Chapter 3

Comparison of CpG island methylation on the mouse and human X chromosomes

3.1 Introduction

The vertebrate genomes are depleted of the dinucleotide CpG, which is only observed at 0.2-0.25 of the frequency expected from base composition (Russell *et al.*, 1976; Swartz *et al.*, 1962). The majority of the CpGs are methylated at the carbon-5-position of the cytosine (Bird and Taggart, 1980; Ehrlich *et al.*, 1982) and their rarity has been attributed to the tendency for 5mCpG to mutate by deamination to TpG (and CpA in the complementary strand) (Coulondre *et al.*, 1978). However, when genomic DNA from a number of vertebrates was digested with 5mCpG-sensitive restriction enzymes, it was found that about 1% of the genome, the HTF (*Hpa*II tiny fragments) fraction, is highly rich in unmethylated

*Hpa*II and *Hha*I sites (Cooper *et al.*, 1983). Bird and colleagues (1985) then mapped three randomly chosen HTFs to 'islands' of DNA that are CpG-rich and unmethylated in a range of mouse tissues. At the same time, several studies identified unmethylated regions at the 5' end of genes (McKeon *et al.*, 1982; Stein *et al.*, 1983) and sequence analysis revealed that these regions are also CpG-rich (Tykocinski and Max, 1984). Having reviewed the mounting evidence, Bird (1986) proposed to define HTF-like sequences simply by their GC content and CpG frequency, and predicted that such HTF-like islands would be unmethylated. These CpG-rich islands were found near the transcription origin of many genes, especially housekeeping genes (Bird, 1986). This trend was confirmed by Larsen and colleagues, whose extensive sequence analysis study found that all known human housekeeping genes and some tissue-specific genes are associated with CpG islands (CGIs) covering their transcriptional start (Larsen *et al.*, 1992).

CGI methylation has been implicated in the transcriptional silencing of genes. Keshet and colleagues (1985) built constructs of the hamster *aprt* and *tk* genes with different portions methylated, and studied the effects of methylation on their expression in transfected mouse cells. They found that methylation in the 5' promoter region, which was identified by other researchers as a CGI (Lewis, 1986; Tykocinski and Max, 1984), is sufficient to inhibit transcription for both genes. Studies of genes on the X chromosome in mammals provided further support for the inhibitory role of 5' CGI methylation. As an effect of XCI, the same gene presents an active and an inactive copy in the same cell environment, and experimental systems have been developed to distinguish the two: human-rodent somatic cell hybrid containing only the active or inactive human X chromosome, and *Mus musculus* x *M. caroli* mouse, in which the *M. caroli* X chromosome is

always inactivated. With the help of a de-methylation agent, 5-azacytidine, researchers induced re-expression of human *HPRT* and mouse *Hprt* genes from the inactive X chromosome (Graves, 1982; Mohandas *et al.*, 1980). Characterisation of methylation profiles of the 5' region of the human *HPRT* and mouse *Hprt* genes revealed that this region is methylated on the inactive X, but unmethylated on the active and re-activated X in females, and as expected, unmethylated on the single active X in males (Lock *et al.*, 1986; Wolf *et al.*, 1984; Yen *et al.*, 1984). Indeed, hypermethylated 5' CGIs, although very rare in mammalian genomes, are commonly seen on the inactive X chromosome (Tribioli *et al.*, 1992). In healthy individuals, hypermethylated 5' CGIs have also been identified for imprinted genes, which only express from one chromosome depending on the parental origin (Stöger *et al.*, 1993; Tremblay *et al.*, 1995). For example, methylation of the 5' region was found in sperms but not oocytes for the maternally expressed *H19* gene in mouse, and this pattern is maintained in the pre-implantation embryo (Tremblay *et al.*, 1995). Disruption of normal CGI methylation states can have serious health consequences. A systematic study of methylation states of more than a thousand CGIs in nearly a hundred primary human tumours found tumour-type-specific aberrant methylation patterns (Costello *et al.*, 2000). Using a candidate gene approach, another report showed that one or more of the 12 key cancer genes investigated had promoter hypermethylation in each of the 15 major tumour types studied, and there exist unique patterns for different tumour types (Esteller *et al.*, 2001). The exact mechanism by which CGI methylation induces transcriptional silencing remains unclear but it is possible that methylation at 5' CGI prevents initiation of transcription by blocking access of key DNA-binding factors (Comb and Goodman, 1990).

In the search for regulatory factors of XCI, CGI methylation, which provides effective in *cis* silencing and can be stably inherited through cell division, became an attractive candidate. Initially, Riggs (1975) proposed a model in which DNA methylation regulates the activity of the X inactivation centre and thus plays a role in initiating XCI. However, the discovery that the opossum *G6PD* gene, X-linked and dosage-compensated like its mouse companion, is hypomethylated at its 5' CGI even in females led researchers to believe that DNA methylation stabilises rather than initiates the process (Kaslow and Migeon, 1987). Study of early embryonic development in mouse confirmed that methylation takes place after the initiation of XCI, supporting the maintenance role of CGI methylation in XCI (Lock *et al.*, 1987).

3.1.1 Identification of CGIs

In their initial description of the HTF islands, Bird and colleagues (1985) found several distinct features of these islands: they are GC-rich, not deficient of CpG (with similar frequencies of CpG and GpC), and unmethylated. These islands, later known as CpG islands, were identified by isolation and analysis of small fragments generated by methylation-sensitive restriction enzyme digestion (Antequera and Bird, 1993; Bird *et al.*, 1985). In his landmark paper, Bird (1986) formulated a simple test for HTF-like sequences based on CpG frequency: if a sequence has GC content of over 50% and has similar numbers of CpG and GpC, both near the expected frequency from the sequence composition, the region is HTF-like and is predicted to be unmethylated. Using this test, he found that CGIs are frequently located at the 5' region of genes, most of which are house-

keeping (Bird, 1986). This strong association with gene was confirmed in larger studies of more CGIs and made CGI a popular marker in the search of novel gene sequences (Antequera and Bird, 1993; Larsen *et al.*, 1992).

With the rapid accumulation of sequence data in public databases, computer algorithms were soon developed to search for CGIs. The first algorithm was proposed by Gardiner-Garden and Frommer (1987) who defined a CGI as a GC-rich region (GC content >50%) greater than 200 bp in length and having an observed CpG / expected CpG ratio (Obs_{CpG}/Exp_{CpG}) of greater than or equal to 0.6. Using these criteria, they screened vertebrate genomic sequences containing genes transcribed by RNA polymerase II in GenBank (a database of all publicly available DNA sequences). Most of the CGIs identified are associated with the 5' end of genes and most of the 5' CGIs extend well into genes. Consistent with the other investigations, they also found all housekeeping genes in their survey to have 5' CGIs. The CGIs are highly CpG-rich, with average GC content over 60 and Obs_{CpG}/Exp_{CpG} over 0.8. In addition, although 200 bp was used as the cut-off size of CGI, most of the CGIs identified are greater than 500 bp.

These criteria have been widely applied in prediction of CGIs, but many repetitive elements, for example human Alus, may also fulfil the criteria, leading to a high rate of false positive predictions (Schmid, 1998). In an attempt to more accurately identify CGIs, more stringent parameters than the original Gardiner-Garden and Frommer definition were favoured in some more recent studies. The Vertebrate Genome Annotation (Vega) database, which contains high quality manual annotation of finished vertebrate genome sequences, uses a higher cut-off size of 400 bp, optimised using known CGI sequences (Gos Micklem, personal communications). A slightly stricter set of parameters were proposed by Takai

and Jones (2002) after analysing CGIs in human chromosomes 21 and 22. They defined a CGI as a stretch of DNA of greater than 500 bp length with a GC content equal to or greater than 55% and Obs_{CpG}/Exp_{CpG} of 0.65 or more. CGIs identified by the above two methods are more likely to be associated with the 5' regions of genes.

3.1.2 Approaches to studying CGI methylation

As DNA methylation is recognised as an important component in gene regulation, the field of methylation is attracting growing interest of researchers around the world, especially in cancer biology. A great range of methodologies to study the phenomenon have been developed in the past two decades. To date, most efforts in resolving states of DNA methylation were enabled by methods based on restriction enzyme digestion or bisulphite conversion. The various methods reviewed below are summarised in Figure 3.1.

In the early days of DNA methylation research, genomic DNA was digested with methylation-sensitive restriction enzymes and Southern blotting was used to provide readout of methylation status. In the first studies to examine *de novo* methylation of 5' CGI as a potential mechanism of oncogenesis, DNA samples from primary retinoblastomas and unaffected controls were digested with *Sac*I and the methylation-sensitive *Hpa*II, and a restriction map was established by Southern hybridisation of a probe covering the 5' region of the retinoblastoma gene (*RB*) (Greger *et al.*, 1989; Sakai *et al.*, 1991). If all *Hpa*II sites are methylated, the probe would detect a large fragment bounded by SacI sites, while smaller fragments would indicate some *Hpa*II sites being unmethylated, thus

cleaved. Using methylation-sensitive restriction enzymes with different recognition sites, a number of CpGs were assayed, and hypermethylation at the *RB* CGI was found in one out of 21 (Greger *et al.*, 1989) and five out of 56 retinoblastomas (Sakai *et al.*, 1991). Employing a similar approach, abnormal hypermethylation in tumours were identified at the 5' CGI of more tumour-suppressor genes, including the von Hippel-Lindau (*VHL*) gene (Herman *et al.*, 1994) and the *p16* gene (Merlo *et al.*, 1995). This approach is simple and powerful for examining DNA methylation patterns, but only a subset of CpGs are suitable for assaying, limited by the variety of methylation-sensitive restriction enzymes. Furthermore, using Southern hybridisation as a detection method relies on availability of suitable probes and demands large quantity of test DNA, which may not always be accessible.

