DYNAMICS OF DNA METHYLATION IN NORMAL AND CLONED SHEEP DEVELOPMENT

Hannah R. Fairburn

Ph.D.

The University of Edinburgh

2002

Declaration

I hereby declare that this thesis has been composed by myself and has not been submitted for any previous degree. Acknowledgment of specific procedures not performed by myself is clearly stated throughout this thesis, otherwise, the work described herein is my own.

Hannah Fairburn

Date: 16/8/02

Abstract

Methylation of the C-5 position of cytosine is a stable, heritable and reversible epigenetic modification of DNA that is widespread in vertebrate animals, plants and fungi. In vertebrates, it is predominantly involved in transcriptional repression, with known roles in genomic imprinting, X inactivation and silencing of repetitive DNA. The levels of DNA methylation in murine embryos undergo significant changes post-fertilisation during an event known as epigenetic reprogramming. This phenomenon has been widely investigated in the mouse and to a lesser extent in other vertebrate species. The purpose of this study was to investigate DNA methylation in sheep (*Ovis aries*), a species of renewed interest since the birth of a female sheep (Dolly) produced by somatic cell nuclear transfer (SCNT). Significantly, it has been hypothesised that the low success rates observed in SCNT may be related to incomplete or inefficient reprogramming of DNA methylation patterns during early stages of development.

The main components of the DNA methylation repression system are the cytosine methyltransferases (maintenance (Dnmt1) and de novo (Dnmt3a and 3b)) and the methyl-CpG specific transcription repressors (MeCPs or MBDs). Cloning of ovine cDNA sequences, and expression analysis of genes encoding proteins involved in DNA methylation, yielded insights into DNA methylation processes in early development and also in adult tissues. In addition, techniques were developed to analyse genomic DNA methylation levels in single preimplantation embryos, and were also used to assess the effects of somatic cell nuclear transfer on DNA methylation reprogramming.

The ovine homologues of DNA methyltransferases *DNMT1*, *DNMT3A* and *DNMT3B*, and the methyl-CpG-binding protein *MBD2*, were cloned fully or partially, using a combination of RT-PCR, cDNA library screening, and 5' RACE methods. Analysis of predicted protein sequences demonstrated high levels of conservation within and out with functional domains. Attempts to clone an oocyte-specific form of *DNMT1*, known to exist in the mouse, were unsuccessful. Instead, a form identical to that isolated from adult tissues was detected in oocytes. These

results imply there is little or no oocyte-specific form of Dnmt1 in sheep oocytes, which may be indicative of species-specific differences during early epigenetic reprogramming events. Analysis of *DNMT1*, *DNMT3A* and *MBD2* expression in adult tissues using Taqman RT-PCR, demonstrated similar patterns to those observed in the human, where previously reported. Qualitative expression analysis of *DNMT1* expression in ovine oocytes and preimplantation embryos of various stages, suggested comparable expression patterns to those observed in the mouse, with depletion of maternal stores observed between the 1-cell and blastocyst stages. Expression analysis of *DNMT3A* demonstrated a reciprocal increase between these stages, and analysis of *DNMT3B* expression indicated a similar pattern, again increasing at the blastocyst stage. *MBD2* was not found to be highly expressed in oocytes or blastocysts.

A methyl-sensitive PCR method, optimised for use in single oocytes and embryos, was employed to assay DNA methylation levels of an ovine satellite sequence, and an ovine SINE element. No significant changes were detected in the satellite sequence of *in vivo* derived embryos between the 8-16 cell and blastocyst stages of development, contrasting with the significant decrease in satellite DNA methylation levels observed in the mouse. However, comparisons between *in vivo*, *in vitro* fertilised (IVF) and SCNT blastocysts demonstrated significant differences, with the IVF and SCNT embryos (both of which undergo equivalent periods of *in vitro* culture) relatively undermethylated within the satellite and SINE element. In addition, SCNT blastocysts demonstrated increased methylation levels relative to their IVF counterparts within the satellite sequence, but not the SINE element, suggesting sequence-specific reprogramming errors.

In conclusion, there is strong evidence of overall conservation of DNA methylation mechanisms between sheep and mice however, there are subtle differences. In addition, SCNT blastocysts appear to show aberrant reprogramming of methylation patterns when compared to their normal counterparts. This may underlie the low success rates observed in animal SCNT. Future work will build on these results, investigating global methylation levels during preimplantation development of normal and cloned embryos, and also determining the precise roles of ovine DNA methyltransferases in methylation reprogramming.

Publications

Fairburn, H.R, Young, L.E. and Hendrich B.D. Epigenetic reprogramming: how now, cloned cow? *Current Biology* 2002 **12** R68-R70. Review.

Young, L.E. & **Fairburn, H.R**. Improving the safety of embryo technologies: possible role of genomic imprinting. *Theriogenology* 2000 **53**(2):627-48. Review.

Fairburn, H.R. & Young, L.E**.** Cloning and expression analysis of DNA methylation enzymes in sheep. *J Reprod Fertil* 2000 Abstract Series **25**, p42

Acknowledgements

I would like to extend my sincere gratitude to Dr. Lorraine Young and Dr. Richard Meehan for their continual guidance and encouragement over the last three years. I would also like to thank Prof. Ian Wilmut for his support and helpful discussions.

Special thanks to Ken Fernandes, for sharing his wealth of practical expertise and enthusiasm. Further thanks to Jane McCracken, John Gardner, Nathalie Beaujean and Jennifer Berger for practical assistance, helpful discussions and making the lab a pleasant place to be.

I would like to thank those who helped with the provision of nuclear transfer embryos (Ali Ainslie, Tricia Ferrier, Linda Harkess and Bill Ritchie), and the staff at the L.A.U. (John Bracken, June Bowering, Tim King and Marjorie Ritchie) for providing *in vivo* embryos. Thanks also to Mike Clinton and Charis Hogg for help with the Taqman PCR and Dave Waddington for assistance with statistical analysis.

I would also like to thank friends and family, especially Tony, whose support and patience over the last few years has been greatly appreciated.

Finally, I would like to thank the BBSRC and Geron for providing financial assistance to support this research.

Contents

List of Figures

List of Tables

Abbreviations

aa	Amino acids
bp	Base pair(s)
BSA	Bovine Serum albumin
cDNA	complementary DNA
dH_2O	Distilled water
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNMT	DNA methyltransferase
dNTP	Deoxynucleotides
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
ES	Embryonic stem (cell)
g	Gramme
ICM	Inner cell mass
IVF	*In vitro* fertilisation
IVM	*In vitro* maturation
Kb	Kilo basepairs
l	litre
LB	Luria-Bertani
μ	Micro
m	Milli
M	Molar (moles/litre)
m^5C	5-Methyl cytosine
MBD	Methyl-binding domain
mRNA	Messenger RNA
n	Nano
p	Pico
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pg	Picogram
pK	Proteinase K
pmol	Picomole
R.I.	Roslin Institute
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribosomal RNA
RT-PCR	Reverse-transcription PCR
SCNT	Somatic cell nuclear transfer
SDS	Sodium Dodecyl Sulphate
SINE	Short interspersed repetitive element
SSC	Sodium chloride sodium citrate
TBE	Tris-borate-ethylenediaminetetaacetic acid
V	Volts
ZGA	Zygotic genome activation

Gene nomenclature

Due to a lack of clarity regarding gene nomenclature for farm animals, this thesis uses the recommended human notation (http://www.gene.ucl.ac.uk/nomenclature/) (Dolling et al., 1997) when describing sheep or cow genes/transcripts and proteins. To summarise, this involves the use of italicised capital letters for the gene/ transcript and non-italicised capital letters for the corresponding protein. Mouse genes or transcripts will be referred to by lower case italicised symbols, with lower case, non-italicised symbols for the protein.

Thesis rationale

DNA methylation is a stable, reversible and heritable modification of DNA with known involvement in a number of biological processes including genomic imprinting, X-inactivation and silencing of repetitive DNA. In each of these processes, it is thought to function to repress transcription. Dynamic changes in DNA methylation patterns occur in murine preimplantation development, as part of the epigenetic reprogramming process that resets gametic chromatin into a form that will support fetal development. The extent of conservation of this reprogramming event in other mammalian species is poorly understood. In addition, it has been hypothesised that DNA methylation patterns may be reprogrammed incompletely in embryos produced by somatic cell nuclear transfer (SCNT).

The aim of this thesis was to initiate characterisation of the DNA methylation system in sheep and in particular, to investigate its dynamics in preimplantation development. The methods used to analyse DNA methylation in normal preimplantation development were then applied to compare the extent of methylation reprogramming in *in vivo* derived, *in vitro* fertilised (IVF) and SCNT embryos.

Chapter 1

Introduction

1.1 Epigenetic modification of DNA

The term "epigenetic" was coined to describe the inheritance of gene expression patterns conferred not by the DNA sequence itself, but by mechanisms regulating the transcription of DNA. The field of epigenetic study was initiated by Conrad Waddington (1942), at a time when many scientists believed physical change in the genome (e.g. the loss of unnecessary genes) was the causal factor for cellular differentiation. He defined epigenetics as "the study of the processes by which genotype give rise to phenotype". This definition remained untouched for several decades, until Robin Holliday proposed two independent definitions encapsulating changes in gene expression occurring at both the adult and developmental stages of growth (Holliday, 1987 & 1994). These terms were combined to form the current day definition of epigenetics as the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence.

Since the initial musings of Waddington, mounting interest in epigenetics has led to the identification of an increasing number of DNA or chromatin modifications and RNA based mechanisms. These include the modification of DNA itself (DNA methylation; reviewed by Bird, 2002), modification of histones (reviewed by Rice & Allis, 2001), higher order chromatin structure (reviewed by Lyko & Paro, 1999; Grewal & Elgin, 2002) and RNA interference (Matzke *et al.*, 2001). Most of these processes are potentially 'epigenetic', but too little is known of their specific requirement in developmental processes to determine whether they fulfil a crucial obligation of a true epigenetic modification, heritability. However, one of these modifications is known to be truly epigenetic in nature: DNA methylation. Of all of the known DNA regulatory

mechanisms, DNA methylation has been most intensely investigated and is known to be stable, reversible (Doerfler *et al.*, 1998) and most importantly, clonally heritable (Bestor & Verdine, 1994).

1.2 DNA methylation

DNA methylation is the covalent modification of DNA involving the addition of a methyl group to the cyclic carbon 5 of a cytosine residue (m^5C) and is a process prevalent in both prokaryotic and eukaryotic organisms. The main role of DNA methylation in prokaryotes is as part of the host defence system, where marking of the bacterial genome with methylated cytosine protects the host genome from destruction by its own restriction endonucleases (reviewed by Hubacek, 1992). DNA methylation is also present in all four groups of eukaryotes (plants, animal, fungi and protists), albeit to different extents. Most invertebrate species have moderate to high levels of m^5C, however, this is not true in all cases. The fruit fly, *Drosophila melanogaster*, contains low levels of m^5C, predominantly at CpT residues (Lyko *et al.*, 2000), and the nematode worm (*Caenorhabditis elegans*) completely lacks detectable m^5C (Simpson *et al.*, 1986). DNA methylation is known to be widespread in plants (reviewed by Martienssen & Colot, 2001), with all species examined demonstrating high m^5C levels. In the case of fungi, some species present methylated genomes (*Neurospora crassa*, *Candida albans*) whereas methylation is undetectable in close relatives such as *Saccharomyces cerevisiae* (Selker, 1998; Russell *et al.*, 1987; Proffit *et al.*, 1984).

Vertebrate species demonstrate high levels of m^5C, dispersed throughout the genome. This contrasts with the distribution observed in invertebrate species, where methylation patterns are generally restricted to non-coding regions. Even amongst vertebrate species, substantial differences in methylation are detectable, for example, rRNA-encoding genes are highly methylated in fish and amphibian species, but are relatively undermethylated in mammals, birds and reptiles (Tweedie *et al.*, 1997). It has been proposed that the differential distribution of DNA methylation amongst vertebrates, and

in eukaryotes in general, may reflect an evolutionary system for the implementation of alternative DNA methylation functions (Colot & Rossignol, 1999).

In mammals, methylated cytosine residues are found predominantly adjacent to guanine residues, within the so-called CpG dinucleotide. In a typical differentiated human cell, 70-80% of CpG dinucleotides are methylated, with most unmethylated CpG residues residing in GC-rich regions termed CpG islands localised within the 5' regions of genes (Bird, 1986). Approximately 60% of human genes are predicted to contain CpG islands, a significant number of which remain unmethylated throughout development and in adult tissues (Antequera & Bird, 1993). Recent studies in mice have also identified methylated cytosine residing in CpA and CpT dinucleotides, although this appears to be restricted to ES cells and the functional significance has not yet been determined (Ramsahoye *et al.*, 2000).

Extensive studies of CpG island methylation in the mouse has demonstrated its involvement in a wide variety of biological processes including X-inactivation (Panning & Jaenisch, 1996), genomic imprinting (Li *et al.*, 1993), carcinogenesis (reviewed by Baylin & Herman, 2000) and ageing (reviewed by Issa, 2000). There is also increasing evidence that DNA methylation is required for the regulation of tissue-specific and developmental stage-specific genes (Jackson-Grusby *et al.*, 2001; Stancheva *et al.*, 2002). In each of the aforementioned processes, DNA methylation is likely to be present in the capacity to repress transcriptional activity.

In addition, cytosine methylation occurs outwith CpG islands, predominantly to silence transposable elements including SINE and LINE sequences, and intracisternal A particle (IAP) elements (Liu *et al.*, 1994; Woodcock *et al.*, 1997; Walsh *et al*, 1998). Presumably, this action is required to silence the promoters of these DNA elements, thus preventing DNA damage by unconstrained transposition, the so-called 'genome defense' model (Yoder *et al.*, 1997). Centromeric satellite DNA is also normally highly

methylated (Miller *et al.*, 1974), most likely reflecting the highly repressed nature of heterochromatic regions.

1.2.1 The establishment and maintenance of DNA methylation patterns

The methylation of DNA is catalysed by a family of enzymes known as the DNA methyltransferases. These enzymes function through a highly conserved catalytic domain containing a set of conserved sequence motifs. The enzyme utilises S-adenosylmethionine (SAM) as a substrate, transferring a methyl group to carbon 5 of the cytosine residue and creating a covalent modification of the DNA. Several DNA methyltransferases with a demonstratable methylating function both *in vitro* and *in vivo* have been identified in mammals. DNA methyltransferase 1 (*Dnmt1*) was the first to be cloned in the mouse (Bestor *et al.*, 1988), and is known as the 'maintenance methyltransferase' due to its preference for methylating hemi-methylated DNA (Bestor 1992) The recognition and methylation of hemi-methylated DNA allows the clonal maintenance of methylation patterns in newly replicated DNA. In addition, Dnmt1 protein has been shown to be localised at the replication focus (Leonhardt *et al.*, 1992), suggesting a direct coupling of DNA methylation with DNA replication. This notion is further backed by the identification of a PCNA (proliferating cell nuclear antigen) binding site within the Dnmt1 protein sequence (Chuang *et al.*, 1997).

The establishment of new methylation patterns is known generically as *de novo* methylation, a process thought to be independent of Dnmt1 activity and thus requiring a different methyltransferase enzyme, or set of enzymes. The recent identification of four novel mammalian methyltransferase genes provided potential candidates for this unique role. One gene encodes the Dnmt2 protein, containing the conserved methyltransferase motifs identified in both prokaryotic and eukaryotic methyltransferases (Yoder *et al.*, 1998). However, this protein has so far failed to demonstrate either maintenance or *de novo* methylating activity *in vitro* and methylation levels in ES cells carrying a targeted deletion of *Dnmt2* demonstrate no reduction in methylation levels (Okano *et al.*, 1998b).

Dnmt3a and *Dnmt3b* are two closely related methyltransferase genes identified through database homology searches for genes containing conserved methyltransferase motifs and, in contrast with Dnmt2, have demonstrated methylating activity *in vitro* (Okano *et al.*, 1998a). More significantly, they did not show preference for hemi-methylated DNA in experiments performed *in vitro*, raising the possibility that they were the long sought after *de novo* methyltransferases. Indeed, further experiments confirmed that both Dnmt3a and Dnmt3b are required for genome-wide *de novo* methylation (Okano *et al.*, 1999). Although Dnmt3a and Dnmt3b are now widely recognised as *de novo* methyltransferases, the mechanisms by which they target specific DNA sequences remains unknown.

A further gene belonging to the *Dnmt3* family, *Dnmt3L*, was recently identified, although the lack of methyltransferase motifs would suggest no direct role in the methylation of DNA (Aapola *et al.*, 2001). Interestingly, targeted disruption of *Dnmt3L* in mice leads to a loss of maternal methylation imprints, implying Dnmt3L may be indirectly involved in the establishment of DNA methylation patterns (Bourc'his *et al.*, 2001). **Figure 1.1** demonstrates the structural relationships between all known mammalian DNA methyltransferases.

In addition to Dnmt3L, there appear to be several other proteins with that are inextricably linked with the maintenance or establishment of methylation patterns but have no obvious methylating ability. The first was identified in the plant *Arabadopsis thaliana* and named *DDM1* (Decrease in DNA Methylation) and was shown to be required for the maintenance of m^5C mostly at repeated sequences (Vongs *et al.*, 1993). Sequence analysis of this gene demonstrated that it does not encode a DNA methyltransferase, but instead encodes a member of the SNF2-like helicase family, members of which are known to be involved in chromatin remodelling (Jeddeloh *et al.*, 1999).

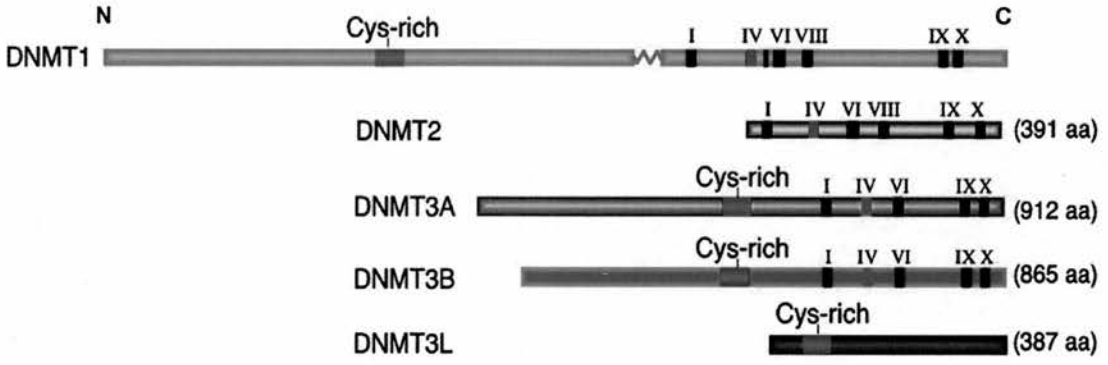


Figure 1.1 The DNA methyltransferases of mammals.

Taken from Bestor (2000). All mammalian DNA methyltransferases contain the canonical methyltransferase motifs within the C-terminal domain except Dnmt3L, which is not thought to perform methylating activities *in vivo*. The N-terminal domain, which is responsible for specific protein-protein interactions, varies between Dnmt1 and the Dnmt3 family, but is completely absent in Dnmt2.

Mutations in a second protein, ATRX, known to cause development abnormalities including mental retardation, facial dysmorphism and α-thalassemia (Gibbons *et al.*, 1995) also demonstrated a similar epigenotype to *DDM1* mutants in mice and humans, with changes in DNA methylation pattern observed in several highly repeated sequences (Gibbons *et al.*, 2000). Interestingly, the ATRX protein, also a member of the SNF2-like helicase family (McDowell *et al.*, 1999), shares a highly conserved zinc finger domain with the DNMT3 family of proteins (Aapola *et al.*, 2000), raising the possibility that various interactions between these proteins could occur to establish or maintain sequence-specific methylation patterns. A third member of the SNF2-like helicase family, Lsh (Lymphoid specific helicase; a misleading name as it is expressed ubiquitously in mouse fetuses (Raabe *et al.*, 2000)), was also recently shown to be required for maintenance of normal methylation levels, particularly in repetitive DNA sequences but also in single-copy genes including β-*globin*, *PgK-2* (phosphoglycerate

kinase) and the imprinted gene, *H19* (Dennis *et al.*, 2001). As with ATRX, the apparent developmental requirement for Lsh and sequence-specific loss of methylation, may indicate a requirement for SNF-helicases in aiding the establishment of region-specific methylation patterns at certain stages of development, perhaps by making the DNA more accessible to DNA methyltransferases.

The processes of *de novo* and maintenance methylation are likely to be interlinked, with individual DNA methyltransferases cooperating to establish and maintain methylation patterns. This has been demonstrated in experiments expressing Dnmt1 and Dnmt3a in transgenic fruitflies (*Drosophila melanogaster*), a species with little detectable m^5C and thus arguably a good system for analysing *in vivo* function of DNA methyltransferases (Lyko *et al.*, 1999). Although this experiment ultimately demonstrated a non-tolerance for DNA methylation and DNMT expression in fruitflies (resulting in a lethal phenotype), interesting observations were made concerning the interaction of DNMTs. Flies co-expressing Dnmt1 and Dnmt3a demonstrated increased genomic m5C levels when compared to flies expressing only Dnmt3a, suggesting significant cooperation of the two enzymes to establish and maintain methylation patterns.

More recently, biochemical analyses of Dnmt3a *in vitro* have demonstrated a hemi-methylating action of Dnmt3a (Yokochi & Robertson, 2002), creating a possible scenario whereby Dnmt1 recognises and completes the methylation of the double stranded DNA template. This concurs with the previously described findings of Lyko *et al.* (1999). The potential interactions between known DNA methyltransferases in relation to the dynamics of DNA methylation are summarised in **Figure 1.2**.

Although specific roles have been defined for each of the DNA methyltransferases identified so far, for example, in establishing or maintaining methylation patterns, the possibility cannot be excluded that these roles are interchangeable in certain circumstances. Indeed, in cancer cells lacking a functional *DNMT1* gene, genomic DNA methylation patterns are stably maintained (Rhee *et al.*, 2000), perhaps demonstrating an

ability of DNMT3A or DNMT3B to recognise hemi-methylated DNA as a substrate. Whether alternatative functions of DNMTs are observed solely as a result of cellular transformation, or whether there is an inherent ability present in all cell types to switch DNMT roles, remains unknown. In addition, the discovery of further mammalian DNA methyltransferases cannot be discounted at this time.

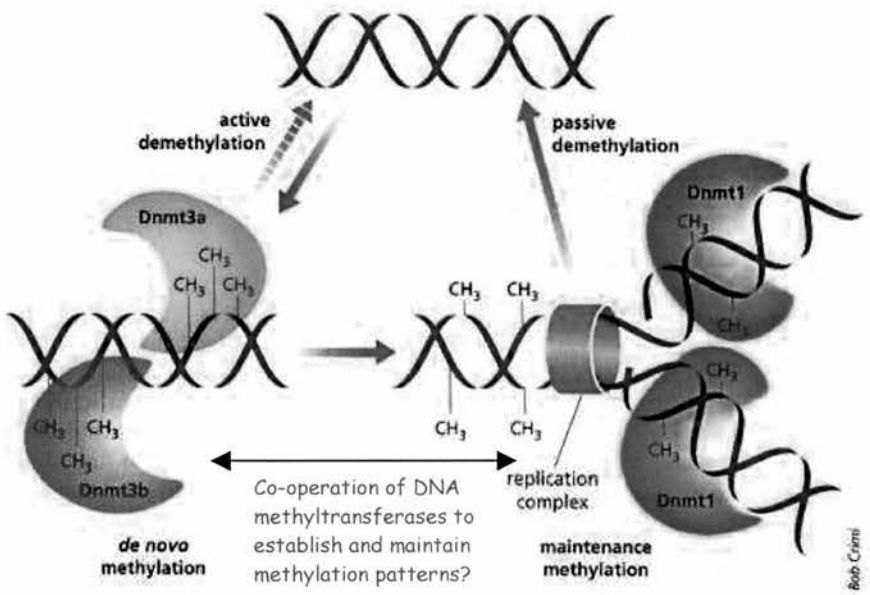


Figure 1.2 Dynamics of DNA methylation pattern

Adapted from Reik *et al.* (1999). *De novo* methyltransferases Dnmt3a and Dnmt3b are generally thought to establish methylation patterns, which are subsequently maintained by Dnmt1. It is possible that the roles of Dnmt1, Dnmt3a and Dnmt3b are at least partially interchangeable.

1.2.2 Demethylation

As DNA methylation is associated with transcriptional repression, it follows that demethylation would be linked to gene activation. Demethylation could arise gradually via passive mechanisms, for example, by preventing maintenance methylation activities through subsequent cell cycles (passive demethylation). Conversely, methyl groups could be stripped from DNA, involving an enzymatic process (active demethylation). The search for an enzyme capable of demethylating DNA (the so-called "demethylase") has been a long and uphill struggle, with plenty of false starts and significant controversy.

Several studies performed in the late 1980s and early 1990s demonstrated that active demethylation can occur within tissue-specific genes (Razin *et al.*, 1986; Paroush *et al.*, 1990; Saluz *et al.*, 1988). The purification of an extract demonstrating active demethylating properties was first reported by Jost (1993), and was isolated from chicken nuclear extracts. Further analysis of the complex comprising the demethylating activity demonstrated the presence of both RNA and proteins, and a preference for hemi-methylated DNA (Jost *et al.*, 1997). More precisely, the mode of action involves the removal of methylated cytosine by m^5C DNA glycosylase activity, a component of which was later identified as a G/T mismatch DNA glycosylase (Zhu *et al.*, 2000).

In 1996, a breakthrough in the search for a mammalian demethylase was apparent, with the discovery of an RNA-associated enzyme activity in rat myoblast cells (Weiss *et al.*, 1996). However, further examination of this ribozyme-like demethylating activity demonstrated no RNA component involvement (Swisher *et al.*, 1998). In 1999, a second report identifying a novel mammalian demethylase protein was released. This putative demethylase was a truncated variant of the methyl-binding protein, MBD2, (named MBD2B) and demonstrated the ability to cleave methyl groups from cytosine residues, releasing methanol as a by-product (Bhattacharya *et al.*, 1999). However, soon after the reported identification of MBD2B as a mammalian demethylase, two independent

publications emerged describing the presence of MBD2B within a transcriptional repressor complex (Ng *et al.*, 1999; Wade *et al.*, 1999). As demethylation is often associated with gene activation, not repression, the authors of the two 'repression' reports sought to recreate the experiments demonstrating the demethylating activity of MBD2B. Unfortunately, these experiments failed and combined with the observation of unchanged global methylation levels in mice carrying a targeted deletion of MBD2 (Hendrich *et al.*, 2001), the demethylating activity observed by Bhattacharya and colleagues remains in doubt.

After years of experiments, the search for a mammalian demethylase is still in progress, with attention returning to m^5C DNA glycosylases. The recent identification of MBD4 as a human protein with G/T mismatch DNA glycosylase activity (Hendrich *et al.*, 1999a) provides a potential candidate. Analysis of recombinant MBD4 confirmed the presence of DNA glyosylase activity, although in contrast with the chicken, this is inhibited by the presence of RNA (Zhu *et al.*, 2000). However, targets of this specific mode of demethylation have yet to be defined and it is also unclear as to whether this protein is involved in the large-scale demethylation events observed during early mouse embryogenesis (discussed in section **1.4.4**).

1.3 DNA methylation and transcriptional repression

Although the specific roles of DNA methylation in eukaryotic organisms vary substantially, they share a common theme of transcriptional repression. Approximately 60-90% of the CpG dinucleotides within the mammalian genome are methylated (Bird, 1986), with the unmethylated CpG dinucleotides normally concentrated in CpG-rich regions termed CpG islands (Bird, 1986). CpG islands tend to be localised within the promoter regions of many RNA polymerase II transcribed genes (Antequera & Bird, 1993). There have been several models proposed over the years to describe how DNA methylation may act to repress transcription; these are illustrated in **Figure 1.3.**

The model currently most favoured is that of DNA methylation-mediated repression through the recruitment of methyl-binding proteins and associated repression complexes (**Fig. 1.3c**). The best described methyl-binding protein MeCP2, was first identified in the late eighties (Meehan *et al.*, 1989). This protein was shown to preferentially bind methylated DNA (methyl-Cp-G), via the methyl-Cp-G binding domain (MBD). MeCP2 also contains a transcriptional repression domain (TRD) that interacts with the co-repressor Sin3a, which in turn, binds histone deacetylase (Nan *et al.*, 1998) ultimately leading to the formation of a silent chromatin state.

Recently, several other methyl-binding proteins (MBD1, MBD2, MBD3 and MBD4) have been identified, each containing the highly conserved methyl-CpG-binding domain (Hendrich & Bird, 1998). All of these proteins, with the exception of MBD3, demonstrate affinity for methylated DNA, but the precise roles of these proteins vary significantly. MBD1 is involved in HDAC-associated transcriptional repression (Ng *et al.*, 2000), as are MBD2 and MBD3. There are also several reports suggesting MBD2 and MBD3 directly interact as components of the methylation-associated repression complex (Hendrich *et al.*, 2001; Zhang *et al.*, 1999; Tatematsu *et al.*, 2000). Furthermore, Mbd2, alongside a second recently identified methyl-binding protein named Kaiso, have been identified as part of the methylation-associated repression complex initially identified as MECP1 (Ng *et al.*, 1999; Prokhortchouk *et al.*, 2001). As mentioned previously, MBD4 has been identified as a thymine glycosylase, involved in DNA repair of deaminated methylated CpG residues (Hendrich *et al.*, 1999a). The extent of involvement of the MBD-containing proteins in global transcriptional repression has yet to be determined. Little is known of the sequences and genes repressed by this particular system.

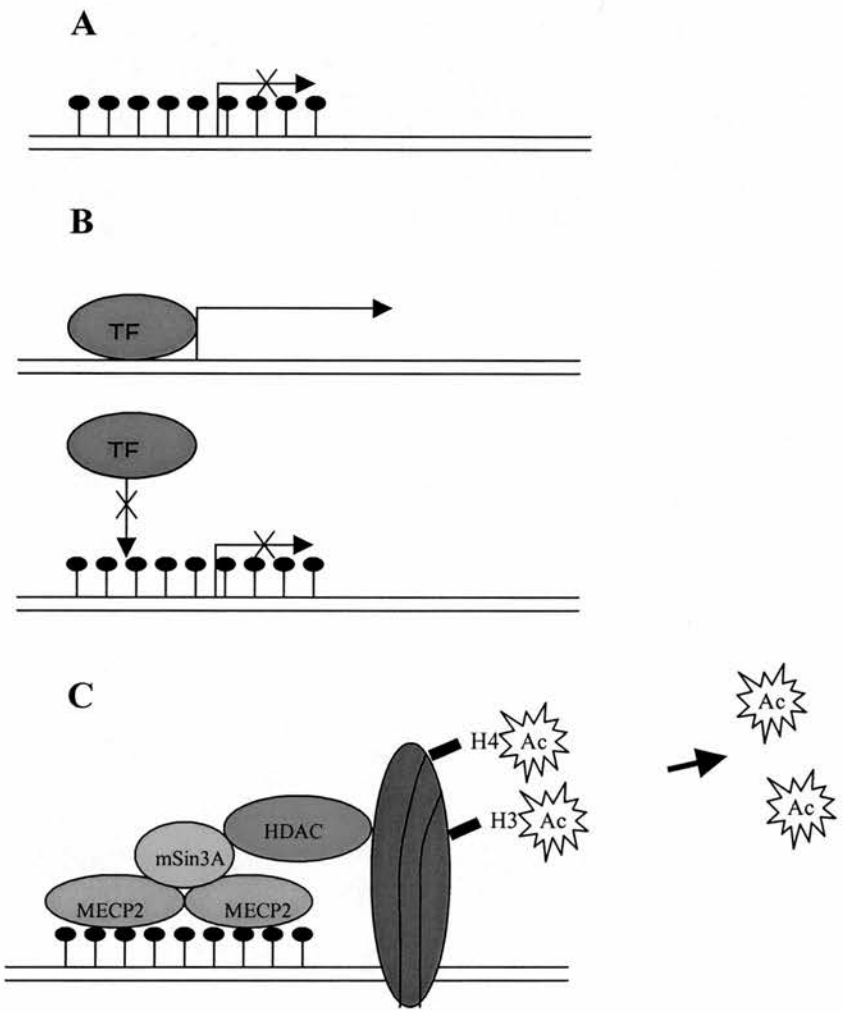


Figure 1.3 Possible modes of transcriptional repression by DNA methylation

(A) Direct repression: DNA methylation of CpG dinucleotides within the promoter region prevents RNA polymerase II from passing through this region, halting transcription. (B) DNA methylation prevents transcription factors (TF), or other proteins associated with transcriptional activation, from binding to the promoter region. (C) Methylated DNA recruits methyl-binding proteins, which in turn recruit transcriptional repression complexes, such as the mSin3A/HDAC complex. This leads to deacteylation of histone tails, and the formation of a compact, inaccessible chromatin structure.

An interesting observation demonstrating an additional level of complexity to the methylation-mediated transcriptional repression mechanism was presented recently. The DNA methyltransferase Dnmt1, was shown to associate with histone deacetylases HDAC1 and HDAC2 (Fuks *et al.*, 2000). This potentially allows the maintenance of transcriptionally repressed chromatin during and immediately after replication. An alternative interpretation suggests Dnmt1 may require HDAC activity for it to function correctly; chromatin remodelling may be required to allow Dnmt1 access to the DNA. A second paper demonstrated that this association was mediated by a co-repressor named DMAP1 (DNMT-associated protein 1; Rountree *et al.*, 2000). More recently, the putative *de novo* methyltransferase, Dnmt3a, has also been found to associate with histone deacteylase (Fuks *et al.*, 2001) and is recruited by a sequence-specific DNA binding factor, RP58. Furthermore, Dnmt3a is capable of transcriptional repression independent of its methylating activity (Bachman *et al.*, 2001: Fuks *et al.*, 2001).

1.4 Specific roles of DNA methylation in gene silencing

Since the initial discovery of DNA methylation, several biological processes have been identified in which DNA methylation plays a significant role in gene silencing. Some of these roles are more well-defined, such as allele-specific silencing of imprinted genes, whereas others remain controversial (tissue-specific gene silencing). Here, the nature of DNA methylation involvement in a number of silencing processes is described.

1.4.1 Genomic Imprinting

Imprinted genes are by definition, those that are only expressed from one allele; either the maternal or paternal allele. Genomic imprinting is a phenomenon found only in placental mammals and is thought to be an evolutionary adaptation to address competing parental contributions to embryo size and survival (the so-called 'parental conflict hypothesis'; Moore & Haig, 1991). Imprinted genes have been identified and studied in several species including the human and sheep; however, the most of our current

knowledge about the mechanisms of genomic imprinting comes from experiments performed using the mouse as a model species.

The epigenetic marks, or 'imprints', required for the regulation of imprinted gene expression, are established in the germlines of each parent and transmitted to the embryo after fertilisation (Latham, 1999). DNA methylation has properties that meet the full requirements for the imprinting mark; it affects gene expression (Nan *et al.*, 1998), it is heritable (Bestor & Verdine, 1994) and is also reversible (Doerfler *et al.*, 1998). DNA methylation has been shown to be essential to imprinting (Li *et al.*, 1993) as mice with a targeted deletion of *Dnmt1* demonstrate inappropriate expression (biallelic expression, or complete silencing) of a number of imprinted genes including *H19, Igf2* and *Igf2r*. However, there is no clear indication of whether DNA methylation establishes the imprinting mark, and differential methylation may be important in stabilising the imprint and regulating gene expression rather than being the primary signal.

In imprinted genes, DNA methylation is generally localised within so-called differentially methylated regions (DMRs). DMRs of imprinted genes are *de novo* methylated or demethylated in a sex-specific manner during gametogenesis, establishing the permanent methylation pattern (Obata & Kono, 2002). In addition, further allele-specific alterations in methylation occur in some imprinted genes during embryogenesis, maintaining imprinted control of gene expression at appropriate developmental stages (Brandeis *et al.*, 1993; Olek & Walter, 1997).

DNA methylation within a DMR can lead to either repression or activation of an imprinted allele. For example, the maternally expressed murine *Igf2r* gene contains two differentially methylated regions, one within the promoter, the second within intron 2 (Stoger *et al.*, 1993). In the maternal allele, DMR1 remains unmethylated, whilst DMR2 is methylated, allowing transcription to occur from the maternal promoter. A reciprocal pattern of methylation is observed on the paternal allele, with direct silencing of the paternal promoter through methylation, and indirect silencing via the production of an

antisense transcript from the unmethylated DMR2, extending into the promoter region (Wutz *et al.*, 1997; Lyle *et al.*, 2000).

DNA methylation outwith DMRs can also play a role in the regulation of imprinted gene expression. The mouse imprinted *Igf2-H19* loci contains two reciprocally imprinted genes that are co-ordinately regulated via a common imprinting control centre (reviewed by Thorvaldsen & Bartolomei, 2000). This imprinting centre is differentially methylated and serves to function as a CTCF-binding insulator (Szabo *et al.*, 2000). When the imprinting centre is unmethylated, CTCF can bind, preventing enhancers downstream of *H19* from interacting with the upstream *Igf2* promoter. When this site is methylated, CTCF cannot bind, and the enhancers are free to interact with the *Igf2* promoter (Hark *et al.*, 2000; Bell & Felsenfeld, 2000). This method of regulation is observed only in endoderm; mesodermal-specific enhancers have yet to be identified and alternative regulation strategies may be involved. **Figure 1.4** summarises the different modes by which DNA methylation can interfere with transcription at the *Igf2-H19* locus, and also within the *Igf2r* gene.

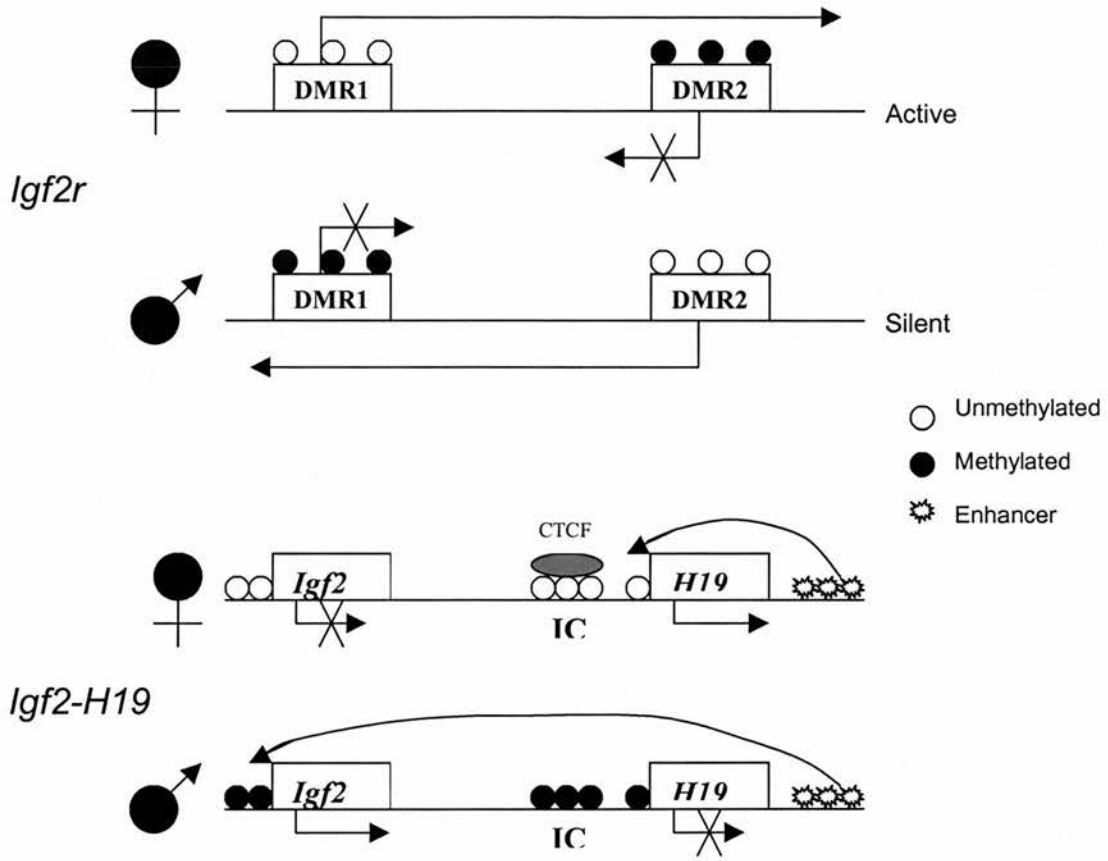


Figure 1.3 How differential DNA methylation can regulate allele-specific gene expression in imprinted genes.

***Igf2r*:** Methylation of the maternal DMR2 prevents production of an antisense transcript, and allows uninhibited transcription from the unmethylated DMR1. The unmethylated paternal DMR2 leads to the transcription of antisense RNA which proceeds into DMR1, aiding the silencing of the paternal *Igf2r* allele.

***Igf2-H19*:** When the *Igf2-H19* imprinting centre is methylated, enhancers downstream of *H19* can act to stimulate transcription of the paternal Igf2 promoter. However, when the imprinting centre is unmethylated, CTCF, an insulator protein, can bind to the site and prevent access of the enhancers to the *Igf2* promoter. The enhancers shown in this diagram are specific to endoderm; mesoderm-specific enhancers have yet to be identified.

1.4.2 X-inactivation

DNA methylation is also involved in the regulation of X-inactivation, specifically as a regulator of *Xist* (inactive X-specific transcript), a gene encoding an RNA that binds to, and coats the X chromosome, thus contributing to X-inactivation (Brown *et al.*, 1991; Brown *et al.*, 1992; Brockdorff *et al.*, 1992). In the murine trophoblast, *Xist* expression is imprinted, with differential methylation observed within the promoter region (demethylated in male germ cells, methylated in female germ cells; Ariel *et al.*, 1995; Zuccotti & Monk, 1995). *Xist* is therefore expressed from the paternal X chromosome, which ultimately leads to its silencing. However, X-inactivation is random in the fetus proper, with either the maternal or paternally-derived X-chromosome inactivated. To reinforce the idea that DNA methylation is necessary for X-inactivation, mice carrying a mutated *Dnmt1* gene demonstrated aberrant patterns of Xist expression, leading to inappropriate X-inactivation activities (Panning & Jaenisch, 1996). In addition to the regulation of imprinted *Xist* expression, DNA methylation may also be required for the regulation of CTCF binding within the *Tsix* (antisense counterpart of *Xist* responsible for imprinted X-inactivation in the placenta) locus, controlling the access of enhancers to the *Xist* promoter (Chao *et al.*, 2002).

It should be noted that in humans, the regulation of X-inactivation appears to differ from the mouse. X-inactivation is not imprinted in the human trophoblast, with random X-inactivation occuring in both the fetus and trophoblast. Thus, the methods by which *Xist* and *Tsix* expression is regulated, would be expected to differ. Indeed, human *TSIX* is known to be truncated in comparison with its murine counterpart (Migeon *et al.*, 2001). The status of imprinted X-inactivation in other mammalian species is currently unknown.

1.4.3 Tissue-specific and developmental stage-specific expression

The reversible and stable nature of DNA methylation makes it an ideal mechanism for the regulation of genes in a tissue-specific or developmental stage-specific manner. Indeed, it was proposed in the 1970s that methylation and demethylation of specific DNA sequences could regulate gene expression during mammalian development (Holliday and Pugh, 1975; Riggs, 1975). DNA methylation has been shown to be essential for normal development, as loss of maintenance DNA methyltransferase (Dnmt1) activity results in embryonic lethality at gastrulation (Li *et al.*, 1992). However, it is not yet known whether embryonic death can be attributed to a widespread disruption of methylation-associated gene-regulation during differentiation, since the observed mis-expression of imprinted genes (Li *et al.*, 1993) or active transcription of endogenous retroviruses (Walsh *et al.*, 1998) may be sufficient to prohibit normal development. In fact, some believe that DNA methylation is only required for the silencing of endogenous retroviruses, and that it silences genes only in specific circumstances, namely imprinted genes and during X-inactivation (Walsh & Bestor, 1999).

Studies investigating the relationship between gene expression and promoter methylation status have been performed in both tissues and the developing embryo, however, these have proved somewhat inconclusive. For example, the mouse skeletal α-actin promoter retains similar levels of methylation in expressing and non-expressing tissues (Warnecke & Clarke, 1999). Furthermore, genes including the oxytocin receptor (Oxtr; normally expressed in myometrium, endometrium, mammary gland and ovaries (Kubota *et al.*, 1996)) and Pax3 (expressed in limb precursors and the central nervous sytem) were not found to be highly methylated in non-expressing tissues (Walsh & Bestor, 1999). On the contrary, methylation of the human ApoA-I gene promoter is found only in non-expressing tissues, consistent with a role in gene silencing (Shemer *et al.*, 1990). The different observations may reflect the use of alternative silencing

mechanisms in tissue-specific gene regulation, although it should be noted that the techniques used to analyse promoter methylation status in the aforementioned studies only looked at limited CpG sites and may not be indicative of the entire promoter region.

It was not until recently that convincing evidence emerged, providing a link between DNA methylation and developmental regulation of tissue-specific or stage-specific genes. Studies using *Xenopus laevis* have demonstrated alterations in promoter methylation in several genes, correlating with changes in developmental expression patterns (Stancheva *et al.*, 2002). The promoters of three genes known to be activated at the midblastula transition (MBT; *TFIIIA*, *Xbra* and *c-Myc II*) were found to be significantly demethylated, whereas repetitive sequences, and promoters of genes known to be inactive at MBT, demonstrated no detectable changes in DNA methylation. Furthermore, depletion of *DNMT1* in Xenopus embryos led to observations of inappropriate promoter methylation in genes required at MBT, and transcriptional release of genes normally silenced at this stage. Although these experiments clearly demonstrate a requirement for DNA methylation in developmental gene regulation in *Xenopus*, experiments to determine whether this role is replicated in other species, particularly mammals are evidently required.

A recent experiment investigating the requirement for DNA methylation in a differentiated cell type provides additional evidence to support the hypothesis that DNA methylation is required for tissue-specific gene regulation. Using murine fibroblast cells carrying a conditional deletion of Dnmt1, Jackson-Grusby and colleagues (2001) demonstrated that loss of Dnmt1 activity resulted in the activation of many non-tissue-specific genes. They also described a 10% overall increase in gene activation, including two placental-specific genes, the B-cell surface antigen 3b, and a gene involved in adipose differentiation. These results imply a role for DNA methylation in regulating tissue-specific expression patterns, although the possibility that epigenetic disregulation occurred as a consequence of an indirect mechanism, cannot be excluded.

1.4.4 Cancer and ageing

Changes in genomic m^5C content are often observed as a result of neoplasia. These alterations are specifically manifested as a genome-wide hypomethylation (loss of methylation) event (Goelz *et al.*, 1985) and hypermethylation (increased methylation) of gene-linked CpG islands (reviewed by Jones & Laird, 1999). Hypermethylation of tumor-supressor gene promoter regions are seen in several cancers, including retinoblastoma (*RB* gene) and renal cancer (von Hippel Landau (VHL) gene), providing a substantial link between methylation and cancer (Stirzaker *et al.*, 1997; Herman *et al.*, 1994). The extent of DNA methylation involvement in silencing tumour suppressor genes has lead some to revise the 'two-hit hypothesis' to include DNA methylation as one of the processes able to inactivate these genes and induce tumorigenesis (Jones & Laird, 1999).

A gradual increase in CpG island methylation, alongside a general decrease in global methylation levels, is also seen as a result of the ageing process (Issa, 2000). For example, an age-related increase in methylation levels within the human oestrogen receptor promoter has been reported, mirroring the inappropriate methylation known to occur in colon cancer (Issa *et al.*, 1994). In addition, the methylation status of the imprinted gene *IGF2*, also undergoes age-related changes, with the differential methylated region DMR1, switching from mono-allelic to biallelic methylation (Issa *et al.*, 1996). As with the changes in the human oestrogen receptor, the increased methylation status of *IGF2* also correlates with neoplastic transformation. These results suggest that certain age-related methylation changes may increase pre-disposition to cancer.

1.5 Mammalian development

Mammalian development shares many common themes, but there are subtle differences between species, including modes of implantation and placentation, the length of gestation and the timing of key events such as compaction and gastrulation. Differences observed at specific stages of development, including preimplantation and peri-implantation development, most likely result from alternative gene expression and thus thus may represent epigenetic differences. Here, a brief outline of events occurring during post-fertilisation sheep development is described, highlighting the known differences between ruminant, primate and rodent development at specific stages of early development.

1.5.1 Pre-implantation development

The events of early (preimplantation) development in mammalian embryos are morphologically similar, although differences in timing (summarised in **Table 1.1**) and cell size are evident. In all species, the zygote undergoes several rounds of cell cleavage, leading to the formation of a blastocyst and followed by attachment of the embryo to the uterine wall. Initially, the mature oocyte is arrested at metaphase II of meiosis, characterised by the formation of the polar body, with fertilisation subsequently triggering the completion of the meiotic process. At this stage, the second polar body (containing a haploid set of chromosomes) is extruded, and DNA synthesis of sperm and egg chromosomes proceeds, forming the male and female pronuclei. When DNA synthesis is complete, the pronuclei fuse and the first cell division is underway. Subsequent cell divisions lead to the formation of 4, 8 and 16 cell embryos. At the 16-cell stage in mouse (32-64 cell stage in sheep), the embryo becomes known as a morula, characterised by cellular compaction and a fluid-filled cavity known as the blastocoel (summarised in **Figure 1.5 (i)**).

The blastocyst is distinguished by the compartmentalisation of two specialised cell-types, a process initiated at the morula stage and signalling the onset of differentiation. The inner cell mass (ICM) contains cells destined to form mainly embryonic tissues and the remaining outer cells form the trophectoderm which contributes to the extra-embryonic tissues in subsequent stages of development. After further expansion of the blastocoel cavity, the embryo 'hatches' out of the zona pellucida (a protein layer of extracellular material deposited by the growing oocyte). From this point on, considerable divergence of mammalian implantation and placentation mechanisms are apparent.

Stage	Sheep	Mouse	Human	Cow	Pig
2-cell	D1	D1	D1.5	D1	D0.6-0.8
4-cell	D2	D1-2 ZGA	D2-2.5 ZGA	D2	D1 ZGA
8-cell	D3 ZGA	D2	D2.5	D3 ZGA	D2.5
Morula	D5 (32-64 cells)	D3 (16-32 cells)	D4 (16-32 cells)	D4 (32-64 cells)?	
Blastocyst	D6-7 (>100 cells)	D3.5 (64 cells)	D6-7 (>80cells)	D7 (>100cells)	D5-6
Implantation	D15	D4	D7-8	D22	D13
End of Gestation	D145-155	D19-20	D252-274	D278-290	D112-115

Table 1.1 Timing of specific developmental stages in the sheep, mouse, human cow and pig

The day at which each species reaches the stage specified is expressed in terms of days post conception (D=dpc). The approximate cell number at the formation of the morula and blastocyst stages in each species is also highlighted. ZGA indicates the onset of zygotic genome activation in each species. The table is a modified version taken from Noden and DeLahunta, 1985).

1.5.2 Implantation and placentation

In the mouse and human, the blastocyst invades the uterine endothelium almost immediately after hatching, however, this is not the case in farm animals. The hatched sheep, pig or cow blastocyst demonstrate a marked change in size and structure, with the trophoblast elongating rapidly to reach a length of several centimetres by days 15, 16 and 17 of gestation respectively (highlighted in **Figure 1.5(ii)**). During the elongation process, the embryo remains free in the uterus, and does not implant into the uterine wall until around day 15 of gestation (an extensive overview of farm animal implantation can be found in King *et al*, 1982). In the mouse and human, implantation is followed by gastrulation and the subsequent differentiation of the ICM into the mesoderm, endoderm and ectoderm, which in turn give rise to the embryo. In contrast, farm animals gastrulate prior to implantation.

The modes of placentation observed in rodents and primates also differs significantly from placentation in farm animals. In the mouse, the outer layer of trophectoderm cells invades the uterine epithelium, basal lamina and endometrium, leading to the formation of the placenta. The placenta subsequently further invades the uterus, with direct contact occurring between the fetal trophoblast and maternal blood vessels. This form of 'invasive' placentation is known as hemochorial placentation and is also seen in humans (rodent and primate placentation is reviewed by Enders & Welsh, 1993).

When the ruminant embryo reaches the time of implantation, the outer trophoblast cells do not invade the maternal endometrium, but instead form attachments at specific sites known as caruncles (as highlighted in **Figure 1.5 (iii)**). At these points of fetal-maternal contact, specialised structures known as cotyledons then form from the chorioallantoic membranes of the developing conceptus. The cotyledons and caruncles are jointly known as placentomes and it is through these structures that most fetal nourishment is obtained. This form of placentation is known as epitheliochorial placentation, and

differs from hemochorial placentation as the placenta and uterus are separated by maternal and placental epithelial layers. Epitheliochorial placentation is also seen in the pig, although there are differences between the placentae of pigs and ruminants. An extensive review of placentation in farm animals can be found in Noden & De Lahunta, 1985.

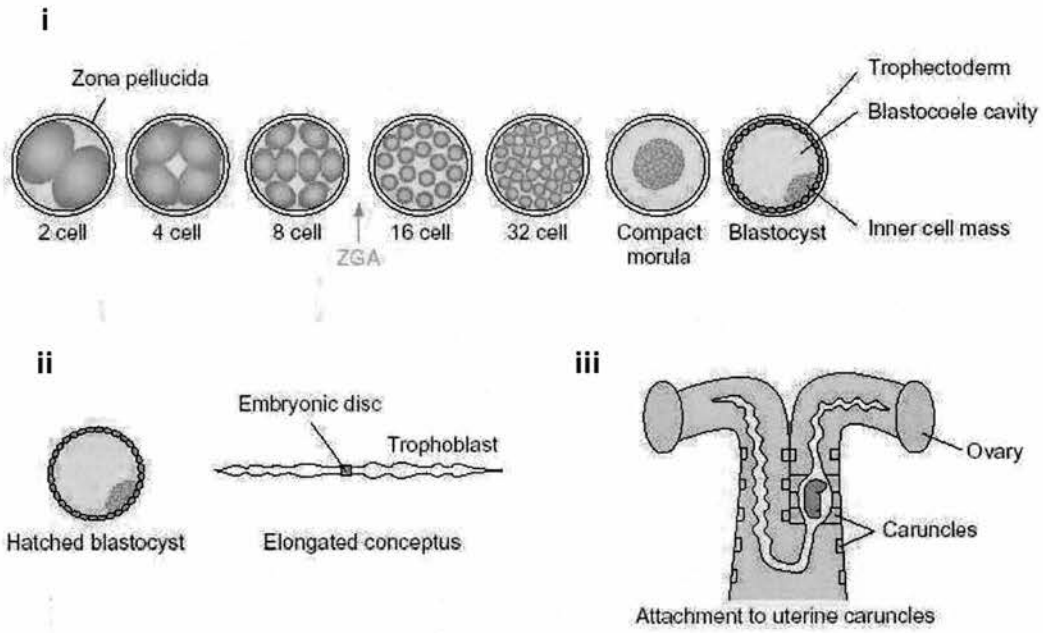


Figure 1.5 Stages of preimplantation development, implantation and placentation in the sheep and cow

Taken from Young *et al.*, 1998a. (i) After fertilisation, the subsequent cleavage rounds occur leading to the formation of the blastocyst. The timing of zygotic genome activation (ZGA) is indicated. (ii) After the blastocyst hatches, the trophoblast elongates forming a 'tapeworm-like' structure, implanting in the uterus at day 15. Gastrulation occurs during elongation, prior to implantation. (iii) Sheep and sheep placentation is non-invasive, with fetal-maternal contact localised at structures known as placentomes, consisting of the caruncle (maternal structure) and cotyledon (embryonic structure).

1.6 Reprogramming of DNA methylation patterns in normal mammalian development

The reprogramming of gametic DNA methylation patterns during preimplantation development is required for the formation of an epigenetically typical embryonic genome capable of sustaining development. The dynamics of DNA methylation patterns during murine preimplantation development have been intensely investigated, with both passive and active mechanisms believed to be involved in reprogramming. However, the methylation reprogramming observed post-fertilisation is a continuation of events initiated when the primordial germ cells (PGCs) of a day 10.5-11.5 mouse fetus enter the genital ridge. At this time, the DNA of PGCs is rapidly demethylated, leading to the erasure of all methylation imprints and re-activation of the inactive X in female germ cells (Tada *et al.*, 1998; Mann, 2001; Tam *et al.*, 1994). It is not yet known what triggers the demethylation event or whether it is an active or passive process. During later stages of gametogenesis, imprints are re-established (as discussed in section 1.4.1) alongside a general increase in DNA methylation levels, with the sperm acquiring higher levels of methylation in comparison to the oocyte (Sanford *et al.*, 1987; Razin & Shemer, 1995).

Shortly after fertilisation and prior to the first cell division, the murine paternal genome undergoes widespread demethylation (Mayer *et al.*, 2000; Oswald *et al.*, 2000), a process thought to be active in mechanism. Whilst the paternal genome is rapidly demethylated, the maternal pronucleus in the fertilised embryo retains its methylation status, and is passively demethylated during the first few embryonic cell divisions, apparently by the failure to maintain methylation patterns (Rougier *et al.*, 1998; Monk *et al.*, 1987). The differential methylation of maternal and paternal genomes is thought to be linked to the establishment of methylation imprints in the zygote (Oswald *et al.*, 2000).

One hypothesis explaining the rapid demethylation of the paternal pronucleus after fertilisation may be linked to chromatin remodelling events occurring at this stage. When the sperm first enters the oocyte, its DNA is packaged in protamines, forming a highly compact chromatin structure. These protamines are progressively exchanged for maternally-stored histones, between the point of sperm entry and the initial phase of DNA replication (Nonchev & Tsanev, 1990). Thus, the maternal and paternal pronuclei are significantly different in terms of gross chromatin structure at the onset of development. During this time, the sperm DNA is rendered 'naked' for a short period, which may be long enough for demethylating agents in the ooplasm to act upon the unprotected DNA. This hypothesis could also account for the protection of the maternal pronucleus (with previously acquired histones) from active demethylation. In addition to this, parthenogenetic embryos do not demonstrate any demethylation post-activation, strongly suggesting that the differences in paternal and maternal chromatin remodelling may be the underlying cause of the observed differential demethylation events (Barton *et al.*, 2001). The passive loss of DNA methylation observed between the 2-cell and blastocyst stage of development is likely to be a result of the retention of maintenance methyltransferase, Dnmt1, in the cytoplasm of oocytes and early stage embryos (Carlson *et al.*, 1992; Cardoso & Leonhardt, 1999).

It should be noted that there are exceptions to the widespread demethylation events, for example, some paternally derived sequences, such as the *H19* paternal allele, are protected from the initial active demethylation (Olek & Walter, 1997; Tremblay *et al.*, 1997). The retention of high levels of methylation in IAP proviral repeats (Walsh *et al.*, 1998) and centromeric satellite sequences (Rougier *et al.*, 1998) is also observed. As Dnmt1 is not thought to be active for most of preimplantation development, the activity of the *de novo* methyltransferases Dnmt3a and Dnmt3b may be responsible for the residual methylating activity occurring at this time. As these proteins have not been assessed in preimplantation embryos as yet, it is not possible to relate the retention or establishment of DNA methylation levels within specific DNA sequences to the activity of these proteins, or to as yet unidentified molecules. After the gametic methylation

patterns have been erased, they must be reprogrammed to allow successful development of the embryo. This re-establishment of methylation patterns, known as *de novo* methylation, may be performed by the putative *de novo* methyltransferases Dnmt3a and Dnmt3b, (Okano *et al.*, 1999) although the precise mechanisms of this process are poorly understood.

The level of conservation of epigenetic reprogramming in other species was not investigated until recently. Studies in two lower vertebrate species, *Xenopus* and Zebrafish, failed to show any significant changes in global methylation levels during the early stages of development (Stancheva *et al.*, 2002; Macleod *et al.*, 1999). The absence of global demethylation in these two species is intriguing, although it is possible that other forms of epigenetic reprogramming, such as chromatin remodelling, may be more important in non-mammalian species. As neither frogs nor fish have imprinted genes, and it has been proposed that post-fertilisation demethylation is required to re-set methylation imprints (Oswald *et al.*, 2000), this event may be strictly a mammalian phenomenon.

Dean *et al.* (2001) have recently shown that methylation reprogramming is conserved to a large extent in three other mammalian species; cow, rat and pig. Using indirect immunoflourescence with an antibody directed against 5-Methyl Cytosine (5-MeC), they demonstrated that active demethylation of the paternal genome occurs shortly after fertilisation in each of these species. Methylation patterns were also analysed in later stages of cow preimplantation development where passive demethylation of the maternal genome was observed, although this did not extend to the blastocyst stage as occurs in the mouse. The timing of *de novo* methylation appears to differ slightly in cow embryos, occurring at the 8-16 cell stage, as opposed to after the blastocyst stage in the mouse. Overall, the transient demethylation and early *de novo* methylation in the cow embryo highlight mammalian species-specific differences in these reprogramming events.

Interestingly, two further papers analysing sequence-specific methylation in cow preimplantation embryos both observed demethylation of euchromatic sequences (including the BOV-B LINE), whilst heterochromatic sequences (including Satellite I remained highly methylated (Bourc'his *et al.*, 2001; Kang *et al.*, 2001a). Furthermore, demethylation extended into the blastocyst stage, with no early *de novo* methylation observed. The discrepancies in observations between the three cow studies are suprising, but most likely reflect the use of different methylation analysis techniques. For example, Dean *et al.* used indirect immunofluorescent methods on whole embryos, whereas Bourc'his *et al.* studied metaphase chromosome spreads. This may have resulted in the preferential binding of the m^5C antibody to different sequences, thus accounting for the observed differences. Moreover, the bisulfite method utilised by Kang *et al.* (2001a) analysed only specific sequences.

The differences observed between species in the timing of key events such as zygotic genome activation (ZGA; the switch from maternal to zygotic control of development), compaction, implantation and gastrulation, are most likely reflected in developmental gene expression and presumably epigenetic processes. Firstly, the number of cells in morulae or blastocysts differ between the embryos of various mammals, reflecting the different timing of compaction. Secondly, the timing of implantation and embryonic morphology at this stage is also highly different. Finally, the initiation of zygotic genome activation (ZGA) occurs at different embryo stages, with significant active transcription occurring at the 2-cell stage in mouse, the 4-cell stage in the human as opposed to the 8-cell stage in the pig and the 8-16 cell stage in the sheep and cow (Telford *et al.*, 1990).

It is possible that DNA methylation is one such epigenetic mechanism involved in regulating the timing of key events in pre- and peri- implantation development. If so, it might be expected that variations in methylation reprogramming exist between mammalian species, reflecting the differences in timing. As mentioned previously, Dean and colleagues (2001) have shown that although active demethylation of paternal

pronuclear DNA is conserved between several mammalian species, the precise reprogramming events occurring beyond this stage (particularly for the cow) may differ substantially from those seen in the mouse. However, as the functional role of DNA methylation reprogramming observed during preimplantation development in processes such as ZGA, compaction and blastocyst formation remains unclear, it is not possible to say whether the observed differences in species-specific reprogramming underlie the timings of key events during preimplantation development.

The reprogramming of DNA methylation patterns is not the only significant epigenetic event occurring during preimplantation development. For example, in the mouse, the male pronucleus demonstrates significantly higher levels of transcriptional activity in the 1-cell embryo, (Ram & Schulz, 1993; Bouniol *et al.*, 1995) representing differences in chromatin structure and accessibility. Studies of histone acetylation in mouse 1-cell embryos demonstrated differences between the maternal and paternal pronuclei. Hyperacetylated histone H4 was shown to be associated with paternal but not maternal chromatin (Adenot *et al.*, 1997), explaining the differential transcriptional activity observed at this time. It is likely that further epigenetic modifications such as histone methylation, are also involved in the control of transcriptional activity and thus the regulation of timings of ZGA, compaction and implantation in mammalian development.

1.7 Somatic cell nuclear transfer

In vitro embryo technologies present major advantages in commercial agriculture, assisted reproduction in humans and towards more efficient mouse embryo production methods for scientific research. In humans, *in vitro* fertilisation (IVF) has been used successfully since the 1970s to assist couples with fertility problems (review of assisted reproduction technologies including IVF can be found in Menezo *et al.*, 2000). More specific techniques such as ICSI (Intra-Cytoplasmic Sperm Injection) have also been developed (reviewed by Mansour, 1998). The state of embryo technologies in the mouse has greatly enhanced the production of mice for research purposes with methods

such as superovulation (hormonal stimulation of oocyte release) allowing optimal production of embryos.

In farm animals, artificial insemination and *in vitro* fertilisation potentially allows for the improved selection and large-scale production of animals with desired genetic traits. In addition, somatic cell nuclear transfer (SCNT), a more recently established embryo technology, is a potentially valuable commodity for the production of genetically modified animals for use as bioreactors, or as organ donors for xenotransplantation. Recent advances have demonstrated the feasibility of these processes, with the production of sheep expressing the human factor IX protein in milk (Schneike *et al.*, 1997) and cloned pigs and sheep carrying a targeted disruption of the alpha1,3-galactosyltransferase gene, encoding a major xenoantigen responsible for tissue rejection when performing xenotransplantation (Dai *et al.*, 2002; Denning *et al.*, 2001).

Further to the aforementioned agricultural advantages, SCNT may also contribute to the success of human stem cell therapies. It is envisaged that human adult cells could be cloned, with ES cells produced from the resultant embryos and differentiated *in vitro* to produce virtually any adult cell type. The resultant cells would be genetically identical to the patient's and would thus remove any possibility of tissue rejection. Diseases such as Parkinson's could be cured by the cloning of a 'healthy' cell type (e.g. skin fibroblast), to create an embryo, and subsequent differentiation of embryonic ES cells into dopaminergic cells. However, such research is in its infancy and currently limited by the low success rates of cloning and inherent problems as described later in this section. Furthermore, the ethical considerations of cloning research in humans, particularly the generation of embryos solely for subsequent ES cell derivation, have been widely debated and hence research is restricted by licensing. Consequently, the use of surplus IVF and ICSI human embryos may be a more practical source of ES cells for basic research at this time.

Currently, the widespread application of livestock embryo SCNT-related technologies is somewhat limited by the very low success rates, reflecting the high incidence of fetal abnormalities and mortality (reviewed by Young & Fairburn, 2000). These may be attributed to incomplete epigenetic reprogramming of somatic nuclei occurring after the SCNT process. In addition to SCNT, *in vitro* embryo production techniques also suffer inherent problems including low success rates and increased fetal size (often associated with fetal mortality) when comparing *in vitro* produced embryos with their *in vivo*-derived counterparts (Young *et al.*, 2001).

Somatic cell nuclear transfer involves the transfer of genetic material from a differentiated nucleus into an enucleated oocyte. The cell is either fused with, or microinjected into the oocyte to initiate development. The first vertebrate nuclear transfer, or 'transplantation' experiments were performed by Briggs & King (1952), in order to determine whether differentiated cells could be reprogrammed to induce development. Using the frog as a model (*Rana pipiens*), blastula nuclei were transplanted into enucleated oocytes, and tadpoles obtained. Further experiments demonstrated that the more advanced or differentiated the nucleus (using cells from a post-gastrulation embryo), the lower the success rates of these transplantation experiments (Briggs & King, 1957), prompting many to believe that reprogramming highly differentiated nuclei was impossible. Almost a decade after the original experiments were performed, it was discovered that differentiated intestinal cells derived from *Xenopus laevis* larvae were indeed able to de-differentiate when transferred to an enucleated oocyte, and sustain normal development to the tadpole stage (Gurdon, 1962).

After these initial successes, attention turned to mammalian nuclear transfer, with micro-injection techniques developed during the mid to late 1970s (Bromhall, 1975). Thus followed almost twenty years of experimental optimisation of the nuclear transfer technique, with little or no success. The persistent failures, combined with the discovery that both the maternal and paternal genomes were required for normal development, led some to believe mammalian cloning was impossible (McGrath & Solter, 1984; Barton *et*

al., 1984). However, this was not true, and success first came in the production of sheep using a cultured embryonic cell line that had undergone several passages (Campbell *et al.*, 1996). The method utilised fused enucleated metaphase II (MII) oocytes with serum-starved donor cells arrested in the G_0 phase of the cell cycle, with subsequent activation by electrical stimulation (outlined in **Figure 1.6**). Nonetheless, it wasn't until the advent of Dolly, that the production of a clone from a fully differentiated cell (adult mammary gland epithelial cell) was shown to be possible (Wilmut *et al.*, 1997). Soon after, the cloning of mice from cumulus cell nuclei was reported, using a microinjection technique for nuclear transfer rather than cell fusion (Wakayama *et al.*, 1998). Many successes in other mammalian species have followed in recent years, although success rates are consistently low. These are summarised in **Table 1.2**.

Species	Donor cell type	Reference	% morula/ blastocysts	No. of live offspring (% of embryos transferred)	No. of surviving offspring	Cloning efficiency
Mouse	Cumulus	Wakayama *et al.*, 1998	-	31 (2.2%)	22	1.3%
	Fibroblast	Ogura *et al.*,2000	21.8%	5 (1.7%)		
Cow	Cumulus	Kato *et al.*, 1998	39.4%	5 (83.3%)	2	5%
	Oviduct	Kato *et al.*, 1998	39.3%	3 (75%)	2	2%
	Uterine	Kato *et al.*, 2000	-	2 (14.2%)	0	2%
	Granulosa	Wells *et al.*, 1999	69.4%	10 (10%)	10	1.8%
	Mammary Gland	Zakhartchenko *et al.*, 1999	16.1%	1 (25%)	1	0.4%
	Muscle	Shiga *et al.*, 1999	14.2%	4 (15.3%)	1	0.8%
	Fibroblast	Zakhartchenko *et al.*, 1999	53.3%	1 (6.25%)	0	1.1%
Sheep	Mammary Gland	Wilmut *et al.*, 1997	10.5%	1 (3.4%)	1	0.4%
Pig	Granulosa	Polejaeva *et al.*, 2000	-	5 (0.8%)	5	0.1%
Goat	Granulosa	Keefer *et al.*, 2000	-	7 (7.3%)	6	?
	Cumulus	Zou *et al.*, 2001	-	3 (1.3%)	1	0.7%
Cat	Cumulus	Shin *et al.*, 2002	-	87 -	1	?
Rabbit	Cumulus	Chesne *et al.*, 2002	-	371 (60.6%)	4	?

Table 1.2 Somatic cell nuclear transfer (SCNT) efficiency in mammals

The table is adapted from the mammalian SCNT efficiency table compiled by Lesley Paterson, R.I; the complete version can be found at http://www.roslin.ac.uk/public/webtablesGR.pdf. The % of morula/blastocysts represents the number of SCNT embryos reaching this specific stage. The Number of Live Offspring (% embryos transferred) refers to the number of live offspring born and is also given as a % of the total number of embryos transferred (normally morula or blastocyts, however some studies transfer 2,4,and 8 cell). No. of surviving offspring refers to the number of offspring still alive at the time the paper went to press, and thus varies between different papers. The cloning efficiency is the number of live offspring expressed as a % of the total number of SCNT embryos.

It is believed that the low success rates of SCNT can be mostly attributed to the failure of the oocyte to completely reprogramme somatic nuclei in order to sustain normal embryonic development. In addition to reprogramming errors (discussed in section **1.7**), there are several elements of the SCNT procedure itself, which may significantly reduce the potential for successful development and ultimately lead to failure of the SCNT process (these are highlighted in **Figure 1.6**). The processes which may introduce additional errors are also common to other *in vitro* embryo production technologies used routinely in sheep and cattle. For example, the source of oocytes used as the recipient could affect the success of SCNT depending on their existing developmental capacity prior to enucleation. As SCNT requires large numbers of recipient oocytes (>100) to gain only a few live offspring, the harvesting of naturally ovulated oocytes is not a practical option in large animals. Instead, either superovulation or *in vitro* maturation of oocytes is commonly used.

In vitro maturation (IVM) of oocytes consists of the collection of oocytes from abattoir slaughtered sheep or sheep ovaries and subsequent culture *in vitro*. Typically, only follicles of 2-5 mm are aspirated (purportedly containing the most developmentally competent oocytes) and the oocytes are subsequently matured for 24 hours in TCM199 medium supplemented with serum, FSH (follicle stimulating hormome), LH (luteinising hormone) and oestradiol (Walker *et al.*, 1996). This encourages oocyte development to the metaphase II stage of meiosis, the stage at which mature oocytes are normally released from the ovary. Superovulation involves the hormonal stimulation of ovaries with GnRH (gonadotrophin-releasing hormone), FSH and LH to release larger than normal numbers of meiotically matured oocytes over a short period of time (reviewed by Davis & Rosenwaks, 2001).

