Chapter 1

Introduction

1.1 Dosage compensation and X chromosome inactivation (XCI)

Most animals and plants adopt sexual reproduction, which creates great genetic variation. Many sexually reproducing species consist of two sexes, male and female. In some cases, like in certain species of fish and reptiles, sex is determined by environmental variables, but in most cases, sex determination is genetic. A system of chromosomal sex-determination has evolved independently in many plant and animal groups, where the two sexes have different chromosomal constitutions (Bull, 1983). One sex, termed the heterogametic sex, has two different sex chromosomes, while the other sex, the homogametic sex, has a pair of identical chromosomes. The X and Y chromosomes (when male is heterogametic), or the Z and W chromosomes (when female is heterogametic), are often morphologically and genetically distinct, shaped by sex chromosome evolution.

1.1 Dosage compensation and X chromosome inactivation (XCI)

In the 1960s, Ohno (1967) first proposed an autosomal origin for sex chromosomes. According to this hypothesis, a pair of proto-sex chromosomes arose from a pair of autosomes when a sex-determining allele evolved on one of them, termed the Y or W chromosome (the XY system is discussed thereafter). Starting from a pair of identical autosomes, the sex chromosomes went onto very different paths. Selection favoured restriction of recombination between the Y-linked sex-determining region and its X homologue to avoid sexual ambiguity. This restriction eventually spread throughout the two chromosomes, leading to degeneration of the Y due to accumulation of mutations. The X chromosome, on the other hand, retained its gene content, which comprises an indispensable part of the genome, and had to accommodate its hemizygous state in male. Thus the need for dosage compensation between the two sexes arose.

Just as chromosomal sex-determination has evolved many times in different lineages, so dosage compensation mechanisms have also evolved multiple times. Typical of evolution, different solutions have been used to solve the same problem: to achieve dosage balance between gene expression from the two sexes, as well as balance between expression from the X and autosomes. Dosage compensation mechanisms have been best studied in the fruit fly *Drosophila melanogaster*, the nematode worm *Caenorhabditis elegans*, and in mouse and human. Each group has evolved a drastically different system to achieve dosage compensation. Drosophila has an XY sex-determination system but without a male-determining gene on the Y. Rather, sex determination results from counting the X:autosome ratio. An X:autosome ratio of 1:1 results in development into a female, and an X:autosome ratio of 2:1 gives rise to a male. So an XO individual in Drosophila is a male, not a female as in human. Dosage compensation is achieved by up-

regulating the expression of genes on the single X in male two-fold (reviewed in Baker *et al.*, 1994; Cline and Meyer, 1996). In *C. elegans*, the Y chromosome is lost completely, and sex is determined by the number of X chromosomes. As in Drosophila, the XO individual is a male, but the XX individual is a hermaphrodite (having both male and female reproductive organs). Dosage compensation is targeted at both X chromosomes in the hermaphrodite, reducing the transcription level of each X by half (reviewed in Cline and Meyer, 1996; Meyer, 2000).

In contrast to the straightforward solutions evolved in flies and worms, a much more complex dosage compensation mechanism has developed in mammals, namely X chromosome inactivation (XCI). The hypothesis was first proposed by Mary Lyon over 40 years ago: in mammals, the dosage equalisation of gene products from the X chromosome is achieved by transcriptionally inactivating one of the two X chromosomes in female cells during embryogenesis (Lyon, 1961). In her landmark paper, Lyon linked together several curious observations about the X chromosome in mammals. Barr and Bertram (1949) had discovered that a 'nucleolar satellite' was always found in female cells but not in male cells; Ohno (1960; 1959) suggested that the 'Barr body', as it was later called, is a highly condensed X chromosome. Around the same time, it was found that female mice with only one X chromosome (XO) are viable and even fertile (Welshons and Russell, 1959), suggesting that one active X chromosome is sufficient for normal development. Lyon's hypothesis explains the mosaic, rather than intermediate, coat colour of female mice heterozygous for X-linked pigment genes: during very early stages of embryonic development, one X chromosome becomes inactivated in each cell, and the progenies of which can only make the normal or mutant form of the pigment, depending on which X is active, giving rise to patches of colours

(see Figure 1.1 for an example in cat).

At the same time, the idea of XCI was also indicated in a paper on mammalian sex chromosomes by Russell (1961). In an attempt to explain the variegated-type position effects, where the recessive allele is manifested in heterozygote as a result of X;autosome translocation, Russell suggested that only one X is active in females and the extra Xs acquire characteristics of heterochromatin (Russell, 1961). Direct evidence in support of the XCI hypothesis came from study of cloned cells. Davidson and colleagues (1963) studied cell cultures from human females heterozygous for the X-linked gene encoding the enzyme glucose-6-phosphate dehydrogenase (G6PD). They found that each cloned culture, originating from single cells, only expressed one of the two alleles. This proved that only one X chromosome is active in each cell, and showed how the inactivation pattern of every cell is faithfully reproduced in its daughter cells.

