Proc. Assoc. Advmt. Anim. Breed. Genet. 18:1-6

IMPRINTED GENES IN MAN AND MOUSE ARE MODEL SYSTEMS IN COMPARATIVE EPIGENOMICS

Martina Paulsen

Saarland University, FR 8.3 Life Sciences, PO Box 151150, D-66041 Saarbruecken, Germany.

SUMMARY

In mammals, epigenetic modifications are key players in gene regulation and genome stability. Consequently, the epigenetic protein machinery and epigenetically modified regulatory elements, such as promoter regions, are highly conserved among mammals. Hence, comparative studies of mammalian epigenomes may help to understand the mechanisms and functions of epigenetic gene regulation. Imprinted genes that are mono-allelically expressed due to allele-specific epigenetic modifications of their regulatory sequences have been recognized as ideal model systems in epigenetics. For this reason, detailed comparative studies in epigenomics were firstly initiated on imprinted genes. These analyses have resulted in the identification of new imprinted genes and regulatory elements and have highlighted complex patterns of conservation that includes not only sequence conservation but also structural elements such as the presence of tandem repeats and retrotransposed elements. Comparative studies on these genes have been extended to other topics such as the conservation of tissue-specific gene expression patterns. These analyses show that the tight conservation of epigenetic regulation of imprinting does not prevent the divergence of tissue-specific gene expression patterns that might be associated with new species-specific functions of imprinted genes.

EPIGENOMICS - RELATIONSHIPS BETWEEN GENETIC INFORMATION AND CHROMATIN STRUCTURE

In eukaryotic species, inheritance of information is not only based on the sequence of the DNA but also on epigenetic modifications of the chromatin. These modifications include modification of histone proteins and modifications of the DNA that do not affect the DNA sequence (Jenuwein and Allis 2001). Both types of modifications determine the structure of the chromatin and thereby, influence the expression of genes. Especially for DNA methylation, it has been shown that methylation patterns once they are established can be transmitted through numerous cell divisions. Nevertheless, especially in mammals, epigenetic modifications are substantially changed during development. Epigenetic reprogramming affects development of germ cells, embryonic stem cell development and differentiation processes, thereby indicating that epigenetic research has a substantial input into stem cell and reproduction research (Hemberger et al. 2009, Dean et al. 2001).

During the last few years, the genomic sequences of numerous mammalian species and an increasing number of human individuals have been made available. In parallel, experimental techniques have been established to analyse epigenetic modifications on a genome-wide level (Weber et al. 2005, Zhang et al. 2009, Roh et al. 2005, Bernstein et al. 2004). Current research in epigenomics focuses on the understanding of interactions between proteins involved in epigenetic processes and regulatory sequences, such as CpG islands in promoter regions, that attract a distinct type of epigenetic modifications. For example, the presence of specific DNA sequence motifs, the structure of the double helix, and overlap with repetitive elements have a strong influence on the chromatin structure of CpG islands (Bock et al. 2006, Bock et al. 2007).

The comparison of genetic and epigenetic features in different species allows, on one hand, the efficient identification of conserved genetic elements involved in epigenetic gene regulation, and on the other hand, highlights epigenetic differences that may restrict the usage of model

organisms, such as the mouse, in fields like epigenetic topics in medical research. On the level of species, it has become clear that mammalian species possess a well-conserved epigenetic machinery that encompasses evolutionary conserved proteins, such as DNA methyltransferases, histone modifying enzymes, and proteins that bind to specific chromatin structures (Yokomine et al. 2006, Bestor 2000). Also regulatory sequences, such as promoters, are highly similar in different species. Nevertheless, comparisons of the human and mouse genomes have shown that these species may exhibit some differences in their epigenetic regulatory elements. For example, CpG islands in the human are longer than in the mouse, and due to the acquisition of lineage-specific Alu elements, the human genome possesses more CpG islands that reside in repetitive elements (Hutter et al. 2009, Zhang et al. 2004). Besides differences between species, differences among (human) individuals receive more focus in epigenetic research, since such these differences might contribute to phenotypic diversity and might represent additional risk factors in human disease.

IMPRINTED GENES ARE MODELS FOR EPIGENETIC GENE REGULATION

In therians, a number of genes are expressed only from one of the two parental chromosomes. These so-called imprinted genes acquire different epigenetic modifications in the parental germlines that are maintained after fertilization. Subsequently, these different epigenetic marks result in silencing of one gene copy whereas the other copy remains active. As the mono-allelic expression of imprinted genes depends solely on the differential epigenetic modifications of the parental gene copies, these genes represent an ideal model system in epigenetics.

