

Dnmt1 and Dnmt3 of the Winter Skate (*Leucoraja ocellata*)

By
Gordon D. Lake

A thesis submitted to the
School of Graduate Studies
in partial fulfillment of the
requirements for the degree of
Master of Science

Department of Biology
Memorial University of Newfoundland

July 2011

St. John's

Newfoundland and Labrador

Abstract

Chromatin management throughout vertebrate development is a stringently controlled and dynamic process. One event involved in that process is the covalent modification of cytosine nucleotides by the addition of a methyl group to the fifth carbon atom of the pyrimidine ring. Enzymes involved in establishing and maintaining DNA methylation patterns (DNA-methyltransferases: Dnmts) have been isolated and characterized in the teleost and mammalian lineages. There are structural similarities between enzymes as well as basic developmental expression patterns but there are fundamental differences in gene number and function in the context of two very different reproductive strategies. Subclass Elasmobranchii displays a wide variety of reproductive strategies and studying genome management in this subclass could provide great insight into the key differences and similarities in DNA methylation observed in the tetrapod and teleost models. I have isolated and characterized the first full length cDNAs of the maintenance and *de novo* methyltransferases (*Dnmt1* and *Dnmt3* respectively) from the Winter Skate (*Leucoraja ocellata*), a member of subclass Elasmobranchii. Evidence is presented for multiple *Dnmt3* splice variants as well as at least two *Dnmt3* retrotransposed pseudogenes. Preliminary experiments indicate that the early developmental methylation dynamics observed in both mammals and teleosts may also be present in *L. ocellata*.

Acknowledgments

I would like to thank the faculty and staff of Memorial University for giving me the opportunity to work towards, and the assistance necessary to complete, this degree. My supervisor Ross McGowan (Biology/Biochemistry Department) provided an excellent learning environment while H el ene Volkoff (Biology/Biochemistry Department) and Ed Yaskowiak (Genomics and Proteomics facility) were always willing to provide input regarding general research methods and experimental protocols. My colleagues and fellow graduate students were ever present throughout the editing and preparation of this thesis. None of my research into subclass Elasmobranchii would have been possible had the Ocean Sciences Centre (Logy Bay, NL, Canada) not maintained *Leucoraja ocellata* specimens. Additional funding, aside from the School of Graduate Studies stipend, was provided by the National Sciences and Engineering Research Council of Canada and also played a large role in the completion of this study. Finally, I would like to thank my family for their patience and support throughout this experience. Without them I'm sure I never would have made it as far as I have.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
List of Tables.....	v
List of Figures.....	vi – viii
List of Abbreviations.....	ix – x
Introduction.....	1 – 18
Materials and Methods.....	19 – 31
Results.....	32 – 68
Discussion.....	69 – 94
Summary and Conclusion.....	95 – 97
References.....	98 – 115

List of Tables

Table 1: Primers used throughout the isolation and characterization of <i>Leucoraja ocellata</i> <i>Dnmt3</i> and <i>Dnmt1</i>	21
Table 2: Comparison of <i>Dnmt1</i> translation start sites in vertebrates.....	39
Table 3: Comparison of <i>Dnmt3</i> translation start sites in vertebrates.....	50
Table 4: Tissue specific differential expression of the <i>L. ocellata</i> <i>Dnmt3</i> 3' transcript variants	53
Table 5: Tally of post transcriptional splice sites.....	63
Table 6: Comparison of <i>L. ocellata</i> <i>Dnmt1</i> to representative vertebrate <i>Dnmt1</i> proteins.....	75