The advent of the polymerase chain reaction (PCR) has provided a powerful tool for molecular biology and several methods combining restriction enzymes with PCR have been employed in methylation analysis. The sensitivity of PCR allowed Singer-Sam and colleagues (1990) to study the timing of DNA methylation on the X chromosome in early mouse embryos, when the embryo consists of fewer than a thousand cells. DNA extracted from individual mouse embryos was digested with *Hpa*II and used as template in PCR amplification with a pair of primers enclosing an *Hpa*II site in the 5' CGI of the X-linked *Pgk1* gene. A PCR product is only generated when the site is protected by methylation. They found that this site became 40% methylated in females at about the time of XCI (Singer-Sam *et al.*, 1990a). An alternative to the candidate approach is methylation fingerprinting, which is popular for studying aberrant methylation patterns in tumours (Frigola *et al.*, 2002; Gonzalgo *et al.*, 1997; Huang *et al.*, 1997). In

methylation-sensitive arbitrarily-primed PCR (MS-AP-PCR), for example, DNA from tumour and control samples are digested with *Rsa*I and *Hpa*II, and arbitrary primers in combination with low stringency reaction conditions are used to amplify random sequences containing methylated sites (Gonzalgo *et al.*, 1997). The primers were designed with a 3' end complementary to the recognition sequence of *Hpa*II to enhance the chance that an amplified product contains *Hpa*II sites, and DNA digested with the methylation-resistant isoschizomer of *Hpa*II, *Msp*I, was used as a control to distinguish between amplified products as a result of methylation at *Hpa*II sites and lack of *Hpa*II sites. These PCR-based restriction techniques are simple, fast, and sensitive, but can still only analyse CpGs located in methylation-sensitive restriction enzyme recognition sites.

The second class of methods 'fixes' the methylation pattern in DNA sequence by selective deamination of unmethylated cytosine to uracil using sodium bisulphite (Frommer *et al.*, 1992). Bisulphite-converted DNA can then be sequenced to give a detailed methylation profile of individual CpG dinucleotides. Bisulphite sequencing gives the most comprehensive methylation profiles and remains the gold standard in DNA methylation studies, but it is also the most laborious and thus limited in the scale of study. This technique is discussed in more detail in Chapter 4. PCR-based alternatives to sequencing have lower resolution but are more rapid: methylation-specific PCR (MSP) primers or unmethylated-DNA-specific PCR (USP) primers can be used to preferentially amplify hyper- or hypomethylated regions (Herman *et al.*, 1996; Kawakami *et al.*, 2004). Where accurate quantification of methylation of individual sites is required, methylation-sensitive single nucleotide primer extension (MS-SNuPE), which uses radioactive C or T, can alternatively be applied (Gonzalgo and Jones, 1997).

In recent years, microarrays have gained popularity as a high-throughput detection method in DNA methylation research. Methylation-specific oligonucleotide (MSO) microarrays were designed to contain sets of oligonucleotides that distinguish between bisulphite-converted TpGs and methylation-protected CpGs, and were used to screen methylation at selected CpGs in tumour samples (Adorján *et al.*, 2002; Gitan *et al.*, 2002). Alternatively, hyper- or hypomethylated fractions may be enriched prior to hybridisation to microarrays. In differential methylation hybridisation (DMH), fragmented genomic DNA are ligated to adaptors followed by digestion with methylation-sensitive restriction enzymes, and the intact hypermethylated fragments are enriched by PCR amplification with primers specific to the adaptors (Huang *et al.*, 1999). Similarly, hypomethylated proportions can be enriched if the genomic DNA is initially fragmented by methylation-sensitive restriction enzyme, and the methylated fragments are eliminated by an enzyme that cleaves DNA containing methylcytosine (Schumacher *et al.*, 2006). More recently, a promising new technique has been developed to enrich methylated DNA by immunoprecipitation. In this methylation DNA immunoprecipitation (MeDIP) approach, genomic DNA is fragmented, denatured, and incubated with a methylcytosine-specific antibody, and methylated DNA fragments are recovered through immunoprecipitation (Weber *et al.*, 2005). MeDIP-enriched DNA and control genomic DNA can be labelled with different fluorescent dyes and hybridised to the same microarray, and methylation levels can be measured by calculating the ratio between the two signals. Combining MeDIP with two different arrays, a whole genome tiling array and a CGI array, Weber and colleagues (2005) confirmed that the inactive X is globally hypomethylated but hypermethylated at CGIs. These array-based methods are very powerful high-throughput

methods that yield a great amount of information in a single experiment and are ideal for whole genome profiling, but they require extensive optimisation and careful statistical analysis, and are limited in resolution.

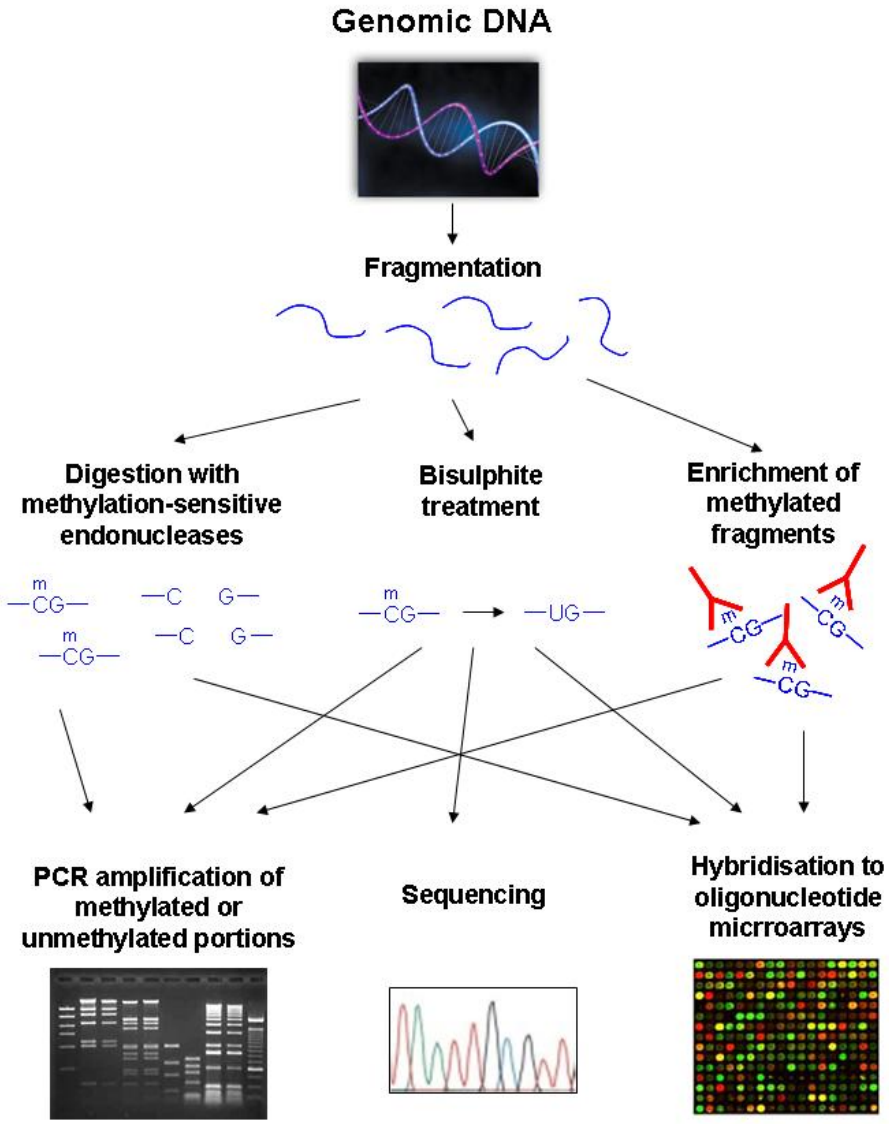

Figure 3.1: Approaches to studying DNA methylation.

3.1.3 Aims

The work described in this chapter had two aims: the first was to assess the use of CGI methylation state as an indication of a gene's XCI status in human; the second was to extend this analysis to the orthologous genes in the mouse in order to see if there are gross differences between the two species that might explain variability in escape from XCI.

