Chapter 5

Investigation of CpG island methylation on the opossum X chromosome

5.1 Introduction

Marsupial and eutherian mammals shared a common ancestor approximately 180 MYA (Murphy *et al.*, 2004). The separation of the marsupial lineage is half way between the split between birds and mammals $\sim$ 310 MYA (Bininda-Emonds *et al.*, 2007; Hedges and Kumar, 2004) and the radiation of the eutherian clades $\sim$ 100 MYA (Bininda-Emonds *et al.*, 2007; Woodburne *et al.*, 2003), making them an important group for studying the biology and evolution of mammals. Although the marsupials share some basic mammalian characteristics with the eutherians, for example giving birth to live young, presence of mammary glands and fur, they also display a great number of differences, the most distinct being

the early developmental stage at birth and the long and complicated lactation period. Such divergence is also apparent in the genetics of the sex chromosomes.

Marsupial mammals share the same XY sex determination system with the eutherians, and also achieve dosage compensation through XCI, but with distinct differences. In both groups, the inactive X chromosome is late replicating and turned into heterochromatin. Key histone modification on the eutherian inactive X chromosome, like hypoacetylation of histone H4, was identified on the marsupial X. However, the initiation and maintenance of XCI appear to be very different in the marsupials. Most noticeably, the master switch gene for random XCI in eutherians, *XIST*, is not present in marsupials, and XCI in marsupials is imprinted, where the paternal X is always silenced. It has also been suggested that methylation of the CGIs on the inactive X chromosome, contributing towards maintenance of XCI in the eutherians mammals, may be absent from the marsupial mammals (Kaslow and Migeon, 1987).

To investigate whether the apparently poor maintenance of XCI states in marsupials (Cooper *et al.*, 1977; Samollow *et al.*, 1987) is associated with lack of CGI methylation, Kaslow and Migeon (1987) studied the 5' CGI of *G6PD* in a North American marsupial, the Virginian opossum (*Didelphis virginiana*). *G6PD* is stably inactivated in human and mouse, and its 5' CGI was found to be methylated on the inactive X in both species. Its XCI appears to be less stable in several marsupial species, where expression of the paternal (inactive) allele of *G6PD* was detected in tissue samples or cultured fibroblast cells (Samollow *et al.*, 1987). Kaslow and Migeon (1987) found this CGI to be hypomethylated in both male and female samples, possibly leading to its incomplete XCI. However, despite lack of methylation, expression of the paternal allele was at best partial, so they

proposed that methylation may not play a role in marsupial XCI, but that it serves to stabilise XCI in eutherians. Later, bisulphite sequencing analysis of the 5' CGI of *G6PD* in an Australian marsupial, the common wallaroo (*Macropus robustus*), confirmed hypomethylation (Loebel and Johnston, 1996). The paternal allele of *G6PD* was previously found to be completely repressed in this species, so the authors speculated that CGI methylation might not be required for maintenance of XCI in marsupials.

At the time of commencement of this study, the notion that 5' CGIs are not methylated on the marsupial X chromosome is based on methylation analysis of only a single gene, *G6PD*. During the course of the study, a new study was described where Hornecker and colleagues (2007) examined CGI methylation of *G6PD* and PGK1 in adult females of the grey short-tailed opossum (*Monodelphis domestica*). Using bisulphite sequencing, they concluded that both CGIs were hypomethylated on the paternal X (also the inactive X), so the previous perception about CGI methylation in marsupial XCI remained unchallenged.

However, there is now evidence for differential methylation at the 5' CGIs of imprinted autosomal genes in marsupials (Smits *et al.*, 2008; Suzuki *et al.*, 2007). Suzuki and colleagues (2007) reported that *PEG10* is imprinted in tammar wallaby, with almost exclusive expression from the paternal allele. Its 5' promoter region, covered by a CGI, was heavily methylated on the maternal allele but unmethylated on the paternal allele. The same CGI also covers the 5' promoter region of an adjacent gene, *SGCE*, but the *SGCE* part of the CGI was completely unmethylated, correlating with the biallelic expression pattern of *SGCE*. A second differentially methylated region associated with an imprinted gene in marsupials was identified at the *H19* promoter and exon 1 in tammar wallaby in a

recent study of the *H19-IGF2* imprinting locus in mammals (Smits *et al.*, 2008). This association between differential methylation and imprinting is similar to the situation in eutherians. On the basis of these observations, and considering the dearth of information on X-linked genes, it remains a possibility that differential methylation of X chromosome genes may be involved in marsupial XCI.

The availability of the first whole genomic sequence of a marsupial (Mikkelsen *et al.*, 2007) provides a great opportunity to examine the methylation states of a greater number of X-linked genes in marsupials and to ascertain the importance of methylation in marsupial XCI.

5.1.1 The grey short-tailed opossum (*M. domestica*)

There are more than 300 species of marsupial mammals currently identified, mainly in Australia, where they are the dominant mammalian group (more than 200 species, including kangaroos, possums and koalas), and in South America (more than 90 species of opossums and shrew opossums) (Wilson and Reeder, 2005). Only one species is found in North America, the Virginian opossum (*D. virginiana*).

The chosen species for the first marsupial genome sequencing project was the grey short-tailed opossum, *M. domestica* (Figure 5.1), a South American marsupial that is used as a model organism in a wide range of biological studies (reviewed in Samollow, 2006). It is small, fast-growing, rapid-breeding, and can easily adapt to standard rodent facilities, so it is probably not surprising that *M. domestica* has been raised in pedigreed lab colonies for over 25 years (VandeBerg, 1989). Some earlier work on marsupial XCI was carried out using the Virginian

opossum, the original 'opossum', but as *M. domestica* gains popularity as an experimental model, *D. virginiana* is used much less. Both *D. virginiana* and *M. domestica* belong to the Didelphidae family. For the purpose of this thesis, the name 'opossum' refers to *M. domestica*.

Figure 5.1: South American, grey short-tailed opossum (*Monodelphis domestica*). Courtesy of Paul Samollow, Southwest Foundation for Biomedical Research, San Antonio. Freely distributed image obtained from `http://www.genome.gov/`

5.1.2 The opossum X chromosome

A high quality draft of the genomic sequence of opossum was described in May 2007 (Mikkelsen *et al.*, 2007). The genome of a partially inbred female opossum was sequenced using the whole genome shotgun approach to 6.8x depth. The

current assembly, monDom5, covers 99% of the estimated euchromatic sequence, and has excellent accuracy and contiguity.

The opossum X chromosome is very small compared to the autosomes. It is the smallest chromosome in opossum, representing only 2% of the total opossum genome, and is approximately 76 Mb in length (Mikkelsen *et al.*, 2007). In comparison, the sizes of X chromosomes in eutherian mammals tend to be comparable with that of the autosomes. The human, mouse, and dog X chromosomes each represent approximately 5% of the genome and are about twice the size of the opossum X (Lindblad-Toh *et al.*, 2005; Ross *et al.*, 2005; Waterston *et al.*, 2002). The size difference reflects the addition of the XAR to the X chromosome after the split between marsupial and eutherian mammals. Comparative mapping confirmed the conservation of synteny between the opossum X chromosome and the XCR of the human X chromosome (Mikkelsen *et al.*, 2007).

The opossum X chromosome is also unique in its sequence composition in relation to its autosomes. In eutherian mammals, the X chromosome has lower (G+C) content and CpG density than the autosomes. The situation in the opossum genome is just the opposite. The (G+C) content of the opossum X is not only higher than that of the autosomes, which have the lowest mean autosomal (G+C) content among all mammals sequenced, but also higher than any other X chromosome sequenced (Mikkelsen *et al.*, 2007). Similarly, despite the extremely low CpG frequency in opossum autosomes (two-fold lower than in other mammals), the CpG density of the opossum X chromosome is comparable with that of the eutherian X chromosomes (Mikkelsen *et al.*, 2007). One hypothesis proposes that the decline of (G+C) content, especially CpG dinucleotides, is balanced by a recombination-mediated (G+C)-biased gene conversion process (Duret 2006).

Consistent with this model, the opossum autosomes have lower recombination rates than those of other sequenced mammals, while the recombination rate of the opossum X was estimated to be higher than that of the autosomes, possibly as a result of their dramatic size difference (Mikkelsen *et al.*, 2007). If at least one recombination event is required for each chromosome pair at meiosis to ensure proper disjunction, then the small size of the X would result in very high recombination rate relative to the autosomes. Interestingly, this is also the opposite of the situation in eutherian mammals, where the X chromosome has a lower recombination rate than the autosomes.

5.1.3 Aims of this chapter

The aim of the work described in this chapter is to explore a large number of CGIs on the opossum X chromosome to investigate the hypothesis that lack of differential methylation between the female X chromosomes is a common feature of marsupial X-linked genes.

5.2 Identification of CGIs on the opossum X chromosome

Information about the opossum genes used in this study was obtained from the Ensembl database (v47) based on the sequence assembly monDom5. Opossum CGIs were predicted (by Val Curwen) using the cpg program with the same parameters that were applied in identifying human and mouse CGIs (section 3.2.2).

There were 448 protein coding genes found on the opossum X chromosome, 144 (32.1%) of which had a predicted CGI within 2 kb of the 5' end. CGI sizes range between 400 bp and 5 kb, with an average size of about 1 kb. Of the 130 CGIs that are each associated with a single gene, 94 overlap with exon1, the vast majority extending into intron1, and the remaining 36 are upstream from the 5' end of the genes. Seven pairs of genes, transcribed in opposite directions, had a CGI covering the 5' regions of both genes. Four of these CGIs overlap with exon1 of one gene and are a short distance upstream from the 5' end of the other gene, two CGIs are upstream from both genes, and one CGI overlaps with exon1 of both genes.

Out of these 144 genes with a predicted 5' CGI on the opossum X chromosome, 110 have putative human orthologues according to Ensembl annotation; 107 of these are also X-linked and are of particular interest for this study. For convenience, in this thesis all opossum genes are referred to using the gene ID of their putative human orthologues (predicted in Ensembl).

5.3 Assessing CGI methylation on the opossum X chromosomes by RPMA

In my previous studies on CGI methylation in human and mouse, RPMA was shown to be a reliable method to assess rapidly the methylation states of a large number of CGIs. Therefore in the initial stage of the project, RPMA was used to examine methylation of a number of CGIs on the opossum X chromosome.

Predicted CGIs on the opossum X chromosome, with 500 bp flanking se-

quences, were extracted from Ensembl, and primers were designed to amplify 150-250 bp products containing multiple *Hpa*II/*Msp*I cleavage sites lying within the predicted CGI (section 3.3.2). Since bisulphite sequencing in human revealed that RPMA may produce a signal in case of low level methylation, three pairs of primers were designed for each CGI wherever possible to increase the predictive power of the method. Primers were tested in PCR amplification using male genomic DNA. For the 41 CGIs targeted (see below for details on selection of candidate CGIs), the expected products were successfully amplified for 37 CGIs: nine of these had three pairs of successful primers, 18 had two, and 10 had only one. All primer combinations, their predicted amplicon sizes, and the number of enclosed CCGG sites are listed in Appendix I.

Genomic DNA was extracted from the liver of one female and one male opossum (section 2.9.1). Restriction enzyme digestion was performed as described in Chapter 3, but a simpler PCR protocol (optimised by Frances Lovell, personal communication, section 2.11.2) was applied.