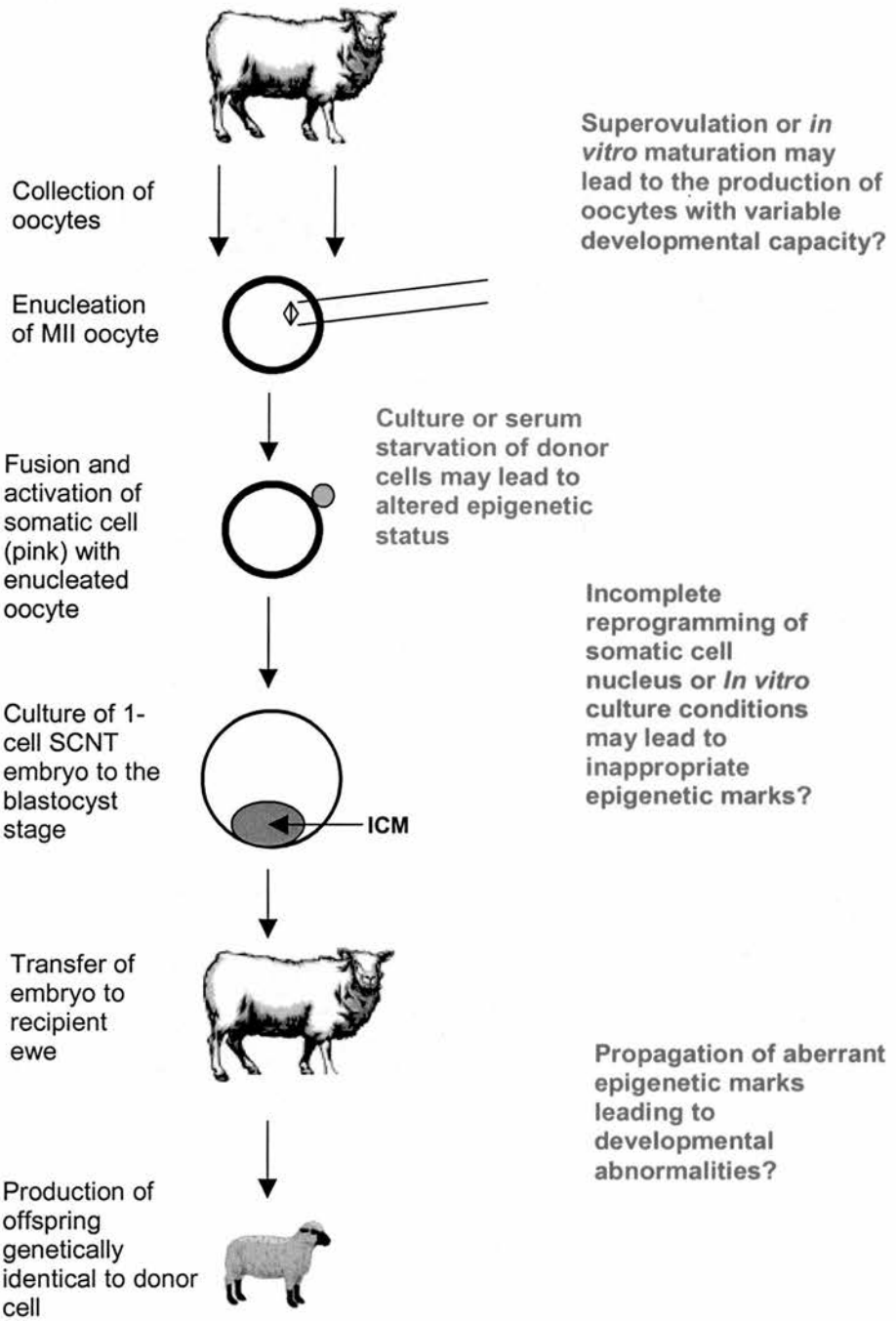


Figure 1.6 Somatic cell nuclear transfer (SCNT) in sheep

The stages of the nuclear transfer process from oocyte collection to the production of live offspring are outlined. Possible reasons for the low success rates of SCNT, including inefficient reprogramming and effects of oocyte source or *in vitro* culture, are highlighted in red.

In vivo, oocyte competence is gained gradually, as the oocyte develops within the follicle. This involves the acquisition of mRNA transcripts and proteins required for control of the initial stages of development. In addition, modifications are made to both DNA and RNA during the final stages of oocyte maturation. For example, mRNA transcripts found in the growing oocyte have shorter than normal polyA tails, inhibiting translation and are only adenylated when translation is required (reviewed by Gandolfi & Gandolfi, 2000). Furthermore, DNA methylation imprints are established at different stages of oocyte growth (Obato & Kono, 2002). A key methylation modifier, Dnmt1o, is also known to be present abundantly in mature mouse oocytes. This is a truncated, oocyte-specific isoform of Dnmt1 predominantly localised in the cytoplasm of mature oocytes and preimplantation embryos (Mertineit *et al.*,1998; Carlsson *et al.*, 1992). However, in earlier stages of oocyte growth, Dnmt1 is found in the nucleus and during later stages is dispersed into the periphery of the ooplasm until localised in a cortical shell (Mertineit *et al.*, 1998). Some of the oocytes acquired by superovulation or *in vitro* maturation could be relatively immature in terms of cytoplasmic content, or even possibly derived from mildly atretic follicles. This could in theory, lead to a heterogeneous pool of oocytes being used as recipients in SCNT experiments, with varying extents of incomplete methylation imprints, untranslatable mRNA transcripts and inappropriate localisation of Dnmt1o.

A second potential source of problems could arise from *in vitro* culture of SCNT embryos. The same culture system used for *in vitro* fertilised embryos is used to culture fused SCNT embryos to the blastocyst stage. Serum free media is favoured (e.g. synthetic oviduct fluid containing amino acids and sheep serum albumin (SOFaaBSA)), as exposure to sera has been shown to contribute to the fetal growth and developmental abnormalities collectively termed as 'Large Offspring Syndrome' (LOS; Young *et al.*, 1998a). This phenotype has been reported in cattle and sheep produced by IVM, IVF and *in vitro* culture, as well as somatic cell nuclear transfer (SCNT; reviewed by Young & Fairburn, 2000). The precise mechanisms or specific factors involved in the induction of this phenotype remain undetermined. However, aberrant methylation of an imprinted

gene, the insulin-like growth factor II receptor (*IGF2R*), has been reported in sheep fetuses with fetal overgrowth (Young *et al.*, 2001). It is possible that inappropriate DNA methylation of other imprinted genes, or other DNA sequences, may also contribute to this phenotype in *in vitro* produced or SCNT embryos. Although the use of serum-free media generally leads to the production of apparently healthy offspring (De Sousa *et al.*, 2001), it cannot be ruled out that this method of embryo culture has no effect on epigenetic status whatsoever.

It should also be noted that the cells used to provide donor nuclei could potentially be epigenetically altered by culture or serum-starvation prior to SCNT. Epigenetic defects in imprinted genes (*Igf2, H19, Igf2r* and *U2af1-rs1*) have been observed as a result of ES-cell culture, leading to the mis-expression of developmentally important genes in embryos derived from these cells (Dean *et al.*, 1998). The effects of serum-starvation on sequence-specific epigenetic status have been analysed in fetal fibroblasts, with no change in methylation levels observed within Satellite I DNA pre- and post- serum starvation (Kang *et al.*, 2001a). However, it possible that some genes or sequences may be more susceptible to epigenetic modification in culture, and thus other epigenetic changes may have remained undetected.

1.8 Epigenetic reprogramming in cloned embryos

The successful production of animals by nuclear transfer has demonstrated that highly methylated and differentiated somatic nuclei can be sufficiently reprogrammed to sustain embryonic development (reviewed by Solter, 2000). However, the extent to which epigenetic reprogramming is conserved in cloned embryo development is only beginning to be examined (reviewed by Fairburn *et al.*, 2002) and inefficient or partial epigenetic reprogramming may be responsible for the low birth rates and developmental abnormalities often seen as a result of the nuclear transfer process (reviewed by Young & Fairburn, 2000). One aspect of interest is the ability of the enucleated oocyte to

recognise and reprogramme methylation patterns of somatic nuclei or other epigenetic modifications

There are a number of epigenetic processes that may be affected by the SCNT process. In the 'normal' situation, maternal factors stored in the oocyte cytoplasm modify paternal and maternal DNA differentially, whereas in cloned embryos, the same factors have to modify a single, highly differentiated somatic nucleus. Most is known about DNA methylation changes occurring post-fertilisation in mouse development, including the differential demethylation of paternal and maternal genomes (Mayer *et al.*, 2000; Oswald *et al.*, 2000) and the maintenance and establishment of methylation imprints (El-Maari *et al.*, 2001; Olek & Walter, 1997; Tremblay *et al.*, 1997). The dynamic changes occurring at this time signify DNA methylation as a candidate process to analyse in SCNT embryos.

Another significant epigenetic change occurring after fertilisation is the protamine-histone exchange (as discussed **section 1.6**). Also mentioned previously (**section 1.6**) was the possibility that active demethylation of the paternal pronucleus observed after fertilisation may be aided by the temporary cytoplasmic exposure of unpackaged paternal DNA during histone-protamine exchange. In fused SCNT 1-cell embryos, the DNA of somatic nuclei is already packaged into somatic histones, and these may not be exchanged to the same extent as protamines (Bordignon *et al.*, 2001). If only a fraction of the DNA is exposed to demethylating agents at this stage, somatic DNA methylation patterns may be maintained, with detrimental effects to later stages of embryonic development.

Further epigenetic modifications such as histone acetylation and histone methylation may also be difficult to reprogramme in SCNT embryos. Interference from the large quantity of proteins associated with a transcriptionally active somatic nucleus may also contribute to reprogramming problems in cloned embryos. However, little is known of general histone modifications in the preimplantation embryo, thus the reprogramming of

DNA methylation remains the most easily examined modifications in cloned embryos at this time.

To address this specific issue, several investigations have been reported recently, analysing the conservation of methylation reprogramming events during cloned embryo development. Dean *et al.* (2001) analysed demethylation events in cloned sheep embryos. Interestingly, the initial demethylation event appeared to be conserved in cloned 1-cell embryos, with loss of methylation in the somatic donor cell nucleus observed shortly after fusion with the enucleated oocyte. Further demethylation was observed at the 2-cell stage at which point demethylation ceased, an event followed by spurious *de novo* methylation in many 4-8 cell stage embryos.

The observations of Dean *et al.* (2001) demonstrated that methylation reprogramming in somatic nuclei is partially conserved in cloned embryos, with different extents of inappropriate methylation seen in individual embryos. However, more detailed information on which sequences are aberrantly methylated in cloned embryos is required to further understand the observed phenotypes and low success rates of cloning. A second investigation using the m^5C antibody on metaphasic chromosome spreads addressed this question, analysing euchromatin and heterochromatin specific regions in normal and cloned cow blastocysts (Bourc'his *et al.,* 2001). When compared to their normal counterparts, cloned cow blastocysts demonstrated increased levels of methylation in centromeric heterochromatin, but comparable levels in euchromatin. In agreement with Dean *et al.* (2001), this study also demonstrated inefficient passive demethylation during the initial embryonic cell divisions, although contrarily, they failed to observe the 'active' demethylation of somatic nuclei after fusion.

Methylation analysis of specific DNA sequences has also been performed using bisulfite treatment and subsequent restriction analysis in cloned cow and porcine embryos (Kang *et al.*, 2001a; Kang *et al.*, 2001b). Analysis of the Satellite I DNA sequence (a major constituent of centromeric heterochromatin) in normal and cloned cow blastocysts

demonstrated increased methylation levels in significant numbers of cloned embryos. Analysis of another repeat sequence in cow cloned embryos, the Bov-B LINE (which normally undergoes demethylation during cow preimplantation development) again demonstrated inefficient reprogramming, with the methylation status remaining unchanged between the 4-8 cell and blastocyst stages. Surprisingly, cloned porcine embryos do not appear to show highly aberrant patterns of methylation, with most cloned blastocysts retaining similar methylation levels to their normal counterparts in the centromeric satellite and the PRE-1 SINE DNA sequences.

Overall, these findings present a picture of partial, but incomplete methylation reprogramming in many cloned embryos. Possible reasons for the lack of complete reprogramming include the inability of the oocyte to efficiently reprogramme highly methylated and differentiated somatic nuclei. However, as the somatic donor nuclei of cloned cow embryos appear to be demethylated shortly after activation (Dean *et al.*, 2001), the oocyte is able to at least partially reprogramme methylation patterns of the somatic nuclei. A second explanation resides within the somatic nucleus, as inappropriate expression and nuclear localisation of somatic Dnmt1 in the embryonic nucleus may be introduced from within the donor cell nucleus. As the normal mouse zygote retains Dnmt1 in the cytoplasm, the introduction of somatic Dnmt1 could result in maintenance of methylation levels either significantly, or partially, thus contributing to the aberrant epigenotype observed in cloned embryos. Alongside Dnmt1, a whole host of other DNA methylation associated proteins (including *de novo* methyltransferases, DNMT-interactors or methyl-binding proteins) perhaps not normally found in the early embryo, could also lead to aberrant methylating activity. Finally, altered methylation levels in cloned embryos could also arise from the early expression of DNA methyltransferases (or indeed other methylation-associated proteins) caused by spurious transcriptional activation of the donor cell nucleus.

1.8.1 Consequences of inefficient methylation reprogramming

Certain fundamental questions pertaining to the specific roles of DNA methylation in development must be addressed before the observations of aberrant methylation reprogramming can be interpreted confidently. For example, are methylation errors in heterochromatic sequences, such as satellite DNA, detrimental to normal development? Undermethylation of minor satellite DNA is observed in patients suffering with ICF syndrome (conferred by mutations in the human *DNMT3B* gene), contributing to a range of developmental abnormalities. And to what extent is aberrant reprogramming tolerated in embryos? It is well known that methylation errors in imprinted genes can be directly linked to the mis-expression of these developmentally important genes, thus perturbing normal development (reviewed by Reik & Walter, 2001). Erroneous expression of this particular subset of genes has been widely postulated as a potential contributory factor to the overgrowth phenotypes observed in cloned animals. It has been suggested that cloned embryos may tolerate a certain extent of inappropriate reprogramming (Rideout *et al.*, 2001) as cloned mice with aberrant expression of imprinted genes appear developmentally normal (Humpherys *et al.*, 2001). It is clear that further studies investigating the long-term effects of aberrant methylation of other developmentally regulated genes and repetitive DNA sequences acquired during pre- and postimplantation development are required to elucidate the functional significance.

1.9 Thesis aims and objectives

The purpose of this study was to investigate DNA methylation in sheep, in particular, its dynamics during the early stages of embryo development. As DNA methylation has not previously been investigated extensively in sheep, isolation and characterisation of components of the DNA methylation repressor system (DNA methyltransferases and methyl-CpG-binding proteins) was initially required. This would allow analysis of sequence homology of these components between several mammalian species, and also to understand the requirement for these proteins in adult sheep tissues by assessing expression levels. Preliminary characterisation of methylation dynamics in normal sheep preimplantation development and events in cloned (SCNT) embryos were then able to be performed for the first time

Specific aims

1. **To clone sheep homologues of genes encoding DNA methyltransferases and the methyl-CpG-binding protein, MBD2 and to evaluate their conservation with mouse and human homologues.**
2. **To investigate expression levels of these genes in adult sheep tissues.**
3. **To determine whether an oocyte-specific form of *DNMT1* exists in sheep oocytes.**
4. **To determine expression patterns of *DNMT1*, *DNMT3A* and *DNMT3B* genes in sheep and mouse oocytes and preimplantation embryos.**
5. **To optimise techniques for DNA methylation analysis of specific sequences in single oocytes and embryos, and analyse changes in methylation levels between different stages of sheep preimplantation development.**
6. **To determine whether DNA methylation of specific sequences is equivalent in *in vivo* produced, *in vitro* fertilised and SCNT embryos.**

Chapter 2

Materials and Methods

2.1 Materials and equipment

Abgene. *London, UK:* Thermo-Start® PCR mastermix, 0.5ml thin-walled PCR tubes.

Ambion (Europe) Ltd. *Huntingdon, Cambridgeshire, UK:* QuantumRNA Classic II 18S standards

Amersham Pharmacia Biotech UK Ltd. *Little Chalfont, Buckinghamshire, UK:* SeeDNA™, Multiprime DNA labelling kit, Hybond-C Nitrocellulose filters, Hybond N+ membrane, [α-^{32}P]dCTP radiolabelled nucleotide, film, Rapid-hyb buffer, Hyperfilm™ MP, First-Strand cDNA Synthesis Kit, 20% SDS (USB).

Applied Biosystems. *Warrington, Cheshire, UK:* ABI PRISM® 7700 Sequence Detection System, Taqman probe synthesis, ABI PRISM™ Optical 96-Well Reaction Plates, MicroAmp® Caps, TaqMan® Universal PCR Master Mix, TaqMan® Ribosomal RNA Control Reagents.

BDH. *Poole, Dorset, UK:* Molecular Biology Grade Agarose, sodium chloride, magnesium sulphate, gelatine, EDTA, sodium phosphate, tri-sodium citrate, Tris base, boric acid, sodium hydroxide.

Beckman Coulter Inc. *Fullerton, CA, USA.* Beckman J2-21M/E refrigerated centrifuge.

BD Biosciences Clontech UK. *Basingstoke, Hampshire, UK:* SMART™ RACE cDNA amplification kit.

BIO-RAD. *Hercules, California, USA*: Multi-Analyst software.

DIFCO. *West Molesey, Surrey, UK:* Yeast extract, Bactotryptone, Casamino acids.

Greiner Labortechnik. *Frickenhausen, Germany*: 15ml centrifuge tubes, 50ml centrifuge tubes.

Hettich-Zentrifugen. *Tuttlingen, Germany:* MIKRO 22R refrigerated benchtop centrifuge, MIKRO 20 benchtop centrifuge.

Hybaid Ltd. *Teddington, Middlesex, UK:* Touchdown thermal cycler, Multiblock System (MBS), RiboLyser™ Cell Disrupter, RiboLyser™ GREEN Kit.

Invitrogen. *Groningen, The Netherlands:* 10mM dNTP's, TOPO-TA Cloning® kit, Agar.

Kendro Laboratory products (Sorvall). *Connecticut, USA:* Sorvall RC-3B refrigerated centrifuge.

Konica Medical Imaging Inc *.New Jersey, USA:* SRX 101A Film processor.

Life Technologies Ltd. *Paisley, UK*: 1 Kb Plus DNA ladder.

MWG Biotech. *Milton Keynes, UK*: Sequencing service, oligonucleotide synthesis.

New England Biolabs (UK) Ltd. *Hitchin, Hertfordshire, UK*: *Hha* I restriction enzyme, *Sss* I (CpG) methylase.

Novagen Inc. *Madison, Wisconsin, USA:* Pellet Paint™ Co-precipitant.

QIAGEN Ltd. *Crawley, West Sussex, UK:* QIAquick Gel Extraction Kit, QIAprep Plasmid Miniprep kit, QIAprep Lambda Midiprep kit, RNeasy Miniprep kit, QIAshredder columns.

Roche Diagnostics Ltd. *Lewes, East Sussex, UK:* *Eco* RI, *Xho* I, *Bam* HI and *Hind* III restriction enzymes, *Taq* Polymerase, Expand HiFi *Taq* polymerase.

Sigma-Aldrich Company Ltd. *Poole, Dorset, UK:* Gel loading solution, Proteinase K, Denhardt's solution, 10X MOPS buffer, Ethidium Bromide, Maltose, Formamide, Ampicillin, casein hydrolysate.

Stratagene. *La Jolla, CA, USA:* Adult Sheep lung cDNA (Unizap) library, SOLR competent cells, XL1-blue MRA (P2) competent cells, UV Stratalinker™ 1800.

UVP Inc. *Upland, California, USA:* High performance transilluminator.

2.2 Buffers and Media

All buffers and solutions described below were made up by H.F apart from 10X TBE, 20X SSC, LB medium and LB agar, which were prepared by M.Hutchison and S.Ewart (Roslin Institute (R.I.)).

10X TBE:	0.9 M Tris-borate (BDH) 0.02 M EDTA (BDH) pH 8.0
20X SSC buffer:	3 M NaCl (BDH) 0.3 M Sodium Citrate (BDH) pH 7.0
Hybridisation buffer:	50% Formamide (Sigma) 5 X SSC (R.I.) 0.01 M EDTA (BDH) 5x Denhardt's solution (Sigma) 0.05 M $NaPO_4$ (BDH; pH 7.0) 1% SDS (UBS)
DNA extraction buffer A:	75 mM NaCl (BDH) 25 mM EDTA pH 8.0
DNA extraction buffer B:	10 mM Tris-HCl (R.I.) 10 mM EDTA 1% SDS pH 8.0

LB Medium (1 litre):	Bactotryptone (DIFCO)	10.0 g
	Yeast extract (DIFCO)	5.0 g
	NaCl	10.0 g
	pH7.0	
LB Agar (1 litre):	Bactotryptone	10.0 g
	Yeast Extract	5.0 g
	NaCl	10.0 g
	Agar (Invitrogen)	15.0 g
NZY Medium (1 litre):	Casein hydrolysate (Sigma)	10.0 g
	NaCl	5.0 g
	Yeast extract	5.0 g
	$MgSO_4.7H_2O$ (BDH)	2.0 g
	pH7.0	
NZY agar (1 litre):	Casein hydrolysate	10.0 g
	NaCl	5.0 g
	Yeast extract	5.0 g
	$MgSO_4.7H_2O$	2.0 g
	Agar	15.0g
NZY top agarose (1 litre):	Casein hydrolysate	10.0 g
	NaCl	5.0 g
	Yeast extract	5.0 g
	$MgSO_4.7H_2O$	2.0 g
	Agarose (BDH)	7.0 g

SM buffer (1 litre):	NaCl	5.8 g
	$MgSO_4.7H_2O$	2.0 g
	1 M Tris-HCl (pH 7.5)	50 ml
	2% gelatine	5 ml

Buffer X2	20 mM Tris-HCl (pH 8.0)
	20 mM EDTA
	200 mM NaCl
	80 mM DTT
	4% SDS
	250 µg/ml Proteinase K

2.3 Biological materials

2.3.1 *In vitro* fertilised/cultured (IVF) embryos

All stages of *in vitro* oocyte maturation, fertilisation and embryo culture were performed by Mr. J. Gardner and Dr. J. McCracken (Roslin Institute (R.I)). A detailed protocol is described in Walker *et al.*, 1996. Immature oocytes were obtained from abattoir-derived ovaries, and the semen pellets were provided by a Texel ram (a kind gift from Edinburgh Genetics). At various stages between 1-cell and blastocyst embryos were selected and subsequently stored at -80°C in 1 µl culture medium.

2.3.2 *In vivo* derived embryo collection

In vivo produced embryos were collected from superovulated and naturally-mated Scottish Blackface ewes. Embryo collections were performed by members of the Large Animal Unit (R.I) under strict accordance with Home office regulations. Embryos were selected at various stages between 1-cell and blastocyst stages and stored at -80°C in 1 µl culture medium. Mouse oocytes and preimplantation embryos were collected from superovulated and naturally-mated mice (B6D2F1) by Mrs M. McGarry (R.I.).

2.3.3 Somatic cell nuclear transfer (SCNT)

Nuclear transfer experiments were performed by Dr. J. McCracken and Mr. W. Ritchie (R.I). A detailed protocol is described in De Sousa *et al.*, 2001. Recipient oocytes were collected from superovulated Scottish Blackface ewes. Donor cells were from a fetal fibroblast cell line BlWF1P5, prepared by Mrs. P. Ferrier (R.I.), and serum starved for 5 days prior to nuclear transfer. Embryos were selected at various stages between 1-cell and blastocyst stages and stored at -80°C in 1 µl culture medium.

2.3.4 Tissue samples

Tissue samples for RNA extraction were collected from a Poll Dorset ewe by members of the Large Animal Unit (R.I).

2.4 General Methods

2.4.1 Polymerase chain reaction (PCR) methods

2.4.1.1 Standard PCR

PCR reactions were performed in 25 µl volumes containing 1x PCR reaction buffer (Roche Diagnostics; final magnesium [Mg^{2+}] concentration 1.5-2.5 mM depending on primers used), 1mM dNTP's (Invitrogen), 1U *Taq* polymerase (Roche Diagnostics), 250 nM each of forward and reverse primer, dH_20 and 1-5 µl of template. PCR reactions were performed on either the Hybaid Touchdown thermal cycler, or the Hybaid Multi-block system. Cycling conditions were dependent on primer pairs and template concentration. Specific annealing temperatures, magnesium conditions and PCR programs used for the various primer sets used can be found in **Table 2.1**. PCR program conditions can be found in **Table 2.2.** Sequences of all primers used in this study can be found in **Table 2.3**

2.4.1.2 Hotstart PCR

The Thermo-Start® PCR mastermix (ABgene) was utilised when hotstart PCR was required to increase specificity. PCR reactions were performed in 25 µl reaction volumes containing 12.5 µl 2x PCR mastermix (final [Mg^{2+}] concentration 1.5 mM; 0.8 mM dNTPs, 1U Thermo-Start® *Taq* polymerase), 250 nM each of forward and reverse primers, dH_20 and 1-5 µl template DNA. Cycling conditions were the same as standard PCR (Tables 1 & 2), except for the initial 94°C denaturing cycle, which was held for 15 minutes to allow for the incremental activation of the Thermo-Start® Taq polymerase. Specific cycling conditions for the various primer sets can be found in **Table 2.1** and **Table 2.2.**

Primer Pair	***Annealing Temperature (°C)***	***Optimal Mg^{2+} concentration (mM)***	***PCR program***
MTRA1F/R	57	2.5	S1
HDNMT3AF/R	60	1.5	S1
DNMT3B2F/R	55	1.5	S1
DMTF/R	56	1.5	S1
MBDF/R	58	2.5	S1
*18SF/R	60	1.5	S2
RX15C/13A	60	2.5	S1
BOVBF/R	58	1.5	TH1
OVSAT1AF/R	62	1.5	TH1
SHE2/SOMDT2	59	2.5	TH2
SOMDT1/SOMDT2	59	2.5	TH2
MOE1/MSER	59	2.5	TH2
MSE1/MSER	59	2.5	TH2

Table 2.1 Optimal PCR conditions for specific primer sets.

The table shows the optimal Magnesium concentrations and annealing temperatures for all standard primers used in this study. PCR programs are described in Table 2.2.

* QuantumRNA Classic II 18S primers (Ambion).

PCR Program	*No. of cycles*	Denaturing		*Annealing*		*Extension*	
		℃	*min:sec*	*℃*	*min:sec*	*℃*	*min:sec*
S1	1	94	5:00	-		-	
	40	94	0:30	55-62	0:30	72	0:45
	1	-		-		72	5:00
S2	1	94	5:00	-		-	
	30	94	0:30	55-62	0:30	72	0:45
	1	-		-		72	5:00
TH1	1	94	15:00	-		-	
	*40	94	0:30	55-62	0:30	72	0:45
	1	-		-		72	5:00
TH2	1	94	15:00	-		-	
	40	94	0:30	55-62	0:30	72	0:45
	1	-		-		72	7:00

Table 2.2 PCR programs

Specific cycling conditions for PCR programs described in this study. Annealing temperatures are specified in Table 2.1. 'S' programs are standard PCR programs utilising Roche Taq Polymerase, 'TH' programs are Hot start PCR programs, using Thermo-Start® PCR mastermix.

* Cycle numbers for quantitative MS-PCR using program TH1 are described in **2.4.12**

Primer	*Forward (5'-3')*
BOVBF	GAAATTAAAAGATGCTGGCTCC
BOVBR	AGTTCAGTCGCTCAGTCGTG
OVSAT1AF	ATTCCTCTCCCGCTGATGCC
OVSAT1AR	GTGCCACCATTCCAAGAGTCC
DNMT3B2F	ACCTGCTGAATTACTCACGCC
DNMT3BR	ATGTCCCTCTTGTCGCCAAC
MTRA1F	CCTTCCTCAGCTACTGCGAC
MTRA1R	ACTGATAGCCCATGCGGAC
MBDF	CGGATGGAAGAAGGAGGAAG
MBDR	ACCTTGCCAACTGAGGCTTG
HDNMT3AF	GGGGACGTCCGCAGCGTCACAC
HDNMT3AR	CAGGGTTGGACTCGAGAAATCGC
RX15C	GCTGCAGCTGCAGCTGCTGCT
RX13A	CTGCACAGCTCTGGGAACTCG
RACE X	GGTGTCCAAGACCGATCCTCCGAAGT
RACE Y	GCCAGATTCGTTAGCGTCGAAGATGGA
RACE Z	GCGTCGAAGATGGACAACCTCTCGTT
SOMDT1	CCTTCTCACTGCCTGACGAT
SOMDT2	CCAGGTAGCCCTCCTCAGAT
SHE2	GCTCAAAGATTTGGAAAGAGATAG
MOE1	CCACTCTTTGCTGCTTGAGA
MSE1	AACAGCTCCAGCCCGAGT
MSER	CCAGGTAGCCTTCCTCAGAT

Table 2.3 Primer sequences

All primers used in this study were designed using the 'Prime' application of GCG10, or the web-based Primer 3 program, and synthesised by MWG Biotech (UK)

2.4.2 Gel electrophoresis protocol

A 20 µl aliquot of PCR reactions/restriction enzyme digests was mixed with 3 µl loading dye (Sigma) and electrophoresed in a 1.5-2% agarose (BDH) gels using 1X TBE containing 0.2 µg/ml ethidium bromide (Sigma) at 8-9 Vcm^{-1}. PCR products were visualized on a UV transilluminator and the image recorded using a video camera system linked to the Multi-analyst software program (BIO-RAD) and gel images were edited in Microsoft Photo Editor Version 3.01. If necessary, band intensities were increased/decreased using the brightness or contrast tools of Photo Editor, but changes were made across the entire gel image, not in single lanes. To quantify PCR products, band intensity was measured using the Multi-Analyst Software.

2.4.3 Subcloning and sequencing of PCR products

Single band PCR products to be sequenced were cut from the agarose gel and purified using the QIAquick Gel Extraction kit (QIAgen) as per manufacturers instructions. The PCR products were then subcloned into the pCR2.1-TOPO vector using the Invitrogen TOPO TA cloning kit, as per manufacturers instructions. Recombinant plasmids were identified using blue-white screening, and 5-10 single white colonies picked for DNA miniprep. The single colonies were grown overnight in a shaking incubator in 3 ml LB broth supplemented with Ampicillin (Sigma; 100 µg/ml), at 37°C. Plasmid DNA was subsequently isolated from the minicultures using the QIAprep plasmid DNA miniprep kit (QIAGEN) as per manufacturers instructions. An aliquot of 1-2 µl of the resultant plasmid DNA was digested with *EcoR* I (see **2.4.4**) and run on a 1.5-2% gel (see **2.4.2**) to confirm insert identity. Approximately 5-10 µg of miniprep DNA was then ethanol precipitated (see **2.4.5**) and the pellet sent to MWG Biotech UK for sequencing. Single strand sequencing was performed using either the M13 universal or reverse primers.

2.4.4 Sequence analysis

The raw data of all sequences generated was checked by analysis of the sequence chromatograms; any sequence of poor quality data was removed. Ambiguous regions of sequence were re-sequenced by MWG Biotech, performing both primer design and sequencing of specified region. For inserts longer than 1.5 kb, sequencing was performed in 500-600 bp reads. Primer walking was employed to gain complete sequence for the insert. The 500-600 bp sequence reads were assembled into a single contiguous sequence using the Staden Sequence Analysis Package. Analysis and editing of generated sequences was performed using the Genetic Computer Group (GCG) Wisconsin Sequence Analysis Package (Version 10 for Unix). Pairwise sequence alignments were generated using the 'Gap' application, and multiple sequence alignments were generated using the 'localpileup' application. Default settings were used in all cases. Multiple sequence alignments were displayed and shaded using the GeneDoc Multiple Sequence Alignment Editor and Shading Utility Version 2.6.001. The Shade options were set to 'Conservation mode', 'Shade to 4 levels' and 'Enable similarity groups'.

2.4.5 Restriction enzyme digestion

Restriction enzyme digestion of DNA was normally performed in a 20 µl volume containing 500 ng-1 µg DNA, 10U restriction enzyme and 2 µl of the appropriate 10X buffer. Incubation was generally performed for 2-4 hr at 37°C. Enzymes were obtained from either Roche or New England Biolabs.

2.4.6 Ethanol Precipitation

To precipitate DNA, 0.1 volumes of 3 M sodium acetate (pH 5.0) and two volumes of 100% ethanol were added to the DNA sample and then frozen at -70°C for 30 min. The DNA sample was then centrifuged at 19280 RCF (Relative Centrifugal Force) for 30

min (MIKRO 22R benchtop refrigerated centrifuge; Hettich Zentrifugen), then the supernatant removed. The pellet was washed using 500 µl of 70% ethanol and the pellet was centrifuged at 19280 RCF for 20 min. The supernatant was removed and the DNA pellet air-dried at room temperature. The DNA pellet was subsequently resuspended in a suitable volume of water.

2.4.7 cDNA library screen

The cDNA library screen was performed using a commercially prepared adult sheep lung cDNA library packaged in the Uni-ZAP® XR vector (Stratagene).

2.4.7.1 Plating of Phage

A fresh stock of XL1-blue MRF (Stratagene) was prepared by inoculating 50 ml LB broth (supplemented with 10mM $MgSO_4$ and 0.2% maltose) with a single colony of these bacteria. The culture was grown overnight, until it reached an optical density of 0.5 at 600nm. The bacteria were pelleted at 483 RCF (Beckman J2-21M/E refrigerated centrifuge) for 10 min, and resuspended in 10mM $MgSO_4$ to an optical density of 0.5 at 600nm.

During preliminary experiments, the XL1-blue MRF bacteria were mixed with various dilutions of the bacteriophage cDNA library to ascertain the dilution required to gain approximately 50000 pfu/plate. The bacteriophage/bacteria mixture was incubated for 20 min at 37°C; 6.5 ml molten top agarose (47°C) was then added and poured onto large (150 mm) NZY agar plates. Plates were incubated for approximately 8 h, or until the plaques showed confluent growth. Once confluent phage density was obtained, 10 large NZY plates were prepared as described for the primary screen. Plates were refrigerated for 2 h to prevent the top agar from sticking to the nitrocellulose filters in later steps.

2.4.7.2 Plaque lifts

Hybond C Nitrocellulose filters (150 mm; APBiotech) were transferred to the NZY plate for 2 min and both filters and plates were marked for orientation using an ink-filled syringe. Duplicate lifts were performed where two filters were subsequently transferred onto phage infected NZY plates for 2 mins and 5 mins respectively. After transfer, filters were denatured by submerging in 1.5 M NaCl/0.5 M NaOH for 2 min, then neutralised for 5 min in 0.2 M Tris-HCl (pH 7.5)/2X SSC. Filters were subsequently blotted on Whatman 3MM paper, then baked at 80°C for 2h. After baking, filters were wetted with 2X SSC prior to hybridisation.

2.4.7.3 Hybridisation

Filters were prewashed at 50°C in 100 ml pre-wash solution containing 5X SSC, 2% SDS and 1 mM EDTA (pH 8.0), in a rotisserie hybridisation oven (also used for subsequent hybridisation and post-hybridisation washes). The pre-wash solution was removed and 10 ml hybridisation buffer (see **2.2**) was added for pre-hybridisation. Filters were prehybridised at 43°C for 1-2 h.

Typically, 100-250 ng of PCR product was used as the DNA probe (PCR product cut out from gel and purified using QIAprep Gel Extraction kit following the manufacturers instructions). The PCR product was labelled with [α-^{32}P] dCTP radiolabelled nucleotide (APBiotech) using the Multiprime™ DNA labelling system (APBiotech) as per manufacturers instructions, then purified using the Nick™ column (APBiotech) as per manufacturers instructions. The labelled probe was denatured at 95°C for 5 min, then cooled on ice. Sufficient denatured probe (1 million counts per ml) was added to 10 ml fresh hybridisation buffer (see **2.2**) in the bottle containing the filters. Hybridisation was performed overnight at 43°C.

2.4.7.4 Post-hybridisation washes

The hybridisation buffer was removed and the following washes performed:

1st wash: 5 min at room temperature, shaking gently, with 200 ml of 2X SSC + 0.5% SDS

2nd wash: 200 ml of 0.5X SSC + 0.1% SDS at 60°C for 1h.

3rd wash: 200 ml of 0.1X SSC + 0.1% SDS at 60°C for 1h.

4th wash: as 3rd wash.

After the final wash, filters were prepared for exposure by blotting dry, then wrapping in SaranWrap™ with care to avoid trapping air bubbles. Filters were placed against Hyperfilm™ MP in an autoradiography cassette and exposed at -70°C. The exposed film was developed using the automated film processor. Positive signals on the film were aligned with the corresponding plate using the ink orientation marks. Positive clones were subsequently isolated as an agar plug, and stored in 0.5 ml SM buffer at 4°C for use in secondary screen.

2.4.7.5 Secondary screen

The bacteriophage were diluted accordingly in order to gain a moderate covering of plaques on the phage infected NZY plates, with plaques well separated. Plaque lifts, hybridisation and washes were performed as described previously (**2.4.7.2**; **2.4.7.3**; **2.4.7.4**). Positive plaques were isolated and stored in 0.5 ml SM buffer at 4°C for use in tertiary screen.

2.4.7.6 Tertiary screen

The bacteriophage were diluted in order to gain isolated plaques. Plaque lift and hybridisation steps were repeated as described before. Positive plaques were isolated and stored in SM buffer at 4°C until required for pBluescript phagemid excision.

2.4.7.7 Excision of pBluescript phagemid

Fresh stocks of XL1-blue MRF and SOLR bacteria (Stratagene) were prepared by inoculating 50 ml LB broth with a single colony. The cultures were grown overnight, until it reached an optical density of 0.5 at 600 nm. The bacterial cultures were then spun down at 483 RCF (Beckman J2-21M/E refrigerated centrifuge) for 10 min, and resuspended in 10 mM $MgSO_4$ to an optical density of 1.0 at 600nm.

To excise the pBluescript phagemid containing the 'positive' cDNA insert from the positive lambda vector, 125 µl of the lambda phage stock was incubated with the fresh 200 µl XL1 blue MRF' bacteria (OD_{600} = 1.0), 125 µl SM buffer and 1 µl ExAssist helper phage (3.75 pfu/ml; ExAssist phage were provided with the Stratagene cDNA library) at 37°C for 15 min. After incubation, 3 ml of LB broth was added to the tube and then incubated for 2-3 h at 37°C with shaking, followed by a further incubation at 65°C for 20 min. The tube was then centrifuged at 2296 RCF (Sorvall RC-3B refrigerated centrifuge) for 20 min and the supernatant decanted into a sterile 15ml centrifuge tube. To plate the excised phagemids, 200 µl of the fresh SOLR cells (OD_{600} = 1.0) was added to 10 µl of the phage supernatant and incubated at 37°C for 15 min. The bacteriophage/SOLR mixture was then plated onto pre-warmed LB-Amp plates and incubated at 37°C overnight. 10 single colonies per initial bacteriophage clone were subsequently picked from the plates and DNA was prepared using the QIAprep plasmid DNA miniprep kit (Stratagene) as per manufacturers instructions.

2.4.8 Rapid amplification of 5' cDNA ends (5' RACE)

5'RACE was performed to gain additional 5' DNMT1 cDNA sequence, after sequence analysis of the cDNA library screen products. The 5'RACE method was performed using the Clontech SMART RACE kit, as per manufacturers instructions. Prior to 5' RACE, RNA was extracted from tissue or oocytes, as described **2.4.11.1; 2.4.11.2.** **Figure 2.1** outlines the modified cDNA synthesis method utilised by the Clontech SMART RACE kit.

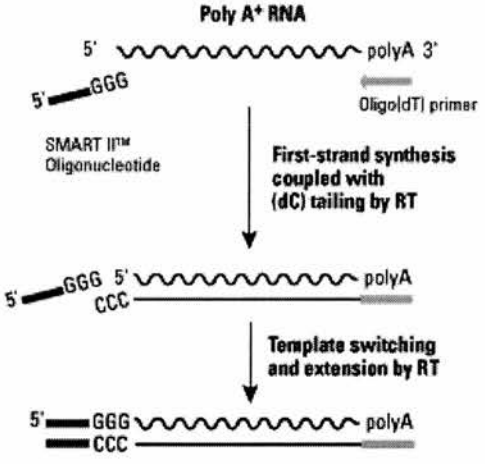


Figure 2.1 The Clontech SMART RACE cDNA amplification protocol.
First-strand synthesis is primed using a gene-specific primer (replacing oligo(dT) primer shown in figure). After reverse transcriptase reaches the end of the mRNA template, it adds several dC residues. The SMART II Oligonucleotide anneals to the tail of the cDNA and serves as an extended template for the reverse transcription (RT).

After first-strand cDNA synthesis, rapid amplification of cDNA ends (SMART RACE PCR) was performed using a gene-specific primer and the universal primer mix (specific to the SMART II oligonucleotide) provided with the kit. Gene-specific primers were required to have an annealing temperature of 65-70°c, to be compatible with the universal primer mix. Touchdown PCR was performed using the Advantage® 2 PCR

Kit (Clontech) provided with the SMART RACE kit, as per manufacturers instructions. Touchdown PCR increases sensitivity and reduces background (or non-specific bands) by initially creating stringent PCR conditions (high annealing temperature) optimal for amplification of gene-specific products only. As the PCR progresses, annealing temperatures are decreased gradually, relaxing conditions, and allowing further amplification of the enriched gene-specific target. Touchdown PCR conditions are described in **Table 2.4**

Temperature °C	**Time (mins:seconds)**	**Number of cycles**
94	1:00	1
94 72	0:20 3:00	5
94 70 72	0:20 0:30 2:30	5
94 68 72	0:20 0:30 2:30	30
72	5:00	1

Table 2.4 PCR conditions used for 5' RACE PCR.

The touchdown PCR program starts with high specificity reflected in the high annealing temperature. With subsequent stages of the PCR, the conditions are relaxed, allowing accumulation of the specific PCR products.

2.4.9 DNA extraction from tissues and cultured cells

Approximately 100 mg of tissue was homogenised in 5 ml DNA extraction buffer A (see **2.2**) using a Polytron homogeniser. After homogenisation, 5 ml of DNA extraction buffer B (see **2.2**) and 1 µg Proteinase K were added, and the homogenate was incubated overnight at 37°C on a slow shaking tray. The DNA was subsequently isolated using a standard phenol:chloroform extraction method. One volume of phenol was added and the solution mixed well, followed by addition of 1 volume of chloroform:isoamyl alcohol (24:1). After thorough mixing, the solution was centrifuged at 2296 RCF (Sorvall RC-3B refrigerated centrifuge) for 20 min. The aqueous layer was subsequently isolated, and the phenol:chloroform extraction repeated. After the second isolation of the aqueous layer, a final extraction was performed using 1 volume of chloroform:isoamyl alcohol (24:1), and the thoroughly mixed solution was centrifuged as described for the previous step. The aqueous phase was again isolated, and 0.2 volumes of 10 M ammonium acetate and 2 volumes of ice-cold 100% ethanol were added. The solution was shaken gently until the DNA precipitated out of solution. The DNA was removed, and washed in 70% ethanol, prior to drying at 37°C. DNA was resuspended in 50-100 µl dH_2O according to desired use.

To extract DNA from cultured cells, the QIAgen DNeasy kit (QIAgen) was used, as per manufacturers instructions. Frozen cell pellets were used as the starting material, and DNA was eluted in 100 µl of dH_2O. The QIAgen DNeasy kit was also used for the isolation of DNA from sperm, with a number of modifications made to the protocol. Briefly, 100 µl of Buffer X2 (see **2.2**) was added to 100 µl of sperm, with the DTT and Proteinase K added to buffer X2 just prior to use. The sample was incubated at 55°C for 1 hour, with inversion of the tube occasionally to ensure the contents dispersed thoroughly. When the sample was dissolved, 200 µl of Buffer AL (provided in Dneasy kit) and 200 µl of ethanol were added to the sample. The mixture was subsequently pipetted onto a DNeasy mini column, and the column centrifuged at 6290 RCF for 1 minute. The flow-through was discarded and the DNeasy column placed into a new 2

ml collection tube. Then 500 µl of Buffer AW1 (provided in DNeasy kit) was added and the column centrifuged as before. Again, flow-through was discarded, and the DNeasy column placed in a fresh collection tube. To wash the DNA a final time, 500 µl of Buffer AW2 (provided in DNeasy kit) was added to the DNeasy column, and the column was subsequently centrifuged at 6290 RCF for 3 mins. The DNeasy column was placed in a 1.3 ml centrifuge tube and 100 µl of dH20 was pipetted directly onto the DNeasy column membrane. The DNA was eluted by centrifugation at 6290 RCF for 1 minute.

2.4.10 Single embryo methylation analysis

2.4.10.1 DNA extraction from single oocytes or embryos

The tube containing the single embryo in 1 µl culture medium was freeze-thawed three times. Freezing was performed at -80°C for periods of 20-30 min, with the embryos defrosted at room temperature between subsequent freezes. The embryo was subsequently digested with Proteinase K (0.1 µg) at 50°C for 1h. Prior to ethanol precipitation, Proteinase K activity was heat-killed at 80°C for 20 min. DNA was ethanol precipitated using either 2 µl SeeDNA™ co-precipitate (APBiotech) or Pellet Paint™ (Novagen) in a standard sodium acetate/ethanol precipitation with the following specifications: 100% ethanol wash was performed at 22130 RCF (MIKRO 22R benchtop refrigerated centrifuge) for 1 h; 70% ethanol wash was performed at 22130 RCF for 30 min. The DNA pellet was resuspended in 6 µl dH_20.

2.4.10.2 Digestion of single embryo DNA

DNA from single oocytes/ embryos was divided into two aliquots of 3 µl. One aliquot was digested with *Hha* I restriction endonuclease (New England Biolabs) in a 19 µl digestion reaction containing 20U *Hha* I, 1X NEBuffer 4 and 1X BSA for 2 h at 37°C. A further 20U *Hha* I was added after two hours (to ensure complete digestion of DNA)

and incubated for an additional two hours. Enzyme activity was subsequently heat-killed at 65°C for 20 min. The second DNA aliquot was incubated in a 20 µl digest reaction containing 1X NEBuffer 4 and 1X BSA for 4 h at 37°C, followed by a 65°C incubation for 20 min.

2.4.10.3 Methyl-sensitive PCR

PCR reactions were performed in a 25 µl volume containing: 1 X PCR buffer (Roche diagnostics; final concentration. 1.5mM), 200 µM dNTP's (Invitrogen), 250nM each primer, 1U *Taq* DNA polymerase (Roche diagnostics) and either 2 µl *Hha I* digested or undigested DNA, or control template. Thermal cycler conditions for specific primers can be found in **Table 2.5**.

Primers/embryo stage	No. of cycles	Denaturing		Annealing		Extension	
		°C	min:sec	°C	min:sec	°C	min:sec
Ovsat1a/blastocyst	1	94	15:00		-		-
	19	94	0:30	62	0:30	72	0:45
	1		-		-	72	7:00
Ovsat1a/morula	1	94	15:00		-		-
	21	94	0:30	62	0:30	72	0:45
	1		-		-	72	7:00
Ovsat1a/8-16 cell	1	94	15:00		-		-
	22	94	0:30	62	0:30	72	0:45
	1		-		-	72	7:00
Ovsat1a/oocyte	1	94	15:00		-		-
	24	94	0:30	62	0:30	72	0:45
	1		-		-	72	7:00
Bovb/blastocyst	1	94	15:00				
	24	94	0:30	58	0:30	72	0:45
	1		-		-	72	7:00

Table 2.5 PCR conditions for primers used in methyl-sensitive PCR.
The table indicates the cycling conditions used for each primer pair at different developmental stages.

2.4.10.4 *In vitro* CpG methylation of plasmid DNA

One μg of plasmid DNA was incubated for 1-2 h at 37°C in a reaction mixture containing 1X SAM, 1X NEBuffer 4 and 5U *Sss* I methylase (NEB), made up to 20 μl with water. *Sss I* methylase activity was subsequently heat-killed at 65°C for 20 min, prior to restriction enzyme digestion.

2.4.11 RNA extraction and cDNA synthesis

2.4.11.1 Tissues and cultured cells

RNA was extracted from tissues using the Hybaid Ribolyser™ system combined with the QIAgen RNeasy kit (QIAgen). The Ribolyser™ technology combines the use of silica and ceramic beads (within the Ribolyser™ tube) with high speed twisting and shaking movements. Friction generated by the collision of these beads allows disruption of the tissue, producing a lysate suitable for subsequent RNA extraction. Briefly, 100mg tissue was placed in a Ribolyser™ tube containing 1.8 ml RNeasy buffer RLT. The tube was then placed in the Ribolyser™ and subjected to a 30 sec pulse of high-speed motion, sufficient to disrupt the cells. One third of the lysate (600µl) was used in the subsequent RNA extraction, using the QIAgen RNeasy kit as per manufacturers instructions. For RNA extraction from pelleted fibroblast cells, the cells were lysed using the QIAshredder kit (according to manufacturers instructions). RNA extraction from the lysed cells was then continued as for tissues, using the QIAgen RNeasy kit. An optional DNase digestion step was performed (according to maunfacturers instrictions) using a DNase kit specifically supplied for use with the QIAgen RNeasy kit. RNA was reverse transcribed using the First strand cDNA synthesis kit (APBiotech) as per manufacturers instructions, starting with 1 µg of total RNA and using the random hexamer ($pd(N)_6$) primer for cDNA synthesis.

2.4.11.2 Single oocytes and embryos

RNA was extracted from single oocytes or preimplantation embryos using the QIAGEN QIAshredder kit and RNeasy Mini kit as per manufacturers instructions. RNA was then ethanol precipitated using Pellet Paint™ coprecipitant (Novagen) as per manufacturers instructions. The RNA was resuspended in 8 µl DEPC-treated H_2O. Reverse transcription was performed immediately after RNA extraction using the First-Strand cDNA Synthesis kit (APBiotech) as per manufacturers instructions, using the random

hexamer ($pd(N)_6$) primer for cDNA synthesis. Embryo and oocyte cDNA was stored at –20°C after preparation.

2.4.12 Taqman® RT-PCR

The Taqman® assay, used in conjunction with the ABI PRISM® 7700 Sequence Detection System (Applied Biosystems), provides a highly sensitive method for quantification of mRNA (Heid et al., 1996). By combining fluorogenic probes with the 5' exonuclease activity of Taq polymerase, real-time monitoring of PCR product accumulation is allowed. **Figure 2.2** demonstrates the Taqman real-time PCR chemistry.

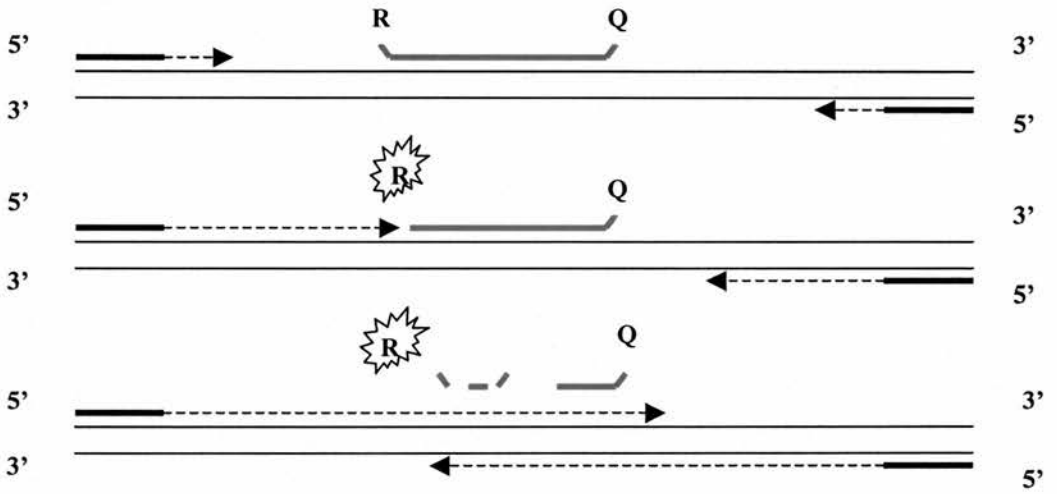


Figure 2.2 The Taq polymerase extends the forward and reverse primers (highlighted as solid black horizontal lines) as normal. The fluorogenic gene-specific probe (red solid horizontal line)has a reporter (R) dye at the 5' end and a quencher (Q) dye at the 3' end. The gene-specific probe anneals to the template between the primers. When the primer extension reaches the probe, the taq exonuclease activity cleaves the reporter dye from the probe. The reporter dye is then released from the quencher, producing a fluorescent signal. As the amount of PCR product generated in each cycle increases, a proportional amount of flourescence is released.

PCR product accumulation is measured as an increase in fluorescence signal detected by the sequence detection software. The fluorescence of the reporter dye is normalised to a passive reference dye included in the PCR mastermix; this is necessary to correct for any fluctuation in fluorescence due to discrepancies in concentration or reaction volume. The PCR cycle at which a statistically significant increase in fluorescence is observed is described as the Ct value, or 'threshold cycle', and correlates with the beginning of the exponential phase of the PCR. The Ct value is the predictive target quantification value. Figure **2.3** shows the typical graphical output of a Taqman® reaction, demonstrating how Ct values are obtained.

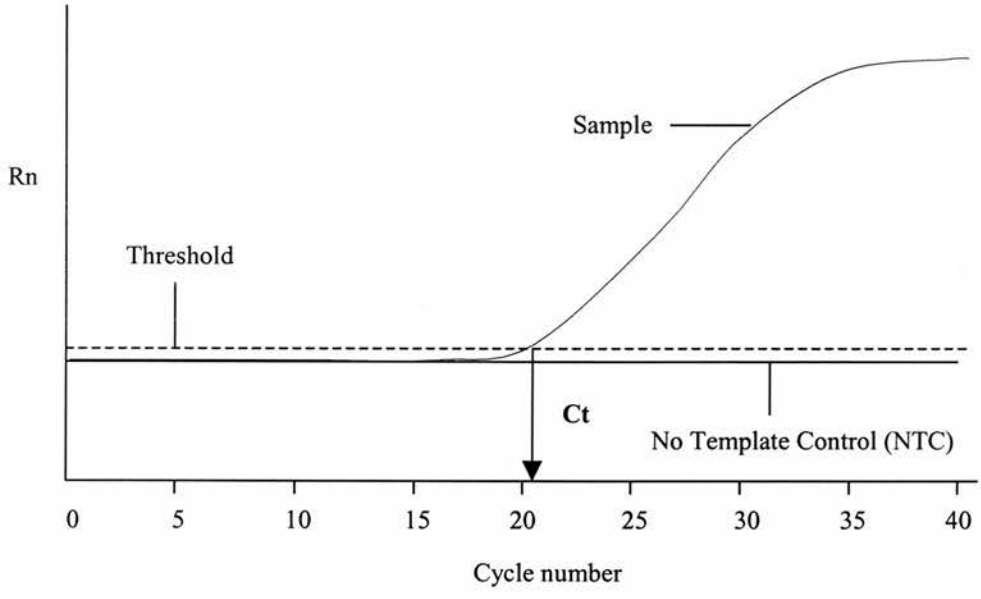


Figure 2.3 Taqman® real-time PCR graphical output. The threshold cycle (Ct) is calculated as the point when significant increases in reporter dye fluorescence (Rn) are observed, rising above the 'background' emission intensity noise. The level of background noise is determined by the emission intensity in the No Template Control (NTC).

2.4.12.1 Relative quantification of gene expression using the Taqman Assay

For analysis of gene expression in adult tissues, it was decided to use 18S ribosomal RNA (rRNA) as an endogenous control. Due to the abundant nature of 18S rRNA in adult tissues, a separate tube method was performed instead of multiplex PCR, to avoid primer competition within a single PCR reaction. The comparative Ct method quantifies gene expression relative to a 'calibrator' tissue, after normalisation of expression against endogenous gene 18S rRNA. In order to validate this method of quantification, it is important that the gene-specific and 18S rRNA primers both amplify target cDNA with the same efficiency. This is achieved by assessing the Ct values over a range of cDNA dilutions. As 18S rRNA is more abundant in tissues than mRNA transcripts, a different, or 'more dilute' range of cDNA dilutions can be used in order to conserve cDNA stocks. This is an acceptable practice for this method of quantification, although it remains essential that the intervals between dilutions remain constant between the two primer sets when evaluating relative efficiency. For example, a range of dilutions 1:10, 1:50 and 1:100 may be used for the gene-specific primers, and for 18S rRNA, dilutions of 1:100, 1:500, 1:1000 would be equivalent. By plotting the log of input (cDNA dilution factor) against the difference in Ct values (ΔCt; gene-of- interest Ct value minus18S rRNA Ct value), equally efficient primers will demonstrate a slope of <0.1. If a slope of >0.1 is observed, 18S primer concentrations can be reduced to limit the primer efficiency to an appropriate level. Once confirmation of equal primer efficiencies is obtained, gene expression can be quantified using the ΔΔCt equation:

$$2^{-\Delta\Delta Ct}$$

When analysing gene expression in a range of tissues, expression values are first normalised against the relative 18S rRNA level (ΔCt). A calibrator tissue is then chosen, usually the tissue with the lowest level of gene expression (or highest ΔCt value) and is designated the 1x sample. The calculation of ΔΔCt involves subtraction by the ΔCt calibrator value; for example if heart tissue is the calibrator value, then all tissues will subtract the heart ΔCt value from their own ΔCt value. The calculation $2^{-\Delta\Delta CT}$ is then completed, and the tissue gene expression levels are expressed as a '*n* fold' increase relative to the calibrator (1x) tissue. The standard deviation of the difference in Ct values (ΔCt) is calculated from the standard deviations of the 'gene-of-interest' and 18S values. The calculation is given below:

$$S = \sqrt{S_1^2 + S_2^2}$$

For each tissue analysed;

S_1 is the standard deviation of the gene-of-interest replicate Ct values.

S_2 is the standard deviation of the 18S replicate Ct values.

2.4.12.2 Taqman® PCR method

To quantify gene expression in tissues, we used the Real Time Quantitative PCR (Taqman® PCR, Applied Biosystems) method utilising the ABI 7700 Sequence Detector (Applied Biosystems). Gene-specific Taqman® primers and probes were designed using Primer Express® software (Applied Biosystems; default parameters). Triplicate reactions were prepared containing 12.5 μl of 2X Universal PCR Mastermix (containing AmpliTaq Gold), 2.5 μl each of 3-9 μM (final concentration 300-900 nm) forward and reverse gene-specific primers or 2.5μl each of 0.25-0.50 μM (final concentration 25-50 nm) 18S forward and reverse primers, 2.5 μl of 2 μM (final concentration 200 nm) gene-specific/18S Taqman® probe, 2.5μl dH_2O and 2.5 μl cDNA. Reactions were performed

in a MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems) and the wells were sealed using MicroAmp® Optical Caps (Applied Biosystem). The initial UNG digestion was performed at 50°C for 2 min, followed by activation of the AmpliTaq Gold at 95°C for 10 min. Thermal cycling then proceeded with 40 cycles of 95°C for 15 sec and 60°C for 1 min. Primer sequences and primer/probe concentrations optimised for relative equal efficiency with 18S are described in **Table 2.6**.

Primer/Probe name	*Sequence 5' – 3'*	*Optimal final primer/probe concentration (nM)*
DT1F	CGCTACTTCCTCTTGGAGAACG	900
DT1R	CGCAGCAGTCAGCTTCAGG	300
DT1 probe	AAACTTCGTCTCCTTCAAGCGCTCCATG	200
DT3AF	CGGCCAGTGCCCTCG	900
DT3AR	CCTGCAATGATCTCTCCATCG	900
DT3A probe	AGAGTCCCTTGCGGGCAGGATTG	200
DT3BF	CTCTTGTCGCCAACCTTCA	900
DT3BR	ATTACTCACGCCCCAAGGA	900
DT3B probe	TGGCCACCACGTTCTCAAACATC	200
MD2F	GCAACCAGTAACCAAAGTCACA	900
MD2R	TGTAGCCTGTTGTCCCAGAA	900
MD2 probe	CCCACAGAGAATGAATGAACAGCCA	200
18SF	-	50/*25
18SR	-	50/*25
18S probe	-	200

Table 2.6 Taqman primer/probe sequences and optimal PCR conditions.

Primer and probe sequences are described for all gene-specific sets; the 18S primer and probe sequences (Applied Biosystems 'ribosomal RNA control reagents') are not disclosed by the manufacturer. Optimal primer and probe concentrations are also described; these conditions allow for relative quantification of gene expression using the 'Comparative Ct' method.

* 18S rRNA primers were used at a final concentration of 25 nm when used in separate tube quantification of DNMT3A expression

Chapter 3

Cloning and sequence analysis of the sheep DNA methyltransferase *DNMT1* coding regions

3.1. Introduction

DNA methyltransferase 1 (Dnmt1) was, until recently, the only DNA (cytosine-5) methyltransferase identified in vertebrate tissues. The *Dnmt1* gene and encoded protein have been studied extensively in the mouse, and also in humans and other vertebrate species, albeit to a lesser extent. The mouse *Dnmt1* gene spans 56 Kb of genomic DNA, is organised into 39 exons and encodes a protein of 1502 amino acids. The human gene is similarly organised, except for the presence of an additional intron (Margot *et al.*, 2000).

Since the initial identification, the structure and function of mouse *Dnmt1* has been widely investigated. The presence of ten sequence motifs observed in prokaryotic 5-cytosine DNA methyltransferases was detected in the mouse protein (Lauster *et al.*, 1989; Posfai *et al.*, 1989), six of these motifs are highly conserved between all DNA cytosine-5 methyltransferases. These motifs encode the enzymatic domain of Dnmt1, and play different roles in the catalysis of DNA methylation. These include interactions such as the folding together of motifs I and X to form the AdoMet (S-Adenosyl-Methionine; the substrate for Dnmt1) binding site. Motif IV contains the Pro-Cys dipeptide catalytic active site, motif VI contains a glutamyl residue required for protonation of the cytosine target, and motif IX is involved in maintaining the target recognition domain structure (located between motifs VIII and IX; Motif interactions are reviewed in Bestor & Verdine, 1994).

The catalytic domain of Dnmt1 is found at the C-terminus of the protein, and is one of three postulated structural compartments (Margot *et al.*, 2000). The other two compartments are the N-terminal domain, containing several functional domains of interest, and a central domain, with poorly characterised function. **Figure 3.1** demonstrates the structural compartmentalisation of Dnmt1 and highlights known functional domains of interest.

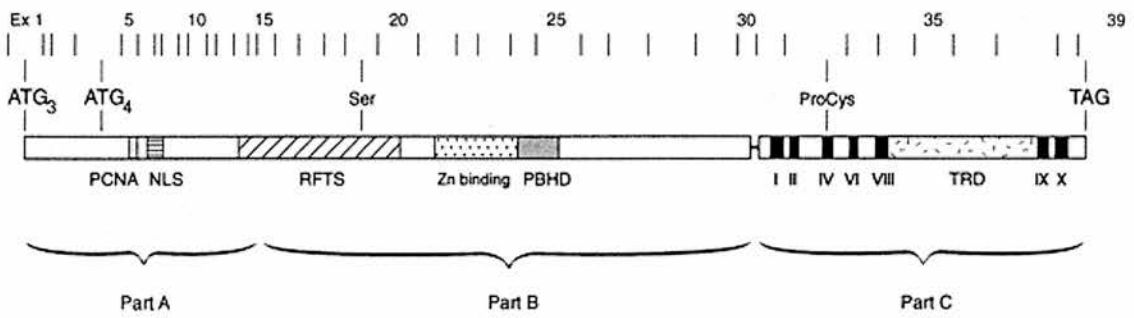


Figure 3.1 The 'Tri-partite' structure of Dnmt1. Taken from Margot *et al.*, 2000. A schematic representation of the three distinct regions (Parts A, B & C) of the mouse Dnmt1 protein. Part A represents the N-terminal domain, part B represents the central domain, and part C represents the catalytic domain. Known functional domains and amino acids of interest are indicated in the diagram; relative exon positions are highlighted above the protein domains. Key: PCNA (Proliferating cell nuclear antigen), NLS (Nuclear localisation signal), RFTS (Replication foci targeting sequence), PBHD (Poly-bromo1 protein homology domain), TRD (Target recognition domain).

As mentioned previously, the C-terminal region of Dnmt1 encodes the catalytic domain, and demonstrates similarities with bacterial Type II DNA cytosine methyltransferases (Bestor, 1988). However, it is the large N-terminal region that differentiates mouse Dnmt1 from bacterial and other vertebrate methyltransferases. The Dnmt1 N-terminus has been suggested to be responsible for conferring a number of specialised functions through its many domains including the preferential recognition of hemi-methylated DNA (Bestor *et al.*, 1992).

The functions of these domains have been mainly characterised using various deletion mutants. The domains include the Replication-fork Targeting (RFT) domain, responsible for targeting Dnmt1 to sites of newly synthesised hemi-methylated DNA (Liu *et al.*, 1998; Leonhardt *et al.*, 1992). Upstream of the RFT domain lies the Nuclear Localisation Signal (NLS), responsible for the import of Dnmt1 into the nucleus (Leonhardt *et al.*, 1992). The Zinc binding domain (Bestor, 1992) lies adjacent to the Polybromo-1 Protein Homologous Domain (PBHD), so-called due to the sequence homology with the Poly-bromo-1 protein (Nicolas & Goodwin, 1996). The function of this domain has so far remained undefined. Protein binding sites identified within the N-terminal domain include the PCNA and Dnmt1-associated protein 1 (DMAP1) binding sites. PCNA is a component of DNA replication protein complexes and is thought to interact with Dnmt1 at the replication fork (Chuang *et al.*, 1997). DMAP1 is involved in the recruitment of histone deacetylase complexes to Dnmt1 (Rountree *et al.*, 2000).

3.1.1 Sex-specific promoters of *Dnmt1*

All forms of Dnmt1 found in mouse somatic cells and tissues are the result of transcriptional initiation from exon 1. In germ cells, alternative forms of Dnmt1 protein are produced by initiation of transcription from sex-specific promoters (Mertineit *et al.*, 1998). In oocytes, an alternative first exon, exon 1o (somatic exon 1 was designated by Mertineit *et al.* as exon 1s; and the 'sperm-specific' exon as 1p), leads to the initiation of translation from ATG_4 in exon 4, forming a truncated Dnmt1o protein. Dnmt1o is extremely abundant in mature oocytes where it is localised at the cytoplasmic periphery of the cell, and is thought to be the sole form of Dnmt1 found in oocytes and preimplantation embryos (Mertineit *et al.*, 1998). In pachytene spermatocytes, transcription is driven from the exon 1p (p = pachytene) promoter and translation of Dnmt1 from this transcript leads to a grossly truncated protein, due to the presence of multiple short open reading frames within exon 1p. **Figure 3.2** demonstrates the genomic organisation of the mouse *Dnmt1,* indicating sex-specific exons. The figure also shows mRNA products generated from the somatic and sex-specific exons, indicating the effect of alternative promoter use.

The function of sex-specific Dnmt1 proteins has been the source of much speculation. Dnmt1p is undetectable in spermatocytes, with Mertineit *et al.* (1998) suggesting this is due to the expected truncated nature of the protein. However, the 'sperm-specific' *Dnmt1* isoform can be translated in other tissues, specifically skeletal muscle, despite the presence of the multiple short reading frames (Aguirre-Arteta *et al.*, 2000). This gives rise to a shorter form of Dnmt1 protein and puts into doubt the sex-specific nature of this *Dnmt1* transcript.

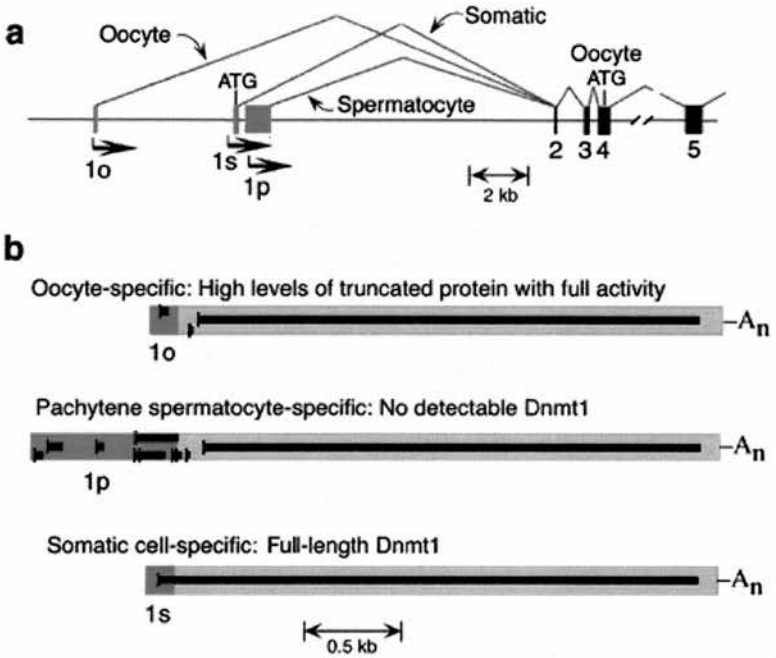


Figure 3.2 Genomic organisation of the mouse *Dnmt1* gene.

Taken from Bestor, 2000. (a) The genomic locations of sex-specific and somatic exons 1o, 1s and 1p in the mouse *Dnmt1* gene are indicated by red boxes. Alternate splicing of these exons and start codons are also highlighted in the diagram. (b) The figure describes the Dnmt1 protein products of sex-specific splicing, highlighting the effects. The dark horizontal bars indicate open reading frames; short vertical bars indicate ATG initiation codons.

The abundant nature and unusual localisation of Dnmt1o in oocytes and preimplantation embryos is certainly intriguing, in particular, the unusual localisation of Dnmt1o when compared to the somatic form, Dnmt1s. Dnmt1s is first observed around the time of implantation, and is abundant in the nuclei of somatic cells and tissues. On the contrary, Dnmt1o is found exclusively in the cytoplasm of preimplantation embryos, with the exception of the 8-cell stage, where it has been observed by immunodetection to enter and subsequently exit the nucleus (Carlson *et al.*, 1992). The reason for this stage-specific nuclear localisation was recently determined through experiments involving

targeted deletions of Exon 1o of the mouse Dnmt1 gene. Mice with homozygous deletions demonstrated methylation errors within certain imprinted genes (Howell *et al.*, 2001). Thus, it appears Dnmt1o is responsible for maintaining DNA methylation patterns within specific imprinted genes solely during the 8–cell stage of development.

The presence of sex-specific Dnmt1 proteins in other vertebrate species so far remains undetermined. A study of *Xenopus laevis* oocytes demonstrated the abundant presence of somatic Dnmt1 but did not rule out the possibility of an oocyte-specific form (Shi *et al.*, 2001). However, Xenopus do not have imprinted genes, so may well not require a protein with such specialised activity. Studies of other vertebrate species known to have imprinted genes are clearly required.

The DNA methyltransferase genes have not previously been cloned or characterised in sheep. The aim of this particular study was to isolate the complete cDNA sequence encoding sheep DNMT1 in order to assess conservation of cDNA and protein sequences between sheep, humans and mice, and also to determine whether an oocyte-specific DNMT1 exists in sheep. Isolation of the *DNMT1* coding region will also allow expression analysis to be performed in adult sheep tissues and preimplantation embryos, to unravel the role of DNA methylation in sheep development. This chapter describes the experiments performed to obtain the full-length *DNMT1* cDNA sequence, the subsequent cDNA sequence analysis and protein prediction, and finally, attempts to identify a sheep oocyte-specific form of *DNMT1*.

3.2 Results

3.2.1. Generation of partial sheep DNMT1 cDNA clones by RT-PCR

Sequence alignments of human and mouse *DNMT1* cDNA sequences (NM001379; X14805) were generated using the 'Gap' application (default parameters) of GCG10. Primers specific to *DNMT1* were designed from the human sequence in regions seen to be highly conserved between the mouse and human. **Table 3.1** lists the primers used and the location within the source sequence. Pooled d125 fetal heart and kidney cDNA was used as the template in a standard PCR reaction performed as described **2.4.1.1.** PCR products were then electrophoresed on a 2% ethidium stained agarose gel as described **2.4.2.** PCR products of expected size (130 bp) were observed and subsequently cloned and sequenced (as described **2.4.3**) to confirm identity.

Gene	*Primer*	*Direction of primer*	*Sequence primer designed from (Genbank accession number)*	*Location of primer*
DNMT1	MTRA1F	Forward	NM001379	3936-3956
	MTRA1R	Reverse	NM001379	4066-4047

Table 3.1 PCR primers used to amplify partial *DNMT1* cDNA sequence in sheep. Primer sequences are given in **Table 2.3**

3.2.2. Isolation of a *DNMT1* cDNA clone from an adult sheep lung cDNA library

The partial cDNA clone generated by RT-PCR corresponding to the sheep *DNMT1* sequence was subsequently used to probe an adult lung sheep cDNA library (obtained from Stratagene). The cDNA library screen was performed as described **2.4.7.** The primary screen identified two positive clones, designated 4 and 6, which were re-plated for the secondary screen. After the secondary screen, several positive plugs were identified on the plates derived from clone 6; clone 4 appeared to be a false positive as it failed to demonstrate binding with the *DNMT1* probe in the second round of library screening. These single plaque colonies were isolated and PCR analysis performed using the MTRA1 primer set as an additional control. PCR products were electrophoresed on a 2% ethidium bromide stained agarose gel (as described **2.4.2)**. The results of this PCR are demonstrated in **Figure 3.3.** A single clone, 6.3a, was chosen for further analysis. Excision of the pBluescript plasmid (containing the cDNA fragment) from the Uni-ZAP® XR vector was subsequently performed (as described **2.4.7.7**) and plasmid DNA minipreps were established from 10 single colonies for clone 6.3 (as described **2.4.3**). PCR analysis using the MTRA1 primer set was again performed to check insert identity of the minipreps. PCR products were electrophoresed on a 2% ethidium bromide stained agarose gel (as described **2.4.2)**, results are shown in **Figure 3.4**. An *EcoR I/Xho* I double digest was also performed to release the insert from the pBluescript vector, and simultaneously estimate insert size for each clone. Digests were electrophoresed on a 1% ethidium bromide stained agarose gel (as described **2.4.2)**, results are shown in **Figure 3.5**

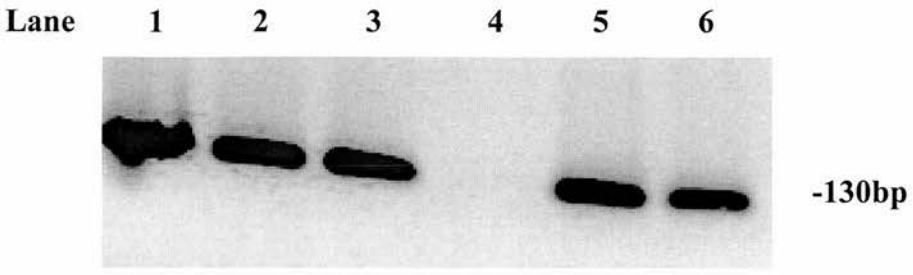


Figure 3.3 PCR analysis of positive plugs isolated after the secondary *DNMT1* cDNA library screen.