Subsequently, Ohno (1967) provided a model for the evolution of XCI in mammals. Following the decay of the Y chromosome, the rate of product output of X-linked genes was doubled to compensate for the hemizygous existence of X-linked genes in the male. Once this step was accomplished, there was no longer the need for two active Xs in the female, and inactivation of one X has evolved. Therefore dosage compensation in mammals includes two separate elements: a) transcriptional silencing of the inactive X (X_i); and b) two-fold upregulation of gene expression on the active X (X_a). The upregulation has only been confirmed in the past 15 years, thanks to advance in gene expression assays (Adler *et al.*, 1997; Lin *et al.*, 2007; Nguyen and Disteche, 2006), but the silencing has been extensively studied over the last half century.

Figure 1.1: Tortoiseshell cat. Manifestation of XCI in coat colour patches. Photo reproduced from Wikimedia Commons (`http://en.wikipedia.org/wiki/Image:Tortoiseshell_tabby_-_TortiePuff.jpg`) under the GNU Free Documentation License.

1.2 The mechanism of XCI

Although many of the details are still poorly understood, decades of active research has now constructed a basic model for the initiation, spreading and maintenance of XCI.

In the somatic cells of eutherians, the choice of the X to be inactivated is random. Early studies suggested a region from where X-inactivation is initiated and spreads across the chromosome (Russell, 1963). A *cis*-acting X-inactivation cen-

tre (XIC) was supported by mouse X;autosome translocations, where inactivation spread from the inactive X into the attached autosome (Russell and Montgomery, 1965), and X chromosome deletions, where inactivation was limited to the XIC-bearing segment (Rastan and Robertson, 1985).

From the region of the human XIC, the *XIST* gene was cloned (Brown *et al.*, 1991). Its expression was only detected in human cells containing the X_i, including female and male cells with multiple X chromosomes, but not seen in cells from XO female or normal male (Brown *et al.*, 1991). By use of somatic cell hybrids containing a single human X chromosome, they confirmed that *XIST* is expressed exclusively from the X_i but not the X_a (Brown *et al.*, 1991). Using the human *XIST* cDNA as a probe, its mouse homologue *Xist* was mapped to the mouse Xic (Borsani *et al.*, 1991). Like its human homologue, expression of the mouse *Xist* is also limited to the inactive X chromosome (Borsani *et al.*, 1991). Its location and unique expression pattern in both human and mouse made *XIST/Xist* the prime candidate as a controlling element of XCI. Both the human and mouse genes contain several tandem repeats and produce large, alternatively spliced, transcripts (Brockdorff *et al.*, 1992; Brown *et al.*, 1992). The matured transcripts lack significant conserved open reading frames (ORF) and localise almost exclusively to the nucleus, suggesting that *XIST/Xist* may not encode a protein but rather function as a structural RNA (Brockdorff *et al.*, 1992; Brown *et al.*, 1992).

The role of XIST as a structural RNA in XCI was strongly supported by the close association of spliced but non-translated XIST RNA with the X_i, observed in fluorescent *in situ* hybridization (FISH) studies (Clemson *et al.*, 1996). More direct evidence of the involvement of *XIST* in XCI came from transgenic studies. The mouse embryonic stem (ES) cell has provided an excellent model to study

the mechanism of XCI. Upon induction to differentiate, the female ES cells carry out random XCI, allowing researchers better access to details of the early events. In addition, these cells can be manipulated to investigate how mutations in individual genes can affect the process of XCI. Using gene targeting to delete large parts of the *Xist* gene, Penny and colleagues (1996) demonstrated that an intact *Xist* is required for *cis* inactivation: in the cells with an X chromosome bearing the null allele of *Xist*, only the X chromosome with the intact *Xist* could be inactivated. Furthermore, *Xist* was shown to be sufficient to induce inactivation when introduced into an autosome (Herzing *et al.*, 1997).

Current evidence suggests that the role of XIC to initiate *cis* inactivation is predominantly carried out by the *XIST* locus, but there are other elements present in the XIC which may play a part in the other role of XIC: the choice of the active X. Since the X chromosome contains many genes important for normal development and proper cell functioning, it is critical to ensure one, and only one, X remains active. In random XCI, the initial step is the random choice of a single future X_a, followed by inactivation of any remaining X chromosome. Before the discovery of *Xist*, Rastan (1983) proposed a simple model, where a *trans*-acting blocking factor randomly blocks the Xic on a single X chromosome and thus protects that chromosome from being silenced. This model was supported by behaviour of embryo-derived (EK) cell lines with X chromosome deletions (Rastan and Robertson, 1985). When cells with a partly-deleted X chromosome were induced to differentiate, either the intact X or the deleted X could be inactivated; but when Xic was included in the deletion, neither X was inactivated (Rastan and Robertson, 1985).