3.2 Identification of target CpG islands

3.2.1 Region of interest

Recent studies revealed that more than 15% of genes on the human X chromosome escape from XCI (Carrel and Willard, 2005), but the distribution of such 'escapees' is not uniform, reflecting their evolutionary history. As described in Chapter 1, the human X chromosome can be divided into five evolutionary strata (Lahn and Page, 1999; Ross *et al.*, 2005): S1 and S2 in the XCR, which is common to X chromosomes in all mammalian groups (Graves, 1995); and S3, S4, S5 in the XAR, which was translocated from an autosome to the sex chromosomes in the eutherian lineage (Figure 1.2) (Graves 1995). These strata show an increasing level of sequence identity between X and Y chromosomes from S1 (lowest) to S5 (highest). At the distal end of the XAR, beyond S5, lies PAR1, where X and Y still recombine. Escapees are very rare in S1 and S2, but constitute a good proportion of genes in S3 and make up the entire set of genes tested in S4, S5 and PAR1, supporting a differential recruitment of X-linked genes into XCI (Carrel and Willard, 2005).

For the purpose of this study, the candidate region needs to contain a large number of inactivated and escapee genes and have orthologues on the mouse X chromosome, so any difference of XCI statuses between human and mouse are most likely to be seen. The first requirement limits the candidate region to XAR. In the XAR, most genes in S4 and S5 do not have orthologues on the mouse X chromosome, and comparison between the human and chicken genomes suggests that these genes were present in the original XAR and were lost in the mouse lineage (Ross *et al.*, 2005). Therefore, I have initially chosen to study genes within S3.

3.2.2 Identification of target human and mouse CGIs

Information about human genes used in this study was obtained from the Vega database (v19, based on the NCBI36 build) for the high quality annotation. There were not enough mouse genes annotated in Vega at the time, so the Ensembl annotation (v38, based on the NCBIM35 build) was used for the mouse genes. CGIs were predicted by Val Curwen using the cpg program (Gos Micklem, unpublished) with the following parameters: a minimum length of 400 bp, a minimal $\mathrm{Obs_{CpG}/Exp_{CpG}}$ ratio of 0.6 and a minimal GC content of 50%.

On the human X chromosome, 147 protein-coding genes were found in S3, 91 (62%) of which had a predicted CGI close to the 5' end. Three pairs of genes, transcribed in opposite directions, had a CGI overlapping with the first exons of both genes, so 88 unique CGIs were available to study. CGI sizes ranged between 400 bp and 3 kb, with an average size of about 1 kb. Two CGIs were upstream from the transcription start site of the genes. The remaining CGIs all overlapped

exon1, the vast majority extending into intron1.

Orthologous genes in mouse for each of the 147 human S3 genes were identified based on Ensembl annotation. Putative orthologues were identified for 131 genes, of which 126 were also X-linked in mouse. Fifty-four of the 126 genes had a predicted 5' CGI. Only one gene, *CXorf23*, has a human orthologue without 5' CGI, and is not suitable for this study. The other 53 genes were all included in this study. CGI sizes ranged between 400 bp and 2 kb, with an average size of about 900 bp. Three CGIs were entirely upstream of the gene start. All the other CGIs cover the entire exon1 and extend into intron 1.

In total, 88 human CGIs and 53 mouse CGIs were selected for this study. Their sequences, with 100 bp flanking sequence on each side, were extracted from the appropriate database (Vega for human, Ensembl for mouse) via Ensembl Perl API using a Perl script (Appendix V).

3.3 Assessing CpG island methylation on mouse and human X chromosomes by Restriction-PCR Methylation Assay (RPMA)

For the purpose of rapidly assaying a large number of CGIs initially, I decided to use a restriction-PCR approach, which I have termed RPMA (restriction PCR methylation assay). RPMA is similar to the method used by Singer-Sam and colleagues (1990a; 1990b) but includes DNA digested by the methylation-resistant *Msp*I to serve as a control for successful digestion. Female and male genomic DNAs are digested with *Msp*I or its methylation-sensitive isoschizomer *Hpa*II,

and then the digested DNA is tested for its ability to act as template for amplification of PCR products containing multiple *Msp*I/*Hpa*II sites from predicted CpG islands. This method was employed by Jegalian and Page (1998) to assay CGI methylation of four X-linked genes in 19 species. This is the only large multi-species XCI study to date, from which the authors concluded that CGI methylation accompanies XCI in a wide range of eutherians and that XCI evolved on a gene-by-gene or cluster-by-cluster basis.

Human and mouse genomic DNA was extracted from female and male cultured fibroblast cells, digested with *Hind*III, *Hpa*II, or *Msp*I, and then used as template for PCR amplification. Since the predicted amplicons contained no *Hind*III cleavage sites, the *Hind*III-digested DNA was expected always to support amplification of PCR products, and thus acted as a positive control. By contrast, DNA digested with the methylation-resistant *Msp*I was expected always to be cleaved between the PCR primers and to fail to yield any PCR products. This allowed control for effectiveness of restriction digestion. Cleavage by the methylation-sensitive *Hpa*II is blocked by methylation of the CpG in the CCGG recognition site, and so a PCR product was only expected when all the CpG dinucleotides in the amplicon were methylated. An outline of the experimental strategy and the expected agarose gel electrophoresis pictures for hypo- and hypermethylated CGIs are shown in Figure 3.2.

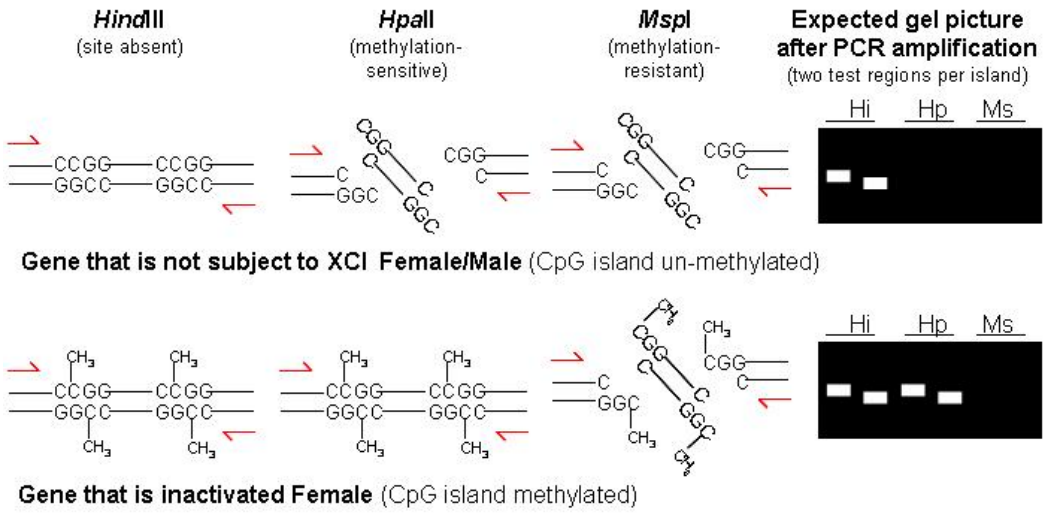

Figure 3.2: Restriction-PCR Methylation Assay (RPMA). Female and male genomic DNAs digested with *Hind*III (*Hi*, positive control), *Hpa*II (*Hp*), or *Msp*I (*Ms*, negative control) were used as templates for PCR amplification. A strong band in the female *Hp* lane (comparable to the female *Hi* lane) is expected if the CGI is hypermethylated and the gene undergoes XCI, whereas absence of product in the female *Hp* lane is expected when the CGI is hypomethylated and the gene escapes from XCI. Wherever possible, two regions were tested in each CGI to give better representation.

3.3.1 Optimisation of experimental conditions

PCR conditions were optimised to ensure successful and specific amplification of the GC-rich sequences. A set of reaction conditions have been previously optimised for amplification of GC-rich region: 5% DMSO was included to prohibit secondary structure formation and to help destabilise the double-helix structure; 50% of the dGTP was substituted with 7-deaza-dGTP to prevent compressions; and a mixture of Advantage^{TM} 2 polymerase mix and AmpliTaq® polymerase was found to facilitate PCR amplification of a wide range of product sizes (credit to Christine Burrows and Dr Tamsin Eades). These PCR conditions were tested on four published primer pairs (Jegalian and Page 1998) using human and mouse

genomic DNA digested with *Hind*III, *Hpa*II, or *Msp*I, but it failed to amplify the *ZFX* CGI in both human and mouse, and gave only very weak signals for the CGIs of *SMCX* and *RPS4X*, again in both human and mouse.

I then tested the effects of a higher primer concentration, a hot-start step of heating the template and primers at 100 °C for 5 minutes, and an alternative cycling profile as used by Jegalian and Page (1998) (with a higher annealing temperature). The higher primer concentration and the hot-start step were found to improve PCR amplification. The alternative cycling profile reduced appearance of non-specific products but also gave weaker signals of the desired products. To find out if the PCR results were affected by the restriction enzyme digestion step, these experiments were repeated on undigested genomic DNA and digested DNA that underwent a heat inactivation of restriction enzymes at 65 °C for 20 minutes. Heat inactivation reduced appearance of non-specific products, but the signals were not as clean and strong as when using undigested DNA as template. On the base of these results, a higher primer concentration and the hot-start step were incorporated into my original PCR conditions for subsequent experiments.