Bacteriophage clone DNA was amplified using MTRA1 primers. Lanes 1-6 represent PCR products amplified from clones 6.1a, 6.1b, 6.2a, 6.2b, 6.3a and 6.3b respectively. The size of the PCR product is indicated to the right of the gel image.

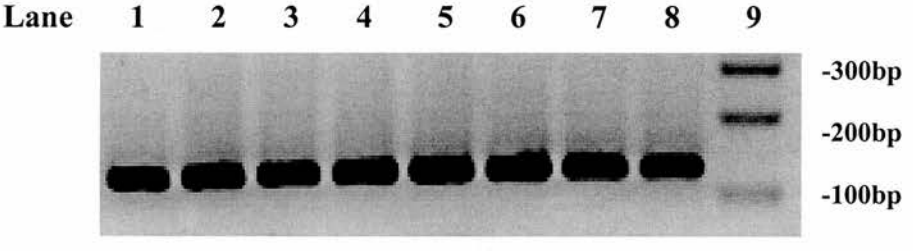


Figure 3.4 PCR analysis of DNA minipreps established from the Uni-ZAP® XR clone 6.3a

DNA minipreps were screened using the MTRA1 primers to confirm identity of the clones. Lanes 1-8 show PCR products amplified in minipreps prepared from 8 single colonies. DNA marker sizes (Lane 9; 1Kb Plus DNA ladder (Life Technologies)) are indicted to the right of the gel image.

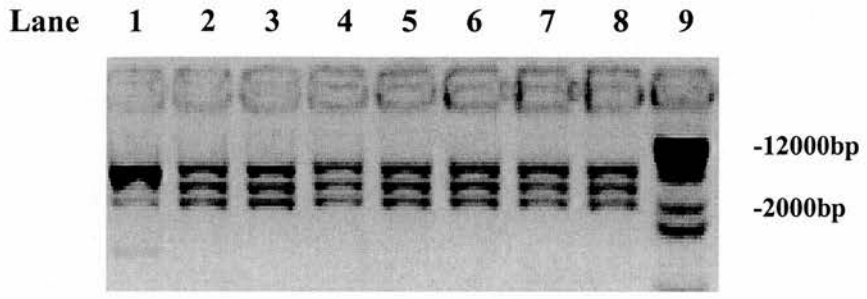


Figure 3.5 *EcoR* I/*Xho* I digestion of DNA minipreps established from Uni-ZAP® XR clone 6.3a.

Lanes 1-8 represent minipreps 1-8 respectively, DNA marker sizes (Lane 9; 1Kb Plus DNA ladder (Life Technologies)) are indicated to the right of the gel image.

A single miniprep originating from clone 6.3a, with an insert of approximately 4.2 Kb, was ethanol precipitated and sequenced (as described **2.4.6**; **2.4.3**). In total, 4243 nucleotides of sequence corresponding to the sheep *DNMT1* cDNA sequences were obtained. The sheep *DNMT1* cDNA sequence corresponded to the nucleotides 1185-5428 of the human *DNMT1* sequence (Accession No. NM001379), including the PolyA tail, but lacking the first 1 Kb of the coding region.

3.2.3 Cloning the 5' regions of the sheep *DNMT1* cDNA sequence

A further 1-1.5 Kb of 5' sequence was required to gain the full length *DNMT1* cDNA clone (known to be approximately 5.2 Kb in the mouse; 5.4 Kb in the human). It was decided to use a 5' RACE technique to clone the remaining sheep *DNMT1* cDNA sequence from adult tissue (heart) RNA. Parallel attempts were made to clone 5' regions of an oocyte-specific sheep *DNMT1*, using pooled oocyte RNA as the starting material instead of heart RNA.

An oligonucleotide, RACE Y, was designed within the initial 250 bp of the sheep *DNMT1* cDNA clone, in order to generate only *DNMT1*-specific cDNA during the modified 5' RACE reverse-transcription process. A control forward primer, RACE X, and an additional reverse primer, RACE Z, were also designed to provide a control for analysis of cDNA synthesised during later stages of the RACE protocol, allowing a method of confirming the specific presence of *DNMT1* cDNA. **Figure 3.6** demonstrates the positions of the RACE primers in the *DNMT1* sequence.

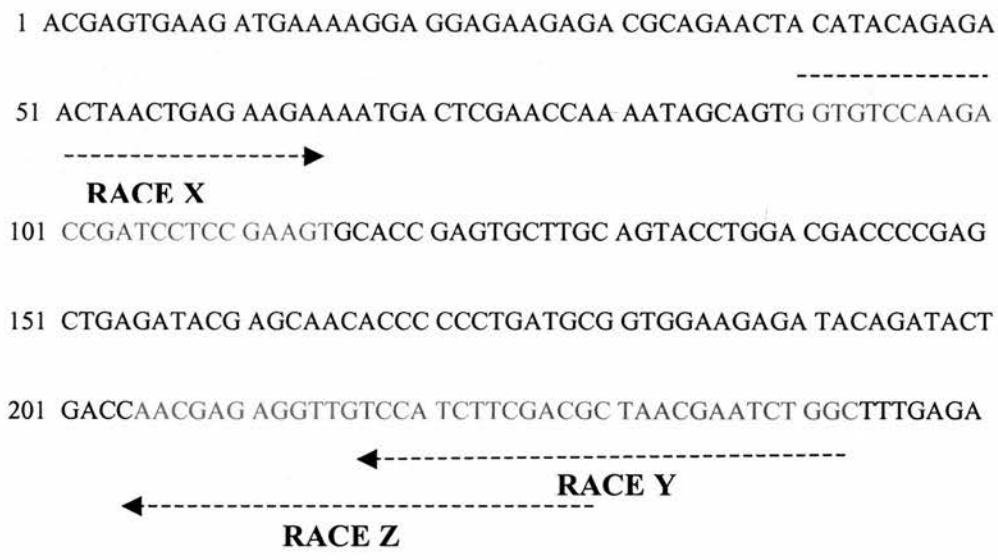


Figure 3.6 Sequence location of RACE primers designed to amplify 5' regions of sheep *DNMT1*

Nucleotide 1 corresponds to the first nucleotide of the *DNMT1* cDNA library screen clone (corresponding to nucleotide 1185 of the human *DNMT1* cDNA sequence). Primer RACE Y, highlighted in blue/purple (purple text indicates overlapping sequence with the RACE Z primer), is the gene-specific primer used to amplify the oocyte and somatic specific 5' ends of the sheep *DNMT1* after the modified (SMART) reverse transcription protocol. The same primer was used to detect DNMT1 products in the post-cDNA synthesis PCR. The primer sequence highlighted in green is the control forward primer RACE X. When RACE X is used in conjunction with RACE Z (highlighted in red/purple), a product of 103 bp is generated allowing detection of *DNMT1* transcripts within the RACE cDNA population.

3.2.3.1 Isolation of the sheep somatic *DNMT1* 5' cDNA sequence

The RACE Y primer was used in the reverse transcription of total RNA isolated from heart tissue (as described **2.4.11.1**) or pooled oocytes (as described **2.4.11.2**) following the SMART RACE protocol (as described **2.4.8**). After synthesis of modified heart and oocyte cDNA, PCR was performed using the RACE Y primer in conjunction with the Universal primer mix (provided in SMART RACE kit), containing a primer specific to the additional 'SMART II oligo' sequence at the 5' end of the cDNA, with conditions described in **2.4.8.** PCR products were electrophoresed on a 1.5% ethidium bromide stained agarose gel (as described **2.4.2**) and PCR products of approximately 1.3 Kb and 0.7 Kb were detected for heart cDNA (designated H1.3). A PCR product of 1.3 Kb was also detected for oocyte cDNA, alongside two smaller products of 0.7 Kb and 0.5 Kb (designated Ooc1.3, Ooc0.7 and Ooc0.5 respectively). The results are shown in **Figure 3.7**. A second PCR was performed on the PCR products as an additional confirmation, using RACE Y and RACE X primers, as described in **2.4.1.2.** PCR products were electrophoresed on a 2% ethidium bromide stained gel as described in **2.4.2**, the results are shown in **Figure 3.8**. PCR products H1.3, Ooc1.3 and Ooc0.7 were then subcloned and sequenced (as described in **2.4.3**).

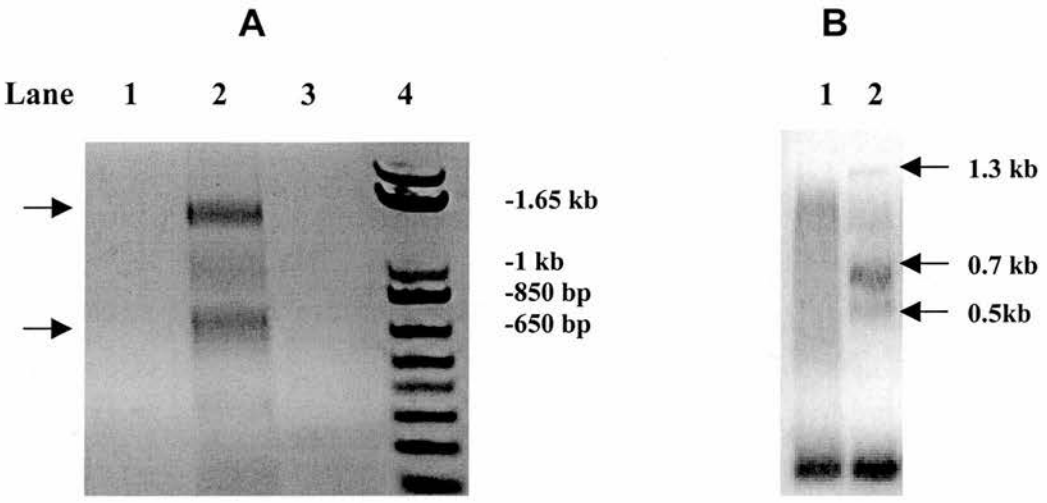


Figure 3.7 RACE PCR products amplified from heart and oocyte cDNA

(A) Lane 1 contains a negative (no template) PCR control. Lane 2 contains RACE PCR products of 1.3 Kb and 0.7Kb (indicated by arrows) generated from modified heart cDNA, Lane 3 demonstrates a failure to amplify *DNMT1*-specific RACE PCR products from modified oocyte cDNA. DNA marker sizes (lane 4) are indicated to the right of the gel image.

(B) The PCR was repeated for oocyte cDNA, using a higher cycle number. Lane 1 contains a negative (no template) PCR control. Lane 2 contains the three RACE PCR products (1.3 Kb, 0.7 Kb and 0.5 Kb) generated from modified oocyte cDNA. RACE PCR product sizes are indicated to the right of the gel image.

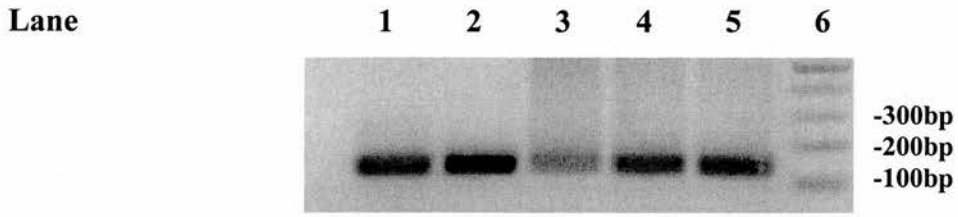
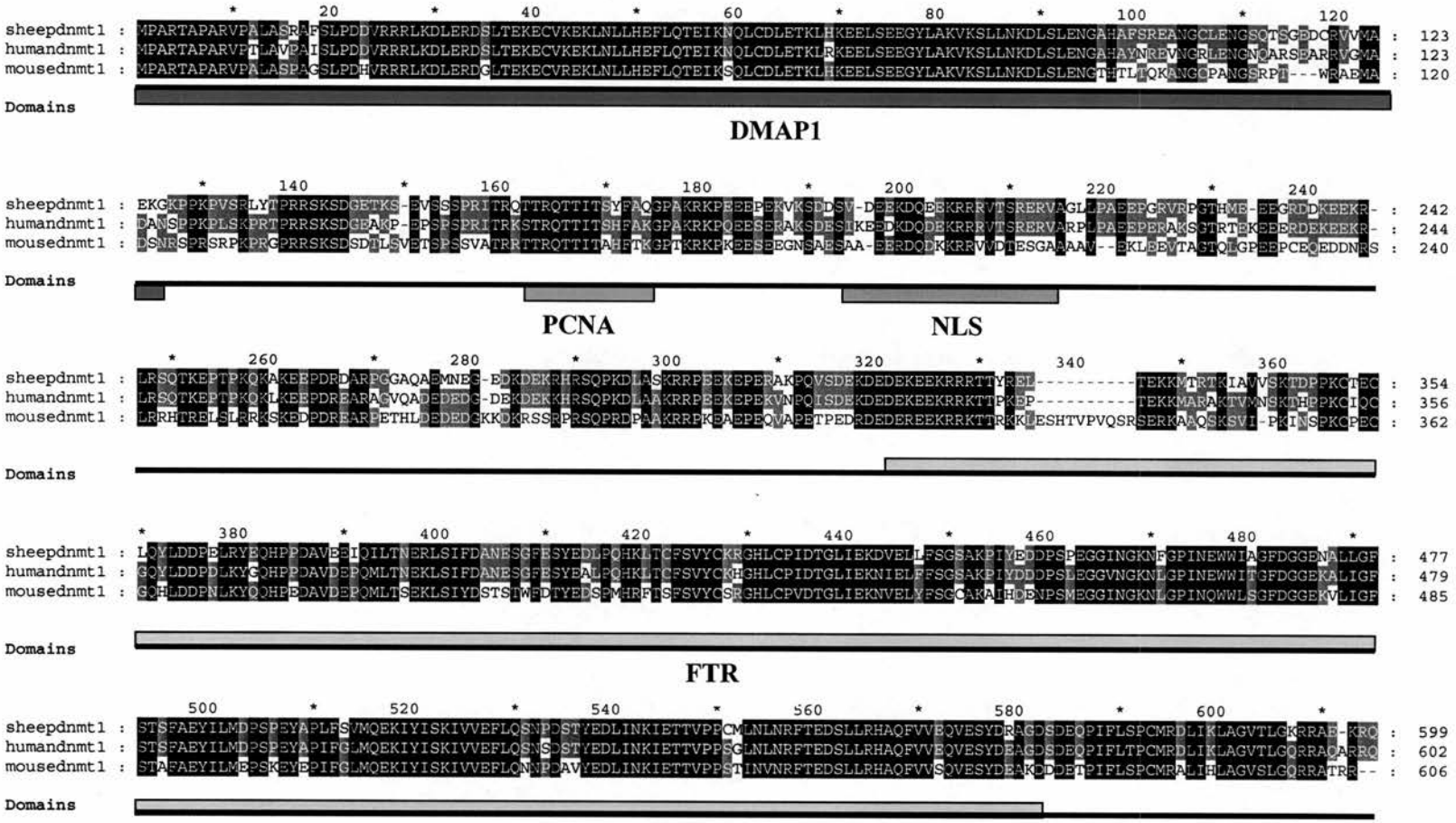
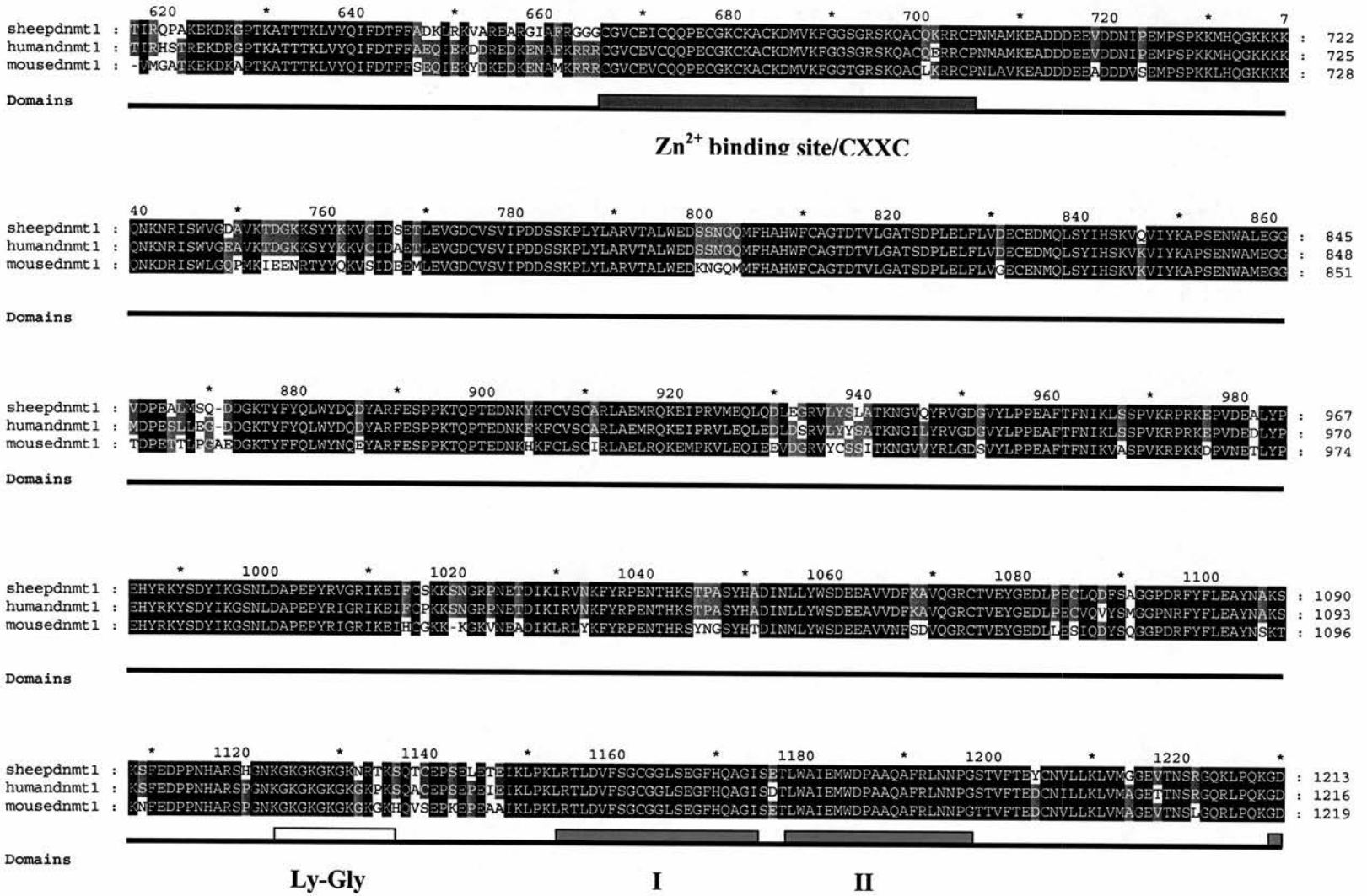


Figure 3.8 PCR analysis of products isolated by 5' RACE using internal primers. Lanes 2-5 contain PCR products amplified from RACE PCR products H1.3, Ooc1.3, Ooc0.7 and Ooc0.5 respectively. Lane 1 is a positive control PCR amplified from kidney cDNA. DNA marker sizes (lane 6) are indicated to the right of the gel image.

The sequencing results confirmed all PCR products isolated and subcloned after the 5' RACE experiment were *DNMT1* cDNA sequences. After sequence editing, an extra 990 bp of 5' *DNMT1* cDNA sequence were found to have been cloned by this method. The 990 bp H1.3 PCR product sequence was joined to the existing sheep *DNMT1* cDNA sequence, giving a total of 5233bp. The complete edited sheep *DNMT1* sequence can be found in **Appendix I**. Pairwise alignments of the entire cloned sheep cDNA sequence with the human and mouse *DNMT1* cDNA sequences were performed using the 'Gap' application (default parameters) of the Wisconsin Package Version 10.2 (GCG10). The sequence identity between the sheep and human *DNMT1* cDNA was found to be relatively high overall (86%), but slightly lower when comparing the sheep and mouse cDNA sequences (75% homology), with the 3' portion of the sequence being the most highly conserved between species. The pairwise sequence comparison of sheep/human and sheep/mouse *DNMT1* cDNA sequences can be found in the **Appendix I**. The complete sheep *DNMT1* cDNA clone sequence was translated using the 'Translate' application (default parameters) of GCG10, and an open-reading frame was identified in frame 1. The open reading frame started at nucleotide 54 and terminated at a stop codon (TAG) located at nucleotides 4887-4889. A multiple sequence alignment was performed using the 'localpileup' program (default parameters) on GCG10, comparing the sheep *DNMT1* protein with the known human (Accession No. NP001370) and mouse (Accession No. NP034196) *DNMT1* protein sequences. Shading was added using the Genedoc program (default settings). **Figure 3.9** shows the multiple alignment of the protein sequences. Analysis of the sheep *DNMT1* protein sequence demonstrated the presence of highly conserved functional domains including a cysteine-rich domain, proliferating cell nuclear antigen (PCNA) binding site, the replication foci targeting sequence, and the methyltransferase motifs (catalytic domain).


sheepdnmt1
humandnmt1
mousednmt1
Domains
DMAP1
PCNA
NLS
FTR


sheepdnmt1
humandnmt1
mousednmt1
Domains
Zn2+ binding site/CXXC
Ly-Gly
I
II

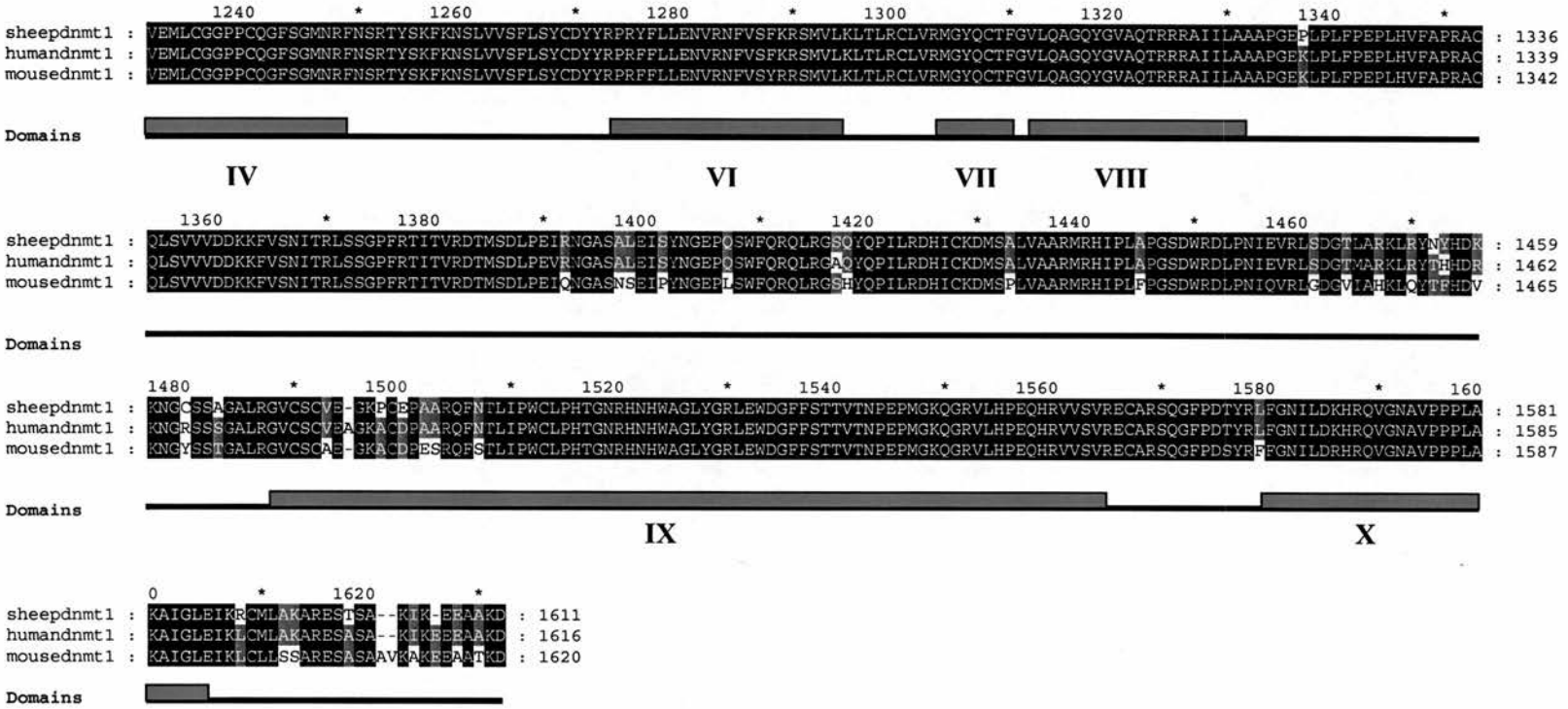


Figure 3.9 Multiple alignment of the sheep, human and mouse, DNMT1 protein sequences. Multiple sequence alignment was performed using the localpileup application on GCG10 and shaded using the default settings of the Genedoc program. The functional domains are highlighted on the line below the sequence alignment, and are identified with the domain name: DNMT1-associated protein 1 (DMAP1) binding site (blue); PCNA binding site (orange) Nuclear localization signal (NLS; purple), Fork-targeting region (FTR; light blue), Zinc binding domain containing the CXXC region (green), Lysine-Glycine rich region (white), and the methyltransferase motifs I, II, IV, VI, VII and VIII (all red).

3.2.3.2 Analysis of sheep oocyte *DNMT1* cDNA sequences

Comparison of cDNA sequences containing predicted exons 1-4 of the H1.3 (somatic) and Ooc1.3 (oocyte) sequences, with known exons 1-4 of mouse somatic and oocyte-specific *DNMT1*, was performed by multiple sequence alignment (Localpileup; GCG10). The multiple sequence alignment is demonstrated in **Figure 3.10** demonstrate that the heart and oocyte *DNMT1* sequences shared high levels of homology with the mouse somatic form of *DNMT1*. However, the *DNMT1* sequence derived from sheep oocytes demonstrated no similarity whatsoever with the alternative exon 1 of the mouse oocyte-specific form.

After sequence editing, the 0.7 Kb PCR product derived from oocyte cDNA (Ooc0.7) appeared to be a truncated form, with identical sequence to nucleotides 735-1232 of the edited sheep *DNMT1* sequence. A similar size product was also observed after RACE PCR on heart cDNA, suggesting the presence of truncated cDNAs possibly due to the inappropriate binding of the Smart II oligonucleotide primer during PCR steps. In support of this argument, the 0.7Kb PCR product amplified in heart cDNA disappeared when PCR conditions were made more stringent.

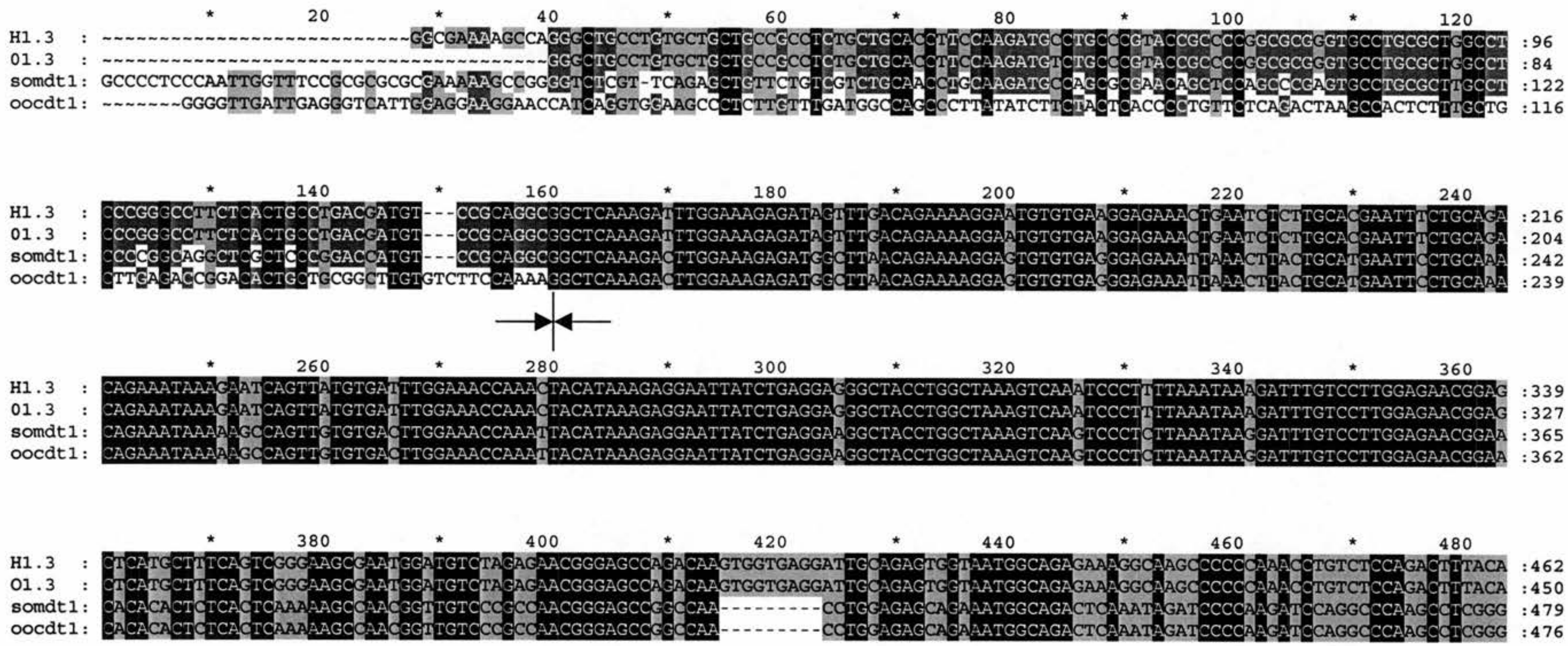


Figure 3.10 Multiple alignment of 5' sheep and mouse DNMT1 cDNA sequences. H1.3 and O1.3 are the sheep heart and oocyte 5' DNMT1 cDNA regions respectively. somdt1 is the mouse somatic form of DNMT1; oocdt1 is the mouse oocyte-specific form of DNMT1. Multiple alignment was performed using localpileup (GCG10) and shading added using default settings of GeneDoc. The two facing arrows denotes the exon1/exon2 boundary.

3.2.4 Investigation of *DNMT1* isoforms in sheep and mouse oocytes.

To confirm the presence of the somatic *DNMT1* in sheep oocytes, and to rule out possible contamination of oocyte RNA prior to RACE experiments, a PCR-based assay was designed. A method that would allow specific exon-detection was designed by using primers within exon 1, (the exon which differs in mouse somatic and oocyte-specific forms of *DNMT1* (primer SOMDT1)), exon 2, (which should amplify all forms of *DNMT1* (primer SHE2)) and a reverse primer in exon 3 (primer SOMDT2). Primer SOMDT2 was designed for use with both SOMDT1 and SHE2 primers, with all primers sharing a similar annealing temperature. **Figure 3.11** demonstrates the predicted exon boundaries within the 5' region of the sheep *DNMT1* cDNA sequence, and exon-specific primer sites. In parallel, a similar experiment was designed for the mouse *DNMT1* sequence. Primers were designed for somatic and oocyte-specific exon 1 (MSE1 and MOE1 respectively) and a reverse primer (MSER) was also designed within exon 3. This was to confirm previous data demonstrating oocyte-specific DNMT1 is the only form of DNMT1 in mouse oocytes and preimplantation embryos (Mertinet *et al.*, 1998).

To determine whether sheep oocytes contained the same form of *DNMT1* as somatic tissues, a PCR was performed using either SOMDT1/SOMDT2 or SHE2/SOMDT2 primer pairs with pooled oocyte cDNA as the template (as described **2.4.1.2**). The SHE2/SOMDT1 primer pair, which in theory detects all forms of *DNMT1*, was used as a positive control. A second PCR was also performed using MOE1/MSER and MSE1/MSER primer pairs, using pooled mouse oocyte cDNA as template (as described **2.4.1.1**). All PCR products were subsequently electrophoresed on a 2% ethidium bromide stained agarose gel (as described **2.4.2**), results are shown in **Figures 3.12** and **3.13.** The intensity of the SOMDT1/SOMDT2 PCR product demonstrated considerable amounts of somatic *DNMT1* transcripts in sheep oocytes, confirming the previous observation of identical forms of *DNMT1* in oocytes and heart tissue. The second PCR, analysing specific exon presence in mouse oocyte cDNA demonstrated no detectable somatic *DNMT1*, with the oocyte-specific form being the only form present.

All *DNMT1* PCR products generated in this particular experiment were digested with a sequence-specific restriction enzyme (*Apo* I), to confirm identity (as described **2.4.5**). Furthermore, the sheep-specific PCR primers failed to amplify the correct size PCR products when using genomic DNA as template, indicating the presence of intron-exon boundaries.

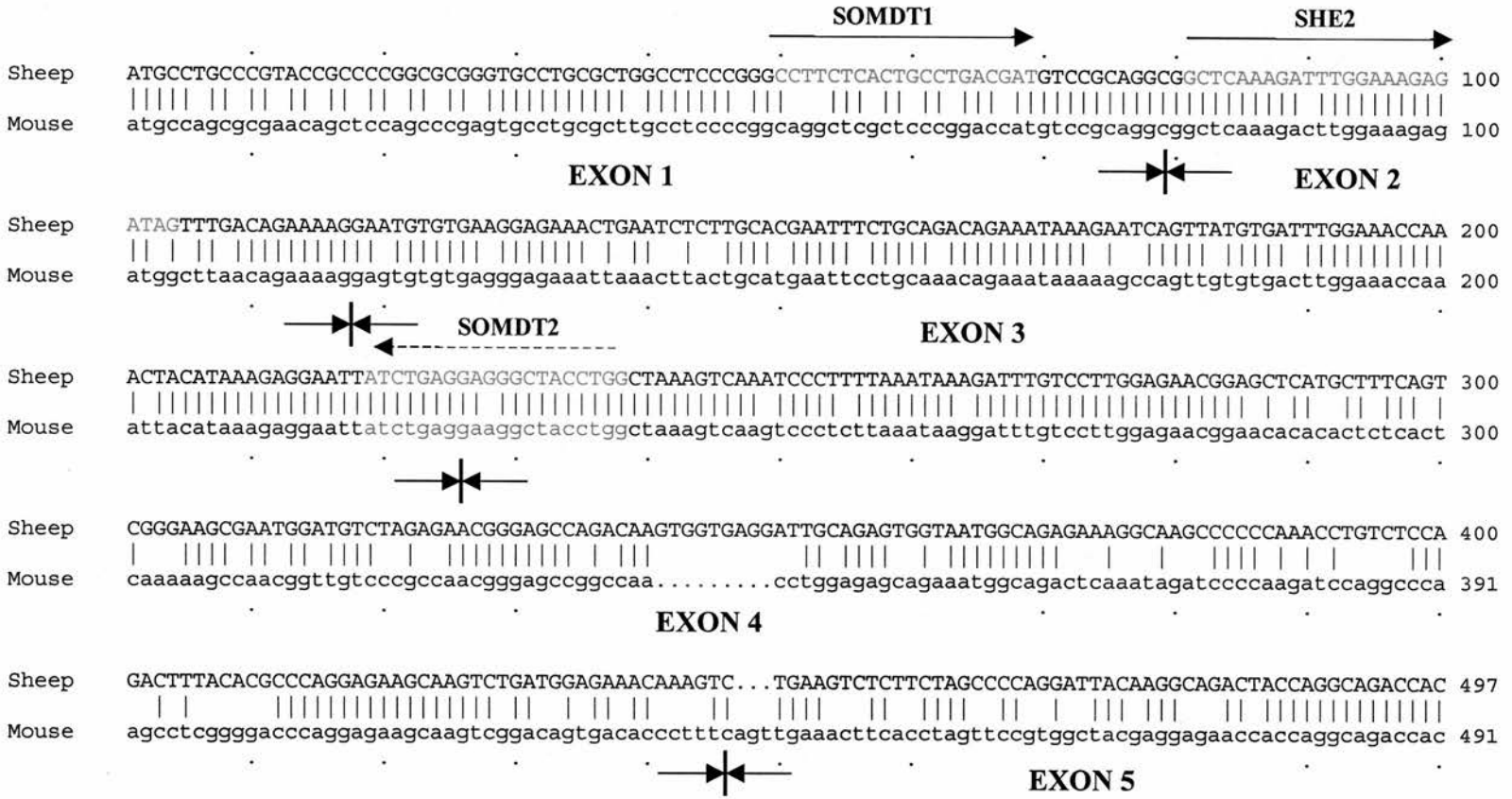


Figure 3.11 Predicted exon boundaries of sheep *DNMT1* and somatic *DNMT1* PCR primer sites.
Known exon boundaries of the mouse *Dnmt1* are indicated by facing arrows, and the sequence aligned with the sheep *DNMT1* sequence. Primer sites used to confirm the presence of somatic *DNMT1* in sheep oocytes are denoted by text highlighted in red and labelled dotted line arrows. Primers designed to detect somatic *Dnmt1* in mouse oocytes are denoted by blue text and labelled dotted line arrows. SOMDT1 and MSE1 are the sheep and mouse exon 1 specific primers and SHE2 is the sheep exon 2 specific primer. SOMDT2 and MSER are the sheep and mouse-specific reverse primers located at the exon3/exon 4 boundary.

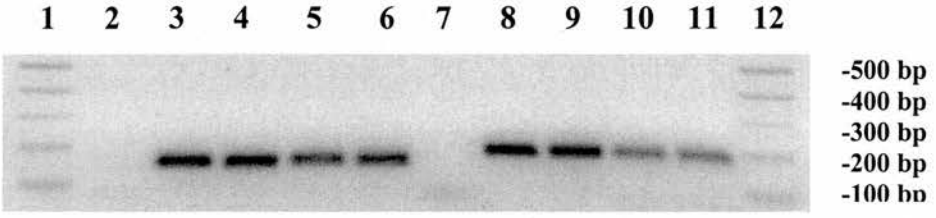


Figure 3.12 Detetection of *DNMT1* in sheep oocytes and heart tissue using exon-specific primers.

Lanes 2-6 represent PCR products generated using SHE2/SOMDT2 primers, lanes 7-11 contain PCR products generated using SOMDT1/SOMDT2 primers. Lanes 3-4 and 8-9 are PCR products amplified from heart cDNA. Lanes 5-6 and 10-11 are PCR products amplified from pooled oocyte cDNA. Lanes 2 and 7 are negative (no template) PCR controls. DNA marker sizes (1 Kb plus DNA ladder, Invitrogen) are indicated to the right of the gel image.

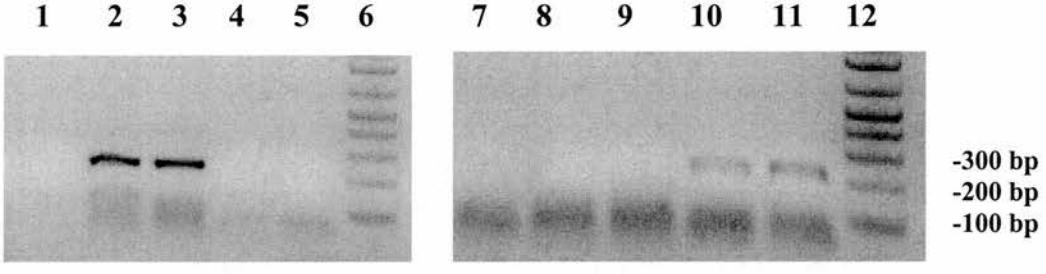


Figure 3.13 Detetection of oocyte- or somatic-specific *DNMT1* in mouse oocytes and heart tissue.

Lanes 1-5 show PCR products generated using the MSE1/MSER primer combination, and lanes 6-10 show PCR products generated using the MOE1/MSER primer pair. Lanes 1/7 are negative controls, lanes 2-3/8-9 are products generated from heart tissue, lanes 4-5/10-11 contain PCR products generated from oocyte cDNA. DNA marker sizes (lanes 6 and 12) are indicated to the right of the gel image.

3.4 Discussion

A 4.2 Kb cDNA clone partially encoding sheep *DNMT1* was isolated from an adult sheep lung cDNA library. A further 1 Kb of cDNA sequence comprising the 5' region of the somatic form of sheep *DNMT1* was cloned using a 5' RACE technique. In total, 5233 bp of sheep *DNMT1* cDNA sequence was obtained. After sequence editing, pairwise alignments were generated comparing sheep *DNMT1* with the human and mouse *DNMT1* cDNA sequences. The sheep cDNA sequence shares greater homology with the human sequence (86%) than the mouse cDNA sequence (77%). Translation of the sheep *DNMT1* sequence from the first methionine residue (nucleotide position 54) identified a long open reading frame terminating with the stop codon (TAG) at nucleotide position 4887. This open reading frame potentially encodes a protein of 1611 amino acid residues with a predicted molecular weight of 182 kDa. In addition to the translated regions, the cDNA clone contained 53 bp of 5' UTR and 346 bp of 3' UTR sequence, including the polyA tail.

It is possible that further 5' UTR *DNMT1* cDNA sequences exist in the sheep, as the 5' RACE method used to clone these regions cannot guarantee attainment of complete cDNA sequences for several reasons. Firstly, to gain complete 5' UTR sequences, the RNA used for cDNA synthesis has to be of the highest quality. Standard procedures were strictly adhered to with respect to the collection and storage of tissues for RNA extraction, but even this cannot prevent some degradation of the RNA. Secondly, the RACE method itself may form slightly truncated products if the reverse transcriptase activity terminates prematurely (before the end of the RNA template). The additional sequence likely to be gained by further attempts at 5' RACE or other cloning methods would be negligible regions of further 5' UTR cDNA sequence, and in the context of this study, adequate *DNMT1* sequence information was already attained.

Multiple sequence alignment comparing the predicted sheep protein sequence with the known human and mouse protein sequences was performed. The predicted sheep DNMT1 protein shares high levels of homology with the human and mouse proteins, particularly within functional domains. The region demonstrating the highest levels of homology is the carboxy terminal, specifically the methyltransferase motifs. Eight of the ten methyltransferase motifs present in prokaryotic cytosine methyltransferases are highly conserved between the sheep, human and mouse, and display identical sequence homology. One of these highly conserved methyltransferase motifs, motif IV, contains the proline-cysteine dipeptide (the catalytic site). Although the N-terminal region of DNMT1 is not as highly conserved as the C-terminal region between species, high levels of homology are apparent within functional domains. These include the cysteine-rich (Zn binding) domain and the replication foci targeting domain. Protein binding sites for PCNA and DMAP1 also retain high levels of homology between the three mammalian species.

The predicted protein sequence for sheep DNMT1 is clearly highly homologous with human and mouse protein sequence, and would therefore be expected to perform similar functions. Analysis of the protein would be required to confirm this, including measurement of protein size and assessment of its activity *in vitro*. The predicted size of DNMT1 could be confirmed by western blot analysis, and assays can also be performed to investigate the specificity of methyltransferases for hemi-methylated or unmethylated substrates. However, it would be unlikely to discover completely different functions of sheep DNMT1 when it is so similar to other mammalian forms of DNMT1.

In addition to isolation of the somatic form of sheep *DNMT1*, parallel attempts were made to clone an oocyte-specific *DNMT1*. The oocyte-specific form of DNMT1 identified in the mouse is identical to the somatic form, except for an alternative first exon (exon 1o). Using 5' RACE on pooled oocyte RNA, a 1 Kb cDNA was amplified corresponding to the 5' region of the *DNMT1* cDNA sequence. However, closer inspection of the cDNA sequence demonstrated 100% homology with the 5' region

amplified from somatic tissue. To rule out the possibility of contamination of the pooled oocyte RNA source, an exon-specific PCR was designed. Primer sites were designed in exon 1 (to amplify somatic form only), exon 2 (to amplify ALL forms of *DNMT1*) and a reverse primer in exon 3. The exon-specific PCR performed on pooled oocyte cDNA confirmed the presence of 'somatic' *DNMT1* in sheep oocytes. A similar exon-specific PCR was designed to amplify the mouse *Dnmt1* as previous experiments confirming the existence of the mouse oocyte-specific Dnmt1 (Mertineit *et al.*, 1998) did not specifically analyse Dnmt1 isoforms in mature oocyte RNA. The PCR showed that there was no somatic form of *Dnmt1* in pooled mouse mature oocytes.

The species-specific differences observed here are intriguing. Although the existence of an sheep oocyte-specific *DNMT1* cannot be ruled out, it is clear that it is not the sole form of DNMT1 found in sheep oocytes. Additional attempts were made to clone an 'oocyte-specific' *DNMT1* in sheep using the primers used to amplify the mouse form. The reverse primer of this pair is highly homologous with the sheep *DNMT1* sequence in exon 2. However, even when conditions were made less stringent, such as reducing primer annealing temperature and altering the magnesium concentrations, no PCR product was observed in sheep oocytes. Attempts to quantify the amount of somatic-specific *DNMT1* in sheep oocytes were difficult on account of the differences observed in primer efficiencies and so were not continued. Moreover, it was decided it would be more informative to analyse presence or absence of DNMT1o at the protein level. Unfortunately, time limitations prevented this particular experiment from being carried out.

The significance of the differences observed between the two species is unclear. The mouse is the only known species expressing DNMT1o, however, there are no reports of investigations performed in other vertebrate species except Xenopus. Xenopus does not appear to have an oocyte-specific form of DNMT1, with significant amounts of somatic DNMT1 observed in Xenopus eggs. The apparent lack of DNMT1o in Xenopus may not be significant, as this species does not have imprinted genes; the subset of genes

methylated by Dnmt1o in the mouse at the 8-cell stage of development (Howell *et al.*, 2001). In contrast, evidence for genomic imprinting in the sheep is strong, and allele-specific gene expression of *IGF2*, *IGF2R*, *MEST* (Feil *et al.*, 1998) *H19* (Hagemann *et al.*, 1998) *DLK1*, *GTL2*, *DAT*, *PEG11*, *antiPEG11* and *MEG8* (Charlier *et al.*, 2001) have been reported. In addition, a more detailed study of sheep *IGF2* imprinting was recently undertaken, demonstrating 'imprinting' characteristics (developmental switch to biallelic expression in the liver; sex-specific differences in recombination rates) shared with the human *IGF2* gene (McLaren & Montgomery, 1999).

Of note, a database search was initiated to investigate the presence of an oocyte-specific exon in other mammalian species known to have imprinted genes, to identify sequences sharing homology with the oocyte-specific exon 1o. Although the entire human chromosome 19 (the location of the human *DNMT1* gene) is available for nucleotide-based searches, no sequence demonstrating homology with the mouse exon 1o was identified. It is possible, however, that any exon 1o sequences may differ significantly between species, or an alternative method of regulation may exist.

If the sheep does not express an oocyte-specific *DNMT1*, then it is unclear as to what the effect may be on imprinted gene methylation during preimplantation development. The unusual localisation and nuclear import/export of mouse DNMT1o may be essential for maintenance of imprinted gene methylation in the mouse, but is this the case in other mammalian species? One fundamental question needs addressing prior to drawing conclusions. If sheep do not express an oocyte-specific *DNMT1*, is the somatic form of DNMT1 able to perform a similar role? The mouse Dnmt1o is a truncated protein, lacking the first 3 exons of the somatic form. This difference in N-terminal regions is clearly important for the specialised properties of oocyte and somatic forms of Dnmt1.

Also worthy of note, is a study in which the truncated form of Dnmt1 protein was isolated in a screen for proteins binding annexin V (Ohsawa *et al.*, 1996). This protein has been postulated to be involved in the sequestration of Dnmt1o at the periphery of the

cytoplasm in mature oocytes (Mertineit *et al.*, 1998). In addition, annexin V has also been shown to undergo inducible nuclear localisation (Barwise and Walker, 1996). As the N-terminal of somatic Dnmt1 is extended in comparison with Dnmt1o, binding with annexin V may be prevented. Indeed, DNMT1o has recently been shown to colocalise with Annexin V in mouse oocytes (Doherty *et al.*, 2002). Another protein whose binding may be altered by the alternative N-terminal regions of Dnmt1o is DMAP1. The DMAP1 binding site is located within the N-terminal region of somatic Dnmt1 (amino acid residues 1-130), and may be prevented in the truncated Dnmt1o if a significant section of the binding site is absent.

Unfortunately, it appears that the 'oocyte specificity' of the truncated Dnmt1 protein is not as clear-cut as previously suggested. The Dnmt1 protein shown to bind Annexin V in the initial screening study was a truncated form (similar to Dnmt1o) found in the brain of a d17 rat (Ohsawa *et al.*, 1996; cDNA sequence accession no. D64060). A search of the GenBank database also yielded a second rat *Dnmt1* cDNA sequence isolated from rat brain, encoding the same truncated form of Dnmt1 (Accession No. AF116345). The two isoforms identified in the rat brain were not identical, with different sequences 5' prime of exon 4. Multiple sequence alignment generated from the two rat brain *Dnmt1* cDNA sequences alongside the oocyte and somatic forms of mouse *Dnmt1* demonstrated no homology between the rat Dnmt1 sequences and exon 1o of the oocyte-specific form of *Dnmt1*. Comparison of the rat *Dnmt1* sequences with the somatic form of *Dnmt1* demonstrated homology commencing from either exon 2 or exon 4. The sequence alignment is presented in **Figure 3.14**. As the predicted protein sequences for both of these isoforms initiate at the same ATG codon as the 'oocyte-specific' Dnmt1, it is likely that this truncated form may be generated by several possible routes of alternate splicing and not solely the result of exon 1o usage.

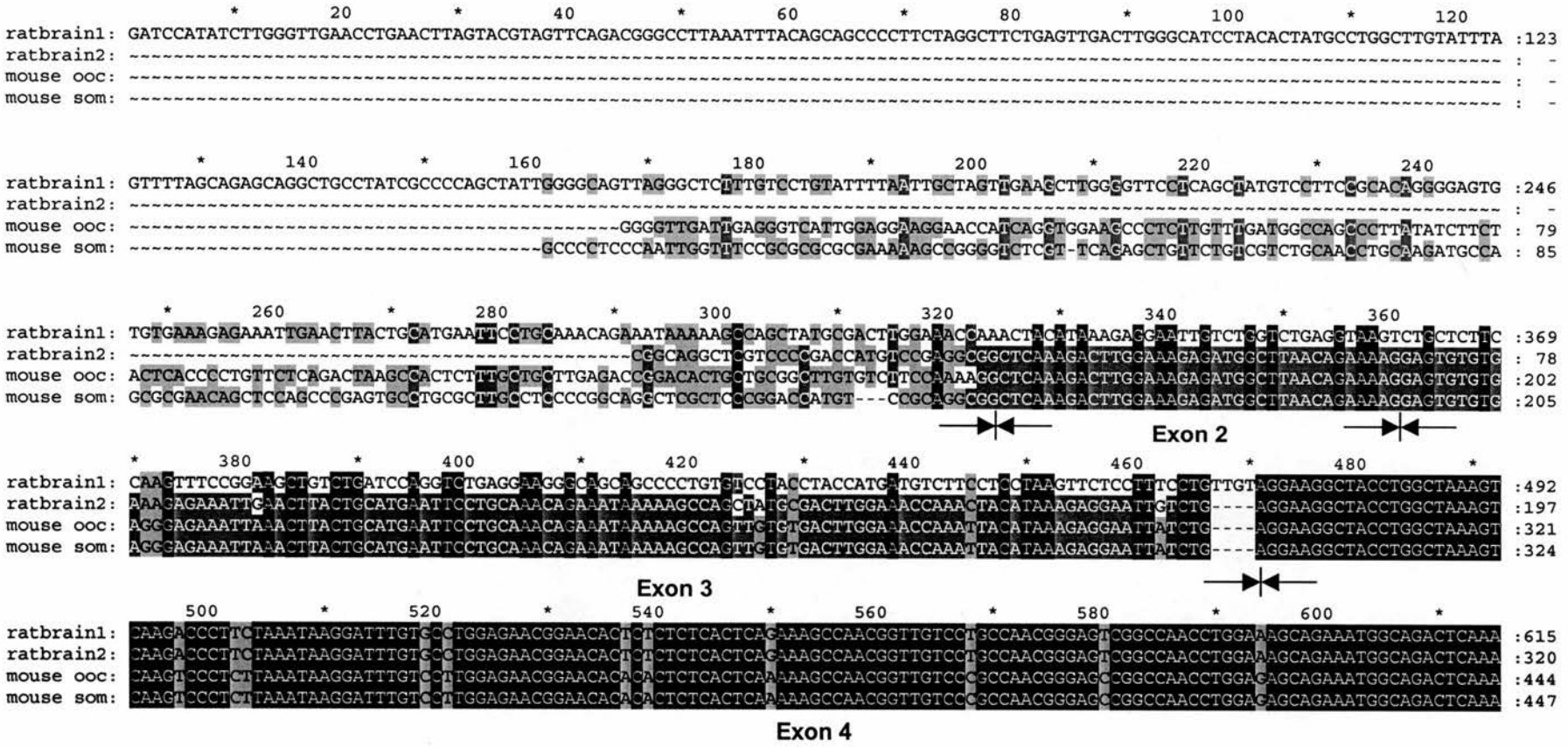


Figure 3.14 Multiple sequence alignment of rat and mouse *Dnmt1* isoforms. The two rat *Dnmt1* isoforms identified in brain are indicated as ratbrain1 (Accession No. AF116345) and ratbrain2 (Accession No. D64060). The mouse oocyte and somatic forms of *Dnmt1* are indicated as 'ooc' and 'som' respectively. Facing arrows indicate exon boundaries.

As the N-terminal regions of the oocyte-specific Dnmt1o appear to encode unique localisation and specific methylating properties, and somatic DNMT1 is the abundant form of DNMT1 in sheep oocytes, then the cytoplasmic localisation observed in mouse oocytes and preimplantation embryos may not be replicated. Instead, localisation may be observed throughout the cytoplasm and nucleus. The consequences of nuclear localisation in terms of methylation reprogramming are significant.

In the mouse, exclusion of Dnmt1 from the nucleus during preimplantation development correlates with the passive demethylation of the genome observed at this stage. If the nuclei of sheep oocytes (and consequently preimplantation embryos) contain significant amounts of DNA methyltransferase activity, demethylation may not be observed to such an extent, if at all, in sheep preimplantation development. Furthermore, the presence of DNMT1 in the nucleus during preimplantation development would potentially allow for the continual maintenance of imprinted gene methylation, eliminating the need for a specialised DNMT1 with stage-specific localisation abilities. This would represent very different reprogramming mechanisms occurring in two mammalian species.

Methylation reprogramming has been investigated to a small extent in the cow (*Bos taurus*; a ruminant species closely related to sheep) and both active and passive demethylation occurs post-fertilisation (Dean *et al.*, 2001). In contrast to mice, however, *de novo* methylation occurs at the 8-16 cell stage, demonstrating significant differences in reprogramming between species. It would be very interesting to establish whether the cow also expresses the somatic form of *DNMT1* in oocytes, or whether the mouse is unique in its unusual method of DNA methylation regulation in embryos.

Still, the possibility of an oocyte-specific DNMT1 in sheep cannot be ruled out. It must be remembered that so far, only the somatic *DNMT1* transcripts have been detected in sheep oocytes, and it may be possible that they are not translated. However, the presence of somatic *DNMT1* in sheep oocytes demonstrates significant species-specific differences with further implications for methylation reprogramming in preimplantation

development. Further work is clearly required to determine the specific localisation of somatic (or oocyte) specific forms of DNMT1 in sheep oocytes and preimplantation embryos.

Overall, the cloning of *DNMT1* in sheep reported here, will allow further analysis of the functions of this protein in sheep. Expression analysis of *DNMT1* in adult tissues and preimplantation embryos will allow insights into the roles *DNMT1* of in sheep development, and is described in chapter 5. Furthermore, *in situ* analysis of *DNMT1* mRNA and immunolocalisation of the protein in preimplantation embryos could also be used to explore the role of this gene in sheep development.

Chapter 4

Cloning and sequence analysis of the sheep *de novo* methyltransferases *DNMT3A* and *DNMT3B*

4.1 Introduction

De novo methylation is the modification of completely unmethylated DNA and is thought to be the process involved in establishing genomic DNA methylation patterns during early development. The first mammalian DNA cytosine-5 methyltransferase to be identified was Dnmt1, a protein predominantly involved in maintaining methylation patterns, although it does encode some residual *de novo* methylating activity (Lei *et al.*, 1996). The existence of separately encoded DNA methyltransferases solely involved in establishing methylation patterns, was first realised after observations of persistent *de novo* methylating activity in ES cells carrying a targeted deletion of *Dnmt1* (Lei *et al.*, 1996). The race to clone these novel DNA methyltransferase genes led to the recent identification of *Dnmt2*, *Dnmt3a* and *Dnmt3b* (Yoder *et al.*, 1998; Okano *et al.*, 1998a); all encoding proteins homologous with bacterial cytosine-5 methyltransferases. Dnmt2, a protein related to pmt1p of fission yeast (*S.pombe*), was almost instantly ruled out as a *de novo* methyltransferase. Firstly, the protein failed to demonstrate any CpG methylating activity *in vitro*, and secondly, ES cells carrying a disrupted *Dnmt2* did not show any alterations in global methylation levels (Okano *et al.*, 1998b).

Dnmt3a and Dnmt3b are thought to be more realistic candidates for mammalian *de novo* methyltransferases for several reasons. Firstly, *in vitro* expression of Dnmt3a and Dnmt3b demonstrated methylating activity, and furthermore, demonstrated no preference for hemi-methylated substrates (Okano *et al.*, 1998a). In addition, *Dnmt3a* and *Dnmt3b* are also highly expressed in undifferentiated ES cells, but downregulated in adult tissues, suggesting developmental roles (Okano *et al.*, 1998a). Dnmt3a is a protein of 908 amino acids, encoded by a 4192 bp transcript

(Okano et al., 1998a). The *Dnmt3b* cDNA sequence is of a similar size (4195 bp), but is alternatively spliced to produce proteins of 859 aa (Dnmt3b1), 840 aa (Dnmt3b2) and 777 aa (Dnmt3b3; Okano et al., 1998a).

The Dnmt3 family of proteins share a very similar C-terminal domain, along with a cysteine-rich region. The C-terminal regions of Dnmt3a and Dnmt3b are not closely related to the one seen in Dnmt1, suggesting different prokaryotic 5-methyl cytosine ancestors (Xie *et al.*, 1999). Furthermore, the cysteine rich region does not share homology with the Dnmt1 cysteine-rich region (Zn-binding domain), but is very similar to a domain found in the ATRX protein, known to be a member of the SNF2/SWI family of chromatin remodelling proteins (Picketts *et al.*, 1996). The Dnmt3a and Dnmt3b N-terminal regions demonstrate significant differences (less than 30% sequence homology) between the two proteins (Xie *et al.*, 1999). The N-terminal regions may encode different roles for the two proteins, possibly mediated by different protein-protein or protein-DNA interactions. **Figure 4.1** highlights the structural differences between the two proteins.

Genetic analysis of *Dnmt3a* and *Dnmt3b* has demonstrated essential roles for these proteins in development. Mice carrying a targeted deletion of *Dnmt3b* die at mid-gestation whilst mice carrying a targeted deletion of *Dnmt3a* die shortly after birth (Okano *et al.*, 1999). Furthermore, the specific roles of Dnmt3a and Dnmt3b in development have also been investigated. Analysis of the mice carrying a targeted deletion of *Dnmt3b* demonstrated a failure to methylate centromeric satellite minor repeats (Okano *et al.*, 1999), and a high percentage of humans suffering with the genetic disorder ICF (Immunodeficiency, Centromeric instability, and Facial anomalies) carry mutations in the *DNMT3B* gene (Hansen *et al.*, 1999; Xu *et al.*, 2000). This suggests a role for Dnmt3b in the *de novo* methylation of specific repeat sequences.

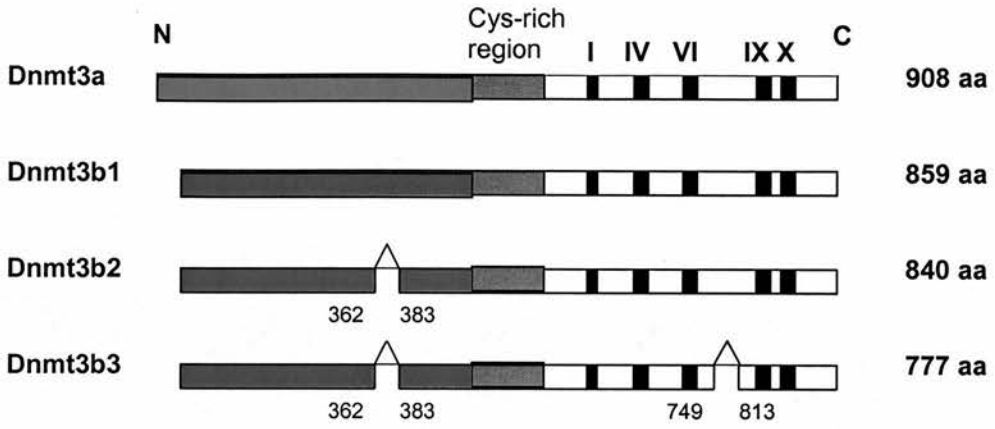


Figure 4.1 Structural comparison of Dnmt3a and Dnmt3b protein domains. Adapted from Okano *et al.*, 1998a. The cysteine rich region of Dnmt3a and Dnmt3b proteins are highlighted in grey, the five most highly conserved methyltransferase motifs are represented by black rectangles (numbered). Splice sites in Dnmt3b2 (amino acids 362 to 383) and Dnmt3b3 (amino acids 362 to 383 and 749 to 813) are also indicated in the diagram. The N-terminal regions of Dnmt3a and Dnmt3b are highlighted in red (Dnmt3a) and blue (Dnmt3b) to demonstrate the differences in protein sequence.

In addition to its known role in CpG methylation, there is increasing evidence suggesting a role for Dnmt3a in the methylation of non-CpG cytosine residues (predominantly CpA dinucleotides) during mouse development. Non-CpG methylation has been observed in ES cells (Ramsahoye *et al.*, 2000), and has also been demonstrated in the mouse neurofibromatosis type 1 gene (*Nf1*) during post-fertilisation stages of development (Haines *et al.*, 2001). Dnmt3a is known to be responsible for the non-CpG methylation in ES cells (Ramsahoye *et al.*, 2000), but as there is no data available for *Dnmt3a* expression in mouse preimplantation embryos and oocytes, it is difficult to ascertain whether it may be involved in methylation of the *Nf1* or other genes.

A closely related protein, Dnmt3L, has recently been shown to be essential for the establishment of imprinted gene methylation in mouse oocytes (Bourc'his *et al.*, 2001). Dnmt3L is very similar to both Dnmt3a and Dnmt3b, containing the cysteine

rich domain shared with the ATRX protein, but it lacks the conserved methyltransferase motifs present in other members of the Dnmt3 family (Aapola *et al.*, 2000). Recent experiments investigating the precise role of the oocyte-specific Dnmt1 (Dnmt1o) have demonstrated that it is not responsible for establishing imprinted gene methylation in oocytes and therefore another methyltransferase must be involved. (Howell *et al.*, 2001) As Dnmt3L itself cannot methylate genes it is possible that it may cooperate with Dnmt3a or Dnmt3b (perhaps via interaction of the shared cysteine rich domain) to establish methylation marks of imprinted genes in mouse oocytes.

The importance of methylation reprogramming in mouse development, and the apparent requirement for *de novo* methyltransferases Dnmt3a and Dnmt3b in normal development is clear. As one of the main aims of this study is to investigate DNA methylation and methylating activities occurring during sheep preimplantation development, the *de novo* methyltransferase genes are key candidates for analysis. The expression patterns of these genes have not been studied in the mouse preimplantation embryo but their expression in ES cells, coupled with the high levels of non-CpG methylation observed in early stage mouse embryos, would suggest significant roles at this stage of development.

This study describes the initial cloning and sequence analysis of the sheep *DNMT3A* and *DNMT3B* genes. Coding regions of the *DNMT3A* and *DNMT3B* genes were cloned using a combination of RT-PCR and cDNA library screening. After verification of cDNA clone identity, sequence analysis was performed and comparisons made with the known human and mouse cDNA and protein sequences to assess levels of conservation between the three species.

4.2 Results

4.2.1 Generation of partial sheep *DNMT3A* and *DNMT3B* cDNA clones by RT-PCR

Primer sequences specific to *DNMT3A* were obtained from Robertson *et al.*, 1999. Pairwise sequence alignments of the human (Accession no. AF176228) and mouse (Accession no. NM_010068) *DNMT3B* cDNA sequence were generated using the 'Gap' application (default parameters) of GCG10. Primers were designed in a region of the human cDNA sequence seen to be highly conserved between the human and mouse and present in all Dnmt3b splice variants. **Table 4.1** lists the primers used to amplify *DNMT3A* and *DNMT3B* cDNA, and the locations within the human *DNMT3A* or *DNMT3B* cDNA sequence. Heart or liver cDNA derived from d125 fetal tissue was used as the template and a standard PCR reaction performed (as described **2.4.1.1**). In addition, PCR reactions using 500 ng genomic DNA as template were also performed, to determine whether the *DNMT3A* or *DNMT3B* primers spanned intronic regions. PCR products were electrophoresed on a 2% ethidium bromide stained agarose gel (as described **2.4.2**), results are shown in **Figure 4.2**. PCR products of expected size (*DNMT3A* 280 bp; *DNMT3B* 101 bp) were observed and subsequently cloned and sequenced (as described **2.4.3**) to confirm identity. Amplification of genomic DNA using the HDNMT3A primers produced a larger PCR product of 1.2 Kb, indicating the primers span an intronic region of approximately 0.9 Kb. The PCR product observed as a result of amplification of genomic DNA with the *DNMT3B2* primers was the same size as the RT-PCR product, indicating the primers do not span an intron.

Gene	*Primer*	*Direction of Primer*	*Genbank Accession No. (human cDNA sequence)*	*Location of primer*
DNMT3A	HDNMT3AF	Forward	AF067972	2282-2303
DNMT3A	HDNMT3AR	Reverse	AF067972	2561-2539
DNMT3B	DNMT3B2F	Forward	AF176228	2269-2290
DNMT3B	DNMT3B2F	Reverse	AF176228	2369-2350

Table 4.1 Table of *DNMT3A* and *DNMT3B* primer design information. Primer name, orientation and location within source cDNA sequence are indicated.

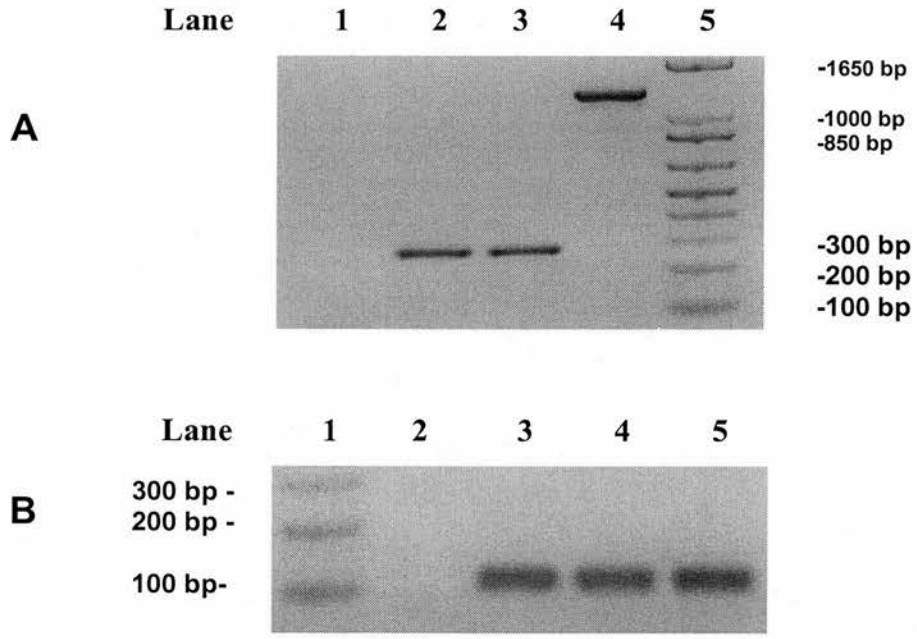


Figure 4.2 Partial sheep *DNMT3A* and *DNMT3B* cDNA clones generated by RT-PCR.

(A) Lane 1 is a negative PCR control with no template. Lanes 2 and 3 are PCR products amplified from d125 fetal heart and liver cDNA respectively. Lane 4 is a PCR product amplified from genomic DNA. DNA marker sizes (Lane 5; I Kb Plus DNA ladder (Invitrogen)) are indicated to the right of the gel image.

(B) Lane 2 is a negative PCR control with no template. Lanes 3 and 4 are PCR products amplified from d125 fetal heart and liver cDNA respectively. Lane 5 is a PCR product amplified from genomic DNA. DNA marker sizes (Lane 1; I Kb Plus DNA ladder (Invitrogen)) are indicated to the left of the gel image.

4.2.2 Isolation of a *DNMT3A* cDNA clone from an adult sheep lung cDNA library

The partial sheep *DNMT3A* cDNA clone generated by RT-PCR was used to probe an adult lung sheep cDNA library obtained from Stratagene (as described **2.4.7**). The primary screen identified a positive clone, designated 9, which was re-plated for the secondary screen. After the secondary screen, a positive signal corresponding to six single plaque colonies (designated 9a-9f respectively) were identified. These single plaque colonies were isolated and PCR analysis performed using the HDNMT3A primer set as an additional control. The results of this PCR are demonstrated in **Figure 4.3.** A single clone, 9D, was chosen for further analysis. Excision of the pBluescript plasmid (containing the cDNA fragment) from the Uni-ZAP® XR vector was subsequently performed (as described **2.4.7.7**) and plasmid DNA prepared from 8 single colonies for clone 9D (as described **2.4.3**). PCR analysis using the HDNMT3A primer set was performed to check insert identity of the plasmid DNA, and an *EcoR I/Xho* I double digest was also performed to release the insert and simultaneously ascertain insert size for lambda clone 9D, results are shown in **Figure 4.4**

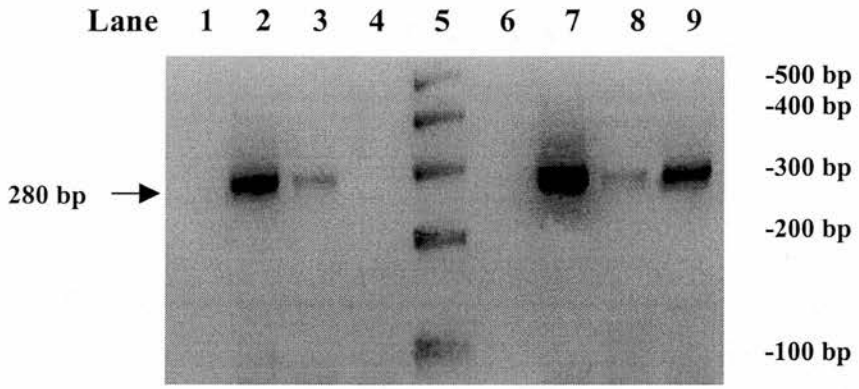


Figure 4.3 PCR analysis of *DNMT3A* positive plugs.

Lane 1 is a minus template control, Lane 2 is a positive PCR control using liver cDNA as template. Lanes 3-4 and 6-9 represent the PCR products amplified from 'positive' plugs 9a-9f respectively. DNA marker sizes (lane 5; 1 Kb Plus DNA ladder, Invitrogen) are indicated to the right of the gel image.

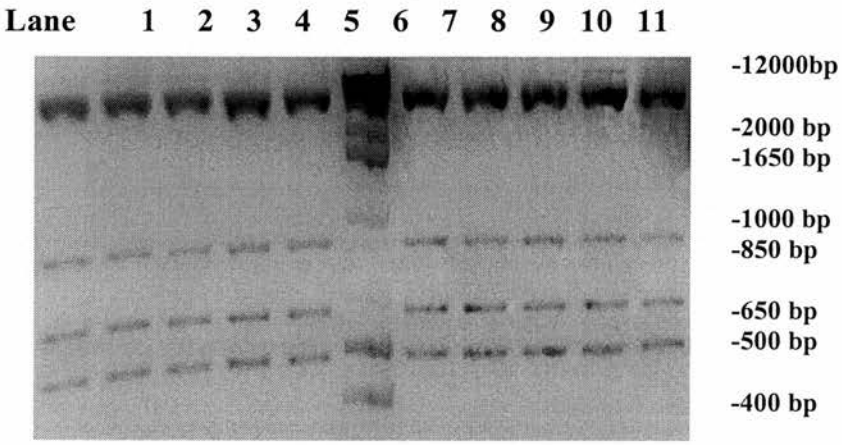


Figure 4.4 *EcoR I/Xho* I digestion of DNA minipreps established from Uni-ZAP® XR clone 9D.

Lanes 1-5 and 7-11 represent minipreps 1-8, DNA marker sizes (lane 6; 1 Kb Plus DNA ladder, Invitrogen) are indicated to the right of the gel image.