The nature of the *trans*-acting blocking factor remains unclear to this day,

but studies in mice have provided several candidates within the Xic that may respond to the blocking factor. The *Xist* gene itself does not seem to affect choice, as mouse ES cells with one *Xist* deleted still behaved as if they had two Xic and underwent inactivation (Penny *et al.*, 1996). At least in mouse, multiple elements 3' to the *Xist* appear to be involved in the complex process of counting and choice. When a 65-kb region 3' to *Xist* exon 6 was deleted, the X chromosome bearing the deletion was always inactivated, even in XO cells, suggesting that this region is needed to repress *Xist* in *cis* (Clerc and Avner, 1998). From this region, a transcript antisense to *Xist* was discovered by RNA FISH using strand-specific probes in mouse ES cells, and was named *Tsix* (Lee and Lu, 1999). The Tsix RNA, spanning the entire length of *Xist*, is specifically expressed from the active X from the onset of XCI and persists only until the establishment of XCI, consistent with a role in blocking *Xist* function. Such a role was supported by examination of female ES cells heterozygous for a deletion of *Tsix*, in which only the mutation-bearing X was always inactivated (Lee and Lu, 1999). However, restoration of *Tsix* in cells carrying the previously described 65-kb region failed to restore random XCI, suggesting existence of additional functional elements within this region (Morey *et al.*, 2001). Further exploration in the remaining region uncovered a second *cis*-acting element, *Xite* (X inactivation intergenic transcription elements), which lies upstream of *Tsix* and appears to act through promotion of *Tsix* persistence (Ogawa and Lee, 2003). A less clearly defined locus that is involved in the choice process is the X controlling element (*Xce*), which has long been found to affect the randomness of XCI (Cattanach and Isaacson, 1967; Cattanach and Williams, 1972). It has been mapped 3' to *Xist*, but its exact location and nature remain undetermined (Chadwick *et al.*, 2006; Simmler

et al., 1993).

In human, the process of choice is even less understood. An orthologous gene of the mouse *Tsix*, *TSIX*, was also found in the human XIC (Migeon *et al.*, 2001). However, despite being an antisense to *XIST*, the human version of *TSIX* was seen expressed only from the inactive X, together with *XIST*, so does not appear to block XIST function (Migeon *et al.*, 2002). The other locus affecting choice in mouse, *Xite*, does not appear to have an orthologue in human so far.

After choosing the future X_a, *Xist* expression from the future X_i is upregulated and the *Xist* transcripts rapidly spread along and associate with, or 'coat', the chromosome in *cis* (Clemson *et al.*, 1996). The silencing itself is achieved by changing the chromatin of the future X_i to a transcriptionally inactive state via a cascade of chromatin modifications. DNA is packed around a special kind of proteins, the histones, to form chromatin. For transcription to happen, the chromatin needs to be in an 'open' state to allow the transcription initiation complex to access and associate with the DNA. In contrast, transcription is inhibited when the chromatin is in a 'closed' state, tightly packed and gives no access to the factors essential for transcription initiation. Several modifications to certain amino acids in the histone, especially in the tails, can change the conformation of the protein and in turn affects the packing of DNA around it. Histone modifications associated with a particular state of the chromatin can be seen as the histone 'signature' for the state. The 'open' chromatin, or euchromatin, usually has the lysines in histones acetylated, whereas the 'close' chromatin, or heterochromatin, normally has these lysines hypoacetylated (Grunstein, 1997).

It has been noticed that the chromatin of X_i carries epigenetic marks distinct from those of the X_a, but closely resembling the features of constitutive hete-

rochromatin, including hypoacetylation of histones H3 and H4, and enrichment of the histone H2A variant macroH2A (Boggs *et al.*, 1996; Costanzi and Pehrson, 1998; Jeppesen and Turner, 1993). Using mouse ES cells as a model, a number of research groups were able to establish the timing of these events in early development (Goto *et al.*, 2002; Heard *et al.*, 2001; reviewed in Plath *et al.*, 2002). A combination of Xist RNA FISH and immunofluorescence staining with specific antibodies also enabled researchers to detect the fine details and interplay of the various types of histone modification (Goto *et al.*, 2002; Heard *et al.*, 2001; reviewed in Plath *et al.*, 2002). Closely following the spread of Xist RNA along the X_i, one of the earliest events is the methylation of histone H3 lysine 9 (H3K9), together with hypoacetylation of H3K9 and loss of methylation at histone H3 lysine 4 (H3K4) (Heard *et al.*, 2001). Hypoacetylation of H4 was observed shortly after modifications of the H3 tail, and silencing of the two genes examined also followed X-chromosome-wide H3K9 methylation (Heard *et al.*, 2001). These histone marks are maintained in adult cells, as shown by studies of the mouse genes encoding phosphoglycerate kinase 1 (*Pgk1*) and hypoxanthine-guanine phosphoribosyltransferase (*Hprt*) (Goto *et al.*, 2002). At both loci, chromatin on the inactive X was marked by H3K9 methylation and hypoacetylation of histones H3 and H4, whereas the opposite pattern of H3K4 methylation and acetylated H3 and H4 was found on the active X (Goto *et al.*, 2002). Enrichment of macroH2A on the X_i occurred later and was not essential for inactivation, suggesting a redundant role in the maintenance of XCI (Csankovszki *et al.*, 1999; Mermoud *et al.*, 1999). As a result of a series of chromatin modifications, the chromatin of the inactive X is turned into tightly packed heterochromatin, manifesting itself as the condensed 'Barr body' during interphase, and closed to transcription.

