# Chapter VI – Elucidating the ancestral imprinting mechanism at IC1 in marsupials

## *6.1 Introduction*

### 6.1.1 Aims of this chapter

This chapter describes the systematic searches for each of the functional elements involved in the imprinting mechanism in the Imprinting Centre 1 (IC1) domain of wallaby. The results presented were obtained from a close collaboration between me and Guillaume Smits (Babraham Institute). This division of labour and the accompanying exchange of ideas greatly contributed to the generation of a more complete picture as described below. Specifically Guillaume performed the reverse transcription polymerase chain reaction (RT-PCR) amplification of wallaby samples and subsequent RACE experiments to determine the wallaby *H19* gene structure. SNPs identified by Guillaume in the wallaby H19 and *IGF2* genes were used by him to determine imprinting status. The characterisation of the wallaby *H19* differentially methylated region in a variety of tissues was also performed by Guillaume. All other experiments and bioinformatic analyses presented were performed by me.

To further our knowledge of the origin of genomic imprinting and the imprinting mechanism in ancestral mammals I have focused on the *H19/IGF2* or IC1 locus, located towards the telomere of the short arm of human chromosome 11 (11p15.5)

and mouse distal chromosome 7 (7qF5). The murine *Igf2* and *H19* genes were amongst the first imprinted genes to be identified (Bartolomei et al. 1991, DeChiara et al. 1991) and as a consequence the IC1 locus has been well characterized in eutherian mammals. In recent years there has also been a growing body of literature regarding the imprinting status of marsupial, monotreme and bird *Igf2* orthologues (Killian et al. 2001, O'Neill et al. 2000, Suzuki et al. 2005). The *Igf2* gene has been shown to be imprinted in the marsupial species *Monodelphis domestica* (grey short-tailed opossum, O'Neill et al. 2000) and *Macropus eugenii* (tammar wallaby, Suzuki et al. 2005) but biallelically expressed in the monotreme species *Ornithorhynchus anatinus* (duck-billed platypus) and *Tachyglossus aculeatus* (short-beaked echidna) as well as the birds such as *Gallus gallus* (chicken, Killian et al. 2001, O'Neill et al. 2000, Weidman et al. 2004). In contrast, the expression status of the long, non-coding RNA (hereafter termed macroRNA) *H19* gene in ancestral vertebrate species has not been elucidated as this non-coding RNA gene has yet to be identified in any non-eutherian species (Paulsen et al. 2005, Yokomine et al. 2005).

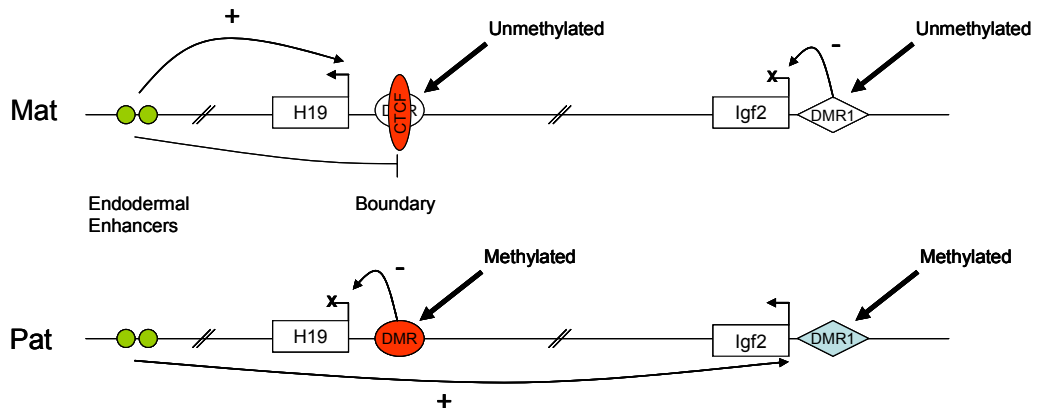


**Figure VI.1. Boundary model of *Igf2/H19* gene regulation.**

**On the maternal allele the lack of methylation at a differentially methylated region (DMR) permits the binding of the insulator protein CTCF which acts as an enhancer blocker, preventing access of downstream endodermal enhancers to the *Igf2* promoter region. These enhancers are therefore able to drive expression of *H19*, a non-coding RNA of unknown function. On the paternal allele the DMR is methylated preventing CTCF binding and therefore permitting access of the enhancers to the *Igf2* promoter. The result is paternal expression of *Igf2*.**

Whilst the integrity of the functional elements controlling *H19* transcription is essential for the correct imprinting of *Igf2* (at least in mice) the *H19* RNA itself is not, as illustrated by experiments to generate transgenic mice lacking the *H19* transcript (Jones et al. 1998, Thorvaldsen et al. 2002). The balancing paternal expression of *Igf2* and maternal expression of *H19*, in eutherian species studied to date, has given rise to the accepted 'boundary model' of imprinting at the IC1 locus (Arney. 2003, Hark et al. 2000 and Figure VI.1). In all tissues studied, eutherian *Igf2* imprinting is dependent upon a differentially methylated region (DMR) located 2 kb 5' of the *H19* gene. This 2.3 kb element (in mouse) is an unmethylated insulator on the maternal chromosome (resulting in *Igf2* insulation and *H19* expression) and a methylated silencer on the paternal chromosome (resulting in *H19* promoter repression and *Igf2* expression, Figure VI.1, Bartolomei et al. 1991, Bell and

Felsenfeld. 2000, Drewell et al. 2000). Our understanding of tissue specific expression at the IC1 locus was enhanced by the identification, in mouse foetal and neonatal liver, of two enhancers located 9 and 11 kb downstream of *H19* shown to be crucial for endodermal expression (Leighton et al. 1995). Recent capturing chromosome conformation (3C, Dekker et al. 2002) studies in liver demonstrated that the maternal *Igf2* allele was insulated from the endodermal enhancers by an epigenetic switch. The formation of chromatin loops caused by the binding of CTCF (CCCTC binding factor) protein to the unmethylated *H19* DMR results in the *Igf2* promoter residing in an inactive chromatin domain, insulated from the endodermal enhancers which lie with *H19* in an active chromatin domain (Kurukuti et al. 2006, Murrell et al. 2004). CTCF is a highly evolutionary conserved protein with 11-zinc finger domains and was found to bind to the prototypical chicken insulator in the β-globin locus (Bell et al. 1999). The precise means by which CTCF binding confers chromatin insulation and therefore segregates distinct regulatory domains is unclear. However, recent studies have shown that other proteins, such as nucleophosmin and CHD8 may bind to CTCF to generate a protein complex tethered to the nuclear matrix (Ishihara et al. 2006, Yusufzai and Felsenfeld. 2004). Binding to the nuclear matrix is thought to result in chromatin loops that may partition regulatory elements and thus prevent their interaction. Such a model in the IC1 locus was proposed by Murrell and colleagues (Murrell et al. 2004).

