the well-known growth factor gene *DLK1* and a large cluster of microRNAs that are expressed in the brain.

Possibly associated to their functions as growth regulators, prominent imprinted genes such as the paternally expressed *IGF2* and *DLK1* genes are deregulated in many different types of tumours (Khoury et al. 2010; Kaneda and Feinberg 2005). Defects in imprinted gene regulation are associated with several growth syndromes in humans, such as Beckwith–Wiedemann syndrome, Prader–Willi syndrome and intrauterine growth restriction (Schofield et al. 1989; Chaillet et al. 1991; Monk and Moore 2004).

Furthermore, imprinted genes appear to affect behaviour: For female mice with a defect in the paternally expressed *Peg1* gene, it was shown that they cared less for their offspring (Lefebvre et al. 1998). In the human, it has been speculated that autistic syndromes might be influenced by parental origin effects (Badcock and Crespi 2006). This might partially be explained by a genetic linkage of autism to a genomic region that neighbours the imprinted Prader–Willi/Angelman syndromes region on human chromosome 15.

### Different responses to natural selection shaped maternally and paternally expressed genes

Due to their mono-allelic expression patterns, imprinted genes are usually seen as functional haploids, i.e. only the active gene copy contributes to the phenotype, whereas the silent copy is irrelevant in functional terms. As a result, imprinted genes are supposed to react stronger against purifying selection, whereas positive selection is predicted to be slowed down. Hence, imprinted genes should be more conserved than biallelically expressed genes.

Several hypotheses have been established that try to explain the evolution of imprinted gene expression from a functional point of view. Among all these hypotheses, the kinship theory is the most prominent one (Moore and Haig 1991). This theory has been inspired by two observations: (1) In the animal kingdom, genomic imprinting is a mechanism of gene regulation that is most prominent in mammals, i.e. in species in which the embryo is in direct contact with the mother and (2) quite a number of imprinted genes are involved in the regulation of embryonic growth. Many paternally expressed genes encode growth factors, whereas among the maternally expressed genes are prominent growth-inhibiting genes.

The kinship theory suggests that the repression of growth-inhibiting genes in the male germ line induces enhanced growth of the embryo. This would result in an increased nutritional demand and consequently in an increased exploitation of the mother's resources, thereby the female's chances on further successful pregnancies would be lowered. In case of polygamous species, this would not necessarily affect the male's interests as it is by no means certain that he would father further offspring with the same female.

Counteracting the paternal silencing of growth inhibiting genes, the repression of growth-enhancing genes in the female germ line would reduce embryonic growth. This would save the female's resource and would increase the chances for further pregnancies.

One prominent feature of the human species is the rather immature state of the neonate and a prolonged postnatal childhood and adolescence. Given that imprinted genes are

regulators of embryonic growth, the observed changes in embryonic development might be related to the evolution of these genes in the human.

### **Paternally and maternally expressed genes show differences in the conservation of protein-encoding sequences**

In a recent study, we have asked if the mono-allelic gene expression patterns or the evolution of special functions have left specific marks in the DNA sequences of imprinted genes. Interestingly, this seems indeed to be the case. We detected very complex patterns of sequence conservation, particularly in their protein-encoding exons of imprinted genes (Hutter et al. 2010a). Whereas paternally expressed genes are conserved at similar level as biallelically expressed genes, maternally expressed genes show a relaxed conservation. On protein level, the divergence of maternally expressed genes is milder. The frequency of non-synonymous mutations does not reach a level that would indicate strong positive selection. Thus, the increased divergence of maternally expressed genes is more likely caused by reduced purifying selection. Although on the first view very puzzling, the different conservation patterns of paternally and maternally expressed genes are possibly caused by the phenotypic consequences of their different expression patterns as described by the kinship theory: The described scenario implies that maternally expressed genes of the embryo have contra-acting effects on the fitness of embryo and on the fitness of the mother who has transmitted the expressed gene copy. In contrast, paternally expressed genes in the embryo do not have a feedback effect on the fitness of the transmitting father. Hence, on average, mutations in paternally expressed genes will have a stronger (unidirectional) influence on the net fitness of carriers than mutations in maternally expressed genes that affect the fitness of females and their offspring in opposite directions. Therefore, purifying selection should affect paternally expressed genes more than maternally expressed ones.