Experiments in mouse demonstrated that, although *XIST* is still expressed for a while after the initiation of XCI and may contribute towards the stabilisation of XCI, it is not essential for the maintenance of XCI (Brown and Willard, 1994; Csankovszki *et al.*, 2001, 1999). During mitosis, *XIST* is not expressed and XIST RNA dissociates from the X_i (Clemson *et al.*, 1996); but the inactivation pattern is faithfully replicated through cell generations. The 'memory' mechanism of XCI maintenance is not clear but DNA methylation has been suggested to play a role (Kaslow and Migeon, 1987). Clusters of CpG-rich sequences, or CpG islands (CGI), are present near the 5' transcription start sites of most housekeeping genes (Bird *et al.*, 1985). CGIs are normally unmethylated to allow gene activity, but CGIs on the inactive X are frequently hypermethylated (Tribioli *et al.*, 1992). Observations in mouse embryos and mouse embryonic stem cells confirmed that CGI methylation is a late event in XCI, only occurring after the cascade of histone modification, consistent with a role in the maintenance of XCI (Keohane *et al.*, 1996; Lock *et al.*, 1987). Once established, CGI methylation patterns can be faithfully reproduced after DNA replication, thanks to the preferential methylation at semi-methylated sites by the maintenance-methylating enzyme, DNA methyltransferase 1 (Dnmt1). In support with the maintenance function, treatments with demethylation agents were shown to induce re-activation of genes on the inactive X (Graves, 1982; Jones *et al.*, 1982). CGI methylation and its involvement in XCI is discussed in more details in Chapter 3.

1.3 Escape from XCI

Much insight into the process of XCI has been gained through studying the expression of genes on the X chromosome. One interesting discovery was the exception to the rule: although most genes on the inactive X are stably silenced, a subset of genes actually escape from XCI and are expressed from both the active and inactive X.

Escapees were first discovered in humans. Very soon after Lyon published her hypothesis of XCI, there was suggestion that the XG blood group gene in human is not inactivated on X_i based on failure to identify cells manifesting the recessive allele in heterozygous females (Gorman *et al.*, 1963). A new test for XCI status was then proposed using blood cells from patients with Lesch-Nyhan syndrome (Fialkow, 1970). The syndrome is caused by a deficiency of the enzyme HPRT, but red blood cells from heterozygous female patients were found to have skewed XCI, where the X carrying the mutant allele of *HPRT* is always inactivated, presumably due to cell selection in early development. Three females, each heterozygous for both *HPRT* and *XG*, and each inherited the normal *HPRT* allele and the recessive *XG* allele from their fathers, had their red cells typed as XG(+ve), indicating that the *XG*(+ve) allele on the inactive X was also expressed (Fialkow, 1970). Later a second escapee, encoding the protein steroid sulfatase (STS), was identified using cloned cell lines, which had the same X inactivated (Shapiro *et al.*, 1979). Interestingly, *STS* is closely linked with *XG* and both are located at the extreme distal short arm of the X chromosome (Xp), suggesting a region of genes that escape from XCI.

Most subsequent escapees were discovered using the somatic cell hybrid method.

When hybrid cells are made by merging rodent and human cells, these hybrids will retain most rodent chromosomes but lose the human chromosomes. Hybrid cell lines retaining human X_a or X_i can be established by growing in suitable selective media (Mohandas *et al.*, 1980). This experimental system provides the means to directly test gene expression from morphologically normal X_a and X_i, and was first used to confirm the expression of *STS* from X_i (Mohandas *et al.*, 1980). Using somatic cell hybrids, other genes at the tip of Xp were found to escape XCI, including *CD99* (Goodfellow *et al.*, 1984), *XE7* (Ellison *et al.*, 1992), and *ANT3* (which encodes adenine nucleotide translocator 3) (Schiebel *et al.*, 1993). However, the identification of *ZFX* (which encodes a zinc-finger protein), located in the more proximal part of Xp, and *RPS4X* (which encodes ribosomal protein S4), located in the long arm of X chromosome (Xq), as escapees suggests that genes escaping inactivation are interspersed among genes subject to inactivation (Fisher *et al.*, 1990; Schneider-Gädicke *et al.*, 1989). Recently, 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. Six hundred and twenty four transcripts, representing ∼95% of all assayable genes on the human X chromosome, were assayed in up to nine cell hybrids, each containing a different X_i. They found about 15% of the transcripts to be expressed from the X_i in all or all but one or two hybrids, thus escaping from XCI. An additional 10% of the transcripts were expressed from the X_i in some hybrids, displaying some degree of escape.