To gain further insights into the very early stages of genomic imprinting evolution and the gene regulatory mechanisms operating to bring about this unusual phenomenon it is important to study extant ancestral mammalian species in which genomic imprinting as been demonstrated. In order to establish the mechanism of imprinting in marsupial species it is therefore of critical importance to look for and

characterize the *H19* non-coding RNA and its regulatory elements in these species. Despite a report in 1996 claiming the fluorescence *in situ* hybridisation mapping of a phage clone containing tammar wallaby *H19* to chromosome 2p (Toder et al. 1996), sequence of this clone was not obtained (Jenny Marshall-Graves, personal communication). The recent whole genome shotgun draft assembly of the grey, short-tailed opossum also failed to reveal the presence of *H19*. To determine whether *H19* does exist in marsupial species BACs mapping to the IC1 orthologous region were sequenced in both tammar wallaby and opossum (chapters III and IV).

As discussed in chapter V there is a striking conservation of both the order of the ECRs and the physical size of the inter ECR regions, particularly between human and wallaby, which permits the ECRs to serve as landmarks between the *Lsp1* gene and the 3' region of *Igf2* (Figure VI.2). In human and mouse, the region between ECR14 and ECR15 (39059 and 31162 bp, respectively) contains two endodermal enhancers (EE1 and EE2, Leighton et al. 1995, Yoo-Warren et al. 1988), the *H19* gene and the differentially methylated region (DMR) which are critical for the reciprocal imprinting of *Igf2* and *H19* genes (Figure VI.2). Unexpectedly, these features were not found by standard BLAST analyses in the finished orthologous wallaby ECR14-15 region sequence which is very similar in length (38186 bp) to the human interval. In addition these features were not readily identifiable in the whole wallaby BAC sequence (174698 bp, CR855994). The wallaby ECR14-15 region was therefore studied in fine detail for the presence of *H19* and associated functional elements.

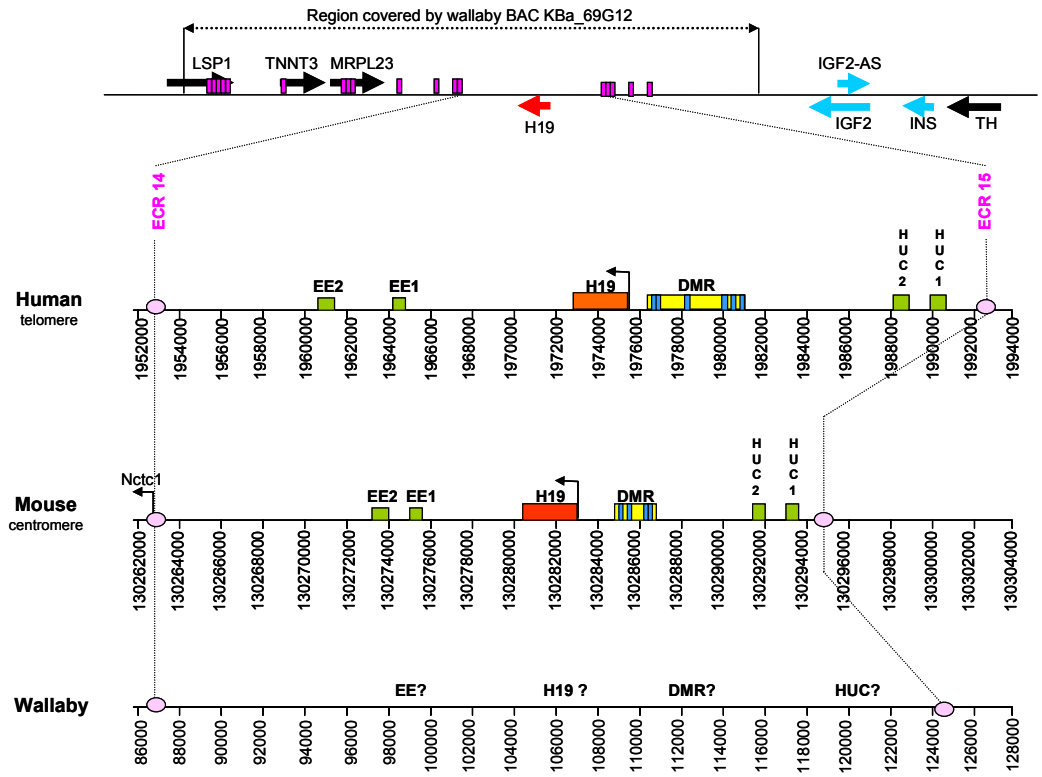


**Figure VI.2. Sequence analysis in the *H19* region.**

**The human region of 11p15.5 encompassing the IC1 locus is presented from telomeric (left) to centromeric (right) end with the genes encoded on the sense strand above the line and those encoded on the antisense strand below the line. The genes depicted by black arrows have not been shown to be imprinted in any species (*LSP1*, *TNNT3*, *MRPL23*, *TH*), the genes depicted by blue arrows (*IGF2*, *IGF2-AS*, *INS*) are imprinted and paternally expressed (at least in some tissues). The maternally expressed *H19* gene is shown in red. The protein coding genes *LSP1*, *TNNT3* and *MRPL23* are all present in the wallaby BAC ME_KBa69G12. ECRs conserved between human, mouse and wallaby are illustrated by pink rectangles. The ECRs 14 and 15 define a region of approximately 40 kb in the wallaby BAC sequence where the *H19* gene and regulatory elements reside in human and mouse.**

## *6.2 Identifying wallaby H19 and establishing its imprinting status*

BLAST 2 sequences (BL2SEQ) analysis was performed at NCBI (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi) using high sensitivity (low specificity) parameters (word size of 10, match and mismatch penalties of +3 and -2

respectively) between the 2.3 kb human *H19* gene sequence (NR_002196) and the 38 kb wallaby sequence between ECRs14 and 15. A 49% match with 13% gaps was observed in the anticipated orthologous region of *H19*. Guillaume Smits (Babraham Institute) subsequently elucidated the gene structure and determined the imprinting status of wallaby *H19* (and *IGF2*). Briefly, RT-PCR primers were designed within a region 1.5 kb 5' to 600 bp 3' of the *H19* candidate sequence. These primers were then used to screen two pouch young wallaby liver samples kindly supplied by Marilyn Renfree (Melbourne). Sequencing of the amplicons and subsequent 5' RACE experiments revealed the structure of the 2.7 kb wallaby *H19* gene (Figure VI.3).