5.3.1 Analysis of CGI methylation on the opossum X chromosome

Initially, I selected nine candidate CGIs to assay: *G6PD*, *HPRT1*, *PGK1*, *SMCX*, *RBM10*, *UBE1*, *USP11*, *SOX3*, and *DIAPH2*. Three genes, *G6PD*, *HPRT1*, and *PGK1*, are housekeeping genes that have been extensively studied in human, mouse, and marsupials. All three are from evolutionary stratum S1, so are among the first genes recruited into XCI and have had a long time to achieve methylation. Allele-specific expression of all three genes and the CGI methylation state of

G6PD have been studied in a number of marsupial species, making them obvious candidates for further methylation studies. Two additional genes, *SOX3* and *DIAPH2*, were chosen from S1. *SOX3* is particularly interesting as it has been suggested to be the X-borne homologue of the mammalian male-sex-determining gene SRY (Foster and Graves, 1994). Therefore in theory it should be the first excluded from recombination and has had the longest time of all X-linked genes to acquire methylation. I assayed CGI methylation for both genes by RPMA in a pilot study preceding work described in Chapter 3. *SOX3* had female-specific CGI methylation in both species, consistent with the gene being X inactivated. *DIAPH2*, a partial escapee in human (Carrel and Willard, 2005), had female-specific CGI methylation in mouse but not in human. To provide a contrast to the above five genes, another four genes were chosen from the younger evolutionary stratum S2; genes on which have been recruited into XCI for a shorter time and a greater proportion escape from XCI. *SMCX* escapes XCI in both human and mouse (Agulnik *et al.*, 1994), and the CGI was found to be hypomethylated in 6 of 11 eutherian orders (12 of 18 species) tested by Jegalian and Page (1999). The *RBM10* (inactivated) - *UBE1* (escapee) - *USP11* (partial escapee) gene cluster in human contains a potential boundary separating inactivated and escapee domains (XCI status from Carrel and Willard (2005)). The orthologous genes on the opossum X form a cluster in the same order, so it is interesting to investigate whether such domains and boundary are also present in opossum. In addition, the marsupial *UBE1* has a homologue on the Y chromosome (Mitchell *et al.*, 1992), so is expected to escape from XCI and free from CGI methylation, thus providing a negative control for methylation analysis.

RPMA results were obtained for seven of these genes, but no PCR primers

were successfully designed for the CGIs of *HPRT1* or *RBM10*. Consistent with expectation, and as in human and mouse, no CGIs showed a pattern indicating hypermethylation in male samples (Table 5.1). However, not all male results showed clear hypomethylation. In some cases, a very faint or faint signal was present in the *Hpa*II lane, which may be caused by low level methylation at the enzyme cleavage sites or incomplete digestion. These faint signals in the *Hpa*II lane were often accompanied by a similar signal in the *Msp*I lane, where cleavage is not affected by methylation, so it is more likely that the *Hpa*II signal has resulted from partial digestion rather than methylation. The opossum DNA was extracted from tissue samples, so may still contain substances that inhibit digestion, for example glycogen. This problem was not seen in human an mouse, possibly due to the different source of the test DNA (from cultured fibroblast cells). Attempts were made to remove this problem, including use of fresh batch of restriction enzymes, digestion with higher concentration of enzymes, re-purification of the DNA samples, and use of spermidine to facilitate digestion, but without success.

In spite of the problems with digestion, giving rise to faint signals in male samples, speculations about methylation states are still possible through comparison between male and female results. In all but one CGIs, the female pattern was identical (4 CGIs) or very similar (2 CGIs) to the male pattern (Table 5.1). The only exception is *PGK1*, which showed a greater level of methylation in the female sample in both amplicons (Table 5.1, Figure 5.2). RPMA results for these seven CGIs are consistent with female-specific methylation being rare on the opossum X chromosome.

In order to explore this apparent lack of sex-specific differential methylation of CGIs on the opossum X chromosome further, I extended the study to another 15

CGIs, selected from the centromeric end of the chromosome. As expected, most CGIs were not hypermethylated in male samples (Table 5.1). As before, faint or very faint signals were frequently found in the *Hpa*II and *Msp*I lanes, indicative of partial digestion. Therefore, no attempt was made to define the methylation state of a CGI, as described in Chapter 3, but the island was simply scored for presence or absence of female-specific methylation.

Four amplicons, three of which contain only one cleavage site, had strong PCR products for all enzymes in both sexes, making them not informative for estimating presence of female-specific methylation. Eleven CGIs exhibited identical (5) or very similar (6) RPMA patterns in both sexes (Table 5.1). One of them, the CGI of *ATP6AP2* had a strong signal in the *Hpa*II lane for both sexes, the *Msp*I lane was clean in the female sample and had only a faint signal in the male sample, suggesting hypermethylation at both female and male islands. Heavier methylation in female was detected for the CGIs associated with *MPP1*, *FAAH2*, and *F8A*$\sim$ (Table 5.1, Figure 5.2). For all three CGIs, the male samples had a very faint signal in the *Hpa*II and *Msp*I lanes, while the female samples had a strong signal in the *Hpa*II lane, but only a very faint signal in the *Msp*I lane. Repeated experiments showed identical or very similar patterns. For the remaining CGI, the only amplicon available was not informative so no conclusion could be made. In general, these results are still consistent with lack of female-specific methylation, but uncovered a small cluster of CGIs with heavier methylation in females.

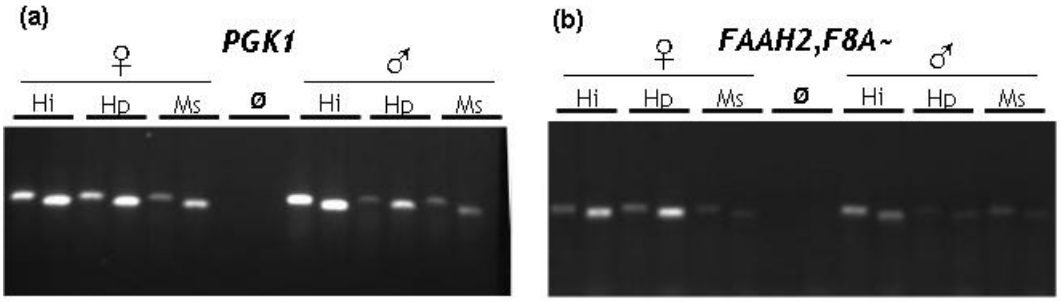

Figure 5.2: RPMA results of three opossum genes with female-specific methylation of CGI. *FAAH2* and *F8A*∼ only had one amplicon each.

Interestingly the *MPP1*, *FAAH2* and *F8A*∼ genes reside in the same neighbourhood of a 1.5 Mb region on the opossum X chromosome, only separated by a few non-CGI-associated genes. Having previously observed higher female methylation associated with the CGI of *PGK1*, I hypothesised that I might have success in detecting additional examples of methylation among its neighbouring genes. I assayed 15 CGIs surrounding *PGK1*. Disregarding the three uninformative amplicons, which had strong signals in the *Msp*I lane, 12 CGIs exhibited identical (6) or very similar (6) methylation patterns in both sexes (Table 5.1). Interestingly, the three CGIs immediately surrounding *PGK1*, *FGF16*, *IAG2*, *ITM2A*, all showed some degree of female-specific methylation (Table 5.1). The CGIs of *FGF16* and *ITM2A* each had two amplicons, one showed no difference between sexes, but the other with strong signal in the *Hpa*II lane in female and no signal in male. The only amplicon of the *IAG2* CGI had a strong signal in the *Hpa*II lane in female and a faint signal in male.

In summary, 37 CGIs were successfully assayed. The majority (73%) showed no methylation differences between the female and male samples (Table 5.1, Figure 5.3). Only seven CGIs, in two distinct clusters, demonstrated greater

level of methylation in the female samples (Table 5.1, Figure 5.3). However, a weak signal was frequently detected in the *Msp*I lanes in both female and male RPMA results, suggesting incomplete digestion and making it difficult to estimate the exact extent of methylation.

Table 5.1: RPMA results of opossum CGIs. '+' = strong band; 'f' = faint band; 'vf' = very faint band; '-' = no visible band. One CGI is shared by each pair of the following: *BCAP31* and *ABCD1*, *IDH3G* and *SSR4*

	RPMA primer pair 1			RPMA primer pair 2			RPMA primer pair 3			Female-specific methylation?
Gene	CCGG sites	Female Hi Hp Ms	Male Hi Hp Ms	CCGG sites	Female Hi Hp Ms	Male Hi Hp Ms	CCGG sites	Female Hi Hp Ms	Male Hi Hp Ms	
USP11	3	+ - -	+ - vf	3	+ - -	+ - -				N
UBE1	3	+ vf vf	+ vf vf	3	+ vf -	+ vf vf				N
SMCX	3	+ vf -	+ vf -	2	+ - -	+ - -				N
G6PD	3	+ vf -	+ vf -	2	+ f f	+ f f				N
SOX3	2	+ vf vf	+ vf vf	2	+ vf vf	+ vf vf				N
DIAPH2	5	+ - -	+ - -	4	+ - -	+ - -				N
PGK1	2	+ + f	+ vf vf	1	+ + f	+ f vf				Y
GDI1	4	+ - -	+ - -	2	+ f -	+ f -				N
ATP6AP1	1	+ + -	+ + f	1	+ + +	+ + +				N
PLXNA3	3	+ - -	+ - -	2	+ f vf	+ vf vf				N
SEPHS2	5	+ - -	+ - -							N
DKC1	2	+ f f	+ f f	2	+ + +	+ + +				N
MPP1	2	+ + vf	+ vf vf							Y
FAAH2	2	+ + vf	+ vf vf							Y
F8A~	1	+ + vf	+ vf vf							Y
TSC22D3	2	+ f -	+ vf vf	5	+ f -	+ f -				N
TMEM164	4	+ - -	+ - -	1	+ + +	+ + +	1	+ + f	+ f f	N
AMMECR1	3	+ vf vf	+ vf vf							N
AR	3	+ - -	+ - -	3	+ vf -	+ vf -	1	+ - -	+ - -	N
MSN	1	+ + +	+ + +							U
BCAP31	2	+ + f	+ + f	3	+ f -	+ f -	2	+ - -	+ - -	N
IDH3G	2	+ vf -	+ - vf	3	+ vf -	+ vf -				N
HTATSF1	1	+ f f	+ f vf	3	+ vf vf	+ vf -	3	+ f -	+ vf -	N
RBMX	2	+ vf -	+ vf -	2	+ - -	+ - -				N
GPR101	2	+ - -	+ - -							N
ZIC3	1	+ + +	+ + +	1	+ + f	+ f f	2	+ vf vf	+ vf vf	N
CHM	2	+ - -	+ - -	2	+ f f	+ f vf	3	+ f -	+ f -	N
DACH2	3	+ - -	+ - -	2	+ f -	+ f -	3	+ f -	+ vf -	N
FGF16	3	+ + vf	+ - -	7	+ - -	+ - -				Y
IAG2	1	+ + f	+ f f							Y
ITM2A	1	+ f f	+ f f	1	+ + -	+ - -				Y
BRWD3	2	+ f f	+ f f	2	+ f f	+ f f	2	+ f f	+ f f	N
ALG13	1	+ vf vf	+ - -	2	+ f vf	+ f -				N
OGT	3	+ - -	+ - -	1	+ f vf	+ f vf				N
TAF1	4	+ - -	+ - -	2	+ - -	+ - -	1	+ + +	+ + +	N
ZMYM3	2	+ - vf	+ - vf							N
SNX12	1	+ f vf	+ vf vf							N

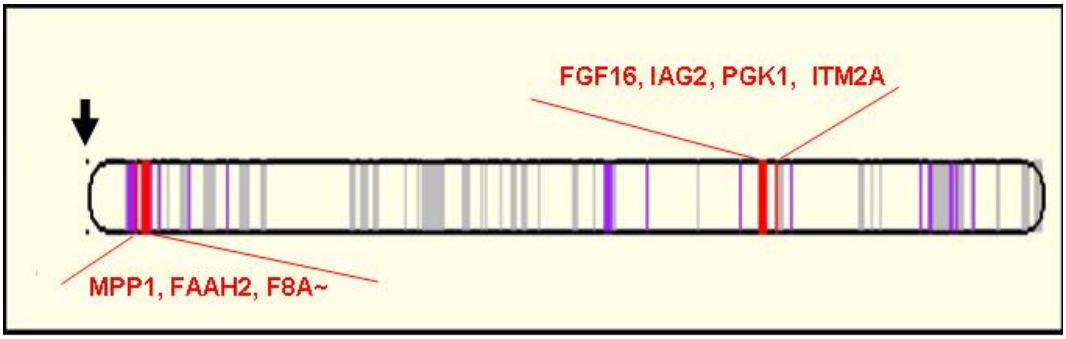

Figure 5.3: Distribution of assayed CGIs on the opossum X chromosome. Locations of all assayable CGIs are indicated by grey bars. The assayed CGIs are coloured: those with female-specific methylation are in red, and the ones without female-specific methylation are in purple. Location of centromere is indicated by arrow.