There are two possible ways that the restriction enzyme digestion step may affect PCR amplification: template availability may be reduced by over-digestion; or chemicals present in the digestion reaction may interfere with PCR conditions. To address the former issue, restriction enzyme digestion conditions were optimised to ensure complete but not over digestion of genomic DNA. Human DNA was digested with *Hind*III, to test for over-digestion, or *Msp*I, to test for incomplete digestion, at 100 units of enzyme per μg DNA concentration for different lengths of time. The PCR results from templates generated by 4-, 8-, or 16-hour digestion were indistinguishable (Figure 3.3a) for all four amplicons, and

for convenience a 16-hour (overnight) digestion was performed for all subsequent experiments. Next the human DNA was digested with *Hpa*II or *Msp*I at four different enzyme concentrations, 10, 20, 50, or 100 units of enzyme per μg DNA, to test for incomplete digestion. In all cases, no PCR products at expected amplicon size were detected, but 10 and 20 units of enzyme per μg DNA resulted in a number of non-specific products for the *SMCX* CGI, so the optimal concentration was decided to be 50 units of enzyme per μg genomic DNA (Figure 3.3b). To examine the possible negative effects of chemicals present in the digestion reaction on PCR, the digested DNA was purified from heat-inactivated digestion reaction by organic solvent extraction followed by ethanol precipitation. PCR using the purified digested DNA gave comparable results with PCR using undigested DNA, so the purification step was used in all subsequent experiments.

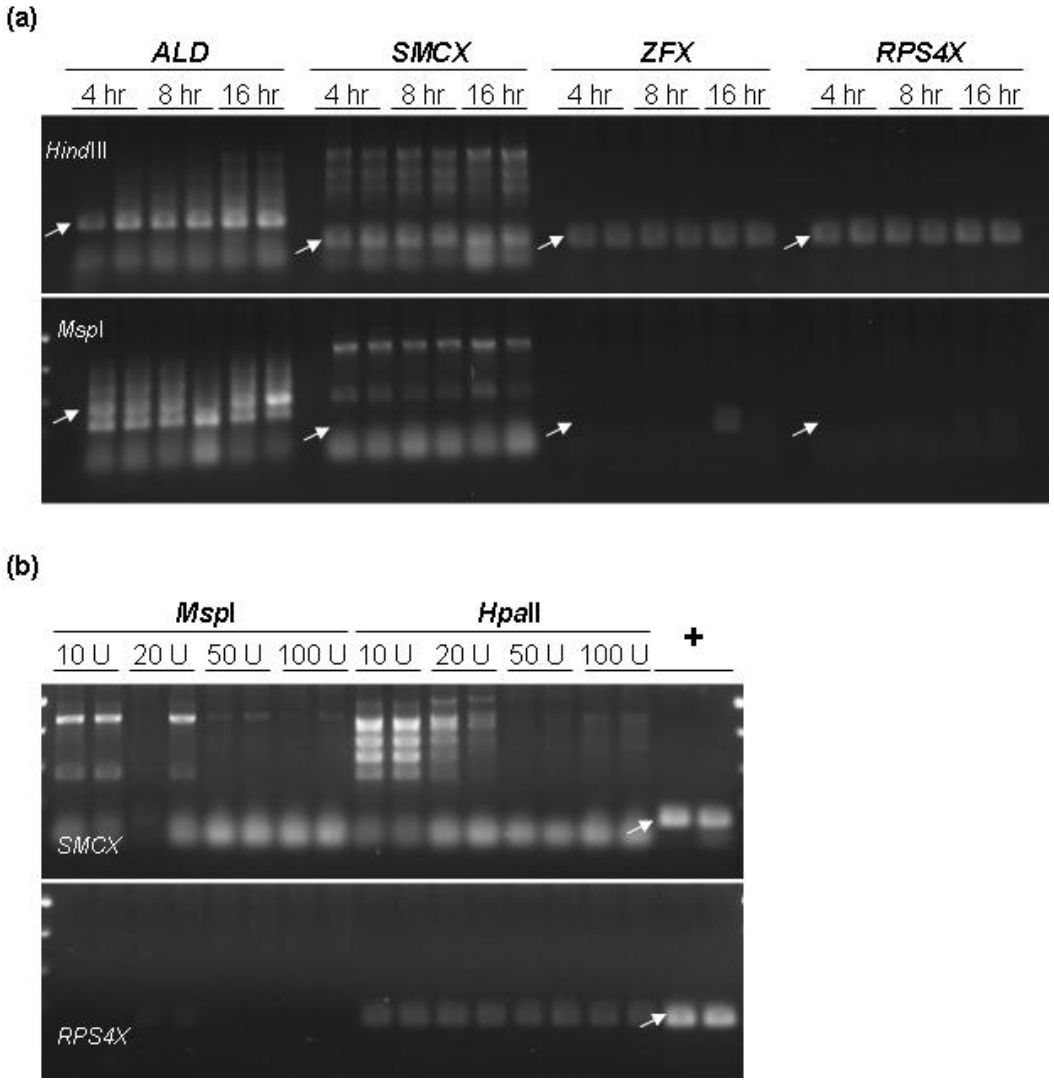

Figure 3.3: Optimisation of restriction enzyme digestion conditions. Expected product sizes are indicated by arrows. a) PCR results from templates generated by 4-, 8-, or 16-hour digestion. A PCR product is expected after *Hind*III but not *Msp*I digestion. b) PCR results from templates generated by digestion with 10, 20, 50, or 100 units of enzyme per μg DNA. '+' are positive controls. No products are expected from either digestion.

3.3.2 Primer design

Primers were designed to amplify 150-250 bp products containing multiple *Hpa*II/*Msp*I cleavage sites (CCGG) lying within the predicted CGI. Since single cleavage site may be protected by rare methylation, leading to over-representation of methylation, and a high number of sites may result in digestion despite hypermethylation, leading to under-representation of methylation, so I aimed at including 2-4 cleavage sites in each amplicon. The amplicon sizes were designed to be larger than those used in Jegalian and Page's study (1998) to ensure good separation of PCR product and primer-dimer in gel electrophoresis. In order to increase the predictive power of the method, two pairs of primers enclosing regions with no overlapping CCGG sites were designed for each CGI. For three of the 88 human CGIs and four of the 53 mouse CGIs, only one pair of primers could be designed. Primers were tested in PCR amplification using male genomic DNA. Only three pairs of human primers failed the test, one of which is the only pair of primers available for a CGI. The expected products were successfully amplified for all other human primers and all mouse primers. Therefore in total, 87 human CGIs and 53 mouse CGIs were possible to assay. All primer combinations, their predicted amplicon sizes, and the number of enclosed CCGG sites are listed in Appendix I.

3.3.3 Analysis of CGI methylation on the human X chromosome

The 87 human CGIs were all successfully assayed in female and male human fibroblasts. Of these, 82 CGIs had data for both amplicons and five others had

results from a single amplicon.

The CGIs on the single X chromosome in male cells are expected to be unmethylated, like the CGIs on autosomes, and serve as a convenient indicator of the reliability of the assay. Indeed, the vast majority of CGIs showed a pattern consistent with hypomethylation in male samples, as expected (Table 3.1, Figure 3.4). For 58 (67%) CGIs, no PCR products were amplified using the *Hpa*II digested material, despite strong PCR amplification from the *Hind*III digested control. Another 16 (18%) CGIs had a complete lack of signal in the *Hpa*II lane for one amplicon, and a faint signal in the *Hpa*II lane for the other amplicon, still in support with overall hypomethylation. Six CGIs had a faint signal in the *Hpa*II lane for both amplicons, possibly indicating low level methylation. Only Seven (8%) CGIs had a signal in the *Hpa*II lane that is comparable with the signal from the *Hind*III control. These are discussed in more detail below. In contrast, many CGIs showed a pattern consistent with hypermethylation in female, confirming that the assay has the power to detect hypermethylation (Table 3.1, Figure 3.4). These observations indicate that the methylation state inferred from the assay data were reliable.

The CGIs that are hypomethylated in male can be divided into three categories according to their methylation patterns in the female samples.

Thirteen (15%) CGIs were clearly hypomethylated in both female and male samples, as demonstrated by the lack of signal in the *Hpa*II lane for both amplicons of each CGI (Figure 3.4, a-c).

Forty-nine (56%) CGIs showed female-specific methylation in at least one amplicon (Figure 3.4, d-f). In the male samples, at least one amplicon had no signal in the *Hpa*II lane, and the other amplicon was either absent (3 cases), or

had no signal (34 cases) or a faint signal (15 cases) in the *Hpa*II lane. In the female samples, 15 CGIs were clearly hypermethylated at both amplicons, each having a strong signal in the *Hpa*II lane indistinguishable from the signal from the *Hind*III control. The lack of signals from the *Msp*I control eliminated the possibility of incomplete digestion. An additional 22 CGIs had one amplicon with the similar clear hypermethylation pattern, and the other amplicon was either absent (3 cases), had a strong signal in the *Hpa*II lane but also a faint signal from the *Msp*I control (10 cases), suggesting some incomplete digestion, or had a faint or no signal in the *Hpa*II lane. For the other 12 CGIs, one or both amplicons had a strong signal in the *Hpa*II lane but also a faint or very faint signal from the *Msp*I control. In one case (*RP11-450P7.3*), one amplicon is hypermethylated in female but hypomethylated in male, and the other amplicon is hypermethylated in both sexes.