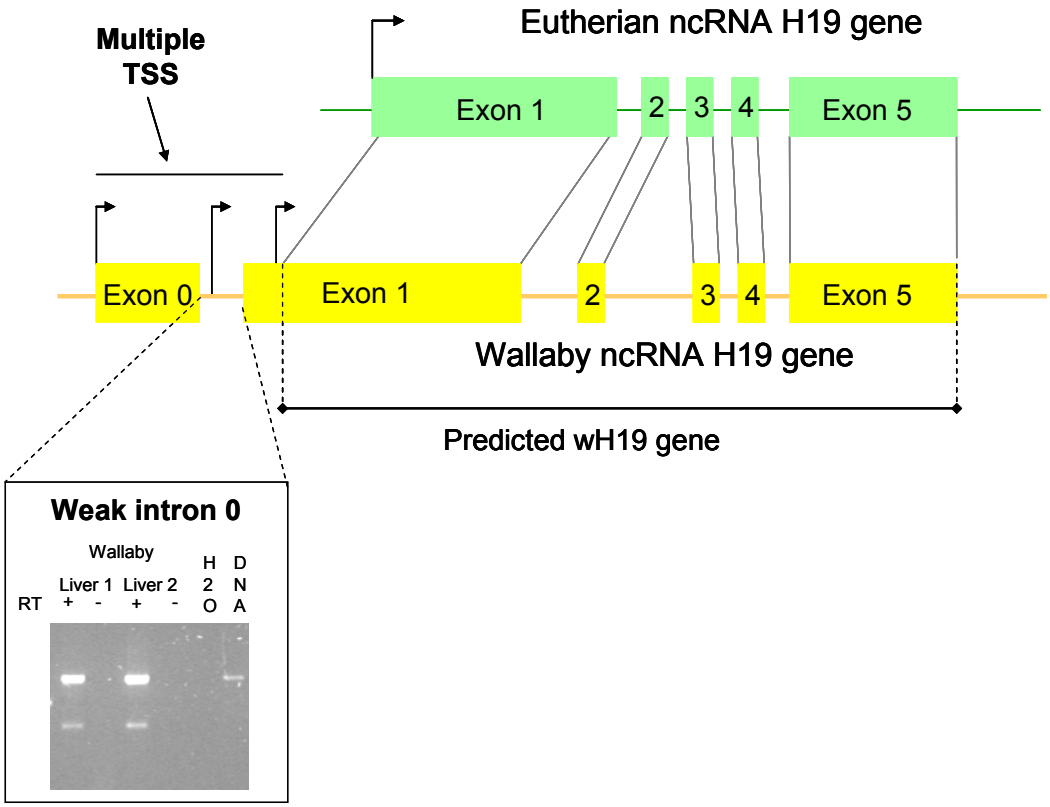


**Figure VI.3. Elucidated structure of the wallaby *H19* gene.**

**Weak amplification products visible upon gel electrophoresis in both wallaby pouch young liver samples following Reverse Transcriptase (RT) PCR indicates the presence of an additional intron (intron 0) in the wallaby *H19* (wH19) gene. TSS, transcription start site; ncRNA, non-coding RNA.**

Despite 148 Myr of parallel evolution the exon/intron structure (but not the sequence) of this non-coding RNA gene is remarkably well conserved between wallaby and eutherians with one notable exception; In wallaby a supplementary 5' exon (exon 0) is separated from exon 1 by a weak intron (intron 0, Figure VI.3). The wallaby *H19* gene has a major transcription start site (TSS) at the beginning of exon 0 and a series of apparently minor TSSs throughout exon 0, intron 0 and the 5' region of exon 1, consistent with the recent definition of a broad promoter (Sandelin et al. 2007).

Further wallaby tissue samples were received from Marilyn Renfree and used to study the expression of wallaby *H19*. Expression was observed in the marsupial placenta (specifically in bilaminar and trilaminar yolk sac), foetal liver, pouch young liver and brain (Guillaume Smits, personal communication). SNPs in the wallaby *H19* and *IGF2* genes were used to determine whether these genes were imprinted in the tissues investigated, and if so, their parent-of-origin. As expected, the *H19* and *IGF2* genes were expressed from the maternal and paternal alleles respectively (Guillaume Smits, personal communication). In every tissue examined, imprinting (maternal expression) of *H19* was seen to be "complete" upon cDNA sequencing (Figure VI.4). In contrast, paternal expression of *IGF2* was complete in liver but "leaky" in brain and placenta (Suzuki et al. 2005). The eutherian (human and mouse) chorionic epithelium (brain) shows bi-allelic expression of *IGF2* and therefore the finding of leaky imprinted expression of wallaby *IGF2* in brain is not entirely surprising. However, eutherian imprinted *IGF2* expression in placenta and yolk sac is complete, at odds with the findings reported here of leaky wallaby *IGF2* expression. One explanation for this observation is that two cell populations (one imprinted and another with bi-allelic expression of *IGF2*) exist in the wallaby placenta, reminiscent of the eutherian brain. The difference in *IGF2* imprinting between wallaby liver and placenta parallels that observed in the eutherian *Kcnq1ot1* locus (Reik and Lewis. 2005). Leaky wallaby *IGF2* expression in the placenta is indicative that the therian placenta was likely not the major driving force for the emergence of *IGF2* imprinting. However, the possibility that wallaby has lost the need for complete *IGF2* imprinting cannot be excluded.

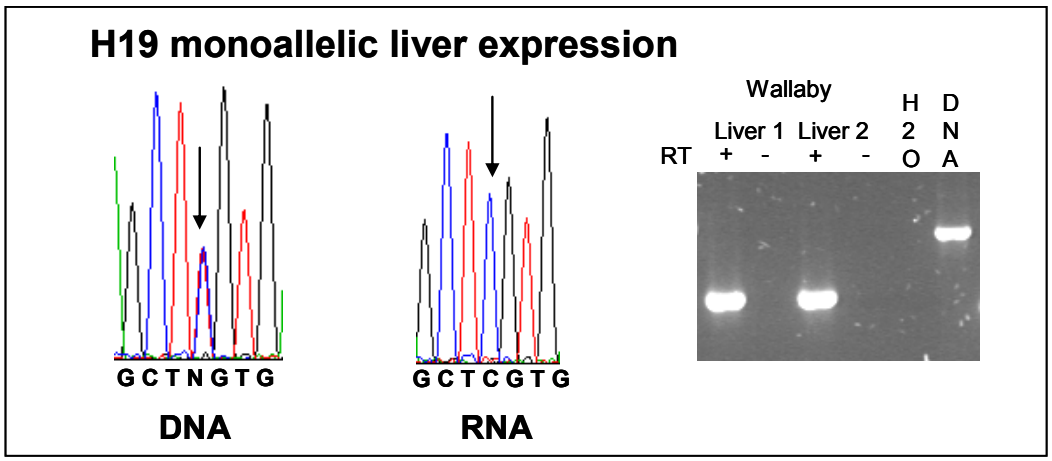


**Figure VI.4. Expression of wallaby *H19*.**

**RT-PCR amplification from two pouch young liver samples is shown (right). A C/T SNP in genomic DNA (DNA) was used to show 'complete' monoallelic expression of *H19*. The cytosine base marked by the arrow in the RNA corresponds to the maternal allele.**