Imprinted gene expression is a specific way of gene regulation that is seen in eutherian species and to some extent also in marsupials, but appears to be absent in other vertebrate species (Killian et al. 2001). Therefore, imprinted genes are intensively investigated in human and mouse, and to some extent also in cattle and marsupials (Gebert et al. 2006). The comparison of these genes in different species has several purposes. Firstly, it aims to identify features of the DNA sequence that are responsible for establishment and maintenance of mono-allelic gene expression. A second goal is the identification of all genes that are prone to be imprinted. Last but not least, the comparison of imprinted genes in different species should help to understand the evolution of imprinting in therian species and should highlight conserved functions of these genes in this clade.

To date, more than mammalian 150 transcripts imprinting effects have been noted (Morison and Reeve 1998, http://igc.otago.ac.nz/home.html). With few exceptions, genes whose allele-specific expression patterns have been analysed in human as well as in mouse are imprinted in both species, indicating strict conservation of imprinting among eutherian species. Most imprinted genes are organized in so-called imprinted regions in the mammalian genome, i.e. imprinted genes are neighboured by other imprinted genes (MRC Harwell, UK, http://www.har.mrc.ac.uk/research/genomic_imprinting/index.html). In these regions, central differentially methylated regions (DMR) control the mono-allelic expression patterns of neighbouring genes. Strict conservation of the overall physical structures of imprinted regions, as indicated by the conserved presence and order of orthologous imprinted genes, suggests that an evolutionary conserved gene arrangement is required for cis allele-specific interactions between genes and regulatory elements (Paulsen et al. 2005, Paulsen et al. 2001). A typical feature of imprinted regions is the presence of evolutionary conserved non-coding RNAs. This includes small RNAs, such as microRNAs and snoRNAs, and longer non-translated transcripts that often represent antisense-transcripts of protein-encoding genes. The lack of conserved non-coding RNA genes in the vicinity of the orthologs of imprinted genes in non-mammalian vertebrates indicates that the evolution of these non-coding RNAs coincided with the evolution of imprinted gene expression in the respective genomic regions (Paulsen et al. 2005, Edwards et al. 2008, Smits et al. 2008). Especially for long imprinted non-translated transcripts, it has been shown that they

mediate cis long-range epigenetic silencing of overlapping or neighbouring protein-encoding genes (Fitzpatrick et al. 2002). Hence, the establishment of non-coding transcripts in these regions might have been a crucial event in the evolution of imprinted gene expression.

REPETITIVE ELEMENTS IN IMPRINTED REGIONS

In mammalian genomes, retrotransposable elements are usually epigenetically silenced, thereby preventing retro-transposition events especially in the germlines. Since LINE1 elements are enriched on eutherian X chromosomes, it has been proposed that repetitive elements might be involved in epigenetic silencing processes (Bailey et al. 2000). Consistent with this hypothesis, unusual densities of repetitive elements have been observed in imprinted regions in human and mouse (Greally 2002, Walter et al. 2006). This affects mostly SINE elements that are under-represented in imprinted regions. However, so far, there is no experimental proof that these elements might indeed exhibit special allele-specific patterns of epigenetic modifications in imprinted regions. Similarly, there are no mouse models available in which such elements have been deleted or amplified in order to show that they attract indeed germ-line specific DNA methylation or histone modification patterns.

The unusual densities and distribution of repetitive elements in imprinted genes have been exploited by bioinformatic approaches addressing the identification of new imprinted genes (Luedi et al. 2005, Luedi et al. 2007) In the meantime, other studies used experimental approaches for the systematic genome-wide discovery of imprinted genes (Nikaido et al. 2003). Unfortunately, these studies showed little overlap of candidates and subsequent studies which tried to validate imprinted gene expression of predicted candidates showed that only few of them were indeed imprinted (Ruf et al. 2007). This indicates that the specific characteristics of retrotransposed elements in imprinted genes are not a feature that efficiently distinguishes imprinted genes from bi-allelically expressed genes.