A single plasmid DNA miniprep originating from the 9D clone, with an insert of approximately 2 Kb, was ethanol precipitated and sequenced (as described **2.4.6**; **2.4.3**). A total of 1850 nucleotides of sequence corresponding to the sheep *DNMT3A* cDNA sequence were obtained and can be found in **Appendix II**. Pairwise alignments of the cloned sheep cDNA sequence and the human or mouse *DNMT3A* cDNA sequences were performed using the 'Gap' application (default parameters) on GCG10. The sheep *DNMT3A* cDNA sequence was found to be highly conserved (94%) when compared to the human homologue (Accession No. AF067692. The pairwise alignment of sheep and human or mouse *DNMT3A* cDNA sequences can be found in **Appendix II**. The sheep *DNMT3A* cDNA sequence was subsequently translated using the translate application in GCG10, and an open reading frame identified in frame 1. This open reading frame terminated at a TAA stop codon located at nucleotides 1427-1429. Multiple protein sequence alignment was performed as described in Chapter 3, comparing the human (Accession No. NP072046), mouse (Accession No. NP031898) and sheep protein sequences. **Figure 4.5** shows the high level of conservation of *DNMT3A* between species at the protein level.

```
                *       20        *       40        *       60        *       80        *      100        *      120
sheep  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
mouse  : ~~~MPSSGPGDTSSSSLEREDDRKEGEEQEENRGKEERQEPSATARKVGRPGRKRKHPPVESSDTPKDPAVTTKSQPMAQDSGPSDLLPNGDLEKRSEPQPEEGSPAAGQKGGAPAEGEG-TET : 120
human  : MPAMPSSGPGDTSSSAAEREEDRKDGEEQEEPRGKEERQEPSTTARKVGRPGRKRKHPPVESGDTPKDPAVISKSPSMAQDSGASELLPNGDLEKRSEPQPEEGSPAGGQKGGAPAEGEGAAET : 124
Domains

              *      140        *      160        *      180        *      200        *      220        *      240
sheep  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
mouse  : PPEASRAVENGCCVTKEGRGASAGEGKEQKQTNIESMKMEGSRGRLRGGLGWESSLRQRPMPRLTFQAGDPYYISKRKRDEWLARWKREAEKKAKVIAVMNAVEENQASGESQKVEEASPPAVQ : 244
human  : LPEASRAVENGCCTPKEGRGAPAEAGKEQKETNIESMKMEGSRGRLRGGLGWESSLRQRPMPRLTFQAGDPYYISKRKRDEWLARWKREAEKKAKVIAGMNAVEENQGPGESQKVEEASPPAVQ : 248
Domains

         *      260        *      280        *      300        *      320        *      340        *      360        *
sheep  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
mouse  : QPTDPASPTVATTPEPVGGDAGDKNATKAADDEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTRWVMWFGDGKFSVVCVEKLMPLSSFCSAFHQATYNKQPMYRKAIYEV : 368
human  : QPTDPASPTVATTPEPVGSDAGDKNATKAGDDEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTRWVMWFGDGKFSVVCVEKLMPLSSFCSAFHQATYNKQPMYRKAIYEV : 372
Domains

             380        *      400        *      420        *      440        *      460        *      480        *
sheep  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~DMWVEPEAAAYAPPPPAKKPRKSTTEKPKVKEIIDERTRERLVYEVRQKCRNIEDICIS :  59
mouse  : LQVASSRAGKLFPACHDSDESDSGKAVEVQNKQMIEWALGGFQPSGPKGLEPPEEEKNPYKEVYTDMWVEPEAAAYAPPPPAKKPRKSTTEKPKVKEIIDERTRERLVYEVRQKCRNIEDICIS : 492
human  : LQVASSRAGKLFPVCHDSDESDTAKAVEVQNKPMIEWALGGFQPSGPKGLEPPEEEKNPYKEVYTDMWVEPEAAAYAPPPPAKKPRKSTAEKPKVKEIIDERTRERLVYEVRQKCRNIEDICIS : 496
Domains
```

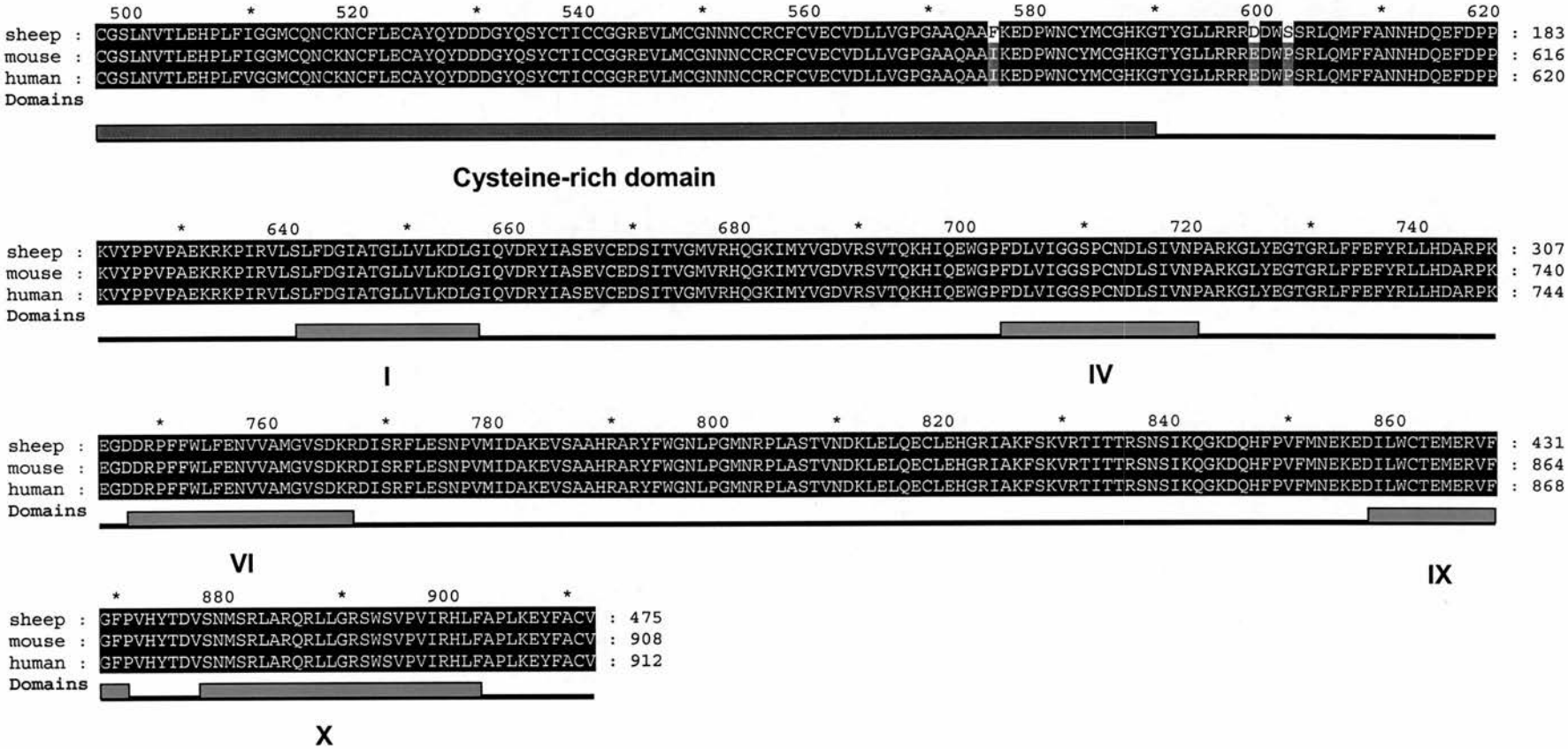


Figure 4.5 Multiple alignment of the sheep, human and mouse DNMT3A protein sequences. Multiple sequence alignment was performed using the localpileup application on GCG10 and shaded was added using the Genedoc program (default settings). The known functional domains are highlighted on the line below the sequence alignment, and are identified with the domain name; Cysteine rich domain (turquoise); highly conserved methyltransferase motifs (red)

4.2.3 Sequence analysis of a partial sheep *DNMT3B* cDNA clone.

The partial *DNMT3B* cDNA clone generated by RT-PCR (refer to **4.2.1**) was used as a probe to screen the adult lung cDNA library (as described **2.4.7**). Several attempts to isolate a larger *DNMT3B* cDNA clone using this strategy failed. As the failure may have been due to inefficiency of the probe (perhaps due to the small size of the probe), library screening was performed using the full-length mouse *Dnmt3b* cDNA (a kind gift from Dr I. Stancheva, University of Edinburgh). Again, several attempts using this method were unsuccessful. Poor representation of the *DNMT3B* cDNA in the library, perhaps due to low expression in the adult sheep lung, may have been one reason the library screen failed, therefore, an alternative strategy was devised. Additional primer pairs were designed over regions seen to be homologous in the human and mouse, however, attempts to amplify *DNMT3B* specific PCR products failed. Ultimately, it was decided to cease attempts to clone a larger partial *DNMT3B* cDNA clone due to time constraints.

The 101bp of *DNMT3B* cDNA sequence amplified by RT-PCR corresponded to nucleotides 2269-2369 human *DNMT3B* sequence (Accession No.AF176228). Due to the high levels of homology shared by *DNMT3A* and *DNMT3B* at the nucleotide level, it was necessary to confirm the identity of the partial sheep '*DNMT3B*' sequence by comparison with both genes. After removal of primer sequences, multiple sequence alignment of the remaining 59 bp of sheep *DNMT3B* sequence and the human *DNMT3B* and *DNMT3A* cDNA sequences (shown in **Figure 4.6**) was performed using the 'localpileup' application (default parameters) of GCG10. Pairwise alignment of the 59 bp sheep *DNMT3B* sequence with the human *DNMT3B* and *DNMT3A* cDNA sequences demonstrated 92% and 81% homology respectively, increasing confidence in the identification of the cDNA sequence as *DNMT3B*. The 59 bp of sheep *DNMT3B* cDNA sequence were subsequently translated using the 'Translate' application of GCG10. Pairwise sequence alignment of the predicted sheep DNMT3B protein sequence with the known human DNMT3B protein sequence (Accession No. NO008823) was performed using the 'Gap' application (default parameters) of GCG10. The protein sequences were highly conserved, with

31/32 amino acids being identical. The alignment of DNMT3B protein sequences is shown in **Figure 4.7**.

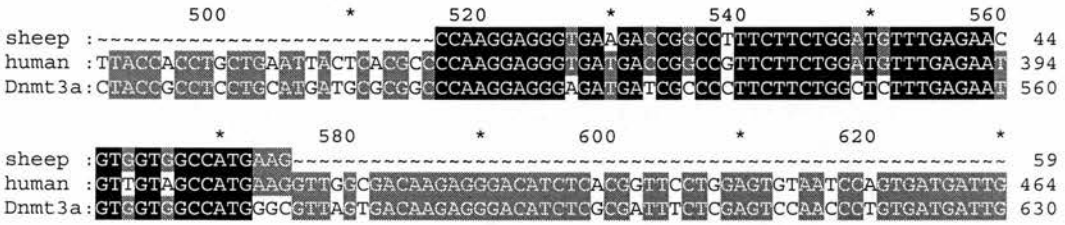


Figure 4.6 Alignment of the sheep DNMT3B cDNA sequence with human DNMT3B and DNMT3A sequences.

Multiple sequence alignment was performed using the 'localpileup' application (default parameters) of GCG10.

Figure 4.7 Alignment of the known sheep and human DNMT3B protein sequences. Pairwise protein sequence comparison was performed using the Gap application (default parameters) of GCG10. Vertical lines represent identical amino acids.

4.3 Discussion

Partial clones encoding the sheep *de novo* methyltransferases *DNMT3A* and *DNMT3B* have been isolated in sheep for the first time. A 1.85 Kb clone corresponding to the 3' region of *DNMT3A* was cloned through a combination of RT-PCR and cDNA library screening. Pairwise sequence alignments of the sheep cDNA sequence with the human or mouse *DNMT3A* sequences were generated, demonstrating very high levels of homology with both the human (94%) and mouse (89%) sequences. The cDNA sequence was translated and an open reading frame identified, terminating at nucleotide 1427 (stop codon TAA). The predicted sheep DNMT3A protein sequence was compared with the human and mouse protein sequences and was found to be highly homologous, with 473 out of 475 amino acid residues (99% homology) identical to the human and mouse sequences. The predicted protein sequence encodes the catalytic domain, containing the five highly conserved methyltransferases motifs, and the cysteine-rich domain within the N-terminal region.

The high level of conservation of DNMT3A between species is very interesting. Between the mouse and human DNMT3A protein sequences, the N-terminal domain is very highly conserved, (95%), more so than the N terminal regions of the mouse and human DNMT1 protein (79%). As the predicted protein sequence of the clone isolated from the sheep lung cDNA library only represents the C-terminal and N-terminal regions in close proximity to the catalytic domain, it is difficult to say whether the protein is completely conserved between sheep and other mammalian species. However, the high levels of conservation between the known DNMT3A sequences observed overall, signifies an important role for DNMT3A in mammals. Indeed, it is known to be essential for normal development in the mouse (Okano *et al.*, 1999), although its specific roles have not been clearly determined.

Attempts to clone *DNMT3B* in sheep resulted in a partial clone containing 59 nucleotides of novel sequence. cDNA library screening was performed to gain a larger cDNA clone using the small *DNMT3B* PCR product as a probe, but this

approach failed repeatedly. Further efforts to gain a larger clone included screening the cDNA library with the complete mouse *Dnmt3b* cDNA sequence, and designing additional PCR primers, but all of these experiments were unsuccessful. Failure of the library screening may have been due to poor representation of *DNMT3B* cDNA within the sheep adult lung cDNA library. Indeed, further experiments demonstrated that *DNMT3B* was undetectable in several adult tissues including lung, although it was found to be expressed in d125 fetal tissues (reported in Chapter 5). Additional attempts to clone further regions of *DNMT3B* using alternative PCR primers and d125 fetal tissue cDNA also failed, suggesting a possible lack of homology between the sheep and human or mouse *DNMT3B* sequences.

The small region of *DNMT3B* isolated in sheep demonstrated high levels of homology with the human sequence at both nucleotide and amino acid level, although it is not possible to determine whether this is indicative of the entire sequence. A comparison of the sheep *DNMT3B* and human *DNMT3A* cDNA sequences was also performed, as the two genes are known to share high levels of DNA sequence homology in the human. Although high levels of homology were observed between the sheep DNMT3B and human DNMT3A sequences (81%), this was lower than that observed between the sheep and human DNMT3B cDNA sequence (92%). This observation increased confidence in the verification of the sheep clone as DNMT3B encoding sequence.

Overall, *DNMT3B* appears to be the least conserved between species of the known functional DNA methyltransferases, with only 73% nucleotide sequence identity between the human and mouse *DNMT3B* cDNA sequences. This compares to 79% sequence identity between human and mouse *DNMT1* cDNA sequences, and 89% sequence identity between human and mouse *DNMT3A* cDNA sequences. This may partially explain why the sheep *DNMT3B* cDNA clone has remained so difficult to isolate.

In addition to cloning *DNMT3A* and *DNMT3B*, the *de novo* methyltransferases with known methylating activity, it would have been valuable to clone the sheep

homologue of the closely related *DNMT3L* gene. However, as DNMT3L was only recently identified (Aapola *et al.*, 2000), and the report describing its initial isolation and characterisation was only available during the latter stages of this study, it was deemed to be too late to initiate a new study. Moreover, *DNMT3L* is not expressed in adult mouse or human lung tissue (Aapola *et al.*, 2000; Aapola *et al.*, 2001), therefore a strategy of cDNA library screening using the lung cDNA library (the only sheep cDNA library commercially available) would not have been feasible.

As with *DNMT3B*, *DNMT3L* does not appear to demonstrate very high levels of sequence identity between species. A comparison of the human *DNMT3L* cDNA sequence and a partial mouse *Dnmt3L* sequence, demonstrated only 69% nucleotide sequence identity, which is lower than that seen between the human and mouse *DNMT3B* cDNA sequences. This observation would suggest potential difficulties in cloning the sheep homologue of *DNMT3L*, similar to those experienced whilst cloning *DNMT3B*. Nevertheless, given its importance in the establishment of DNA methylation marks in imprinted genes in mouse germline, attempts should certainly be made to clone *DNMT3L* and to assess its role in sheep development in the future.

The identification and substantial cloning of *DNMT3A* in sheep has demonstrated high levels of conservation of this protein between species and has also generated substantial sequence to facilitate further analysis of its roles in sheep development. In contrast, only a small region of *DNMT3B* was cloned in sheep, but sufficient sequence was obtained to confirm its identity. Although it would have been of interest to clone a larger region of *DNMT3B*, sufficient sequence was determined for expression analysis in sheep development, a primary interest of this thesis, which is discussed in the following chapter.

Chapter 5

Expression of DNA methyltransferase genes in sheep tissues and preimplantation embryos

5.1 Introduction

The expression patterns of DNA methyltransferase genes in human and mouse tissues have been well characterised, though less is known about events during preimplantation development, particularly for the putative *de novo* methyltransferases *DNMT3A* and *DNMT3B*. In the human, *DNMT1* is expressed ubiquitously in most adult tissues (Yen *et al.*, 1992), reflecting its role in maintaining methylation patterns during DNA replication and cell division. *DNMT3A* is also expressed ubiquitously in adult tissues, although at slightly lower levels than *DNMT1,* and *DNMT3B* transcripts are detected at low levels in comparison to both *DNMT1* and *DNMT3A* (Xie *et al.*, 1999). In adult mouse tissues, DNA methyltransferases demonstrate expression patterns similar to those seen in the human, with ubiquitous expression of *Dnmt1* (Bestor *et al.*, 1988) and downregulation of *Dnmt3a* and *Dnmt3b* in most adult tissues relative to levels in undifferentiated ES cells (Okano *et al.*, 1998a). It is worthy of note, that inappropriate upregulation of these three DNA methyltransferases is often associated with cellular neoplasticity (Robertson *et al.*, 1999; Kanai *et al.*, 2001).

The initial discovery that DNA methylation levels change significantly during mouse preimplantation development (Monk *et al.*, 1987) led to almost immediate investigation of Dnmt1 protein level and localisation in oocytes and preimplantation embryos (Monk *et al.*, 1991; Carlson *et al.,* 1992), however, analysis of *Dnmt1* mRNA levels was largely ignored. Dnmt1 protein levels are exceptionally high in

oocytes, with 15000-fold more Dnmt1 compared to fibroblasts cells (Carlson *et al.*, 1992). Dnmt1 levels then decline gradually during preimplantation development, with blastocysts retaining one quarter of the amount seen in oocytes (Carlson *et al.*, 1992). Dnmt1 protein also undergoes strict regulation of cellular localisation during preimplantation development, to aid the correct reprogramming of methylation patterns at this time (Carlson *et al.*, 1992; Cardoso & Leonhardt, 1999)

Previous studies of *de novo* methylation have focused on the role of the *de novo* methyltransferases Dnmt3a and Dnmt3b during postimplantation development. Mice carrying targeted disruptions of these genes died either at mid-gestation (*Dnmt3b*) or four weeks post-partum (*Dnmt3a*; Okano *et al.*, 1999). Furthermore, $Dnmt3a^{-/-}$, $Dnmt3b^{-/-}$ double mutant homozygous embryos were abnormally small, demonstrated a failure to form somites and died before d11.5. As the phenotype in double mutant homozygous mice was considerably more severe than either the *Dnmt3a* or *Dnmt3b* single mutant homozygotes, these proteins evidently have overlapping functions during embryogenesis. Although Dnmt3a and Dnmt3b have been shown to be essential for normal post-gastrulation development, any possible requirement in preimplantation embryos remains undetermined.

DNA methylation patterns are dynamic during the early stages of mouse development, and regulation of DNA methyltransferase genes is required to ensure that gamete specific methylation patterns are erased and subsequently re-established to support embryonic growth and differentiation. At a time when the majority of methylation patterns are being reprogrammed, imprinted genes often maintain or establish methylation patterns post-fertilisation (Tremblay *et al.*, 1997; Chaillet *et al.*, 1991; El-Maarri *et al.*, 2001). The mechanism by which this *de novo* or maintenance methylation occurs during oocyte maturation and preimplantation development still remains undetermined. Bisulfite analysis of the maternal imprint at the *Snrpn* locus in oocytes carrying a targeted deletion of the oocyte-specific Dnmt1 exon1o, ruled out DNA methylation by Dnmt1, as DNA methylation was

maintained (Howell *et al.*, 2001). It is possible that either of the two *de novo* methyltransferases (*Dnmt3a* or *Dnmt3b*) may be involved in this process, and analysis of *Dnmt3a* and *Dnmt3b* expression in oocytes may help determine any possible roles. Indeed, a closely related protein, Dnmt3L has been shown to be required for the establishment of maternal imprints (Bourc'his *et al.*, 2001), and could possibly interact with either Dnmt3a or Dnmt3b.

Imprinted genes are not the only DNA sequences that are *de novo* or continually methylated during preimplantation development. In the mouse, centromeric satellite DNA and IAP repeats retain high levels of methylation during the initial stages of preimplantation development (Rougier *et al.*, 1998; Walsh *et al.*, 1998). As Dnmt1o is actively retained in the cytoplasm during the first 3 cell divisions (Carlson *et al.*, 1992; Cardoso & Leonhardt, 1999), another DNA methyltransferase must be responsible for any methylating activity occurring at this time. Again, expression analysis of *Dnmt3a* and *Dnmt3b* may help to determine whether these genes are required for the methylation of specific DNA sequences during the early stages of preimplantation development.

In previous chapters, the cloning and sequence characterisation of sheep *DNMT1, DNMT3A* and *DNMT3B* cDNA sequences was described, with high levels of conservation observed. The aim of this particular study was to investigate the expression patterns of *DNMT1*, *DNMT3A* and *DNMT3B* in adult sheep tissues, and also to determine their presence in sheep oocytes and preimplantation embryos. This is the first reported study of *DNMT3A* and *DNMT3B* expression in mammalian preimplantation development. In addition, the levels of *DNMT1*, *DNMT3A* and *DNMT3B* expression were analysed in cultured fetal fibroblast cells, the cells used as donor cells in somatic cell nuclear transfer (SCNT) experiments described in Chapter 7. As Chapter 3 demonstrated different *DNMT1* isoforms in sheep and mouse oocytes with potentially significant consequences for methylation reprogramming, expression of the *Dnmt1, Dnmt3a* and *Dnmt3b* genes was also

analysed in mouse oocytes and preimplantation embryos, in order to investigate any species-specific differences.

5.2 Methodology

For quantitative analysis of *DNMT1*, *DNMT3A* and *DNMT3B* gene expression in adult tissues and fibroblast cells, Taqman® real-time PCR was used. The Taqman® chemistry, methodology and methods of RNA quantification are outlined in Chapter 2 (**2.4.12**). The method employed for this particular study allows quantification of RNA relative to a 'calibrator' tissue sample; the equation used to calculate relative expression levels of the DNMT genes is shown below.

$$2^{-\Delta\Delta Ct}$$

When analysing gene expression in a range of tissues, expression values were first normalised against the relative 18S level (ΔCt). A calibrator tissue is then chosen, usually the tissue with the lowest level of gene expression (or highest ΔCt value) and designated the 1x sample. The calculation of ΔΔCt involves subtraction by the ΔCt calibrator value; for example if heart tissue is the calibrator, then all tissues will subtract the heart ΔCt value from their own ΔCt value.

For the analysis of *DNMT1*, *DNMT3A* and *DNMT3B* transcript levels in oocytes and preimplantation embryos, single oocytes or embryos were used rather than pooled. This allowed for the detection of biological variation within a group of embryos of the same stage; a particularly useful molecular tool when comparing heterogeneous embryos derived from embryo technologies. Although in this particular study, time limitations prevented gene expression comparison in embryos derived from different

embryo production methods, existing techniques permitting gene expression analysis in single blastocysts (Young *et al.*, 1998b) were optimised for use in earlier stage *in vivo*-derived embryos and oocytes.

It was initially hoped that Taqman® methodology could be employed to analyse *DNMT1*, *DNMT3A* and *DNMT3B* transcript levels in single sheep or mouse oocytes and preimplantation embryos. However, given the specified requirement for relatively high levels of expression (PCR exponential phase must be between 18 and 32 cycles) and the small amount of starting material, it was not possible to quantify transcript levels using this technique. Instead, a method allowing simultaneous comparison of *DNMT1*, *DNMT3A* and *DNMT3B* transcript levels in oocytes and preimplantation embryos of various stages was employed.

To validate the detection or absence of gene-specific transcripts in oocytes and embryos, 18S was used as an internal control for the cDNA integrity and PCR process. As 18S is itself subject to changes in gene expression levels throughout preimplantation development (Bilodeau-Goeseels & Schultz, 1997), as are other housekeeping genes, a perfect internal reference for comparing *DNMT* levels between different stages is not available. The RT-PCR method was designed so that amplification of all genes (*DNMT1*, *DNMT3A*, *DNMT3B* and 18S) at all embryo stages was performed simultaneously and PCR products were electrophoresed on the same agarose gel. Thus individual stage-specific variations in transcript levels could be compared between different genes. The method also allowed comparison of expression patterns for any one specific gene in the mouse and sheep.

5.3 Results

5.3.1 Expression of *DNMT1* in adult tissues and fetal fibroblast cells

To quantify *DNMT1* expression in adult tissues and fetal fibroblast cells, Taqman® real-time RT-PCR was used. The 'Comparative Ct' method of relative quantification (described in **2.4.12**) was employed, with 18S acting as the endogenous control. Primers and probe specific to the sheep *DNMT1* cDNA sequence were designed (as described **2.12.3)** and can be found in **Table 2.6**. For the template, RNA was extracted from heart, liver, kidney, lung, ovary and spleen tissue and 1μg of each tissue RNA sample was subsequently reverse transcribed (as described **2.4.11.1**). RNA was also extracted from pelleted fetal fibroblast cells and 1μg of RNA was reverse transcribed (as described **2.4.11.1**) In parallel, 1μg of each tissue/cell RNA sample was carried through the reverse transcription procedure, minus the reverse transcriptase enzyme, and designated as the tissue 'no RT control'. Any amplification from these samples in subsequent PCR reactions would be from DNA, and thus the no RT control acts as a measure of significant DNA contamination. In addition, a blank 'no sample control' was simultaneously carried through all RNA extraction/reverse transcription steps to ensure reagents and plastic consumables were not contaminated.

A preliminary Taqman® PCR using the *DNMT1*-specific primers and probe was performed to check the no RT controls were not significantly contaminated in relation to the relative cDNA sample, and also to check the no sample control for general contamination. The Taqman® PCR was performed (as described **2.12.3**) with optimal *DNMT1* primer/probe conditions as described in **Table 2.6**. Template was either the no sample control, or a 1:10 dilution of tissue cDNA stock (PCR optimised for this amount to conserve cDNA stocks) or corresponding tissue no RT control. The Taqman® PCR was optimised for a 1:10 cDNA dilution to ensure all *DNMT* and *MBD2* genes could be analysed within a single tissue cDNA sample. A

'no template control' was also performed to verify the contamination-free status of the Taqman® PCR reagents. The average Ct values for the triplicate reactions performed for each sample were calculated and are demonstrated in **Table 5.1**. The no template control (NTC) and no sample control (NSC) displayed Ct values of 40, indicating no amplification of PCR products after 40 cycles, and hence demonstrating no contamination of tissue samples occurred during RNA extraction/cDNA synthesis steps or the subsequent PCR. Most of the no RT samples demonstrated Ct values of <40, indicating DNA contamination of the RNA samples, despite DNase treatment (see **Table 5.1**). As the differences between no RT samples and cDNA samples of respective tissues were significant (6-10 cycles difference), and each cycle represents a 2-fold difference in expression, the contribution of DNA to the total expression level measured would be 1/64 ($1/2^6$) at a maximum (Heart tissue) and 1/1024 ($1/2^{10}$) at a minimum (fibroblasts). This 0.1-1.0% bias was deemed acceptable for the purpose of this study.

The no RT and NSC controls were also checked with the 18S primer/probe set. As the DNA contamination seen in the no RT controls (analysed using the *DNMT1* primers/probe) was considered negligible, the tissue no RT controls were pooled, and a Taqman® PCR performed to compare the Ct value of pooled no RT controls with pooled tissue cDNA (as described **2.15.3**; optimal primer/probe conditions described in Table **2.5**). The no RT control gave an average Ct value of 29.1, compared with a value of 16.82 from pooled cDNA. The difference in cycle numbers corresponds to a 4000-fold (2^{12}) difference in expression, meaning that DNA contamination would only contribute 1/4000 (0.025%) to the overall expression level at a maximum. The NSC gave a Ct value of >35 when used in the 18S Taqman PCR, demonstrating no significant contamination. As the 18S ribosomal RNA reagents will amplify 18S across a wide range of species, it is difficult to gain a NSC with absolutely no 18S contamination, and Ct values of 35-40 are widely accepted as the norm.

Sample	cDNA	no RT
NTC	40	-
NSC	40	-
Heart	31.86	37.91
Kidney	29.95	39.41
Liver	31.27	40
Lung	28.56	37.97
Ovary	28.34	37.18
Spleen	27.93	37.84
Fibroblast	29.45	40

Table 5.1 Average Ct values for tissue cDNA samples and controls after amplification with *DNMT1*-specific primers and probe (pre-quantitation control experiment).

The 'no sample' control and 'no template' control are indicated as NSC and NTC respectively. Ct values are shown in the 'cDNA' or 'no RT' columns and are the average of three triplicate PCR reactions performed simultaneously.

Prior to quantification of *DNMT1* in tissues, an additional experiment was required to measure the relative efficiencies of the *DNMT1* and 18S primer/probe sets over an equivalent and suitable working range of cDNA dilutions. This was required in order to validate the Comparative Ct method of quantification, which relies on the relative efficiencies of the 'gene-of-interest' and 'endogenous control' primer/probe sets to be equivalent. The Taqman® PCR reaction to assess relative primer efficiency was performed (as described **2.4.12.3**), with the *DNMT1* and 18S PCR reactions present on the same plate. The cDNA dilutions were prepared using pooled cDNA from all tissue samples. The optimal primer and probe conditions used are described in **Table 2.6** (18S final primer concentration, 50nm) To measure the relative efficiencies of the two primer/probe sets, ΔCt values were plotted against the log of input cDNA dilution, and a line of best fit generated. The slope of the line was 0.14, demonstrating relative primer efficiencies within the accepted range (<0.1). The graph showing the relative linear efficiency of the *DNMT1* and 18S primer/probe sets is presented in **Figure 5.1**.

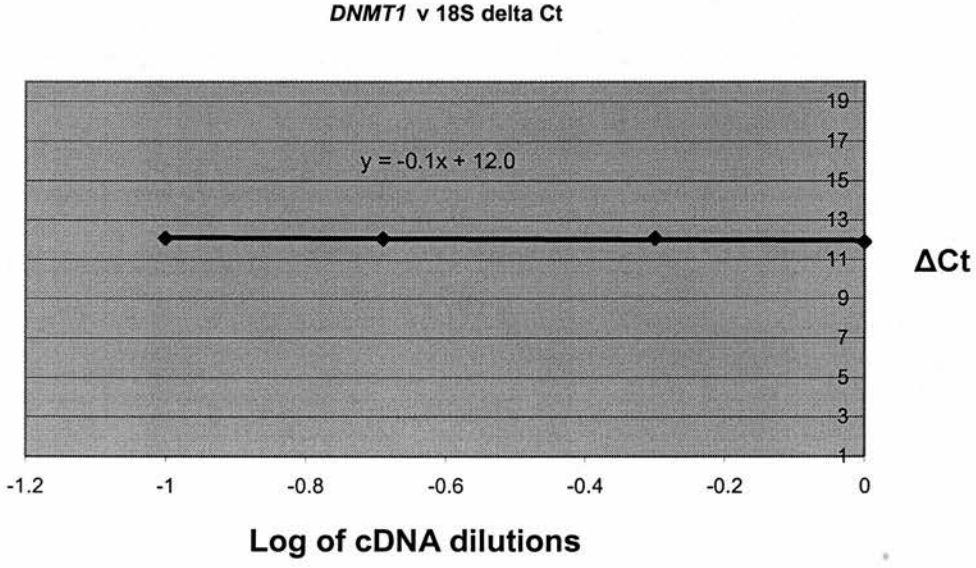


Figure 5.1 Evaluation of *DNMT1* and 18S relative primer efficiency.

The Log of input cDNA dilutions are indicated on the x axis. The ΔCt values (*DNMT1* Ct minus 18S Ct) are indicated on the x axis. Over a working range of cDNA dilutions, the *DNMT1*/18S ΔCt values produced a line with slope of 0.1, demonstrating acceptable relative primer efficiencies.

After confirmation that *DNMT1* and 18S primer/probe sets performed with equal efficiency, quantification of *DNMT1* expression was performed. Dilutions of 1:10 and 1:100 for each tissue cDNA were prepared for expression analysis of *DNMT1* and 18S respectively. The Taqman® PCR was performed (as described **2.12.3**) with the *MBD2* and 18S PCR reactions present on the same plate. Optimal primer/probe conditions were determined and are described in **Table 2.5** (18S primer concentration, 50 nm). The PCR was repeated, and the mean of the two ΔCT values for each tissue was calculated and used in subsequent quantification of *DNMT1*. The tissue found to have the lowest levels of *DNMT1* expression was liver, and was therefore designated the 'calibrator' tissue (1-fold). The levels of *DNMT1* expression in the other tissues and cells are therefore expressed as *n-fold* increase relative to liver. The results of the quantification experiment are shown in **Table 5.2**.

The result of the quantification experiment demonstrated expression of *DNMT1* in all tissues analysed, and also in the fetal fibroblast cells. Expression was highest in spleen and lung, whereas heart and kidney demonstrated levels similar to liver. *DNMT1* was found to be moderately expressed in fetal fibroblasts relative to expression observed in the other analysed tissues.

Tissue	*DNMT1* Ct	*18S* Ct	ΔCt	Average ΔCt	ΔΔCt ($\Delta Ct-\Delta Ct_{Liver}$)	DNMT1 Rel. to Liver
Liver (1)	30.03	16.31	13.72	13.94+/- 0.31	0+/- 0.31	1
Liver (2)	30.33	16.17	14.16			(0.81-1.24)
Heart (1)	30.54	18.18	12.36	12.91+/- 0.77	-1.03+/- 0.77	2.04
Heart (2)	31.79	18.32	13.47			(1.18-3.5)
Kidney (1)	28.81	16.09	12.72	12.97+/- 0.35	-0.97+/- 0.35	1.96
Kidney (2)	29.36	16.15	13.21			(1.54-2.5)
Lung (1)	27.45	16.42	11.03	11.42+/- 0.54	-2.53+/- 0.54	5.76
Lung (2)	28.42	16.62	11.80			(3.96-8.37)
Ovary (1)	27.22	16.21	11.01	11.48+/- 0.66	-2.46+/- 0.66	5.50
Ovary (2)	28.19	16.24	11.95			(3.48-8.69)
Spleen (1)	26.75	16.03	10.70	10.92+/- 0.31	-3.02+/- 0.31	8.11
Spleen (2)	27.27	16.13	11.14			(6.54-10.06)
Fibroblast (1)	28.43	16.54	11.89	11.96+/- 0.10	-1.98+/- 0.1	3.94
Fibroblast (2)	28.66	16.63	12.03			(3.68-4.23)

Table 5.2 Relative quantitation of *DNMT1* using the Comparative Ct method.

Data was collected from the two separate experiments and designated as experiment (1) or experiment (2) in the table. The *DNMT1* and 18S Ct values are the mean Ct values gained in triplicate PCR reactions. ΔCt equates to *DNMT1* Ct minus 18S Ct. The average ΔCt value of the two experiments, and the standard deviation (s.d) of the two ΔCt values is shown. The ΔΔCt calculation involved the subtraction by the chosen calibrator tissue, liver. *DNMT1* expression is thus determined as *n-fold* relative to liver. The range of expression given for each tissue (in brackets) was determined by solving the eqation $\mathbf{2}^{-\Delta\Delta Ct}$, using ΔΔCt + s.d or ΔΔCt –s.d

5.3.2 Expression of *DNMT3A* in adult tissues and fetal fibroblast cells

Quantification of *DNMT3A* in adult tissues was performed using the Comparative Ct method as described for *DNMT1*. Primers and probe were designed for the sheep *DNMT3A* cDNA sequence, as described **2.12.3**; primer and probe sequences can be found in **Table 2.6**. Sufficient template for *DNMT3A* expression analysis remained in the cDNA samples described in **5.2.1**. As the *DNMT3A* primer/probe set span a relatively large intron (~1.2 Kb, see **4.2.1**), amplification of contaminating DNA should not contribute to the observed PCR signal. To demonstrate this, analysis of two randomly chosen tissue samples (heart and lung) was performed, alongside analysis of the 'no sample control' (NSC) to confirm the previous observation of no contamination when using the *DNMT1* primer/probe set. The Taqman® PCR was performed as described in **2.12.3**; optimal *DNMT3A* primer/probe conditions are described in **Table 2.6.** Template was either a 1:10 dilution of tissue cDNA or corresponding no RT control, or a 1:10 dilution of 'no sample control'. The results are shown in **Table 5.3**.

Sample	Average Ct
NSC	40
Heart	30.13
Heart no RT	40
Lung	27.3
Lung no RT	40

Table 5.3 Average Ct values for tissue cDNA and RT-PCR controls after *DNMT3A*-specific amplification.

'No sample control' is indicated as NSC. Average Ct values are the mean of the Ct values gained from the three triplicate PCR reactions performed for each sample.

As with the *DNMT1*, evaluation of the relative efficiencies of *DNMT3A* and 18S primers was measured over an equivalent range of cDNA dilutions prior to quantification experiments. The Taqman® PCR reaction to assess relative primer efficiency was performed (as described **2.4.12.3**), with both *DNMT3A* and 18S PCR reactions present on the same plate. Initially, the 18S primers were used at a final concentration of 50 nm, but analysis of the relative primer efficiencies produced a line with slope >0.1. The Taqman® PCR was repeated, limiting the final 18S primer concentration to 25 nm. In this case, plotting of the ΔCt values against the log of input cDNA dilution generated a line of best fit with slope <0.1. The graph showing the relative linear efficiency of the two primer/probe sets is presented in **Figure 5.2**.

After confirmation of relative equal efficiency of the *DNMT3A* and 18S primer/probe sets, quantification of *DNMT3A* gene expression in a range of tissues was performed. Dilutions of 1:10 and 1:100 for each tissue cDNA sample were prepared for expression analysis of *DNMT3A* and 18S respectively. The Taqman® PCR was performed (as described **2.12.3**) with both the *DNMT3A* and 18S PCR reactions performed simultaneously on the same plate. Optimal primer/probe conditions are described in **Table 2.6** (18S primer concentration, 25 nm). The PCR was repeated, and the mean of the two ΔCt values for each tissue was calculated and used in subsequent quantification of *DNMT3A (*as described **2.12***)*. The tissue with the lowest levels of *DNMT3A* expression was liver, and was therefore designated the calibrator tissue. The levels of expression of *DNMT3A* are expressed as *n-fold* relative to liver and results of the quantification experiment are shown in **Table 5.4**

Analysis of *DNMT3A* in tissues and fibroblast cells demonstrated high levels of expression in the lung, ovary and fibroblast. Expression levels were lowest in the liver, heart and kidney, and moderate levels of expression were seen in the spleen.

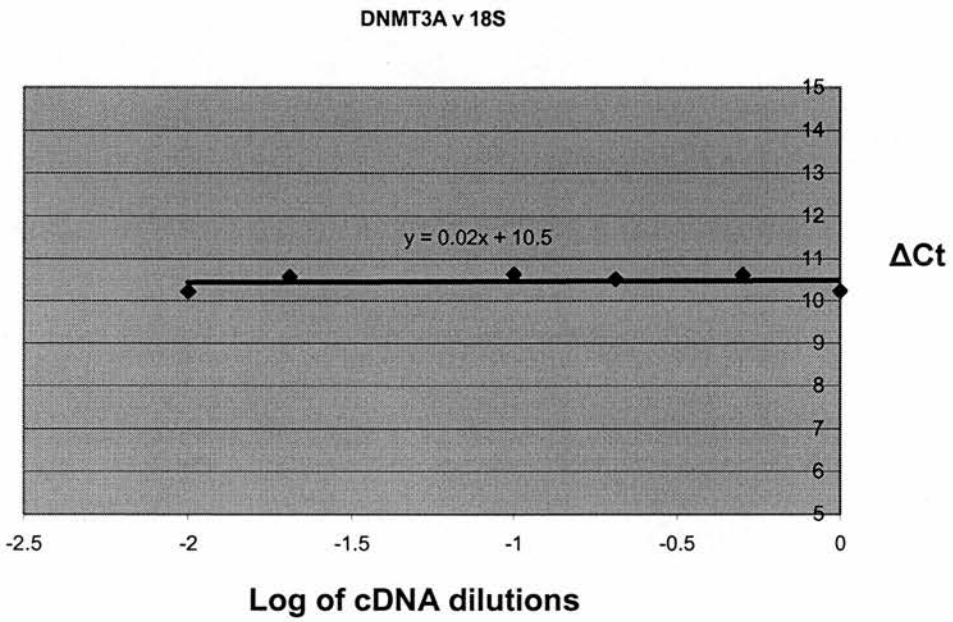


Figure 5.2 Evaluation of *DNMT3A* and 18S relative primer efficiency.

The Log of input cDNA dilutions are indicated on the x axis. ΔCt values (*DNMT3A* Ct minus 18S Ct) are indicated on the y axis. Over a working range of cDNA dilutions, the *DNMT3A*/18S ΔCt values produced a line with slope 0.02, demonstrating acceptable relative primer efficiencies.

Tissue	*DNMT3A* Ct	*18S* Ct	ΔCt	Average ΔCt	ΔΔCt ($\Delta Ct-\Delta Ct_{Liver}$)	*DNMT3A* Rel. to Liver
Liver (1)	28.3	16.41	11.89	11.49+/- 0.57	0+/- 0.57	1
Liver (2)	27.65	16.56	11.10			(0.67-1.48)
Heart (1)	29.56	18.82	10.74	10.29+/- 0.64	-1.2+/- 0.64	2.30
Heart (2)	28.41	18.58	9.83			(1.47-3.58)
Kidney (1)	27.54	16.34	11.20	10.76+/- 0.63	-0.73+/- 0.63	1.66
Kidney (2)	26.3	15.99	10.31			(1.07-2.57)
Lung (1)	26.41	16.67	9.74	9.15+/- 0.83	-2.34+/- 0.83	5.06
Lung (2)	25.36	16.8	8.56			(2.85-9.00)
Ovary (1)	26.07	16.3	9.77	9.05+/- 1.04	-2.44+/- 1.04	5.43
Ovary (2)	24.83	16.52	8.31			(2.64-11.16)
Spleen (1)	26.36	16.01	10.35	9.72+/- 0.9	-1.77+/- 0.9	3.41
Spleen (2)	25.09	16.01	9.08			(1.83-6.36)
Fibroblast (1)	26.64	17.18	9.46	8.98+/- 0.69	-2.51+/- 0.69	5.7
Fibroblast (2)	25.55	17.06	8.49			(3.53-9.19)

Table 5.4 Relative quantitation of *DNMT3A* using the Comparative Ct method.

Data was collected from two separate experiments and designated experiment (1) or experiment (2) in the table. ΔCt equates to *DNMT3A* Ct minus 18S Ct. The average ΔCt value of the two experiments, and the standard deviation (s.d.) of the two ΔCt values is also shown. The ΔΔCt calculation involved the subtraction by the chosen calibrator tissue, liver. *DNMT3A* expression is thus determined as *n-fold* relative to liver. The range of expression given for each tissue (in brackets) was determined by solving the equation $\mathbf{2}^{-\Delta\Delta Ct}$ using ΔΔCt + s.d or ΔΔCt - s.d.

5.3.3 Expression of *DNMT3B* in adult tissues and fetal fibroblast cells

Quantification of *DNMT3B* expression in adult tissues and fetal fibroblast cells, was performed using Taqman® real-time RT-PCR as previously described for *DNMT1* and *DNMT3A*. Primers and probe were designed for the sheep *DNMT3B* cDNA sequence, as described **2.12.3**; primer and probe sequences can be found in **Table 2.6**. Sufficient template for *DNMT3B* expression analysis remained in the cDNA samples described in **5.2.1**. Preliminary Taqman® RT-PCR results using 1:10 dilution of tissue cDNA demonstrated no detectable *DNMT3B* transcripts in adult tissues and fetal fibroblast cells after 40 cycles of PCR (Taqman® RT-PCR performed as described **2.4.12.3**; optimal primer and probe conditions described **Table 2.6**). Therefore it was not possible to perform *DNMT3B* quantification in these tissues using this method. The Taqman® RT-PCR was repeated to confirm the results, and average Ct values of a typical *DNMT3B* expression analysis experiment are shown in **Table 5.5**. As a positive control, simultaneous amplification of the same tissue cDNA samples using 18S Taqman® primers and probes was performed on the same plate (as described **2.4.12.3**).

To further confirm these observations, *DNMT3B* expression was analysed in adult tissues using standard RT-PCR methods. A 1:10 dilution of adult tissue cDNA was used in a PCR reaction, using the DNMT3B2 primer pair (as described **2.4.1**; primer-specific PCR conditions are described in **Tables 2.1** and **2.2**) As a positive control, 1:10 dilutions of fetal d125 fetal heart, kidney or liver cDNA (synthesised from 1 μg of RNA using the same cDNA synthesis method as described for adult tissues (see **2.4.11.1**) were used as template for the PCR reaction. D125 tissue cDNA has previously been used successfully as a positive control for DNMT3B2 PCR reactions (see **Figure 4.2**, Chapter 4), and in addition, pooled d125 fetal cDNA was used to optimise the *DNMT3B* Taqman® primer and probe sets, with Ct values <40 observed. PCR products were electrophoresed in a 2% ethidium-bromide

stained agarose gel (as described **2.4.2**); the results of this PCR are shown in **Figure 5.3.**

The standard DNMT3B2 PCR demonstrated no PCR amplification in any of the adult tissues after 40 cycles; however, amplification was observed from the d125 fetal tissue cDNA samples. This clearly indicates *DNMT3B* is expressed in d125 fetal tissues, but is not detectable in adult tissues after 40 cycles of PCR, and is thus either absent, or expressed at very low levels.

Sample	Ct value
NTC	40
Heart	40
Kidney	39.7
Liver	39.6
Lung	39.8
Ovary	39
Spleen	40
Fibroblast	40

Table 5.5 ***DNMT3B*** **expression in adult sheep tissues.**

'No Template control' is indicated as NTC. Average Ct values are the mean of the Ct values gained from the three triplicate PCR reactions performed for each sample.

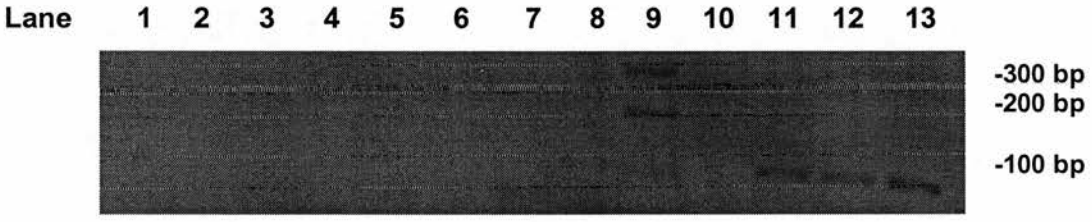


Figure 5.4 ***DNMT3B*** **expression in adult and d125 fetal tissues.**

Lanes 2-8 are PCR reactions containing adult heart, kidney, liver, lung, ovary, spleen and fetal fibroblast cDNA respectively. Lanes 11-13 demonstrate PCR products amplified from d125 fetal heart, kidney and liver cDNA respectively. Lanes 1 and 10 are negative (no template) controls. DNA marker sizes (Lane 9; 1 Kb Plus DNA ladder (Invitrogen Life Technologies)) are indicated to the right of the gel image.

5.3.4 Qualitative analysis of *DNMT1*, *DNMT3A* and *DNMT3B* transcripts in sheep and mouse oocytes and preimplantation embryos

To detect DNA methyltransferase transcripts in single oocytes and preimplantation embryos, a non-quantitative PCR method was employed. 18S was used as an endogenous control, confirming the presence of embryonic derived cDNA in the case of failure to amplify PCR products for the DNMT-specific primers. Briefly, RNA was extracted from three single oocytes/embryos of the following developmental stages: MII oocyte, 2-cell, 8-16 cell, morula and blastocyst, and subsequently reverse transcribed (as described **2.4.11.2**). In addition, a no embryo control (NEC) was carried through all RNA extraction/cDNA synthesis steps to rule out any possible contamination from reagents or plastic consumables. A no RT control was also performed, with the RNA extracted from a single blastocyst taken through cDNA synthesis minus the reverse transcriptase. Four PCR reactions per single oocyte/embryo were performed using the MTRA1, HDNMT3A, DNMT3B2 and Classic II 18S primers (as described **2.4.1**; primer-specific PCR conditions are described in **Tables 2.1** and **2.2**), and PCR reactions for all genes and embryonic stages were performed simultaneously. As template, 2μl of cDNA was used for the 18S PCR and 4μl for the *DNMT1*, *DNMT3A* or *DNMT3B* PCR reactions. For the 'no embryo control' or no RT control, 4μl of cDNA was used as template, and as a positive control, 2μl of d125 fetal liver cDNA was used. All PCR products were electrophoresed on the same 2.5% ethidium bromide stained agarose gel (as described **2.4.2**), results are shown in **Figure 5.4 (A)** and are representative of two independent experiments. PCR amplification from the NEC control demonstrated no contamination; a very faint band was observed as a result of PCR amplification of the no RT control using 18S primers, demonstrating a slight but tolerable contamination.

A parallel experiment was performed analysing *Dnmt1*, *Dnmt3a* and *Dnmt3b* transcript levels in mouse embryos. This experiment was performed firstly, to demonstrate the applicability of the PCR method to single embryo/oocyte analysis in the mouse, where the amount of starting RNA is even lower than in sheep. Secondly, analysis of *Dnmt3a* and *Dnmt3b* transcript levels has not previously been described in mouse oocytes or preimplantation embryos. RNA was extracted from oocytes and embryos and reverse transcribed (as described **2.4.11.2**), with the same controls (NEC and no RT) as described for sheep embryos. PCR reactions were performed as described for single sheep embryos (**5.2.4**), using the same primer-specific PCR conditions (as described **Tables 2.1** and **2.2**). Template (samples and controls) was as described for DNMT expression analysis in sheep embryos (**5.2.4**). All PCR products were electrophoresed on the same 2.5 % ethidium bromide stained agarose gel (as described **2.4.2**), results are shown in **Figure 5.4 (B)**. The results are representative of two independent experiments. The MTRA1, HDNMT3A and DNMT3B2 PCR products amplified from mouse-derived cDNA were subcloned and sequenced (DNMT3B2), as described **2.4.3**, or digested with sequence-specific restriction enzymes (MTRA1 (*Hpa* II) and HDNMT3A (*Taq* I)) to verify insert identity (as described **2.4.5**).

In the sheep, the apparent increase of 18S at the 2-cell stage, relative to the oocyte, could be explained by incomplete lysis of the oocyte, possibly due to a difficulty in rupturing the zona pellucida of the sheep oocyte at this stage. The same observation was made when the RNA extraction, cDNA synthesis and PCR were repeated. *DNMT1* was detected at all stages analysed, with apparently higher levels of expression in oocytes and 2-cell embryos followed by an apparent decrease at the 8-16 cell stage. This probably corresponds to the destruction of maternally-inherited RNA. *DNMT3A* and *DNMT3B* were also detected at all stages analysed, again with an apparent decrease at the 8-16 cell stage (more evident with *DNMT3A*); however, *DNMT3A* transcript levels appeared to increase at the blastocyst stage.

In the mouse, similar patterns were observed. *Dnmt1* was detected at all stages, with high levels in the oocyte, an apparent reduction at the 2-cell stage and similar levels observed henceforward. *Dnmt3a* was also highly expressed in oocytes, relative to *Dnmt3b,* and detectable at all stages, although the apparent increase seen after the onset of zygotic genome activation (2-cell stage in the mouse) was not as marked as in the sheep. *Dnmt3b* was detected at all preimplantation embryo stages analysed.

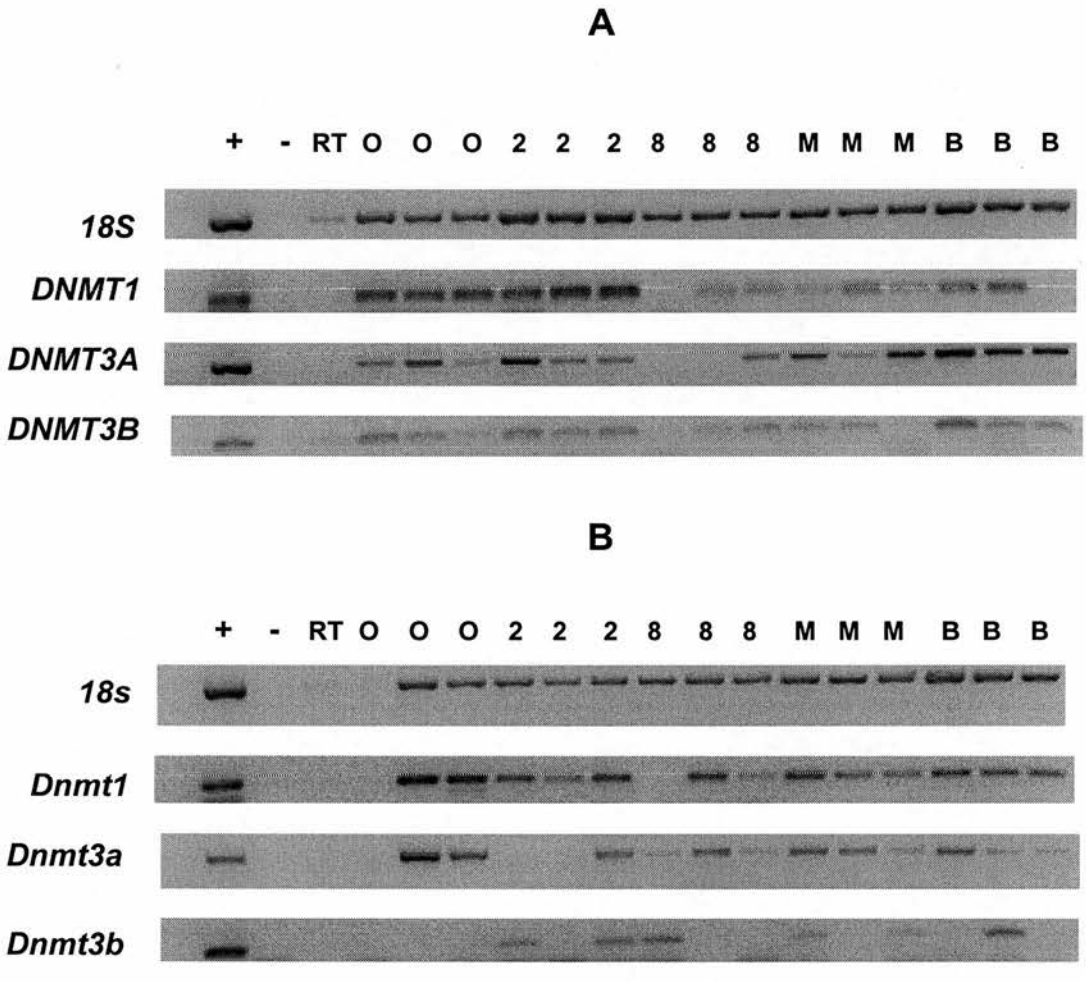


Figure 5.4 ***DNMT1*, *DNMT3A* and *DNMT3B* transcripts in single sheep and mouse oocytes or preimplantation embryos.**

Results for sheep oocytes and preimplantation embryos are shown in gel image **A**; results for the mouse are shown in gel image **B**. Gene-specific PCR products are indicated to the left of the gel image. Lanes are labelled with the origin of the template cDNA and are as follows; + (positive control; liver cDNA); - (no embryo control); RT (no RT control); O (oocyte cDNA); 2 (2-cell embryo); 8 (8-16 cell embryo); M (morula stage embryo) and B (blastocyst stage embryo). The four gene-specific PCR products amplified from cDNA derived from a single oocyte/embryo are shown within the same lane column.

5.4 Discussion

Semi-quantitative Taqman® real-time RT-PCR experiments demonstrated expression of *DNMT1* in all adult sheep tissues and cell types analysed. Expression levels were measured and were found to be highest in spleen (8.1-fold increase), lung (5.8-fold increase) and ovary (5.5-fold increase), moderate in fetal fibroblast cells (3.9-fold increase), and lowest in heart (2-fold increase) and kidney (2-fold increase) relative to liver. These observations fit with the requirement for high levels of *DNMT1* expression in tissues with high cellular turnover, in order to methylate newly replicated DNA and mirror the patterns observed in the human and mouse where known (Yen *et al.*, 1992; Bestor *et al.*, 1988). The high levels of expression in ovary tissue may reflect the acquisition of large stores of Dnmt1 in growing oocytes (Mertineit *et al.*, 1998) in addition to the high cell turnover.

Analysis of *DNMT3A* expression in sheep adult tissue samples, using the same quantification method as for *DNMT1*, demonstrated high levels of expression in fetal fibroblasts (5.7-fold increase), ovary (5.4-fold increase) and lung (5.1-fold increase), moderate in spleen (3.4-fold increase), and lowest in heart (2.3-fold increase) and kidney (1.6-fold increase) relative to liver. The relatively high levels of *DNMT3A* expression observed in ovary tissue was particularly interesting, and may indicate an important role for *DNMT3A* in oocyte growth. To further support this notion, analysis of *DNMT3A* in mature oocytes and early stage sheep preimplantation embryos demonstrated significant levels of expression (see **5.3.4**).

DNMT3A expression in sheep tissues was very similar to patterns observed in both the human and mouse. *DNMT3A* was also found to be highly expressed in sheep fetal fibroblasts (cells derived from d25 fetuses). This observation correlates with findings in the mouse, where *Dnmt3a* is ubiquitously expressed in the developing fetus, but downregulated in adult tissue relative to undifferentiated ES cells (Okano *et al.*, 1999; Okano *et al.*, 1998a).

DNMT3B expression analysis indicated very low levels, or potentially absent expression in all adult tissues and fetal fibroblast cells. However, *DNMT3B* was found to be expressed in heart, kidney and liver tissue from d125 fetuses. The expression of *de novo* methyltransferases in adult sheep tissues described here reflects findings in the human, where *DNMT3A* is expressed at high levels in most tissues (although at slightly lower levels than *DNMT1*), and *DNMT3B* is expressed at much lower levels than either *DNMT1* or *DNMT3A* (Okano *et al.*, 1999). Adult human tissues in which *DNMT3B* is detectable include testis, thyroid and bone marrow (Okano *et a*l., 1999), tissue types that were not available for analysis in this study.

It was not possible to accurately determine whether *DNMT1* or *DNMT3A* is the most abundant DNA methyltransferase transcript in adult sheep tissues. The Taqman® real-time PCR quantification technique requires gene-specific primers/probes of equal efficiency and the *DNMT3A* primer/probe were found to be more efficient than those designed for *DNMT1*. However, Ct values for both genes appeared to be within a similar range, suggesting comparable levels of expression, although different tissue-specific patterns of expression were observed for each gene.

Analysis of *DNMT1*, *DNMT3A* and *DNMT3B* transcripts in sheep oocytes and preimplantation embryos was also performed. *DNMT1* was shown to be expressed in oocytes and 2-cell embryos, followed by an apparent decrease at the 8-16 cell stage, most likely corresponding to the degradation of maternally-derived transcripts. As RT-PCR only measures steady-state transcripts, it could not be determined in this experiment as to whether *DNMT1* is actively transcribed from the embryonic genome at the morula or blastocyst stages, or whether detection of transcripts is due to residual maternal stores of *DNMT1*. The relative abundance of *DNMT1* transcripts in sheep oocytes and early preimplantation embryos cannot be disputed, demonstrating significant similarities with the mouse. It remains to be determined whether this similarity extends to protein levels, and due to the apparent

lack of an oocyte-specific *DNMT1* in sheep, it would be surprising to see the unusual nuclear trafficking of DNMT1, as observed in the mouse preimplantation embryo (Carlsson *et al.*, 1992)

DNMT3A was shown to be expressed in the oocyte, followed by an apparent decrease to the 8-16 cell stage. However, in contrast with *DNMT1*, there is a marked increase in *DNMT3A* at the blastocyst level, perhaps suggesting transcription of *DNMT3A* from the embryonic genome has occurred at, or prior to, this stage. *DNMT3B* was detected in oocytes and all embryo stages analysed, although, as with *DNMT1*, it is unclear as to the origin of transcripts at the later stages of preimplantation development.

Analysis of *Dnmt1*, *Dnmt3a* and *Dnmt3b* expression in mouse oocytes and preimplantation embryos allowed direct comparison with the sheep. Previous results have demonstrated significant species-specific differences in *DNMT1* isoform expression in mouse and sheep oocytes (Chapter 3) and it was of particular interest to see if species-specific expression patterns of other *de novo* methyltransferases also exist. The *Dnmt1* transcript was present at all embryo stages and was apparently high in oocytes and 2-cell stage embryos, with levels decreasing gradually from this point onwards. *Dnmt3a* transcripts levels were similar to *Dnmt1* in oocytes, with a gradual decrease through later stages, and no marked increase at the morula/blastocyst stage, as observed in the sheep. *Dnmt3b* transcripts were detected at all stages of mouse preimplantation development, with no clear transcriptional origin during later stages.

This is the first reported investigation of *de novo* methyltransferase expression in mammalian oocytes or preimplantation embryos. It is likely from these findings, that *DNMT3A* in particular, may have an active role in *de novo* methylation in oocytes or early stages of preimplantation development. There are known *de novo* methylation activities occurring in the growing mouse oocyte, with the

establishment of methylation imprints occurring at this time (Obato & Kono, 2002). Furthermore, *de novo* methylation may also occur post-fertilisation in the absence of Dnmt1 (Rougier *et al.*, 1998; Walsh *et al.*, 1998). The observation in this thesis that Dnmt3a is apparently highly expressed in both sheep and mouse oocytes make this gene a potential candidate for establishing maternal imprints in the oocyte. Examination of its expression in different stages of oocyte growth would deduce any such role. As both Dnmt3a and Dnmt3b are expressed in early stage preimplantation embryos, either protein could be involved in *de novo* methylating activities occurring at this time.

It was noticed that transcript levels were sometimes inconsistent between sheep or mouse oocytes and embryos of the same stage. This apparent variability of gene transcript levels in similar stage embryos may be biologically significant, although it is also possible that differences could be due to unequal PCR amplification between tubes. The small amounts of template used in individual PCR reactions (cDNA comprising 1/4 of a single oocyte or embryo's RNA) could certainly account for inconsistent amplification between PCR tubes. To further back this hypothesis, inconsistencies in gene transcript levels were more obvious in oocytes and earlier stage preimplantation embryos, when RNA isolation may be problematic. The inconsistencies may also reflect the efficiency of the primer pair used for amplification; for example, *DNMT1* transcript levels appear to be more consistent between same stage embryos than *DNMT3A* or *DNMT3B* transcript levels. Thus, the *DNMT1* primer set (MTRA1) may be more efficient than those used to amplify *DNMT3A* and *DNMT3B*. These differences were reproduced in repetitions of the mouse and sheep PCR experiments. This apparent difference in primer efficiency also justifies why it was deemed inappropriate to assign significance to the differences observed between *DNMT1*, *DNMT3A* and *DNMT3B* transcript levels within a single embryo. Overall, the analysis of larger numbers of oocytes and embryos will allow more confident interpretation of these results.

The analysis of *DNMT* expression levels in fetal fibroblasts presented an opportunity to investigate the quantities of *DNMT* transcripts in cells used as SCNT donors. Inappropriate transfer of the 'somatic' form of Dnmt1 within a somatic cell donor nucleus has been postulated as a possible reason for increased methylation levels observed in SCNT produced sheep blastocysts (Kang *et al.*, 2001; Bourc'his *et al.*, 2001; Dean *et al.*, 2001). This could indeed be an important factor, when considering the observed long retention of mouse Dnmt1 in oocyte cytoplasm (Cardoso & Leonhardt, 1999; Carlson *et al.*, 1992). Indeed, expression levels of *DNMT1* were found to be moderately high in sheep fetal fibroblast cells, the cells commonly used as nuclear donors for SCNT at Roslin Institute (R.I) and for the SCNT embryos analysed in this thesis (reported in Chapter 7). If the expression levels of *DNMT1* in sheep fetal fibroblast cells are indicative of protein levels, this could lead to the introduction of 'somatic' DNMT1 into nuclei of SCNT-derived preimplantation embryos. However, as sheep oocytes already contain large quantities of 'somatic' *DNMT1* mRNA, and presumably DNMT1 protein, introduction of large quantities of DNMT1 from the SCNT donor nuclei may have little, or no consequence. It may, however, be a potential source of aberrant methylation reprogramming in other mammalian species.

The significance of high levels of *DNMT3A* expression in fibroblast cells, in relation to SCNT, is unclear. If DNMT3A protein levels are similarly high in fetal fibroblast nuclei, the developing SCNT embryo may be inappropriately exposed to large quantities of this protein. *DNMT3A* mRNA appears to be present in significant quantities in the sheep and mouse oocyte and therefore the protein may also be present in significant amounts. However, as nothing is known of the roles or localisation of DNMT3A in oocytes or preimplantation embryos, presumptions cannot be made regarding the effects of an SCNT-caused increase in protein levels.

Chapter 6

Cloning and characterisation of the gene encoding the methyl-CpG-binding protein MBD2 in sheep

6.1 Introduction

Methyl-CpG-binding protein 2 (Mbd2) is a mammalian protein that demonstrates affinity for methylated DNA, and shares homology with other known methyl-CpG binding proteins (Hendrich & Bird, 1998) including Methyl-CpG Binding Protein 2 (MeCP2). This protein, alongside Mbd1, Mbd3 and Mbd4, was initially identified during a database search for proteins containing the methyl-CpG binding domain first identified in the transcriptional repressor, MeCP2 (Nan *et al.*, 1993). The methyl-binding proteins interact with methylated DNA via the highly conserved MBD (Methyl-CpG-binding domain) motif, and all MBD-containing proteins, aside from Mbd3, demonstrate high affinity for methylated DNA (Hendrich & Bird, 1998). These MBD-containing proteins are of significant interest due to their potential involvement in methylation-mediated transcriptional repression. Of the four recently identified MBD proteins, Mbd2 is best characterised, and more interestingly, has generated controversy over its putative roles in response to the DNA methylation signal

Structurally, the human and mouse *MBD2* genes are similarly organised, comprising 7 exons and a sex-specific exon located in intron 2 with undetermined function (Hendrich *et al.*, 1999b). In addition, the mouse has two forms of Mbd2. Mbd2a encodes a protein of 414 aa whilst Mbd2b encodes a smaller protein of 262 aa (Hendrich & Bird, 1998). Mbd2b is closely related to Mbd3 and shares significant sequence identity. **Figure 6.1** highlights the structures of Mbd2 and Mbd3, including the alternative Mbd2 splice variant.

Figure 6.1 Structural comparison of Mbd2 and Mbd3 proteins.
All three proteins share the highly conserved MBD domain, highlighted as a black rectangle. The size of each protein is indicated at the right hand side of the diagram.

As mentioned previously, the involvement of Mbd2 in transcriptional repression has been studied in the mouse, but has not been a straightforward case. In early 1998, a paper identified the human MBD2B as a protein with inherent demethylating activity, the elusive mammalian CpG 'demethylase' (Bhattacharya *et al.*, 1999). Previous studies aiming to identify mammalian demethylases had failed to isolate any convincing candidate genes (Weiss *et al.*, 1996; Swisher *et al.*, 1998). As demethylation is a fundamental event in reprogramming methylation patterns during post-fertilisation development, the identification of MBD2B as a demethylase was a potentially significant discovery. However, subsequent reports implicated Mbd2 in transcriptional repression, as part of the MeCP1-HDAC (Histone deacetylase) complex involved in silencing methylated DNA (Ng *et al.*, 1999). As demethylation is a process expected to be associated with gene activation, not gene-silencing, this discovery cast doubts over the initial 'demethylase' study. Moreover, attempts to re-create the experiments described by Bhattacharya and colleagues, demonstrating the demethylating activity of Mbd2, failed when attempted in another laboratory (Ng *et al.*, 1999; Boeke *et al.*, 2000).

In the light of these observations, it was suggested that Mbd2 could be a protein with dual silencing and activating activities, possibly acting as a molecular switch

between active and silenced chromatin states (Ng *et al.*, 1999). However, recent experiments have ruled out the involvement of Mbd2 in active demethylation during mouse development. Santos and colleagues (2001) demonstrated normal demethylating activities in embryos derived from mouse oocytes carrying a targeted deletion of Mbd2. Moreover, mice carrying a targeted deletion of Mbd2 develop to term and demonstrate no difference in global methylation levels when compared to their normal counterparts (Hendrich *et al.*, 2001).

Mbd2 is expressed in a wide range of adult tissues, but is significantly down-regulated in ES cells (Hendrich & Bird, 1998), Even though ES cells generally maintain high levels of genomic methylation, ES cells carrying a targeted deletion of *Dnmt1* are viable (Lei *et al.*, 1996), demonstrating a non-requirement for the maintenance of DNA methylation and subsequently may not require Mbd2 activity. The requirement for Mbd2 in preimplantation development is unknown and no investigations into *Mbd2* expression during preimplantation development have been reported to date.

During the initial stages of this project, the identification of Mbd2 as a demethylase provided a candidate gene for analysis in sheep development. By the time reports implicating Mbd2 in transcriptional repression became available, substantial progress had been made in the cloning of the sheep homologue of MBD2. Given the new information, it was decided to continue with analysis of Mbd2 as an interesting component of the sheep DNA methylation repressor system.

In this study, the partial cloning of a sheep homologue of *MBD2*, using a combination of RT-PCR and cDNA library screening is described. Pairwise and multiple sequence alignments were generated analysing the conservation of the *MBD2* sequence between sheep, humans and mice, and assessing similarities with other closely related methyl-CpG binding proteins. Expression analysis of the *MBD2* gene in adult sheep tissues was performed using Taqman® real-time RT-PCR to determine its requirement in differentiated tissues. Analysis of *MBD2* transcript levels was also performed in sheep preimplantation embryos, being the first reported

study to investigate *MBD2* expression during mammalian preimplantation development.

6.2 Results

6.2.1 Generation of partial sheep *MBD2* clones by RT-PCR

Pairwise alignments of the human (Accession No. AF072242) and mouse (Accession No. NM010773) *MBD2* cDNA sequences were generated and primers were designed within highly conserved regions between the human and mouse cDNA sequences. Partial cDNA clones were generated from pooled heart and kidney cDNA in a standard PCR reaction performed (as described **2.4.1.1**). PCR products were electrophoresed in a 2% ethidium bromide stained agarose gel (as described **2.4.2**), results are shown in **Figure 6.2.** PCR products of expected size were observed (MBD 116 bp; DMT 168 bp) and were subsequently cloned and sequenced (as described **2.4.3**) to confirm identity. In addition to amplification of PCR products from cDNA, genomic DNA was used as the template to indicate the presence of exon-intron boundaries within the partial cDNA sequence. PCR reactions were performed as previously described. The MBD primers generated a 400 bp product from genomic DNA, indicating the primers spanned a small intron of approximately 280 bp. No PCR product was detected for DMT primers on genomic DNA, suggesting the presence of a larger intron within the region. **Table 6.1** lists the primers used and the location within the source sequence.

Gene	*Primer*	*Direction of primer*	*Sequence primer designed from (Genbank accession number)*	*Location of primer*
MBD2	MBDF	Forward	AF072242	700-719
MBD2	MBDR	Reverse	AF072242	815-796
MBD2	DMTF	Forward	AF072242	1040-1058
MBD2	DMTR	Reverse	AF072242	1208-1190

Table 6.1 PCR primers used to amplify partial MBD2 cDNA sequences in sheep.

The primer sequences can be found in **Table 2.3**.

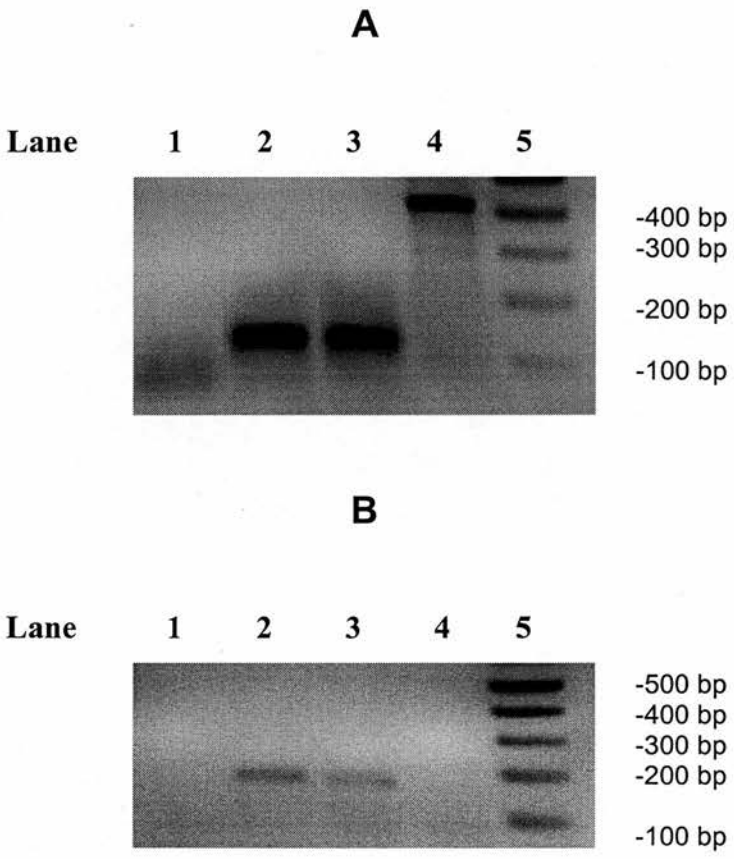


Figure 6.2 Amplification of partial *MBD2* cDNA clones by RT-PCR.