## *6.3 Identifying wallaby micro RNA (miR-675) within exon 1 of the H19 gene*

MicroRNAs (miRNAs) are short (18-25 nucleotides (nt)) non-coding RNA sequences that have been shown to regulate a wide range of biological processes including vertebrate development (reviewed in Zhao and Srivastava. 2007). Cellular miRNAs are initially expressed as part of one arm of an approximately 80-nt RNA hairpin derived from a longer primary miRNA transcript transcribed by RNA polymerase II (Cullen. 2004). Primary miRNA hairpins have a characteristic structure, consisting of an approximately 30-base-pair (bp) imperfect stem, a large terminal loop, and flanking unstructured RNA sequences, that is both necessary and sufficient for recognition by the nuclear RNase III enzyme Drosha acting in concert with its cofactor DGCR8 (Han et al. 2004, Zeng et al. 2005). Drosha then cleaves approximately 22 bp down the stem to excise the approximately 60-nt pre-miRNA intermediate. The cleavage of the primary miRNA by Drosha processing results in

at least three RNA fragments and therefore miRNAs are rarely found within the exons of coding transcripts. Instead, miRNA hairpins are generally located in introns or in the exons of non-coding RNAs (Cullen. 2004). After nuclear export of the pre-miRNA intermediate, a second RNase III enzyme called Dicer cleaves at the stem-loop junction to generate the mature miRNA (Hutvagner et al. 2001, Ketting et al. 2001). The mature miRNA is then incorporated into the RNA-induced silencing complex (RISC), where it acts as a guide RNA to direct RISC to complementary mRNA species for degradation or translational inhibition (Hammond et al. 2000, Martinez et al. 2002, Schwarz et al. 2002).

Mineno and colleagues (Mineno et al. 2006) used massively parallel signature sequencing (Brenner et al. 2000) to identify novel miRNA species from whole embryos of mouse. A total of 390 miRNAs were identified using this approach, of which 195 (50%) were previously known. The novel miRNAs were deposited into the public miRBase (Griffiths-Jones et al. 2006). One of these novel murine miRNAs (miR-675) was independently identified and further characterized in mouse and human by Cai and Cullen (Cai and Cullen. 2007). In mouse there are two mature 22 nt miRNAs lying within exon 1 of the *H19* gene (Positions chr7:142386502-142386523 and 142386468-142386489 in NCBI build 36 of mouse). A single mature 23 nt human miRNA lies at position chr11:1974606-1974628 in NCBI build 36 and position 1014-1036 of human *H19* RNA (NR_002196.1) The *H19* gene itself was therefore demonstrated to be the primary miRNA transcript in humans and mice and sheds some light on the elusive function of *H19*.


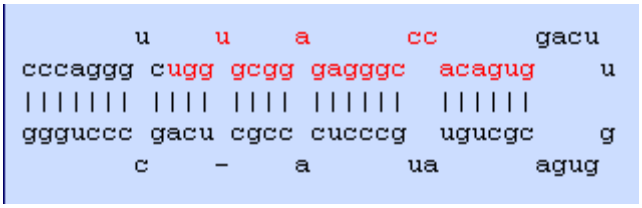


**Figure VI.5. *Mus musculus* (top) and *Homo sapiens* (bottom) miR-675 stem-loop sequences.**

**As illustrated in miRBase. Mature miR-675 sequences are highlighted in red.**

The approximately 80 nt miR-675 precursor is highly conserved between human and mouse (Figure VI.5). Alignments of derived eutherian miR-675 sequences were performed by Sam Griffiths-Jones (University of Manchester). The most highly conserved sequences from this alignment were used to search the wallaby BAC sequence (CR855994) using the 'DNA analysis' feature of ACeDB. This analysis revealed the presence of a candidate wallaby miR-675 lying on the reverse strand at position 109093-109178 of the BAC sequence (CR855994). This location is consistent with that of the wallaby *H19* gene (Figure VI.2).

## *6.4 Opossum H19 and miR-675*

In order to assess whether the findings above are specific to wallaby or can be applied more generally to marsupial mammals BACs representing the IC1 domain of opossum were mapped and sequenced (see chapters III and IV). The unfinished consensus sequence for BAC VMRC18-490C6 was exported from a gap4 database and a BLASTN database created (chapter II). This database was searched with 2260

bp of the wallaby *H19* sequence which identified an approximately 70% nucleotide match to a short (1610 bp) opossum sequence contig. Next the same opossum BAC sequence was searched with the wallaby miR-675 sequence which revealed a 95% (82/86 bp) identity. Predicted mature miRNA sequences were identical in the two marsupial species. The eutherian and marsupial alignments of miR-675 are shown in Figure VI.6.

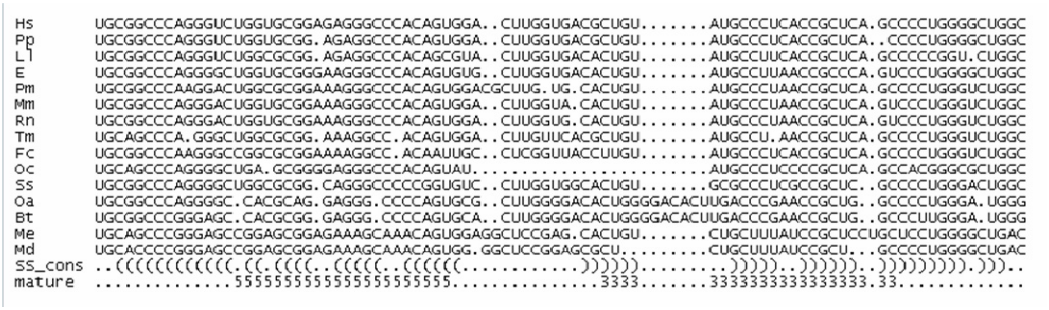


**Figure VI.6. Sequence alignment of therian miR-675 sequences.**

**Sequence alignment produced by Sam Griffiths-Jones using the STOCKHOLM 1.0 software package. The *Monodelphis domestica* derived miR-675 sequence from BAC VMRC18-490C6 sequence was manually aligned. Hs, *Homo sapiens* (human); Pp, *Pongo pygmaeus* (orangutan); Ll, *Lynx lynx* (Eurasian lynx); E, *Elephantidae* (elephant); Pm, *Peromyscus maniculatus* (deer mouse); Mm, *Mus musculus* (house mouse); Rn, *Rattus norvegicus* (Norway rat); Tm, *Thomomys monticola* (mountain pocket gopher); Fc, *Felis catus* (domestic cat); Oc, *Oryctolagus cuniculus* (domestic rabbit); Ss, *Sus scrofa* (wild boar); Oa, *Ovis aries* (domestic sheep); Bt, *Bos taurus* (domestic cow); Me, *Macropus eugenii* (tammar wallaby); Md, *Monodelphis domestica* (opossum).**

## *6.5 Identification and characterization of the H19 differentially methylated region in wallaby*

An analysis of the CpG content of the wallaby *H19* gene and flanking sequences demonstrated a high density of CpG dinucleotides extending from 4 kb 5' to exon 0

to the 3' end of exon 1 (Figure VI.7). Differential methylation of this region was confirmed in liver, yolk sac, brain and primary ear fibroblasts by Guillaume Smits and colleagues (Babraham Institute, Cambridge). The region was fully methylated in testis and therefore a paternal germ-line DMR as in eutherians (Olek and Walter. 1997, Tremblay et al. 1997).