5.4 Assessing CGI methylation on the opossum X chromosomes by bisulphite sequencing

The data described indicate that most CGIs on the female opossum X chromosomes are hypomethylated. This is consistent with the previous picture for other marsupial species but extends the basis for this conclusion well beyond the study of a single gene. Most significantly, the RPMA analysis has uncovered a number of apparent examples of higher female methylation. If confirmed, this would be the first demonstration of this phenomenon in marsupial and would open up the possibility of a role for methylation in marsupial XCI that was not previously appreciated. Owing to the possibility of incomplete digestion, as well as the limited resolution of RPMA, it was necessary to confirm these findings using a more accurate method. To this end, six CGIs that showed differential methylation by RPMA, together with two CGIs that are associated with well-studied genes in

marsupials (*HPRT1* and *G6PD*), were examined by bisulphite sequencing. It was not possible to design primers for the *MPP1* CGI, which also showed differential methylation by RPMA. Bisulphite sequencing was performed as described in Chapter 4. All primer combinations and their predicted amplicon sizes are listed in Appendix II.

The CGIs of *G6PD*, *HPRT1*, *FAAH2*, *FGF16*, *IAG2*, *PGK1*, and *ITM2A* were successfully bisulphite sequenced in both female and male samples (Figures 5.4-5.10). *F8A*∼ results were only available for the female sample (Figure 5.11). One or two amplicons were sequenced for each CGI, and at least 40 individual molecules, each representing the island of an individual chromosome, were sequenced for each amplicon. Some amplicons contain CpG dinucleotides outside the actual island, these CpGs were not counted in the methylation statistics presented below.

Most CGI molecules were clearly hypomethylated in the male samples. For five CGIs, the majority of island molecules were completely free from methylation, and the remaining ones had only one or two CpGs methylated (Figures 5.4-5.8). The CGI of *ITM2A* had a small proportion of island molecules with three out of 22 CpGs methylated (Figure 5.10). For the CGI of *PGK1*, a small proportion of island molecules had nearly identical low level methylation, but most other island molecules had no methylated CpG dinucleotides (Figure 5.9).

The CGIs of *G6PD* and *HPRT1* were hypomethylated in both sexes (Figures 5.4, 5.5). Like their male companions, the female CGIs had the majority of island molecules completely unmethylated, apart from a small number with one or two methylated CpGs. Only one molecule of the *HPRT1* CGI (n=77) and six molecules of the *G6PD* CGI (n=155) had more than two CpGs methylated in

females.

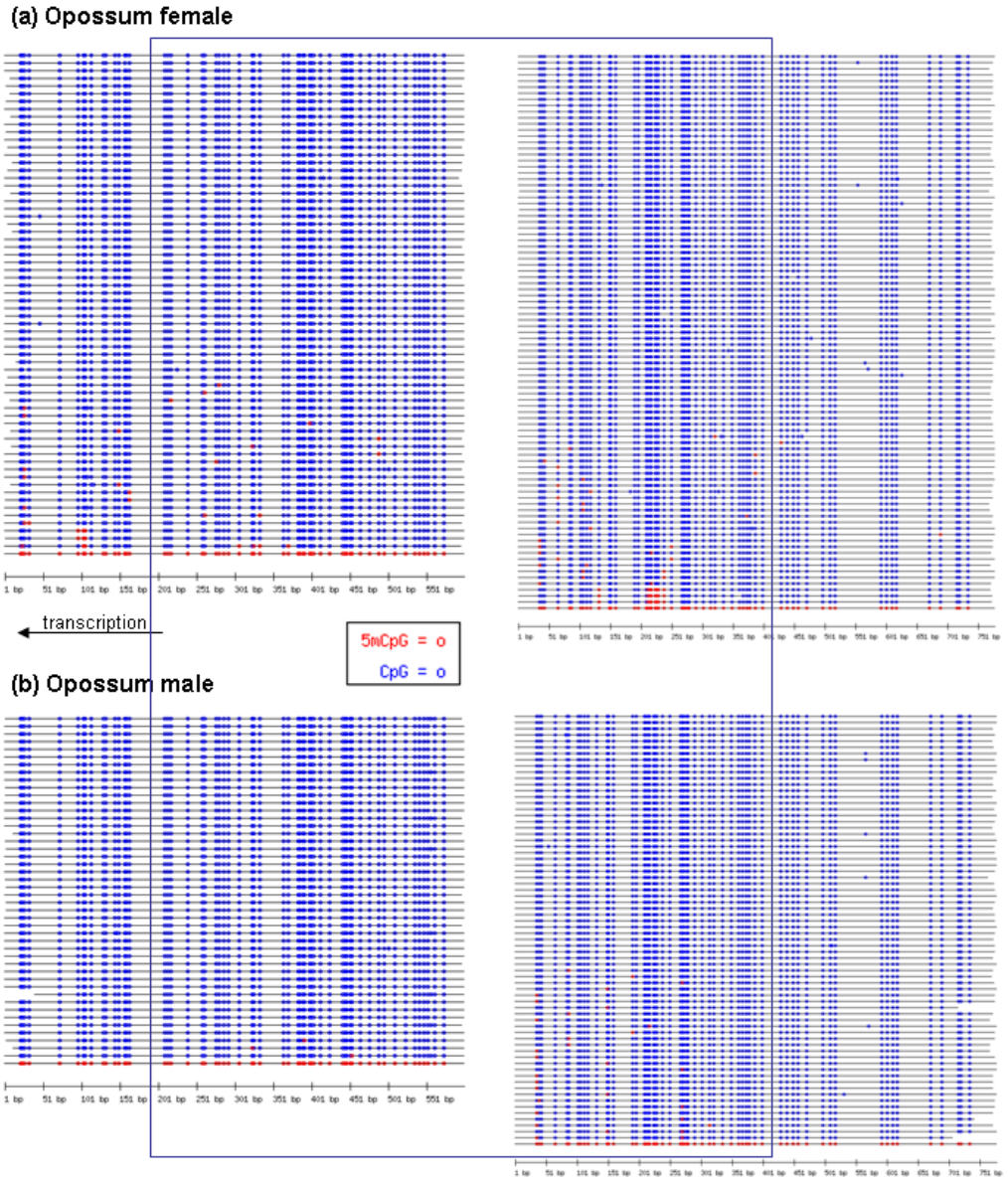

Figure 5.4: CGI methylation profiles of opossum *G6PD*. This CGI was shown to lack female-specific methylation in RPMA. The bisulphite sequenced fragments are overlapping and cover the first two thirds of the CGI. The overlapped region is indicated by the blue box. The CGI is upstream of the gene and direction of transcription is indicated by arrow. For detailed annotation see Figure 4.3.

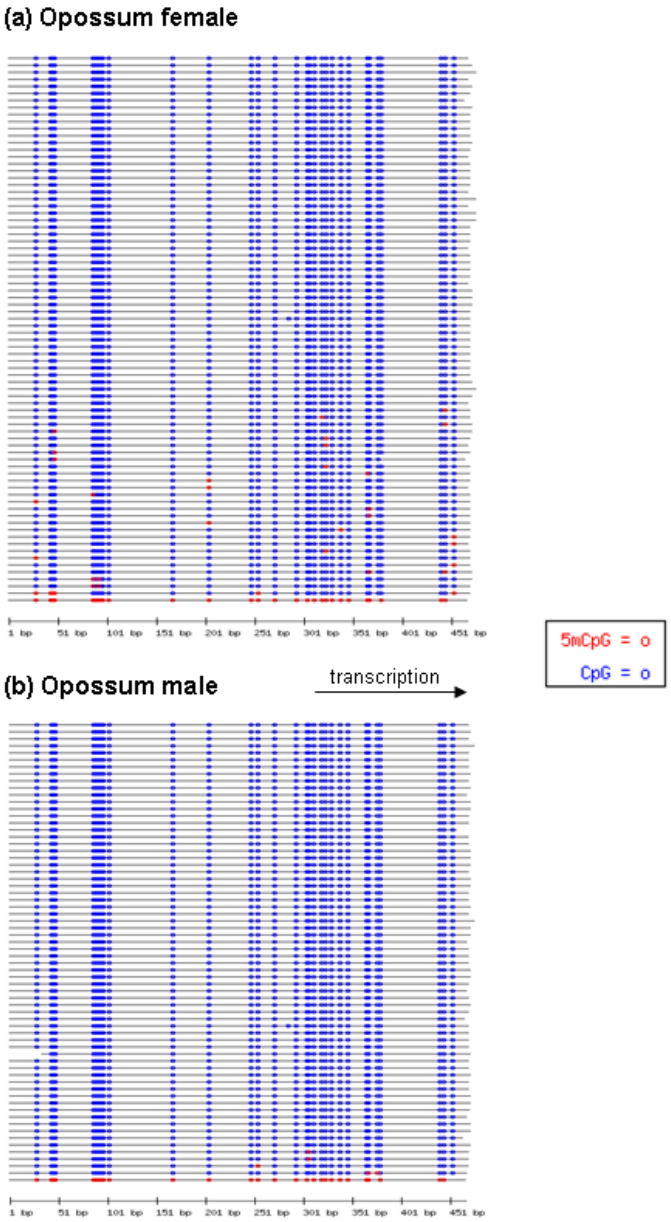

Figure 5.5: CGI methylation profiles of opossum *HPRT1*. It was not possible to assay this CGI by RPMA. The bisulphite sequenced fragment covers the first two thirds of the CGI and is upstream of the transcription start site. Direction of transcription is indicated by arrow. For detailed annotation see Figure 4.3.

As predicted by RPMA, the CGIs of *FAAH2*, *FGF16*, *IAG2*, *PGK1*, and

ITM2A all showed clearly heavier methylation in the female samples (Figures 5.6-5.10). However, unlike the mouse and human female CGIs which contained distinct hypo- and hypermethylated portions, there was a continuum of methylation densities for all opossum female CGIs and the methylation densities were generally low. Therefore it was not easy to decide which molecules might have come from the active X, and which from the inactive X. Since the single X in male can be regarded as an equivalent as the active X in female, the methylation pattern of each male CGI was used as benchmark for methylation level of X_a. The female molecules with higher level of methylation than the most heavily methylated molecule from male are presumably derived from the X_i.

Using this approach, *FAAH2* and *FGF16* exhibited a CGI methylation pattern most typical of more heavily methylated female X_i and unmethylated male X / female X_a (Figures 5.6, 5.7). Methylation was scarce in the male samples, but around half of the island molecules in the female samples had greater than male level of methylation. A similar CGI methylation pattern was observed for *PGK1* and *ITM2A*, but their male samples had a very small number of island molecules with more elevated methylation, so a smaller portion of female island molecules had greater than male level of methylation (Figures 5.9, 5.10). In the case of *IAG2*, the female CGI was still more methylated than the male CGI, but had lower proportion of methylated molecules and lower levels of methylation in the methylated molecule compared to the above islands (Figure 5.8). Around half of the island molecules in *F8A*$\sim$ CGI female samples were also methylated (Figure 5.11). Although the male results were not available for comparison, given the methylation patterns in other male CGIs it is most likely that the *F8A*$\sim$ CGI was also hypomethylated in males.

It should be noted that, overall, the extent of methylation is still low in females, even for the most highly methylated molecules. On the other hand, a clear difference between the sexes was observed, not just for the level of methylation, but also in the pattern of methylation. There are regions that were clearly methylated in female island molecules that were classified as being from the X_i, but were devoid of methylation in the male (see Figures 5.6 and 5.9 for example).