Gel image **A** demonstrates PCR products amplified using the MBD primers, image **B** demonstrates PCR products amplified using the DMT primers. In both images, lane 1 is a no template (negative) PCR control and lanes 2 and 3 are PCR products amplified from d125 fetal heart and liver cDNA respectively. Lane 4 is the PCR product amplified from genomic DNA. DNA marker sizes (Lane 5) are indicated to the right of the gel image.

6.2.2 Isolation of partial *MBD2* clones from an adult sheep lung cDNA library

The partial cDNA clones generated by the MBD and DMT primers were used to probe an adult sheep lung cDNA library purchased from Stratagene (as described **2.4.7)**. The primary screen identified three positive clones, 4a, 5a and 6a, which were re-plated for the secondary screen. After the secondary screen, three positive plaques were identified from the initial positive clones 4a (4a.1, 4a.2 and 4a.3) and one positive plaque from the positive clone 5a (5a.1). These 5 single colonies were isolated and the pBluescript phagemids excised from the Uni-ZAP® XR vector (as described **2.4.7.7**). DNA minipreps were established from 4 single colonies for each positive clone and digested with *EcoR* I and *Xho* I to release and ascertain the vector insert size (as shown in **Figure 6.3)** Clones 4a.1-4a.3 were found to be identical, containing an insert of approximately 0.8 Kb, clone 5a.1 contained an insert of approximately 1.3 Kb.

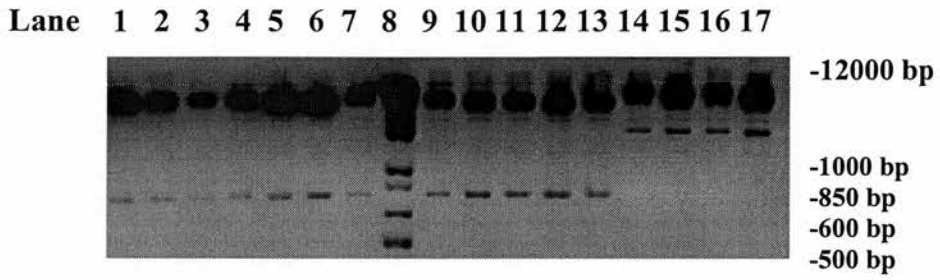


Figure 6.3 *EcoR I/Xho* I digestion of DNA minipreps prepared from the Uni-ZAP® XR clones 4a.1, 4a.2, 4a.3 and 5a.1

Lanes 1-4 represent minipreps established from clone 4a.1, lanes 5-7 and 9 represent minipreps established from clone 4a.2, lanes 10-13 represent minipreps established from clone 4a.3 and lanes 14-17 represent minipreps established from clone 5a.1. DNA marker sizes are indicated to the right of the gel image.

A single miniprep originating from the clone 4a.2, and a single miniprep originating from clone 5a.1 were ethanol precipitated and sequenced as described (**2.4.6; 2.4.3**). After sequence editing of the two clones, and assembly using the Staden fragment assembly program, 1224 bp of sequence corresponding to the sheep *MBD2* cDNA sequence was obtained, with substantial overlapping of the two clones. The sequence contained the PolyA tail and corresponded to nucleotides 749 to 1964 of the human MBD2 sequence (Accession No. AF072242). The partial *MBD2* cDNA sequence can be found in **Appendix III**. Pairwise alignment of the sheep and human or mouse cDNA sequences were performed using the 'Gap' application (default parameters) of GCG10. The partial sheep MBD2 cDNA sequence was found to be 90% homologous to the human *MBD2* sequence and 87% homologous to the mouse MBD2 cDNA sequence (Accession No. NM010773). The complete pairwise alignment for the sheep and human *MBD2* cDNA sequences can be found in **Appendix III**. Translation of the partial MBD2 sequence was performed using the 'translate' application on GCG10, and an open reading frame was identified, terminating at a TAA stop codon at nucleotide 724. A multiple protein sequence alignment was generated using the 'localpileup' application (default parameters) of GCG10, comparing the predicted sheep MBD2 protein sequence with the known human MBD2 (Accession No. AAD56597) and closely related human MBD3 (Accession No. NP003917) protein sequences. Shading of the multiple sequence alignment was added using the GeneDoc program (default settings). The multiple sequence alignment is shown in **Figure 6.4** and highlights the conserved methyl-CpG binding domain (MBD).

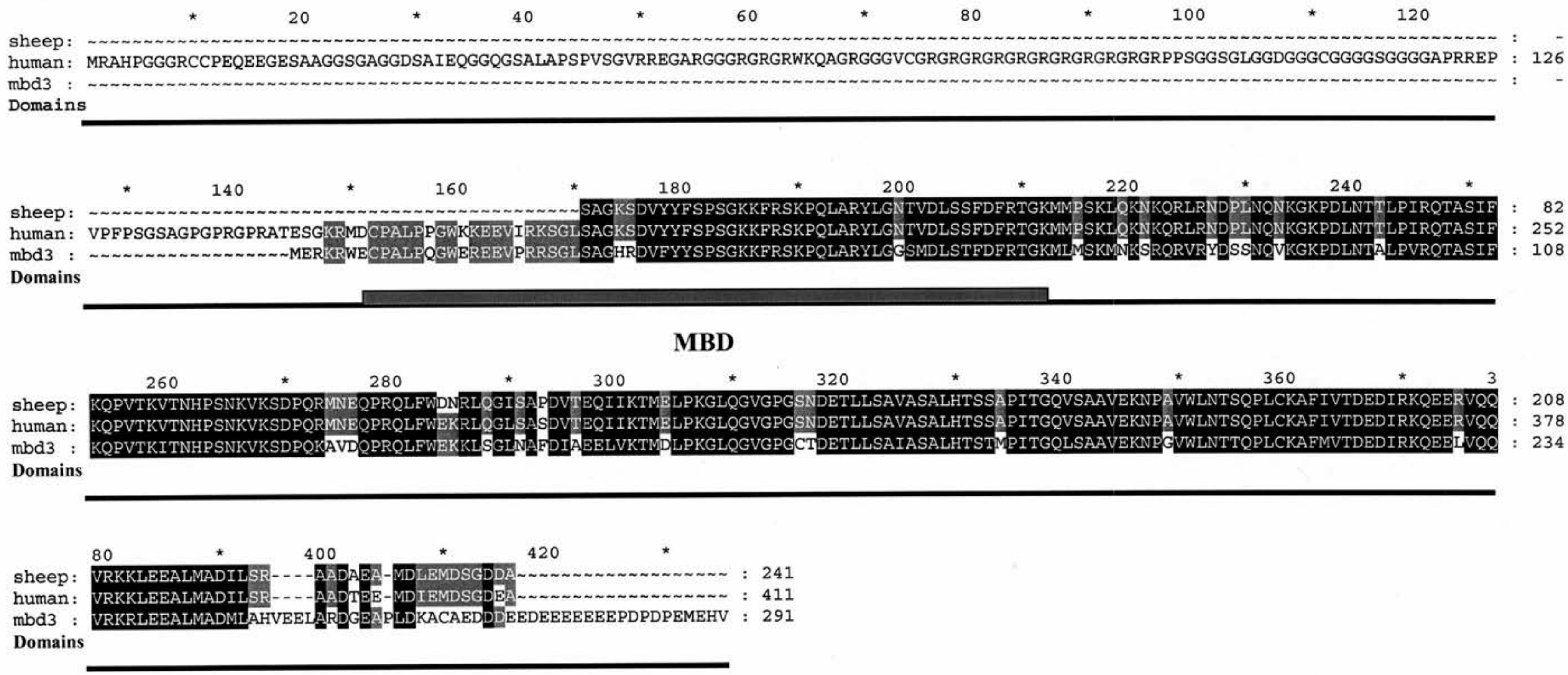


Figure 6.4 Multiple alignment of the sheep MBD2 and the human MBD2 and MBD3 protein sequences. Multiple sequence alignment was performed using the localpileup application on GCG10 and shaded was added using the Genedoc program (default settings). The functional domains are highlighted on the black line below the sequence alignment. The methyl-binding domain is represented by the labeled green rectangle.

6.2.3 Expression of *MBD2* in adult sheep tissues and fetal fibroblast cells

To quantify *MBD2* gene expression in adult sheep tissues and differentiated cells, Taqman® real-time PCR employing the Comparative Ct method of relative quantification was used, with 18S rRNA acting as the endogenous control. *MBD2*-specific primers and Taqman® probe were designed from the sheep *MBD2* sequence (as described **2.4.12.3)**; primer and probe sequences can be found in **Table 2.6**. Sufficient template for *MBD2* expression analysis remained in the cDNA samples derived from heart, kidney, lung, liver, ovary, spleen and fetal fibroblast cells described in the previous chapter (for analysis of *DNMT1*, *DNMT3a* and *DNMT3b* expression). As the no RT controls for each of these tissue/cell samples were checked for DNA contamination in the previous chapter, with no significant contribution of DNA contamination to the PCR signal demonstrated, it was decided to compare a pooled no RT control (of all the adult sheep tissue controls) with a pooled tissue cDNA sample, to prevent wastage of tissue cDNA stocks and the expensive Taqman reagents and consumables. The Ct values for *MBD2* expression in pooled cDNA tissue and the pooled no RT control are demonstrated in **Table 6.2**.

Well	Dye	Sample	Ct
1	FAM	NTC	40
2	FAM	NTC	40
3	FAM	NTC	40
4	FAM	cDNA	23.68
5	FAM	cDNA	23.07
6	FAM	cDNA	23.81
7	FAM	no RT	32.86
8	FAM	no RT	32.65
9	FAM	no RT	32.11

Table 6.2 Average Ct values for tissue cDNA and RT-PCR controls after MBD2-specific amplification.

Samples are clearly designated as cDNA (pooled cDNA), no RT (pooled no RT RNA samples) and NTC (no template control).

The no RT sample Ct value of 32 indicated some DNA contamination, but as this value is significantly higher (9-10 cycles = ~500-1000 fold difference in expression and corresponds to a 0.1-0.2% bias) than the Ct value of the pooled cDNA sample itself, it was deemed acceptable.

Prior to quantification experiments, the relative efficiencies of the *MBD2* primers/probe and 18S primers/probe sets were measured over a suitable working range of cDNA dilutions (using pooled cDNA from all tissue samples). The Taqman PCR reaction to assess relative primer efficiency was performed (as described **2.4.12.3)**, with all reactions performed simultaneously on the same plate. The optimal primer and probe conditions used for both genes are described in **Table 2.6**. ΔCt values were plotted against the log of input cDNA dilution and a line of best fit generated. The slope of the line was 0.08, demonstrating relative primer efficiencies within the accepted range (<0.1). The graph showing the relative linear efficiency of the two primer/probe sets in presented in **Figure 6.5**

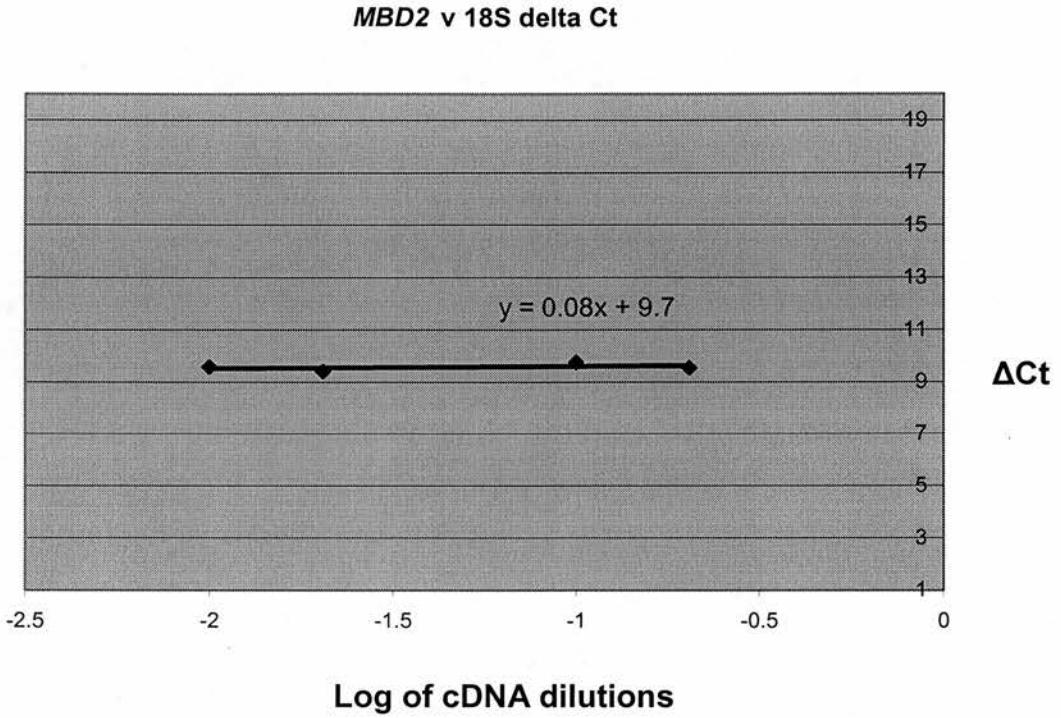


Figure 6.5 Evaluation of *MBD2* and 18S relative primer efficiency.

The Log of input cDNA dilutions are indicated on the x axis, and the delta Ct values (*MBD2* Ct minus 18S Ct) are indicated on the y axis. Over a suitable working range of cDNA dilutions, the *MBD2*/18S delta Ct values produced a line with slope of 0.08. This is <1.0 and therefore demonstrates acceptable relative primer efficencies.

After confirmation of equal relative efficiency of the *MBD2* and 18S primer/probe sets, quantification of gene expression in a range of tissues was performed. Dilutions of 1:10 and 1:100 for each tissue cDNA were prepared for expression analysis of *MBD2* and 18S respectively. The Taqman PCR was performed (as described **2.12.3)** with the *MBD2* and 18S PCR reactions performed simultaneously on the same plate. Optimal primer/probe conditions are described in **Table 2.6**. The PCR was repeated, and the mean of the two ΔCt values for each tissue was calculated and used in subsequent quantification of *MBD2* by the comparative Ct method (as described **2.12.2)**. The tissue with lowest expression was liver, and was therefore designated the 'calibrator' tissue (1 fold). The levels of expression of *MBD2* in the other tissues analysed are therefore expressed as *n-fold* increase relative to liver. The results of the quantification experiment are shown in **Table 6.3**

The Taqman® PCR results demonstrated expression of *MBD2* in all adult tissues analysed and the fetal fibroblast cells. *MBD2* expression is high in lung and spleen relative to liver, whereas heart and kidney tissue demonstrate similar levels to liver. Ovary tissue and fibroblast cells show moderate levels of expression.

Tissue	*MBD2* Ct	18S Ct	ΔCt	Average ΔCt	ΔΔCt ($\Delta Ct-\Delta Ct_{Liver}$)	MBD2 Rel. to Liver
Liver (1)	25.87	16.31	9.56	9.77+/- 0.29	0+/- 0.29	1
Liver (2)	26.15	16.17	9.98			(0.82-1.22)
Heart (1)	27.65	18.18	9.47	9.68+/- 0.29	-0.09+/- 0.29	1.16
Heart (2)	28.205	18.32	9.88			(0.87-1.3)
Kidney (1)	25.12	16.09	9.03	9.18+/- 0.21	-0.59+/- 0.21	1.5
Kidney (2)	25.48	16.15	9.33			(1.3-1.74)
Lung (1)	23.78	16.42	7.36	7.46+/- 0.13	-2.31+/- 0.13	4.95
Lung (2)	24.17	16.62	7.55			(4.53-5.42)
Ovary (1)	24.97	16.21	8.76	8.96+/- 0.28	-0.81+/- 0.28	1.75
Ovary (2)	25.04	16.24	9.16			(1.44-2.13)
Spleen (1)	23.19	16.03	7.16	7.6+/- 0.62	-2.17+/- 0.62	4.5
Spleen (2)	24.17	16.13	8.04			(2.93-6.92)
Fibroblast (1)	24.7	16.54	8.16	8.34+/- 0.25	-1.43+/- 0.25	2.69
Fibroblast (2)	25.14	16.63	8.51			(2.27-3.2)

Table 6.3 Relative quantitation of *MBD2* using the Taqman® Comparative Ct method.

Data collected from the two separate experiments is designated as experiment (1) or experiment (2) in the table. The *MBD2* and 18S Ct values are the mean Ct values gained in triplicate PCR reactions. ΔCt demonstrates the average ΔCt value of the two experiments, and the standard deviation (s.d.) of the two Ct values is shown. The ΔΔCt calculation involved the subtraction by the chosen calibrator tissue, liver. *MBD2* expression levels are determined relative to liver. The range of expression given for each tissue (in brackets) was determined by solving the equation $\mathbf{2}^{-\Delta\Delta Ct}$, using ΔΔCt + s.d or ΔΔCt – s.d.

6.2.4 *MBD2* expression in preimplantation embryos.

To assess the importance of *MBD2* expression in preimplantation development, RT-PCR was performed on cDNA derived from either single 2-cell or morula stage embryos. These stages were chosen as they represent stages pre- and post- zygotic genome activation. RNA was extracted from 5 single *in vitro* fertilised and cultured 2-cell embryos or morulae, and reverse-transcribed (as described **2.4.11.2**). In addition, a 'blank' sample (No Embryo Control) was carried through the RNA extraction and cDNA synthesis steps, to ensure the reagents and plastic consumables were not contaminated. Half of the single embryo cDNA was used as template in the subsequent PCR reaction using the 'MBD' *MBD2*-specific primers (as described **2.4.1.1**) that would in theory amplify all forms of *MBD2*. As the MBD primers span an intron, DNase treatment of the embryos was not required as any DNA contamination would produce an alternative PCR product. Primer-specific PCR conditions are described in **Tables 2.1** and **2.2**. PCR products were electrophoresed in a 2% ethidium bromide stained agarose gel (as described **2.4.2**), results are shown in **Figure 6.6**.

Results of the PCR indicated the presence of *MBD2* mRNA in 2-cell and morula stage sheep embryos, albeit barely detectable at both stages.

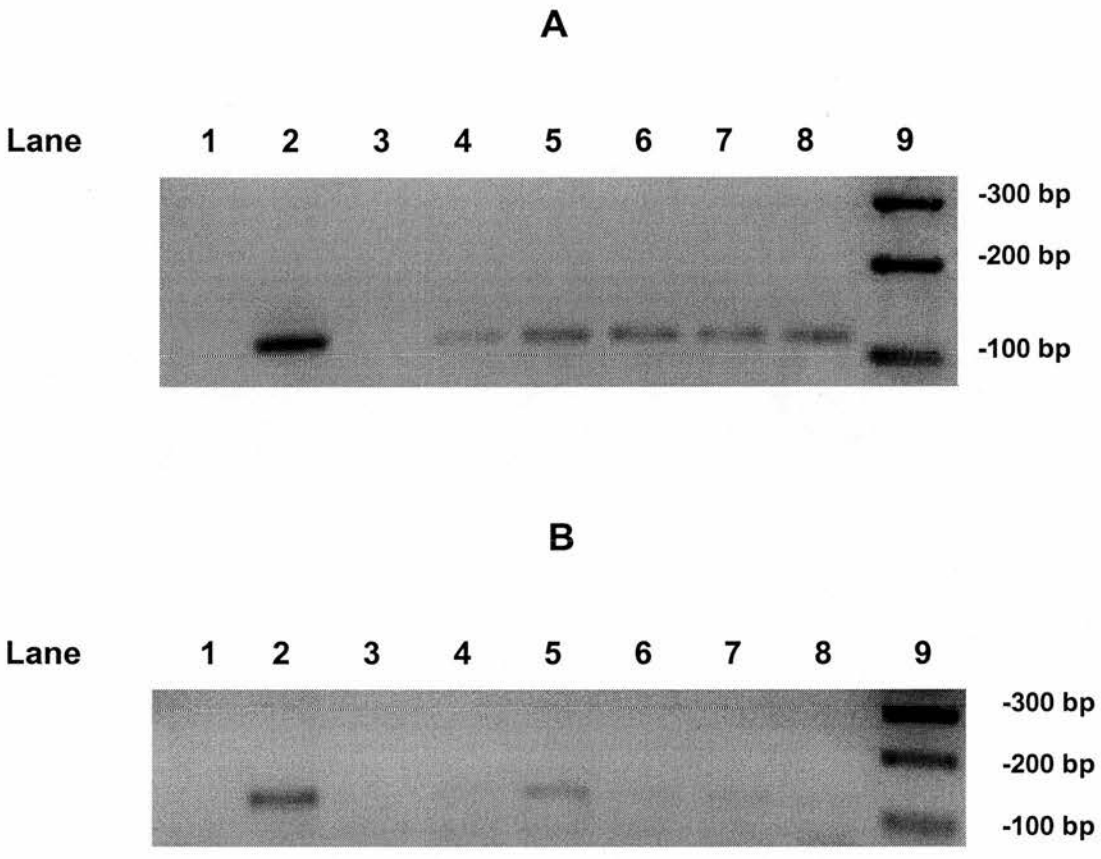


Figure 6.6 ***MBD2*** **mRNA transcripts in single 2-cell and morula stage sheep embryos.**

Gel **A** shows PCR products generated from 2-cell embryos, Gel **B** shows PCR products generated from morula stage embryos. In both images, Lane 1 is a negative 'no template control' PCR reaction. Lane 2 is a positive control PCR product amplified from 2 µl of d125 fetal kidney cDNA. Lane 3 is a second negative control, with the No Embryo Control used as template in the PCR reaction. Lanes 4-8 are PCR products amplified from cDNA derived from 5 single 2-cell stage embryos (gel image **A**) or 5 single morula stage embryos (gel image **B**). DNA marker sizes (lane 9; 1Kb plus DNA ladder (Invitrogen)) are indicated to the right of the gel images.

6.3 Discussion

This study has described the isolation of a sheep partial cDNA clone homologous to the *MBD2* gene through a combination of RT-PCR and cDNA library screening. The cDNA clone corresponds to nucleotides 749 to 1964 of the human cDNA sequence, and includes the 3' UTR and PolyA tail. Pairwise alignments of the partial sheep *MBD2* cDNA clone with the known human and mouse *MBD2* sequences demonstrated high levels of conservation. The sheep *MBD2* clone shared 90% homology with the human *MBD2* cDNA sequence, and 87% sequence identity with the mouse *Mbd2* cDNA. Conceptual translation of the sheep MBD2 cDNA sequence identified an open reading frame, terminating at a TAA codon located at nucleotides 724-726. The predicted sheep MBD2 protein sequence was compared with the human MBD2 and MBD3 sequences in a multiple sequence alignment. The predicted sheep MBD2 protein appears to be very highly conserved with the human MBD2, with 235/241 amino acid residue sequence identity (98% homology).

As the full-length cDNA sequence was not gained for the sheep homologue of *MBD2*, it is possible that the sheep clone isolated from the cDNA library could correspond to either Mbd2a or Mbd2b. These isoforms are identical aside from the extended N-terminal region of Mbd2a and both perform simliar functions. However, as the main aim of this study was to gain sufficient sequence to be confident of the sheep Mbd2 identity, to assess general conservation levels and subsequently perform expression analysis, it was decided not to pursue full-length cloning of *MBD2*.

Comparison of the MBD2 sequences with MBD3 demonstrates high levels of conservation between these two closely related methyl-CpG binding proteins. It was important to rule out the possibility that the clone isolated from the sheep lung cDNA library could actually be Mbd3, as the predicted sheep MBD2 protein sequence was of a similar size to MBD3 and the human MBD2B and MBD3 proteins are known to share 70% homology (Hendrich & Bird, 1998). Protein sequence comparison of the sheep MBD2 and human MBD3 demonstrates high

levels of conservation within the methyl-binding domain, but the extreme c-terminal regions are very different, with the sheep MBD2 demonstrating strong similarities with the human MBD2.

The similarity of Mbd2 and Mbd3 at the protein level does not denote identical roles. Firstly, the two proteins differ in their ability to bind methylated DNA (Mbd2 can bind methylated DNA, Mbd3 does not) even though they both contain the MBD motif (Hendrich & Bird, 1998). In terms of development, Mbd3 appears to be essential, with mice carrying a targeted deletion of *Mbd3* failing to progress through early stages of embryogenesis (Hendrich *et al.*, 2001). On the contrary, mice carrying a targeted deletion of *Mbd2* are viable, although demonstrate maternal-related behavioural defects (Hendrich *et al.*, 2001). However, a certain extent of cooperation between these two proteins has been demonstrated, with biochemical analysis indicating an indirect interaction between the two proteins, such as Mbd2 recruiting the Mbd3-containing NuRD complex (involved in transcriptional repression) to methylated DNA (Zhang *et al.*, 1999).

To further characterise the sheep homologue of *MBD2*, gene expression was quantified in sheep adult tissues using real-time PCR methodology, and measured in relation to liver, the tissue demonstrating lowest levels of expression. All adult tissues and fetal fibroblast cells analysed demonstrated *MBD2* expression, with levels varying between tissues. Lung and spleen tissue demonstrated the highest levels of expression, with a 4-fold increase relative to liver. Heart and kidney demonstrated levels comparable to liver, whilst ovary tissue and fetal fibroblasts demonstrated moderate levels of expression. In relation to Mbd2 expression observed in the mouse, the expression patterns in sheep appear to be similar. Spleen and lung demonstrate high levels of expression in both species, whilst liver demonstrates low levels, suggesting functional conservation. *MBD2* expression data for human tissues was not available for comparison.

Comparisons of *MBD2* and *DNMT1* expression in adult tissues were also made (*DNMT1* expression analysis results can be found in Chapter 5), with similar

patterns of expression observed. Tissues with high levels of *MBD2* also demonstrated high levels of *DNMT1* and vice versa. As Mbd2 binds to and mediates the transcriptional repression of methylated DNA, differentiated tissues with high levels of methylation and cellular turnover (indicated by high DNMT1 expression levels) may require similarly high levels of Mbd2. A more direct interaction of Dnmt1 and Mbd2 has also been suggested by Tatematsu and colleagues (2000) with a MBD2/MBD3 complex recruiting Dnmt1 to the replication fork, allowing maintenance of transcriptionally repressed chromatin after DNA replication. Comparison of *DNMT1* and *MBD2* expression in ovary tissue provided the one exception to the rule. Although the ovary demonstrates high levels of Dnmt1 expression, a significant proportion of this may be located in the growing or non-growing oocytes. These cells are not dividing, and the Dnmt1 is laid down as a store (non-functioning), thus Mbd2 may not be required to 'interpret' the methylation signal as for differentiated or somatic cells.

In addition to analysis of adult tissues, *MBD2* transcript levels were also analysed in sheep preimplantation stage embryos to determine its requirement during early stages of development. Embryos from the 2-cell and morula stages of development were chosen for analysis due to their timing pre- and post- zygotic genome activation at 8-16 cell stage. *MBD2* was detected in all 2-cell embryos analysed, albeit at apparently low levels. *MBD2* was barely detectable in morulae, and analysis of blastocyst cDNA using an alternative primer pair (the DMT primers; data not shown) resulted in a similar observation.

Since *MBD2* is observed at the 2-cell stage, before embryonic genome activation, it must have been present as a maternal store in oocytes. This suggests a possible requirement for this protein during the late stages of oocyte maturation or the very early stages of preimplantation development. The barely detectable presence of *MBD2* at the morula and blastocyst stages would suggest *MBD2* is not expressed from the embryonic genome prior to the blastocyst stage; detected transcripts are most likely from residual maternal stores. The very low levels of *MBD2* seen at the morula/blastocyst stages of development fits well with observations made in the

mouse where MBD2 is downregulated in ES cells, or cells lacking the requirement for DNA methylation. As MBD2 is a methylation-mediated transcriptional repressor, it would most likely not be required in preimplantation development when methylation patterns are undergoing significant changes, and totipotency is required.

By cloning and characterising a component of the DNA methylation repressor system, further insights have been gained into the role of DNA methylation in sheep. As shown with the DNA methyltransferases, the *MBD2* gene and predicted protein sequence are highly conserved between mammals, as are indeed the expression patterns in adult tissues. These results have also indicated a possible requirement for *MBD2* during oocyte growth and the early stages of sheep preimplantation development.

Chapter 7

Analysis of genomic DNA methylation levels in normal preimplantation development: comparisons with cloned (SCNT) and *in vitro* fertilised (IVF) embryos

7.1 Introduction

DNA methylation levels are known to change significantly during mouse preimplantation development (Monk *et al.*, 1987; Howlett & Reik, 1991; Rougier *et al.*, 1998), reprogramming gamete-specific methylation patterns to support embryonic development. More recently, investigations performed in several other mammalian species (cow, pig and rat) indicated conservation to a certain extent of methylation reprogramming events during preimplantation development (Dean *et al.*, 2001). All species investigated demonstrate an active demethylation of the paternal genome post-fertilisation, with different extents of further demethylation and *de novo* methylation occurring during later stages of preimplantation development

The recent surge of interest in somatic cell nuclear transfer (SCNT) technology has led many to question the epigenetic reprogrammability of an adult nucleus when fused with an enucleated oocyte. The low success rates of SCNT and high incidence rates of fetal overgrowth, fetal abnormalities and pre-natal death (reviewed in Young & Fairburn, 1999), are widely blamed on the failure of the oocyte to correctly reprogram the somatic cell nuclei, thus preventing normal embryonic development. As DNA methylation is a key epigenetic modification undergoing significant changes post-fertilisation, it is possible that 'aberrant' methylation patterns acquired or maintained in the very early stages of development after SCNT may be recapitulated during subsequent stages of

development, and hence contribute to the unusual phenotypes observed in animals generated using this technology. Indeed, recent studies have demonstrated incorrect methylation patterns in cloned cow and pig preimplantation embryos (Dean *et al.*, 2001; Bourc'his *et al.*, 2001; Kang *et al.*, 2001a and Kang *et al.*, 2001b).

A major aim of this thesis was to determine whether the DNA methylation events described in mouse preimplantation development are conserved in sheep. An important application of these studies is to determine whether DNA methylation patterns of SCNT-derived embryos are equivalent to those seen in embryos produced *in vivo* or by IVF. Techniques were developed to analyse DNA methylation status of repetitive DNA sequences in single oocytes and preimplantation embryos. These methods were then employed to compare DNA methylation status in *in vivo*-derived, IVF or SCNT produced embryos. This chapter describes the design of the DNA methylation assay, observations made of DNA methylation status in 'normal' embryos and comparisons of DNA methylation status made between embryos derived from different embryo production methods.

7.2 Methodology

7.2.1 Methods of DNA methylation analysis

There are several commonly reported methods of performing genomic DNA methylation analysis for specific DNA sequences. One of the simplest methods involves a combined approach of restriction enzyme analysis of DNA with methylation sensitive isoschizomeric enzymes (e.g., *Hpa* II/*Msp* I) and PCR (Singer-Sam *et al.*, 1990), to analyse the methylation status of specific cytosine residues within a known DNA sequence. This technique has been widely used in the past, but is limited by the number of residues analysed in a single experiment. Currently, the most commonly used techniques involve bisulfite treatment of DNA, a process that converts unmethylated cytosine residues to uracil (Frommer *et al.*, 1992). The methylation status of multiple specific CpG dinucleotides can then be assessed using a number of different methods. Bisulfite treated DNA can be amplified by PCR to obtain specific regions of genes, and these PCR products are then cloned and sequenced to determine the methylation status of the cytosine residues within the sequence. This method is labour intensive as it requires the sequencing of many clones to differentiate between the normal variability within a population of cells and possible incomplete conversion of bisulfite treated DNA.

An alternative method using bisulfite treatment involves the design of PCR primers that will amplify only 'converted' (unmethylated) DNA, allowing the assessment of the methylation status for a number of CpG dinucleotides within the primer sequence itself. Bisulfite treatment of DNA can also be combined with restriction enzyme analysis, using enzymes that will only recognise the converted DNA sequence (COBRA; Xiong & Laird, 1997). However, all bisulfite techniques have inherent problems. Attainment of complete conversion of the sequence of interest is clearly desirable, but is known to be problematic, especially within CpG rich sequences and DNA containing unusual secondary structures (Rother *et al.*, 1995). Quantification of methylation levels can also

be problematic when using primers to amplify converted C-rich sequences. It has been demonstrated that primers used for this form of methylation analysis do not always amplify methylated and unmethylated DNA proportionately (Warnecke *et al.*, 1997).

For this particular study it was desirable to perform genomic methylation analysis on single embryos and oocytes to take into account the variability expected within a population. By pooling embryos, the variations that exist between individual embryos are not measurable and therefore the data provides a less accurate picture of the typical situation. Thus, greater numbers of embryos are required for methylation analysis. Instead, a method combining the use of a methyl-sensitive restriction enzyme *Hha* I (digests only unmethylated cytosine residues) and PCR, was adapted to allow quantification of methylation status for repetitive DNA sequences in single embryos and oocytes.

7.2.2 Selection of sheep repetitive sequences for methylation analysis

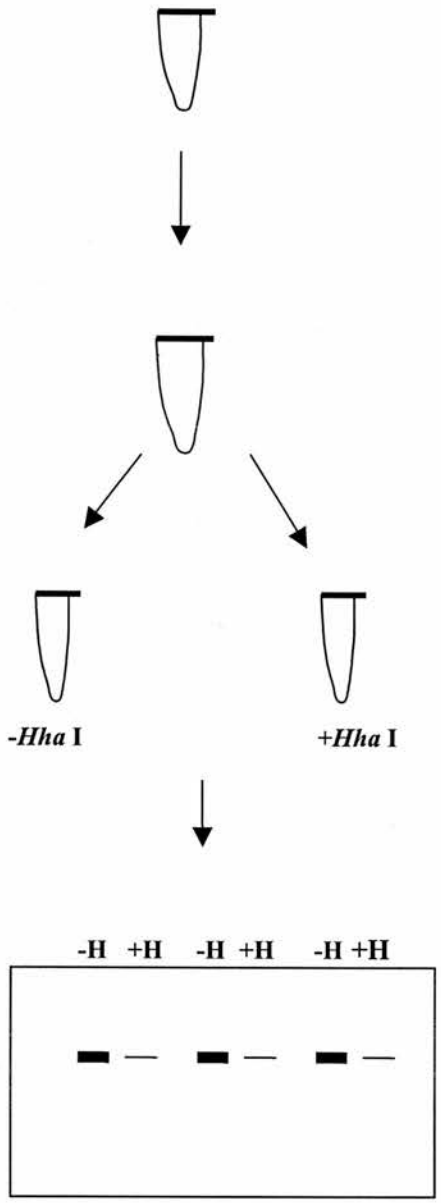
Two repetitive sequences were chosen for analysis; the sheep Satellite I sequence and an sheep homologue of the sheep SINE element BOV-2 (henceforth referred to as OV-2). The sheep Satellite I is found predominantly in centromeric regions of acrocentric chromosomes, but is also found to a much lesser extent on metacentric and sex chromosomes (Kurnit *et al.*,1978; Novak 1984; Burkin *et al.*, 1996). The BOV-2 SINE element is known to have approximately 100,000 copies (Lenstra *et al.*, 1993) and is dispersed throughout the genome. The OV-2 SINE has not been characterised in terms of copy number and location, although the ability to amplify the OV-2 SINE element from small amounts of DNA at very low PCR cycle numbers (24 cycles when using $1/10^{th}$ of a blastocyst's DNA) would indicate a relatively high copy number. SINE elements are a form of non-autonomous transposable element known to habituate multiple locations throughout the genome (Prak & Kazazian, 2000). The different locations of these two sequences allowed determination of methylation levels of two very different repetitive sequences within the sheep genome. This was aimed to indicate

whether different compartments of the genome are methylated at different levels during sheep preimplantation development, giving us an insight into the global methylation events occurring at this time.

Finally, DNA methylation analysis of a single CpG dinucleotide within a sequence will not give an accurate representation of the methylation status of the entire Satellite I or SINE element DNA sequence. However, it will allow comparison of the methylation status of the single site to be made between different embryonic stages and different embryo treatments.

7.2.3 Methyl-sensitive PCR (MS-PCR) method

The method chosen for analysis of OV-2 SINE and the sheep Satellite I DNA sequences utilised the design of primers spanning a single *Hha* I site for methylation analysis. The method was developed in order that each embryo had a 'standard' (the undigested DNA PCR product, representing 100% methylation) to directly compare the *Hha* I digested PCR product with. **Figure 7.1** outlines the complete method from DNA extraction to the quantification of the resultant PCR products.



DNA extracted from single embryos by repeat freeze-thawing and proteinase K digestion.

DNA ethanol precipitated using SeeDNA™ (APBiotech) and resuspended in 6µl dH_20.

DNA divided into two aliquots; one digested with *Hha* I (+*Hha* I; 20µl volume), the second aliquot remains undigested (-*Hha* I; *Hha* I-free 20µl digest).

Sequence-specific MS-PCR: 4 replicate PCR reactions per digested/undigested single embryo DNA were performed, each using 2 µl template DNA. The intensity of the *Hha* I-digested DNA PCR product was directly compared to the undigested DNA PCR product (representing 100% methylation), to ascertain the % methylation of the *Hha* I site within the DNA sequence.

Figure 7.1 Outline of methyl-sensitive PCR method.

7.3 Development and optimisation of MS-PCR assay

7.3.1 Restriction mapping and primer design for the sheep Satellite I sequence

The sheep Satellite I sequence (Accession No. Z18540), was mapped for *Hha* I sites using the Map function on GCG10 and primers spanning a single *Hha* I site were designed using the Prime function of GCG10. **Figure 7.2** shows the DNA sequence of the sheep Satellite I, highlighting the *Hha* I site used for methylation analysis, and the position of MS-PCR primers. Initial PCR products generated using the Satellite-specific primers (OVSAT1A), were subcloned and sequenced (as described **2.4.3**) to verify insert identity.

7.3.2 Cloning, restriction mapping and primer design for the OV-2 SINE DNA sequence

The OV-2 SINE element was isolated during a genomic library screen intended to clone a differential methylated region of the sheep *IGF2R* gene (an experiment performed independently of this thesis). Sequencing of a 4 Kb sub-clone isolated during this screen demonstrated the presence of a 556 bp sequence highly homologous to the BOV-2 SINE (Accession No. X64125.1) element. The OV-2 sequence was restriction mapped using the 'Map' application of GCG10 identifying the characteristic *Pst* I sites flanking the SINE element, and a single *Hha* I site for potential use in later methylation analysis experiments. The 'Prime' application of GCG10 was used to identify PCR primer positions flanking the *Hha* I site and the primers named BOVB forward and reverse. The sequence obtained is demonstrated in **Figure 7.3** and highlights the *Pst* I and *Hha* I cutting sites, and the positions of PCR primers. The initial PCR products generated by the BOVB primers were cloned and sequenced (as described **2.4.3**), to confirm sequence identity.

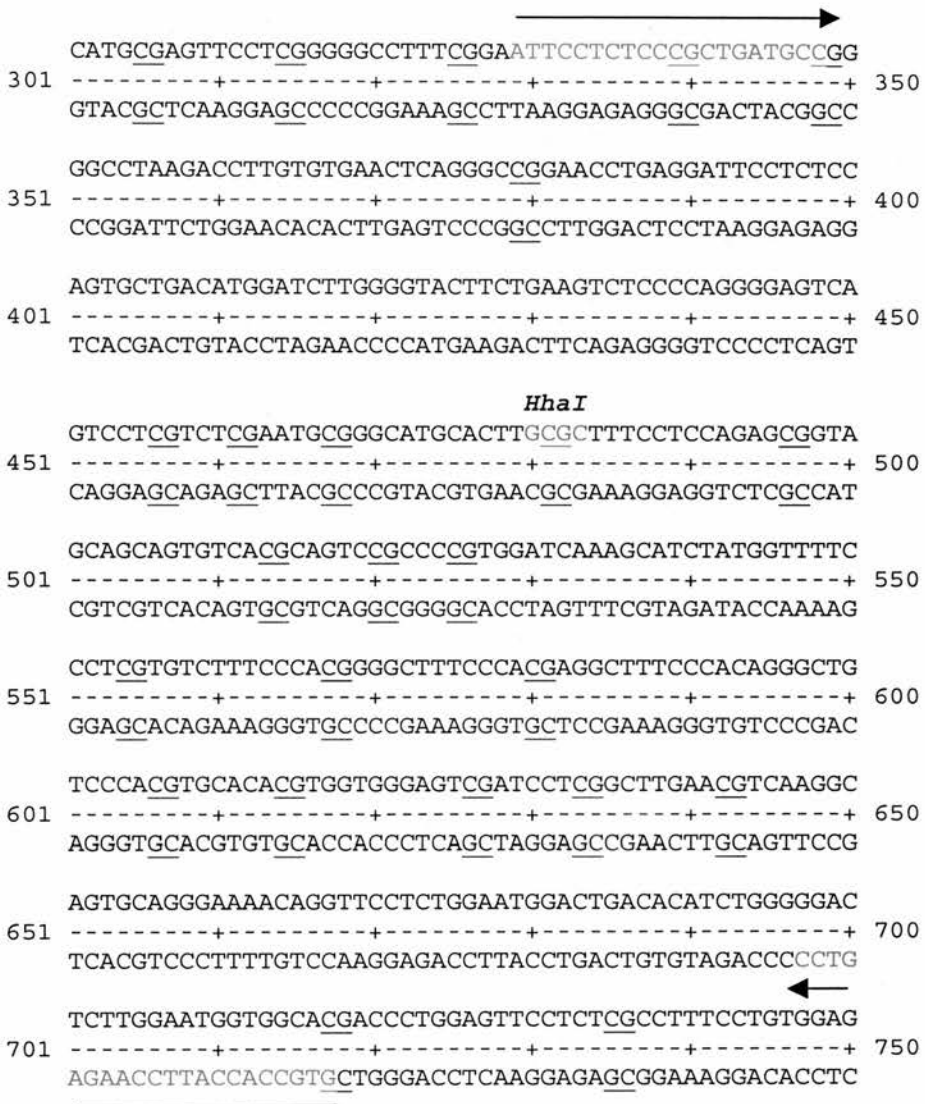


Figure 7.2 Sheep Satellite I DNA sequence highlighting *Hha* I enzyme cutting sites and positions of OVSAT1A MS-PCR primers. The *Hha* I restriction enzyme cutting sites are highlighted in blue text. The MS-PCR primer (OVSAT1A) positions are highlighted in red text and labelled with arrows. Other CpG sites are underlined, indicating the density of CpG dinucleotides within the Satellite I DNA sequence.

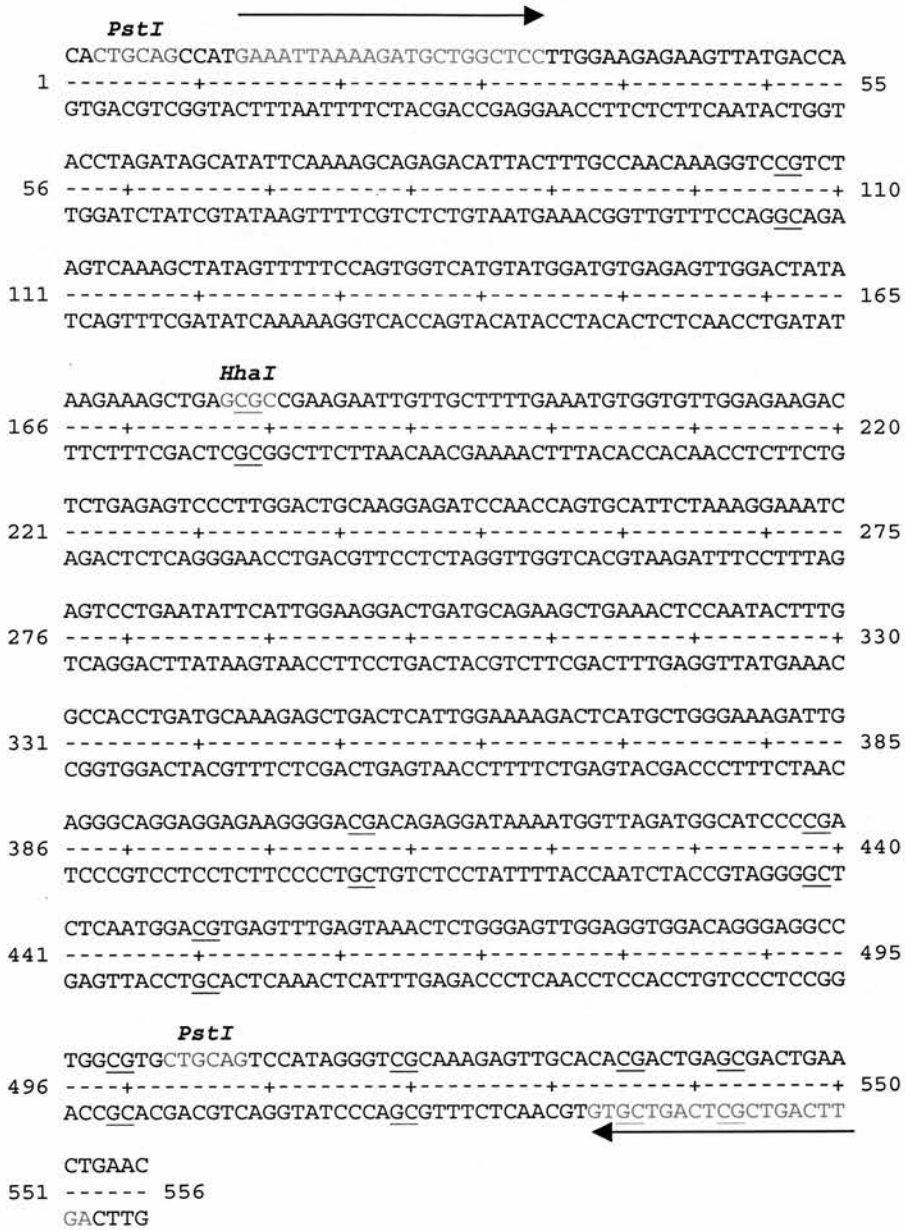


Figure 7.3 OV-2 DNA sequence indicating *Hha* I cutting sites and positions of BOVB MS-PCR primers. The *Pst* I and *Hha* I restriction enzyme cutting sites are highlighted in blue text. The MS-PCR primer (BOVB) positions are highlighted in red text and labelled with arrows. Other CpG sites are underlined, indicating the density of CpG dinucleotides within the Satellite I DNA sequence.

7.3.3 MS-PCR controls and assessment of reproducibility

Several controls were performed for each MS-PCR experiment relating to various steps of the method. Firstly, a No Embryo Control (NEC; dH_20) was initiated at the DNA extraction step and carried through the entire method, to ensure the reagents and plastic consumables were not contaminated. Controls were also required to ensure that the *Hha* I enzyme (present in excess), completely digested the unmethylated DNA. An unmethylated PCR product amplified by either the OVSAT1A1A or BOVB primers, and then gel extracted and purified (see **2.4.3**), was used for this particular control. To ensure the *Hha* I enzyme was not digesting methylated DNA, a plasmid containing either the Satellite I or OV-2 SINE PCR product was methylated *in vitro* using *Sss* I methylase (as described **2.4.10.4)**. Both the unmethylated and methylated controls were digested alongside the embryo/oocyte DNA and subsequently used as template in a PCR reaction with the appropriate primer pair. The PCR products were visually checked to confirm complete digestion of unmethylated controls and no digestion of methylated plasmid DNA. Controls established for Satellite I and OV-2 SINE MS-PCR analysis are demonstrated in **Figure 7.4**.

It was decided to perform 3-4 PCR replicates per digested/undigested single embryo DNA sample for each primer set, and the reproducibility of the MS-PCR method was found to be very high. **Figure 7.5** demonstrates the reproducibility for PCR replicates (performed at 19 cycles) for a single blastocyst undergoing methylation analysis of the Satellite I and OV-2 SINE sequences. Data derived from single replicates was only removed when a clear difference was observed between triplicate/quadruplicate values, for example, due to pipetting errors.

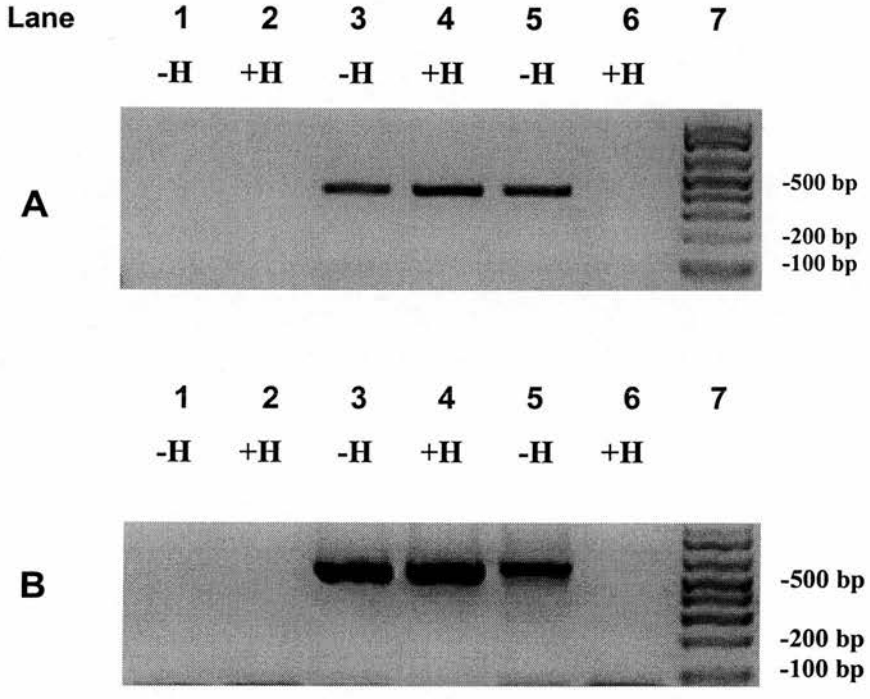


Figure 7.4 PCR analysis of controls used for (A) Satellite I or (B) OV-2 SINE methylation analysis.

The –H or +H above each lane indicates the absence or presence (respectively) of *Hha I* in the digestion step of the MS-PCR method. Lanes 1 and 2 are PCR reactions demonstrating no amplification of the No Embryo Control. Lanes 3 and 4 are PCR products generated from the methylated control template. Lanes 5 and 6 are PCR products generated from the unmethylated controls template. DNA marker sizes (Lane 7; 1 Kb Plus DNA ladder; Invitrogen) are indicated to the right of the gel image.

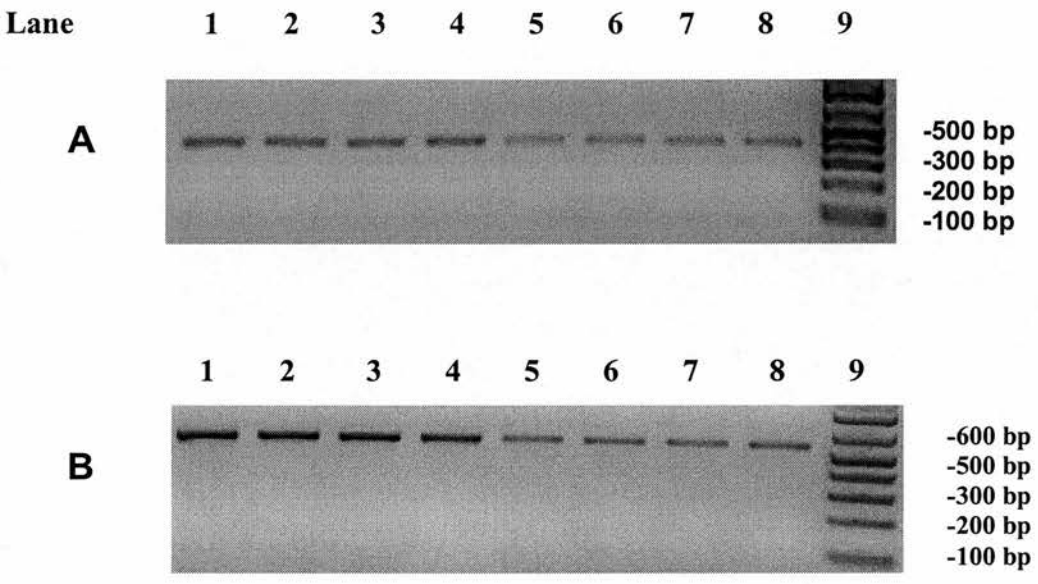
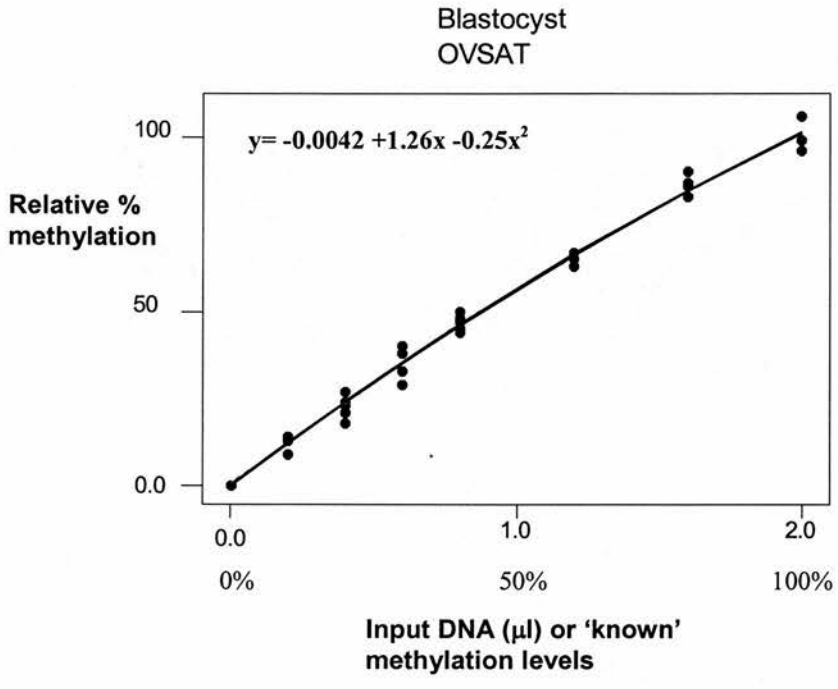


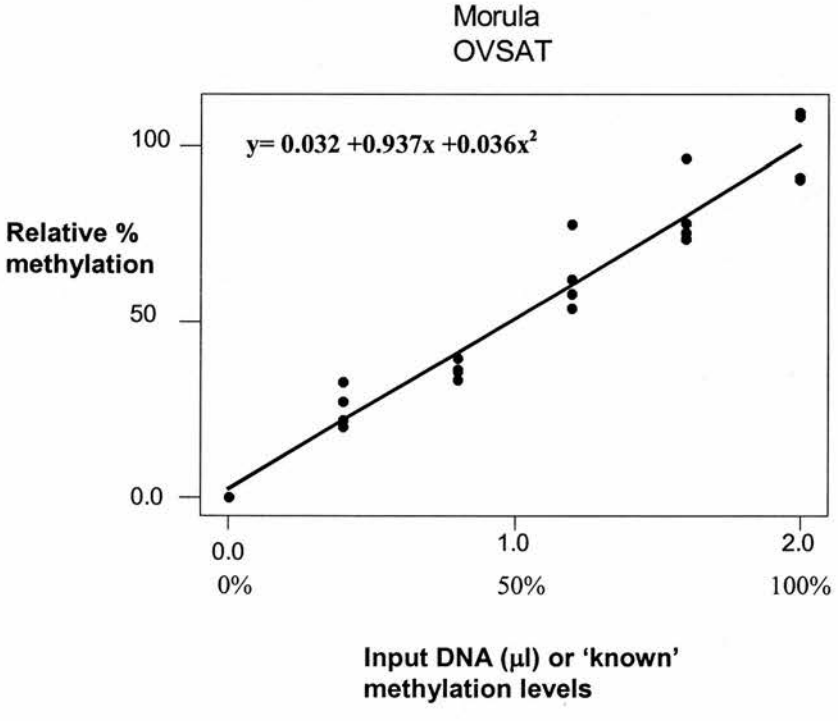
Figure 7.5 PCR replicates for DNA derived from a single blastocyst using the Satellite 1 (OVSAT1A; gel image A) or OV-2 SINE (BOVB; gel image B) primers. Lanes 1-4 are PCR products generated from undigested blastocyst DNA. Lanes 5-8 contain PCR products generated from *Hha* I digested blastocyst DNA. DNA marker sizes (Lane 9; 1 Kb plus DNA ladder; Invitrogen), are indicated to the right of the gel image.

7.3.4 Generation of standard curves for MS-PCR analysis

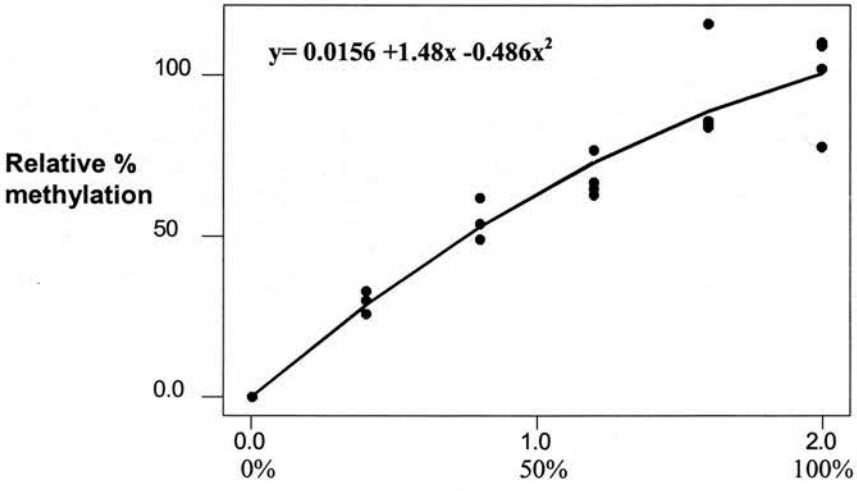
To define the linearity of the MS-PCR amplification within the PCR cycle used for quantification, standard curves were established to analyse DNA standards representing various levels of methylation. Standard curves were generated for both sets of primers, and for each embryo stage analysed, using DNA extracted and pooled from 3-4 oocytes, 8-16 cell embryos, morulae or blastocysts, as appropriate (DNA extraction performed as described **2.4.10.1**). The PCR was performed at the same cycle number as the embryo MS-PCR, using standards containing known quantities of undigested DNA equivalent to 2μl embryo/oocyte DNA (the amount of template used in a MS-PCR reaction) down to 0.4μl, at 0.4μl intervals and with 4 replicates per specified DNA quantity. As the DNA template was undigested, each DNA standard represented a known methylation level from 100% (2 μl DNA) down to 20% methylated (0.4μl DNA). To predict the observed methylation level value from the known standard, the intensity of the PCR products was divided by the '100% methylation' or 2μl standard PCR product intensity, thus replicating the method of quantification used for individual embryo analysis. The observed values were plotted against the predicted methylation level values to generate the standard curves specific to each embryo stage and primer set, and are shown in **Figure 7.6**.

Standard curves generated using OVSAT1A primers for all embryo stages/oocytes, were found to be reproducible, and the curves were of a quadratic nature. There was a clear correlation between decreasing amounts of template, and increased curvature of standard curve lines. In addition, as template levels decreased, the standard deviation between replicates of the same DNA standard increased, suggesting a loss of PCR sensitivity at lower DNA concentrations. Generation of standard curves using BOVB primers for 8-16 cell stage embryos and oocytes failed to provide reproducible standard curves with simple linear or quadratic properties, suggesting the BOVB primers were inefficient when used to amplify template of decreased concentration. Therefore, the small amount of MS-PCR data collected for these stages was ignored.


Blastocyst
OVSAT
y= -0.0042 +1.26x -0.25x²
Relative % methylation
100
50
0.0
0.0
1.0
2.0
0%
50%
100%
Input DNA (µl) or 'known' methylation levels

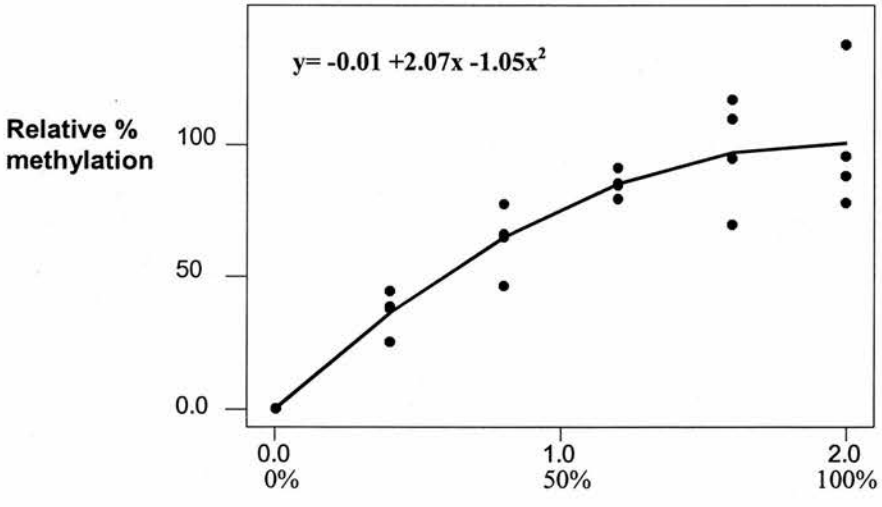

Morula
OVSAT
y= 0.032 +0.937x +0.036x²
Relative % methylation
100
50
0.0
0.0
1.0
2.0
0%
50%
100%
Input DNA (µl) or 'known' methylation levels

8-16 cell
OVSAT



Input DNA (µl) or 'known' methylation levels

Oocyte
OVSAT



Input DNA (µl) or 'known' methylation levels

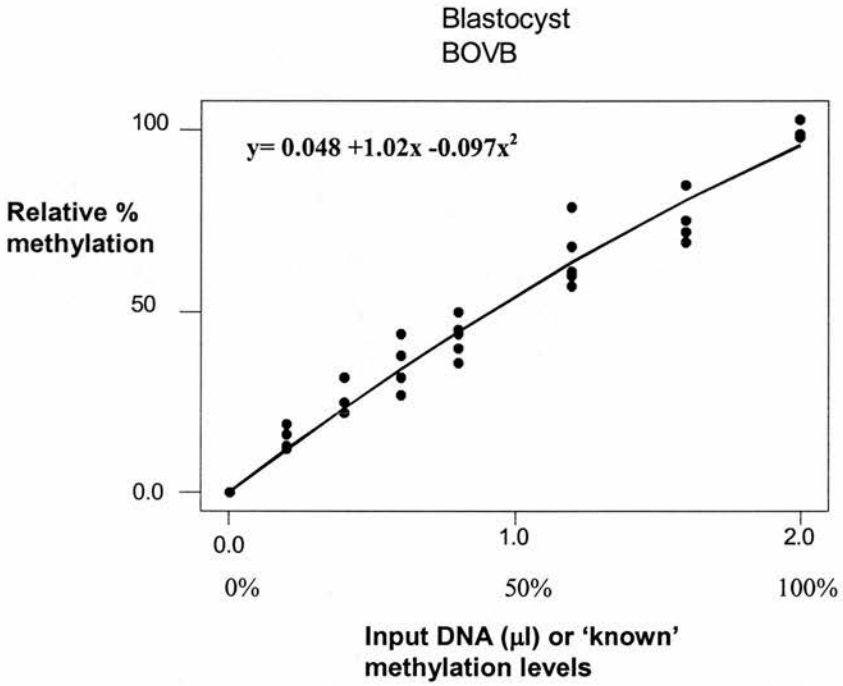


Figure 7.6 Standard curves for calculating methylation status of Satellite I (OVSAT1A) and OV-2 SINE (BOVB).

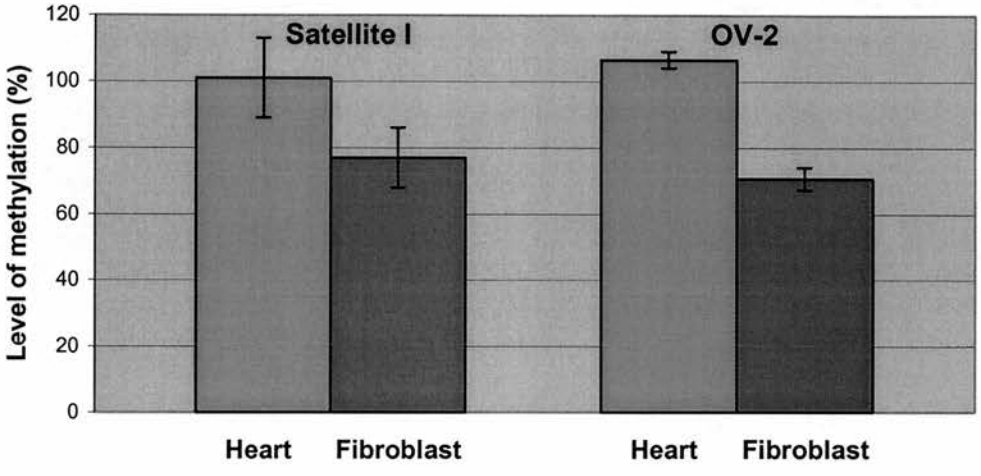
Input DNA for the MS-PCR reaction is indicated on the x-axis, with the 'known' % methylation levels also shown. For generation of the blastocyst standard curves for both OVSAT1A and BOVB primers, two additional standards of 0.6 µl (30% methylated) and 0.2µl (10%) were used. The 'observed' % methylation values are indicated on the y-axis.

7.4 Results

7.4.1 Methylation analysis of sheep heart and fetal fibroblast DNA

The DNA methylation status of the Satellite I and OV-2 SINE sequences were analysed to ascertain methylation levels for a fully differentiated tissue (adult heart), and also for the cells used in SCNT (fetal fibroblast cells). DNA was extracted from heart tissue and cultured fetal fibroblast cells (cell line BlWF1P5), as described **2.4.9**. Two (heart), or three (fetal fibroblast), DNA samples per tissue/cell type were analysed and the equivalent amount of DNA as in one blastocyst (~1 ng), was split into aliquots for incubation with or without *Hha* I (as described **2.4.10.2)**. The PCR reactions were performed at the same cycle number as for blastocyst DNA (as described **2.4.10.3**) and the methylation values obtained from direct comparison of PCR product intensities were standardised against the pooled blastocyst-generated standard curves (refer to **Figure 7.6**). The mean of three values for each tissue sample was calculated and the results are demonstrated in **Figure 7.7**.

DNA methylation levels were found to be similar for both Satellite I and OV-2 SINE DNA sequences in the cell and tissue types analysed. Methylation levels were highest in heart tissue (Satellite I: 101%; OV-2: 107%), and moderately lower in fibroblast cells (Satellite I: 77%; OV-2: 71%). The values gained for heart tissue are both higher than 100%, an impossible event, and the values demonstrate the potential range of error in calculating methylation values using this method. The raw PCR data and calculated methylation values for the DNA samples can be found in **Table A4.1, Appendix IV**.



Tissue/cell type	DNA sequence	Mean	S.E of Mean
Heart	Satellite I	101	12
	OV-2 SINE	106.5	2.5
Fetal Fibroblast	Satellite I	77	9
	OV-2 SINE	70.7	3.38

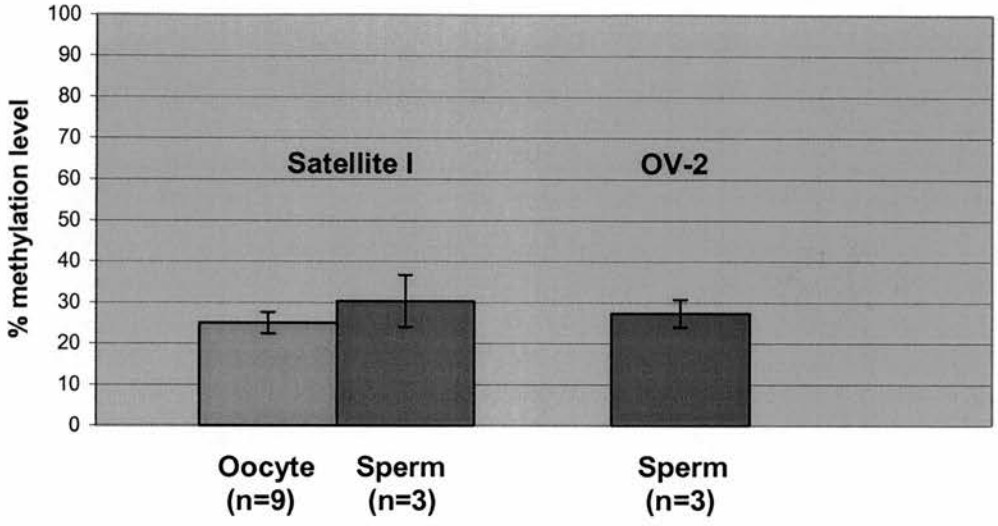
Figure 7.7 DNA methylation levels of Satellite I and OV-2 SINE DNA sequences in Heart tissue and Fetal fibroblast cells.

Methylation level of heart tissue and fetal fibroblast cells are indicated as percentage values and are representative of three samples for fetal fibroblast DNA, and two samples for heart DNA. The mean of methylation levels for each tissue type was calculated, and the standard error of the mean is indicated as bi-directional error bars. The numerical values of the mean and standard error calculated for each sample type are shown in the table below the chart.

7.4.2 Methylation analysis of gametic DNA

DNA was extracted from 10 single superovulated MII oocytes, as described **2.4.10.1.** The DNA was split into aliquots for incubation with or without *Hha* I, then MS-PCR analysis of the Satellite I DNA sequence was performed (as described **2.4.10.2** and **2.4.10.3**). Previous experiments had demonstrated non-reproducible results using the BOVB primers on very small amounts of DNA (less than the equivalent of blastocyst DNA), therefore, analysis of the OV-2 SINE element was not performed. Satellite I methylation values were normalised against the relevant standard curve (refer to **Figure 7.6**). DNA was extracted from frozen sperm (as described **2.4.9**) and three DNA samples with the equivalent amount of DNA as in one blastocyst (~1ng), were removed from the original sample. The three 1 ng DNA samples were split into aliquots for digestion with or without *Hha* I, as described **2.4.10.2.** MS-PCR was performed (as described **2.4.10.2**; **2.4.10.3**), at the same cycle number as for blastocyst DNA and the percentage gained from direct comparison of PCR product intensities was standardised against the relevant blastocyst-generated standard curve. Analysis of both the Satellite I and OV-2 SINE sequences was performed on sperm DNA.

The mean methylation levels of Satellite I for each gamete type was calculated and the results are shown in **Figure 7.8**. The raw PCR data and calculated methylation values for individual samples of the different gamete types can be found in **Table A4.2**, **Appendix IV**. Oocytes and sperm demonstrated similar levels of Satellite I DNA methylation (25% and 30% respectively), and analysis of the methylation status of the OV-2 SINE element in sperm also demonstrated relatively low levels of methylation (~27.3%).



DNA source	DNA sequence	Mean	S.E of Mean
Oocyte	Satellite I	25	2.6
Sperm	Satellite I	30.3	6.3
Sperm	OV-2	27.3	3.4

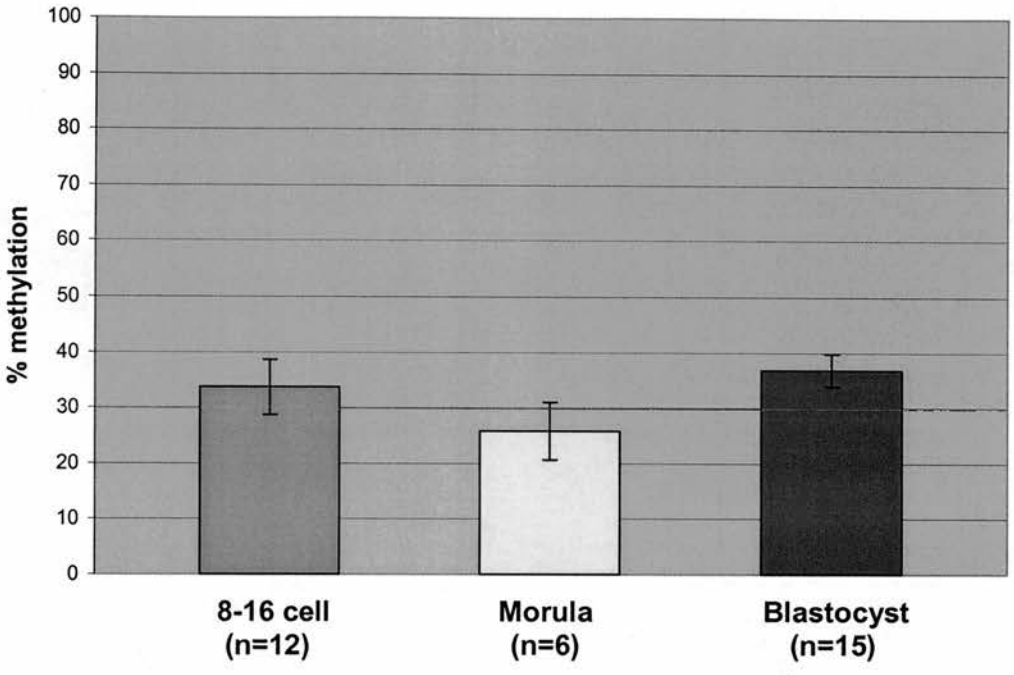
Figure 7.8 DNA methylation levels of Satellite I and OV-2 DNA sequences in gametes.