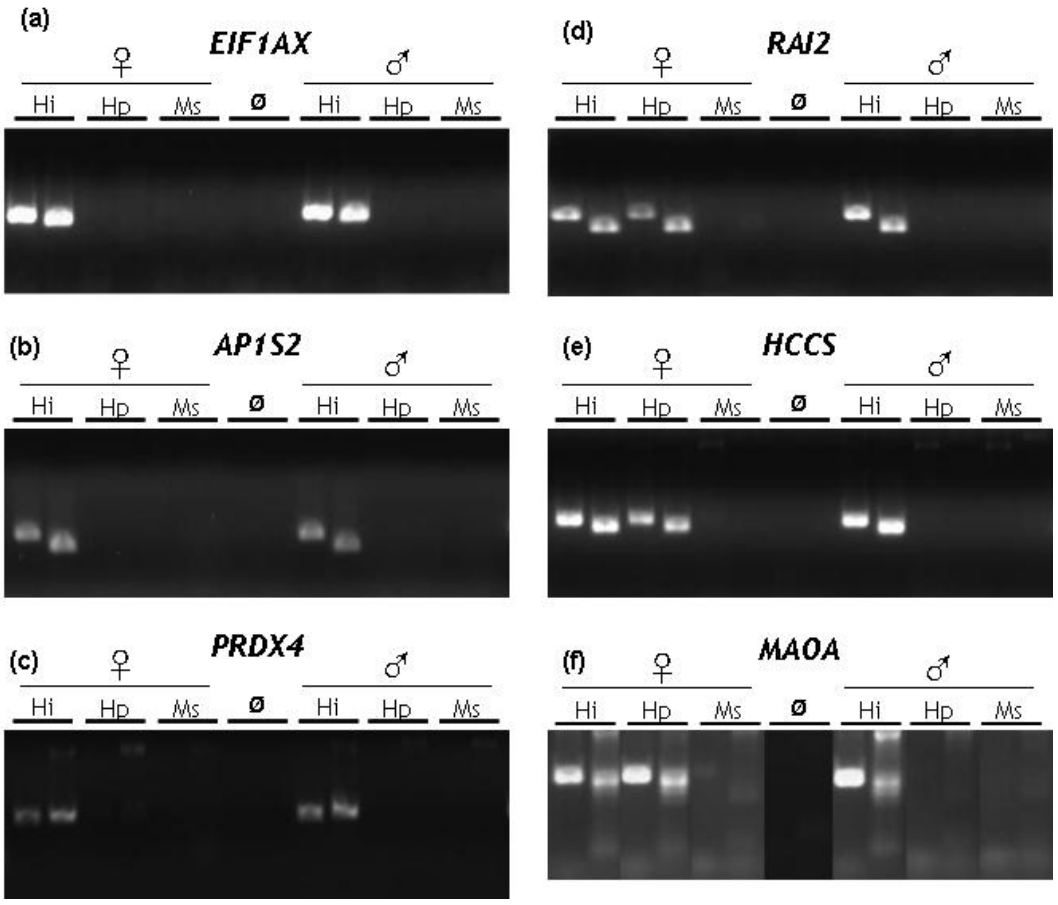

Figure 3.4: RPMA results of human CGIs with hypomethylation or female-specific methylation. (a-c) are examples of CGIs with hypomethylation in both female and male samples. (d-f) are examples of CGIs with female-specific methylation. Lanes in (f) have been re-organised as the original loading pattern had the two amplicons on separate gels.

In 19 cases (22%), the methylation state of the CGI was less well defined. In these cases faint bands were observed in the female *Hpa*II lane (Figure 3.5, a-c). The RPMA approach could be indicating an intermediate state of methylation of these CGIs but would not be able to distinguish between the two possible situations, which are a) a small percentage of heavily methylated CGIs, and b) an intermediate level of methylation in a higher percentage of CGIs. These cases

can be further divided into two broad categories: for eight CGIs, one amplicon was hypomethylated but the other was difficult to score (Figure 3.5a); for another 11 CGIs, the methylation state could not be decided for any amplicon (Figures 3.5, b-c). RPMA was repeated for most of the intermediate cases and the results were either identical or very similar, indicating that this is not an experimental artefact but an accurate representation of an intermediate state.

In the remaining six cases, both male and female CGIs are hypermethylated (Figure 3.5d).

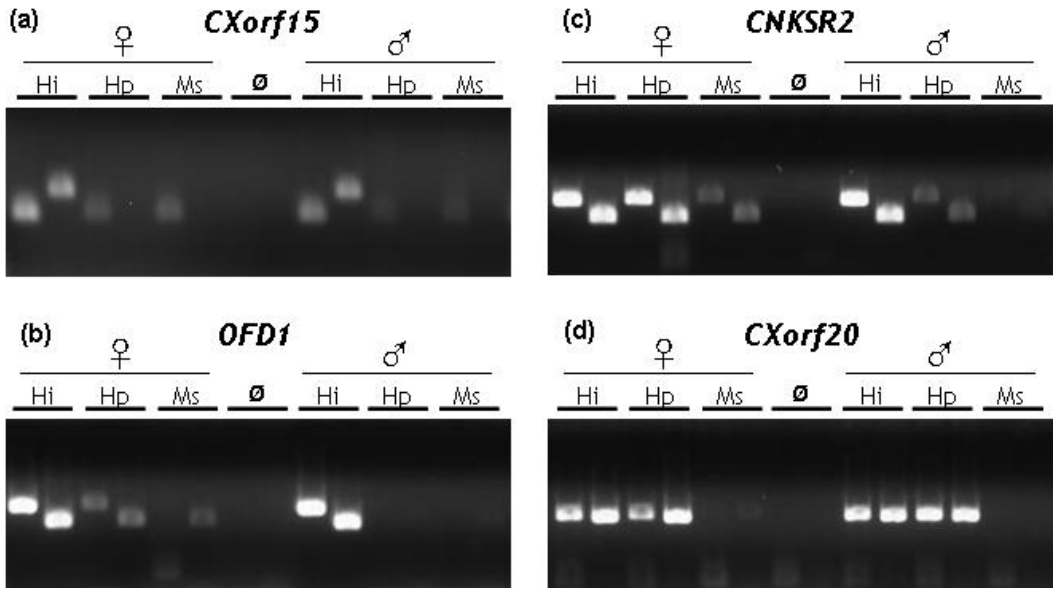

Figure 3.5: RPMA results of human CGIs with other methylation patterns. (a-c) are examples of CGIs for which the methylation state is difficult to define. (d) is an example of hypermethylation in both female and male samples.

Table 3.1: RPMA results of human CGIs. '+' = strong band; 'f' = faint band; 'vf' = very faint band; '-' = no visible band. In the 'XCI state' columns, the 'prediction' column contains XCI status prediction based on RPMA results, and the 'Lit' column contains XCI status recorded in Carrel and Willard (2005), where the numbers of somatic cell hybrids retaining X_i that showed gene expression is presented. XCI status is colour-coded: pink for X-inactivated, violet for escapee, and yellow for undetermined. One CGI is shared by each pair of the following: *RP1-93D11.1* and *RP13-928P6.1*, *RP13-314C10.2* and *RP11-793H5.3*, *OFD* and *TRAPPC2*.