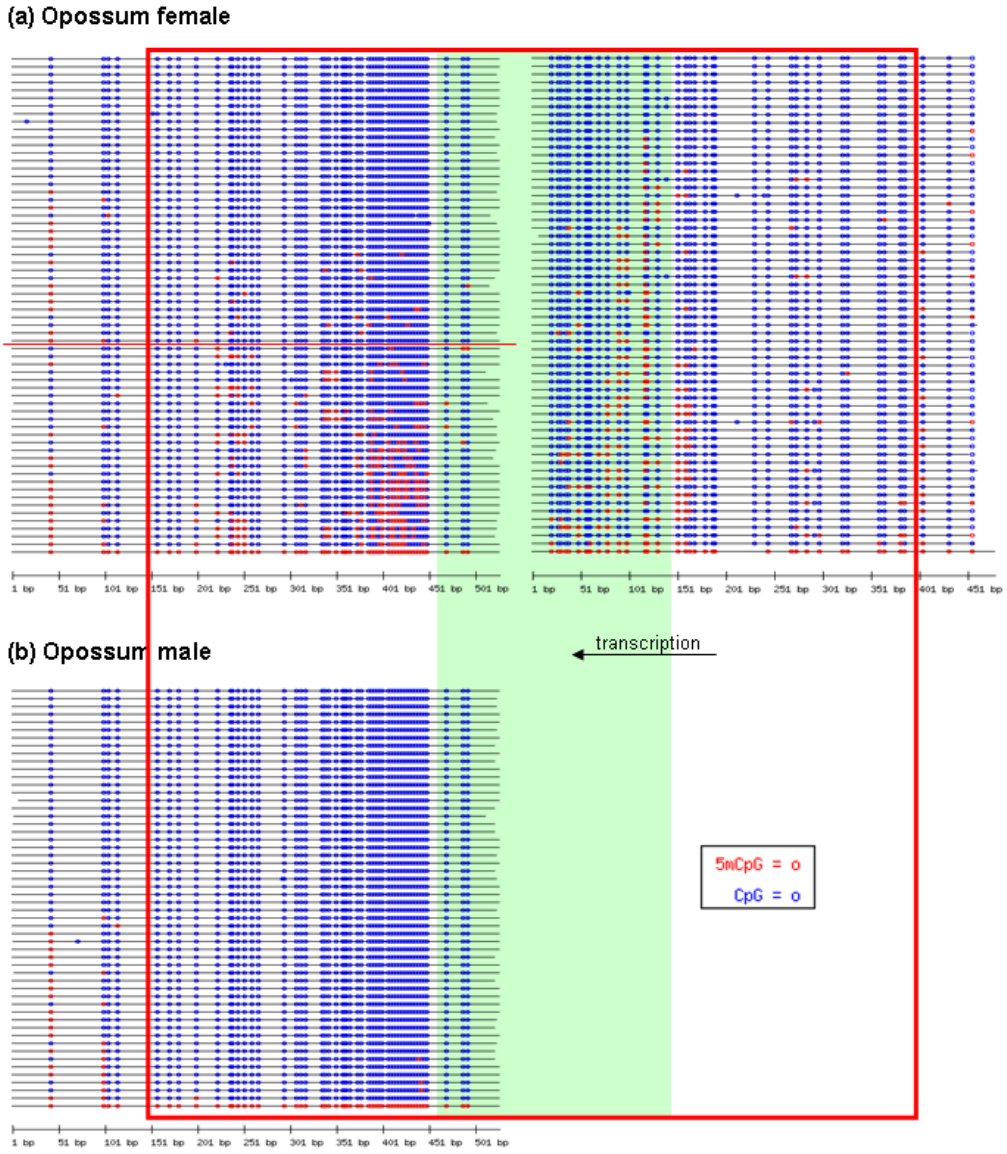

Figure 5.6: CGI methylation profiles of opossum *FAAH2*. This CGI was shown to have heavier methylation in females in RPMA . The bisulphite sequenced fragments are overlapping (but with no overlapping CpGs) and cover the entire CGI, but one amplicon only had data from female samples. The island region is indicated by the red box and the exon is shadowed in green. Female molecules potentially derived from the X_a or X_i are divided by the horizontal red line based on methylation levels in male. Direction of transcription is indicated by arrow. For detailed annotation see Figure 4.3.

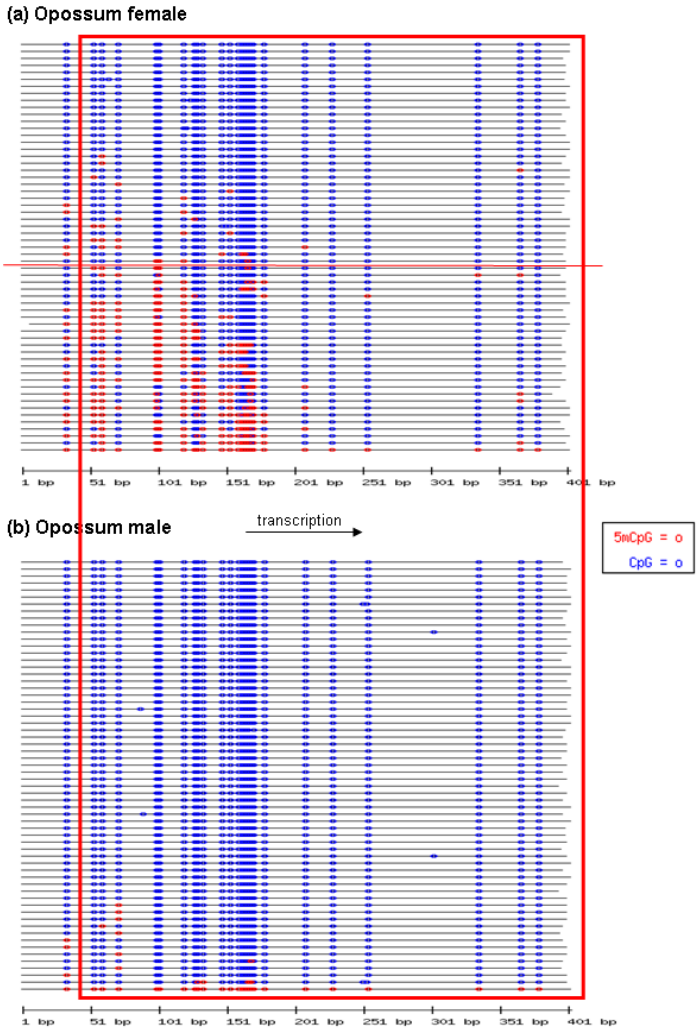

Figure 5.7: CGI methylation profiles of opossum *FGF16*. This CGI was shown to have heavier methylation in females in RPMA. The bisulphite sequenced fragment covers the first quarter of the CGI, and is upstream of the transcription start site. The island region is indicated by the red box . Female molecules potentially derived from the X_a or X_i are divided by the horizontal red line based on methylation levels in male. Direction of transcription is indicated by arrow. For detailed annotation see Figure 4.3.

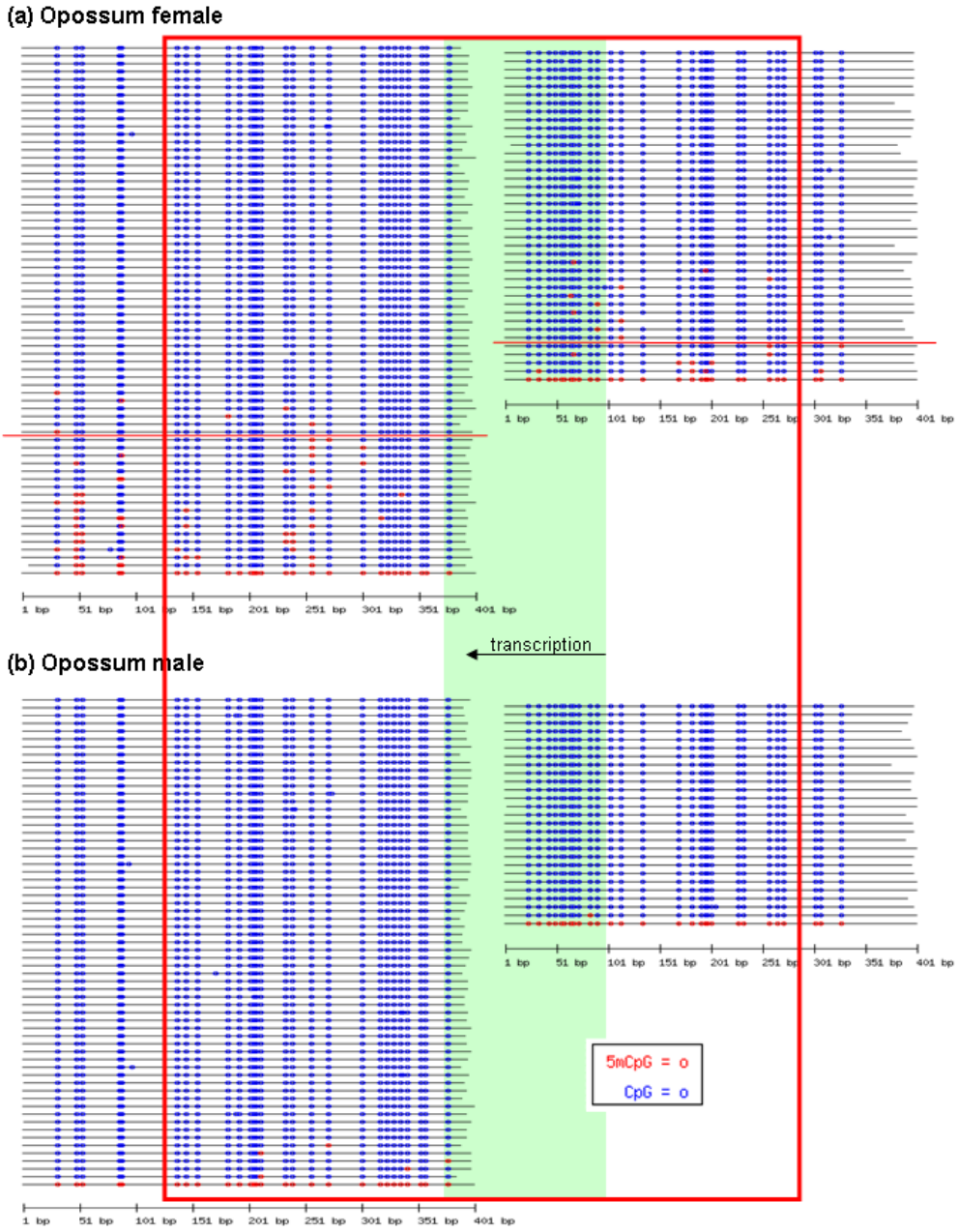

Figure 5.8: CGI methylation profiles of opossum *IAG2*. This CGI was shown to have heavier methylation in females in RPMA . The bisulphite sequenced fragments are overlapping (but with no overlapping CpGs) and cover the entire CGI. The island region is indicated by the red box and the exon is shadowed in green. Female molecules potentially derived from the X_a or X_i are divided by the horizontal red line based on methylation levels in male. Direction of transcription is indicated by arrow. For detailed annotation see Figure 4.3.