List of Figures

Figure 1: Catalytic mechanism of DNA (cytosine-5) methyltransferases.....	4
Figure 2: Schematic of <i>L. ocellata Dnmt1</i> cDNA	23
Figure 3: Schematic of <i>L. ocellata Dnmt3</i> cDNA.....	26
Figure 4: SMARTer RACE RT-PCR products of the 3' and 5' ends of the <i>L. ocellata Dnmt3</i>	28
Figure 5: Methylated cytosine is elevated in the testis of <i>L. ocellata</i> when compared to ovary and somatic tissues.....	33
Figure 6: The 5' end of <i>L. ocellata Dnmt1</i> generated by 5' RACE RT-PCR.....	35
Figure 7: <i>Leucoraja ocellata DNA-methyltransferase 1</i> cDNA sequence.....	36 - 37
Figure 8: <i>Leucoraja ocellata DNA-methyltransferase 1</i> amino acid sequence.....	38
Figure 9: Dnmt1 functional domains in representative vertebrate models	41

Figure 10: Catalytic motifs present in the <i>L. ocellata</i> Dnmt1 C-5 cytosine-specific DNA methyltransferase domain	42
Figure 11: <i>Leucoraja ocellata</i> DNA-methyltransferase 3 cDNA sequence.....	44
Figure 12: <i>Leucoraja ocellata</i> DNA-methyltransferase 3 amino acid sequence.....	45
Figure 13: Dnmt3 functional domains in representative vertebrate models.....	46 - 47
Figure 14: Catalytic motifs in the long 3' variant <i>L. ocellata</i> Dnmt3 C-terminus.....	48
Figure 15: Both the 703 bp and 514 bp 3' transcript variants of the <i>L. ocellata</i> Dnmt3 are present in four different tissue types	52
Figure 16: The long 5' end of the <i>L. ocellata</i> Dnmt3 transcript matches with both the long and short 3' ends.....	54
Figure 17: The short 5' end of the <i>L. ocellata</i> Dnmt3 transcript matches with both the long and short 3' ends.....	56
Figure 18: Both the long and short 3' ends detected in the <i>L. ocellata</i> Dnmt3 transcript are present at the genomic level.....	58

Figure 19: 705 bp of <i>L. ocellata</i> gDNA amplified using primers #3 and #4.....	59
Figure 20: 514 bp of <i>L. ocellata</i> gDNA amplified using primers #3 and #4.....	60
Figure 21: Alignment of <i>L. ocellata</i> 3' region of cDNA and gDNA.....	61
Figure 22: Alignment of <i>L. ocellata</i> spliced 3' region of cDNA and gDNA.....	62
Figure 23: Alignment of <i>L. ocellata Dnmt3</i> 3' region with <i>Dnmt3</i> cDNA of other vertebrate models.....	64 – 68
Figure 24: <i>L. ocellata</i> Dnmt1 amino acid sequence is more similar to that of <i>D. rerio</i> dnmt1 than other representative vertebrate maintenance methyltransferases	76
Figure 25: <i>L. ocellata</i> Dnmt3 amino acid sequence is more similar to that of <i>D. rerio</i> dnmt4 than other representative vertebrate <i>de novo</i> methyltransferases.....	81
Figure 26: Schematic of <i>L. ocellata Dnmt3</i> transcripts illustrating the combinations of 5' and 3' splice variants.....	90

List of Abbreviations

aa.....	amino acid(s)
AdoMet.....	S-adenosyl L-methionine
BAH.....	Bromo-Adjacent Homology
BLAST.....	Basic Local Alignment Search Tool
bp.....	base pair(s)
CDD.....	Conserved Domain Database
cDNA.....	complementary DNA
CNS.....	Central Nervous System
DMAPI.....	DNA Methyltransferase I-Associated Protein
DNA.....	Deoxyribonucleic Acid
Dnmt.....	DNA-methyltransferase
ES.....	Embryonic Stem
EST.....	Expressed Sequence Tag
EtBr.....	Ethidium Bromide
gDNA.....	genomic DNA
HDAC.....	Histone Deacetylase
hpf.....	hours post fertilization
ICF.....	Immunodeficiency-centromeric instability-facial anomalies syndrome
ICM.....	Inner Cell Mass
MBDs.....	Methyl-CpG Binding Domain proteins