The escapees are not randomly distributed along the human X, but show very strong position effect (Carrel and Willard, 2005). For example, genes on the short arm of the X chromosome appear to be more susceptible to escaping from XCI,

perhaps reflecting the evolutionary history of the sex chromosomes. As described in section 1.1, XCI arose as one of the major consequences of mammalian sex chromosome evolution. By comparing present day X-Y homologues in human, Lahn and Page (1999) proposed that human sex chromosome evolution was punctuated by at least four events, as these genes can be divided into four distinct groups based on the degree of divergence between the X and Y copies. Moreover, genes within each group are clustered together on the X chromosome, and these clusters form an orderly sequence corresponding to the degrees of divergence. Lahn and Page (1999) suggested that the most likely event to account for this phenomenon was chromosomal inversion: each time a large block of chromosome was inverted on the Y and ceased to recombine with the X, followed by decay of genes within this block on the Y and recruitment of these genes into XCI on the X. They termed these blocks evolutionary strata, stratum 1 (S1) being the oldest, and stratum 4 (S4) being the youngest. The sequencing of the human X chromosome provided further support to the evolutionary strata, and helped to detect finer detail of the evolutionary events that created the younger strata (Ross *et al.*, 2005). As a result, the boundaries between the strata were refined and S4 was subdivided into S4 and S5 (Ross *et al.*, 2005). This model of sex chromosome evolution is consistent with the distribution of genes that escape XCI: escapees are only sparse in the older strata, S1 and S2, but abundant in the younger strata, S3-5, with S4 and S5 containing almost entirely escapees (Figure 1.2). At the tip of the short arm lies the short arm pseudoautosomal region (PAR1), where X-Y homology is maintained by an obligatory recombination in male meiosis and all genes escape from XCI.

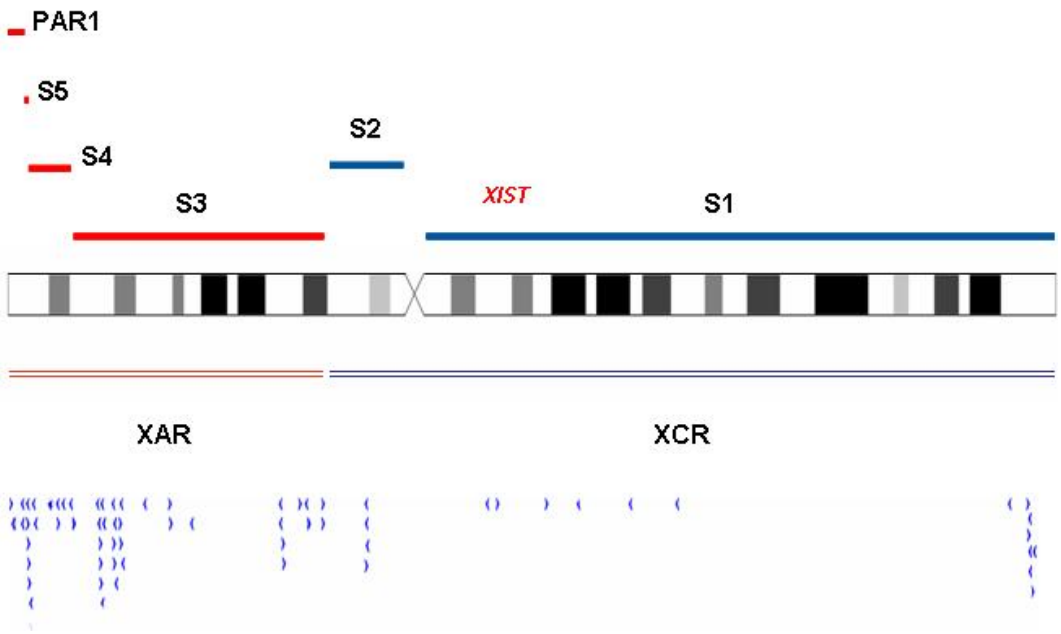

Figure 1.2: Evolutionary strata of human X chromosome. S1-5 are the five evolutionary strata. PAR1 = pseudoautosomal region; XCR = X conserved region; XAR = X added region (added to X and Y from an autosome prior to radiation of the eutherian mammals). Blue triangles represent the positions and orientations of genes that escape from X chromosome inactivation.

Escaping from XCI is not unique to human. The discovery of human escapees led researchers to test the XCI status of their mouse homologues. A convenient experimental system has been developed for mouse to distinguish between gene expression from the active X and the inactive X, using female mouse carrying Searle's X;autosome translocation, T(X;16)16H (thereafter shortened to T16H), in which the normal X is preferentially inactivated (Lyon *et al.*, 1964). By crossing T16H mice with another strain, the F1 female progeny would not only have non-random XCI, but also provide plenty of polymorphisms between the two X chromosomes for examination of allele-specific expression at almost every locus (Adler *et al.*, 1991; Ashworth *et al.*, 1991). Using this system, several human

genes that escape from XCI were found to have mouse orthologues that are also escapees (Agulnik *et al.*, 1994; Ehrmann *et al.*, 1998; Greenfield *et al.*, 1998; Wu *et al.*, 1994). However, more frequently, a gene that escapes from XCI in human is subject to inactivation in mouse (Adler *et al.*, 1991; Ashworth *et al.*, 1991; Kay *et al.*, 1991; reviewed in Disteche *et al.*, 2002). Although a chromosomal-wide XCI profile is not available for mouse, a comparison of XCI status of a homologous gene cluster in human and mouse was in strong support of more complete XCI in mouse (Tsuchiya *et al.*, 2004). When genes surrounding a known escapee in both human and mouse, *SMCX/Smcx* ('selected mouse cDNA on the X', which is now known to encode a histone demethylase), were examined for their XCI status, four additional escapees were found in human but *Smcx* remained the only escapee in this region in mouse.