SPECIAL FEATURES OF IMPRINTED CPG ISLANDS AND DIFFERENTIALLY METHYLATED REGIONS

Allele-specific gene regulation in imprinted regions is mediated by a few DMRs. These elements often contain direct repeats and tend to be CpG rich, i.e. they frequently overlap with CpG islands. Though CpG islands are key elements in epigenetic gene regulation and, therefore, are believed to be strictly conserved, the CpG islands of imprinted genes show some species-specific features. Due to the depletion of SINE elements, imprinted genes in human possess less CpG islands than randomly selected genes (Hutter et al. 2006). A similar effect is not seen in the mouse, since in this species, repetitive elements rarely overlap with CpG islands. In the mouse, many imprinted genes possess intronic CpG islands that may serve as promoters of antisense transcripts, whereas a similar enrichment does not reach statistical significance in the human.

Nevertheless, in both species, CpG islands of imprinted genes are enriched in direct repeats. Interestingly, the repeated motifs are highly divergent, in a way that the CpG islands or DMRs of different genes contain different repeated motifs (Hutter et al. 2006). In addition, they are not conserved in orthologous CpG islands or DMRs in different species (Paulsen et al. 2005). Hence, the presence of direct repeats in the CpG islands and DMRs of imprinted region represents a conservation rather of DNA structure than of DNA sequence. Though the DNA sequences of DMRs are obviously not highly conserved, the presence of repeats seems to be sufficient for conserved regulatory functions of these elements. It has been shown for several DMRs that the allele-specific patterns of DNA methylation are conserved in several species. Targeted deletions of DMRs in mouse and deletions of DMRs in human patients suffering imprinting disorders show that these elements fulfill conserved regulatory functions in both species, and for some DMRs, it

has been shown that their functions (for example, as promoters of non-coding transcripts) are conserved between mouse and human (Lee et al. 1999, Fitzpatrick et al. 2002).

TISSUE-SPECIFIC EXPRESSION PATTERNS OF IMPRINTED GENES INDICATE FUNCTIONAL DIVERGENCE

The evolution of imprinting in therians is believed to be related to the evolution of the placenta as a permeable interface between the embryo and its mother, and imprinted genes are supposed to function predominantly in regulation of embryonic growth and nutrient supply during prenatal development. However, some imprinted genes are only weakly expressed in the placenta or are expressed at pronounced levels in other organs during postnatal stages. In order to address conservation of tissue-specific expression patterns of imprinted genes, we have evaluated tissue-specific micro-array expression data of these genes in human and mouse, thereby showing that strong inter-species conservation of tissue-specific expression is restricted to few imprinted genes and few tissues (Steinhoff et al. 2009). Among these are organs such as adrenal gland, pancreas and pituitary that are involved in endocrinal functions. This suggests that a major evolutionary conserved function of these genes is indeed in regulating growth and nutrient uptake. The identification of conserved binding sites for tissue-specific transcription factors in the promoter regions of genes that are expressed in the corresponding tissues highlights the usefulness of comparative approaches for the detection of tissue-specific regulatory elements.

The pronounced expression of imprinted genes at postnatal stages suggests that postnatal functional aspects of imprinted gene expression will be an interesting field for future research. In addition, the unexpected evolutionary divergence in expression patterns of imprinted genes indicates that these genes are capable of establishing species-specific functions. Hence, future comparative investigations on imprinted genes might also require additional mammalian species as alternative model systems.

REFERENCES

Bailey, J.A., l. Carrel, A. Chakravarti, E. E. Eichler. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci U S A* **97**: 6634.

Bernstein, B.E., E.L. Humphrey, C.L. Liu, and S.L. Schreiber. 2004. The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Methods Enzymol* **376**: 349.

Bestor, T.H. 2000. The DNA methyltransferases of mammals. *Hum Mol Genet* **9**: 2395.

Dean, W., F. Santos, M. Stojkovic, V. Zakhartchenko, J. Walter, E. Wolf, and W. Reik. 2001. Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos**.** *Proc Natl Acad Sci U S A* **98**: 13734.

Edwards, C.A., A.J. Mungall, L. Matthews, E. Ryder, D.J. Gray, A.J. Pask, G. Shaw, J.A. Graves, J. Rogers, I. Dunham, M.B. Renfree, and A.C. Ferguson-Smith. 2008. The evolution of the DLK1-DIO3 imprinted domain in mammals. *PLoS Biol* **6**: e135.

Fitzpatrick G.V., P.D. Soloway, M.J. Higgins. 2002. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet.* **32**:426.