	CGI	RPMA primer pair 1			RPMA primer pair 2			Female-	XCI state	
Gene	size	CCGG	Female	Male	CCGG	Female	Male	specific		
	(bp)	sites	Hi Hp Ms	Hi Hp Ms	sites	Hi Hp Ms	Hi Hp Ms	methylation?	prediction	Lit.
Female-specific methylation in at least one amplicon										
HCCS	481	5	+ + -	+ - -	2	+ + -	+ - -	Y	i	0 / 9
MSL3L1	1151	3	+ + -	+ - -	6	+ + -	+ - -	Y	i	3 / 9
PDZK10	669	3	+ + -	+ - -	3	+ + -	+ - -	Y	i	2 / 9
RP1-93D11.1	742	2	+ + -	+ - -	4	+ + -	+ - -	Y	i	0 / 9
RAI2	1260	6	+ + -	+ - -	3	+ + -	+ - -	Y	i	0 / 9
SCML2	1250	2	+ + -	+ - -	2	+ + -	+ - -	Y	i	0 / 9
MBTPS2	612	4	+ + -	+ - -	2	+ + -	+ - -	Y	i	0 / 9
ACOT9	539	3	+ + -	+ - -	5	+ + -	+ - -	Y	i	2 / 9
PDK3	930	2	+ + -	+ - -	2	+ + -	+ - -	Y	i	0 / 9
POLA	653	2	+ + -	+ - -	3	+ + -	+ - -	Y	i	0 / 9
TSPAN7	989	3	+ + -	+ - -	2	+ + -	+ f -	Y	i	1 / 9
MID1IP1	1857	2	+ + -	+ - -	3	+ + -	+ - -	Y	i	0 / 9
CHST7	1847	3	+ + -	+ - -	4	+ + -	+ - -	Y	i	0 / 9
PHF16	1417	5	+ + -	+ - -	4	+ + -	+ - -	Y	i	0 / 9
RGN	517	3	+ + -	+ - -	2	+ + -	+ - -	Y	i	1 / 9
FAM51A1	565	2	+ + -	+ f -	5	+ - -	+ - -	Y	i	9 / 9
PDHA1	752	2	+ + -	+ - -	7	+ - -	+ - -	Y	i	0 / 6
PTCHD1	2496	3	+ + -	+ - -	3	+ - -	+ - -	Y	i	N/A
BCOR	662	2	+ - -	+ - -	1	+ + -	+ f -	Y	i	2 / 9
SLC9A7	955	2	+ + -	+ - -	7	+ - -	+ - -	Y	i	1 / 9
PRPS2	1176	2	+ f -	+ - -	3	+ + -	+ - -	Y	i	0 / 5
ATP6AP2	965	1	+ + -	+ - -	2	+ f -	+ - -	Y	i	1 / 9
CASK	1151	1	+ + -	+ - -	3	+ f -	+ - -	Y	i	0 / 6
TMSB4X	1491	4	+ vf -	+ - -	1	+ + -	+ f -	Y	i	N/A
KLHL15	1258	5	+ + -	+ - -				Y	i	0 / 9
LANCL3	1077	3	+ + -	+ - -				Y	i	1 / 9
RPGR	977	8	+ + -	+ - -	1	fail	fail	Y	i	0 / 6
RP11-393H10.2	1146	2	+ + vf	+ vf -	3	+ + -	+ - -	Y	i	0 / 9
RPS6KA3	2683	3	+ + -	+ - -	2	+ + vf	+ - -	Y	i	0 / 9
RP11-450P7.3	3047	3	+ + -	+ - -	2	+ + f	+ + f	Y	i	N/A
RP13-314C10.2	724	3	+ + -	+ - -	3	+ + f	+ f -	Y	i	N/A
ARX	1047	2	+ + -	+ - -	2	+ + f	+ f -	Y	i	N/A
PRRG1	1087	1	+ + vf	+ vf vf	1	+ + -	+ - -	Y	i	0 / 9
XK	614	2	+ + f	+ f vf	4	+ + -	+ - -	Y	i	N/A
TCTE1L	676	3	+ + -	+ - -	2	+ + f	+ - -	Y	i	0 / 9
MAOA	1115	2	+ + vf	+ - -	5	+ + -	+ - -	Y	i	9 / 9
ZNF673	588	1	+ + f	+ - -	1	+ + -	+ - -	Y	i	0 / 9
ARHGAP6	2210	4	+ f -	+ - -	2	+ + vf	+ - -	Y	i	1 / 9
PIGA	638	2	+ + vf	+ - -	5	+ f -	+ - -	Y	i	0 / 9
REPS2	868	2	+ + +	+ vf vf	2	+ + vf	+ - -	Y	i	2 / 9
NHS	2657	3	+ + vf	+ - -	2	+ + vf	+ - -	Y	i	3 / 9
SCML1	1594	2	+ + vf	+ - -	3	+ + vf	+ - -	Y	i	0 / 9
PHKA2	1240	2	+ + f	+ vf vf	4	+ f -	+ - -	Y	i	2 / 9
MAP3K15	848	2	+ + f	+ - -	3	+ f vf	+ - -	Y	i	N/A
SH3KBP1	1529	1	+ + vf	+ - -	2	+ + vf	+ - -	Y	i	4 / 9
SAT	668	5	+ + f	+ - -	2	+ + vf	+ - -	Y	i	1 / 9
NR0B1	1228	1	+ + vf	+ f -	2	+ vf vf	+ - -	Y	i	N/A
GK	1129	2	+ + f	+ - -	4	+ + vf	+ - -	Y	i	0 / 9
SRPX	458	2	+ + f	+ - -	2	+ - -	+ - -	Y	i	0 / 9

(Table 3.1 continued)

Gene	CGI size (bp)	RPMA primer pair 1 CCGG sites	Female Hi Hp Ms	Male Hi Hp Ms	RPMA primer pair 2 CCGG sites	Female Hi Hp Ms	Male Hi Hp Ms	Female-specific methylation?	XCI state prediction	Lit.
Hypomethylated in both female and male										
GPM6B	910	3	+ - -	+ - -	3	+ - -	+ - -	N	e	8 / 9
CA5B	584	4	+ - -	+ - -	3	+ - -	+ - -	N	e	9 / 9
AP1S2	1546	3	+ - -	+ - -	5	+ - -	+ - -	N	e	9 / 9
CTPS2	1012	3	+ - -	+ - -	2	+ - -	+ - -	N	e	9 / 9
EIF1AX	1536	3	+ - -	+ - -	3	+ - -	+ - -	N	e	9 / 9
PRDX4	639	4	+ - -	+ - -	4	+ - -	+ - -	N	e	0 / 9
ZFX	1482	2	+ - -	+ - -	3	+ - -	+ - -	N	e	9 / 9
Cxorf38	0	3	+ - -	+ - -	3	+ - -	+ - -	N	e	N/A
CRSP2	941	3	+ - -	+ - -	5	+ - -	+ - -	N	e	6 / 6
USP9X	2246	5	+ - -	+ - -	5	+ - -	+ - -	N	e	9 / 9
UTX	1750	2	+ - -	+ - -	2	+ - -	+ - -	N	e	9 / 9
RP11-30A20.3	519	4	+ - -	+ - -	2	+ vf -	+ - -	N	e	9 / 9
U2AF1L2	633	1	+ vf -	+ - -	1	+ - -	+ - -	N	e	N/A
Methylation state difficult to define										
SMS	1348	2	+ f -	+ - -	6	+ - -	+ - -	U	U	0 / 6
RS1	446	5	+ f -	+ - -	4	+ vf -	+ - -	U	U	2 / 9
DDX3X	2423	4	+ f -	+ - -	2	+ f -	+ f f	U	U	9 / 9
PIR	440	3	+ - -	+ - -	2	+ f vf	+ f -	U	U	9 / 9
CXorf15	1549	1	+ vf vf	+ - -	3	+ - -	+ - -	U	U	9 / 9
RBBP7	1507	1	+ + +	+ f f	4	+ - -	+ - -	U	U	9 / 9
GPR64	1156	7	+ - -	+ - -	6	fail	fail	U	U	N/A
MAOB	527	4	+ - -	+ - -	2	+ f f	+ f f	U	U	N/A
EGFL6	572	2	+ f vf	+ - -	3	+ f f	+ - -	U	U	N/A
RAB9A	499	4	+ f f	+ f f	1	+ f -	+ - -	U	U	9 / 9
OFD1	894	2	+ f -	+ - -	3	+ f vf	+ - -	U	U	6 / 6
CA5BL	968	1	+ - f	+ f vf	1	+ - f	+ f f	U	U	9 / 9
SYAP1	480	1	+ + +	+ f -	2	+ f f	+ f -	U	U	9 / 9
CDKL5	1305	2	+ vf vf	+ - -	3	+ vf -	+ - -	U	U	2 / 9
EIF2S3	617	2	+ vf -	+ vf -	1	+ f -	+ f -	U	U	9 / 9
TMEM47	1219	3	+ f vf	+ - -	3	fail	fail	U	U	0 / 9
CLCN4	561	1	+ + -	+ vf -	3	+ f -	+ vf -	U	U	5 / 9
CNKSR2	1572	2	+ + vf	+ vf -	2	+ + vf	+ vf -	U	U	N/A
RP2	933	2	+ + -	+ f -	1	+ + -	+ f -	U	U	0 / 9
Hypermethylated in both female and male										
CXorf20	837	3	+ + -	+ + -	3	+ + -	+ + -	N/A	N/A	N/A
SIAH1L	446	2	+ + f	+ + f	1	+ + f	+ + f	N/A	N/A	N/A
YY2	563	1	+ + f	+ + f	1	+ + f	+ + f	N/A	N/A	N/A
FTHL17	548	1	+ + +	+ + +	3	+ + f	+ + f	N/A	N/A	N/A
RP1-154K9.1	662	1	+ + f	+ + f	2	+ + vf	+ + vf	N/A	N/A	N/A
MKRN4	482	4	+ + vf	+ + -	6	+ + -	+ + -	N/A	N/A	4 / 4

3.3.4 Analysis of CGI methylation on the mouse X chromosome

All 53 mouse CGIs were successfully assayed by RPMA in female and male mouse embryonic fibroblasts (MEF). Of these, 49 CGIs had data for both amplicons and four others had results from a single amplicon.

Similar to, but more extreme than the situation in human samples, almost all CGIs were hypomethylated in male samples (Table 3.2, Figure 3.6). For 44 (83%) CGIs, no PCR products were amplified using the *Hpa*II digested material. Another seven (13%) CGIs had a complete lack of signal in the *Hpa*II lane for one amplicon, and a faint or very faint signal in the *Hpa*II lane for the other amplicon. Only two CGIs had a signal in the *Hpa*II lane that is comparable with the signal from the *Hind*III control.