DNA methylation levels of oocyte or sperm DNA are indicated as percentage values on the y-axis. The numbers of oocytes/pooled sperm DNA samples used are indicated on the x-axis. The mean of methylation levels for each gamete type was calculated and the standard error of the mean is indicated as bi-directional error bars. The numerical values of the mean and standard error of the mean are shown in the table below the chart.

7.4.3 DNA methylation levels of Satellite I in multi-stage *in vivo* derived embryos

DNA was extracted from *in vivo*-derived embryos at the following stages (as described **2.4.10.1**): 8-16 cell (15), morula (8) and blastocyst (15). The DNA was subsequently incubated with or without *Hha* I (as described **2.4.10.2**), and then subjected to PCR amplification, with embryo stage and primer-specific conditions (as described **2.4.10.3** and **Table 2.5**). The methylation values obtained from direct comparison of PCR product intensities were standardised against the relevant standard curves (refer to **7.6**).

After removal of single embryo data, due to defective PCR amplification or complete failure of the PCR (no embryo in tube), the number of embryos providing acceptable MS-PCR data were as follows: 8-16 cell (12), morula (6) and blastocysts (15). The mean of methylation levels of Satellite I for each embryo stage was calculated and the results are shown in **Figure 7.9**. The raw PCR data and methylation values for individual embryos can be found in **Tables A4.3-A4.5**, **Appendix IV**. Variations in methylation levels of individual embryos at the same embryo stage were evident, sometimes deviating substantially from the mean. The mean methylation level for each embryo stage demonstrated a slight decrease in DNA methylation levels between the 8-16 cell and morula stages, followed a by a slight increase at the blastocysts stage. However, statistical analysis using student's t-tests (95% C.I.), demonstrated the differences in methylation levels observed between different embryo stages were not statistically significant ($P>0.05$). This may be due to the small numbers of embryos analysed at the morula stage.



Embryo stage	Mean	S.E of Mean
8-16 cell	33.7	5.0
Morula	25.8	5.2
Blastocyst	36.8	2.9

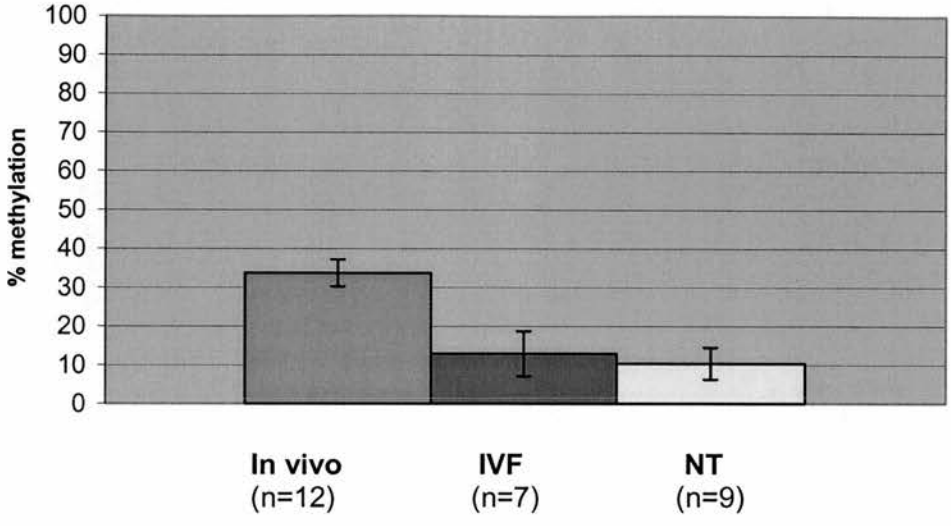
Figure 7.9 DNA methylation levels of Satellite I during preimplantation stages of development.

DNA methylation levels of embryos are indicated as percentage values on the y-axis. The stage and number of embryos used are indicated on the x-axis. The mean of methylation levels from embryos of a single stage was calculated, and the standard error of the mean is indicated as bi-directional error bars on the chart. The numerical values of the mean and standard error of the mean are shown in the table below the chart.

7.4.4 Comparison of DNA methylation levels in *in vivo*, IVF and SCNT derived sheep embryos

MS-PCR analysis of Satellite I DNA was performed (as described **2.4.10**) on the following embryos: IVF blastocysts (18), NT blastocysts (25), IVF 8-16 cell embryos (10) and 8-16 cell NT embryos (10). The numbers of embryos chosen reflects availability. The results already gained for *in vivo*-derived embryos (blastocysts (15) and 8-16 cell embryos (12)) were used for comparison with the results generated from IVF and NT embryos of the same developmental stage. It was possible to combine results collected in separate experiments, as methylation levels of single embryos are measured relative to an 'internal' standard, then undergo standardisation using the appropriate standard curve. MS-PCR of the OV-2 SINE element was also collected (as described **2.4.10**) for the following embryos; *in vivo*-derived blastocyst (10), IVF blastocyst (10) and NT blastocyst (20). The methylation values obtained from direct comparison of PCR product intensities were standardised against the relevant standard curves (refer to **7.6**).

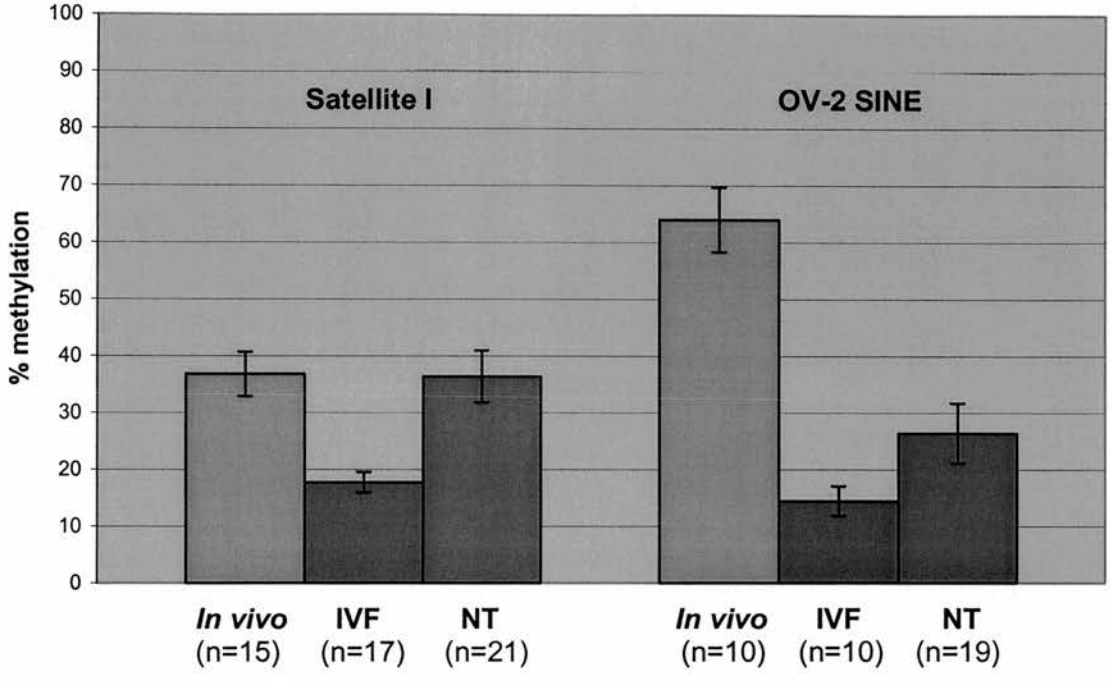
After removal of single embryo data due to defective amplification, or failure of the PCR, the number of embryos providing acceptable MS-PCR data was as follows: IVF blastocysts (17), NT blastocysts (21), IVF 8-16 cell embryos (7) and NT 8-16 cell embryos (9). All of the embryos used for OV-2 SINE methylation analysis provided acceptable data except one NT blastocyst. The mean of methylation levels of the Satellite I or OV-2 SINE sequence was calculated for each group of same-stage embryos produced by the same treatment, the results are shown in **Figures 7.10** (8-16 cell data) and **Figure 7.11** (blastocyst data). The raw PCR data and methylation values for all of the individual embryos can be found in **Tables A4.3**, **A4.5** and **A4.6**, **Appendix IV**.



Embryo source	Mean	S.E of Mean
In vivo	33.7	3.5
IVF	12.9	5.8
NT	10.3	4.1

Figure 7.10 Comparison of Satellite I DNA methylation levels in *in vivo*, IVF and SCNT derived 8-16 cell embryos.

DNA methylation levels of embryos are indicated as percentage values on the y-axis. The source and number of embryos used are indicated on the x-axis. The mean of Satellite I DNA methylation levels from embryos of a single source was calculated, and the standard error of the mean is indicated as bi-directional error bars on the chart. The numerical values of the mean and standard error of the mean are shown in the table below the chart.



Source of Embryo	DNA sequence	Mean	S.E of Mean
In vivo	Satellite I	36.8	3.9
	OV-2	64	5.7
IVF	Satellite I	17.7	1.8
	OV-2	14.4	2.6
NT	Satellite I	36.4	4.6
	OV-2	26.4	5.3

Figure 7.11 Comparison of Satellite I and OV-2 SINE DNA methylation levels in *in vivo*, IVF and NT derived blastocysts

DNA methylation levels of embryos are indicated as percentage values on the y-axis. The source and number of embryos used are indicated on the x-axis. The mean of sequence-specific DNA methylation levels (Satellite I or OV-2) from embryos of a single source was calculated, and the standard error of the mean is indicated as bi-directional error bars on the chart. The numerical values of the mean and standard error of the mean are shown in the table below the chart.

Methylation levels of the Satellite I DNA sequence differed substantially between the groups of 8-16 cell embryos derived from different embryo technologies. Methylation levels were highest in *in vivo*-derived embryos (33.7%), whilst embryos produced by IVF or SCNT demonstrated very low levels of methylation at the *Hha* I site (10-12%). In blastocysts, methylation levels of Satellite I also differed between embryos derived from the different embryo technologies, but did not correlate with the differences seen at the 8-16 cell stage. *In vivo*-derived and SCNT blastocysts demonstrated similar DNA methylation levels (~36%), whilst IVF blastocysts demonstrated low levels of DNA methylation (14.4%). Analysis of the OV-2 SINE DNA sequence in blastocysts also demonstrated differences between embryo groups. *In vivo* blastocysts demonstrated high levels of methylation (64%), whilst IVF and SCNT-derived embryos demonstrated lower levels of methylation (14.4% and 26.4% respectively).

To assess whether the differences observed were statistically significant, one-way analysis of variance (ANOVA) was performed. As the methylation data collected from individual embryos did not demonstrate equal variance between the treatment groups, a square root transformation was performed to achieve approximate equal variance. The one-way ANOVA was performed post-transformation. To calculate the statistical significance of methylation level differences observed between different treatment groups, pairwise comparisons were made using the Tukey simultaneous test (95% confidence intervals). P values were calculated from T-values generated from the Tukey tests, using a t-distribution table. **Figures 7.12, 7.13 and 7.14** demonstrate the output of the one-way ANOVA for Satellite I methylation levels in 8-16 cell embryos, Satellite I in blastocysts and OV-2 SINE in blastocysts respectively.

```
C5 = 1 subtracted from:

C5      Lower     Center     Upper  ---+---------+---------+---------+---
2      -4.319    -2.426    -0.533   (---------*--------)
3      -4.541    -2.786    -1.031  (--------*--------)
                                   ---+---------+---------+---------+---
                                    -4.0      -2.0       0.0       2.0


C5 = 2 subtracted from:

C5      Lower     Center     Upper  ---+---------+---------+---------+---
3      -2.365   -0.3597     1.646             (---------*---------)
                                   ---+---------+---------+---------+---
                                    -4.0      -2.0       0.0       2.0


Tukey Simultaneous Tests
Response Variable C6
All Pairwise Comparisons among Levels of C5

C5 = 1 subtracted from:

Level  Difference       SE of
C5       of Means  Difference   T-Value   P-Value
2          -2.426      0.7604    -3.191    <0.005
3          -2.786      0.7050    -3.951    <0.001

C5 = 2 subtracted from:

Level  Difference       SE of
C5       of Means  Difference   T-Value    P-Value
3         -0.3597      0.8057   -0.4464    >0.300
```

Figure 7.12 Statistical analysis of differences in Satellite I DNA methylation level sobserved between *in vivo*, IVF and SCNT derived 8-16 cell embryos.

Level 1 are the *in vivo*-derived embryos, level 2 are the IVF produced embryos, and level 3 are the SCNT embryos. The Minitab program did not calculate p-values; T-values were used to calculate P-values using a t-distribution table. The number of degrees of freedom was 25 (n-2).

```
C28 = 1 subtracted from:

C28      Lower     Center      Upper  ---------+---------+---------+-------
2       -3.017    -1.879    -0.7406  (------*------)
3       -1.309    -0.223     0.8634             (------*-----)
                                     ---------+---------+---------+-------
                                            -1.6       0.0       1.6


C28 = 2 subtracted from:

C28      Lower     Center      Upper  ---------+---------+---------+-------
3       0.6078     1.656      2.704                       (-----*------)
                                     ---------+---------+---------+-------
                                            -1.6       0.0       1.6


Tukey Simultaneous Tests
Response Variable C29
All Pairwise Comparisons among Levels of C28

C28 = 1 subtracted from:

Level    Difference       SE of
C28        of Means  Difference   T-Value    P-Value
2           -1.879      0.4706    -3.992    <0.001
3           -0.223      0.4491    -0.496    >0.300

C28 = 2 subtracted from:

Level    Difference       SE of
C28        of Means  Difference   T-Value    P-Value
3             1.656      0.4334     3.821    <0.001
```

Figure 7.13 Statistical analysis of differences in Satellite I DNA methylation level observed between *in vivo*, IVF and SCNT derived blastocysts.

Level 1 are *in vivo*-derived embryos, level 2 are the IVF produced embryos, and level 3 are the SCNT embryos. The Minitab program did not calculate p-values; T-values were used to calculate P-values using a t-distribution table. The number of degrees of freedom was 50 (n-2).

```
C19 = 1 subtracted from:

C19      Lower     Center      Upper  ---+---------+---------+---------+---
2       -5.630    -4.005    -2.379  (------*-----)
3       -4.903    -3.483    -2.063     (-----*-----)
                                    ---+---------+---------+---------+---
                                     -5.0       -2.5       0.0       2.5


C19 = 2 subtracted from:

C19      Lower     Center      Upper  ---+---------+---------+---------+---
3      -0.8980    0.5219     1.942                        (-----*-----)
                                    ---+---------+---------+---------+---
                                     -5.0       -2.5       0.0       2.5


Tukey Simultaneous Tests
Response Variable C20
All Pairwise Comparisons among Levels of C19

C19 = 1 subtracted from:

Level    Difference       SE of
C19        of Means  Difference   T-Value    P-Value
2            -4.005      0.6644    -6.028     0.0000
3            -3.483      0.5804    -6.001     0.0000

C19 = 2 subtracted from:

Level    Difference       SE of
C19        of Means  Difference   T-Value    P-Value
3            0.5219      0.5804     0.8993    >0.100
```

Figure 7.14 Statistical analysis of differences in OV-2 SINE DNA methylation level observed between *in vivo*, IVF and SCNT derived blastocysts.

Level 1 are *in vivo*-derived embryos, level 2 are the IVF produced embryos, and levels 3 are the SCNT embryos. The Minitab program did not calculate p-values; T-values were used to calculate P-values using a t-distribution table. The number of degrees of freedom was 50 (n-2).

For the Satellite I sequence of 8-16 cell embryos, statistically significant differences in methylation levels were observed between *in vivo*-derived and IVF embryos, and *in vivo* and SCNT embryos. There was no difference observed between embryos produced by IVF or SCNT. Methylation levels of the Satellite I sequence of blastocysts was significantly different between *in vivo* and IVF embryos, and IVF and SCNT embryos. There was no significant difference observed between *in vivo*-derived and SCNT blastocysts. Finally, methylation levels of the OV-2 SINE sequence were significantly different between *in vivo* and IVF embryos, *in vivo*-derived and SCNT embryos, but not between IVF and SCNT blastocysts.

Analysis of methylation data from single blastocysts provided additional information (individual embryo data can be found in **Tables A4.5-A4.6, Appendix IV**). When analysing Satellite I methylation data of SCNT blastocysts, 9/21 embryos demonstrated methylation levels comparable to IVF blastocysts (10-30%). The remaining embryos demonstrated substantially increased methylation levels (30-74%). Considering that IVF and SCNT embryos are comparable in terms of the equivalent period of *in vitro* embryo culture, it would appear that some embryos are correctly reprogrammed, whilst other embryos maintain high levels of methylation. When performing a similar comparison of the OV-2 SINE element, many of the SCNT embryos retain similar methylation levels to IVF derived blastocysts, whilst only 2/20 embryos demonstrate significantly higher levels of methylation (52-73%).

7.5 Discussion

A technique to analyse the methylation status of repetitive DNA sequences in tissues, cells, single oocytes and embryos was devised and employed to assess DNA methylation status in embryos produced by different embryo technologies. The method was found to be reproducible for Satellite I DNA methylation analysis for all embryo stages and oocytes, but not reproducible when analysing the methylation status of the OV-2 SINE element at specific embryo stages.

The DNA methylation status of a *Hha* I site within the Satellite I and OV-2 SINE sequences was first analysed in heart tissue and fetal fibroblast cells. Adult heart tissue DNA was used to obtain DNA methylation values for a fully differentiated tissue, and fetal fibroblast cell DNA was analysed to ascertain methylation levels of the Satellite I and OV-2 SINE sequences in SCNT donor cells. The *Hha* I sites within the Satellite I and OV-2 SINE sequences were found to be highly methylated in heart DNA (mean of >100%), with fetal fibroblast cells demonstrating slightly lower mean levels of methylation (77% Satellite I; 71% OV-2 SINE). The high levels of DNA methylation observed in heart tissue are possibly indicative of the degree of cellular differentiation of this somatic cell type. The slightly lower levels of methylation of Satellite I and OV-2 SINE in fetal fibroblast DNA may be due to the 'less differentiated' state of fibroblast cells.

The DNA methylation status of the *Hha* I site within the sheep Satellite I sequence was analysed in oocytes and sperm. As the OV-2 SINE MS-PCR method was not reproducible in oocytes, the methylation status of the *Hha* I site within this repetitive sequence was analysed only in sperm. Both sperm and oocytes demonstrated relatively low mean levels of methylation at the *Hha* I site of Satellite I (25% and 30% respectively), and the *Hha* I site within the OV-2 SINE sequence in sperm presented comparable levels of methylation (27%).

In general, investigations of DNA methylation levels of repetitive DNA sequences in mouse oocytes and spermatazoa have demonstrated high levels of methylation in sperm, and undermethylation in oocytes (Sanford *et al.*, 1987, Howlett & Reik, 1991). The observation of undermethylated Satellite I DNA sequences in sheep oocytes fits well with these findings, however, the relatively low levels of DNA methylation seen in sheep sperm Satellite I and OV-2 SINE sequences is in disagreement with previous observations in the mouse. There are several possible explanations for this apparent difference. Firstly, only a few repetitive sequences (including the L1 LINE, MUP (Major urinary protein) and IAP (Intracisternal A particle) sequences) have been analysed in mouse gametes, and may not be representative of all genes and DNA sequences (Sanford *et al.*, 1987; Howlett & Reik, 1991). Secondly, the MS-PCR analysis of OV-2 SINE and Satellite I only analyses the methylation status of a single CpG site, which may not be representative of the entire DNA sequence. Closer inspection of the *Hha* I site in Satellite I demonstrates localisation within a cluster of CpG residues, whereas the OV-2 SINE *Hha* I site is isolated.

DNA methylation status of the *Hha* I Satellite I sequence was compared in sheep preimplantation embryos of different stages and was not found to vary significantly between 8-16 cell, morula and blastocyst stages. A slight decrease was observed at the morula stage, although this may be compounded by the small number of embryos analysed at this stage (n = 6). In the mouse, centromeric heterochromatin (containing Satellite DNA) is not demethylated until the blastocyst stage of preimplantation development. The results collected from sheep embryos suggest no demethylation of Satellite I DNA at the *Hha* I site occurring between the 8-16 cell and blastocyst stages.

It is possible that demethylation may occur prior to the 8-16 cell stage in sheep embryos, but analysis of sheep 2-cell and 4-cell embryos was not possible due to the combined problem of time constraints and difficulties in obtaining significant numbers of *in vivo*-derived embryos. However, taking into consideration that the relatively low methylation values of oocyte and sperm Satellite I DNA (25% and 30% respectively), are

comparable to those seen in 8-16 cell, morula and blastocyst stage embryos, any demethylation occurring at the single CpG site prior to the 8-16 cell stage would most likely have to follow a post-fertilisation *de novo* methylation event.

Direct comparison of Satellite I DNA methylation status was performed in 8-16 cell embryos or blastocysts produced *in vivo*, by IVF or SCNT. In addition, analysis of the OV-2 SINE was also performed for individual blastocysts produced by the three embryo technologies. At the 8-16 cell stage, the DNA methylation status of the single *Hha* I site in Satellite I was comparable in IVF or SCNT-derived embryos (means of 3% and 10% respectively), but *in vivo* embryos demonstrated significantly higher methylation status (mean of 34%). In blastocysts, the DNA methylation status of Satellite I was high in *in vivo*-derived embryos (mean of 36.8%), with comparable levels seen in SCNT blastocysts (mean of 36.4%). The DNA methylation status of Satellite I was considerably lower in IVF produced embryos (mean of 17.7%), when compared to *in vivo*-derived and SCNT produced blastocysts. Differences in OV-2 SINE DNA methylation status were also observed between *in vivo*-derived, IVF and SCNT blastocysts, with *in vivo*-derived embryos demonstrating the highest levels of methylation (mean of 64%), when compared to IVF and SCNT produced blastocysts (means of 14.4% and 26.4% respectively).

The apparent difference in methylation status observed between *in vivo*-derived and IVF embryos is intriguing, as both of these embryo production techniques generally produce a high proportion of 'normal' offspring (De Sousa *et al.*, 2001). When comparing the DNA methylation status of Satellite I or OV-2 SINE in embryos produced by different techniques (*in vivo*-derived, IVF or SCNT produced), the following points should be taken into consideration. Firstly, SCNT embryos are comparable to IVF produced embryos in terms of *in vitro* embryo culture, although they are not wholly comparable, as they are produced using oocytes of different origin (superovulated MII oocytes (SCNT) or *in vitro* matured (IVM) oocytes (IVF)), and may also develop to the blastocyst stage at different rates. It cannot be assumed that superovulated oocytes

would be comparable to *in vitro* matured (IVM) oocytes in respect of methylation status and the required complement of 'methylation-reprogramming factors'. The period of embryo culture common to IVF and SCNT embryos may also lead to differences in sequence-specific DNA methylation status, as observed in large offspring syndrome (Young *et al.*, 2001).

If comparisons of Satellite I methylation status are made between *in vivo*-derived and SCNT embryos, the following methylation reprogramming scenario is apparent. A significant demethylation of the SCNT embryos occurs prior to the 8-16 cell stage, followed by an increase in DNA methylation to levels equivalent to those observed in *in vivo*-derived embryos at the blastocyst stage. This would require hypomethylation of DNA in 8-16 cell SCNT embryos, whilst their *in vivo* counterparts remain moderately methylated. Recent studies investigating DNA methylation patterns in cloned sheep embryos all described abnormally high methylation levels in embryonic nuclei, purportedly reflecting the difficulties in reprogramming somatic nuclei methylation patterns (Kang *et al.*, 2001a; Dean *et al.*, 2001; Bourc'his *et al.*, 2001). It would be most unexpected to observe undermethylated DNA as a result of SCNT and this may therefore be an effect of in vitro culture, rather than aberrant reprogramming. In this respect, comparison of SCNT embryos with IVF embryos may be more suitable. At the 8-16 cell stage, the methylation status of the *Hha* I site of Satellite I in IVF and SCNT produced embryos are very similar, however, at the blastocyst stage, Satellite I methylation levels in both IVF and SCNT embryos have increased, with a more marked increase in SCNT blastocysts. This produces a statistically significant difference between the two groups of embryos ($P<0.001$).

Overall, this creates a picture of correct methylation reprogramming up to the 8-16 cell stage, followed by establishment of aberrant methylation patterns prior to the blastocyst stage. These results agree in kind with those of Dean *et al.* (2001), whom observed partial demethylation up to the 8-cell stage in cloned embryos, followed by spurious *de novo* methylation. Closer inspection of Satellite I DNA methylation values for

individual embryos demonstrated that a proportion of SCNT embryos (9/21) retained methylation levels similar to their IVF produced counterparts (10-30%), whilst the remainder demonstrated increased methylation levels (30-70%). This would suggest some SCNT blastocysts are reprogrammed correctly, but a larger proportion fail to recapitulate 'normal' DNA methylation patterns, and may partially account for the low success rates observed in large animal SCNT.

When analysing DNA methylation status of the OV-2 SINE sequence observed between IVF and SCNT produced blastocysts, the observed mean levels of methylation (14.4% and 26.4% respectively), were not found to be statistically significant ($P>0.1$). When inspecting individual blastocyst DNA methylation values, only 2/20 SCNT blastocysts demonstrated significantly higher levels of methylation than their IVF produced counterparts. This contrasts with the observations made for Satellite I methylation levels, where a large proportion of the SCNT blastocysts demonstrated aberrant Satellite I methylation when compared with IVF blastocysts. This suggests sequence-specific methylation reprogramming errors occurring in SCNT produced embryos. These results agree in kind with the observations of Bourc'his *et al.*, where DNA methylation levels appear to be increased in heterochromatin regions (the location of Satellite I) of chromosomes derived from SCNT blastocysts relative to those derived from IVF blastocysts, but are similar in euchromatin regions (where the OV-2 SINE sequence is likely to reside).

It must be noted that the differences in DNA methylation levels observed between *in vivo*-derived, SCNT and IVF blastocysts may also be influenced by embryo collection methods. *In vivo* blastocysts were collected from superovulated and naturally mated ewes, and it is notoriously difficult to synchronise ovulation in sheep, although PMSG (pregnant mare serum gonadotrophin) was administered to control this as much as possible. In any one collection, blastocysts of different stages may be observed (e.g. early, expanded or hatching), as the timing of ovulation and fertilisation may differ by several hours. IVF and SCNT produced blastocysts are potentially more comparable in

this respect, as they are fertilised or fused in the space of 3-4 hours, cultured *in vitro*, and frozen down at the time of expansion. However, for the 'stage' of blastocyst collected to influence the results in terms of methylation, significant methylation changes (demethylation/*de novo* methylation) must occur at the blastocyst stage in sheep, an as yet undetermined event.

Also, worthy of note was the observation that Satellite I and OV-2 SINE methylation values for individual *in vivo*-derived embryos of the same stage varied substantially. For example, methylation of the *Hha* I site of Satellite I in 8-16 cell embryos varied from 18-80%. This may reflect methylation differences between 8-cell and 16-cell embryos, although other possibilities exist. When taking into consideration that not all *in vivo*-derived embryos will be able to sustain normal development, (i.e. not all 8-16 cell embryos will progress to later stages of development), it is possible that this may be indicated by widely varying methylation levels. If this hypothesis is true, the later stages of development should demonstrate smaller variation within a same-stage group of embryos. Indeed, there was a marked decrease in individual embryo variation between 8-16 cell embryos and blastocysts. However, a certain extent of variation within a population of single-stage embryos is still likely, reflecting both naturally occurring variation and that induced experimentally.

Overall, the DNA methylation data generated from *in vivo*-derived, IVF and SCNT produced embryos is not clear-cut due to the aforementioned experimental constraints. However, valuable preliminary information has been obtained. Sequence-specific inconsistencies in reprogramming are apparent when comparing IVF and SCNT embryos. Furthermore, it is possible that *in vitro* culture may lead to the hypomethylation of repetitive DNA sequences, reflecting previously described anomalies in imprinted gene methylation status of *in vitro* cultured sheep and mouse embryos (Young *et al.*, 2001; Doherty *et al.*, 2000). Analysis of global methylation reprogramming in normal and cloned preimplantation embryos (using indirect

immunoflourescence) alongside examination of the reprogramming capacity of oocytes derived from different sources, will help to address the issues raised in this initial study.

Chapter 8

Dynamics of DNA methylation in normal and cloned sheep development: summary and discussion

8.1 Aims of thesis

The primary aim of this thesis was to initiate a study of DNA methylation in sheep, in particular, its dynamics during the early stages of normal and cloned embryo development.

Specific aims were as follows:

1. To clone sheep homologues of genes encoding DNA methyltransferases and methyl-CpG-binding proteins and evaluate their conservation with mouse and human homologues.
2. To investigate expression levels of these genes in adult sheep tissues.
3. To determine whether an oocyte-specific form of *DNMT1* exists in sheep oocytes.
4. To determine expression patterns of *DNMT1*, *DNMT3A* and *DNMT3B* genes in oocytes and preimplantation embryos.
5. To optimise techniques for DNA methylation analysis of specific sequences in single oocytes and embryos, and quantify changes in methylation levels between different stages of sheep preimplantation development.
6. To determine whether DNA methylation of specific sequences is equivalent in *in vivo* produced, *in vitro* fertilised and SCNT embryos.

8.2 Summary of key results

The sheep homologues of DNA methyltransferases *DNMT1*, *DNMT3A* and *DNMT3B*, and the methyl-CpG-binding protein *MBD2*, were cloned fully or partially, using a combination of RT-PCR, cDNA library screening, and 5' RACE methods. Analysis of predicted protein sequences demonstrated high levels of conservation within and outwith functional domains.

Attempts to identify an oocyte-specific form of *DNMT1*, known to exist in the mouse, were unsuccessful. Instead, a form identical to that isolated from adult tissues was detected in oocytes. These results imply there is little or no oocyte-specific form of DNMT1 in sheep oocytes, which may be indicative of species-specific differences during early epigenetic reprogramming events.

Analysis of *DNMT1*, *DNMT3A*, *DNMT3B* and *MBD2* expression in adult tissues using Taqman® RT-PCR, demonstrated similar patterns to those observed in the human, where previously reported. Qualitative expression analysis of *DNMT1* expression in sheep oocytes and preimplantation embryos of various stages, suggested comparable expression patterns to those observed in the mouse, with depletion of maternal stores observed between the 1-cell and blastocyst stages. Expression analysis of *DNMT3A* demonstrated a reciprocal increase between these stages in the sheep but not the mouse, and analysis of *DNMT3B* expression indicated similar patterns in sheep and mouse. *MBD2* was not found to be highly expressed in 2-cell or morula stage embryos.

A methyl-sensitive PCR method, optimised for use in single oocytes and embryos, was employed to assay DNA methylation status of a sheep Satellite sequence, and a sheep SINE element. Oocytes and sperm demonstrated similar Satellite I DNA methylation status. No significant changes were detected in the Satellite sequence of *in vivo* derived embryos between the 8-16 cell and blastocyst stages of development, contrasting with the significant decrease in Satellite I DNA methylation levels observed in the mouse.

However, comparisons between *in vivo*, *in vitro* fertilised (IVF) and SCNT blastocysts demonstrated differences, with the IVF and SCNT embryos (both of which undergo equivalent periods of *in vitro* culture) relatively undermethylated within Satellite I and the OV-2 SINE element. In addition, SCNT blastocysts appeared to demonstrate increased methylation levels relative to their IVF counterparts within the Satellite I sequence, but not the OV-2 SINE element, suggesting sequence-specific reprogramming errors.

8.3 Key components of the DNA methylation repressor system in sheep: an insight into functional conservation

By cloning complete or partial cDNA sequences encoding sheep homologues of DNA methyltransferases and a methyl-CpG-binding protein, insights into the general levels of conservation of the sheep DNA methylation machinery were obtained. Firstly, all cDNA sequences cloned in sheep demonstrated high levels of homology with human and mouse sequences, with sheep sequences consistently demonstrating more homology with their human counterparts. This similarity reflects evolutionary events, as sheep genomes are more closely related to the genomes of primates than those of rodents (Burt *et al.*, 1999). Conceptual translation of the sheep *DNMT1* and *DNMT3A* cDNA sequences demonstrated functional integrity of key motifs and domains (identified in both human and mouse protein sequences) within predicted protein sequences. This would further suggest conservation of the activities and functions of these DNA methyltransferases within mammalian species. Unfortunately, only limited *DNMT3B* cDNA sequence (51 bp) was obtained, therefore suggestions for the function or activity of the encoded protein in sheep could not be made confidently.

The finding that identical forms of *DNMT1* exist in sheep-derived somatic cells and oocytes was of significant interest. Whether the apparent lack of an oocyte-specific isoforms extends to humans, remains undetermined. As mentioned previously, the genomes of primates are closely related to ruminants in evolutionary terms, and in this respect, the human genome may lack oocyte-specific *DNMT1* sequences. Indeed, database searching of human chromosome 19 (the location of *DNMT1*) failed to identify sequences sharing homology with mouse *Dnmt1* exon1o. However, this does not completely disprove the existence of variant DNMT1 proteins in humans (or sheep), as similarly truncated Dnmt1 isoforms have been identified in rat brain, and these are evidently produced using splicing strategies that do not require an alternative 1st exon. Investigations of DNMT1 isoforms (RNA and protein) in other mammals are required to

determine whether Dnmt1 proteins lacking the N-terminal sequences of the somatic form are restricted to rodent species.

To further characterise the components of DNA methylation machinery cloned in sheep, expression analysis of mRNA transcripts in adult tissues was performed. *DNMT1* was found to be ubiquitous in all tissues examined, with tissue-specific expression patterns demonstrating similarity with those in the human where reported. *DNMT3A* was also expressed ubiquitously in adult tissues, and *DNMT3B* was not detectable in any of the adult sheep tissues analysed, but was found to be expressed in d125 fetal tissues. Again, these patterns are very similar to those reported in the human (Xie *et al.*, 1999).

All three DNA methyltransferases were found to be expressed in d125 fetal tissues (data presented only for *DNMT3B* in this thesis), demonstrating a requirement in developing fetuses. The significant downregulation of *DNMT3B* in adult sheep tissues relative to d125 fetal tissue clearly demonstrates an important role for this gene in fetal development. Indeed, *Dnmt3b* is essential for normal development in the mouse, with lethality occurring at mid-gestation in embryos carrying a targeted deletion of the gene (Okano *et al.*, 1999).

In the mouse, *Dnmt3a* is also downregulated in adult tissues relative to undifferentiated cells (Okano *et al.*, 1998a), but this is not observed in adult sheep (described in this thesis) or human tissues (Xie *et al.*, 1999), with ubiquitous expression comparable to *DNMT1* in both species. The differences observed imply important roles for *DNMT3A* in both adult tissues and embryos of sheep and humans, but mice evidently only require significant *Dnmt3a* expression in developing systems. This discrepancy may represent alternative roles for DNMT3A in different mammalian species. As the specific sequences targeted by DNMT3A are not yet clearly defined in either the mouse or human, it is not possible to speculate how these roles may differ.

MBD2 was shown to be expressed ubiquitously in adult sheep tissues, and mimicked tissue-specific patterns of *DNMT1* expression in all tissues except ovary. Expression data for *MBD2* is not available for adult human tissues, but *MBD2* expression patterns did not differ substantially from those observed in the mouse (Hendrich & Bird, 1998). *MBD2* was found to be expressed abundantly in all tissues, with levels significantly higher than any of the DNA methyltransferases. This would suggest a significant role in methylation-mediated transcriptional repression for this gene, although a certain extent of functional redundancy for this protein has been demonstrated in the mouse (Hendrich *et al.*, 2001).

The similar tissue-specific patterns of *MBD2* and *DNMT1* lend further weight to the recent observation of direct interaction between *DNMT1* and *MBD2* (Tatematsu *et al.*, 2001). A simple explanation of this observation would be that interaction between *DNMT1* and *MBD2* provides a direct link between DNA methylation and immediate transcriptional silencing in actively dividing cells, an idea reinforced by the previous demonstration of the interaction of DNA methyltransferases with histone-deacetylase silencing complexes (Fuks *et al.*, 2000; Rountree *et al.*, 2000).

The apparent high levels of conservation of DNA methylation machinery components in sheep, especially with the human, demonstrates the important nature of DNA methylation in several distinct classes of mammalian species (primates, rodents and ruminants). In addition, the tissue-specific expression patterns reported in this thesis, and the conservation of functional domains within sheep *DNMT* and *MBD2* sequences, suggest the function and specific roles of DNA methylation in adult sheep tissues would be similar to known roles in other mammalian species. However, little is known about the methylation status of genes or DNA sequences in sheep, aside from information acquired in this thesis and a previous study of a differentially methylated region (DMR) within the sheep *IGF2R* gene (Young *et al.*, 2001). Studies of other imprinted and non-imprinted gene promoter regions and their correlation with gene expression, will

determine the extent of involvement of DNA methylation in transcriptional repression in sheep.

8.4 Dynamics of DNA methylation in normal sheep preimplantation development

Methylation analysis of a single CpG site in sheep Satellite I DNA between the 8-16 cell and blastocyst stages of preimplantation development, demonstrated no significant changes in DNA methylation levels. The DNA methylation status of this site was also found to be similar in mature oocytes and sperm. Assuming no increases in DNA methylation occurs between fertilisation and the 8-16 cell stage, and that the single site is representative of the entire Satellite I DNA methylation status, this would suggest no significant demethylation of Satellite I occurs between fertilisation and blastocyst formation.

This finding contrasts with observations made in the mouse, where pericentromeric heterochromatin (containing Satellite DNA) is known to remain highly methylated until just prior to the blastocyst stage (Rougier *et al.*, 1998). However, the DNA methylation status of pericentromeric Satellite I DNA is not indicative of all DNA sequences, coding or non-coding. Demethylation appears to takes place to different extents, with some sequences demethylated rapidly after fertilisation, whilst others remaining methylated throughout preimplantation development. In addition, differential demethylation of the maternal and paternal genomes takes place shortly after fertilisation (Mayer *et al.*, 2000; Oswald *et al.*, 2000).

The MS-PCR technique used in this study did not allow assessment of general genomic methylation levels, only sequence-specific. Therefore, a clear picture of global methylation changes and in particular, observations of any differential demethylation events, was not permitable. However, parallel to this study, immunolocalisation techniques using an antibody directed against m^5C were employed to assess changes in global methylation patterns in sheep preimplantation embryos (N. Beaujean, R.I.). These experiments demonstrated that the global demethylation observed in mouse embryos does not occur to the same extent in sheep, and furthermore, the genome

appears to be *de novo* methylated after the 8-16 cell stage of development. Secondly, the differential demethylation of paternal and maternal genomes observed post-fertilisation in the mouse, rat, pig and cow (Mayer *et al.*, 2000; Dean *et al.*, 2001) does not appear to occur in the sheep.

These findings are exceptionally interesting, and have major implications for theories devised to explain the maternal and paternal differences in demethylation occurring post-fertilisation. The rapid and active demethylation of the paternal genome is thought to be required for the re-setting (or removal) of paternal methylation imprints, whereas, the maternal genome retains DNA methylation levels, with maternal methylation imprints remaining intact (Oswald *et al.*, 2000). This theory is supported by the observation that most maternally imprinted genes require DNA methylation to silence paternal alleles, whereas paternally imprinted genes are often associated with anti-sense transcripts to silence maternal alleles (Reik & Walter, 2001). The imprinting status of several genes in sheep has been confirmed, including *IGF2*, *IGF2R* and *H19* (McLaren & Montgomery, 2000; Feil *et al.*, 1998). So if sheep have imprinted genes, and differential demethylation of paternal genomes is required for the correct re-setting of paternal imprints, why is there no differential demethylation of paternal genomes in post-fertilisation sheep embryos? In this respect, it is likely that sheep have evolved an alternative developmental mechanism for correctly establishing imprints. During the course of this study, it was noticed that sheep gametes shared similar methylation status within Satellite I, and a comparably low methylation status was also observed in the OV-2 SINE element in sperm DNA. This contrasts with the generally high methylation status of DNA sequences in mouse sperm (Sanford *et al.*, 1987; Howlett & Reik, 1991). If the relatively low methylation status of Satellite I and OV-2 SINE are indicative of the entire sperm genome, then this may explain why no paternal demethylation is required post-fertilisation.

Analysis of DNA methyltransferases transcripts in sheep preimplantation embryos can lend possible answers to the unusual observations made in sheep. Firstly, a significant

finding in this thesis was that sheep express an identical form of *DNMT1* in both somatic cells and oocytes. This contrasts with the mouse, where an 'oocyte-specific' truncated form of Dnmt1 (Dnmt1o) is found in oocytes and preimplantation embryos. In the mouse, transcription and translation of this alternative Dnmt1 isoform leads to unusual localisation of the protein in the cytoplasm of oocytes and preimplantation embryos (Carlsson *et al.*, 1992; Cardoso & Leonhardt, 1998), and therefore does not interfere with the processive demethylation of the embryonic genome. The presence of somatic *DNMT1* in sheep oocytes, if translated, could result in the nuclear localisation of DNMT1 protein, thus maintaining methylation levels of both maternal and paternal genomes, and explaining the lack of differential demethylation in sheep preimplantation development. Further to this, *DNMT1* transcripts were found to be abundant in sheep oocytes, leading to the possibility of translation of significant amounts of somatic DNMT1 protein, contributing further to the maintenance of methylation levels.

In the mouse, the 'oocyte-specific' Dnmt1 protein enters the nucleus at the 8-cell stage of development, where it maintains methylation imprints for one cell cycle, then returns to the cytoplasm (Howell *et al.*, 2001). This leads to an unusual pattern of imprinting, with some cells retaining methylation imprints and others not. This presumably could also lead to a general variation in the DNA methylation status between individual cells. Differences in DNA methylation exist between the embryonic and extraembryonic lineages (Chapman *et al.*, 1984; Rossant *et al.*, 1986) and it has been tentatively suggested that the action of the oocyte-specific Dnmt1 at the 8-cell stage (the stage prior to compaction in the mouse) may contribute to the establishment of these differences (Dean & Ferguson-Smith, 2001).

Again, why this elegant mechanism is potentially absent in sheep is perplexing. The continual presence of *DNMT1* transcripts in sheep preimplantation embryos has been demonstrated in all stages from 2-cell embryos to blastocysts. If this correlates to the presence of somatic protein, then DNMT1 would most likely be localised in the nucleus throughout preimplantation development, albeit to varying extents. If DNA methylation

is required for marking inner cell mass and trophectoderm lineages, has the sheep evolved an alternative mechanism of performing this? A process which may also be inextricably linked to the differential methylation of the two cell lineages is X-inactivation. In the mouse, X-inactivation requires DNA methylation in embryonic (ICM derived) tissues, but not for the imprinted X-inactivation event occurring in extraembryonic (trophectoderm-derived) tissues (Sado *et al.*, 2000). In humans, random X-inactivation occurs in both the fetus and trophectoderm, however, it is not known if this is linked to the (unknown) methylation status of extraembryonic and embryonic lineages in the human peri-implantation embryo. The methylating mechanisms in human oocytes and preimplantation embryos also remain undetermined. In order to determine whether regulation of DNMT1 activity is linked with lineage-specific methylation in the sheep embryo, analysis of DNMT1 protein forms in sheep oocytes and embryos must be made to confirm the apparent presence of somatic DNMT1.

In addition to DNMT1, the presence of *de novo* methyltransferase transcripts in sheep oocytes and preimplantation embryos was also performed, in order to understand the DNA methylation changes occurring at this time. *Dnmt* transcripts were also analysed in mouse oocytes and preimplantation embryos, in order to define any significant species-specific differences. As mentioned previously, an increase in methylation most likely corresponding to *de novo* methylation was observed at the 8-16 cell stage in sheep (N. Beaujean, unpublished data). This correlates with the early *de novo* methylation observed in cow embryos (Dean *et al.*, 2001), and contrasts with findings in the mouse, where *de novo* methylation occurs post-implantation.

Analysis of *DNMT3A* transcripts in sheep oocytes and embryos demonstrated abundant transcripts in oocytes, followed by an apparent decrease and culminating in an increase at the morula/blastocyst stage. Analysis of *Dnmt3a* transcripts in the mouse demonstrated a similar pattern, although no increase in transcripts was apparent at the blastocyst stage. This is the first reported demonstration of *de novo* methyltransferase expression in mammalian preimplantation development. These results may suggest a

correlation between *DNMT3A* expression and the early onset of *de novo* methylation in the sheep preimplantation embryo. Unfortunately, expression data for *DNMT3B* was not conclusive in terms of such correlations, most likely due to experimental constraints.

The overall significance of the unusual methylation dynamics observed in the sheep is unclear. However, the unexpected observed changes in DNA methylation are reinforced by experimental evidence demonstrating species-specific DNA methyltransferase regulation. The question of why sheep have evolved such different mechanisms in relation to other mammalian species, including other closely related ruminants, remains to be determined.

Finally, the general function of DNA methylation in mammalian preimplantation development remains largely undetermined. It is likely that the differential methylation of maternal and paternal genomes characterised in the mouse are associated with the erasure and establishment of imprints. However, the involvement of DNA methylation in global transcriptional repression during preimplantation development remains unclear. Studies correlating expression and the methylation status of gene promoters in preimplantation embryos have been largely inconclusive (Walsh & Bestor, 1999). Furthermore, embryonic stem (ES) cells do not appear to require DNA methylation to survive unless induced to differentiate (Li *et al.*, 1992), and key components of the DNA methylation repression system, including MeCP1, MeCP2 and MBD2 are barely detectable in ES cells when compared to adult tissues (Meehan *et al.*, 1992; Hendrich & Bird, 1998). This would suggest DNA methylation is not extensively required for transcriptional repression of stage-specific genes in mammalian preimplantation embryos, but is required in subsequent stages as differentiation occurs. This contrasts with findings in Xenopus, where direct correlations between DNA methylation and the expression of stage-specific genes has been demonstrated (Stancheva *et al.*, 2002). Further work is clearly required to determine the precise requirements for DNA methylation in mammalian preimplantation development.

8.5 DNA methylation in cloned (SCNT) embryo development and the effects of *in vitro* embryo culture

Methods established to analyse DNA methylation in single oocytes and preimplantation embryos were also used to make comparisons between *in vivo*, IVF and SCNT-derived embryos. Although unavoidable experimental constraints prevented the drawing of firm conclusions, valuable information was obtained from these preliminary results.

To recap, it was demonstrated that sequence-specific reprogramming errors appeared to occur in SCNT blastocysts when compared with their IVF-derived counterparts. This manifested as increased levels of methylation in Satellite I sequence (heterochromatin), but not the OV-2 SINE (euchromatin). Furthermore, these reprogramming errors would appear to have occurred after the 8-16 cell stage of development, and thus may have resulted from inappropriate genome activation, rather than incomplete erasure of somatic methylation patterns. Also, comparisons of *in vivo*-derived and IVF-produced embryos appeared to show DNA methylation differences induced by *in vitro* embryo culture.

Several studies of reprogramming in large animal species have now been described, with differing extents of reprogramming observed (Dean *et al.*, 2001; Bourc'his *et al.*, 2001; Kang *et al.*, 2001a; Kang *et al.*, 2001b; Kang *et al.*, 2002). In addition, there was general disagreement on whether the active demethylation of DNA took place post-fusion, with Dean *et al.* (2001) observing a partial demethylation immediately after fusion, whilst the remaining studies failed to observe this. The reasons for this may lie in the different methods used by the different groups. Firstly, each of the cow studies used different cytological or molecular techniques to analyse DNA methylation patterns, which may have resulted in the analysis of different DNA sequences in each study. Also, there was no uniformity regarding the choice or treatment of donor cells used for SCNT. Dean *et al.* (2001) used non-serum starved fetal fibroblasts whilst Kang *et al.* (2001a; 2002) and the experiments described in this thesis were performed with serum-starved fetal fibroblasts. In contrast, Bourc'his *et al.* (2001) used serum-starved adult

skin fibroblasts. Finally, the numbers of embryos used for the cytological studies, were very low and no statistical analysis of differences observed in methylation were described. When taking into consideration these points, the overall picture of methylation reprogramming in large animal clones is far from clear.

More specific reprogramming analysis in mouse SCNT embryos has been recently described (Humpherys *et al.*, 2001; Inoue *et al.*, 2002). Both of these studies focused on the expression of imprinted genes in both the fetus and placenta, as mis-expression of these genes causes phenotypes similar to the fetal overgrowth and placental defects often observed as a result of SCNT (reviewed in Young & Fairburn, 2000). The study by Humpherys *et al.* (2001), described a range of imprinted gene expression anomalies in clones produced from ES cells. However, abnormal expression patterns were also observed in the ES clones from which donor nuclei were derived, leading to the suggestion that intrinsic errors within the donor cell nuclei were more likely to cause aberrant expression rather than inefficient reprogramming (Rideout *et al.*, 2001). Indeed, it has previously been shown that *in vitro* culture of ES cells can lead to aberrant methylation and expression of imprinted genes (Dean *et al.*, 1998).

The study described by Inoue *et al.* (2002) analysed imprinted gene expression in mice cloned from somatic cells, demonstrating comparable expression of imprinted genes in normal and cloned fetuses and no fetal overgrowth. Analysis of the unusually large placenta of these cloned fetuses, however, demonstrated aberrant expression of specific imprinted and non-imprinted genes, including *Peg1/Mest*, *Igf2bp* and *Vegfr2*. This would suggest firstly, that imprints in somatic nuclei are more faithfully capitulated than those in ES cells after nuclear transfer. Secondly, gene expression appears to be dysregulated on a wide scale in the placenta of cloned mice, as opposed to the fetus. As many genes involved in placental development will be tightly repressed in somatic lineages, it is feasible that this repression is difficult to reverse after nuclear transfer, resulting in maintained silencing of these important genes. The finding that methylation patterns are abnormal in the trophectodermal cells of sheep NT blastocysts would

correlate with this hypothesis (Kang *et al.*, 2002). The high incidence of placental defects associated with clones of several mammalian species provides further support to this theory (De Sousa *et al.*, 2001; Humpherys *et al.*, 2001; Hill *et al.*, 2000). It would clearly be of interest to differentiate the fetal abnormalities in cloned mammals caused by placental defects, and those that occur independently.

The variations in aberrant gene expression of clones generated from ES cells or somatic cells raises the question of cell reprogrammability. The apparent observation that embryos cloned from ES-cells were more likely to develop abnormally than somatic clones was surprising, as the undifferentiated, totipotent nature of ES cells should make them easier to reprogram. However, as mentioned previously, these abnormalities are thought to be attributed to epigenetic instability in ES cells, rather than reprogramming errors. To date, studies of epigenetic reprogramming in large animals have analysed clones generated using fetal or skin fibroblasts (Kang *et al.*, 2001a; Bourc'his *et al.*, 2001; Dean *et al.*, 2001; work described in this thesis). It is very possible that the methylation reprogramming errors reported in these studies are not representative of all cell types. For example, would a highly specialised heart cell be more difficult to reprogram than a fetal fibroblast? Results described in this thesis found both Satellite I and OV-2 SINE sequences to be more highly methylated in heart cells than fetal fibroblasts. If this reflects the overall status of these cell types, are heart cells more highly methylated and thus more difficult to reprogram than fibroblasts? The number of different somatic cell types used to produce cloned animals to date is limited (refer to Table 1.2) and generally reflects the use of less specialised cell types for SCNT. A recent report described the cloning of mice from highly differentiated B- and T-cells (Hochedlinger & Jaenisch, 2002). However, this required a modified process, with isolation of ES cells from initially cloned blastocysts, and subsequent injection into a tetraploid blastocyst to produce embryos. This, and the very low success rate (10 times lower than when using fibroblasts or cumulus cells as donors) would suggest that terminally differentiated cells are inherently difficult to reprogram.

The studies of DNA methylation and gene expression patterns are only the tip of the iceberg in terms of investigating aberrant reprogramming in clones. As well as the oocyte failing to recognise and demethylate the chromatin of somatic nuclei, the donor cell may also carry a host of proteins that could potentially interfere with normal reprogramming mechanisms. For example, several components of the DNA methylation repression system, including MeCP1 and MeCP2 are barely detectable in ES cells (Meehan *et al.*, 1992; Hendrich & Bird, 1998) and work described in this thesis demonstrated barely detectable levels of *MBD2* in the preimplantation embryo. As these genes are expressed at relatively high levels in adult tissues, the SCNT process could introduce large amounts of these proteins into the fused embryo. In addition to this, aberrant expression of 'somatic'-specific genes could also occur within the fused SCNT embryo, or be activated at an inappropriate stage of development. Considering the observations made in SCNT-derived sheep embryos, where methylation levels increased after the 8-16 cell stage relative to their 'normal' counterparts, premature activation of a *de novo* methyltransferase gene may be responsible. Finally, it should be reiterated at this point, that DNA methylation is not the only epigenetic modification likely to be reprogrammed incorrectly. Studies of other modifications such as histone methylation or changes in chromatin structure will give further insights to the global reprogramming events in cloned embryos.

A final note

Assumptions that the mouse is a representative model of mammalian epigenetic reprogramming must now be faltering. Even though differential demethylation of paternal and maternal genomes is conserved in most mammalian species examined (Dean *et al.*, 2001), it cannot be assumed that other epigenetic regulation mechanisms are identical in all mammals. This is also highlighted by previous and continuing attempts to clone various mammalian species, where species-specific epigenetic differences may contribute to the ease of cloning. In light of this, any claims that one or another species is a good model for epigenetic reprogramming in the human cannot be regarded as accurate. The lessons learned from sheep should be taken into heed and widespread studies of epigenetic reprogramming mechanisms should be initiated in a diverse range of mammalian species.

8.6 Future work

The work described in this thesis and parallel experiments using an antibody directed against m^5C (performed by N. Beaujean), have demonstrated unexpected methylation changes in sheep preimplantation development. To further investigate this, techniques such as bisulfite sequencing should be employed to gain more detailed sequence-specific methylation data at different stages of development. As well as assessing the methylation status of multiple CpG sites within Satellite I DNA sequences (to see if the levels observed in this study are indicative of the entire sequence), the methylation status of other repetitive sequences, or promoters of developmentally-regulated genes could be analysed. Furthermore, the methylation status of differentially methylated regions of imprinted genes could be analysed at different stages of development to see if they are maintained, demethylated or *de novo* methylated at specific stages. Similar experiments should also be performed using IVF or SCNT-derived preimplantation embryos, allowing more detailed analysis of the sequence-specific reprogramming errors and *in vitro* culture effects hinted at in this study.

The presence of somatic *DNMT1* in mature sheep oocytes also presented a potentially significant species-specific difference. It must be a high priority to ascertain whether these transcripts are translated in the oocyte, corresponding to an abundant presence of somatic DNMT1 protein in the early preimplantation embryo. Western blots should be performed to analyse any protein variants present in adult tissues and oocytes/preimplantation embryos. In addition, the localisation of DNMT1 protein in oocytes and preimplantation embryos should be analysed by immunolocalisation techniques to see if the presence of somatic *DNMT1* correlates with nuclear localisation, or whether the sheep has evolved an alternative form of DNMT1 regulation.

It would also be of great interest to see if the somatic *DNMT1* is present in mature cow oocytes. The differential demethylation of paternal and maternal genomes is observed in

the cow; it would interesting to see if the methylation differences so far observed between the cow and sheep extend to the regulation of DNMT1.

To investigate the role of aberrant methylating activity in cloned embryo development, techniques should be developed to assess *DNMT* transcript or protein levels in cloned embryos. Quantitative RT-PCR would allow relative expression levels of *DNMT*s to be compared in normal and cloned embryos, whilst immunolocalisation studies would allow determination of whether somatic nuclei introduce large quantities of DNMTs into the embryo, or whether the genes are inappropriately expressed from the cloned embryo genome.

Finally, the comparison of *in vivo* derived, *in vitro* fertilised and SCNT-derived embryos was limited by experimental constraints. As well as effects of *in vitro* culture, the oocytes used in these methods may differ considerably in terms of developmental capacity depending on the source (superovulation, *in vivo* maturation, *in vitro* maturation) and the stage of follicle the oocyte was derived from. Transcript levels of DNA methyltransferases and associated proteins could be analysed (using RT-PCR) in oocytes derived from different sources, to see if levels differ substantially. Also, as imprints are known to be established in growing and maturing oocytes, the methylation status of key imprinted genes should be analysed in oocytes derived from different sources to see if this may also contribute to poor developmental capacity.

References

Aapola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M, Shintani A, Minoshima S, Krohn K, Antonarakis SE, Shimizu N, Kudoh J, Peterson P (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 65:293-298

Aapola U, Lyle R, Krohn K, Antonarakis SE, Peterson P (2001). Isolation and initial characterization of the mouse Dnmt3l gene. Cytogenet Cell Genet 92:122-126

Adenot PG, Mercier Y, Renard JP, Thompson EM (1997). Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. Development JID - 8701744 124:4615-4625

Aguirre-Arteta AM, Grunewald I, Cardoso MC, Leonhardt H (2000). Expression of an alternative Dnmt1 isoform during muscle differentiation. Cell Growth Differ 11:551-559

Antequera F, Bird A (1993). Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A 90:11995-11999

Ariel M, Robinson E, McCarrey JR, Cedar H (1995). Gamete-specific methylation correlates with imprinting of the murine Xist gene. Nat Genet 9:312-315

Bachman KE, Rountree MR, Baylin SB (2001). Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. J Biol Chem 276:32282-32287

Barton SC, Arney KL, Shi W, Niveleau A, Fundele R, Surani MA, Haaf T (2001). Genome-wide methylation patterns in normal and uniparental early mouse embryos. Hum Mol Genet 10:2983-2987

Barton SC, Surani MA, Norris ML (1984). Role of paternal and maternal genomes in mouse development. Nature 311:374-376

Barwise JL, Walker JH (1996). Annexins II, IV, V and VI relocate in response to rises in intracellular calcium in human foreskin fibroblasts. J Cell Sci 109 (Pt 1):247-255

Baylin SB, Herman JG (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet 16:168-174

Bell AC, Felsenfeld G (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405:482-485

Bestor TH (1988). Cloning of a mammalian DNA methyltransferase. Gene 74:9-12

Bestor TH (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. EMBO J 11:2611-2617

Bestor TH (2000). The DNA methyltransferases of mammals. Hum Mol Genet 9:2395-2402

Bestor TH, Verdine GL (1994). DNA methyltransferases. Curr Opin Cell Biol 6:380-389

Bhattacharya SK, Ramchandani S, Cervoni N, Szyf M (1999). A mammalian protein with specific demethylase activity for mCpG DNA. Nature 397:579-583

Bird A (2002). DNA methylation patterns and epigenetic memory. Genes Dev 16:6-21

Bird AP (1986). CpG-rich islands and the function of DNA methylation. Nature 321:209-213

Boeke J, Ammerpohl O, Kegel S, Moehren U, Renkawitz R (2000). The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. J Biol Chem 275:34963-34967

Bordignon V, Clarke HJ, Smith LC (2001). Factors controlling the loss of immunoreactive somatic histone H1 from blastomere nuclei in oocyte cytoplasm: a potential marker of nuclear reprogramming. Dev Biol 233:192-203

Bouniol C, Nguyen E, Debey P (1995). Endogenous transcription occurs at the 1-cell stage in the mouse embryo. Exp Cell Res 218:57-62

Bourc'his D, Le Bourhis D, Patin D, Niveleau A, Comizzoli P, Renard J, Viegas-Pequignot E (2001a). Delayed and incomplete reprogramming of chromosome methylation patterns in bovine cloned embryos. Curr Biol 11:1542-1546

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001b). Dnmt3L and the Establishment of Maternal Genomic Imprints. Science 294:2536-2539

Brandeis M, Kafri T, Ariel M, Chaillet JR, McCarrey J, Razin A, Cedar H (1993). The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. EMBO J 12:3669-3677

Briggs R, King TJ (1952). Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. Proc Natl Acad Sci U S A 38:455-463

Briggs R, King TJ (1957). Changes in the nuclei of differentiating endoderm cells as revealed by nuclear transplantation. J Morphol 100:269-312

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71(3):515-26

Bromhall JD (1975). Nuclear transplantation in the rabbit egg. Nature 258:719-722

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349(6304):38-4

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 371(3):527-42

Burkin DJ, Broad TE, Jones C (1996). The chromosomal distribution and organization of sheep satellite I and II centromeric DNA using characterized sheep-hamster somatic cell hybrids. Chromosome Res 4:49-55

Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, Morrice DR, Paton IR, Smith J, Windsor D, Sazanov A, Fries R, Waddington D (1999). The dynamics of chromosome evolution in birds and mammals. Nature 402:411-413

Campbell KH, McWhir J, Ritchie WA, Wilmut I (1996). Sheep cloned by nuclear transfer from a cultured cell line. Nature 380:64-66

Cardoso MC, Leonhardt H (1999). DNA methyltransferase is actively retained in the cytoplasm during early development. J Cell Biol 147:25-32

Carlson LL, Page AW, Bestor TH (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev 6:2536-2541

Chaillet JR, Vogt TF, Beier DR, Leder P (1991). Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell 66:77-83

Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. Science 295:345-347

Chapman V, Forrester L, Sanford J, Hastie N, Rossant J (1984). Cell lineage-specific undermethylation of mouse repetitive DNA. Nature 307:284-286

Charlier C, Segers K, Wagenaar D, Karim L, Berghmans S, Jaillon O, Shay T, Weissenbach J, Cockett N, Gyapay G, Georges M (2001). Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. Genome Res 11:850-862

Chesne P, Adenot PG, Viglietta C, Baratte M, Boulanger L, Renard JP (2002). Cloned rabbits produced by nuclear transfer from adult somatic cells. Nat Biotechnol 20:366-369

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277:1996-2000

Colot, V. and Rossignol, J-L (1999). Eukaryotic DNA methylation as an evolutionary device. Bioessays 21:402-411.

Dai Y, Vaught TD, Boone J, Chen SH, Phelps CJ, Ball S, Monahan JA, Jobst PM, McCreath KJ, Lamborn AE, Cowell-Lucero JL, Wells KD, Colman A, Polejaeva IA, Ayares DL (2002). Targeted disruption of the alpha1,3-galactosyltransferase gene in cloned pigs. Nat Biotechnol 20:251-255

Davis OK, Rosenwaks Z (2001). Superovulation strategies for assisted reproductive technologies. Semin Reprod Med 19:207-212

De Sousa PA, King T, Harkness L, Young LE, Walker SK, Wilmut I (2001). Evaluation of gestational deficiencies in cloned sheep fetuses and placentae. Biol Reprod 65:23-30

Dean W, Bowden L, Aitchison A, Klose J, Moore T, Meneses JJ, Reik W, Feil R (1998). Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. Development 125:2273-2282

Dean W, Ferguson-Smith A (2001). Genomic imprinting: Mother maintains methylation marks. Curr Biol 11:R527-R530

Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E, Reik W (2001). Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. Proc Natl Acad Sci U S A 98:13734-13738

Denning C, Burl S, Ainslie A, Bracken J, Dinnyes A, Fletcher J, King T, Ritchie M, Ritchie WA, Rollo M, De Sousa P, Travers A, Wilmut I, Clark AJ (2001). Deletion of the alpha(1,3)galactosyl transferase (GGTA1) gene and the prion protein (PrP) gene in sheep. Nat Biotechnol 19:559-562

Dennis K, Fan T, Geiman T, Yan Q, Muegge K (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev 15:2940-2944

Doerfler W, Weisshaar B, Hoeveler A, Knebel D, Muller U, Dobrzanski P, Lichtenberg U, Achten S, Hermann R (1988). Promoter inhibition by DNA methylation: a reversible signal. Gene 74:129-133

Doherty AS, Bartolomei MS, Schultz RM (2002). Regulation of stage-specific nuclear translocation of Dnmt1o during preimplantation mouse development. Dev Biol 242:255-266

Doherty AS, Mann MR, Tremblay KD, Bartolomei MS, Schultz RM (2000). Differential effects of culture on imprinted H19 expression in the preimplantation mouse embryo. Biol Reprod 62:1526-1535

El-Maarri O, Buiting K, Peery EG, Kroisel PM, Balaban B, Wagner K, Urman B, Heyd J, Lich C, Brannan CI, Walter J, Horsthemke B (2001). Maternal methylation imprints on human chromosome 15 are established during or after fertilization. Nat Genet 27:341-344

Enders AC, Welsh AO (1993). Structural interactions of trophoblast and uterus during hemochorial placenta formation. J Exp Zool 266:578-587

Fairburn HR, Young LE, Hendrich BD (2002). Epigenetic reprogramming: how now, cloned cow? Curr Biol 12:R68-R70

Feil R, Khosla S, Cappai P, Loi P (1998). Genomic imprinting in ruminants: allele-specific gene expression in parthenogenetic sheep. Mamm Genome 9:831-834

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 89:1827-1831

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet 24:88-91

Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T, (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J 20:2536-2544

Gandolfi TA, Gandolfi F (2001). The maternal legacy to the embryo: cytoplasmic components and their effects on early development. Theriogenology 55:1255-1276

Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H, Higgs DR (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet 24:368-371

Gibbons RJ, Picketts DJ, Villard L, Higgs DR (1995). Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell 80:837-845

Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP (1985). Hypomethylation of DNA from benign and malignant human colon neoplasms. Science 228:187-190

Grewal SI, Elgin SC (2002). Heterochromatin: new possibilities for the inheritance of structure. Curr Opin Genet Dev 12:178-187

Gurdon JB (1962). The transplantation of living cell nuclei. J Embryol Exp Morphol 10:622-640

Hagemann LJ, Peterson AJ, Weilert LL, Lee RS, Tervit HR (1998). In vitro and early in vivo development of sheep gynogenones and putative androgenones. Mol Reprod Dev 50:154-162

Haines TR, Rodenhiser DI, Ainsworth PJ (2001). Allele-Specific Non-CpG Methylation of the Nf1 Gene during Early Mouse Development. Dev Biol 240:585-598

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A 96:14412-14417

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM (2000_. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405:486-489

Hendrich B, Bird A (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18:6538-6547

Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev 15:710-723

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A (1999a). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401:301-304

Hendrich B, Abbott C, McQueen H, Chambers D, Cross S, Bird A (1999b). Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. Mamm Genome 10:906-912

Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarra JR, Linehan WM (1994). Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A 91:9700-9704

Hill JR, Burghardt RC, Jones K, Long CR, Looney CR, Shin T, Spencer TE, Thompson JA, Winger QA, Westhusin ME (2000). Evidence for placental abnormality as the major cause of mortality in first-trimester somatic cell cloned bovine fetuses. Biol Reprod 63:1787-1794

Hochedlinger K, Jaenisch R (2002). Monoclonal mice generated by nuclear transfer from mature B and T donor cells. Nature 415:1035-1038

Holliday R (1987). The inheritance of epigenetic defects. Science 238:163-170

Holliday R (1994). Epigenetics: an overview. Dev Genet 15:453-457

Holliday R, Pugh JE (1975). DNA modification mechanisms and gene activity during development. Science 187:226-232

Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 104:829-838

Howlett, S. K. and Reik, W (1991) Methylation levels of maternal and paternal genomes during preimplantation development. Development 113, 119-127

Hubacek J (1992). Biological function of DNA methylation. Folia Microbiol (Praha) 37:323-329

Humpherys D, Eggan K, Akutsu H, Hochedlinger K, Rideout WM, Biniszkiewicz D, Yanagimachi R, Jaenisch R (2001). Epigenetic instability in ES cells and cloned mice. Science 293:95-97

Inoue K, Kohda T, Lee J, Ogonuki N, Mochida K, Noguchi Y, Tanemura K, Kaneko-Ishino T, Ishino F, Ogura A (2002). Faithful expression of imprinted genes in cloned mice. Science 295:297

Issa JP (2000). CpG-island methylation in aging and cancer. Curr Top Microbiol Immunol 249:101-118

Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB (1994). Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. Nat Genet 7:536-540

Issa JP, Vertino PM, Boehm CD, Newsham IF, Baylin SB (1996). Switch from monoallelic to biallelic human IGF2 promoter methylation during aging and carcinogenesis. Proc Natl Acad Sci U S A 93:11757-11762

Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E, Jaenisch R (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet 27:31-39

Jeddeloh JA, Stokes TL, Richards EJ (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet 22:94-97

Jones PA, Laird PW (1999). Cancer epigenetics comes of age. Nat Genet 21:163-167

Jost JP (1993). Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. Proc Natl Acad Sci U S A 90:4684-4688

Jost JP, Fremont M, Siegmann M, Hofsteenge J (1997). The RNA moiety of chick embryo 5-methylcytosine- DNA glycosylase targets DNA demethylation. Nucleic Acids Res 25:4545-4550

Kanai Y, Ushijima S, Kondo Y, Nakanishi Y, Hirohashi S (2001). DNA methyltransferase expression and DNA methylation of CPG islands and peri-centromeric satellite regions in human colorectal and stomach cancers. Int J Cancer 91:205-212

Kang YK, Koo DB, Park JS, Choi YH, Chung AS, Lee KK, Han YM (2001a). Aberrant methylation of donor genome in cloned bovine embryos. Nat Genet 28:173-177

Kang YK, Koo DB, Park JS, Choi YH, Kim HN, Chang WK, Lee KK, Han YM (2001b). Typical demethylation events in cloned pig embryos. clues on species-specific differences in epigenetic reprogramming of a cloned donor genome. J Biol Chem 276:39980-39984

Kang YK, Park JS, Koo DB, Choi YH, Kim SU, Lee KK, Han YM (2002). Limited demethylation leaves mosaic-type methylation states in cloned bovine pre-implantation embryos. EMBO J 21:1092-1100

Kato Y, Tani T, Sotomaru Y, Kurokawa K, Kato J, Doguchi H, Yasue H, Tsunoda Y (1998). Eight calves cloned from somatic cells of a single adult. Science 282:2095-2098

Kato Y, Tani T, Tsunoda Y (2000). Cloning of calves from various somatic cell types of male and female adult, newborn and fetal cows. J Reprod Fertil 120:231-237

Keefer CL, Bhatia B, Lazaris A, Begin I, Kafidi N, Bilodeau A, Wang B, Zhou FJ, Downey BR, Baldassare H, Karatzas CN (2002). Efficient production of viable goat offspring following nuclear transfer using adult somatic cells. Biology of Reproduction 218, Suppl. 1:218(Abstract)

King GJ, Atkinson BA, Robertson HA (1982). Implantation and early placentation in domestic ungulates. J Reprod Fertil Suppl 31:17-30

Kubota Y, Kimura T, Hashimoto K, Tokugawa Y, Nobunaga K, Azuma C, Saji F, Murata Y (1996). Structure and expression of the mouse oxytocin receptor gene. Mol Cell Endocrinol 124:25-32

Kurnit DM, Brown FL, Maio JJ (1978). Mammalian repetitive DNA sequences in a stable Robertsonian system. Characterization, in situ hybridizations, and cross-species hybridizations of repetitive DNAs in calf, sheep, and goat chromosomes. Cytogenet Cell Genet 21:145-167

Latham KE (1999). Epigenetic modification and imprinting of the mammalian genome during development. Curr Top Dev Biol 43:1-49

Lauster R, Trautner TA, Noyer-Weidner M (1989). Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. J Mol Biol 206:305-312

Lee J, Inoue K, Ono R, Ogonuki N, Kohda T, Kaneko-Ishino T, Ogura A, Ishino F (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development 129:1807-1817

Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122:3195-3205

Lenstra JA, van Boxtel JA, Zwaagstra KA, Schwerin M (1993). Short interspersed nuclear element (SINE) sequences of the Bovidae. Anim Genet 24:33-39

Leonhardt H, Page AW, Weier HU, Bestor TH (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71:865-873

Li E, Beard C, Jaenisch R (1993). Role for DNA methylation in genomic imprinting. Nature 366:362-365

Li E, Bestor TH, Jaenisch R (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915-926

Liu WM, Maraia RJ, Rubin CM, Schmid CW (1994). Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. Nucleic Acids Res 22:1087-1095

Liu Y, Oakeley EJ, Sun L, Jost JP (1998). Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. Nucleic Acids Res 26:1038-1045

Lyko F, Paro R (1999). Chromosomal elements conferring epigenetic inheritance. Bioessays 21:824-832

Lyko F, Ramsahoye BH, Jaenisch R (2000). DNA methylation in Drosophila melanogaster. Nature 408:538-540

Lyko F, Ramsahoye BH, Kashevsky H, Tudor M, Mastrangelo MA, Orr-Weaver TL, Jaenisch R (1999). Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in Drosophila. Nat Genet 23:363-366

Lyle R, Watanabe D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP (2000). The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nat Genet 25:19-21

Macleod, D., Clark, V. H., and Bird, A (1999). Absence of genome-wide changes in DNA methylation during development of the zebrafish. Nature Genetics 23, 139-140.

Mann JR, (2001). Imprinting in the germ line. Stem Cells 19:287-294

Mansour R (1998). Intracytoplasmic sperm injection: a state of the art technique. Hum Reprod Update 4:43-56

Margot JB, Aguirre-Arteta AM, Di Giacco BV, Pradhan S, Roberts RJ, Cardoso MC, Leonhardt H (2000). Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. J Mol Biol 297:293-300

Martienssen RA, Colot V (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science 293:1070-1074

Matzke M, Matzke AJ, Kooter JM (2001). RNA: guiding gene silencing. Science 293:1080-1083

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T (2000). Demethylation of the zygotic paternal genome. Nature 403:501-502

McDowell TL, Gibbons RJ, Sutherland H, O'Rourke DM, Bickmore WA, Pombo A, Turley H, Gatter K, Picketts DJ, Buckle VJ, Chapman L, Rhodes D, Higgs DR (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. Proc Natl Acad Sci U S A 96:13983-13988

McGrath J, Solter D (1984). Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro. Science 226:1317-1319

McLaren RJ, Montgomery GW (1999). Genomic imprinting of the insulin-like growth factor 2 gene in sheep. Mamm Genome 10:588-591

Meehan RR, Lewis JD, Bird A (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res 20:5085-5092

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell 58:499-507

Menezo YJ, Veiga A, Pouly JL (2000). Assisted reproductive technology (ART) in humans: facts and uncertainties. Theriogenology 53:599-610

Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development 125:889-897

Migeon BR, Chowdhury AK, Dunston JA, McIntosh I (2001). Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation. Am J Hum Genet 69:951-960

Miller OJ, Schnedl W, Allen J, Erlanger BF (1974). 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature 251:636-637

Monk, M., Adams, R. L. P., and Rinaldi, A (1991). Decrease in DNA methylase activity during preimplantation development in the mouse. Development 112, 189-192.