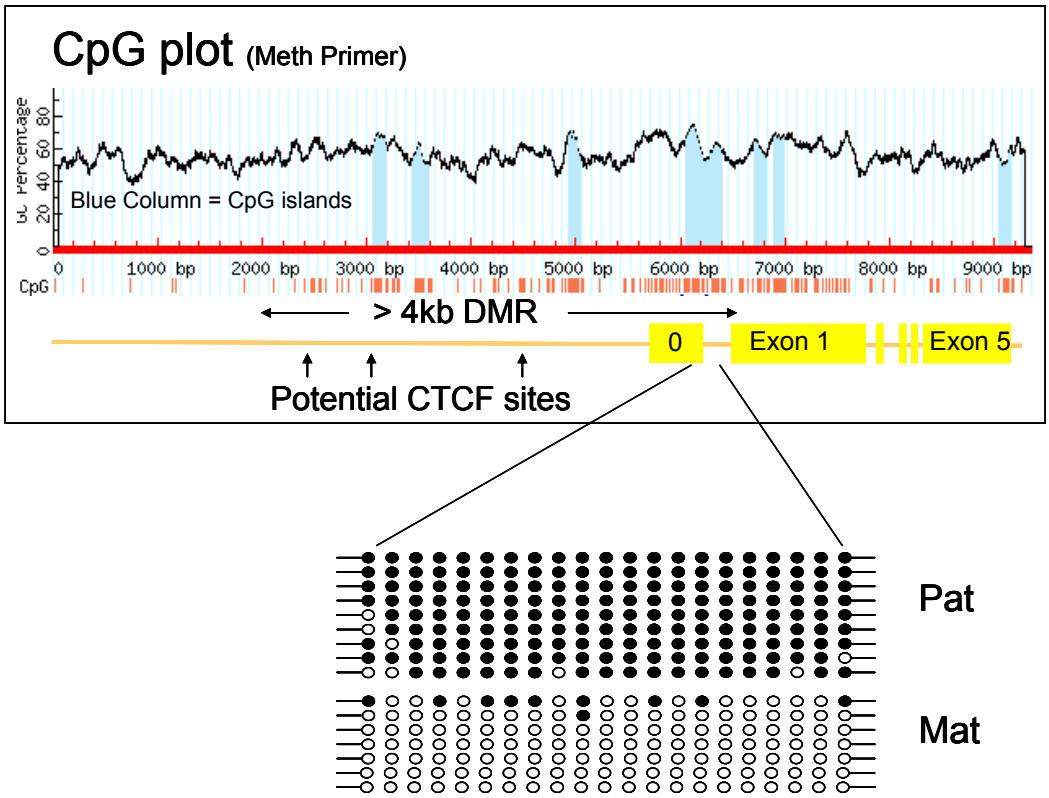


**Figure VI.7. Identification of the wallaby *H19* DMR.**

**The C+G content of the wallaby *H19* gene and upstream region is depicted on the graph. Blue shading beneath the C+G plot indicates CpG islands. Individual CpG dinucleotides are shown by red vertical bars. Bisulphite sequencing of TA clones from the CpG island overlapping with exon 0 reveal differential methylation of this region. Filled circles, methylated; open circles unmethylated. Pat, paternal allele; Mat, maternal allele.**

To address whether the wallaby DMR might function as an insulator, as is the case in eutherian mammals, I searched for potential CTCF binding sites. The 14 bp

consensus sequence CCGCGNGGNGGCAG (Bell and Felsenfeld. 2000) was searched against the wallaby BAC sequence CR855994 in ACeDB. Allowing for two mismatches a total of 25 matches was obtained including a cluster of 3 lying within the previously defined DMR, positions -4.2, -3.4 and -2.1 kb from the major H19 TSS. These putative CTCF binding sites lay within a 48 bp sequence motif (Figure VI.8 and Kim et al. 2007, Xie et al. 2007).

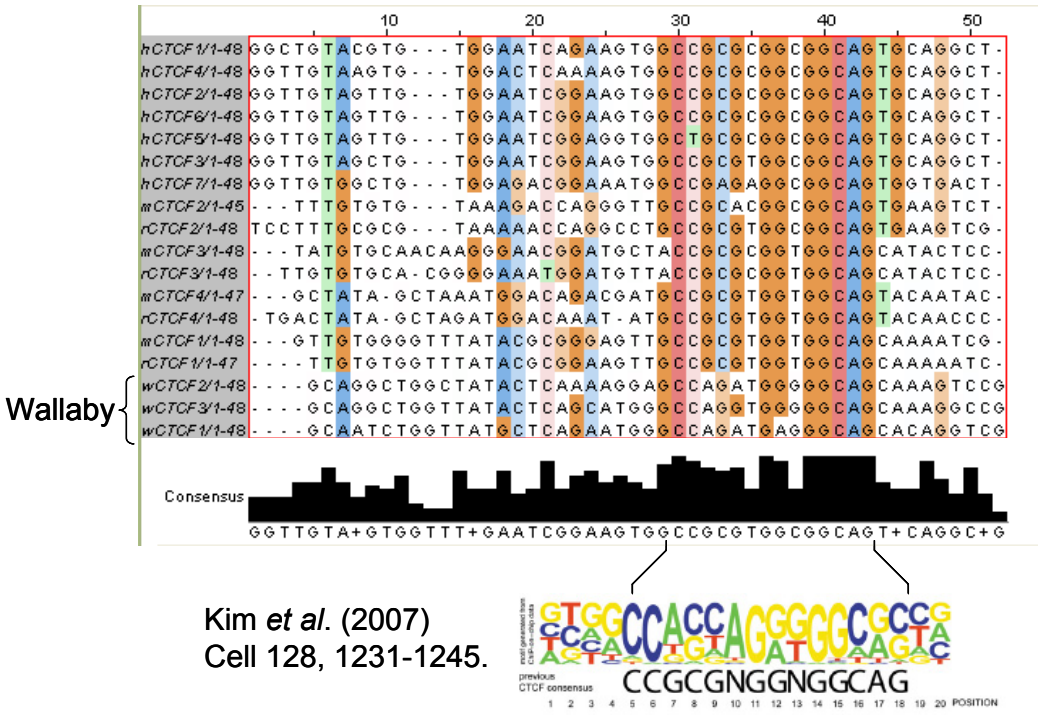


**Figure VI.8. Identifying potential CTCF binding sites in wallaby.**

**ClustalW (version 1.83) sequence alignment of human CTCF (hCTCF), mouse CTCF (mCTCF), rat CTCF (rCTCF) and wallaby CTCF (wCTCF) binding sites in the *H19* DMR region. Alignments are viewed using JalView software** (Clamp et al. 2004)**. A conserved 20 bp CTCF binding site logo is reproduced from** (Kim et al. 2007)**. Under the logo is the original 14 bp consensus** (Bell and Felsenfeld. 2000)**.**