It remains unclear why genes that escape would differ in different mammalian species and little is known about the XCI status of X-linked genes in other mammals. An initial multi-species comparison study was carried out using the 5' CGI methylation state as an indicator of the gene's XCI status (Jegalian and Page, 1998). According to a previous study of methylation states of 28 CGIs on the human X chromosome, methylation of the CGIs of the inactivated genes and demethylation of the CGIs of the active genes appeared to be a general rule. Jegalian and Page (1998) tested CGI methylation of *ALD* (which encodes the adrenoleukodystrophy protein), a gene known to undergo XCI in human and is presumably also silenced in other eutherian species, in 18 species representing nine eutherian orders. In all cases, methylation was observed in female, where *ALD* was expected to be silenced on the inactive X, and never observed in male, where the single X is active, in support with the correlation between CGI methy-

lation and the gene's activity status. They then tested CGI methylation of three known escapees in human in up to 19 species. Interestingly, each gene was shown to escape in a unique group of species, and no gene escapes in all species. Using the same technique, a second study compared the XCI status of seven X-linked genes in four mammals, and found a similar species-specific variation in escaping from XCI (Yen *et al.*, 2007).

Genes that escape from XCI provide a unique opportunity to study the mechanisms of propagation and maintenance of inactivation. Individual escapees were found to lack epigenetic features of inactivated genes, such as histone modification and DNA methylation (reviewed in Disteche, 2002). However, at the genetic level, no universal 'signature' of inactivated or escapee genes has been identified. No significant difference has been found between CGIs associated with inactivated and escapee genes. Previous studies of X;autosome translocation and XIC transgene showed that XCI can spread along the autosome, suggesting that it does not require genomic elements unique to the X chromosome for the propagation of XCI (Lee and Jaenisch, 1997; Rastan and Robertson, 1985). However, it was also noticed that inactivation of the autosome is not as efficient as inactivation of the X chromosome, which led to the hypothesis that there are 'way stations', possibly made of repeated sequences, along the X chromosome to facilitate spreading of the inactivation signals (Gartler and Riggs, 1983). In 1998, Lyon suggested that the LINE1 (L1 long interspersed elements) repeat family may be a good candidate for the 'way stations', as they are particularly concentrated on the X chromosome in both human and mouse, and previous X;autosome translocation and XIC transgene studies were consistent with XCI signals travelling better in LINE-rich regions. A systematic analysis of the distribution of LINE1 elements

in the human genome found a two-fold enrichment of LINE1 on the X chromosome compared to autosomes (Bailey *et al.*, 2000). They also found a significant increase of LINE1 content around XIC and significantly reduced LINE1 contents in the escape regions.

Escapees may occur from either failure to establish or failure to maintain inactivation. Current evidence is far from conclusive but does favour the latter scenario. Expression analysis of a known escapee, *Smcx*, in single cells from embryonic and adult mouse tissues suggested that this gene is inactivated in early embryonic development followed by re-activation (Lingenfelter *et al.*, 1998), perhaps due to progressive loss of CGI methylation (Disteche *et al.*, 2002). It is also not clear at which level inactivation is regulated. The multi-species comparison study carried out by Jegalian and Page (1998) suggested a gene-by-gene or cluster-by-cluster development of XCI following Y-gene decay. The presence of blocks of escapees in human indicates existence of chromatin domains of different XCI status (Carrel *et al.*, 1999), but studies in mouse showed that the escapee domain can be as small as a single gene (Tsuchiya *et al.*, 2004). The CTCF protein has been implicated in the enhancer-blocking function of chromatin insulators (Bell *et al.*, 1999; Lobanenkov *et al.*, 1990; Mukhopadhyay *et al.*, 2004). In a recent study, CTCF binding was detected at the 5' end of several escape genes adjacent to inactivated genes, but not genes embedded in large escaping domains (Filippova *et al.*, 2005).

1.4 XCI in other mammals

Most of our current knowledge about XCI comes from studies in human and mouse, both belonging to the extant mammal group Eutheria, or the 'placental' mammals. There are two other extant groups of mammals, the Metatheria, or marsupials, and the Prototheria, or monotremes. For convenience, the names eutherians, marsupials and monotremes are used in this thesis. Diverged from the monotremes $\sim$ 240 million years ago (MYA) (Murphy *et al.*, 2004), the eutherians and marsupials are together known as Theria, and the two lineages split $\sim$ 180 MYA (Murphy *et al.*, 2004). Marsupial and monotreme mammals have many very distinct features, some almost non-mammal like, yet they share some fundamental mammalian characteristics with the eutherians. Comparisons between these most distantly related mammals give us an insight of the rules and variations of mammalian genetics, and allow us to reconstruct the evolutionary past of the mammal lineage.