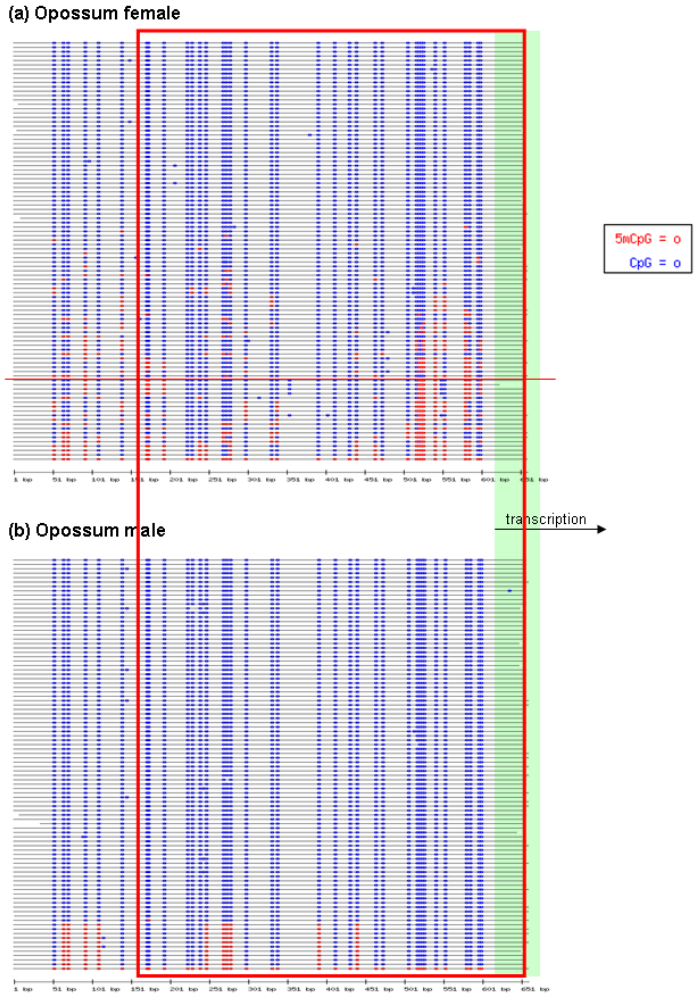

Figure 5.9: CGI methylation profiles of opossum *PGK1*. This CGI was shown to have heavier methylation in females in RPMA . The bisulphite sequenced fragment covers the entire CGI. The island region is indicated by the red box and the exon is shadowed in green. Female molecules potentially derived from the X_a or X_i are divided by the horizontal red line based on methylation levels in male. Direction of transcription is indicated by arrow. For detailed annotation see Figure 4.3.

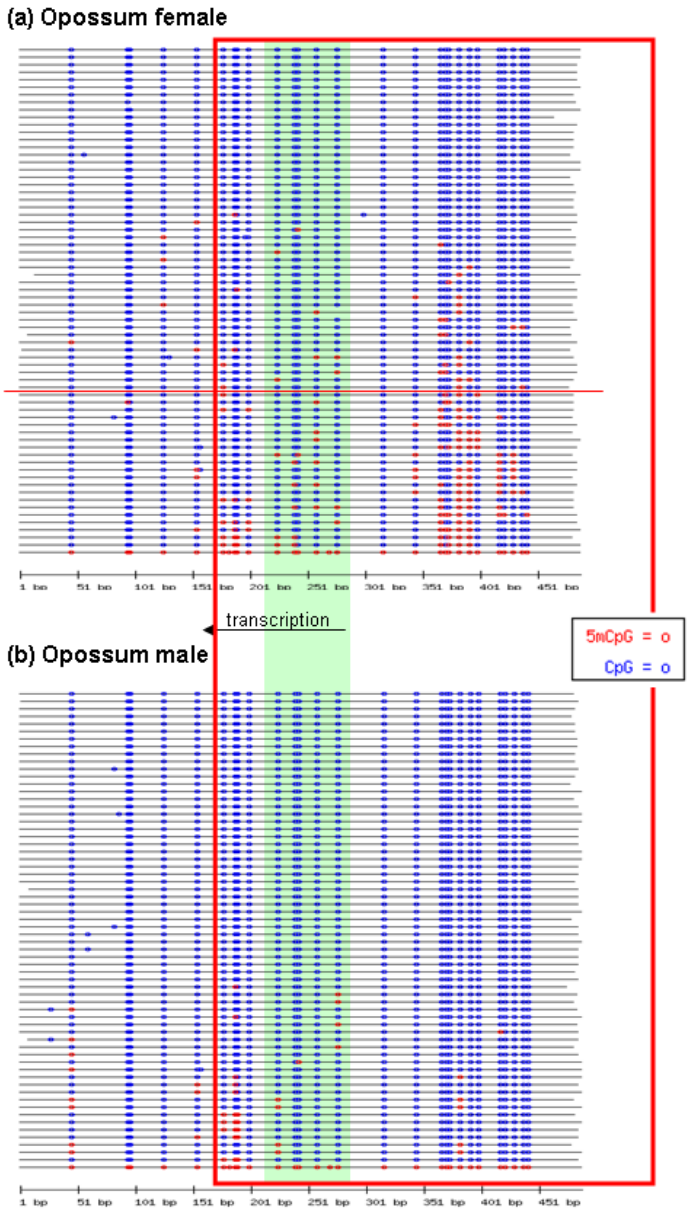

Figure 5.10: CGI methylation profiles of opossum *ITM2A*. This CGI was shown to have heavier methylation in females in RPMA . The bisulphite sequenced fragment covers the first half of the CGI. The island region is indicated by the red box and the exon is shadowed in green. Female molecules potentially derived from the X_a or X_i are divided by the horizontal red line based on methylation levels in male. Direction of transcription is indicated by arrow. For detailed annotation see Figure 4.3.

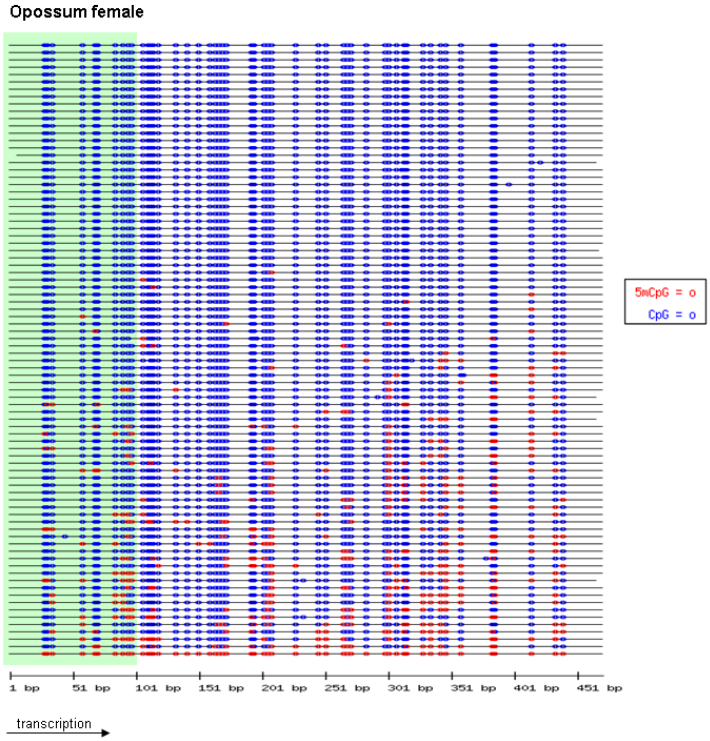

Figure 5.11: CGI methylation profiles of opossum *F8A*∼. This CGI was shown to have heavier methylation in females in RPMA. Bisulphite sequencing was only successful using the female samples. The bisulphite sequenced fragment covers a quarter of the island in the second half of the CGI. The exon is shadowed in green. Direction of transcription is indicated by arrow. For detailed annotation see Figure 4.3.

5.5 Comparison of methylation status across multiple tissues

Following confirmation of the elevation in CGI methylation levels on some female opossum X chromosomes, I decided to investigate whether or not such methylation is limited to the liver. Therefore, all eight CGIs successfully assayed using liver samples were examined by bisulphite sequencing in three additional tissue

samples from the same individuals: heart, kidney, and spleen. The CGIs of *FGF16*, *IAG2*, *PGK1*, and *ITM2A* were successfully assayed in all three tissues (except *PGK1* in male kidney); the CGI of *HPRT1* was successfully assayed in kidney and spleen; but the CGI of *FAAH2* failed in all tissues. For *FA8*∼ results were produced from kidney and spleen, and, as previously seen for liver, only female samples were successfully assayed. For convenience of comparison, methylation density was calculated for each CpG to produce an average methylation level plot for each CGI (Figure 5.12).

Methylation patterns of all test CGIs in the other tissues were similar to those found in liver. The CGIs of *G6PD* and *HPRT1* were hypomethylated in both sexes in all tissues (Figures 5.13 and 5.14). The CGIs of *FGF16*, *PGK1*, and *ITM2A* showed differential methylation consistent with the female X_i being more heavily methylated than the male X / female X_a (Figures 5.15-5.17). The CGI of *F8A*∼ shared a similar methylation pattern in female samples with these three, and is likely to have sex-specific differential methylation although no firm conclusion can be made due to the lack of male results (Figure 5.19). Heavier methylation in female was also observed in the *IAG2* CGI across all tissues, but the methylation was always at very low levels (Figure 5.18).

It is noted from Figures 5.15-5.19 that the distribution of methylation densities along a CGI is not uniform. Interestingly, in all methylated female samples, the methylation density peaks were located around the same positions in all tissues. In general, the methylation levels were slightly higher in liver and spleen than in heart and kidney, but this difference in methylation densities does not affect the distribution of methylation densities in any single tissue, only altering the height of density peaks.

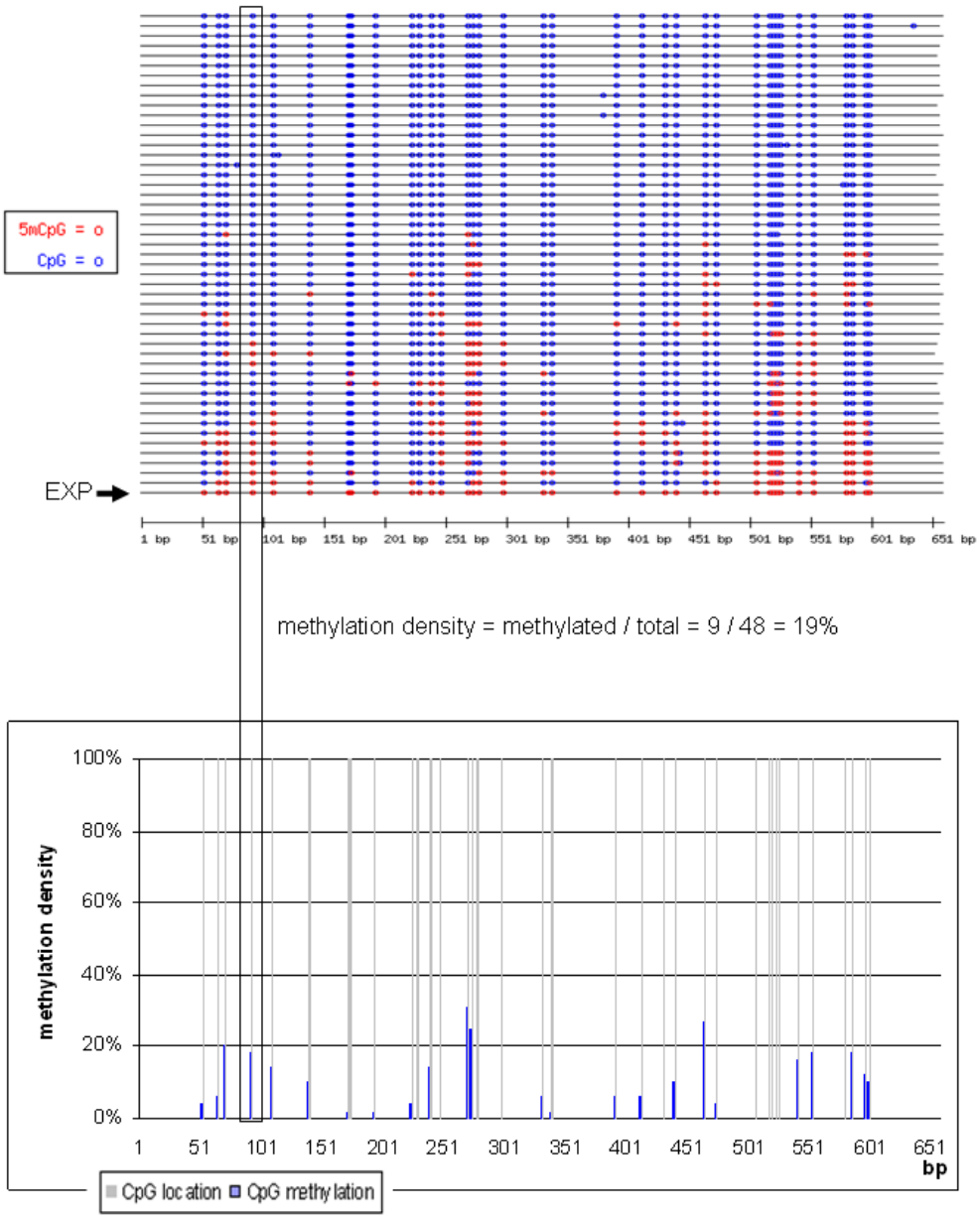

Figure 5.12: Graphical representation of average methylation levels in individual CGI. Methylation density was calculated for each CpG as the percentage of molecules methylated at this CpG out of the total number of molecules where this CpG is present. Methylation densities are shown in a bar chart where the grey bars indicate locations of expected CpGs according to the reference sequence ('EXP') and blue bars represent methylation levels. Methylation density calculation for an example CpG is demonstrated in the figure. Note that the 'EXP' sequence was excluded in the calculation.