Monk, M., Boubelik, M., and Lehnert, S (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development 99, 371-382.

Moore T, Haig D (1991). Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet. 7, 45-9

Nan X, Meehan RR, Bird A (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res 21:4886-4892

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393:386-389

Ng HH, Jeppesen P, Bird A (2000). Active repression of methylated genes by the chromosomal protein MBD1. Mol Cell Biol 20:1394-1406

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet 23:58-61

Nicolas RH, Goodwin GH (1996). Molecular cloning of polybromo, a nuclear protein containing multiple domains including five bromodomains, a truncated HMG-box, and two repeats of a novel domain. Gene 175:233-240

Noden DM, De Lahunta A (1985). The embryology of domestic animals. Developmental mechanisms and malformations, Baltimore/London: Williams & Wilkins

Nonchev S, Tsanev R (1990). Protamine-histone replacement and DNA replication in the male mouse pronucleus. Mol Reprod Dev 25:72-76

Novak U (1984). Structure and properties of a highly repetitive DNA sequence in sheep. Nucleic Acids Res 12:2343-2350

Obata Y, Kono T (2002). Maternal primary imprinting is established at a specific time for each gene throughout oocyte growth. J Biol Chem 277:5285-5289

Ogura A, Inoue K, Takano K, Wakayama T, Yanagimachi R (2000). Birth of mice after nuclear transfer by electrofusion using tail tip cells. Mol Reprod Dev 57:55-59

Ohsawa K, Imai Y, Ito D, Kohsaka S (1996). Molecular cloning and characterization of annexin V-binding proteins with highly hydrophilic peptide structure. J Neurochem 67:89-97

Okano, M, Xie, S., and Li, E (1998a). Cloning and characterization of a novel mammalian DNA (cytosine-5) methyltransferases. Nature Genetics 19, 219-220.

Okano M, Bell DW, Haber DA, Li E, 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247-257

Okano M, Xie S, Li E (1998b). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res 26:2536-2540

Olek A, Walter J (1997). The pre-implantation ontogeny of the H19 methylation imprint. Nat Genet 17:275-276

Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, Walter J (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol 10:475-478

Panning B, Jaenisch R (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev 10:1991-2002

Paroush Z, Keshet I, Yisraeli J, Cedar H (1990). Dynamics of demethylation and activation of the alpha-actin gene in myoblasts. Cell 63:1229-1237

Paulsen M, Takada S, Youngson NA, Benchaib M, Charlier C, Segers K, Georges M, Ferguson-Smith AC (2001). Comparative sequence analysis of the imprinted Dlk1-Gtl2 locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the Igf2-H19 region. Genome Res:2085-2094

Polejaeva IA, Chen SH, Vaught TD, Page RL, Mullins J, Ball S, Dai Y, Boone J, Walker S, Ayares DL, Colman A, Campbell KH (2000). Cloned pigs produced by nuclear transfer from adult somatic cells. Nature 407:86-90

Posfai J, Bhagwat AS, Posfai G, Roberts RJ (1989). Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res 17:2421-2435

Prak ET, Kazazian HHJ (2000). Mobile elements and the human genome. Nat Rev Genet 1:134-144

Proffitt JH, Davie JR, Swinton D, Hattman S (1984). 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol 4:985-988

Prokhortchouk A, Hendrich B, Jorgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev 15:1613-1618

Raabe EH, Abdurrahman L, Behbehani G, Arceci RJ (2001). An SNF2 factor involved in mammalian development and cellular proliferation. Dev Dyn 221:92-105

Ram PT, Schultz RM (1993). Reporter gene expression in G2 of the 1-cell mouse embryo. Dev Biol 156:552-556

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97:5237-5242

Razin A, Shemer R (1995). DNA methylation in early development. Hum Mol Genet 4 Spec No:1751-1755

Razin A, Szyf M, Kafri T, Roll M, Giloh H, Scarpa S, Carotti D, Cantoni GL (1986). Replacement of 5-methylcytosine by cytosine: a possible mechanism for transient DNA demethylation during differentiation. Proc Natl Acad Sci U S A 83:2827-2831

Reik W, Kelsey G, Walter J (1999). Dissecting de novo methylation. Nat Genet 23:380-382

Reik W, Walter J (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet 2:21-32

Rhee I, Jair KW, Yen RW, Lengauer C, Herman JG, Kinzler KW, Vogelstein B, Baylin SB, Schuebel KE (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. Nature 404:1003-1007

Rice JC, Allis CD (2001). Code of silence. Nature 414:258-261

Rideout WM, Eggan K, Jaenisch R (2001). Nuclear cloning and epigenetic reprogramming of the genome. Science 293:1093-1098

Riggs AD (1975). X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14:9-25

Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, Jones PA (1999). The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res 27:2291-2298

Rossant J, Sanford JP, Chapman VM, Andrews GK (1986). Undermethylation of structural gene sequences in extraembryonic lineages of the mouse. Dev Biol 117:567-573

Rother KI, Silke J, Georgiev O, Schaffner W, Matsuo K (1995). Influence of DNA sequence and methylation status on bisulfite conversion of cytosine residues. Anal Biochem 231:263-265

Rougier N, Bourc'his D, Gomes DM, Niveleau A, Plachot M, Paldi A, Viegas-Pequignot E (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes Dev 12:2108-2113

Rountree MR, Bachman KE, Baylin SB (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet 25:269-277

Russell PJ, Welsch JA, Rachlin EM, McCloskey JA (1987). Different levels of DNA methylation in yeast and mycelial forms of Candida albicans. J Bacteriol 169:4393-4395

Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E (2000. X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol 225:294-303

Saluz HP, Feavers IM, Jiricny J, Jost JP (1988). Genomic sequencing and in vivo footprinting of an expression-specific DNase I-hypersensitive site of avian vitellogenin II promoter reveal a demethylation of a mCpG and a change in specific interactions of proteins with DNA. Proc Natl Acad Sci U S A 85:6697-6700

Sanford, J. P., Clark, H. J., Chapman, V. M., and Rossant, J (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes & Development 1, 1039-1046

Schnieke, A. E., Kind, A. J., Ritchie, W. A., Mycock, K., Scott, A. R., Ritchie, M., Wilmut, I., Colman, A., and Campbell, K. H. S (1997). Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. Science 278, 2130-2133.

Selker EU (1998). Trichostatin A causes selective loss of DNA methylation in Neurospora. Proc Natl Acad Sci U S A 95:9430-9435

Shemer R, Walsh A, Eisenberg S, Breslow JL, Razin A (1990). Tissue-specific methylation patterns and expression of the human apolipoprotein AI gene. J Biol Chem 265:1010-1015

Shiga K, Fujita T, Hirose K, Sasae Y, Nagai T (1999). Production of calves by transfer of nuclei from cultured somatic cells obtained from Japanese black bulls. Theriogenology 52:527-535

Shin T, Kraemer D, Pryor J, Liu L, Rugila J, Howe L, Buck S, Murphy K, Lyons L, Westhusin M (2002). A cat cloned by nuclear transplantation. Nature 415:859

Simpson VJ, Johnson TE, Hammen RF (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res 14:6711-6719

Singer-Sam J, LeBon JM, Tanguay RL, Riggs AD (1990). A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. Nucleic Acids Res 18:687

Solter D (2000). Mammalian cloning: advances and limitations. Nat Rev Genet 1:199-207

Stancheva I, El-Maarri O, Walter J, Niveleau A, Meehan RR (2002). DNA methylation at promoter regions regulates the timing of gene activation in Xenopus laevis embryos. Dev Biol 243:155-165

Stirzaker C, Millar DS, Paul CL, Warnecke PM, Harrison J, Vincent PC, Frommer M, Clark SJ (1997). Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. Cancer Res 57:2229-2237

Stoger R, Kubicka P, Liu CG, Kafri T, Razin A, Cedar H, Barlow DP (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. Cell 73:61-71

Swisher JF, Rand E, Cedar H, Marie PA (1998). Analysis of putative RNase sensitivity and protease insensitivity of demethylation activity in extracts from rat myoblasts. Nucleic Acids Res 26:5573-5580

Szabo P, Tang SH, Rentsendorj A, Pfeifer GP, Mann JR (2000). Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr Biol 10:607-610

Tada T, Tada M, Hilton K, Barton SC, Sado T, Takagi N, Surani MA (1998_. Epigenotype switching of imprintable loci in embryonic germ cells. Dev Genes Evol 207:551-561

Tam PP, Zhou SX, Tan SS (1994). X-chromosome activity of the mouse primordial germ cells revealed by the expression of an X-linked lacZ transgene. Development 120:2925-2932

Tatematsu KI, Yamazaki T, Ishikawa F (2000). MBD2-MBD3 complex binds to hemi-methylated DNA and forms a complex containing DNMT1 at the replication foci in late S phase. Genes Cells 5:677-688

Telford NA, Watson AJ, Schultz GA (1990). Transition from maternal to embryonic control in early mammalian development: a comparison of several species. Mol Reprod Dev 26:90-100

Thorvaldsen JL, Bartolomei MS (2000). Molecular biology. Mothers setting boundaries. Science 288:2145-2146

Tremblay KD, Duran KL, Bartolomei MS (1997). A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. Mol Cell Biol 17:4322-4329

Tweedie S, Charlton J, Clark V, Bird A (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol 17:1469-1475

Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993). Arabidopsis thaliana DNA methylation mutants. Science 260:1926-1928

Waddington CH (1942). The epigenotype. Endeavour 1:18

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet 23:62-66

Wakayama T, Perry AC, Zuccotti M, Johnson KR, Yanagimachi R (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature 394:369-374

Walker SK, Hill JL, Kleemann DO, Nancarrow CD (1996). Development of ovine embryos in synthetic oviductal fluid containing amino acids at oviductal fluid concentrations. Biol Reprod 55:703-708

Walsh CP, Bestor TH (1999). Cytosine methylation and mammalian development. Genes Dev 13:26-34

Walsh CP, Chaillet JR, Bestor TH (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet 20:116-117

Warnecke PM, Clark SJ (1999). DNA methylation profile of the mouse skeletal alpha-actin promoter during development and differentiation. Mol Cell Biol 19:164-172

Weiss A, Keshet I, Razin A, Cedar H (1996). DNA demethylation in vitro: involvement of RNA. Cell 86:709-718

Wells DN, Misica PM, Tervit HR (1999). Production of cloned calves following nuclear transfer with cultured adult mural granulosa cells. Biol Reprod 60:996-1005

Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH (1997). Viable offspring derived from fetal and adult mammalian cells. Nature 385:810-813

Woodcock DM, Lawler CB, Linsenmeyer ME, Doherty JP, Warren WD (1997.)Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. J Biol Chem 272:7810-7816

Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature 389:745-749

Xie S, Wang Z, Okano M, Nogami M, Li Y, He WW, Okumura K, Li E (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. Gene 236:87-95

Xiong Z, Laird PW (1997). COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res 25:2532-2534

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402:187-191

Yen, R-W. C., Vertino, P. M., Nelkin, B. D., Yu, J. J., El-Deiry, W., Cumaraswamy, A., Lennon, G. G., Trask, B. J., Celano, P., and Baylin, S. B (1992). Isolation and

characterization of the cDNA encoding human DNA methyltransferase. Nucleic Acids Research 20, 2287-2291.

Yoder JA, Bestor TH (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast Hum Mol Genet 7:279-284

Yoder JA, Walsh CP, Bestor TH (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335-340

Yokochi T, Robertson KD (2002). Preferential methylation of unmethylated DNA by mammalian de novo DNA methyltransferase Dnmt3a. J Biol Chem 277:11735-11745

Young, LE., Sinclair, K. D., and Wilmut, I (1998a). Large offspring syndrome in cattle and sheep. Reviews of Reproduction 3, 155-163.

Young LE, Butterwith SC, Wilmut I (1998b). A novel method for quantifying mRNA levels in single embryos. Theriogenology 49:192(Abstract)

Young LE, Fairburn HR (2000). Improving the safety of embryo technologies: possible role of genomic imprinting. Theriogenology 53:627-648

Young LE, Fernandes K, McEvoy TG, Butterwith SC, Gutierrez CG, Carolan C, Broadbent PJ, Robinson JJ, Wilmut I, Sinclair KD (2001). Epigenetic change in IGF2R is associated with fetal overgrowth after sheep embryo culture. Nat Genet 27:153-154

Zakhartchenko V, Alberio R, Stojkovic M, Prelle K, Schernthaner W, Stojkovic P, Wenigerkind H, Wanke R, Duchler M, Steinborn R, Mueller M, Brem G, Wolf E (1999). Adult cloning in cattle: potential of nuclei from a permanent cell line and from primary cultures. Mol Reprod Dev 54:264-272

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev 13:1924-1935

Zhu B, Zheng Y, Hess D, Angliker H, Schwarz S, Siegmann M, Thiry S, Jost JP (2000). 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. Proc Natl Acad Sci U S A 97:5135-5139

Zou X, Chen Y, Wang Y, Luo J, Zhang Q, Zhang X, Yang Y, Ju H, Shen Y, Lao W, Xu S, Du M (2001). Production of cloned goats from enucleated oocytes injected with cumulus cell nuclei or fused with cumulus cells. Cloning 3:31-37

Zuccotti M, Monk M (1995). Methylation of the mouse Xist gene in sperm and eggs correlates with imprinted Xist expression and paternal X-inactivation. Nat Genet 9:316-320

Appendix I

Sheep *DNMT1* cDNA sequence. Length: 5233 bp

```
   1  GGCGAAAAGC CAGGGCTGCC TGTGCTGCTG CCGCCTCTGC TGCACCTTCC
  51  AAGATGCCTG CCCGTACCGC CCCGGCGCGG GTGCCTGCGC TGGCCTCCCG
 101  GGCCTTCTCA CTGCCTGACG ATGTCCGCAG GCGGCTCAAA GATTTGGAAA
 151  GAGATAGTTT GACAGAAAAG GAATGTGTGA AGGAGAAACT GAATCTCTTG
 201  CACGAATTTC TGCAGACAGA AATAAAGAAT CAGTTATGTG ATTTGGAAAC
 251  CAAACTACAT AAAGAGGAAT TATCTGAGGA GGGCTACCTG GCTAAAGTCA
 301  AATCCCTTTT AAATAAAGAT TTGTCCTTGG AGAACGGAGC TCATGCTTTC
 351  AGTCGGGAAG CGAATGGATG TCTAGAGAAC GGGAGCCAGA CAAGTGGTGA
 401  GGATTGCAGA GTGGTAATGG CAGAGAAAGG CAAGCCCCCC AAACCTGTCT
 451  CCAGACTTTA CACGCCCAGG AGAAGCAAGT CTGATGGAGA AACAAAGTCT
 501  GAAGTCTCTT CTAGCCCCAG GATTACAAGG CAGACTACCA GGCAGACCAC
 551  CATCACATCT TATTTTGCAC AGGGCCCTGC CAAACGAAAA CCCGAGGAAG
 601  AACCTGAAAA AGTGAAGTCA GACGATTCTG TTGATGAAGA AAAAGACCAG
 651  GAGGAAAAGA GACGTCGAGT TACATCCAGA GAACGAGTTG CTGGGCTGCT
 701  CCCTGCAGAA GAACCAGGAA GAGTAAGACC AGGAACACAC ATGGAAGAAG
 751  AAGGAAGAGA TGATAAAGAA GAAAAGAGAC TCAGAAGTCA AACCAAAGAA
 801  CCGACACCTA AACAGAAAGC TAAGGAGGAG CCAGACAGAG ATGCGAGGCC
 851  TGGAGGAGCT CAGGCTGAAA TGAATGAAGG AGAAGACAAA GATGAAAAGA
 901  GGCACAGAAG TCAACCCAAA GATCTAGCTA GCAAGCGGAG ACCAGAAGAG
 951  AAAGAGCCTG AAAGAGCAAA ACCACAAGTC TCTGATGAGA AAGATGAAGA
1001  TGAAAAGGAG GAGAAGAGAC GCAGAACTAC ATACAGAGAA CTAACTGAGA
1051  AGAAAATGAC TCGAACCAAA ATAGCAGTGG TGTCCAAGAC CGATCCTCCG
1101  AAGTGCACCG AGTGCTTGCA GTACCTGGAC GACCCCGAGC TGAGATACGA
1151  GCAACACCCC CCTGATGCGG TGGAAGAGAT ACAGATACTG ACCAACGAGA
1201  GGTTGTCCAT CTTCGACGCT AACGAATCTG GCTTTGAGAG TTATGAGGAT
1251  TTGCCCCAAC ACAAACTAAC CTGCTTCAGC GTGTACTGTA AACGCGGCCA
1301  CCTTTGCCCG ATCGACACCG GCCTCATTGA GAAAGATGTC GAGCTCCTCT
```

1351 TTTCTGGTTC AGCAAAGCCG ATATACGAGG ATGACCCATC TCCCGAAGGT

1401 GGTATTAATG GCAAAAATTT TGGGCCCATA AACGAATGGT GGATTGCTGG

1451 TTTTGATGGA GGTGAAAACG CTCTTCTTGG CTTCAGCACC TCATTTGCCG

1501 AGTATATCTT GATGGATCCC AGCCCAGAGT ACGCACCACT ATTCAGCGTG

1551 ATGCAGGAGA AGATCTACAT AAGTAAGATA GTGGTTGAGT TCCTGCAGAG

1601 CAACCCTGAC TCCACCTACG AAGACCTGAT CAATAAGATT GAGACCACTG

1651 TTCCTCCTTG TATGCTCAAC TTGAATCGAT TCACAGAGGA TTCTCTCCTG

1701 CGGCATGCCC AGTTCGTGGT GGAGCAAGTA GAGAGTTATG ATCGGGCTGG

1751 GGACAGTGAC GAGCAGCCCA TCTTCCTGAG CCCCTGCATG AGAGACCTCA

1801 TCAAGCTGGC CGGGGTCACC CTGGGAAAAA GGCGAGCCGA AAAGCGGCAG

1851 ACCATCCGGC AACCCGCCAA GGAGAAGGAC AAGGGCCCCA CCAAGGCCAC

1901 CACCACCAAG CTGGTCTACC AGATCTTCGA CACTTTCTTT GCGGACAAAT

1951 TGAGGAAGGT TGCCAGGGAG GCCAGGGGGA TTGCTTTCAG GGGCGGGGGT

2001 TGTGGCGTCT GTGAGATTTG TCAGCAGCCC GAGTGTGGAA AGTGTAAGGC

2051 CTGTAAGGAT ATGGTTAAAT TTGGTGGTAG TGGACGGAGC AAGCAGGCTT

2101 GCCAAAAGAG GAGGTGTCCC AACATGGCCA TGAAGGAGGC AGATGACGAT

2151 GAGGAAGTGG ATGACAATAT TCCAGAGATG CCATCACCCA AAAAGATGCA

2201 CCAGGGGAAG AAAAAGAAGC AGAACAAGAA TCGGATCTCT TGGGTTGGCG

2251 ATGCCGTCAA GACTGACGGG AAGAAGAGTT ACTACAAGAA GGTATGCATT

2301 GACTCGGAAA CCCTGGAAGT GGGGGACTGT GTTTCTGTAA TTCCAGACGA

2351 CTCTTCAAAA CCACTGTATC TAGCAAGGGT CACGGCGCTG TGGGAGGACA

2401 GCAGCAATGG GCAGATGTTC CACGCCCACT GGTTCTGTGC TGGGACGGAC

2451 ACGGTCCTCG GGGCCACTTC GGACCCCCTG GAGCTATTCC TGGTTGACGA

2501 ATGTGAGGAT ATGCAGCTCT CGTACATCCA CAGCAAGGTG CAGGTCATCT

2551 ATAAGGCCCC CTCGGAGAAC TGGGCCTTGG AGGGAGGCGT GGACCCCGAG

2601 GCCCTGATGT CGCAGGACGA CGGGAAGACC TACTTCTATC AGCTGTGGTA

2651 CGACCAAGAC TACGCGAGAT TTGAGTCCCC TCCGAAAACT CAGCCAACAG

2701 AGGACAACAA GTACAAGTTC TGTGTGAGCT GTGCACGTCT GGCCGAAATG

2751 AGGCAGAAGG AAATCCCCAG GGTCATGGAG CAGCTCCAGG ACCTGGAAGG

2801 CCGTGTCCTC TATAGCCTTG CCACCAAGAA CGGCGTCCAG TACCGGGTGG

2851 GCGATGGCGT GTACCTCCCT CCCGAGGCCT TCACCTTCAA CATCAAGCTG

2901 TCCAGTCCTG TGAAACGCCC CCGGAAGGAG CCTGTGGACG AAGCTCTGTA

2951 TCCAGAGCAC TACCGGAAGT ACTCCGACTA CATCAAGGGC AGCAACCTGG

3001 ATGCCCCCGA GCCCTACCGT GTTGGCCGCA TAAAGGAGAT CTTCTGCAGC

3051 AAGAAGAGCA ACGGCCGGCC CAATGAGACA GACATCAAGA TCAGGGTCAA

3101 CAAGTTCTAC AGGCCGGAGA ACACACACAA GTCTACCCCA GCCAGTTACC

3151 ATGCAGACAT CAACCTGCTC TACTGGAGCG ATGAGGAGGC CGTGGTGGAC

3201 TTCAAGGCTG TGCAGGGCCG CTGCACCGTG GAGTACGGAG AGGACCTGCC

3251 TGAGTGCCTC CAGGACTTCT CTGCTGGTGG CCCCGATCGC TTCTACTTTC

3301 TCGAGGCCTA TAACGCCAAG AGCAAAAGCT TTGAAGATCC TCCGAACCAC

3351 GCCCGGAGCC ATGGAAATAA AGGGAAGGGG AAGGGGAAAG GAAAAAACAG

3401 GACGAAATCT CAGACGTGTG AGCCGAGTGA ACTGGAGACA GAAATCAAAC

3451 TGCCAAAGCT GCGGACCCTG GACGTGTTTT CCGGCTGTGG GGGACTGTCA

3501 GAAGGCTTCC ACCAAGCAGG CATCTCGGAG ACACTTTGGG CCATCGAGAT

3551 GTGGGACCCT GCGGCCCAGG CGTTCCGGCT CAACAACCCT GGGTCCACGG

3601 TGTTCACAGA GTACTGCAAC GTCCTGCTGA AGCTGGTCAT GGGTGGGGAG

3651 GTGACCAACT CCCGCGGCCA GAAGCTGCCT CAGAAGGGAG ACGTGGAGAT

3701 GCTGTGCGGT GGGCCGCCCT GCCAGGGCTT CAGTGGCATG AACCGCTTCA

3751 ACTCTCGAAC CTACTCCAAG TTCAAGAACT CCCTGGTGGT CTCCTTCCTC

3801 AGCTACTGTG ACTACTACCG GCCCCGCTAC TTCCTCTTGG AGAACGTTCG

3851 AAACTTCGTC TCCTTCAAGC GCTCCATGGT CCTGAAGCTG ACGCTGCGCT

3901 GCCTGGTCCG CATGGGCTAC CAGTGCACCT TCGGCGTGCT GCAGGCTGGT

3951 CAGTACGGCG TGGCCCAGAC TCGGAGGCGA GCCATCATCC TGGCTGCAGC

4001 CCCTGGGGAG CCACTCCCTC TGTTCCCCGA GCCGTTGCAC GTGTTCGCGC

4051 CCCGGGCCTG CCAGCTGAGC GTCGTGGTGG ACGACAAGAA GTTTGTCAGC

4101 AACATCACCA GGTTGAGCTC AGGTCCCTTC CGAACCATCA CCGTGCGGGA

4151 CACCATGTCC GACCTCCCCG AGATCCGGAA TGGGGCCTCG GCACTGGAGA

4201 TCTCATACAA CGGGGAGCCC CAGTCTTGGT TCCAGAGGCA GCTCCGGGGC

4251 TCGCAGTACC AGCCCATCCT CAGGGATCAT ATTTGCAAGG ACATGAGCGC

4301 CTTGGTGGCT GCCCGCATGC GGCACATCCC CCTGGCCCCG GGCTCGGACT

4351 GGCGTGACCT GCCCAACATT GAGGTGCGGC TCTCCGATGG CACCCTGGCC

4401 CGGAAGCTGC GGTACAACTA CCACGACAAG AAGAACGGCT GCAGCAGCGC

4451 CGGCGCCCTC CGCGGGGTCT GCTCCTGTGT GGAAGGCAAG CCCTGTGAGC

4501 CTGCGGCCCG ACAGTTCAAC ACCCTTATTC CCTGGTGCCT GCCCCACACT

4551 GGGAACAGGC ACAACCACTG GGCTGGCCTC TACGGGCGTC TTGAGTGGGA

4601 CGGCTTCTTC AGCACAACTG TCACCAACCC TGAGCCCATG GGTAAGCAGG

4651 GCCGCGTGCT CCACCCCGAG CAGCACCGAG TGGTGAGCGT GCGGGAGTGT

4701 GCCCGCTCCC AGGGCTTCCC TGACACCTAT CGGCTGTTCG GCAACATCCT

4751 GGACAAGCAC CGGCAGGTGG GTAATGCTGT GCCGCCACCA CTGGCCAAAG

4801 CCATCGGCTT GGAGATCAAG CGCTGCATGT TGGCCAAAGC GCGTGAGAGC

4851 ACCTCAGCTA AAATCAAGGA GGAGGCTGCC AAGGACTAGC TCTCTCCTCC

4901 TATCACCCAT GTTTCTGCCA CCAGAGATCC CCAGCGTGCA CTGATATTGT

4951 TGTATTTTCA CATGTCAATC AGTCAGTTCA GGTGTGTCGT ATGCGGTGTT

5001 TGTGGCCTTG GCTGACATGA AGCTCTTCAG TGAGATTTCC TATCGGCTAA

5051 TTTGACTTAG TGATCAAACT GTGCAGTACT TTGTCCATTC TGGATTTTAA

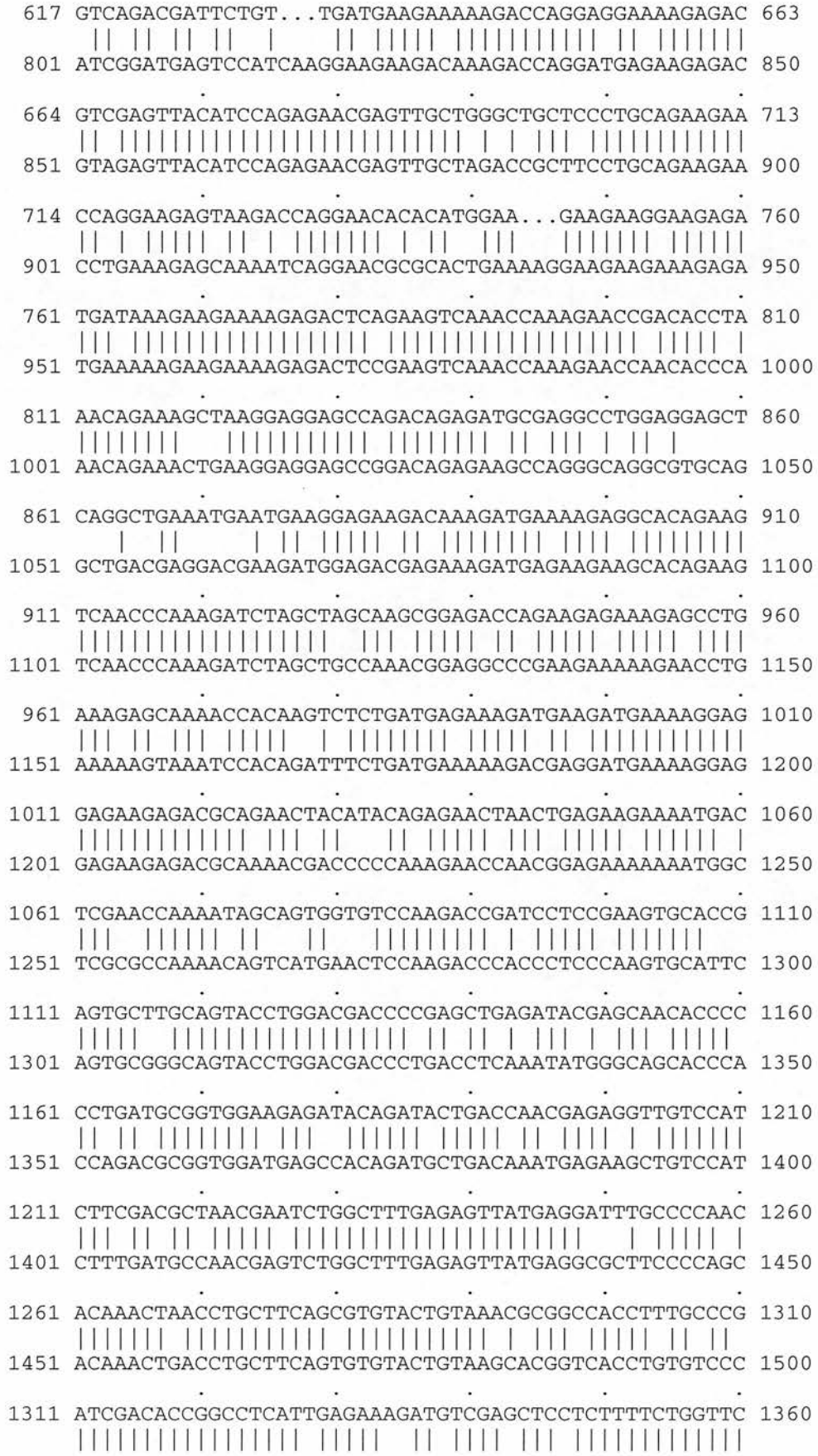
5101 AAGTTTTTTA TTACGCATTA TATCAAATTT ACCACTGTAT GAAGTGGAAA

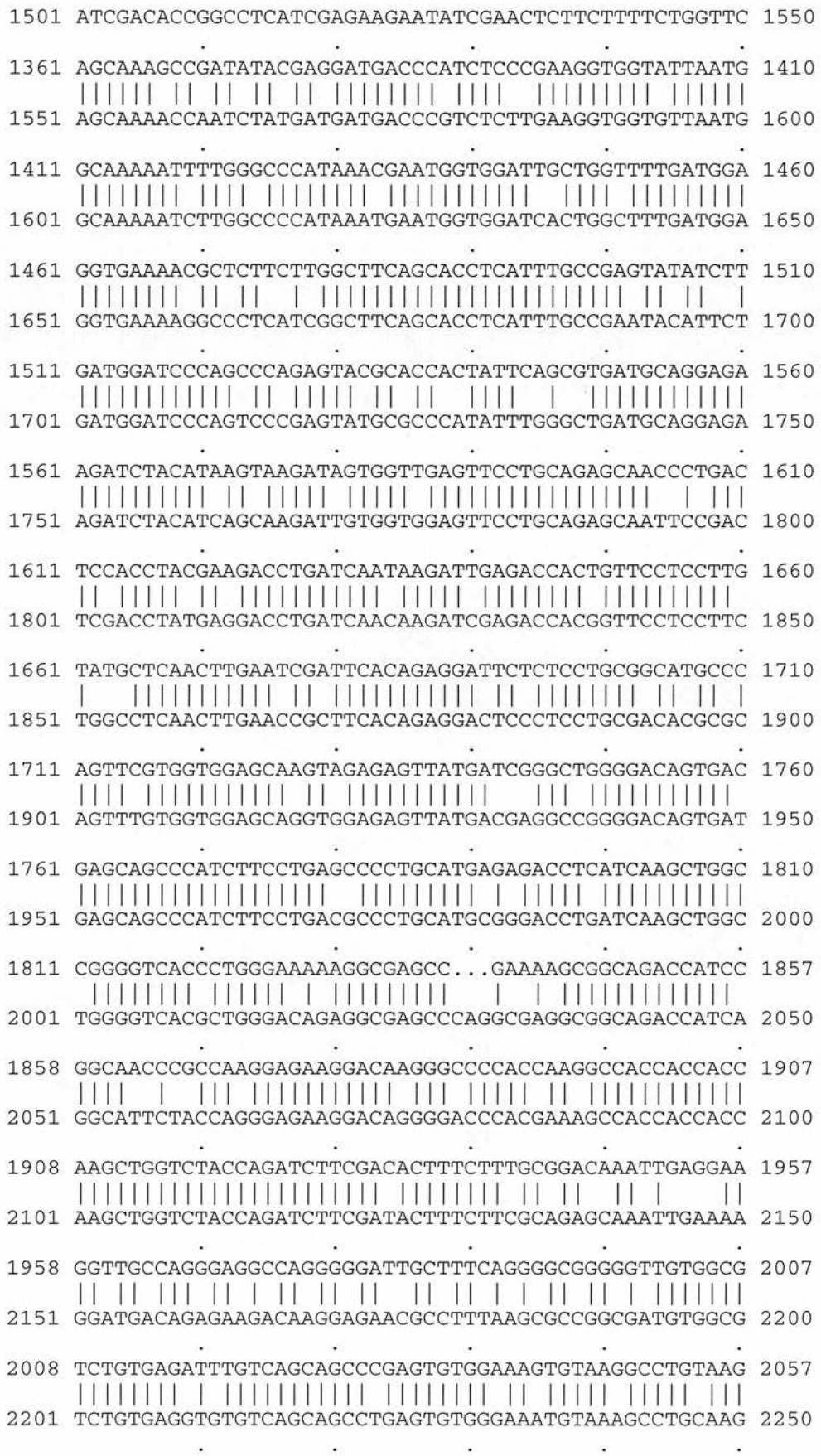
5151 TTAAGACTTT ATGTAGTTTT TATATGTTGT AATATTTCTT CAAATAAATC

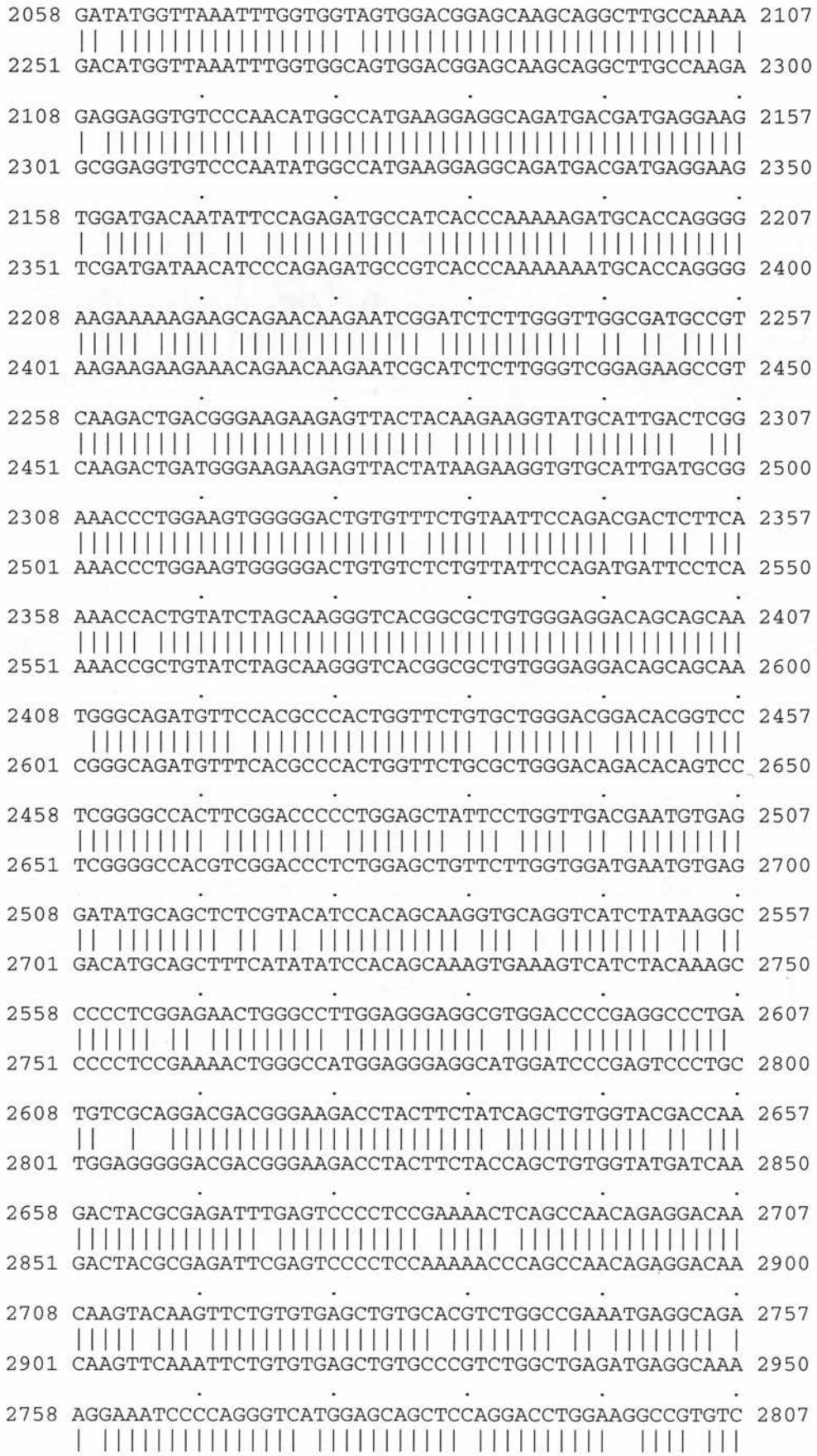
5201 TCTCCTATAA ATCAAAAAAA AAAAAAAAAA AAA

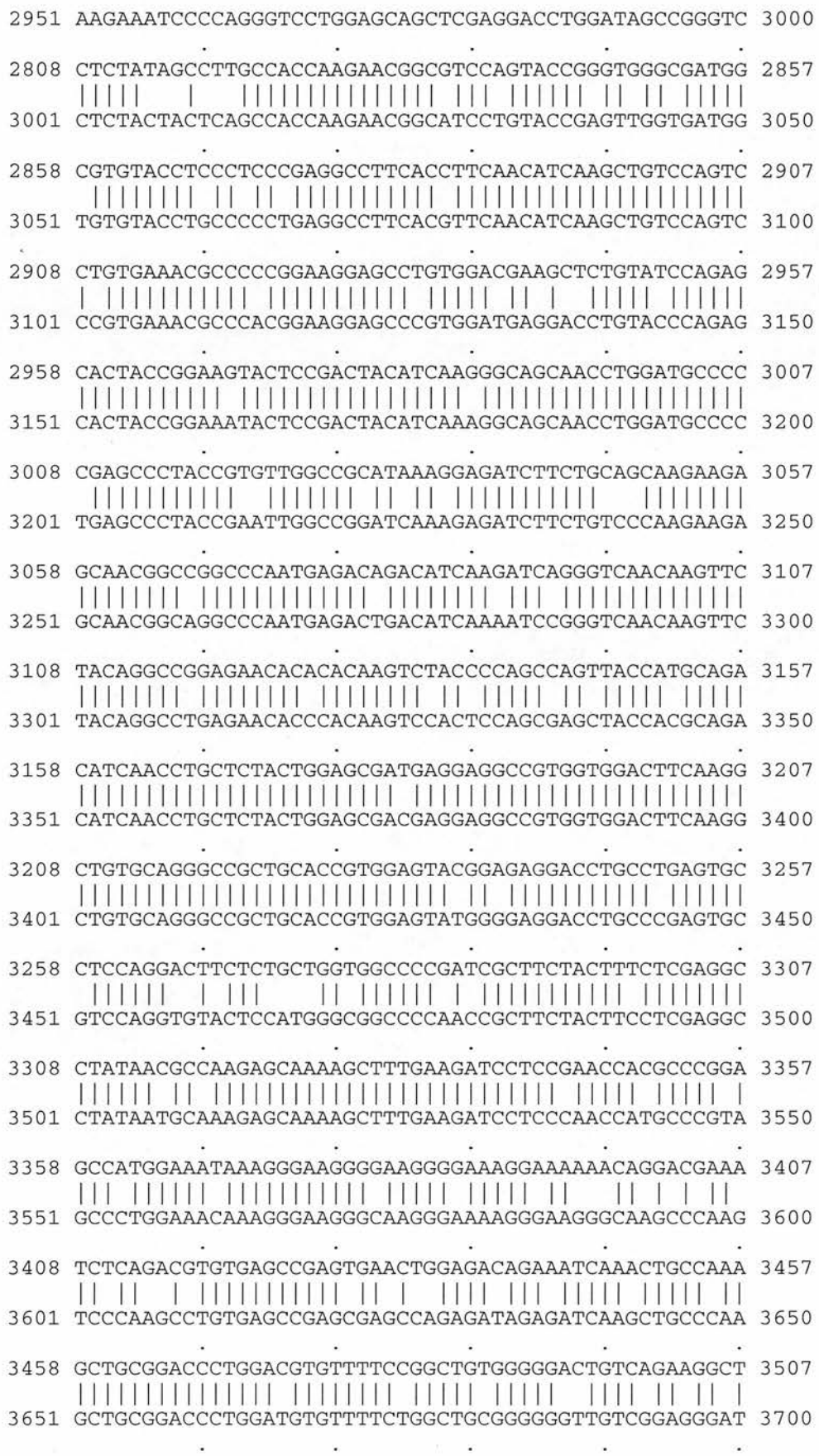
Pairwise alignment of sheep and human (Accession No. NM001379) *DNMT1* cDNA sequences (generated using Gap; GCG10)

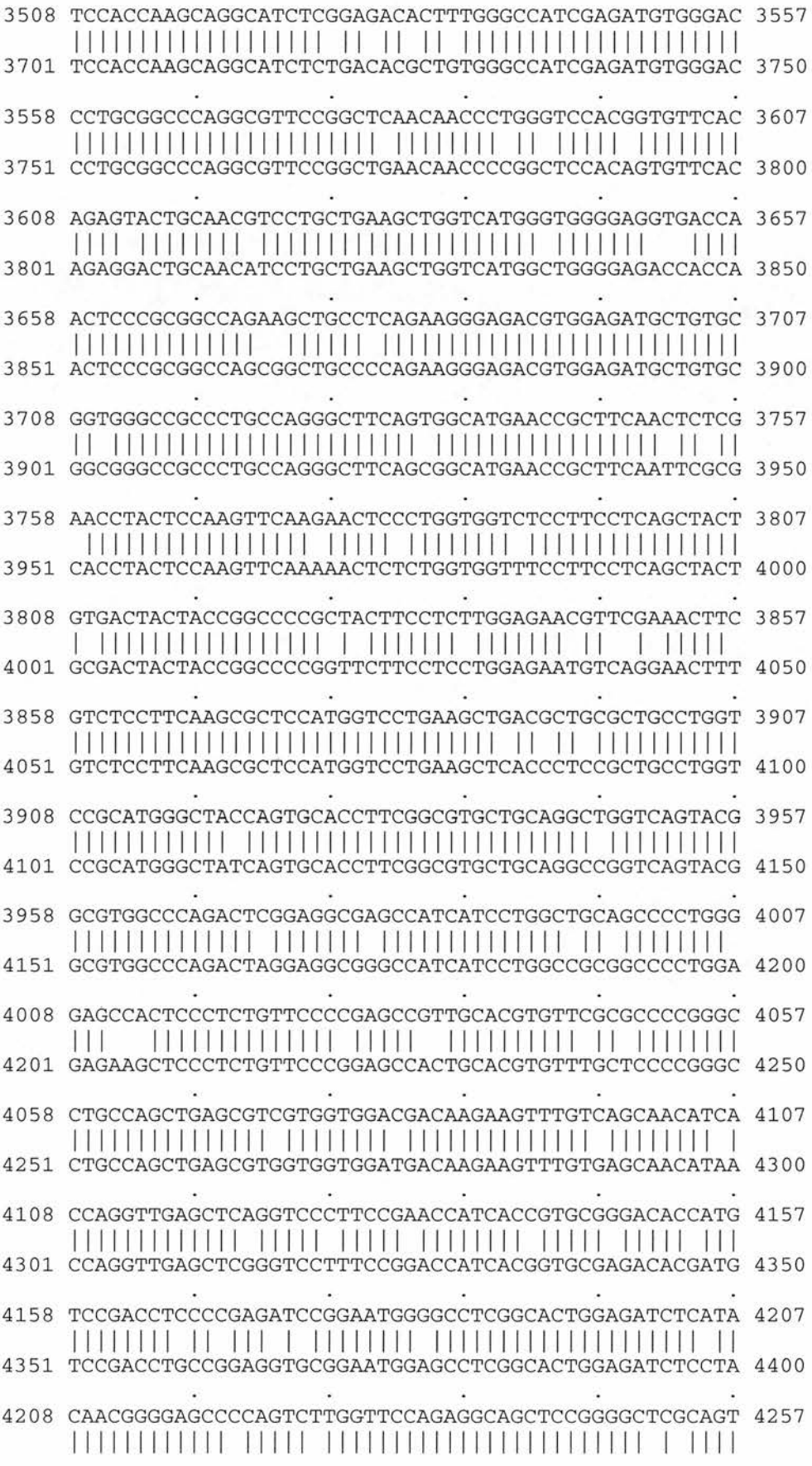
```
                                                                  .
                    .         .         .         .         .
Sheep     1 ..............................GGCGAAAAGCCAGGGCTG  18
                                                  |||||||||| ||||
Human   151 CTCCGCTCATCGCCCCTCCCCATCGGTTTCCGCGCGAAAAGCCGGGGCGC 200
                    .         .         .         .         .
           19 C..CTGTGCTGCTGCCGCCTCTGCTGCACCTTCCAAGATGCCTGCCCGTA  66
              |  |  ||| || ||||| ||||||| | | ||| ||||||| || ||||
          201 CTGCGCTGCCGCCGCCGCGTCTGCTGAAGCCTCCGAGATGCCGGCGCGTA 250
                    .         .         .         .         .
           67 CCGCCCCGGCGCGGGTGCCTGCGCTGGCCTCCCGGGCCTTCTCACTGCCT 116
              ||||||| || ||||||||  | ||||||  || |||| |||| |||||
          251 CCGCCCCAGCCCGGGTGCCCACACTGGCCGTCCCGGCCATCTCGCTGCCC 300
                    .         .         .         .         .
          117 GACGATGTCCGCAGGCGGCTCAAAGATTTGGAAAGAGATAGTTTGACAGA 166
              |||||||||||||||||||||||||||||||||||| || || |||||
          301 GACGATGTCCGCAGGCGGCTCAAAGATTTGGAAAGAGACAGCTTAACAGA 350
                    .         .         .         .         .
          167 AAAGGAATGTGTGAAGGAGAAACTGAATCTCTTGCACGAATTTCTGCAGA 216
              |||||||||||||||||||| |||||||||||||||||||||||||||| |
          351 AAAGGAATGTGTGAAGGAGAAATTGAATCTCTTGCACGAATTTCTGCAAA 400
                    .         .         .         .         .
          217 CAGAAATAAAGAATCAGTTATGTGATTTGGAAACCAAACTACATAAAGAG 266
              |||||||||||||||||||||||| |||||||||||| ||| ||||||
          401 CAGAAATAAAGAATCAGTTATGTGACTTGGAAACCAAATTACGTAAAGAA 450
                    .         .         .         .         .
          267 GAATTATCTGAGGAGGGCTACCTGGCTAAAGTCAAATCCCTTTTAAATAA 316
              |||||||| ||||||||||||||||||||||||||||||||||||||||||
          451 GAATTATCCGAGGAGGGCTACCTGGCTAAAGTCAAATCCCTTTTAAATAA 500
                    .         .         .         .         .
          317 AGATTTGTCCTTGGAGAACGGAGCTCATGCTTTCAGTCGGGAAGCGAATG 366
              |||||||||||||||||||| |||||||||| ||  ||||||| |||||
          501 AGATTTGTCCTTGGAGAACGGTGCTCATGCTTACAACCGGGAAGTGAATG 550
                    .         .         .         .         .
          367 GATGTCTAGAGAACGGGAGCCAGACAAGTGGTGAGGATTGCAGAGTGGTA 416
              || ||||||| ||||||| |||  ||||  |||| |   | ||||||| |
          551 GACGTCTAGAAAACGGGAACCAAGCAAGAAGTGAAGCCCGTAGAGTGGGA 600
                    .         .         .         .         .
          417 ATGGCAGAGAAAGGCAAGCCCCCCAAACCTGTCTCCAGACTTTACACGCC 466
              ||||||||       ||  |||||||||||  | |||| || |  ||||||
          601 ATGGCAGATGCCAACAGCCCCCCCAAACCCCTTTCCAAACCTCGCACGCC 650
                    .         .         .         .         .
          467 CAGGAGAAGCAAGTCTGATGGAGAAACAAAGTCTGAAGTCTCTTCTAGCC 516
              |||||| |||||||| ||||||||  | ||| |||||   ||  ||||||
          651 CAGGAGGAGCAAGTCCGATGGAGAGGCTAAGCCTGAACCTTCACCTAGCC 700
                    .         .         .         .         .
          517 CCAGGATTACAAGGCAGACTACCAGGCAGACCACCATCACATCTTATTTT 566
              |||||||||||||| | |  |||||||| |||||||||||||| |||||
          701 CCAGGATTACAAGGAAAAGCACCAGGCAAACCACCATCACATCTCATTTT 750
                    .         .         .         .         .
          567 GCACAGGGCCCTGCCAAACGAAAACCCGAGGAAGAACCTGAAAAAGTGAA 616
              ||| |||||||||||||| |||||  |||||||  |||||| ||  ||
          751 GCAAAGGGCCCTGCCAAACGGAAACCTCAGGAAGAGTCTGAAAGAGCCAA 800
                    .         .         .         .         .
```

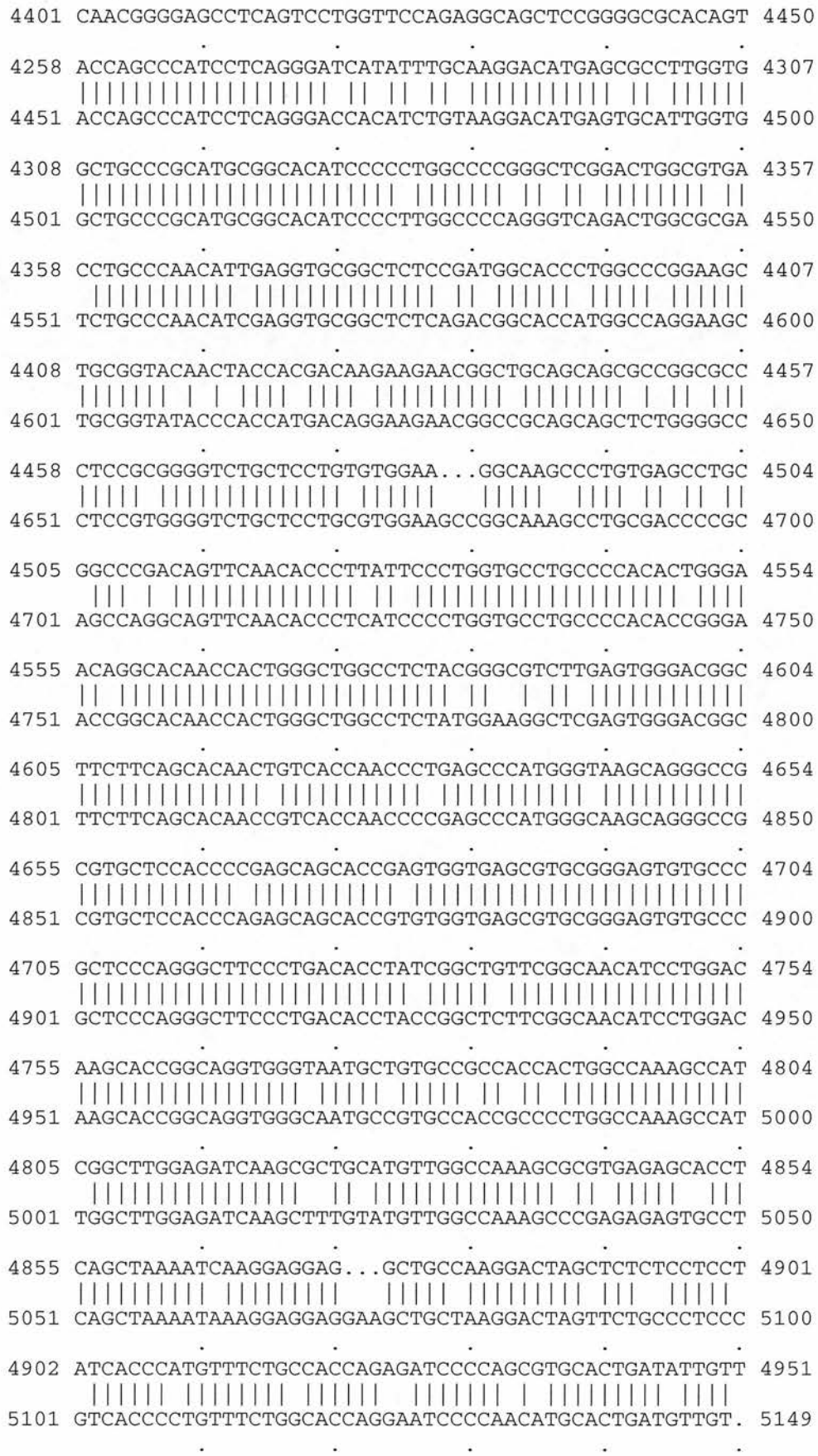





2058 GATATGGTTAAATTTGGTGGTAGTGGACGGAGCAAGCAGGCTTGCCAAAA 2107
2251 GACATGGTTAAATTTGGTGGCAGTGGACGGAGCAAGCAGGCTTGCCAAGA 2300
2108 GAGGAGGTGTCCCAACATGGCCATGAAGGAGGCAGATGACGATGAGGAAG 2157
2301 GCGGAGGTGTCCCAATATGGCCATGAAGGAGGCAGATGACGATGAGGAAG 2350
2158 TGGATGACAATATTCCAGAGATGCCATCACCCAAAAAGATGCACCAGGGG 2207
2351 TCGATGATAACATCCCAGAGATGCCGTCACCCAAAAAAATGCACCAGGGG 2400
2208 AAGAAAAAGAAGCAGAACAAGAATCGGATCTCTTGGGTTGGCGATGCCGT 2257
2401 AAGAAGAAGAAACAGAACAAGAATCGCATCTCTTGGGTCGGAGAAGCCGT 2450
2258 CAAGACTGACGGGAAGAAGAGTTACTACAAGAAGGTATGCATTGACTCGG 2307
2451 CAAGACTGATGGGAAGAAGAGTTACTATAAGAAGGTGTGCATTGATGCGG 2500
2308 AAACCCTGGAAGTGGGGGACTGTGTTTCTGTAATTCCAGACGACTCTTCA 2357
2501 AAACCCTGGAAGTGGGGGACTGTGTCTCTGTTATTCCAGATGATTCCTCA 2550
2358 AAACCACTGTATCTAGCAAGGGTCACGGCGCTGTGGGAGGACAGCAGCAA 2407
2551 AAACCGCTGTATCTAGCAAGGGTCACGGCGCTGTGGGAGGACAGCAGCAA 2600
2408 TGGGCAGATGTTCCACGCCCACTGGTTCTGTGCTGGGACGGACACGGTCC 2457
2601 CGGGCAGATGTTTCACGCCCACTGGTTCTGCGCTGGGACAGACACAGTCC 2650
2458 TCGGGGCCACTTCGGACCCCCTGGAGCTATTCCTGGTTGACGAATGTGAG 2507
2651 TCGGGGCCACGTCGGACCCTCTGGAGCTGTTCTTGGTGGATGAATGTGAG 2700
2508 GATATGCAGCTCTCGTACATCCACAGCAAGGTGCAGGTCATCTATAAGGC 2557
2701 GACATGCAGCTTTCATATATCCACAGCAAAGTGAAAGTCATCTACAAAGC 2750
2558 CCCCTCGGAGAACTGGGCCTTGGAGGGAGGCGTGGACCCCGAGGCCCTGA 2607
2751 CCCCTCCGAAAACTGGGCCATGGAGGGAGGCATGGATCCCGAGTCCCTGC 2800
2608 TGTCGCAGGACGACGGGAAGACCTACTTCTATCAGCTGTGGTACGACCAA 2657
2801 TGGAGGGGGACGACGGGAAGACCTACTTCTACCAGCTGTGGTATGATCAA 2850
2658 GACTACGCGAGATTTGAGTCCCCTCCGAAAACTCAGCCAACAGAGGACAA 2707
2851 GACTACGCGAGATTCGAGTCCCCTCCAAAAACCCAGCCAACAGAGGACAA 2900
2708 CAAGTACAAGTTCTGTGTGAGCTGTGCACGTCTGGCCGAAATGAGGCAGA 2757
2901 CAAGTTCAAATTCTGTGTGAGCTGTGCCCGTCTGGCTGAGATGAGGCAAA 2950
2758 AGGAAATCCCCAGGGTCATGGAGCAGCTCCAGGACCTGGAAGGCCGTGTC 2807



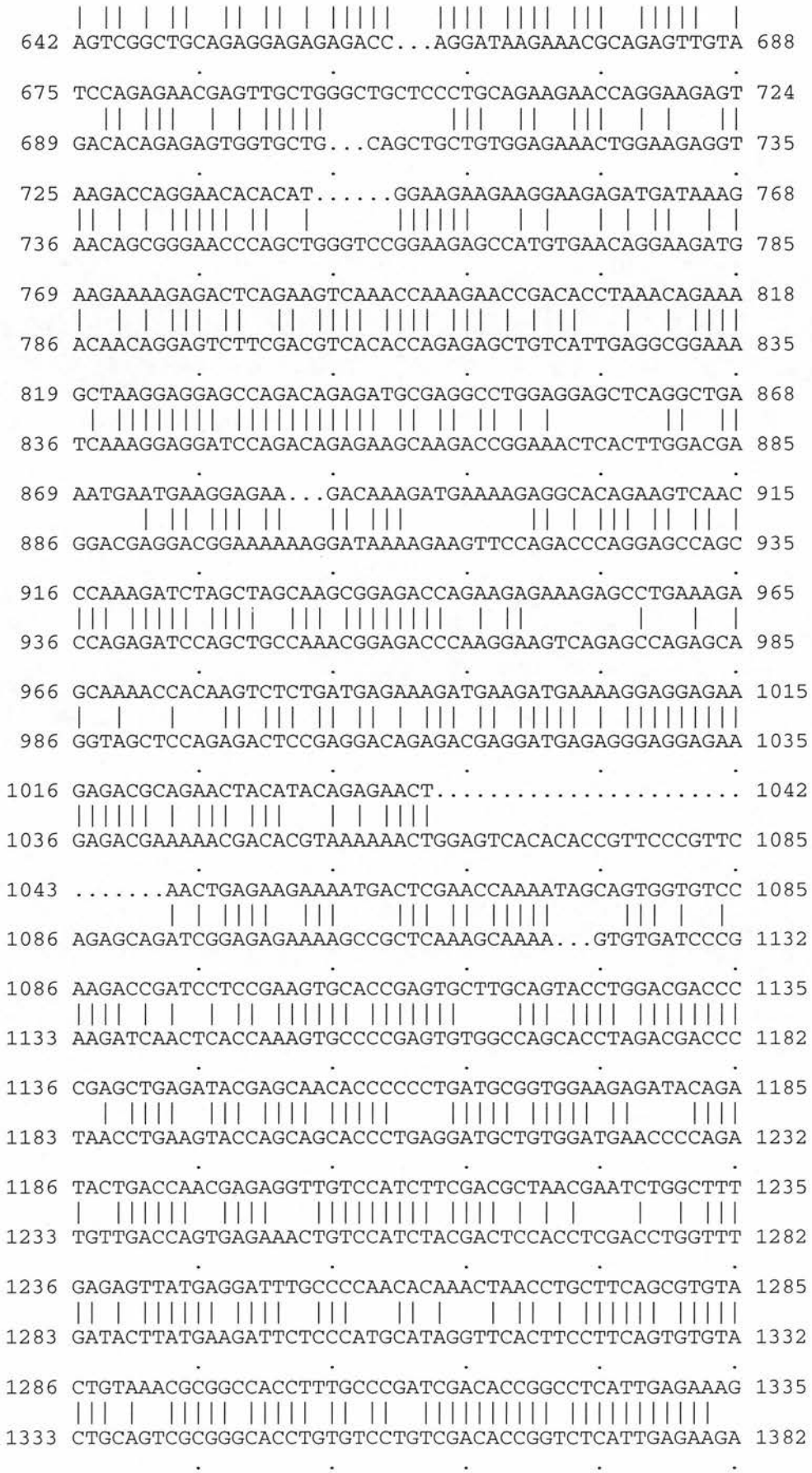


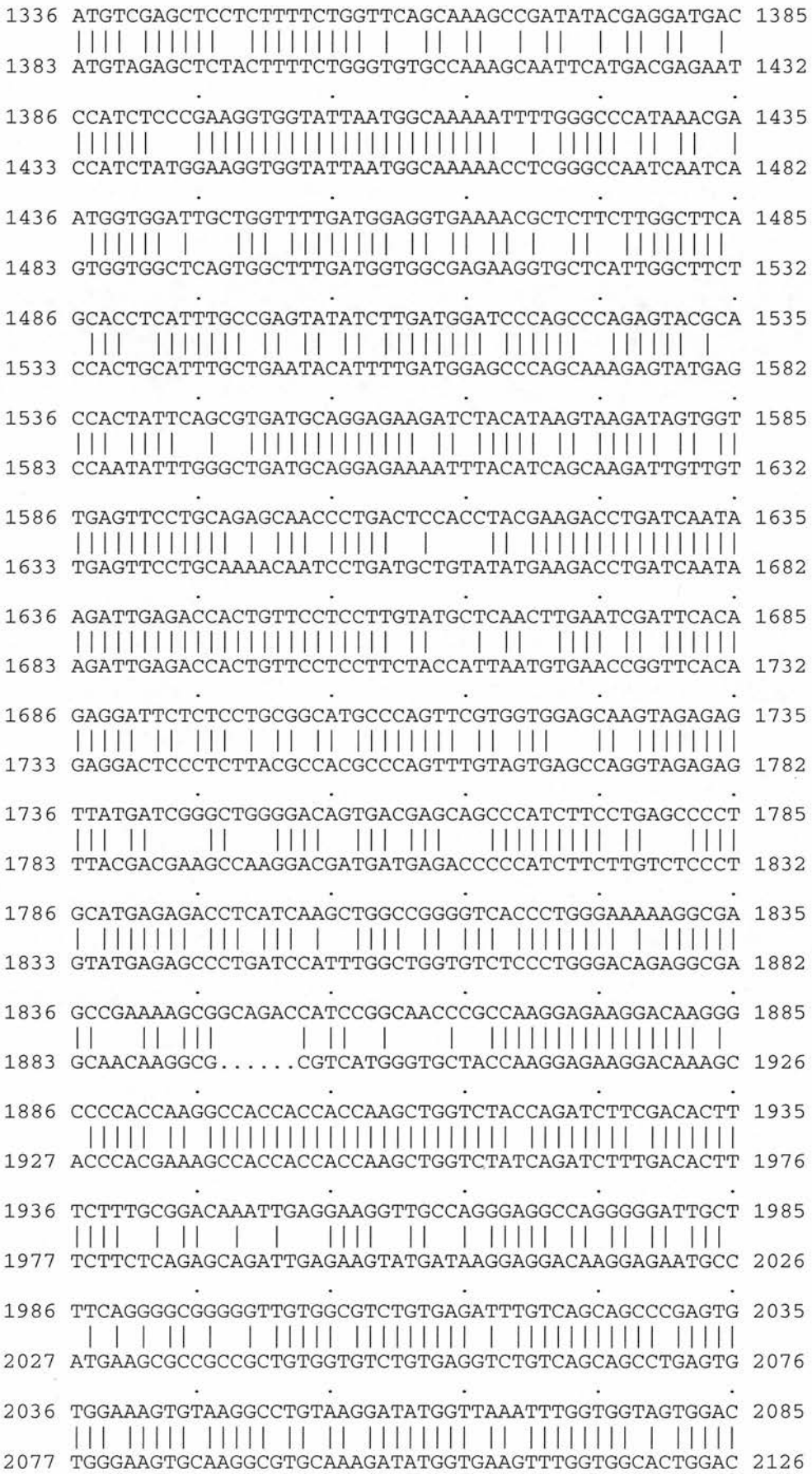


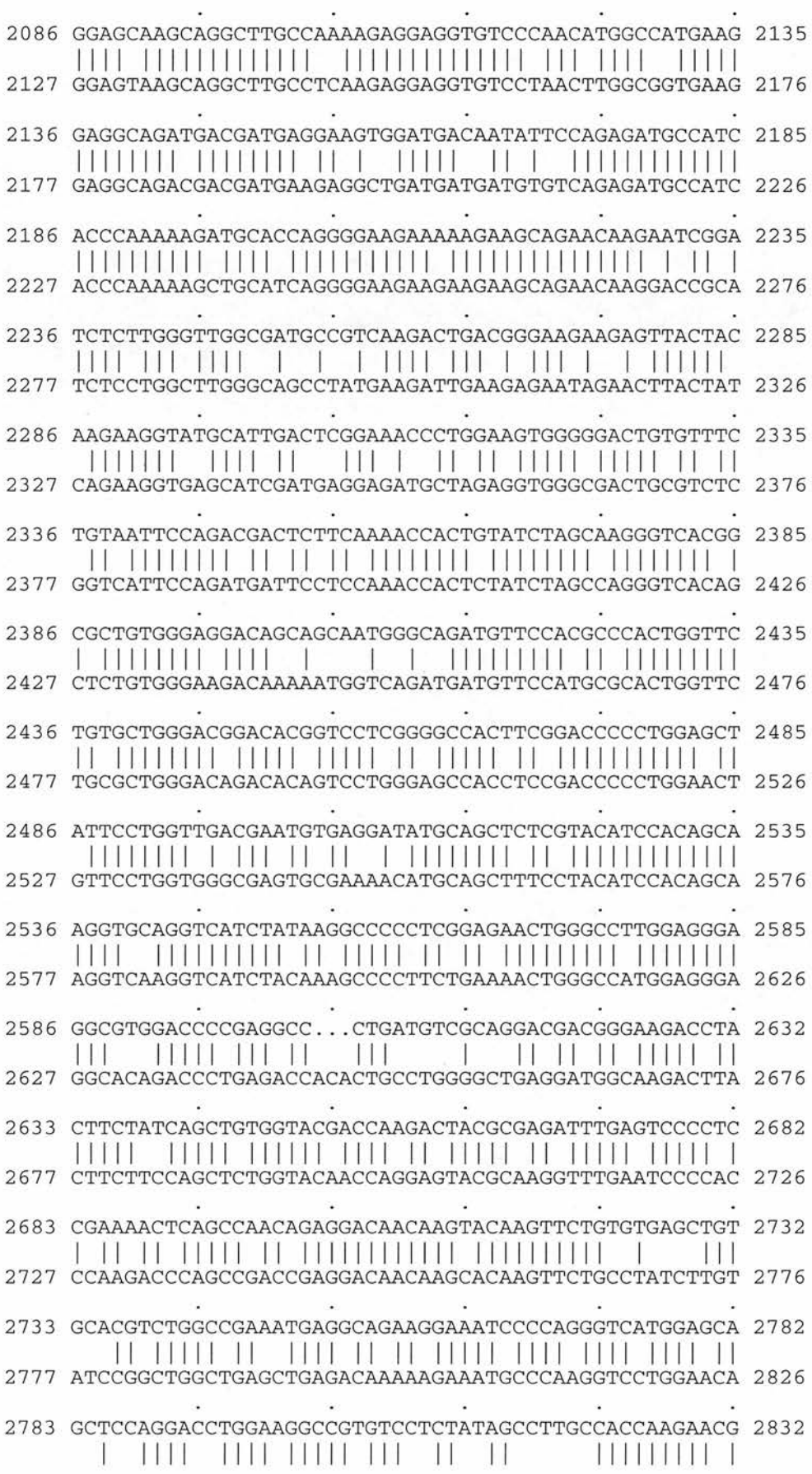
```
4952 GTATTTTCACATGTCAATCAGTCAGTTCAGGTGTGTCGTATGCGGTGTTT 5001
        ||||  |||||||||||  |||  |||||  |||||  |||   |||||||
5150 .GTTTTTAACATGTCAATCTGTCCGTTCACATGTGTGGTACATGGTGTTT 5198
             .          .          .          .          .
5002 GTGGCCTTGGCTGACATGAAGCTCTTCAGTGAGATT.TCCTATCGGCTAA 5050
     ||||||||||||||||||||||||  ||   |||||  ||   |  ||||  ||||
5199 GTGGCCTTGGCTGACATGAAGCTGTTGTGTGAGGTTCGCTTATCAACTAA 5248
             .          .          .          .          .
5051 TTTGACTTAGTGATCAAACTGTGCAGTACTTTGTCCATTCTGGATTTTAA 5100
       |||  |||||||||||  ||||||||||||||||  ||||||||||||||||
5249 ..TGATTTAGTGATCAAATTGTGCAGTACTTTGTGCATTCTGGATTTTAA 5296
             .          .          .          .          .
5101 AAGTTTTTTATTACGCATTATATCAAATTTACCACTGTATGAAGTGGAAA 5150
     ||||||||||||  ||||||||||||||  ||||||||||||  ||||||||
5297 AAGTTTTTTATTATGCATTATATCAAATCTACCACTGTATG.AGTGGAAA 5345
             .          .          .          .          .
5151 TTAAGACTTTATGTAGTTTTTATATGTTGTAATATTTCTTCAAATAAATC 5200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
5346 TTAAGACTTTATGTAGTTTTTATATGTTGTAATATTTCTTCAAATAAATC 5395
             .          .          .
5201 TCTCCTATAAATCAAAAAAAAAAAAAAAAAAAA...... 5233
     ||||||||||  ||||||||||||||||||||
5396 TCTCCTATAAACCAAAAAAAAAAAAAAAAAAAAAAAA 5434
```

Pairwise alignment of Sheep and mouse (Accession No. X14805) *DNMT1* cDNA sequences generated using Gap (GCG10)