Interestingly, the increased divergence of maternally expressed genes is prominent in rodents, whereas in the human, a divergence is less evident (Hutter et al. 2010a). On one hand, this might indicate that imprinted genes played species-specific roles during evolution. On the other hand, the human genome is slowly evolving due to long generation times. In comparison to rodents, mutation rates in the human are low. Therefore, unique patterns in the conservation of human imprinted genes might be less visible. As more human genomes become sequenced and the sequence data quality for other primates continuously increases, a reinvestigation of this topic addressing specif-

ically the input of imprinted genes on human evolution is certainly of high interest.

### **Conservation of noncoding sequences in imprinted regions**

Gene function is not only determined by the sequences and structural properties of encoded proteins or RNAs but also by tissue-specific and temporal expression patterns. Imprinted genes are expressed in a broad spectrum of postnatal tissues in human and mouse. In these tissues, imprinted gene expression patterns are conserved at similar levels as those of other genes (Steinhoff et al. 2009). For embryonic stages, the conservation of expression patterns is difficult to access because information for human embryonic stages is scarce.

Responsible for gene-specific expression patterns are DNA elements in the promoter regions and additional regulatory elements, such as enhancer or silencers, that reside in some distance to the transcriptional start site. Furthermore, special elements in introns or in the untranslated regions of mRNAs play a role in posttranscriptional mRNA splicing or influence RNA stability. These different types of regulatory elements are often conserved between different mammalian species. Hence, high sequence conservation outside of protein-encoding exons can help to identify regulatory elements that control conserved gene regulation on transcriptional or posttranscriptional level.

Comparing imprinted to other autosomal genes, it becomes evident that there are only few differences in sequence conservation outside of protein-encoding exons (Hutter et al. 2010b). For example, at exon–intron boundaries, paternally expressed genes possess more conserved elements than maternally expressed genes. This is probably related to the higher conservation of protein-encoding sequences of paternally expressed genes.

Some differences in conservation patterns are visible in intergenic regions, where conserved elements, particularly of maternally expressed genes, are shorter and less well conserved than those of other autosomal genes (Hutter et al. 2010b). As the regulatory capacities of conserved elements in maternally expressed genes have not been systematically analysed yet, it remains unclear if reduced conservation in intergenic regions is associated with evolutionary divergence of gene regulation.

### **Multiple roles of CpG methylation in the evolution of imprinted genes**

In mammalian species, the gross of cytosine methylation occurs at CpG positions. In contrast to histone modifica-

tions that are not directly linked to the DNA, cytosine methylation appears to have a strong direct influence on the evolution of DNA sequences: In case of cytosine deamination, an unmethylated cytosine is converted into uracil, whereas a methylated cytosine is converted into thymidine. Related to the fact that uracil is a foreign base in DNA, uracil-containing mismatches are probably more efficiently repaired than thymidine-containing mismatches. Due to this phenomenon, genomic regions whose DNA is methylated in the germ lines become gradually CpG poor, whereas unmethylated regions, such as CpG islands, keep their CpG content (Bird 1986). Furthermore, especially CpG islands in promoter regions act as regulatory elements and are therefore subjected to a strong purifying selection that supports a conservation of CpG positions.

Although genome-wide DNA methylation data are now available, there is still little information about (allele-specific) methylation patterns of imprinted genes. Nevertheless, it is generally believed that allele-specific DNA methylation patterns of imprinted genes are restricted to well-defined differentially methylated regions (DMRs) that act as central regulatory elements. Interestingly, outside of CpG islands, the conserved elements of imprinted genes possess an elevated CpG content that appears to be the result of reduced CpG to TpG deamination (Hutter et al. 2010b). It is tempting to speculate that the elevated CpG content points towards reduced CpG methylation levels. However, as described above, due to their mono-allelic expression, imprinted genes differ from non-imprinted genes in their reaction to natural selection. Hence, the high CpG content of imprinted genes might be the result of strong purifying selection acting on CpG positions. Consequently, the here raised questions call for more detailed investigations of allele-specific DNA methylation patterns outside known DMRs. As more and more datasets become available that allow the linkage between SNPs and methylated/unmethylated neighbouring CpG positions, such investigations might soon be possible.

One consequence of an elevated CpG content of imprinted genes is apparently the higher frequency of annotated CpG islands (Hutter et al. 2010b). Especially introns of imprinted genes harbour more CpG islands than intron of other autosomal genes. This observation is of particular interest as some prominent DMRs lie in intronic regions where they serve as promoters of long antisense transcripts. These antisense RNAs are mono-allelically expressed and mediate silencing of the host gene and also of neighbouring genes on the same chromosome (Mancini-DiNardo et al. 2006; Sleutels et al. 2002). Hence, the increased numbers of intronic CpG islands might represent promoters of antisense transcripts that act in epigenetic processes.