## *6.6 Testing the wallaby DMR for insulator barrier activity*

Chromatin insulators, also known as boundary elements, partition the genome into independent chromosomal domains to control individual gene expression. The boundaries produced by these insulators have the ability to block the interaction between enhancer(s) and promoter when positioned between them (enhancer-blocking activity). They also have the ability to block repressive chromatin effects on flanking regions (barrier activity, Bell et al. 2001, Gaszner and Felsenfeld. 2006, West et al. 2002). Although these properties of insulation could be separate, the precise mechanisms remain to be elucidated. Most known vertebrate insulators have been found to contain unique binding sites for the protein CTCF, which is essential for the insulation activity, especially the enhancer-blocking function (Ohlsson et al. 2001). CTCF-dependent insulators have been particularly well characterized in the chicken β-globin locus (Bell et al. 1999) and in the imprinted *IGF2/H19* locus in mouse and human (Bell and Felsenfeld. 2000, Hark et al. 2000).

To assess whether the wallaby *H19* DMR can function as an insulator the transient transfection vector system recently developed by Ishihara and colleagues (Ishihara et al. 2006) was used. This vector ('pIHLE', Figure VI.9) contains the firefly luciferase reporter gene under the control of the mouse *H19* promoter (-818 to +6 relative to the *H19* transcription start site) and stimulated by a downstream SV40 enhancer. Fragments to be tested were cloned into the *Bam*HI site between the luciferase reporter gene and enhancer (Figure VI.9). If a DNA fragment tested has enhancer blocking activity then it will block the interaction between the SV40 enhancer and *H19* promoter and therefore a reduction in luciferase gene expression will result. In contrast cloned fragments with no insulator function should have no effect on luciferase expression.

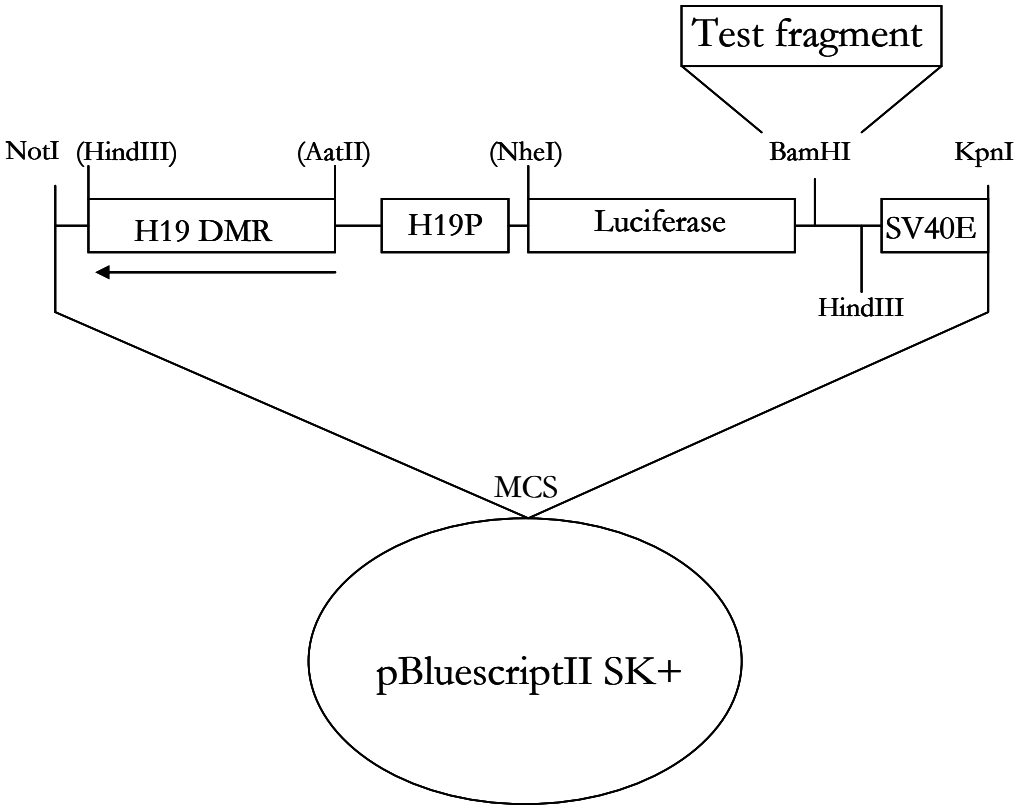


**Figure VI.9. Testing for insulator function.**

**The pIHLE vector was a kind gift of Ko Ishihara and Mitsuyoshi Nakao. *H19* DMR, mouse 1.8 kb insulator; H19P, mouse *H19* promoter; Luciferase, firefly luciferase reporter gene from pGL3-Basic (Promega) vector; SV40E, SV40 enhancer; MCS, multiple cloning site. Test and control insulator sequences were cloned into the *Bam*HI site between luciferase reporter and SV40 enhancer sequences.**

Purified DNA from plasmid constructs were co-transfected with a renilla luciferase expressing plasmid (pRL-CMV, for internal normalisation) in human hepatocellular liver carcinoma cell line (HepG2) cells. HepG2 cells were used because both *IGF2* and *H19* genes are transcribed at high levels in liver. Indeed the name *H19* comes from the fact that it was clone number 19 from a foetal hepatic library. Luciferase levels were detected in a luminometer and insulator activity was indicated by a decrease in relative luciferase expression levels when compared with the pIHLE

vector containing random sequence insertions. The pIHLE vector containing the 1.8 kb mouse H19 DMR insulator (Ishihara et al. 2006) cloned into the insulator site showed a greater than 5-fold (>80%) reduction in luciferase expression (Figure VI.10), compared to the 60% reduction observed by Ishihara and colleagues (Ishihara et al. 2006). One possible explanation for the improvement in insulation in this study is that liver (HepG2) cells were transfected in contrast to HeLa cells used by Ishihara and co-workers. *H19* RNA is strongly expressed in HepG2 cells and therefore the *H19* promoter driving luciferase expression is in a more native context. The pIHLE vector containing the 2.3 kb candidate wallaby insulator sequence demonstrated a 40% reduction in luciferase expression (Figure VI.10). To demonstrate whether the apparent insulator effect could be the result of a position effect i.e. increasing the distance between the SV40 enhancer and *H19* promoter, random 2.3-2.4 kb wallaby fragments were also cloned into the *Bam*HI site of the pIHLE vector. Only a marginal decrease in relative luciferase expression was observed indicating that the 2.3 kb cloned DNA fragments have little or no effect on luciferase activity (Figure VI.10).