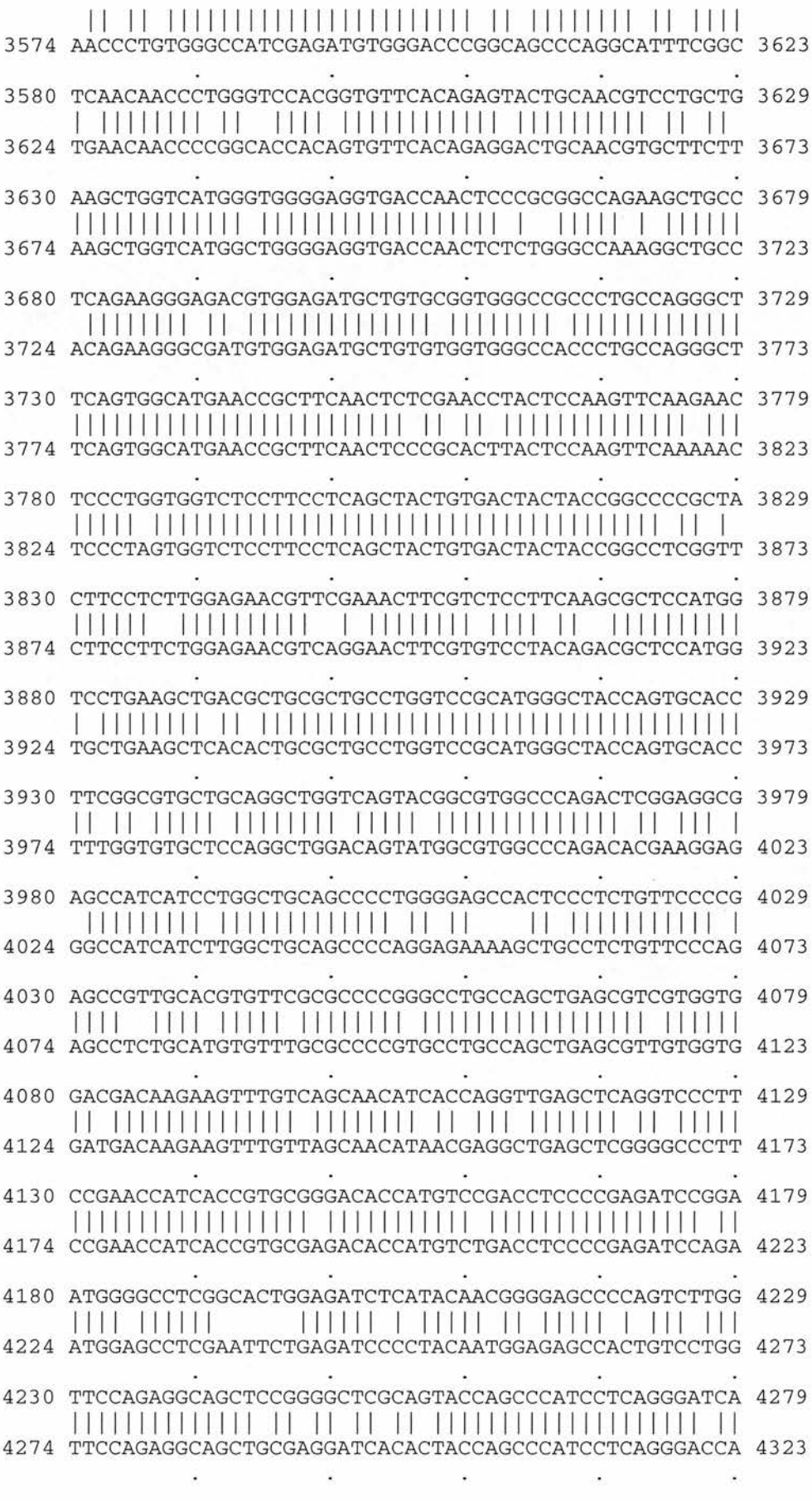
```
                 .         .         .         .         .
Sheep     1 ........................GGCGAAAAGCCAGGGCTGCCTGTG  24
                                     |   ||||||| |||   | |
Mouse     1 GCCCCTCCCAATTGGTTTCCGCGCGCGCGAAAAAGCCGGGGTCTCGTTCA  50

                 .         .         .         .         .
       25 CTGCTGCCGCCTCTGCTGCACCTTCCAAGATGCCTGCCCGTACCGCCCCG  74
            ||||     ||  ||||| | | |||||||||| || || || || ||
       51 GAGCTGTTCTGTCGTCTGCAACCTGCAAGATGCCAGCGCGAACAGCTCCA 100

                 .         .         .         .         .
       75 GCGCGGGTGCCTGCGCTGGCCTCCCGGGCCTTCTCACTGCCTGACGATGT 124
          || || |||||||||||| ||||||| |||   ||| || || ||| ||||
      101 GCCCGAGTGCCTGCGCTTGCCTCCCCGGCAGGCTCGCTCCCGGACCATGT 150

                 .         .         .         .         .
      125 CCGCAGGCGGCTCAAAGATTTGGAAAGAGATAGTTTGACAGAAAAGGAAT 174
          |||||||||||||||||| |||||||||||| | || |||||||||||  |
      151 CCGCAGGCGGCTCAAAGACTTGGAAAGAGATGGCTTAACAGAAAAGGAGT 200

                 .         .         .         .         .
      175 GTGTGAAGGAGAAACTGAATCTCTTGCACGAATTTCTGCAGACAGAAATA 224
          ||||||  |||||||  | ||  |  |||| |||||  |||||  |||||||||
      201 GTGTGAGGGAGAAATTAAACTTACTGCATGAATTCCTGCAAACAGAAATA 250

                 .         .         .         .         .
      225 AAGAATCAGTTATGTGATTTGGAAACCAAACTACATAAAGAGGAATTATC 274
          ||  |  |||||  |||||  |||||||||||| ||||||||||||||||||||
      251 AAAAGCCAGTTGTGTGACTTGGAAACCAAATTACATAAAGAGGAATTATC 300

                 .         .         .         .         .
      275 TGAGGAGGGCTACCTGGCTAAAGTCAAATCCCTTTTAAATAAAGATTTGT 324
          ||||||  |||||||||||||||||||||  |||||  ||||||||  |||||||
      301 TGAGGAAGGCTACCTGGCTAAAGTCAAGTCCCTCTTAAATAAGGATTTGT 350

                 .         .         .         .         .
      325 CCTTGGAGAACGGAGCTCATGCTTTCAGTCGGGAAGCGAATGGATGTCTA 374
          |||||||||||||  | ||  || ||| ||   ||||  || || ||||
      351 CCTTGGAGAACGGAACACACACTCTCACTCAAAAAGCCAACGGTTGTCCC 400

                 .         .         .         .         .
      375 GAGAACGGGAGCCAGACAAGTGGTGAGGATTGCAGAGTGGTAATGGCAGA 424
          |  ||||||||||  | |||           || ||||  | |||||||||
      401 GCCAACGGGAGCCGGCCAA.........CCTGGAGAGCAGAAATGGCAGA 441

                 .         .         .         .         .
      425 GAAAGGCAAGCCCCCCAAACCTGTCTCCAGACTTTACACGCCCAGGAGAA 474
             |   |   |||| | | |     |||  | |      ||||||||||
      442 CTCAAATAGATCCCCAAGATCCAGGCCCAAGCCTCGGGGACCCAGGAGAA 491

                 .         .         .         .         .
      475 GCAAGTCTGATGGAGAAACAAAGTCTGAAGTCTCTTCTAGCCCCAGGATT 524
          |||||||  ||  | || ||    | ||||  ||  ||||  ||  |  |
      492 GCAAGTCGGACAGTGACACCCTTTTTGAAACTTCACCTAGTTCCGTGGCT 541

                 .         .         .         .         .
      525 ACAAGGCAGACTACCAGGCAGACCACCATCACATCTTATTTTGCACAGGG 574
          ||  |||   ||  ||||||||||||||||||||  || | ||  |  ||||
      542 ACGAGGAGAACCACCAGGCAGACCACCATCACGGCTCACTTCACGAAGGG 591

                 .         .         .         .         .
      575 CCCTGCCAAACGAAAACCCGAGGAAGAACCTGAAAAAGTGAAGTCAGACG 624
          |||  | |||||  ||||||  |||||||  | |||  |  |  |||  ||  |  |
      592 CCCCACTAAACGGAAACCCAAGGAAGAGTCGGAAGAGGGGAACTCGGCTG 641

                 .         .         .         .         .
      625 ATTCTGTTGATGAAGAAAAAGACCAGGAGGAAAAGAGACGTCGAGTTACA 674
```

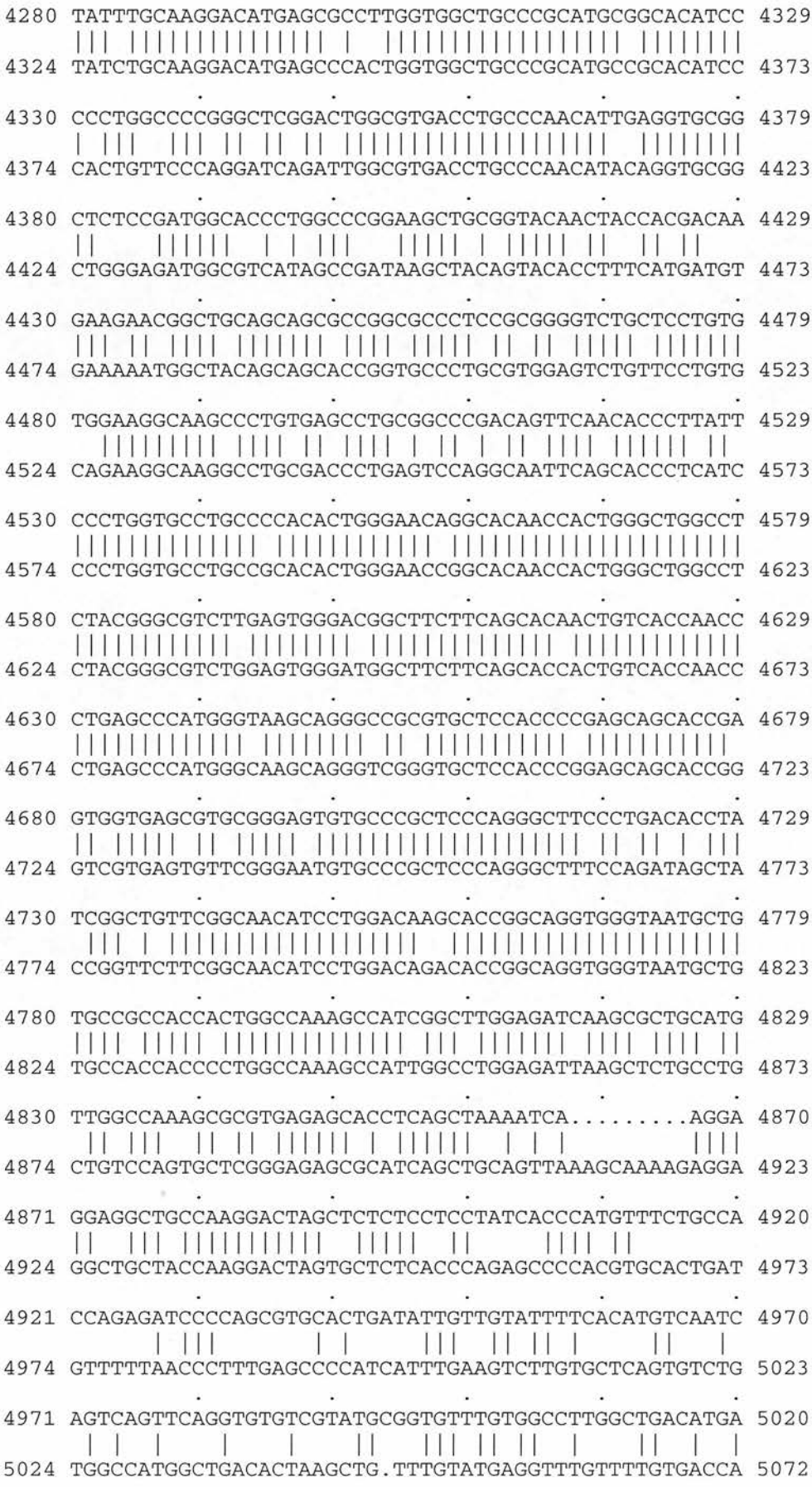






```
2827 AATTGAGGAGGTGGATGGCCGGGTCTACTGCAGTTCCATCACCAAGAATG 2876
2833 GCGTCCAGTACCGGGTGGGCGATGGCGTGTACCTCCCTCCCGAGGCCTTC 2882
2877 GTGTTTGCTACCGACTGGGTGACAGTGTGTACCTTCCTCGGGAGGCCTTT 2926
2883 ACCTTCAACATCAAGCTGTCCAGTCCTGTGAAACGCCCCCGGAAGGAGCC 2932
2927 ACTTTCAACATCAAAGTGGCTAGCCCCGTGAAACGCCCAAAGAAGGATCC 2976
2933 TGTGGACGA.AGCTCTGTATCCAGAG..CACTACCGGAAGTACTCCGACT 2979
2977 TGTGAAAGAGAACCCTGTACCCAGAGACACATACAGAAAGTATTCTGACT 3026
2980 ACATCAAGGGCAGCAACCTGGATGCCCCCGAGCCCTACCGTGTTGGCCGC 3029
3027 ACATCAAGGGGAGAAACCTGGATGCTCCAGAGCCCTATCGCATCGGTCGG 3076
3030 ATAAAGGAGATCTTCTGCAGCAAGAAGAGCAACGGCCGGCCCAATGAGAC 3079
3077 ATAAAAGAGATCCACTGTGGCAAGAAG...AAAGGCAAGGTGAACGAGGC 3123
3080 AGACATCAAGATCAGGGTCAACAAGTTCTACAGGCCGGAGAACACACACA 3129
3124 AGACATCAAGCTGAGGCTCTACAAGTTCTACAGGCCTGAGAATACCCACA 3173
3130 AGTCTACCCCAGCCAGTTACCATGCAGACATCAACCTGCTCTACTGGAGC 3179
3174 GGTCCTACAACGGATCCTATCACACTGACATCAACATGCTTTGCTGGAGC 3223
3180 GATGAGGAGGCCGTGGTGGACTTCAAGGCTGTGCAGGGCCGCTGCACCGT 3229
3224 GACGAGGAAGCTGTGGTGAACTTCAGCGACGTGCAGGGCCGCTGTACCGT 3273
3230 GGAGTACGGAGAGGACCTGCCTGAGTGCCTCCAGGACTTCTCTGCTGGTG 3279
3274 GGAGTACCGGGAAGACCTACTTGAGAGCATCCAGGATTATTCACAAGGGG 3323
3280 GCCCCGATCGCTTCTACTTTCTCGAGGCCTATAACGCCAAGAGCAAAAGC 3329
3324 GCCCTGACCGCTTCTACTTCCTCGAGGCCTACAATTCAAAGACCAAGAAC 3373
3330 TTTGAAGATCCTCCGAACCACGCCCGGAGCCATGGAAATAAAGGGAAGGG 3379
3374 TTTGAAGACCCACCAAACCATGCCCGCAGCCCTGGGAACAAAGGGAAAGG 3423
3380 GAAGGGGAAAGGAAAAAACAGGACGAAATCTCAGACGTGTGAGCCGAGTG 3429
3424 GAAGGGGAAAGGGAAGGGGAAGGGGAAGCATCAGGTGTCAGAGCCCAAAG 3473
3430 AACTGGAGACAGAAATCAAACTGCCAAAGCTGCGGACCCTGGACGTGTTT 3479
3474 AGCCTGAGGCAGCCATCAAACTGCCCAAGCTCCGGACCCTGGATGTGTTT 3523
3480 TCCGGCTGTGGGGGACTGTCAGAAGGCTTCCACCAAGCAGGCATCTCGGA 3529
3524 TCCGGCTGTGGAGGGTTATCGGAAGGATTCCACCAAGCAGGCATCTCGGA 3573
3530 GACACTTTGGGCCATCGAGATGTGGGACCCTGCGGCCCAGGCGTTCCGGC 3579
```





```
                .         .         .         .         .
5021 AGCTCTTCAGTGAGATTTCCTATCGGCTAATTTGACTTAGTGATCAAACT 5070
     ||||  |  |||    | | |  || |    ||    ||           |
5073 AGCTGTGTAGTACTTTGTGCATTCTGAATTTTAAGGTT..........T 5111
                .         .         .         .         .
5071 GTGCAGTACTTTGTCCATTCTGGATTTTAAAAGTTTTTTATTACGCATTA 5120
      |    |  ||||       | | |||      ||| ||||   | |||
5112 TTTTTTTTGTTTGGTTTGGTTTGGTTTGGTTTTTTTCTTATCCTGTATTC 5161
                .         .         .         .         .
5121 TATCAAATTTACCACTGTATGAAGTGGAAATTAAGACTTTATGTAGTTTT 5170
     ||||| || | |||||||   | |||| || | |||||| |||||| |||
5162 TATCAGATCTGCCACTGT.GCAGGTGGCAAGTGAGACTTGATGTAG.TTT 5209
                .         .         .         .         .
5171 TATATGTTGTAATATTTCTTC.AAATAAATCTCTCCTATAAATCAAAAAA 5219
     |||||||||||||||||||| ||||||| | || || | || |||||||
5210 TATATGTTGTAATATTTCTTCAAAATAAAGCGCTTCTGTCAAGCAAAAAA 5259
                .
5220 AAAAAAAAAAAAA 5233
     ||||||||
5260 AAAAAAAA..... 5267
```

Appendix II

Sheep *DNMT3A* cDNA sequence. Length: 1848 bp

```
   1  GACATGTGGG TTGAACCTGA GGCAGCTGCC TATGCGCCGC CCCCACCAGC
  51  CAAAAAGCCC CGAAAGAGCA CAACGGAGAA GCCTAAGGTC AAGGAGATCA
 101  TTGATGAGCG CACAAGAGAG CGGCTGGTGT ACGAGGTACG GCAGAAGTGC
 151  CGGAACATCG AGGACATCTG CATCTCTTGT GGGAGCCTCA ACGTCACCTT
 201  GGAACACCCT CTCTTCATCG GAGGAATGTG CCAAAACTGC AAGAACTGCT
 251  TCCTGGAATG CGCGTACCAG TACGATGACG ACGGCTATCA GTCCTACTGC
 301  ACCATCTGCT GCGGGGGGCG CGAGGTGCTC ATGTGTGGGA ACAACAATTG
 351  CTGCAGGTGC TTTTGCGTGG AGTGTGTGGA TCTCTTGGTG GGGCCAGGGG
 401  CTGCGCAGGC AGCATTCAAG GAAGACCCCT GGAACTGCTA CATGTGCGGG
 451  CACAAGGGTA CCTACGGGCT GCTGCGGCGG CGGGACGACT GGTCGTCTCG
 501  GCTCCAGATG TTCTTCGCCA ACAACCATGA CCAGGAATTC GACCCTCCGA
 551  AGGTTTACCC ACCTGTCCCA GCCGAGAAGA GGAAGCCCAT CCGGGTGCTG
 601  TCTCTATTCG ATGGAATCGC TACAGGGCTT CTGGTGCTGA AGGACTTGGG
 651  CATTCAGGTG GACCGCTACA TCGCCTCCGA GGTGTGTGAG GACTCCATCA
 701  CAGTGGGCAT GGTGCGGCAC CAGGGGAAGA TCATGTACGT CGGGGACGTC
 751  CGCAGCGTTA CACAGAAGCA TATCCAGGAG TGGGGCCCGT TCGATCTGGT
 801  GATTGGGGGC AGTCCCTGCA ATGATCTCTC CATCGTCAAT CCTGCCCGCA
 851  AGGGACTCTA CGAGGGCACT GGCCGGCTCT TCTTTGAGTT CTACCGCCTC
 901  CTGCATGATG CGCGGCCCAA GGAGGGAGAT GACCGCCCCT TCTTCTGGCT
 951  CTTTGAGAAT GTGGTGGCCA TGGGCGTTAG TGACAAGAGG GACATCTCGC
1001  GATTTCTCGA GTCCAACCCT GTGATGATTG ATGCCAAAGA AGTGTCAGCT
1051  GCGCACAGGG CCCGCTACTT CTGGGGGAAC CTTCCTGGTA TGAACAGGCC
1101  ATTGGCATCC ACTGTGAATG ATAAGCTGGA GCTGCAGGAG TGTCTGGAGC
1151  ACGGCCGAAT AGCCAAGTTC AGCAAAGTGA GGACCATTAC TACTAGGTCG
1201  AACTCCATAA AGCAGGGCAA AGACCAGCAT TTCCCCGTCT TCATGAATGA
1251  GAAAGAGGAC ATCTTATGGT GCACTGAAAT GGAAAGGGTG TTTGGCTTCC
```

1301 CTGTCCACTA TACCGACGTC TCCAACATGA GCCGCTTGGC GAGGCAGAGA

1351 CTGCTGGGCC GGTCGTGGAG CGTGCCGGTC ATCCGCCACC TCTTCGCTCC

1401 GCTGAAGGAA TATTTTGCTT GTGTGTAAGG GACGTGAGGG CAAATCGAGG

1451 TAGTGACACA AGGTTAAATA GACAAACAAA AAACACAAAA CACAACAAAA

1501 CACCAACAAC GTGAGGATGG AGAGAAGTAT CAGCACCCAG AAGAGAAAAA

1551 GGAATTTAAA ACAAAACAAA CAAACCCCAC AGAGGCGGAA ATTCCGGAGG

1601 GCTTTGCCTT GCAAAAAGGG TTGGACATCA TCTCCTGATT TTTCGATGTT

1651 AATCTTCAGT CCTATTTAAA AACAAAACCA AGCTCCCTCC CTACCCCCAC

1701 TACCCCCTTT TTTTCTGGTC AAACCTTTGT TTTCTACTCT TTTCAGAGGG

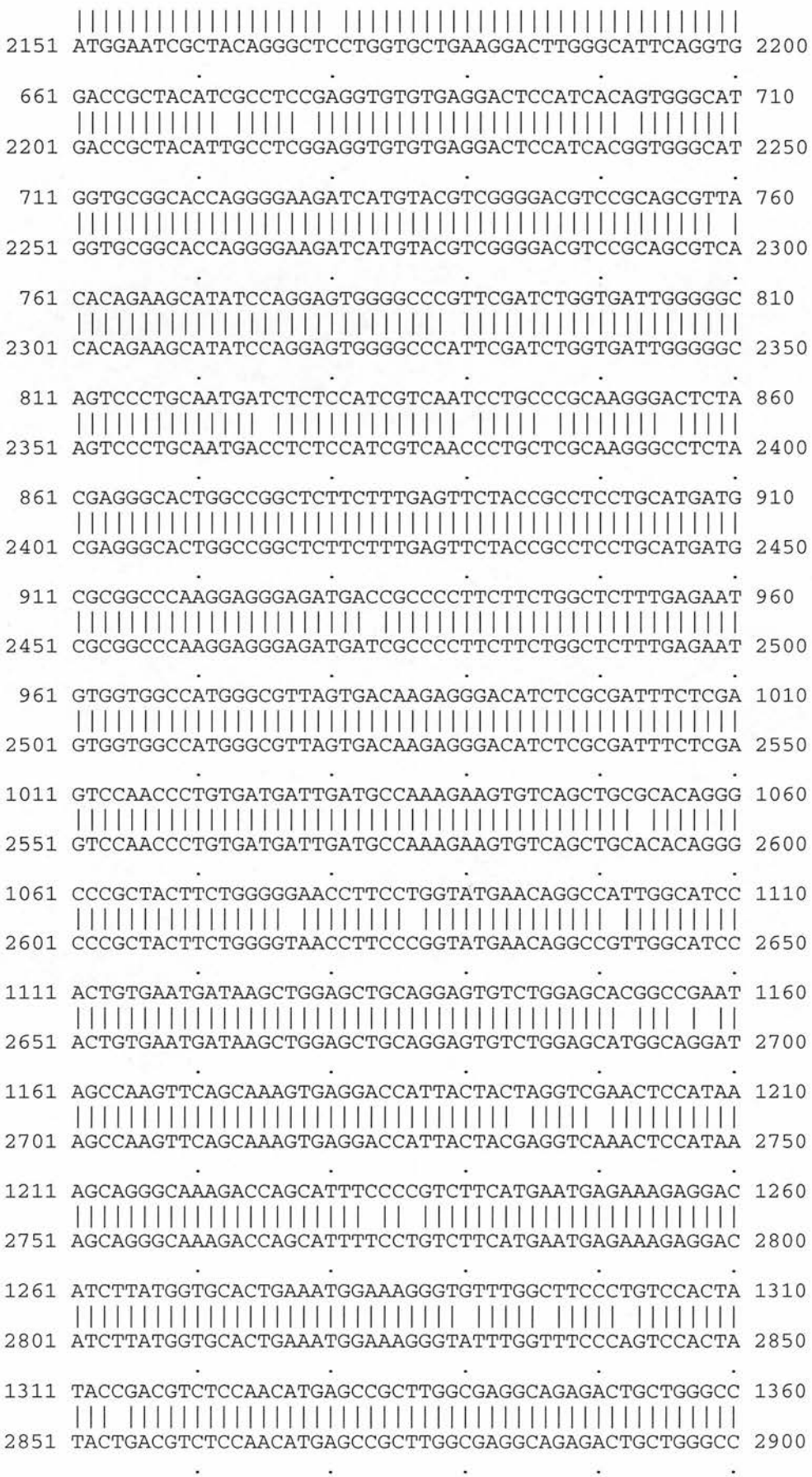
1751 GTTTTCTGTT TGTTTGGGTT TTTGTTTCTT GCTGTGACTG AAACAAGAGG

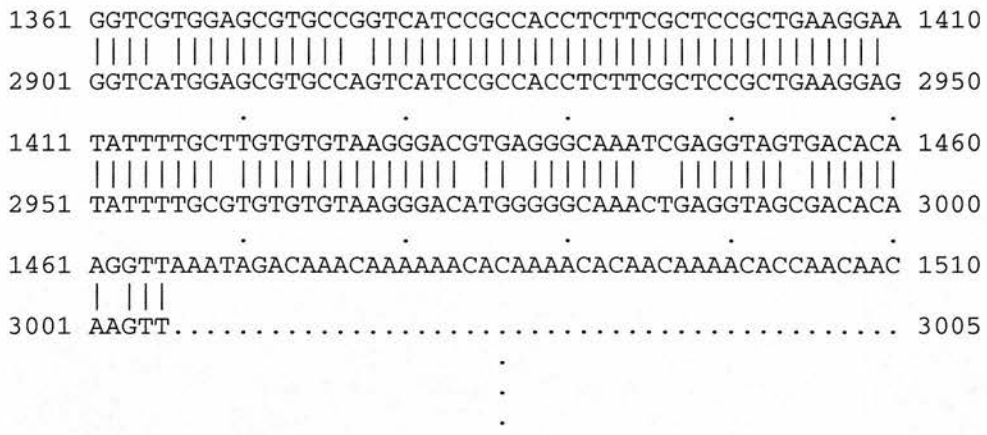
1801 GTTTATTGCA GCAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAA

Pairwise alignment of sheep and human (Accession no. AF067972)

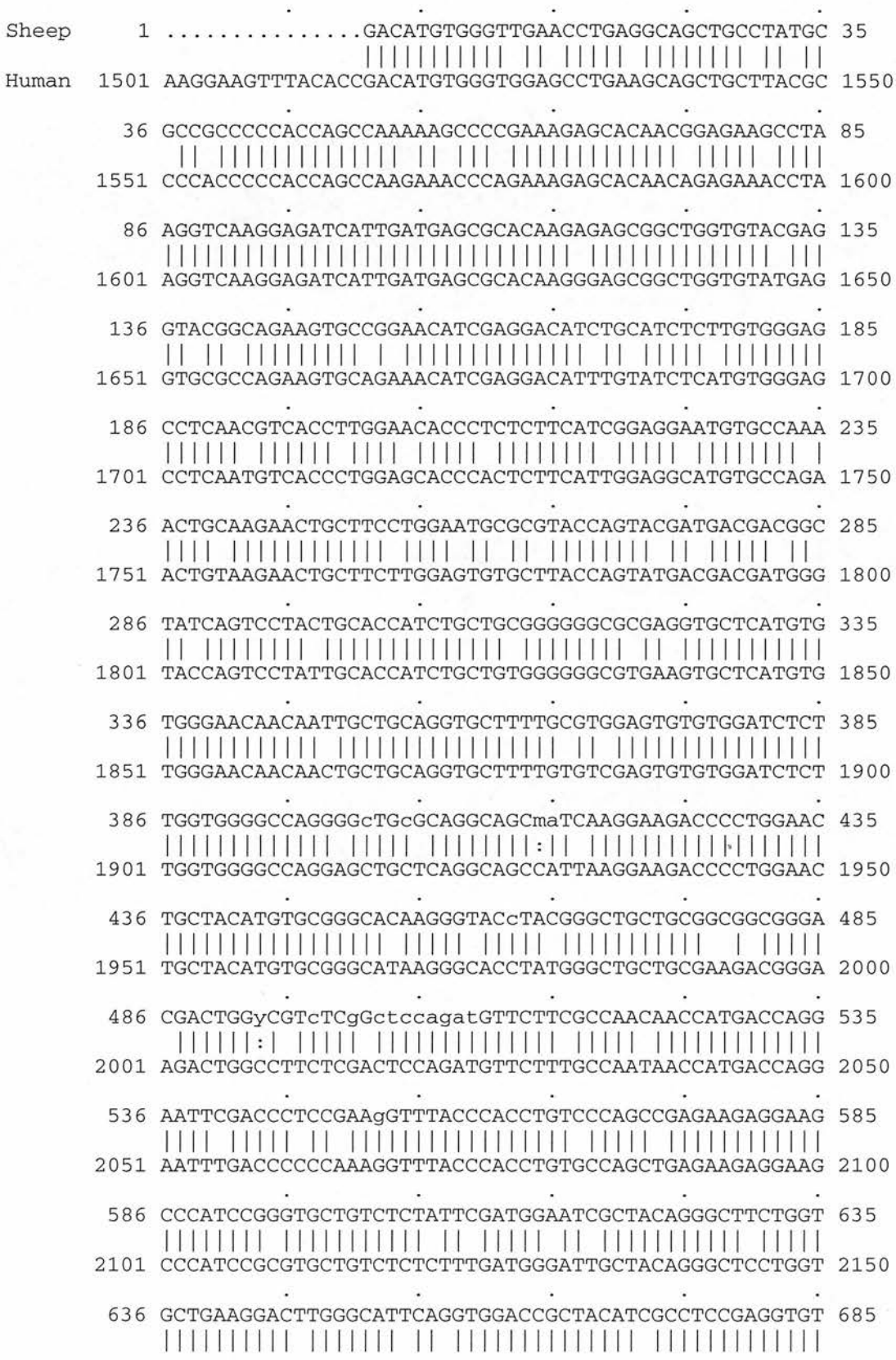
DNMT3A cDNA sequences generated using Gap (GCG10)

```
Sheep     1 ....................................GACATGTGGG 10
                                                 ||||||||||
Human  1501 ACCAGAAGAAGAGAAGAATCCCTACAAAGAAGTGTACACGGACATGTGGG 1550

          11 TTGAACCTGAGGCAGCTGCCTATGCGCCGCCCCCACCAGCCAAAAAGCCC 60
             | |||||||||||||||||||| || || || ||||||||||||||||||
        1551 TGGAACCTGAGGCAGCTGCCTACGCACCACCTCCACCAGCCAAAAAGCCC 1600

          61 CGAAAGAGCACAACGGAGAAGCCTAAGGTCAAGGAGATCATTGATGAGCG 110
             || ||||||||| |||||||||| |||||||||||||| ||||||||||
        1601 CGGAAGAGCACAGCGGAGAAGCCCAAGGTCAAGGAGATTATTGATGAGCG 1650

         111 CACAAGAGAGCGGCTGGTGTACGAGGTACGGCAGAAGTGCCGGAACATCG 160
             |||||||||||||||||||||||||| |||||||||||||||||||| |
        1651 CACAAGAGAGCGGCTGGTGTACGAGGTGCGGCAGAAGTGCCGGAACATTG 1700

         161 AGGACATCTGCATCTCTTGTGGGAGCCTCAACGTCACCTTGGAACACCCT 210
             ||||||||||||||| |||||||||||||| || ||| ||||||||||
        1701 AGGACATCTGCATCTCCTGTGGGAGCCTCAATGTTACCCTGGAACACCCC 1750

         211 CTCTTCATCGGAGGAATGTGCCAAAACTGCAAGAACTGCTTCCTGGAATG 260
             |||||| | |||||||||||||||||||||||||||||||| ||||| ||
        1751 CTCTTCGTTGGAGGAATGTGCCAAAACTGCAAGAACTGCTTTCTGGAGTG 1800

         261 CGCGTACCAGTACGATGACGACGGCTATCAGTCCTACTGCACCATCTGCT 310
              |||||||||||||| |||||||||| |||||||||||||||||||||
        1801 TGCGTACCAGTACGACGACGACGGCTACCAGTCCTACTGCACCATCTGCT 1850

         311 GCGGGGGGCGCGAGGTGCTCATGTGTGGGAACAACAATTGCTGCAGGTGC 360
             | ||||| || |||||||||||||| || |||||||| ||||||||||||
        1851 GTGGGGGCCGTGAGGTGCTCATGTGCGGAAACAACAACTGCTGCAGGTGC 1900

         361 TTTTGCGTGGAGTGTGTGGATCTCTTGGTGGGGCCAGGGGCTGCGCAGGC 410
             |||||||||||||||||| |||||||||||||| |||||||| |||||
        1901 TTTTGCGTGGAGTGTGTGGACCTCTTGGTGGGGCCGGGGGCTGCCCAGGC 1950

         411 AGCATTCAAGGAAGACCCCTGGAACTGCTACATGTGCGGGCACAAGGGTA 460
             |||  | ||||||||||||||||||||||||||||||||||||||||||
        1951 AGCCATTAAGGAAGACCCCTGGAACTGCTACATGTGCGGGCACAAGGGTA 2000

         461 CCTACGGGCTGCTGCGGCGGCGGGACGACTGGTCGTCTCGGCTCCAGATG 510
             |||||||||||||||||||| || |||||| | || ||||||||||||
        2001 CCTACGGGCTGCTGCGGCGGCGAGAGGACTGGCCCTCCCGGCTCCAGATG 2050

         511 TTCTTCGCCAACAACCATGACCAGGAATTCGACCCTCCGAAGGTTTACCC 560
             |||||||| || ||||| |||||||||| |||||||| ||||||||||
        2051 TTCTTCGCTAATAACCACGACCAGGAATTTGACCCTCCAAAGGTTTACCC 2100

         561 ACCTGTCCCAGCCGAGAAGAGGAAGCCCATCCGGGTGCTGTCTCTATTCG 610
             |||||||||||| |||||||||||||||||||||||||||||| || |
        2101 ACCTGTCCCAGCTGAGAAGAGGAAGCCCATCCGGGTGCTGTCTCTCTTTG 2150

         611 ATGGAATCGCTACAGGGCTTCTGGTGCTGAAGGACTTGGGCATTCAGGTG 660
```

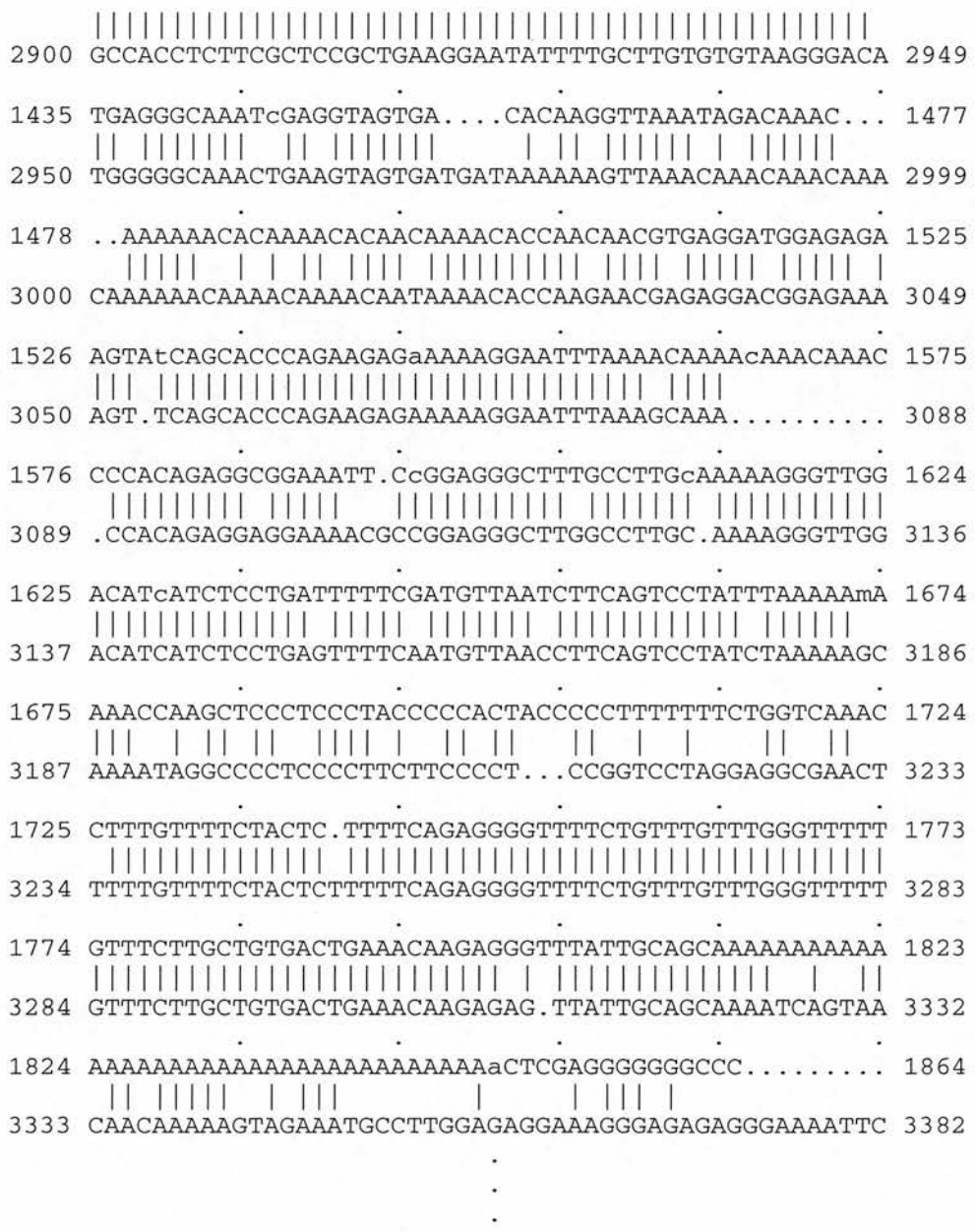



1361 GGTCGTGGAGCGTGCCGGTCATCCGCCACCTCTTCGCTCCGCTGAAGGAA 1410
2901 GGTCATGGAGCGTGCCAGTCATCCGCCACCTCTTCGCTCCGCTGAAGGAG 2950
1411 TATTTTGCTTGTGTGTAAGGGACGTGAGGGCAAATCGAGGTAGTGACACA 1460
2951 TATTTTGCGTGTGTGTAAGGGACATGGGGGCAAACTGAGGTAGCGACACA 3000
1461 AGGTTAAATAGACAAACAAAAAACACAAAACACAACAAAACACCAACAAC 1510
3001 AAGTT 3005

Pairwise alignment of the sheep and mouse (Accession No. AF068625) *DNMT3A* cDNA sequences generated using Gap (GCG10)



```
2151 GCTGAAGGACCTGGGCATCCAAGTGGACCGCTACATTGCCTCCGAGGTGT 2200
          .         .         .         .         .
 686 GTGAGGACTCCATCACAGTGGGCATGGTGCGGCACCAGGGGAAGATCATG 735
     |||||||||||||||| ||||||||||||||||||||||| |||||||||
2201 GTGAGGACTCCATCACGGTGGGCATGGTGCGGCACCAGGGAAAGATCATG 2250
          .         .         .         .         .
 736 TACGTCGGGGACGTCCGCAGCGTTACACAGAAGCATATCCAGGAGTGGGG 785
     |||||||||||||||||||||||| ||||||||||||||||||||||||||
2251 TACGTCGGGGACGTCCGCAGCGTCACACAGAAGCATATCCAGGAGTGGGG 2300
          .         .         .         .         .
 786 CCCGTTCGATCTGGTGATTGGGGGCAGTCCCTGCAATGATCTCTCCATCG 835
     ||| ||||| |||||||||| |||||||||||||||| |||||||| |
2301 CCCATTCGACCTGGTGATTGGAGGCAGTCCCTGCAATGACCTCTCCATTG 2350
          .         .         .         .         .
 836 TCAATcCTGCCCGCAAgGGACTCTACGAGGGCACTGGCCGGCTCTTCTTT 885
     |||| |||||||||||||||| || ||||| |||||||| |||||||||
2351 TCAACCCTGCCCGCAAGGGACTTTATGAGGGTACTGGCCGCCTCTTCTTT 2400
          .         .         .         .         .
 886 GAGTTCTACCGCCTCCTGCATGATGCGCGGCCCAAGGAGGGAGATGACCG 935
     |||||||||||||||||||||||||||||||||||||||||||||||| ||
2401 GAGTTCTACCGCCTCCTGCATGATGCGCGGCCCAAGGAGGGAGATGATCG 2450
          .         .         .         .         .
 936 CCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCATGGGCGTTAGTGACA 985
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2451 CCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCATGGGCGTTAGTGACA 2500
          .         .         .         .         .
 986 AGAGGGACATCTCGCGATTTCTCGAGTCCAaCCCTGTGATGATTGATGCC 1035
     |||||||||||||||||||| ||||| ||||| |||||||||||| |||
2501 AGAGGGACATCTCGCGATTTCTTGAGTCTAACCCCGTGATGATTGACGCC 2550
          .         .         .         .         .
1036 AAAGAAGTGT.CAGcTGCGCACAGGGCCCGCTACTTCTGGGGGAACCTTC 1084
     |||||||||||| ||||| |||||||||||| |||||||||||| |||||||
2551 AAAGAAGTGT.CTGCTGCACACAGGGCCCGTTACTTCTGGGGTAACCTTC 2599
          .         .         .         .         .
1085 CTGGTATGAACAGGCCATTGGCATCCACTGTGAATGATAAGCTGGAGCTG 1134
     |||| |||||||||||| ||||||||||||||||||||||||||||||||
2600 CTGGCATGAACAGGCCTTTGGCATCCACTGTGAATGATAAGCTGGAGCTG 2649
          .         .         .         .         .
1135 CAGGAGTGTCTGGAGCACGGCCGAATAGCCAAGTTCAGCAAAGTGAGGAC 1184
     || |||||||||||||||||| ||||||||||||||||||||||||||||||
2650 CAAGAGTGTCTGGAGCACGGCAGAATAGCCAAGTTCAGCAAAGTGAGGAC 2699
          .         .         .         .         .
1185 CATTACTACTAGGTCGAACTCCATAAAGCAGGGCAAAGACCAGCATTTCC 1234
     |||||| || ||||| ||||| ||||||||||||||||||||||||||||||
2700 CATTACCACCAGGTCAAACTCTATAAAGCAGGGCAAAGACCAGCATTTCC 2749
          .         .         .         .         .
1235 CCGTCTTCatgartgAGAAAGAGGACaTCTTATGGTGCACTGAAATGGAA 1284
     ||||||||||||: ||||| ||||||||| | ||||||||||||||||||
2750 CCGTCTTCATGAACGAGAAGGAGGACATCCTGTGGTGCACTGAAATGGAA 2799
          .         .         .         .         .
1285 AGGGTGTTTGGCTTCCCTGTCCACTATACCGACGTCTCCAACATGaGCCG 1334
     |||||||||||||||| |||||||| || ||||||||||||||||||||
2800 AGGGTGTTTGGCTTCCCCGTCCACTACACAGACGTCTCCAACATGAGCCG 2849
          .         .         .         .         .
1335 CTTGGCGAGGCAGAGACTGCTGGGCCGGTCGTGGAGCGTGCCGGTCATCC 1384
     |||||||||||||||||||||||| ||||||||||||||||||||||||
2850 CTTGGCGAGGCAGAGACTGCTGGGCCGATCGTGGAGCGTGCCGGTCATCC 2899
          .         .         .         .         .
1385 GCCACCTCTtcgcTCCGCTGAAGGAATATTTTGCTTGTGTGTAAGGGACG 1434
```



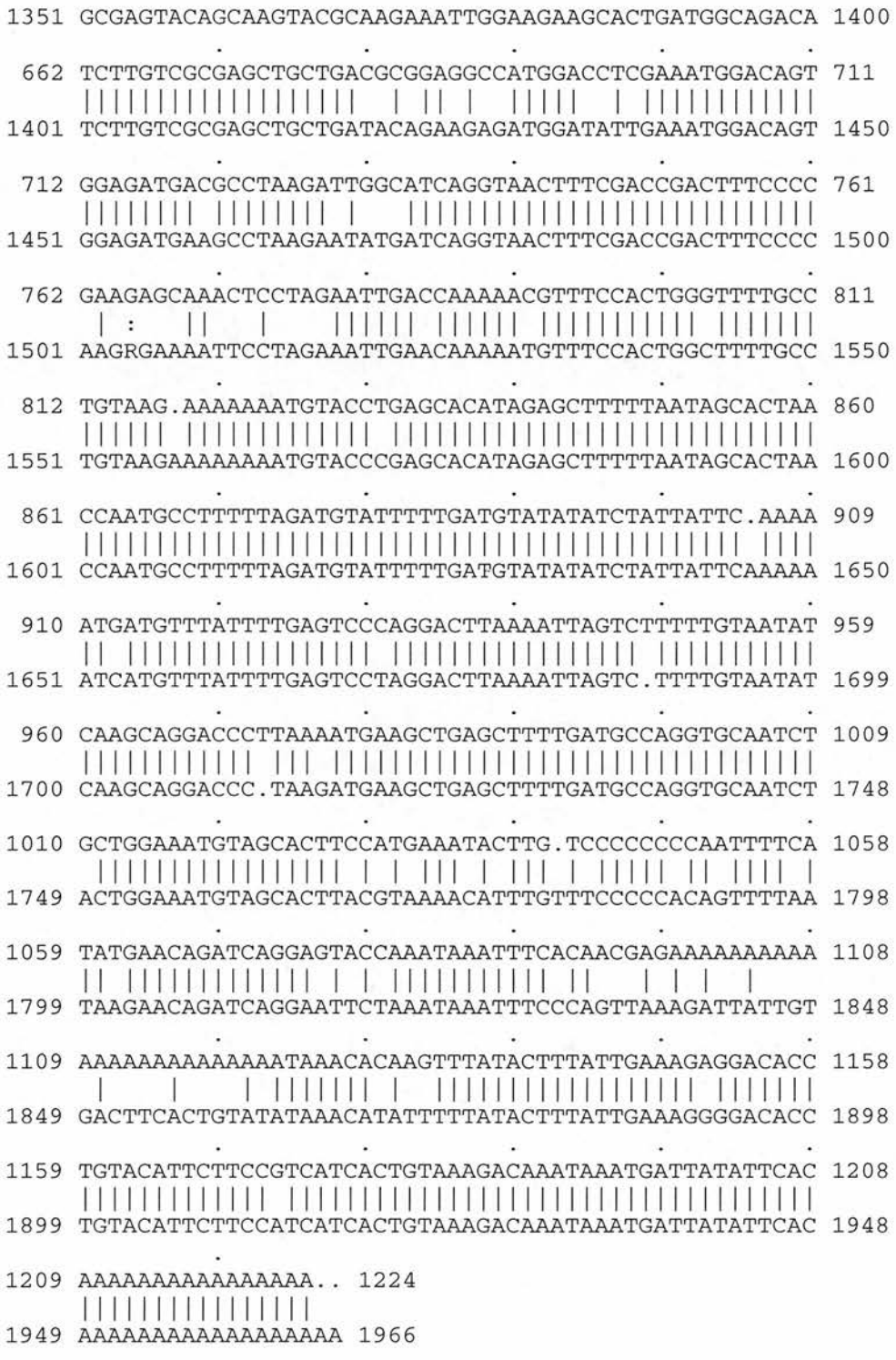
Appendix III

Sheep *MBD2* cDNA sequence. Length: 1224

```
   1  AGCGCCGGCA AGAGCGATGT CTACTACTTC AGTCCGAGCG GTAAGAAGTT
  51  CAGAAGCAAA CCTCAGTTGG CAAGGTATCT GGGAAATACT GTTGACCTCA
 101  GCAGTTTTGA CTTCAGAACT GGAAAGATGA TGCCTAGTAA ATTACAGAAG
 151  AACAAACAGA GACTGCGGAA TGATCCTCTG AATCAGAATA AGGGTAAACC
 201  AGACTTGAAT ACAACATTGC CAATTAGACA AACAGCATCT ATTTTCAAGC
 251  AACCAGTAAC CAAAGTCACA AATCATCCTA GTAATAAAGT GAAATCCGAC
 301  CCACAGAGAA TGAATGAACA GCCACGTCAG CTTTTCTGGG ACAACAGGCT
 351  ACAAGGAATT AGTGCACCAG ATGTAACAGA ACAAATTATA AAAACCATGG
 401  AACTCCCTAA AGGTCTTCAA GGAGTTGGTC CGGGTAGTAA CGATGAGACC
 451  CTTTTATCTG CTGTCGCCAG TGCTTTGCAC ACAAGCTCTG CACCCATCAC
 501  AGGGCAAGTC TCTGCTGCTG TGGAAAAGAA CCCCGCTGTG TGGCTGAACA
 551  CATCTCAACC CCTCTGCAAG GCTTTCATTG TCACAGATGA AGACATTAGG
 601  AAACAGGAAG AGCGAGTTCA GCAAGTACGT AAGAAATTGG AGGAGGCACT
 651  GATGGCAGAC ATCTTGTCGC GAGCTGCTGA CGCGGAGGCC ATGGACCTCG
 701  AAATGGACAG TGGAGATGAC GCCTAAGATT GGCATCAGGT AACTTTCGAC
 751  CGACTTTCCC CGAAGAGCAA ACTCCTAGAA TTGACCAAAA ACGTTTCCAC
 801  TGGGTTTTGC CTGTAAGAAA AAAATGTACC TGAGCACATA GAGCTTTTTA
 851  ATAGCACTAA CCAATGCCTT TTTAGATGTA TTTTTGATGT ATATATCTAT
 901  TATTCAAAAA TGATGTTTAT TTTGAGTCCC AGGACTTAAA ATTAGTCTTT
 951  TTGTAATATC AAGCAGGACC CTTAAAATGA AGCTGAGCTT TTGATGCCAG
1001  GTGCAATCTG CTGGAAATGT AGCACTTCCA TGAAATACTT GTCCCCCCCC
1051  AATTTTCATA TGAACAGATC AGGAGTACCA AATAAATTTC ACAACGAGAA
1101  AAAAAAAAAA AAAAAAAAAA AATAAACACA AGTTTATACT TTATTGAAAG
1151  AGGACACCTG TACATTCTTC CGTCATCACT GTAAAGACAA ATAAATGATT
1201  ATATTCACAA AAAAAAAAAA AAAA
```

Pairwise alignment of sheep and human (Accession No. AF072242) *MBD2* cDNA sequences generated using Gap (GCG10)

```
                 .         .         .         .         .
Sheep      1 ....................................AGCGCCGGCAA  11
                                                 || || |||||
Human    701 GGATGGAAGAAGGAGGAAGTGATCCGAAAATCTGGGCTAAGTGCTGGCAA 750

                 .         .         .         .         .
          12 GAGCGATGTCTACTACTTCAGTCCGAGCGGTAAGAAGTTCAGAAGCAAAC  61
             ||||||||||||||||||||||| || ||||||||||||||||||||| |
         751 GAGCGATGTCTACTACTTCAGTCCAAGTGGTAAGAAGTTCAGAAGCAAGC 800

                 .         .         .         .         .
          62 CTCAGTTGGCAAGGTATCTGGGAAATACTGTTGACCTCAGCAGTTTTGAC 111
             ||||||||||||||| |||||||||||||||| ||||||||||||||
         801 CTCAGTTGGCAAGGTACCTGGGAAATACTGTTGATCTCAGCAGTTTTGAC 850

                 .         .         .         .         .
         112 TTCAGAACTGGAAAGATGATGCCTAGTAAATTACAGAAGAACAAACAGAG 161
             ||||||||||||||||||||||||||||||||||||||||||||||||||
         851 TTCAGAACTGGAAAGATGATGCCTAGTAAATTACAGAAGAACAAACAGAG 900

                 .         .         .         .         .
         162 ACTGCGGAATGATCCTCTGAATCAGAATAAGGGTAAACCAGACTTGAATA 211
             |||||| || |||||||| ||||| ||||||||||||||||||||||||
         901 ACTGCGAAACGATCCTCTCAATCAAAATAAGGGTAAACCAGACTTGAATA 950

                 .         .         .         .         .
         212 CAACATTGCCAATTAGACAAACAGCATCTATTTTCAAGCAACCAGTAACC 261
             |||||||||||||||||||||||||| |||||||| ||||| ||||||
         951 CAACATTGCCAATTAGACAAACAGCATCAATTTTCAAACAACCGGTAACC 1000

                 .         .         .         .         .
         262 AAAGTCACAAATCATCCTAGTAATAAAGTGAAATCCGACCCACAGAGAAT 311
             |||||||||||||||||||||||||||||||||| ||||||||  ||||
        1001 AAAGTCACAAATCATCCTAGTAATAAAGTGAAATCAGACCCACAACGAAT 1050

                 .         .         .         .         .
         312 GAATGAACAGCCACGTCAGCTTTTCTGGGACAACAGGCTACAAGGAATTA 361
             |||||||||||||||||||||||||||| || |||||||||||| |||
        1051 GAATGAACAGCCACGTCAGCTTTTCTGGGAGAAGAGGCTACAAGGACTTA 1100

                 .         .         .         .         .
         362 GTGCACCAGATGTAACAGAACAAATTATAAAAACCATGGAACTCCCTAAA 411
             |||||||||||||||||||||||||||||||||||||||| || |||
        1101 GTGCATCAGATGTAACAGAACAAATTATAAAAACCATGGAACTACCCAAA 1150

                 .         .         .         .         .
         412 GGTCTTCAAGGAGTTGGTCCGGGTAGTAACGATGAGACCCTTTTATCTGC 461
             |||||||||||||||||||| ||||| || ||||||||||||||||||||
        1151 GGTCTTCAAGGAGTTGGTCCAGGTAGCAATGATGAGACCCTTTTATCTGC 1200

                 .         .         .         .         .
         462 TGTCGCCAGTGCTTTGCACACAAGCTCTGCACCCATCACAGGGCAAGTCT 511
             ||| |||||||||||||||||||||||| || ||||||||||||||||
        1201 TGTTGCCAGTGCTTTGCACACAAGCTCTGCGCCAATCACAGGGCAAGTCT 1250

                 .         .         .         .         .
         512 CTGCTGCTGTGGAAAAGAACCCCGCTGTGTGGCTGAACACATCTCAACCC 561
             | |||||||||||||||||||| ||||| ||||| ||||||||||||||
        1251 CCGCTGCTGTGGAAAAGAACCCTGCTGTTTGGCTTAACACATCTCAACCC 1300

                 .         .         .         .         .
         562 CTCTGCAAGGCTTTCATTGTCACAGATGAAGACATTAGGAAACAGGAAGA 611
             |||||||| ||||| |||||||||||||||||| ||||||||||||||
        1301 CTCTGCAAAGCTTTTATTGTCACAGATGAAGACATCAGGAAACAGGAAGA 1350

                 .         .         .         .         .
         612 GCGAGTTCAGCAAGTACGTAAGAAATTGGAGGAGGCACTGATGGCAGACA 661
             |||||| |||||||||| |||||||||| || ||||||||||||||||
```



Appendix IV

The following pages show raw data acquired during the methyl-sensitive PCR experiments, and subsequently used to calculate DNA methylation levels.
Each table is clearly labelled with embryo stage/tissue type and the specific DNA sequence under analysis. The following rules apply to each table:

Sample source is indicated in the first column.
Volumes 1-4 (or Vol 1-3) are band intensities of undigested DNA PCR products, with the average indicated as Average 1.
Volumes 5-8 (or Vol 4-6) are band intensities of *Hha* I digested PCR products, with the average indicated as Average 2.
% meth is Average 2 divided by Average1.
Real % meth is the transformed % meth value (transformed using relevant standard curve as described in Chapter 7).
Band intensity values in **Bold** type were not included when calculating the average, as they deviated substantially from the mean. These values generally correlated with the presence of an unusually intense PCR product band, or complete absence of a PCR product, indicating manual error.

Satellite I

Tissue	Vol 1	Vol 2	Vol 3	Average 1	Vol 4	Vol 5	Vol 6	Average 2	% meth	real % meth
H1	157.16	186.54	199.25	180.9833	189.39	187.36	123.6	166.7833	92	**89**
H2	215.69	174.63	199.94	196.7533	229.5	212.44	211.03	217.6567	110	**113**
FF1	134.24	111.44	121.72	122.4667	76.11	96.01	97.9	90.00667	73	**67**
FF2	129.03	115.53	132.42	125.66	132.09	112.44	120.94	121.8233	97	**95**
FF3	75.66	73.1	72.92	73.89333	53.38	60.37	52.21	55.32	75	**69**
OV-2 SINE										
H1	81.64	82.82	72.5	78.98667	89.58	78.46	79.26	82.43333	105	**109**
H2	29.6	**13.89**	32.24	30.92	33.86	25.41	33.44	30.90333	100	**104**
FF1	60.74	48.61	59.04	56.13	37.08	45.19	42.95	41.74	74	**73**
FF2	102.67	79.09	70.11	83.95667	61.6	59.7	68.65	63.31667	76	**75**
FF3	41.34	45.12	38.17	41.54333	25.57	29.93	27.12	27.54	66	**64**

H= heart
F=fetal fibroblast

Oocyte DNA sequence: Satellite I

Oocyte	Vol 1	Vol 2	Vol 3	Vol 4	Average 1	Vol 5	Vol 6	Vol 7	Vol 8	Average 2	%meth	real % met
O1	69.18	72.76	64.47	63.34	67.4375	23.87	30.27	28.23	24.97	26.835	40	**22**
O2	82.51	55.18	72.98	98.71	77.345	56.74	54.36	46.18	47.75	51.2575	66	**41**
O3	72.51	92.43	73.41	82.55	80.225	37.64	38.19	37.93	44.52	39.57	49	**28**
O4	129.77	125.26	141.96	128.31	131.325	70.23	73.55	68.57	84.18	74.1325	56	**33**
O5	110.83	95.81	91.44	99.06	99.285	33.62	40.79	39.12	33.22	36.6875	37	**21**
O6	50.63	54.14	59.14	50.43	53.585	18.67	16.56	13.47	18.1	16.7	31	**17**
O7	137.03	128.68	131.19	121.14	129.51	42.38	39.67	48.94	49.04	45.0075	35	**19**
O8	79.89	66.63	77.04	81.9	76.365	35.17	34.17	26.39	38.38	33.5275	44	**25**
O9	152.43	130	128.23	131.88	135.635	53.06	41.37	47.37	47.41	47.3025	35	**19**

Sperm

Satellite I

	Vol 1	Vol 2	Vol 3	Average 1	Vol 4	Vol 5	Vol 6	Average 2	% meth	real % meth
SP1	114.44	112.9	101.94	109.76	37.71	39.56	50.45	42.573333	39	**34**
SP2	82.24	89.47	85.62	85.776667	18.23	14.93	23	18.72	22	**18**
SP3	42.18	33.7	30.02	35.3	13.98	11.48	22.61	16.023333	45	**39**

OV-2 SINE

SP1	65.56	75.62	75.88	72.353333	32.46	27.95	22.76	27.723333	38	**34**
SP2	38.73	53.91	72.15	54.93	13.26	17.74	15.21	15.403333	28	**23**
SP3	63.74	58.47	71.26	64.49	17.24	12.57	27.6	19.136667	30	**25**

Embryo stage: 8-16 cell DNA sequence: Satellite I

Embryo	Vol 1	Vol 2	Vol 3	Vol4	Average 1	Vol 5	Vol 6	Vol 7	Vol 8	Average 2	% meth. real	% met
8.1	77.62	92.76	101.69	-	90.69	19.18	18.93	19.07	-	19.06	21	**14**
8.2	419.64	519.93	510.58	-	483.383333	252.64	284.95	252.31	-	263.3	54	**41**
8.3	196.98	143.98	178.59	-	173.183333	79.15	72.56	101.95	-	84.5533333	49	**36**
8.4	200.7	218.07	209.48	-	209.416667	123.45	105.06	187.58	-	138.696667	66	**53**
8.5	**612.48**	145.36	157.77	-	151.565	88.17	110.87	164.73	-	121.256667	80	**69**
8.6	187.07	128.9	120.75	-	145.573333	33.46	21.77	23.12	-	26.1166667	18	**12**
8.7	**580.07**	295.73	300.31	-	298.02	86.69	84.32	83.87	-	84.96	29	**20**
8.8	161.63	162.16	184.11	-	169.3	77.42	95.66	79.84	-	84.3066667	50	**37**
8.9	197.39	227.49	216.31	-	213.73	51.64	58.92	56.76	-	55.7733333	26	**18**
8..10	77.15	78.16	73.76	-	76.3566667	30.28	17.05	24.2	-	23.8433333	31	**21**
8..11	184.54	223.71	179.68	-	195.976667	129.77	91.48	97.02	-	106.09	54	**41**
8..12	81.67	89.37	69.94	-	80.3266667	30.04	60.12	43.44	-	44.5333333	55	**42**
NT8.1	264.68	209.69	185.65	-	220.006667	25.47	19.89	23.77	-	23.0433333	10	**6**
NT8.2	162.99	191.77	171.64	-	175.466667	14.34	10.34	7.06	-	10.58	6	**3**
NT8.3	380.01	386.4	365.83	-	377.413333	12.88	14.92	21.6	-	16.4666667	4	**2**
NT8.4	**11.87**	445.7	410.19	-	427.945	290.91	187.96	186.5	-	221.79	52	**40**
NT8.5	100.99	97.36	100.7	-	99.6833333	30.66	29.12	26.94	-	28.9066667	29	**20**
NT8.6	98.92	92.45	96.4	-	95.9233333	9.12	8.26	9.29	-	8.89	9	**5**
NT8.7	84.35	96.3	98.1	-	92.9166667	**0**	8.25	3.32	-	5.785	6	**3**
NT8.8	433.72	369.96	348.95	-	384.21	82.4	41.82	63.21	-	62.4766667	16	**10**
NT8.9	403.57	384.18	382.24	-	389.996667	33.73	26.01	21.98	-	27.24	7	**4**
IVF8.1	297.39	270.4	275.69	352.38	298.965	33.46	19.11	22.38	26.7	25.4125	9	**5**
IVF8.2	149.83	151.45	132.16	175.79	152.3075	13.02	14.38	31.42	11.16	17.495	11	**7**
IVF8.3	185.15	184.15	168.86	151.97	172.5325	98.39	113.81	109.26	93.16	103.655	60	**47**
IVF8.4	227.29	205.77	273.25	200.24	226.6375	6.33	13.12	8.34	6.52	8.5775	4	**2**
IVF8.5	**0**	220.23	281.01	358.05	286.43	31.71	26.76	29.36	21.54	27.3425	10	**6**
IVF8.6	106.75	85.9	87.92	70.15	87.68	13.11	12.84	16.23	14.42	14.15	16	**10**
IVF8.7	372.98	289.23	**61.8**	222.4	294.87	46.26	70.93	62.93	55.69	58.9525	20	**13**

Embryo stage: morula DNA sequence: Satellite I

EMBRYO	Vol1	Vol 2	Vol 3	Vol 4	Average 1	Vol 5	Vol 6	Vol 7	Vol 8	Average 2	%meth	real % met
M1	235.73	252.91	213.81	201.54	225.9975	43.79	51.25	54.19	34.15	45.845	20	**18**
M2	**89.66**	163.16	173.52	180.13	172.27	23.02	26.42	26.17	28.44	26.0125	15	**13**
M3	163.91	135.6	158.45	163.34	155.325	30.2	43.08	42.52	37.68	38.37	25	**23**
M4	124.91	131.76	134.04	93.55	121.065	23.67	26.27	24.18	**10.44**	24.7066667	20	**18**
M5	85.52	92.62	98.45	82.68	89.8175	42.87	42.85	41.55	43.19	42.615	47	**46**
M6	67.94	54.98	65.61	60.32	62.2125	26.86	20.8	24.29	22.13	23.52	38	**37**

Embryo stage: blastocyst DNA sequence: Satellite I

Embryo	Vol 1	Vol 2	Vol 3	Vol 4	Average 1	Vol 5	Vol 6	Vol 7	Vol 8	Average 2	% meth.	real % meth
B1	468.01	543.59	450.16	450.73	478.1225	159.82	133.06	125.3	137.41	138.8975	29	**25**
B2	727.82	663.44	695.06	699.63	696.4875	211.92	215.54	217.93	249.01	223.6	32	**27**
B3	748.54	720.82	673.38	640.77	695.8775	288.1	285.91	290.17	281.77	286.4875	41	**35**
B4	773.55	727.73	731.49	756.61	747.345	214.91	215.09	210.72	190.99	207.9275	32	**27**
B5	813.83	800.05	813.59	844.93	818.1	288.95	234.83	265.35	248.31	259.36	47	**41**
B6	297.84	292.66	302.23	318.43	302.79	137.55	138.42	142.12	151.49	142.395	47	**41**
B7	413.19	407.23	378.92	405.14	401.12	254.78	231.1	231.29	269.29	246.615	61	**55**
B8	486.44	434.99	388.87	416.41	431.6775	217.28	207.09	219.41	199.04	210.705	49	**43**
B9	748.46	643.24	704.62	682.11	694.6075	330.12	366.03	384.65	381.41	365.5525	53	**47**
B10	188.53	169.53	175.57	161.67	173.825	69.06	65.83	65.92	51.32	63.0325	36	**31**
B11	165.31	149.26	186.8	178.76	170.0325	109.83	115.13	108.52	112.57	111.5125	66	**60**
B12	188.09	184.59	166.78	148.87	172.0825	59.32	55.41	60.36	48.2	55.8225	32	**27**
B13	215.3	182.27	211.28	213.52	205.5925	44.07	46.49	48.55	56.33	48.86	24	**20**
B14	195.78	189.36	179.53	144.68	177.3375	76.76	79.73	86.38	78.2	80.2675	45	**39**
B15	111.47	120.79	95.89	119.89	112.01	47.17	43.56	42.83	43.8	44.34	40	**34**
IVFB1	115.84	137.19	147.05	196.86	149.235	38.05	41.07	33.32	41.48	38.48	26	**22**
IVFB2	81.39	92.24	**215.99**	77.57	83.7333333	30.68	30.64	26.9	24.58	28.2	34	**29**
IVFB3	138.37	159.78	158.44	168.76	156.3375	13.1	17.02	23.9	20.87	18.7225	12	**10**
IVFB4	74.49	93.94	90.72	96.45	88.9	17.26	22.63	23.72	23.38	21.7475	24	**20**
IVFB5	69.17	108.57	101.05	79.65	89.61	23.8	21	22.03	22.85	22.42	25	**21**
IVFB6	200.74	202.65	189.57	212.74	201.425	56.74	45.85	52.79	42.32	49.425	24	**20**
IVFB6	187.57	213.87	186.59	158.77	186.7	48.75	50.85	53.68	50.3	50.895	27	**23**
IVFB8	254.39	221.76	224.91	220.62	230.42	16.07	16.2	27.24	20.34	19.9625	9	**8**
IVFB9	252.61	260.83	257.66	213.81	246.2275	37.89	41.12	39.35	30.75	37.2775	15	**13**
IVFB10	318.12	291.46	384.13	346.7	335.1025	47.26	33.26	25.51	26.91	33.235	10	**8**
IVFB11	228.88	254.96	276.98	258.44	254.815	38.21	39.45	41.53	36.66	38.9625	15	**13**
IVFB12	155.45	160.12	165.3	151.67	158.135	22.97	23.38	28.1	17.7	23.0375	15	**13**
IVFB13	276.3	257.13	254.62	230.5	254.6375	100.03	103.66	95.32	90.67	97.42	38	**33**
IVFB14	174.69	136.19	160.6	188.21	164.9225	24.81	21.95	24.29	19.41	22.615	14	**12**

IVFB15	142.51	118.33	135.2	117.82	128.465	21.62	18.28	12.1	17.48	17.37	14	**12**
IVFB16	95.96	109.27	118.73	92.71	104.1675	26.67	19.7	28.67	19.22	23.565	23	**19**
IVFB17	137.57	136.99	**214.54**	120.06	131.54	44.27	35.53	34.84	38.1	38.185	29	**25**

Embryo	**Vol 1**	**Vol 2**	**Vol 3**	**Vol 4**	**Average 1**	**Vol 5**	**Vol 6**	**Vol 7**	**Vol 8**	**Average 2**	**% meth.**	**real % meth**
NTB1	460.35	415.97	371.73	410.34	414.5975	171.82	169.48	181.96	157.51	170.1925	41	**35**
NTB2	836.59	746.51	705.56	816.32	776.245	243.77	216.41	214.82	212.07	221.7675	29	**25**
NTB3	187.54	156.736	164.84	156.75	166.4665	103.93	91.8	98.41	95.25	97.3475	58	**52**
NTB4	456.13	352.64	424.35	391.44	406.14	249.9	237.78	232.05	202.55	230.57	57	**51**
NTB5	185.57	230.9	234.93	235.94	221.835	67.75	64.75	67.88	47.81	62.0475	28	**24**
NTB6	146.33	145.92	159.87	168.79	155.2275	60.7	72.05	66.1	72.42	67.8175	44	**38**
NTB7	118.97	109.49	119.52	112.84	115.205	65.64	52.99	53.55	58.31	57.6225	50	**44**
NTB8	405.05	394.77	389.25	363.39	388.115	128.25	127.11	121.36	121.76	124.62	32	**27**
NTB9	486.66	370.04	453.08	457.83	441.9025	205.82	231.57	233.47	230.01	225.2175	51	**45**
NTB10	304.75	246.91	244.8	298.55	273.7525	229.72	185.72	190.67	178.37	196.12	72	**66**
NTB11	252.86	186.37	241.06	222.38	225.6675	150.11	152.11	174.65	132.97	152.46	68	**62**
NTB12	242.96	244.52	221.46	**119.84**	236.313333	23.38	21.28	24.85	26.17	23.92	10	**8**
NTB13	297.55	290.57	280.38	215.75	271.0625	40.23	41.88	47.22	45.84	43.7925	16	**13**
NTB14	133.86	136.71	134.8	145.72	137.7725	17.75	25.56	22.93	22.72	22.24	16	**13**
NTB15	254.61	331.59	328.27	341.79	314.065	268.74	257.53	271.75	251.35	262.3425	84	**80**
NTB16	124.1	106.95	**173.96**	112.23	114.426667	34.01	29.83	26.63	22.16	28.1575	25	**21**
NTB17	110.63	100.95	125.51	102.6	109.9225	42.37	34.13	41.08	49.73	41.8275	38	**33**
NTB18	479.96	528.84	532.55	545.8	521.7875	101.97	111.96	112.42	116.65	110.75	21	**18**
NTB19	175.27	175.66	177.05	166.91	173.7225	63.36	74.29	67.28	76.1	70.2575	40	**34**
NTB20	327.93	338.66	266.65	302.8	309.01	25.94	28.78	27.56	24.2	26.62	9	**8**
NTB21	104.14	116.08	127.16	136.83	121.0525	93.79	81.82	91.96	**51.76**	89.19	74	**68**

Embryo stage: blastocyst DNA sequence: OV-2 SINE

Embryo	Vol 1	Vol 2	Vol 3	Vol 4	Average 1	Vol 5	Vol 6	Vol 7	Vol 8	Average 2	% meth.	real % meth
B1	571.79	590.08	584.79	574.31	580.2425	409.43	400.16	422.72	296.86	382.2925	66	**64**
B2	694.81	742.84	630.19	595.28	665.78	330.6	336.53	337.32	344.97	337.355	51	**47**
B3	595.03	601.05	633.92	646.5	619.125	395.14	381.21	311.32	317.23	351.225	57	**54**
B4	604.87	651.07	639.87	635.76	632.8925	498.18	508.7	543.88	481.67	508.1075	80	**80**
B5	131.07	129.89	145.43	140.12	136.6275	94.32	87.45	89.81	86.49	89.5175	66	**64**
B6	160.05	137.98	154.2	165.78	154.5025	113.65	143.73	121.65	138.95	129.495	84	**84**
B7	92.04	107.17	100.79	96.58	99.145	68.92	73.88	74.85	38.06	63.9275	64	**62**
B8	165.22	135.96	135.97	154.69	147.96	109.48	98.75	120.78	85.28	103.5725	70	**68**
B9	182.34	249.02	266.38	235.71	233.3625	113.79	119.49	108.84	112.68	113.7	49	**45**
B10	174.8	163.16	160.72	178.63	169.3275	53.38	41.53	54.65	32.97	45.6325	27	**22**
IVFB1	166.95	134.74	136.35	183.22	155.315	48.28	36.37	34.49	27.75	36.7225	24	**19**
IVFB2	125.65	96.7	149.54	114.71	121.65	29.56	27.23	28.76	52.21	34.44	28	**23**
IVFB3	174.29	163.11	149.7	133.71	155.2025	17.8	26.75	22.49	16.39	20.8575	13	**8**
IVFB4	120.93	100.14	93.45	107.33	105.4625	35.72	34.28	30.91	25.35	31.565	30	**25**
IVFB5	113.52	104.83	108.06	123.85	112.565	19.11	16.01	18.79	21.48	18.8475	17	**12**
IVFB6	166.4	188.14	181.66	152.52	172.18	44.85	48.99	32.4	35.85	40.5225	24	**19**
IVFB7	165.43	276.39	203.63	289.86	233.8275	27.89	17.12	13.87	13.85	18.1825	8	**3**
IVFB8	246.14	236.07	272.64	285.08	259.9825	20.75	12.18	18.2	9.98	15.2775	6	**1**
IVFB9	202.84	148.83	170.71	130.55	163.2325	44.16	29.25	35.39	32.56	35.34	22	**17**
IVFB10	82.59	79.81	87.95	52.53	75.72	23.38	14.64	16.28	13.78	17.02	22	**17**
NTB1	450.22	329.26	372.41	274.51	356.6	51.44	33.11	42.81	37.69	41.2625	12	**7**
NTB2	707.92	721.44	661.9	600.17	672.8575	144.57	145.2	136.64	141.3	141.9275	21	**16**
NTB3	654.73	731.36	699.69	702.85	697.1575	71.42	74.81	94.12	112.44	88.1975	13	**8**
NTB4	774.74	563.8	559.08	615.31	628.2325	178.77	162.6	159.59	259.33	190.0725	30	**25**
NTB5	169.92	165.93	140.7	171.83	162.095	16.13	25.8	24.46	23.74	22.5325	14	**9**
NTB6	153.78	166.53	221.52	198.24	185.0175	32.85	16.42	21.08	20.94	22.8225	12	**7**
NTB7	426.54	462.28	404.01	459.9	438.1825	179.34	274.27	266.13	191.15	227.7225	52	**49**
NTB8	83.25	112.56	148.5	147.24	122.8875	**0**	20.77	33.53	22.42	25.5733333	21	**16**
NTB9	86.91	99.28	104.16	113.54	100.9725	25.53	23.35	23.05	27.37	24.825	25	**20**

Embryo	Vol 1	Vol 2	Vol 3	Vol 4	Average 1	Vol 5	Vol 6	Vol 7	Vol 8	Average 2	% meth.	real % meth
NTB10	242.99	252.17	203.86	205.62	226.16	36.28	33.84	42.36	45.43	39.4775	17	**12**
NTB11	462.12	416.81	434.54	428.45	435.48	170.14	143.47	153.31	151.69	154.6525	26	**21**
NTB12	181.4	197.11	188.95	198.23	191.4225	64.87	62.19	46.53	57.56	57.7875	30	**25**
NTB13	358.08	315.05	321.9	350.76	336.4475	35.24	35.11	38.79	34.86	36	10	**5**
NTB15	162.08	140.99	173.41	165.04	160.38	117.71	109.82	111.48	130.64	117.4125	73	**72**
NTB16	344.75	458.99	401.88	462.33	416.9875	32.39	26.61	30.79	28.04	29.4575	7	**2**
NTB17	485.49	293.16	399.89	381.5	390.01	88.88	90.95	100.87	105.1	96.45	25	**20**
NTB18	213.96	225.51	220.21	231.93	222.9025	59.79	55.2	59.88	56.82	57.9225	26	**21**
NTB19	405.49	392.86	397.58	416.53	403.115	80.26	62.51	71.03	72.29	71.5225	18	**13**
NTB20	114.6	142.23	128.09	83.24	117.04	34.78	25.2	**6.63**	29.4	29.7933333	25	**20**