### **Conserved repeat structures—the unconventional conservation of differentially methylated regions**

Allele-specific expression of imprinted genes is controlled by DMRs. Some of these DMRs acquire their allele-specific methylation marks already in the germ lines. In imprinted regions, these so-called germ line DMRs serve as central regulatory elements. After fertilization, germ line DMRs interact with other regulatory elements in cis and induce the establishment of secondary epigenetic marks, for example, in the promoter regions of neighbouring genes (Lin et al. 2003; Fitzpatrick et al. 2002).

In order to act as regulatory elements, such DMRs need to be differentially recognized and epigenetically modified in the germ lines. Furthermore, special sequence features of the DMRs might be important for the maintenance of epigenetic imprints after fertilization. Hence, special DNA elements within the DMRs or in close vicinity to them should guide the epigenetic machinery during establishment and maintenance of the imprints. As imprinted gene expression appears to be highly conserved among different mammalian species (Steinhoff et al. 2009), one might expect a tight sequence conservation in DMRs. Interestingly, this is not the case. Instead, the prominent feature of DMRs is the presence of tandem repeat arrays. It has been speculated that these structural elements are important for the acquisition of germ line-specific DNA methylation patterns (Paulsen et al. 2005; Hutter et al. 2006). This hypothesis is based on the observation that repetitive elements in general and also transgenes that are present in multiple copies are heavily DNA methylated (Neumann et al. 1995).

The presence of tandem repeats in the orthologous DMRs of different species is often conserved, but numbers and sequences of repeated motifs are usually rather divergent. This indicates that for regulatory function, secondary structures rather than specific DNA motifs are of importance. Interestingly, tandem repeats have been found in maternally as well as paternally methylated DMRs; hence, tandem repeats are not a feature that induces DNA methylation specifically in only one of the parental germ lines.

### **Imprinted genes show short interspersed transposable element depletion**

Although the tandem repeats of DMRs are indeed a prominent feature of DMRs, there are indications that they are not the only elements important for establishment and maintenance of epigenetic imprints. For this reason, the search for typical sequence features addresses also the sequence environment of DMRs. Here, retrotransposed elements have been regarded as interesting candidates. For

example, a high concentration of LINE1 elements on the mammalian X chromosome suggests that LINE1 elements might be involved in the X chromosome inactivation process in females (Bailey et al. 2000). Also, imprinted genes support the idea of a functional role of repetitive elements in mammalian epigenomes. In imprinted genes, the LINE1 element content is only mildly elevated. Instead, these genes are depleted of short interspersed transposable elements, to which the primate-specific Alu elements belong (Walter et al. 2006; Greally 2002). In the human, about 30% of CpG lie within Alu elements. Hence, enrichment or depletion of repetitive elements might change the CpG content of a genomic region dramatically. This might have various consequences for the epigenetic status of regulatory elements in neighbouring DNA segments: Repetitive elements might attract DNA methyltransferases to neighbouring CpG islands, thereby hypermethylation of these regulatory elements might be induced. In a different scenario, a drastic reduction in CpG content due to the absence of CpG-rich repetitive elements might result in a more efficient methylation of the remaining CpGs. Considering that the male and female germ lines differ in their DNA methylation capacity, this might result in increased or decreased DNA methylation in the affected genomic regions in one germ line but not in the other.

## Conclusions

The here described studies show that imprinted genes possess sequence properties that distinguish them from other mammalian genes. The different functions of paternally and maternally expressed genes are reflected in differences in the protein-encoding sequences. Furthermore, imprinted genes show an enrichment of CpGs in highly conserved elements and possess more CpG islands than other genes. In addition, tandem repeats in the DMRs and unusual densities of retrotransposed elements are typical features that indicate a close relation between evolution of DNA sequence and epigenetic gene regulation in imprinted regions.