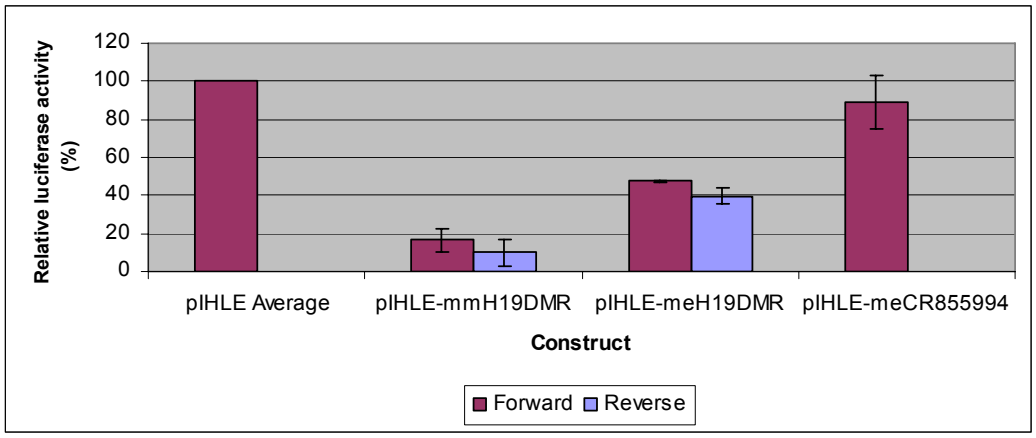


**Figure VI.10. Insulator activity of the wallaby *H19* DMR.**

**Luciferase levels were normalized against the pIHLE vector with no insert. The pIHLE-mmH19DMR vector contains the mouse *H19* DMR with known enhancer-blocker activity. The pIHLE-meH19DMR vector contains a 2.3 kb wallaby sequence encompassing 3 putative CTCF binding sites. The pIHLE-meCR855994 bar represents an average of two independent constructs containing different randomly cloned 2.3 kb PCR products amplified from wallaby BAC sequence CR855994. Mouse and wallaby *H19* DMR constructs were cloned in both forward (claret) and reverse (blue) orientations.**

## *6.7 Searching for wallaby endodermal enhancers*

The *IGF2* gene is imprinted (paternally expressed) in wallaby tissues; yolk sac placenta, foetal and pouch young livers (Guillaume Smits personal communication). We have also established the presence of wallaby *H19* and the upstream DMR which has CTCF sites and insulator (enhancer-blocker) activity analogous to the situation in human and mouse. According to the 'boundary model' the enhancer-blocking activity of the CTCF occupied *H19* DMR on the maternal allele prevents access of the endodermal enhancers to the *IGF2* promoter. So could endodermal enhancers be identified in wallaby? To address this question I designed 20, approximately 3 kb wallaby PCR amplicons, overlapping by 1 kb, from across the 38 kb wallaby region between ECR14 and ECR15 (Figure VI.2). The PCR amplified

products were Gateway® cloned, in both forward and reverse orientations, in the pGL3-promoter modified vector for transient transfection in human HepG2 cells (chapter II). To establish whether we might expect to observe wallaby sequences with enhancer activity in human cells (i.e. cross-species effects) it was important to include positive controls. A literature search in PubMed for "wallaby OR marsupial AND enhancer" returned 20 results, however, none reported the identification of wallaby endodermal enhancers. I therefore elected to use the known mouse endodermal enhancer 1 (EE1) element from the *H19* downstream region. In human HepG2 cells this mouse enhancer increased relative luciferase activity by approximately 16-fold (log2 of 4) over the empty Gateway® modified pGL3-Promoter construct (Figure VI.11). To control for size effects a 3 kb human tile containing both endodermal enhancers EE1 and EE2 (positive control) and a neighbouring, but non-overlapping, 3 kb tile (negative control) was cloned and transfected. The positive control shows that cloning relatively large (3 kb) fragments into an already large plasmid (approximately 8 kb) has little, if any, effect on detecting enhancer activities (Figure VI.11). Indeed there is an apparent additive effect of cloning EE1 and EE2 together compared with the enhancer activities of EE1 and EE2 cloned independently (chapter V). None of the 20 wallaby tiles showed enhancer activity in human HepG2 cells (Figure VI.11). Whether this is the result of sequence divergence between wallaby and human rendering the human gene regulatory machinery incapable of recognising wallaby enhancers is unknown. However, despite having last shared a common ancestor with mouse over 90 Myr ago, mouse EE1 behaves as a potent enhancer in human HepG2 cells with enhancer activity equal to that of the human EEs. Perhaps a more likely explanation for the lack of EE activity is that the enhancers are not located in this region of conserved synteny but lie in a more remote location.

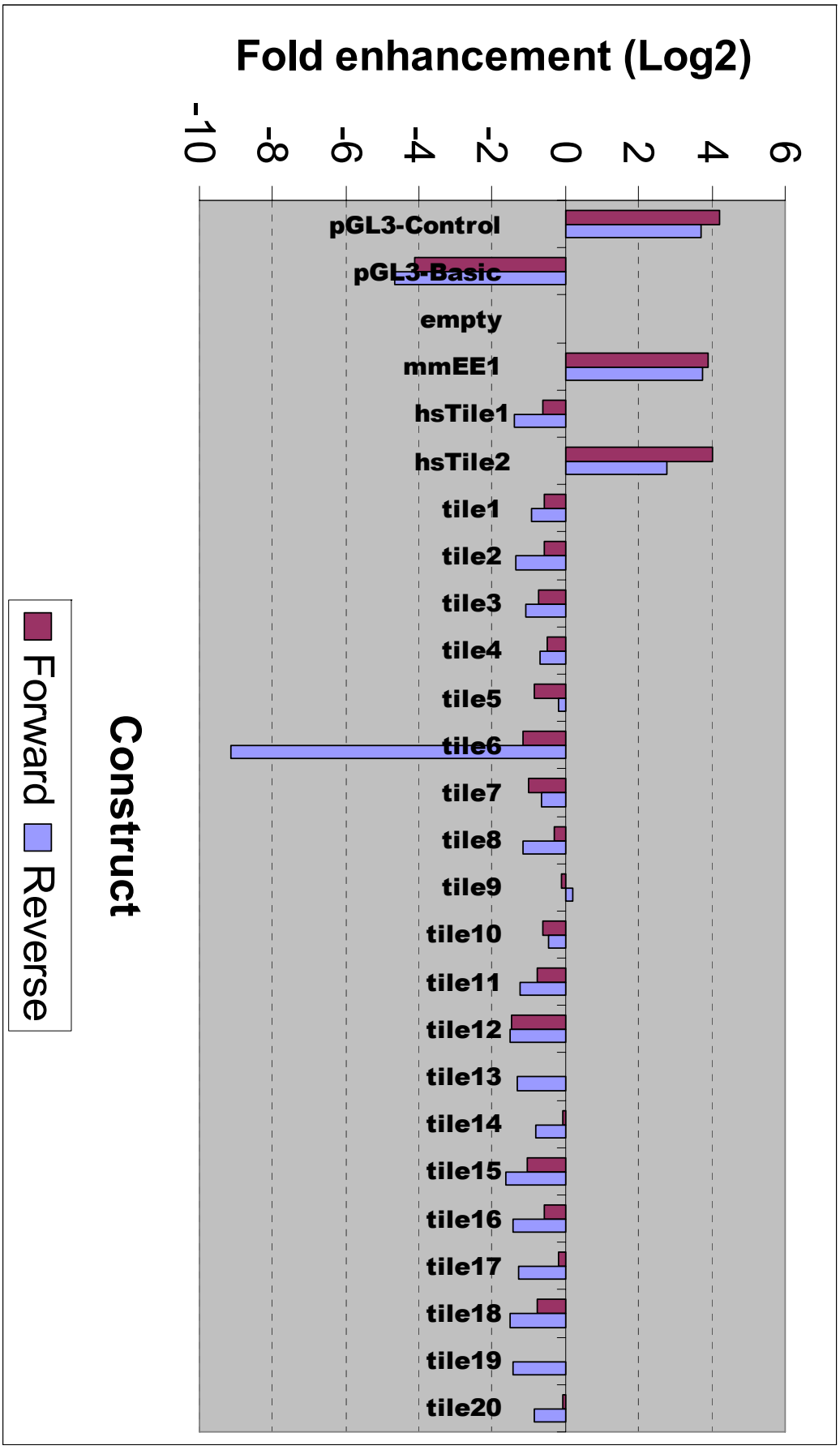
Fold enhancement (Log2)
-10
-8
-6
-4
-2
0
2
4
6
pGL3-Control
pGL3-Basic
empty
mmEE1
hsTile1
hsTile2
tile1
tile2
tile3
tile4
tile5
tile6
tile7
tile8
tile9
tile10
tile11
tile12
tile13
tile14
tile15
tile16
tile17
tile18
tile19
tile20
Construct
Forward
Reverse