However, the methylation patterns in mouse female samples are very different from that observed in human samples. Only three CGIs, covering the 5' regions of the genes *Utx*, *Eif2s3x*, and *Ddx3x*, were clearly hypomethylated in both female and male samples (Figure 3.6, a-c). The CGIs of the orthologous genes in human appeared to be either hypomethylated (*UTX*) or having low level methylation (*EIF2S3*, *DDX3X*) in female.

In contrast, the vast majority of mouse CGIs (48 out of 53) exhibited female-specific methylation in at least one amplicon (Figure 3.6, d-f), and most (35 out of 48) had both amplicons methylated in female samples. Eight CGIs had one amplicon clearly hypermethylated, and the other amplicon was either absent (1 case), had a strong signal in the *Hpa*II lane but also a very faint signal from the *Msp*I control (2 cases), or had a faint signal in the *Hpa*II lane (2 cases). For

the other five CGIs, all amplicons had a strong signal in the *Hpa*II lane but also a faint or very faint signal from the *Msp*I control. Interestingly, the patterns suggesting intermediate levels of methylation, observed in a considerable fraction of human CGIs, were never seen in mouse CGIs.

In the remaining two cases, the CGIs showed some degrees of methylation in both female and male samples.

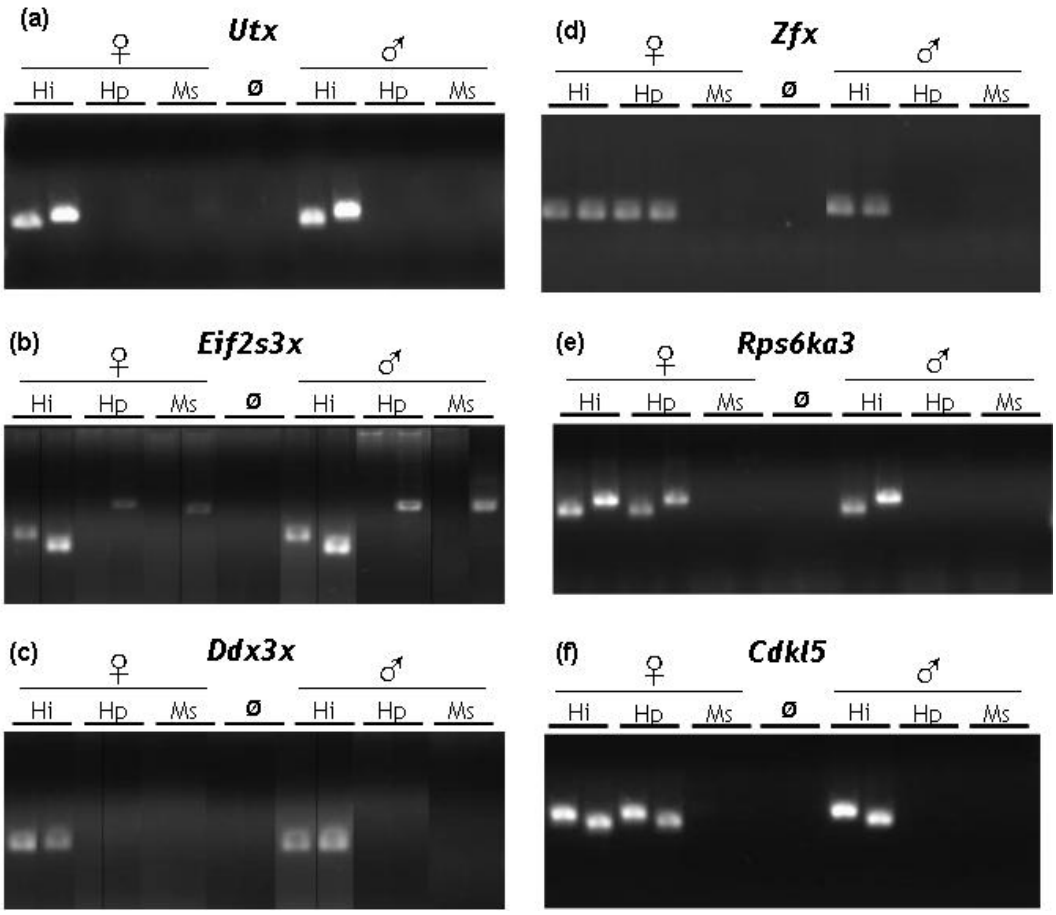

Figure 3.6: RPMA results of mouse CGIs with hypomethylation or female-specific methylation of CGI. (a-c) are examples of CGIs with hypomethylation in both female and male samples. (d-f) are examples of CGIs with female-specific methylation. Lanes in (b) and (c) have been re-organised as the original loading pattern had the two amplicons on separate gels.

Table 3.2: RPMA results of mouse CGIs. '+' = strong band; 'f' = faint band; 'vf' = very faint band; '-' = no visible band. The 'XCI state prediction' is based on RPMA results. For most mouse CGIs, no record of XCI status can be found in literature. XCI status is colour-coded: pink for X-inactivated and violet for escapee.

Gene	Human orthologue	CGI size (bp)	RPMA primer pair 1			RPMA primer pair 2			Female-specific methylation?	XCI state prediction
			CCGG sites	Female Hi Hp Ms	Male Hi Hp Ms	CCGG sites	Female Hi Hp Ms	Male Hi Hp Ms		
Female-specific methylation in at least one amplicon										
Hccs	HCCS	729	3	+ + -	+ - -	3	+ + -	+ f -	Y	i
Msl3l1	MSL3L1	1029	2	+ + -	+ vf -	3	+ + -	+ - -	Y	i
A630018P17Rik	RP11-30A20.3	661	3	+ + -	+ - -	2	+ + -	+ - -	Y	i
Rab9	RAB9A	504	2	+ + -	+ - -	2	+ + -	+ - -	Y	i
Ofd1	OFD1	846	2	+ + -	+ vf -	3	+ + -	+ - -	Y	i
Gpm6b	GPM6B	641	4	+ + -	+ - -	4	+ + -	+ vf -	Y	i
Piga	PIGA	527	2	+ + -	+ - -	6	+ + -	+ - -	Y	i
U2af1-rs2	U2AF1L2	1089	2	+ + -	+ - -	3	+ + -	+ - -	Y	i
Ap1s2	AP1S2	1112	3	+ + -	+ - -	3	+ + -	+ - -	Y	i
4932441K18Rik	CXorf15	1152	2	+ + -	+ - -	2	+ + -	+ - -	Y	i
Rbbp7	RBBP7	1362	3	+ + -	+ - -	2	+ + -	+ - -	Y	i
Reps2	REPS2	516	3	+ + -	+ - -	2	+ + -	+ - -	Y	i
ENSMUSG00000067171	NHS	1354	2	+ + -	+ - -	3	+ + -	+ - -	Y	i
Rai2	RAI2	1016	2	+ + -	+ - -	2	+ + -	+ - -	Y	i
Cdk15	CDKL5	899	2	+ + -	+ - -	2	+ + -	+ - -	Y	i
Phka2	PHKA2	769	2	+ + -	+ - -	3	+ + -	+ - -	Y	i
Gpr64	GPR64	463	3	+ + -	+ - -	3	+ + -	+ - -	Y	i
Pdha1	PDHA1	489	6	+ + -	+ - -	2	+ + -	+ - -	Y	i
ENSMUSG00000031303	MAP3K15	728	3	+ + -	+ - -	6	+ + -	+ - -	Y	i
Sh3kbp1	SH3KBP1	999	3	+ + -	+ - -	2	+ + -	+ - -	Y	i
Eif1ay	EIF1AX	783	4	+ + -	+ - -	3	+ + -	+ - -	Y	i
Rps6ka3	RPS6KA3	1155	3	+ + -	+ - -	4	+ + -	+ - -	Y	i
Cnksr2	CNKSR2	1447	2	+ + -	+ - -	5	+ + -	+ - -	Y	i
Mbtps2	MBTPS2	730	3	+ + -	+ - -	7	+ + -	+ - -	Y	i
Sms	SMS	1001	2	+ + -	+ - -	2	+ + -	+ - -	Y	i
Ptchd1	PTCHD1	1174	3	+ + -	+ - -	2	+ + -	+ - -	Y	i
Zfx	ZFX	1590	2	+ + -	+ - -	3	+ + -	+ - -	Y	i
Siah1b	SIAH1L	686	3	+ + -	+ - -	8	+ + -	+ - -	Y	i
Mid1ip1	MID1IP1	1532	2	+ + -	+ - -	2	+ + -	+ - -	Y	i
Atp6ap2	ATP6AP2	744	3	+ + -	+ - -	1	+ + -	+ - -	Y	i
Usp9x	USP9X	1660	1	+ + -	+ - -	4	+ + -	+ - -	Y	i
Cask	CASK	1174	2	+ + -	+ - -	7	+ + -	+ - -	Y	i
Maoa	MAOA	504	2	+ + -	+ - -	3	+ + -	+ - -	Y	i
Slc9a7	SLC9A7	550	3	+ + -	+ - -	2	+ + -	+ - -	Y	i
Phf16	PHF16	1017	2	+ + -	+ - -	5	+ + -	+ - -	Y	i
Arhgap6	ARHGAP6	463	3	+ + -	+ - -	2	+ + -	+ - -	Y	i
2900002G04Rik	RP11-393H10.2	1286	2	+ + -	+ - -	4	+ f -	+ - -	Y	i
Prdx4	PRDX4	409	3	+ + -	+ - -				Y	i
Tmem47	TMEM47	793	5	+ f -	+ - -	2	+ + -	+ - -	Y	i
Crsp2	CRSP2	974	2	+ + -	+ - -	2	+ f f	+ f f	Y	i
Gyk	GK	659	1	+ + vf	+ - -	2	+ + -	+ - -	Y	i
Lancl3	LANCL3	1078	2	+ + -	+ - -	1	+ + vf	+ f -	Y	i
Prps2	PRPS2	679	2	+ + vf	+ - -	2	+ + vf	+ - -	Y	i
Tmsb4x	TMSB4X	2054	1	+ + f	+ - -	1	+ + f	+ - -	Y	i
Syap1	SYAP1	594	1	+ + vf	+ - -				Y	i
Pdk3	PDK3	497	4	+ + vf	+ - -				Y	i
Pola1	POLA	485	3	+ + vf	+ - -				Y	i
Hypomethylated in both female and male										
Eif2s3x	EIF2S3	651	1	+ - -	+ - -	2	+ - -	+ - -	N	e
Ddx3x	DDX3X	1274	2	+ - -	+ - -	3	+ - -	+ - -	N	e
Utx	UTX	1739	3	+ - -	+ - -	2	+ - -	+ - -	N	e
Hypermethylated in both female and male										
Xkh	XK	652	1	+ + vf	+ f vf	1	+ + vf	+ + -	N/A	N/A
Bcor	BCOR	2118	3	+ + -	+ + -	2	+ + -	+ + -	N/A	N/A