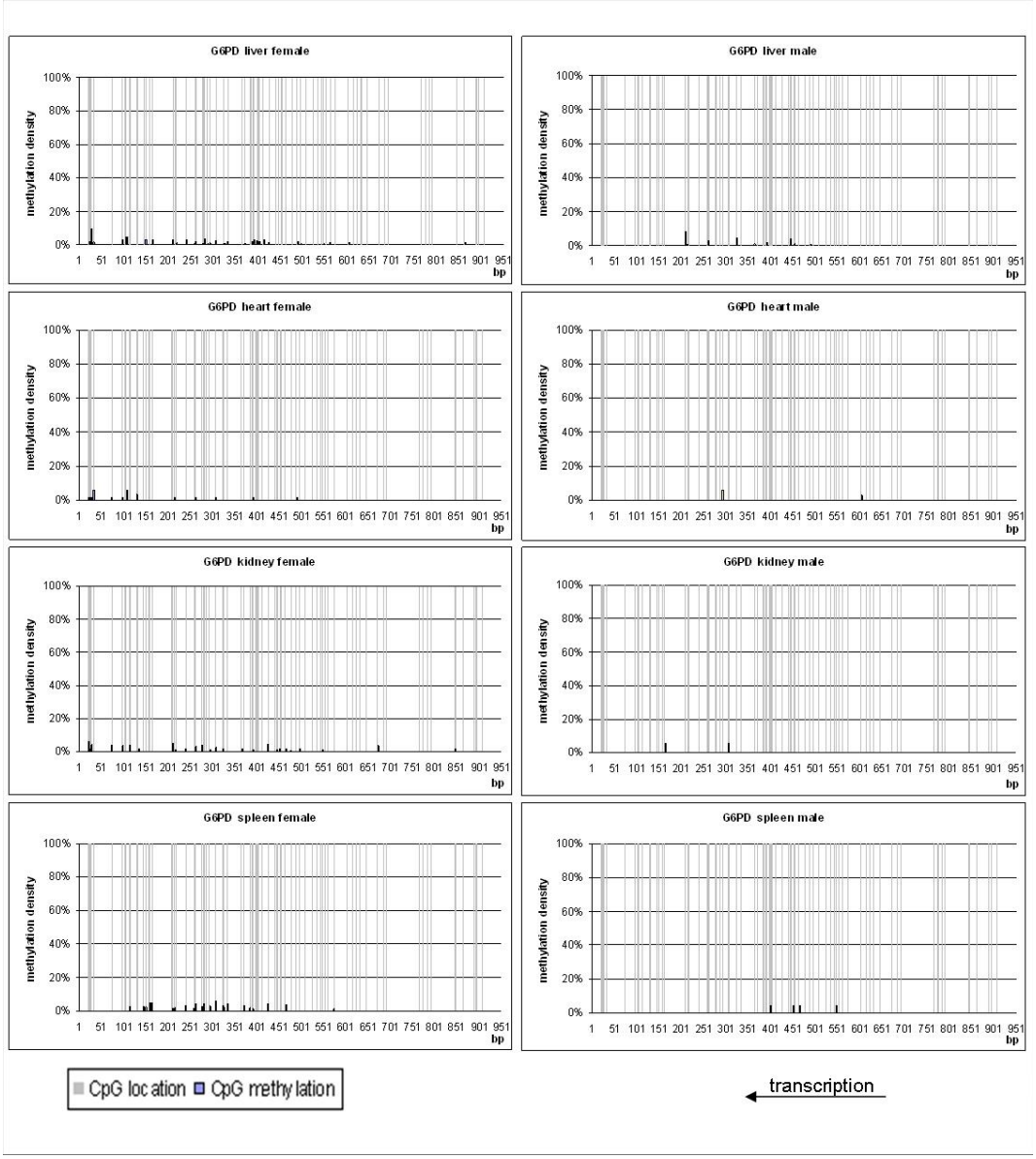

Figure 5.13: CGI methylation profiles of opossum *G6PD* in four tissues. Methylation levels are average of two overlapping sequences, the first covering 1-599 bp, and the second covering 175-951 bp. For female heart and male kidney samples, only the first sequence was available. For male heart and spleen samples, only the second sequence was available. Direction of transcription is indicated by arrow. For detailed annotation see Figure 5.12.

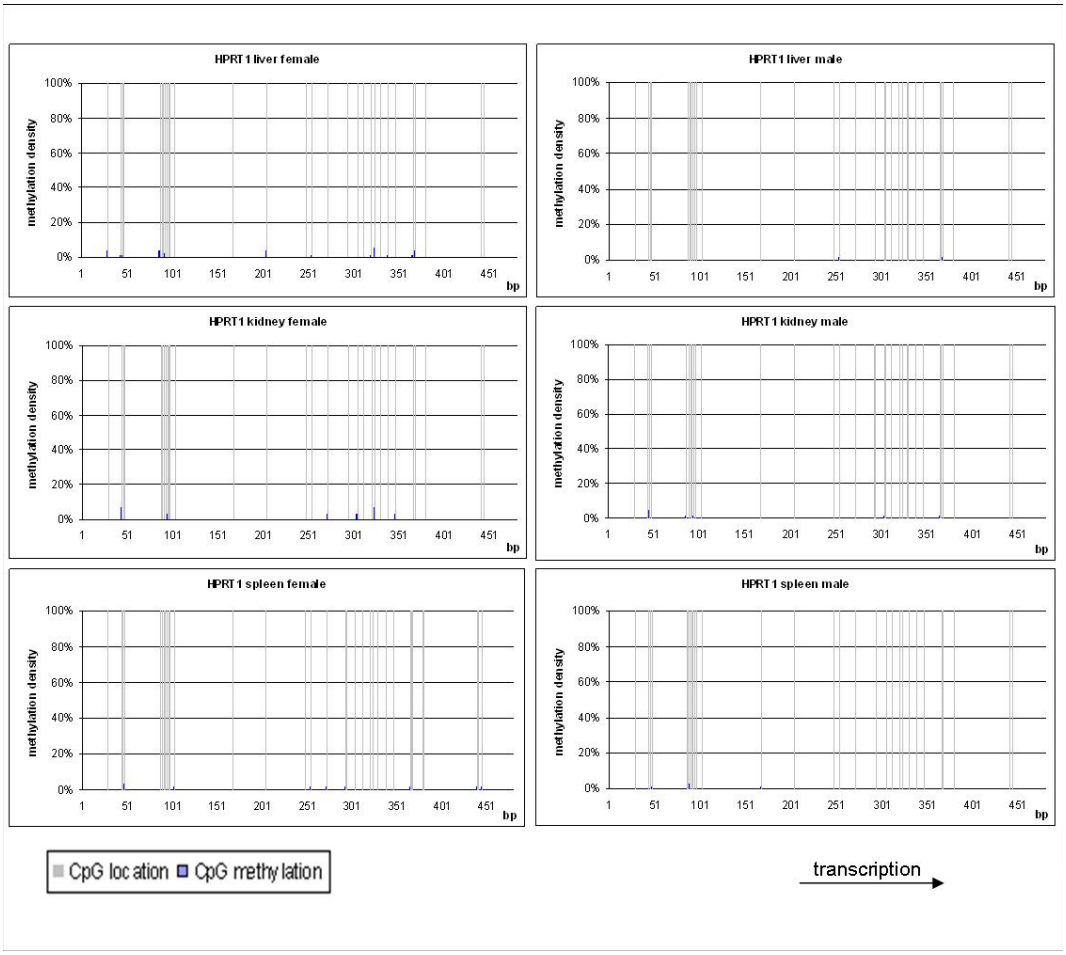

Figure 5.14: CGI methylation profiles of opossum *HPRT1* in three tissues. Direction of transcription is indicated by arrow. For detailed annotation see Figure 5.12.

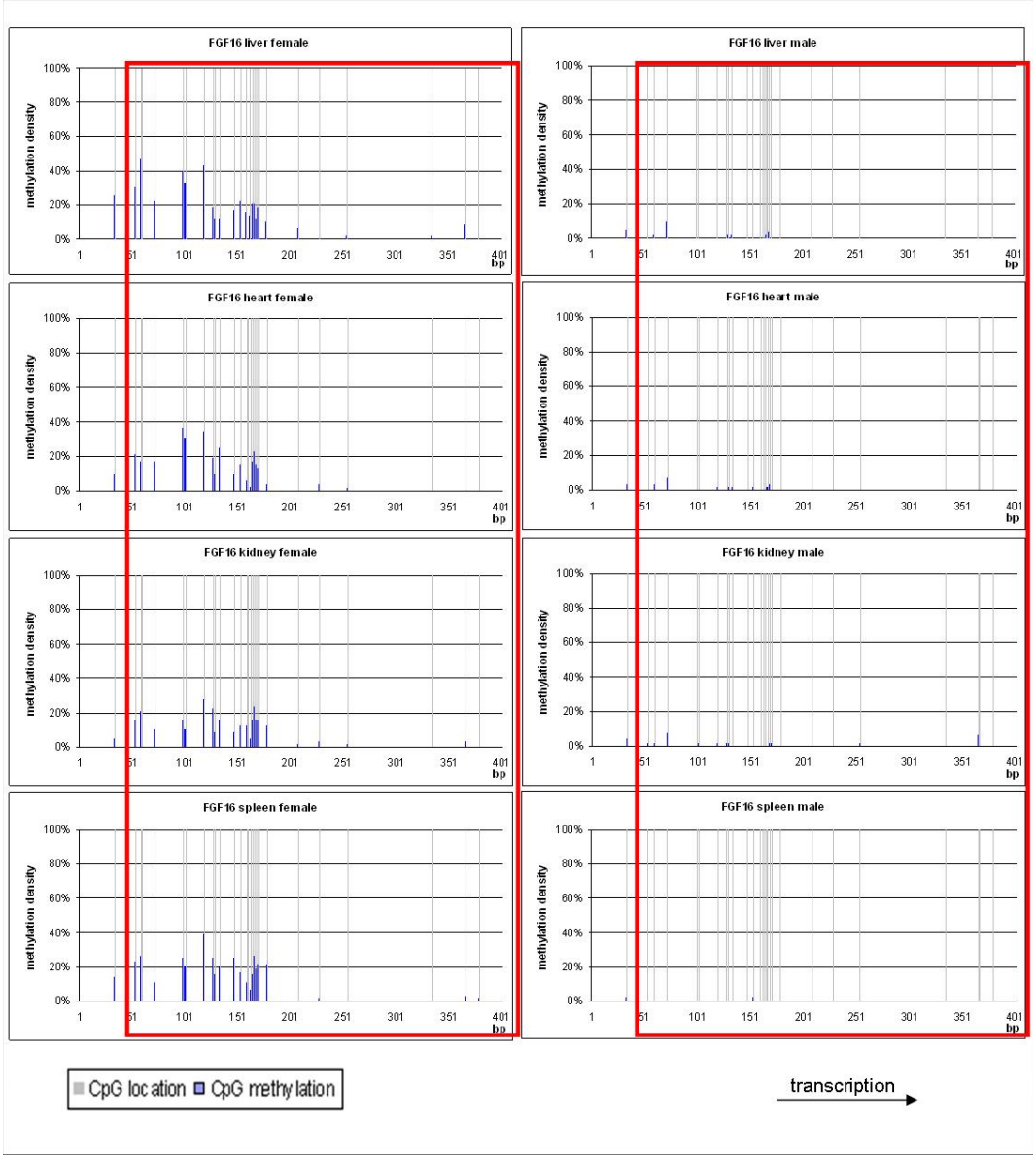

Figure 5.15: CGI methylation profiles of opossum *FGF16* in four tissues. The island region is indicated by the red box. Direction of transcription is indicated by arrow. For detailed annotation see Figure 5.12.