**Figure VI.11. Testing wallaby tiles between ECRs 14 and 15 for enhancer activity in human HepG2 cells.**

**Human and wallaby 3 kb tiles were cloned into pGL3-Promoter vectors in both forward (claret) and reverse (blue) orientations. For each construct transfected the luciferase activities from four technical replicates were averaged and normalized against a construct in which no insert was cloned into the Gateway cassette (labeled 'empty' in the chart). The pGL3-Control vector contains both SV40 enhancer and promoter for optimal expression of firefly luciferase. The pGL3-Basic vector contains neither promoter nor enhancer elements. Empty, modified pGL3-Promoter vector containing the Gateway cassette. mmEE1, *Mus musculus* endodermal enhancer (EE) 1 element cloned into the Gateway cassette. hsTile1, 3 kb human fragment with no known enhancer elements (negative control). hsTile2, 3 kb human sequence containing both EE 1 and 2 (positive control). Tile1-20, wallaby 3 kb overlapping tiles from the ECR14-15 region. Fold enhancement over the empty construct is plotted on a log2 scale.**

## *6.8 Discussion*

Taken together the results presented in this chapter demonstrate that, with the possible exception of endodermal enhancers, all the major hallmarks of the eutherian *H19-IGF2* imprinting system are present in marsupials and must therefore have been present in the therian ancestor in existence prior to the divergence of marsupials and eutherians. One of the few apparent differences between the marsupial and eutherian IC1 domains is the absence of endodermal enhancers in the region of conserved synteny between ECR14 and ECR15. As discussed above this could be a sensitivity problem with the transient transfection assay in that the sequence divergence between wallaby and human gives a false negative result. Alternatively the enhancer position may be less well conserved between therians than, for example, the *H19* DMR. It should be noted that ECR11 and ECR12h, lying 40 kb telomeric of *H19* do show significant enhancer activity in HepG2 cells (chapter V) and therefore these endodermal enhancers could be enhancing wallaby *IGF2* expression on the paternal allele.

Every imprinted gene cluster characterized to date harbours at least one non-coding RNA and yet the function of many remains elusive (Pauler and Barlow. 2006). The processing of miR-675 from exon 1 of the *H19* pri-miRNA precursor RNA is evidently not highly efficient because full-length *H19* transcripts are abundant in developing embryos (Bartolomei et al. 1991). miRNAs themselves are subject to developmental regulation (Obernosterer et al. 2006, Thomson et al. 2006) and therefore it is possible that processing of miRNAs from *H19* varies both spatially (different tissue types) and temporally (different developmental stages). If the full-length *H19* transcript does have a function, which would seem likely given the

evolutionary conservation of the exon/intron structure, then it is possible that miR-675 serves to affect this function.

Kinship theory (Haig. 2004) predicts that maternally expressed genes such as *H19* will generally act to suppress excessive foetal growth stimulated by paternally expressed growth factors such as *IGF2*. It will therefore be interesting to see whether the post-transcriptional targets of miR-675 include growth factors, their receptors or downstream pathway members. A search for predicted microRNA targets for mouse miR-675-5p in the TargetScan (release 4.0, July 2007) database (Grimson et al. 2007) reveals 8 conserved targets; CALN1, OSR2, RNF165, WIT1, TLX3, MLL2, SLC22A3 and EGR3. Some caution must be exercised when interpreting the results of miRNA target prediction before experimental evidence exists. However, it is of note that Wilms tumour upstream neighbour 1 is in the list. This intronless gene lies at 11p13 in a head to tail fashion with the Wilms tumour 1 (WT1) gene and shares a bidirectional promoter with it. WT1 transcripts have been reported to be imprinted (Jinno et al. 1994, Mitsuya et al. 1997, Dallosso et al. 2004). Furthermore WIT1 hypomethylation has been implicated in chemoresistant acute myeloid leukaemia (Plass et al. 1999).

The absence of *IGF2* imprinting in monotreme mammals makes it of particular interest to search for the features discussed above. As described in chapters III and IV physical BAC mapping and sequencing in the IC1 (and IC2) region of platypus has been hampered by clone coverage and stability. As a consequence a BAC spanning ECR14 to ECR15 has yet to be identified. Likewise, no whole genome shotgun contig(s) have been identified from this orthologous region. New platypus genomic resources such as markers and/or BAC libraries may be required to extend existing sequences in this region. A BAC library is also now available for the short-

beaked echidna (*Tachyglossus aculeatus*). Further work will be required to determine whether the IC1 region can be mapped and sequenced in this monotreme.

To conclude, the *H19-IGF2* imprinting system must have been present in the therian ancestor making it the most ancient mammalian epigenetic system identified to date. The marsupial *H19* gene is also the most ancient macroRNA known and for at least 148 million years has served as a miRNA precursor. These findings should help in ascribing elusive function to the *H19* gene.