3.4 Discussion

In this chapter, I wish to use the methylation state of the 5' CGI as an indicator to predict the genes' XCI status. To this end, CGI methylation of 90 human X-linked genes, known to include both inactivated and escapee genes, were examined in cultured fibroblast cells using a rapid method, RPMA. An attempt was made to predict the gene's XCI status based on the methylation profile of its CGI: hypomethylated CGIs were assumed to associate with escapee genes and hypermethylated CGIs were associated with inactivated genes. Since hypomethylation of a single CCGG site, thus cleavage by *Hpa*II is sufficient to prevent PCR amplification, a prediction of escapee was only made when both amplicons in one CGI gave a pattern consistent with hypomethylation. In total, a prediction was made for 64 (71%) genes.

Over half of the human CGIs were found to have female-specific methylation (Table 3.1, Figure 3.4), and the 51 associated genes (57% of all assayed genes) were predicted to be subject to XCI. Previously, Carrel and Willard (2005) created an inactivation profile for the whole human X chromosome by analysing gene expression from human X_a and X_i in a panel of rodent/human somatic cell hybrids. From the S3 region, 128 genes were assayed in their study and 83 (65%) were found to be effectively silenced (expression detected from no X_i hybrid or up to three X_i hybrids). This proportion of silenced genes based on expression is similar to the estimation based on CGI methylation in this study. Moreover, 42 out of the 51 genes with a prediction of inactivation in this thesis were examined in the study by Carrel and Willard (2005). Thirty-nine out of the 42 were shown to be inactivated. *MAOA* (which encodes monoamine oxidase type

A) (Figure 3.4, f), one of those whose inactivation state disagreed, was found to escape XCI in the previous study but should be subject to XCI according to the methylation test. In Carrel and Willard's study, the XCI status of *MAOA* was also tested using a quantitative SNP assay in primary human cells, and interestingly, the expression from X_i never exceeded 15% in any cell line tested, so *MAOA* may still be relatively effectively silenced. In addition, several other studies, using diverse methods including CGI methylation analysis, expression analysis by RT-PCR, and allele-specific expression analysis by SNP genotyping, all found evidence supporting inactivation of *MAOA* (Benjamin *et al.*, 2000; Hendriks *et al.*, 1992; Nordquist and Oreland, 2006).

Notably, a considerable fraction of human CGIs appeared to be hypomethylated. Thirteen CGIs were clearly hypomethylated at both amplicons in both female and male samples (Table 3.1, Figure 3.4). If absence of CGI methylation is a reliable indicator of poor inactivation, these genes are predicted to escape from XCI in the fibroblast cells used. Alternatively, these genes may not rely on methylation for maintaining stable inactivation. Comparison to published data favours the former possibility: of these 13 genes with hypomethylated CGIs, 11 were also assayed by Carrel and Willard (2005), and ten were found to escape from XCI (expression detected in all, or in one case, 8 out 9, X_i hybrids). The only disagreement was *PRDX4*, for which no other XCI data are available in literature.

In total, 64 genes had a prediction of their XCI status made based on clear CGI methylation patterns as assayed by RPMA. Of these, 53 genes were assayed in cell hybrids in the previous study, and for 49 genes (92%), observations from the previous study strongly support predictions made in current study. Therefore,

CGI methylation as measured by RPMA can serve as a good indicator of the gene's XCI status.

Previous researchers have suggested that XCI in mouse is more complete than XCI in human (Disteche, 1995). To investigate whether this difference in XCI is reflected in difference in CGI methylation on the X chromosome, I extended the methylation analysis to the mouse orthologues of human S3 genes. Of the 53 mouse genes assayed, only the three known escapees, *Utx* (Greenfield *et al.*, 1998), *Eif2s3x* (Ehrmann *et al.*, 1998), and *Ddx3x* (Disteche *et al.*, 2002), gave a pattern consistent with escape from XCI (Figure 3.6, a-c). Forty-eight genes were shown to have a profile consistent with inactivation (Table 3.2, Figure 3.6, d-f), including *Zfx*, which is known to be silenced (Adler *et al.*, 1991; Ashworth *et al.*, 1991). XCI statuses for most of the 48 genes have not been reported in literature previously, but available evidence suggests that the vast majority of genes on the mouse X chromosome undergo normal XCI (Disteche, 1995), which is consistent with my finding.

Comparing the CGI methylation profiles of the orthologous human and mouse genes reveals a much higher proportion of escapees in human. The same trend has also been observed by other researchers. Whereas more than 15% of genes on the human X chromosome were found to escape XCI (Carrel and Willard, 2005), less than ten escapees have been identified on the mouse X chromosome so far (Disteche *et al.*, 2002), so it might not be surprising that I failed to discover any novel escapees in mouse. Although an extensive XCI profiling has not been done for the entire mouse X chromosome, many of the genes that escape from XCI in human were found subject to inactivation in mouse, suggesting more complete XCI in mouse (Disteche, 1995). This observation is also supported by

comparative analysis of a conserved gene cluster around *SMCX/Smcx*, a known escapee in both human and mouse (Tsuchiya *et al.*, 2004). Whereas there is a large escapee domain in human, *Smcx* is the only escapee in the region in mouse. Similarly, I found the genes surrounding the three mouse escapees to all have a CGI methylation profile consistent with inactivation, in contrast to the situation in human where the escape domains are larger.

Using the rapid RPMA approach, I confirmed that a much higher percentage of genes are silenced on the mouse X_i than on the human X_i, and identified a large number of mouse genes that undergo XCI, most of which is novel finding.

It is well known that 5' CGIs are generally unmethylated in somatic tissues in healthy individuals. Even in the case of XCI or imprinting, the CGI is only methylated at one of the two alleles. Therefore it is curious that in my study, six human CGIs (Figure 3.5d) and one mouse CGI displayed identical hypermethylation in both female and male samples. This methylation pattern is more commonly observed in non-5' CGIs. A closer examination of the exact locations of the assayed sites revealed that most sites are either intronic (as the case of Figure 3.5d) or in the exon of a single-exon gene. Interestingly, the only other human CGI associated with a single-exon gene had one amplicon hypermethylated in both sexes while the other amplicon had normal female-specific methylation. Moreover, in an additional case in mouse, one assayed site is intronic and was hypermethylated in male, while the other site is located towards the end of exon1 and showed low level methylation. It is very possible that these CGIs, or the assayed portions of them, are not critical in transcriptional control.

Another interesting difference between human and mouse comes from the unclassifiable cases. The remaining 19 CGIs in human mostly had faint signals

in female, and sometimes in male, *Hpa*II lane, suggesting intermediate levels of methylation (Figure 3.5, a-c). These ambiguous methylation patterns make interpretation of XCI status difficult, but interestingly, were never seen in any mouse CGIs. Unfortunately, the RPMA approach is limited in its ability to examine details of methylation. At this stage, it was not possible to distinguish between patterns formed by a small proportion of hypermethylated CGIs and that by a large proportion of intermediately methylated CGIs. To test which situation gave rise to these intermediate methylation patterns, more detailed methylation information, ideally at single CpG resolution, needs to be obtained.

In this chapter, I have demonstrated that RPMA is a rapid method of predicting XCI status for a large fraction of genes with a CGI. RPMA results are consistent with much more complete gene silencing in mouse XCI than in human. In addition, there is a class of intermediately methylated CGIs unique to the human species and I shall carry out further study of the detailed methylation patterns for a representative set of genes in the next chapter.