There are three orders and more than 200 species of marsupial mammals, most found in Australia, with one order found in South America, and only one species in North America (Wilson and Reeder, 2005). Like the eutherians, the marsupials are fur-bearing animals, give birth to live young, and feed the young with milk secreted from their mammary glands, but they have very distinct strategy of reproduction. Although they give live birth, the marsupials lack long gestation periods like the eutherians. Whereas eutherian young are born relatively developed, the marsupial young are born at a very immature stage and complete development attached to a teat, often protected in a pouch (the marsupium).

This kind of variation on a common mammalian theme is also seen in the ge-

netics of the sex chromosome. Marsupial and eutherian mammals share the same XY sex determination system. Comparative studies showed that genes located on the long arm of the human X chromosome are also X-linked in marsupials, indicating a common origin of the sex chromosomes in therian mammals (Spencer *et al.*, 1991b). However, the homology is limited to a subset of genes on the eutherian X chromosome - genes located on the distal short arm of the human X were found to be autosomal in marsupials (Spencer *et al.*, 1991a). Also, the marsupial X chromosome, representing 2-3% of the haploid genome, is much smaller than the eutherian X, which makes up 5% of the haploid genome (Graves and Watson, 1991). It has been postulated that the marsupial X chromosome may reflect the ancient X in the common ancestor of therian mammals, whereas a translocation from autosome added a large block of genes to the eutherian X before the radiation of eutherian mammals (Graves, 1995). The older and younger portions of present day eutherian X have been termed the X conserved region (XCR) and the X added region (XAR) (Figure 1.2). The marsupial Y chromosome, even smaller than the eutherian Y, has also lost most of its gene contents, thus the need for dosage compensation.

Like the eutherians, marsupial mammals also achieve dosage compensation through XCI (Cooper *et al.*, 1971; Richardson *et al.*, 1971; Sharman, 1971), but the exact mechanism of XCI appears to be different. In the marsupials XCI is imprinted, where the paternal X is always silenced (Cooper *et al.*, 1971; Richardson *et al.*, 1971; Sharman, 1971). In the eutherians, XCI is random in somatic cells, and imprinted XCI has only been observed in mouse and bovine extraembryonic tissues (Takagi and Sasaki, 1975; Xue *et al.*, 2002). There has therefore been speculation that imprinted XCI is the ancestral form of XCI in mammals (Huynh and

Lee, 2005). At the molecular level, the master switch gene that initiates random XCI in eutherians, *XIST*, is not present on the marsupial X (Hore *et al.*, 2007; Mikkelsen *et al.*, 2007). A recent search for sequences homologous to the human and mouse *XIST* and the flanking region in a number of vertebrates suggested that this region is disrupted in marsupials and monotremes (Hore *et al.*, 2007). This study also proposed that this unstable region underwent great expansion in the early eutherian evolution and provided start materials for the evolution of the XIC. The recently completed draft genomic sequences of a marsupial (short-tailed opossum) and a monotreme (platypus) have also confirmed the lack of *XIST* on both X chromosomes (Mikkelsen *et al.*, 2007; Warren *et al.*, 2008). The molecular events leading to the initiation and spreading of XCI in marsupials is not understood, but a link has been suggested to the meiotic sex chromosome inactivation (MSCI) in spermatogenesis (Hornecker *et al.*, 2007; Namekawa *et al.*, 2007). MSCI, previously only found in eutherian mammals, was recently identified in the grey short-tailed opossum (Hornecker *et al.*, 2007; Namekawa *et al.*, 2007). In opossum MSCI, the only X chromosome is silenced before fusing with the Y chromosome through a dense plate, and the silencing stably persists after meiosis, suggesting that the inactive state of the paternal X in a zygote may be inherited from the state in the gamete (Hornecker *et al.*, 2007; Namekawa *et al.*, 2007).

Furthermore, XCI in the marsupials appears to be less stable than XCI in the eutherians. Whereas in human and mouse only a small proportion of genes escape XCI (Carrel and Willard, 2005; Disteche *et al.*, 2002), expression of the repressed paternal allele was detected for all examined genes in marsupials (Cooper *et al.*, 1993). Studies of X-linked electrophoretic variants, *i.e.* allozymes, in a number of