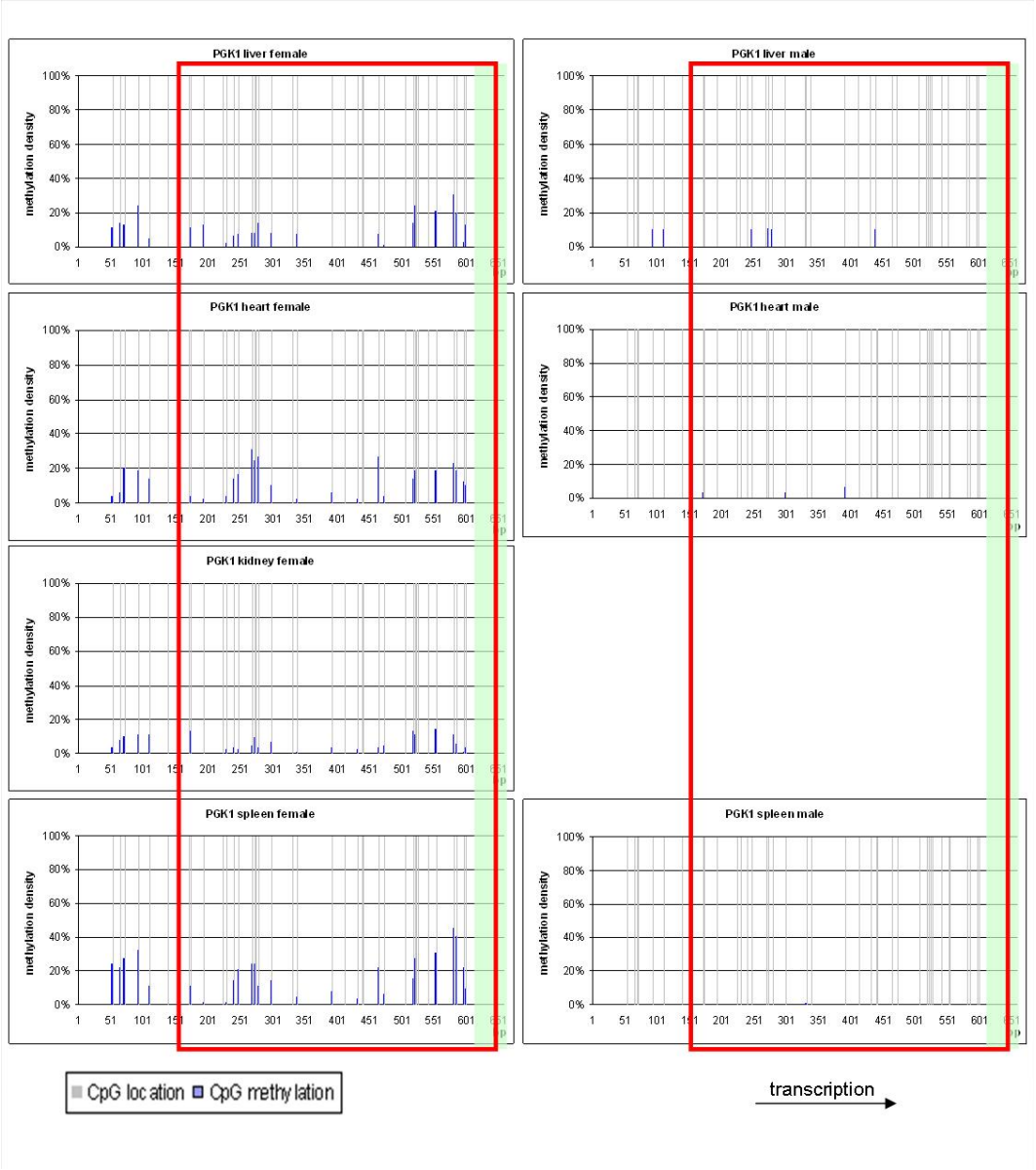

Figure 5.16: CGI methylation profiles of opossum *PGK1* in four tissues. Bisulphite sequencing using the male kidney sample was unsuccessful. The island region is indicated by the red box and the exon is shadowed in green. Direction of transcription is indicated by arrow. For detailed annotation see Figure 5.12.

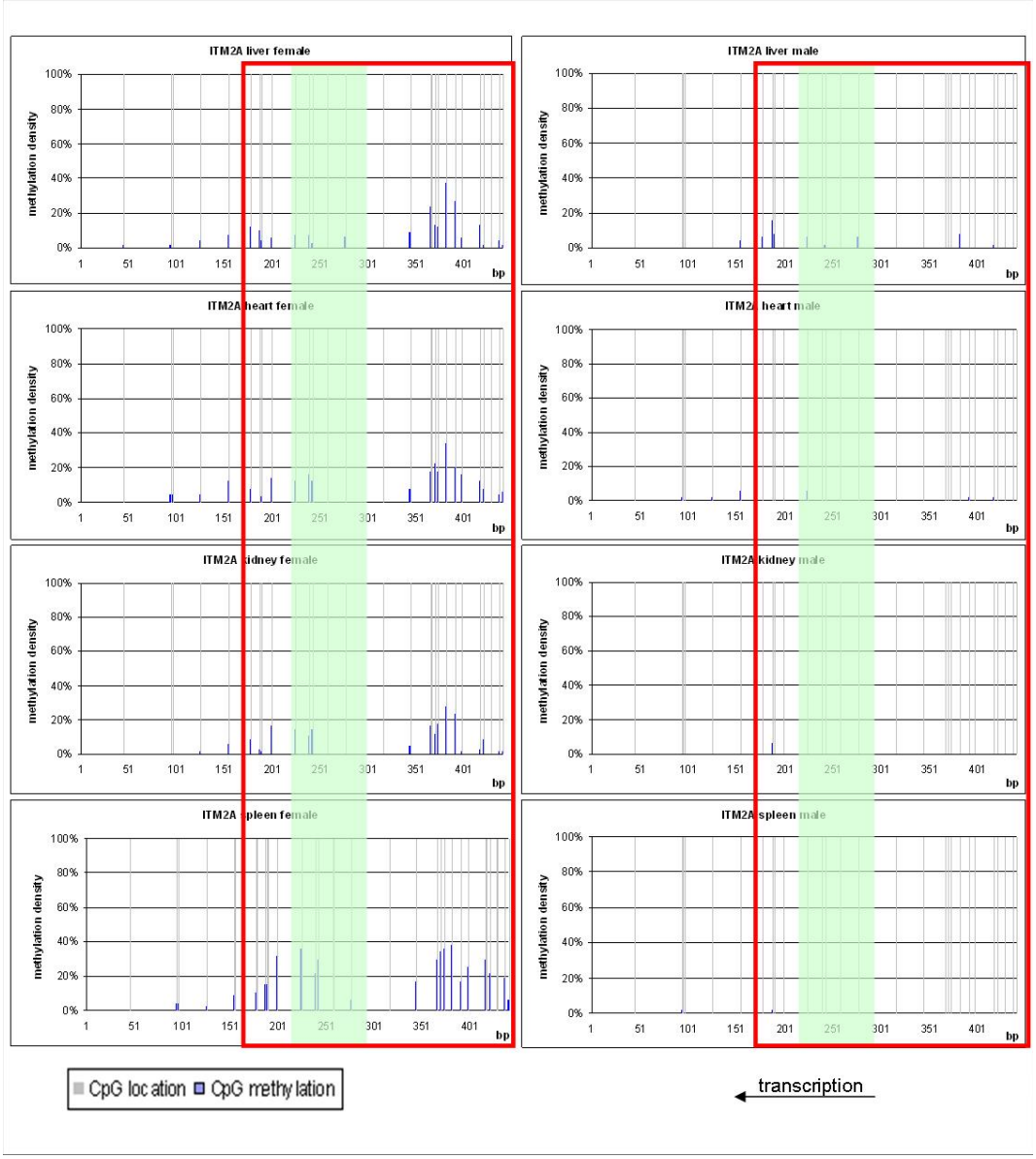

Figure 5.17: CGI methylation profiles of opossum *ITM2A* in four tissues. The island region is indicated by the red box and the exon is shadowed in green. Direction of transcription is indicated by arrow. For detailed annotation see Figure 5.12.

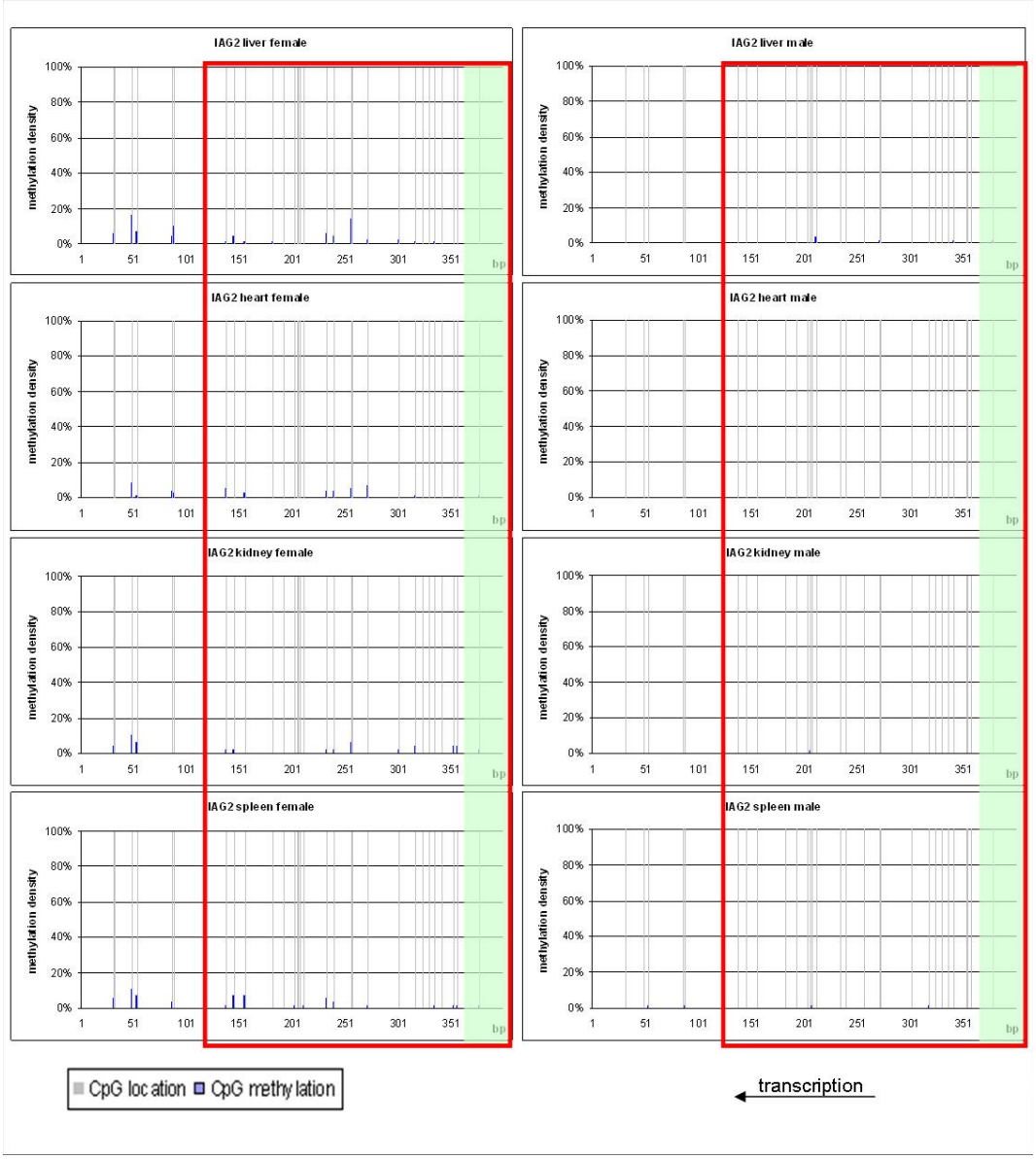

Figure 5.18: CGI methylation profiles of opossum *IAG2* in four tissues. The island region is indicated by the red box and the exon is shadowed in green. Direction of transcription is indicated by arrow. For detailed annotation see Figure 5.12.