With the expansion of epigenomics and systems biology as modern research fields in life science, investigations on the interaction between epigenetic gene regulation and phenotypic aspects will become very interesting topics. As in the evolution of imprinted genes complex interactions of functional and epigenetic factors are clearly evident, these genes will be interesting models for projects that address such issues. In the future, the upcoming wave of sequence data will allow more detailed investigations of the evolution of imprinted genes in human. For example, it will be very interesting to study individual-specific genetic variability in imprinted genes and its association with individual-specific

imprinting effects. Furthermore, a detailed analysis of imprinted genes will facilitate to identify more imprinted genes and may also help to identify genes that are sensitive to aberrant epigenetic modifications that might be caused during germ cell development.

**Acknowledgement** Computational studies of imprinted genes in the Paulsen laboratory are funded by the Deutsche Forschungsgemeinschaft DFG (PA750/3-1).

**Conflict of interest** The author declares that she has no conflict of interest.

## References

- Anway MD, Cupp AS, Uzumcu M, Skinner MK (2005) Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* 308:1466–1469
- Badcock C, Crespi B (2006) Imbalanced genomic imprinting in brain development: an evolutionary basis for the aetiology of autism. *J Evol Biol* 19:1007–1032
- Bailey JA, Carrel L, Chakravarti A, Eichler EE (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci USA* 97:6634–6639
- Bird AP (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321:209–213
- Chaillet JR, Knoll JH, Horsthemke B, Lalande M (1991) The syntenic relationship between the critical deletion region for the Prader-Willi/Angelman syndromes and proximal mouse chromosome 7. *Genomics* 11:773–776
- Davies W, Isles AR, Wilkinson LS (2005) Imprinted gene expression in the brain. *Neurosci Biobehav Rev* 29:421–430
- Fitzpatrick GV, Soloway PD, Higgins MJ (2002) Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet* 32:426–431
- Greally JM (2002) Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci USA* 99:327–332
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH et al (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722
- Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329:643–648
- Hutter B, Bieg M, Helms V, Paulsen M (2010a) Divergence of imprinted genes during mammalian evolution. *BMC Evol Biol* 10:10
- Hutter B, Bieg M, Helms V, Paulsen M (2010b) Imprinted genes show unique patterns of sequence conservation. *BMC Genomics* 11:649
- Hutter B, Helms V, Paulsen M (2006) Tandem repeats in the CpG islands of imprinted genes. *Genomics* 88:323–332
- Kaneda A, Feinberg AP (2005) Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk. *Cancer Res* 65:11236–11240
- Khoury H, Suarez-Saiz F, Wu S, Minden MD (2010) An upstream insulator regulates DLK1 imprinting in AML. *Blood* 115 (11):2260–2263
- Kobayashi H, Hiura H, John RM, Sato A, Otsu E, Kobayashi N, Suzuki R, Suzuki F, Hayashi C, Utsunomiya T et al (2009) DNA methylation errors at imprinted loci after assisted conception originate in the parental sperm. *Eur J Hum Genet* 17:1582–1591

- Lefebvre L, Viville S, Barton SC, Ishino F, Keverne EB, Surani MA (1998) Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene *Mest*. *Nat Genet* 20:163–169
- Lin SP, Youngson N, Takada S, Seitz H, Reik W, Paulsen M, Cavaille J, Ferguson-Smith ACF (2003) Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12. *Nat Genet* 35:97–102
- Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM (2006) Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. *Genes Dev* 20:1268–1282
- Monk D, Moore GE (2004) Intrauterine growth restriction—genetic causes and consequences. *Semin Fetal Neonatal Med* 9:371–378
- Moore T, Haig D (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet* 7:45–49
- Neumann B, Kubicka P, Barlow DP (1995) Characteristics of imprinted genes. *Nat Genet* 9:12–13
- Paulsen M, Khare T, Burgard C, Tierling S, Walter J (2005) Evolution of the Beckwith–Wiedemann syndrome region in vertebrates. *Genome Res* 15:146–153
- Owen CM, Segars JH (2009) Imprinting disorders and assisted reproductive technology. *Semin Reprod Med* 27(5):417–428
- Schofield PN, Lindham S, Engström W (1989) Analysis of gene dosage on chromosome 11 in children suffering from Beckwith–Wiedemann syndrome. *Eur J Pediatr* 148:320–324
- Sleutels F, Zwart R, Barlow DP (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415:810–813
- Steinhoff C, Paulsen M, Kielbasa S, Walter J, Vingron M (2009) Expression profile and transcription factor binding site exploration of imprinted genes in human and mouse. *BMC Genomics* 10:15
- Walter J, Hutter B, Khare T, Paulsen M (2006) Repetitive elements in imprinted genes. *Cytogenet Genome Res* 113:109–115