marsupial species showed that the expression of paternal allele is very common but highly variable, displaying different patterns for different genes, tissues, and species. One of the most widely studied X-linked loci is *G6PD*, which undergoes stable random inactivation in human and mouse. In the marsupials, the paternal allele of *G6PD* is preferentially inactivated, but various degrees of paternal expression were seen in almost all tissues examined in the Virginian opossum, *Didelphis virginiana* (Samollow *et al.*, 1987). An even higher extent of de-repression of the paternal allele of *G6PD* was found in cultured cells arising from a number of Virginian opossum tissues (Migeon *et al.*, 1989). Again the extent of paternal expression varied greatly among different tissues and even different cell types in the same tissue, from no expression to full expression (Migeon *et al.*, 1989). Another widely studied gene, *PGK1*, also X-inactivated in human and mouse, displays a very different pattern in marsupials from that of *G6PD*. Cooper and colleagues (1977) observed partial expression of the paternal allele of *PGK1* in two Australian marsupials, the eastern grey kangaroo (*Macropus giganteus*) and the pretty-face wallaby (*M. parryi*), but Samollow and colleagues (1987) found only maternal expression in the American marsupial Virginian opossum. The unstable XCI in marsupials has been attributed to lack of maintenance by DNA methylation. In eutherian mammals, the 5' CGIs are usually methylated on the inactive, but not the active, X chromosome (Tribioli *et al.*, 1992). In marsupials limited evidence suggested that CGIs are not methylated on either X chromosome (Kaslow and Migeon, 1987; Loebel and Johnston, 1996). CGI methylation on the marsupial X chromosome is discussed in more detail in Chapter 5.

In human and mouse, XCI is established *via* a cascade of histone modifications, including hypoacetylation of core histones H3 and H4 (Boggs *et al.*, 1996;

Jeppesen and Turner, 1993). To exploit whether the same mechanism is employed in the marsupials, Wakefield and colleagues (1997) labelled metaphase chromosomes of Tammar wallaby fibroblasts with an antibody against acetylated histone H4. The autosomes were all brightly labelled, but the two X chromosomes showed a striking difference in the extent of labelling. While one X in female cells and the single X in male cells were brightly labelled, like the autosomes, the other X in female cells showed significantly weaker labelling, consistent with the inactive X being underacetylated. To date, histone hypoacetylation remains the only molecular feature known to be common to both eutherian and marsupial XCI.

The monotreme mammals, which are the most distantly related to the eutherians, share even less similarities with them. There are currently only five known species of monotremes, the duck-bill platypus, and four species of echidnas, all found in Australia (Wilson and Reeder, 2005). Like the marsupials, the monotremes are also most famous for their distinct mode of reproduction, and are sometimes referred to as the 'egg-laying mammals'. Despite the reptilian-like birth, the young are hatched at a less developed stage than birds and reptiles, and typical of mammals, rely on the milk of their mothers during the initial period of their life. As monotreme females lack nipples, the young suck milk directly through the abdominal skin.

The genetics of monotreme mammals reflects their unique evolutionary position. Although monotremes employ an XY sex determination system, their sex chromosomes are highly unusual, consisting of multiple sex chromosomes that form a translocation chain during meiosis (Rens *et al.*, 2004). The monotreme sex chromosomes appear to have different evolutionary origins from those of the marsupials and eutherians. X-linked genes in the therian mammals map to auto-

somes in platypus, while considerable homology is shared between the platypus X chromosome and the chicken Z chromosome (Veyrunes *et al.*, 2008). In addition, the male-determining gene common to the therian mammals, *SRY*, is not present in the monotremes (Wallis *et al.*, 2007). It is not clear whether X-linked genes in monotreme mammals are dosage compensated, but given the different origin of the monotreme sex chromosomes, there is no obvious reason to expect that XCI will occur as the mechanism in monotremes.

1.5 Aims of the thesis

Genes that escape from XCI present a unique opportunity to improve our understanding of the extraordinary biological process of X chromosome inactivation. To date, the XCI status has only been characterised for genes on the human X chromosome and a small number of genes on the mouse X chromosome. It is desirable to catalogue the XCI status of a greater number of genes in multiple species, as comparisons between different species will greatly help in understanding how the landscape of silencing has evolved.

To achieve this, it is critical to have a method to assess the XCI status of individual genes in multiple species. Currently, the vast majority of XCI data in human were generated by the somatic hybrid method and most studies in mouse have made use of the T16H translocation. Both experimental systems are difficult to establish and not suitable for extending to other species. In human and mouse, CGI methylation on the inactive X has been implicated in the maintenance of XCI and has been used previously to indicate a gene's XCI status. Therefore, I have chosen to study the 5' CGI methylation states of a larger number of genes

as surrogate for their XCI status.

The four main aims in this thesis are:

1. To test whether CGI methylation serves as a good indicator for a gene's XCI status by comparison of methylation state with known XCI profile for a number of genes in human and mouse (Chapter 3).

2. To search for differences in CGI methylation of X-linked genes between human and mouse; and to explore whether such differences, if existing, are parallel to the differences in XCI profiles between the two species (Chapters 3 and 4).

3. To investigate the existence or absence of CGI methylation of X-linked genes in marsupials, as the small amount of available evidence points to lack of methylation on the X_i (Chapter 5).

4. To compare characteristics of CGIs in the homologous region of X chromosome between human, mouse, and opossum (Chapter 6).