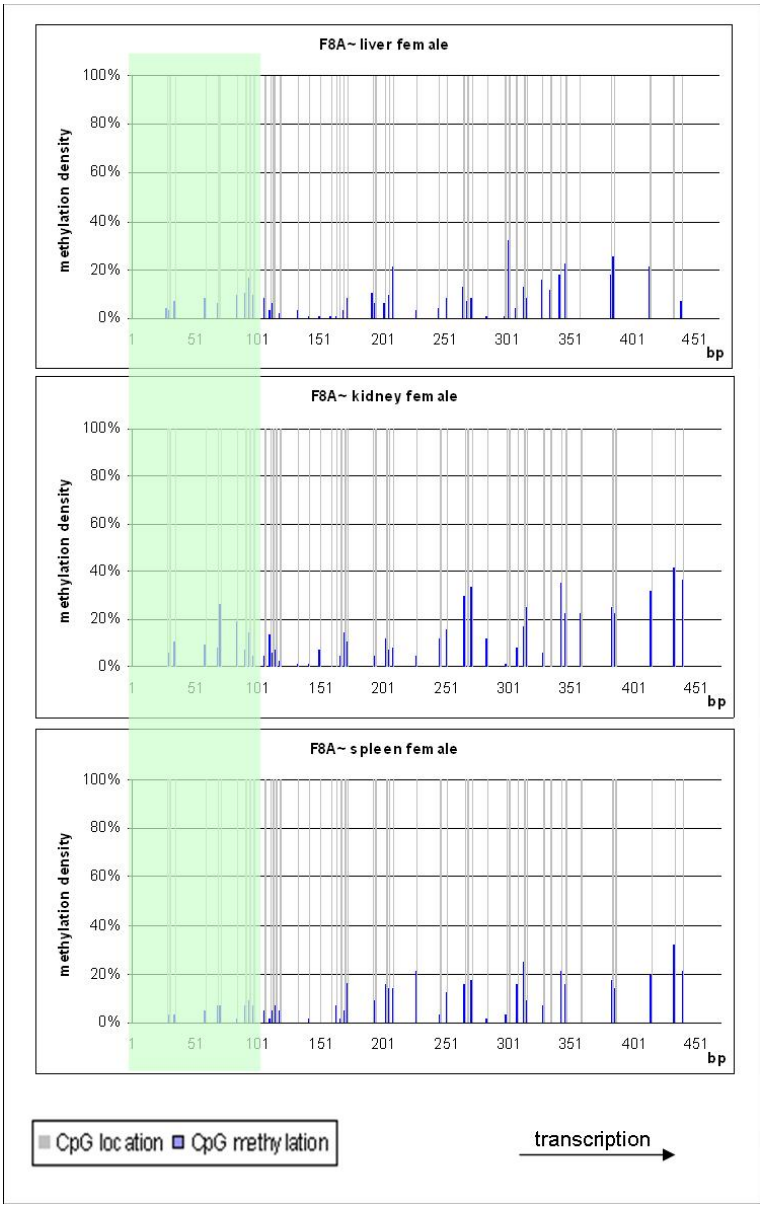

Figure 5.19: CGI methylation profiles of opossum *F8A∼* in three tissues. Bisulphite sequencing was only successful using the female samples. The exon is shadowed in green. Direction of transcription is indicated by arrow. For detailed annotation see Figure 5.12.

5.6 Discussion

Before this study, the only two methylation studies of marsupial X-linked 5' CGIs were both in *G6PD*. Using methylation-sensitive restriction enzymes, Kaslow and Migeon (1987) found no sex difference in methylation of the CGI at the 5' end of *G6PD* in the Virginian opossum. Loebel and Johnston (1993) made the same observation in a distantly related marsupial species, the common wallaroo, using both methylation-sensitive restriction enzymes and bisulphite sequencing analysis. The study described in this chapter greatly extends our knowledge of this area by including almost a third of all CGIs of X-linked protein coding genes in the grey short-tailed opossum. Methylation analysis is described of DNA from female and male opossum using RPMA analysis for 37 CGIs. For eight of these, bisulphite sequencing analysis was performed to gain methylation profiles of single base pair resolution.

In this study, methylation of the 5' CGI of *G6PD* was examined in another marsupial species, the grey short-tailed opossum. Bisulphite sequencing analysis confirmed hypomethylation in both female and male samples in four different tissues, consistent with the previous findings in other marsupial species (Kaslow and Migeon, 1987; Loebel and Johnston, 1993). Similar to the situation in wallaroo, the majority of CpGs were completely unmethylated in all clones sequenced (Loebel and Johnston, 1993). Therefore, if the analysis in this study was also limited to this gene, I would have come to the same conclusion as previous researchers and suggest that methylation may not be a feature of CGIs on the marsupial X chromosome. However, the wealth of other information generated from the current study has presented a different picture.

For a set of seven genes, comprising nearly a fifth of all CGIs examined in this study, clear differential methylation was observed between females and males. Bisulphite sequencing analysis of the CGIs of *FAAH2*, *FGF16*, *IAG2*, *PGK1*, and *ITM2A* confirmed the sex-specific difference. While the male samples were notably hypomethylated, up to half of the clones in each female sample, presumably representing the inactive X chromosomes, were methylated to various extents (Figures 5.6-5.10), just like the CGIs of X-inactivated genes in human and mouse. This is the first observation of female-specific methylation on any marsupial X chromosome. It has demonstrated the risk of relying on only a single observation to derive a general rule.

The presence of higher CGI methylation for X-linked genes in opossum females opens up the possibility that methylation may play a role in maintaining XCI in some marsupial genes. It is noted that the methylation densities in opossum CGIs were much lower than the ones found in human and mouse, but it is not known what level of methylation is needed to silence any gene. Maybe some of the methylation densities observed in this study is already enough to have an effect. It is also possible that a lower level of methylation can stabilise silencing in conjunction with other epigenetic modifications. Having found heavier methylation in females for not one, but a group of genes in opossum, it is highly likely that female-specific methylation will be identified for additional genes in opossum and in other marsupial species. Some of these genes might have high enough methylation to repress gene expression. If methylation does play a role in marsupial XCI, it probably has not become as widespread in marsupials as in eutherians, so that many CGIs still remain unmethylated even on the inactive X.

It has been noted that XCI in marsupials is not always stable and expression

of paternal alleles have been observed in at least one species for all genes studied (reviewed in Cooper *et al.*, 1993). Previous researchers thus proposed that DNA methylation 'locks in' the inactivation state and that lack of CGI methylation on the marsupial X chromosome might be responsible for this leaky XCI (Kaslow and Migeon, 1987). Later evidence confirmed that DNA methylation is a late event in the XCI process (Lock *et al.*, 1987), but it has not been possible to test whether CGI methylation protects a marsupial gene from 'leaky' XCI, since no methylated CGI was recognised before this study. It would therefore be of interest to compare the newly uncovered methylation profiles with the effectiveness of XCI of the relevant genes. Allele-specific expression of three genes, *G6PD*, *HPRT*, and *PGK1*, has been assessed in a related species, *D. virginiana*. Interestingly, the CGI methylation profiles of these genes in *M. domestica* correlates with their paternal allele expression in *D. virginiana*. The CGI of *G6PD* was hypomethylated in both sexes in *M. domestica*; and expression of the paternal allele was detected in tissues as well as in cultured fibroblasts of *D. virginiana* females (Migeon *et al.*, 1989; Samollow *et al.*, 1989). In contrast, CGI methylation was observed here in *M. domestica* female samples for *PGK1*; and no paternal allele expression was found in either tissues or cultured fibroblasts of *D. virginiana* (Samollow 1986, Samollow 1989). *HPRT* has a similar CGI methylation profile to that of *G6PD*. If the correlation observed for *G6PD* and *PGK1* extends to *HPRT1*, expression of the paternal allele in *D. virginiana* females is expected. Although the situation *in vivo* is unknown, it has been demonstrated that the paternal allele of *HPRT* from *D. virginiana* becomes reactivated in cultured fibroblasts, which is indicative of unstable XCI (Migeon *et al.*, 1989).

During the course of this study, Hornecker and colleagues (2007) described the

first study of CGI methylation of *G6PD* and *PGK1* in *M. domestica*. Bisulphite sequencing analysis was performed on both islands in a number of tissues, including liver, which is used in my study. Similar to my observation, they also found the CGI of *G6PD* to be hypomethylated in both sexes. However, their found hypomethylation of the *PGK1* CGI in both sexes whereas my study showed clear female-specific methylation for the same island. There are a number of possible explanations for the discrepancy between the two studies. Firstly, they only assayed a small number of CpGs in a small number of molecules (for *PGK1* in liver, 17 CpGs in 6 molecules, compared with 36 CpGs in 95 molecules in the current study), so might have missed low level methylation by chance. However, their accompanying expression study, in which biallelic expression was found to be common for *PGK1*, does support lack of methylation at this CGI. It is also possible that there is big inter-individual variation in CGI methylation on the opossum X chromosome, and methylation in some individuals is too low to affect expression. The fact that *G6PD* is stably inactivated in some marsupial species despite lack of CGI methylation indicates operation of other maintenance mechanism. Therefore, if methylation plays a role in maintenance of XCI in marsupials, it might form a dispensable layer of 'extra insurance', and allows for greater extents of plasticity.

In recent years there has been increased interest in the role DNA methylation plays in regulating tissue-specific gene expression. In his initial report of CGI, Bird (1985) noted that CGIs remain unmethylated in tissue types and developmental stages where a gene is not expressed, and concluded that it is unlikely for DNA methylation to have a regulatory role in tissue-specific gene expression. More recently, an increasing number of tissue-specific differentially methylated

regions (tDMRs) have been identified and the question has been raised again. A large-scale systematic study by Eckhardt and colleagues (2006), contributing towards the International Epigenome Project, found that only 17% of the analysed genes had tDMRs located in their 5' UTRs and that two-thirds of these tDMRs did not correlate with gene expression. I found the general levels of methylation vary among different tissues for all seven CGIs examined, but not substantially (Figures 5.15-5.19). In contrast, the methylation density distribution patterns remain remarkably stable across different tissue types (Figures 5.15-5.19). Based on these results, methylation of X-linked CGIs in opossum may not play a role in tissue-specific expression control.

This study has provided the first evidence that methylation may have a role in XCI in marsupials, at least for a subset of genes. Greater understanding of the significance of CGI methylation in marsupial XCI will come from further study of expression and methylation patterns of X-linked genes in several different species.