Gebert, C., C. Wrenzycki, D. Herrmann, D. Groger, R. Reinhardt, P. Hajkova, A. Lucas-Hahn, J. Carnwath, H. Lehrach, and H. Niemann. 2006. The bovine IGF2 gene is differentially methylated in oocyte and sperm DNA. *Genomics* **88**: 222.

Greally, J.M. 2002. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci U S A* **99**: 327.

Hemberger, M., W. Dean, and W. Reik. 2009. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat Rev Mol Cell Biol* **10**: 526.

Hutter, B., V. Helms, and M. Paulsen. 2006. Tandem repeats in the CpG islands of imprinted genes. *Genomics* **88**: 323.

Hutter, B., M. Paulsen, and V. Helms. 2009. Identifying CpG Islands by Different Computational Techniques. *Omics*.

Jenuwein, T. and C.D. Allis. 2001. Translating the histone code. *Science* **293**: 1074.

Killian, J.K., C.M. Nolan, N. Stewart, B.L. Munday, N.A. Andersen, S. Nicol, and R.L. Jirtle. 2001. Monotreme IGF2 expression and ancestral origin of genomic imprinting. *J Exp Zool* **291**: 205.

Lee M.P., M.R. DeBaun, K. Mitsuya, H.L. Galonek, S. Brandenburg, M. Oshimura, and A.P. Feinberg. 1999. Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc Natl Acad Sci U S A* **96**:5203.

Luedi, P.P., A.J. Hartemink, and R.L. Jirtle. 2005. Genome-wide prediction of imprinted murine genes. *Genome Res* **15**: 875.

Luedi, P.P., F.S. Dietrich, J.R. Weidman, J.M. Bosko, R.L. Jirtle, and A.J. Hartemink. 2007. Computational and experimental identification of novel human imprinted genes. *Genome Res* **17**: 1723.

Morison, I.M. and A.E. Reeve. 1998. A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Hum Mol Genet* **7**: 1599.

Nikaido, I., C. Saito, Y. Mizuno, M. Meguro, H. Bono, M. Kadomura, T. Kono, G.A. Morris, P.A. Lyons, M. Oshimura, Y. Hayashizaki, and Y. Okazaki. 2003. Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. *Genome Res* **13**: 1402.

Paulsen, M., S. Takada, N.A. Youngson, M. Benchaib, C. Charlier, K. Segers, M. Georges, and A.C. Ferguson-Smith. 2001. Comparative sequence analysis of the imprinted Dlk1-Gtl2 locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the Igf2-H19 region. *Genome Res* **11**: 2085.

Paulsen, M., T. Khare, C. Burgard, S. Tierling, and J. Walter. 2005. Evolution of the Beckwith-Wiedemann syndrome region in vertebrates. *Genome Res* **15**: 146.

Roh, T.Y., S. Cuddapah, and K. Zhao. 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* **19**: 542.

Ruf, N., S. Bahring, D. Galetzka, G. Pliushch, F.C. Luft, P. Nurnberg, T. Haaf, G. Kelsey, and U. Zechner. 2007. Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. *Hum Mol Genet* **16**: 2591.

Smits, G., A.J. Mungall, S. Griffiths-Jones, P. Smith, D. Beury, L. Matthews, J. Rogers, A.J. Pask, G. Shaw, J.L. VandeBerg, and J.R. McCarrey, SAVOIR Consortium, M.B. Renfree, W. Reik, I. Dunham. 2008. Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. *Nat Genet*. **40**:971.

Steinhoff, C., M. Paulsen, S. Kielbasa, J. Walter, and M. Vingron. 2009. Expression profile and transcription factor binding site exploration of imprinted genes in human and mouse. *BMC Genomics* **10**: 144.

Walter, J., B. Hutter, T. Khare and M. Paulsen. 2006. Repetitive elements in imprinted genes. *Cytogenetic and Genome Research* **113**: 109.

Weber, M., J.J. Davies, D. Wittig, E.J. Oakeley, M. Haase, W.L. Lam, and D. Schubeler. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37**: 853.

Yokomine, T., K. Hata, M. Tsudzuki, and H. Sasaki. 2006. Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. Cytogenet *Genome Res* **113**: 75.

Zhang, Z., N. Carriero, and M. Gerstein. 2004. Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends Genet* **20**: 62.

Zhang, Y., C. Rohde, S. Tierling, H. Stamerjohanns, R. Reinhardt, J. Walter, and A. Jeltsch. 2009. DNA methylation analysis by bisulfite conversion, cloning, and sequencing of individual clones. *Methods Mol Biol* **507**: 177.