MeCP2.....	Methylated DNA Binding Protein 2
mRNA.....	messenger RNA
mya.....	million years ago
NCBI.....	National Center for Biotechnology Information
NLS.....	Nuclear Localization Signal
nt.....	nucleotide(s)
ORF.....	Open Reading Frame
ORC1.....	Origin Recognition Complex 1
PHD.....	Plant Homeo Domain
PWWP.....	Proline-Tryptophan-Tryptophan-Proline motif
RFD.....	Replication Foci Domain
RNA.....	Ribonucleic Acid
RT-PCR.....	Reverse Transcription – Polymerase Chain Reaction
TRD.....	Target Recognition Domain
UPM.....	Universal Primer Mix
UTR.....	Untranslated Region
UV.....	Ultraviolet

Introduction

While it is the DNA sequence of genes that dictate how cells are able to respond to and interact with their environment it is the higher order arrangement of chromatin that ultimately controls gene expression spatially and temporally. The complex packaging of the genome within eukaryotic nuclei can be classified as constitutive heterochromatin (centromeres, telomeres, satellite repeats and transposable elements), facultative heterochromatin (transcriptionally silenced DNA) and euchromatin (potentially transcriptionally active DNA). Constitutive heterochromatin remains consistent from cell to cell of a given species while the transcribed regions of the genome are revealed or concealed depending on the stage of development and fate of the cell. This dynamic nature is achieved through chromatin modification and remodelling.

Chromatin remodelling proteins shift nucleosomes and control the organization of chromatin fibres as well as the nuclear scaffold. These remodelling proteins work in concert with proteins that carry out covalent modifications of the histones around which the DNA is coiled as well as the DNA itself [as reviewed in Li 2002]. These modifications, or epigenetic markers (*epi*: over, above – *genetic*: of or relating to genes), exert a heritable influence over the genetic material that does not involve altering the base sequence. Such heritable genomic control can be seen in transposon silencing, gene imprinting (mono-allelic expression dependant on the parental origin of the gene) and mammalian X-chromosome silencing.

DNA methylation is an example of a covalent chromatin modification involving the attachment of a methyl-group to the fifth carbon position of cytosine residues. The modified base 5-methylcytosine (m^5C) can be found in the genome of all vertebrates and flowering plants as well as some fungi, invertebrates and bacterial species [Goll and Bestor 2005]. The mammalian genome, for example, contains $\sim 3 \times 10^7$ m^5C residues, most of which are incorporated into CpG dinucleotides [Bestor 2000]. CpG dinucleotides are most abundant throughout repetitive DNA elements such as pericentric DNA and telomeres as well as sequences harbouring transposons [Ma *et al.* 2005]. The modified m^5C nucleotides of CpGs can act as targets for methyl-CpG binding domain proteins (MBDs). MBDs selectively recognize methylated CpG dinucleotides and further recruit histone deacetylases (HDACs) which play an active role in chromatin remodelling [Jones *et al.* 1998]. Regions of the vertebrate genome exhibiting rich cytosine and guanine content (>55%) are known as CpG islands and range from 0.4 – 3 Kb in size. Such islands can be found associated with 76% of human gene promoter regions [as reviewed by Goll and Bestor 2005]. Though these islands, associated with tissue-specific and housekeeping genes, possess an abundance of methylatable targets, they typically remain unmethylated or lightly and variably methylated in all tissues. It is not yet fully understood if cytosine modification is inhibited or reversed at these island locations; however, the lack of m^5C does eliminate the potential for m^5C deamination to thymine resulting in a promoter region point mutation.

DNA methylation in depth

The nucleotide base cytosine, in its normal genetic conformation, base-pairs with guanine of the oppositely oriented nucleic acid strand. In such an arrangement, cytosine is engaged in three hydrogen bonds stemming from its 4-amino-2-oxo groups as well as a nitrogen atom bound to the third carbon atom of its pyrimidine structure. The enzymatic interaction necessary to covalently link a methyl-group to the fifth carbon atom, therefore, requires cytosine to disengage from its hydrogen bonds and rotate about its sugar-phosphate component so that it is extended out from the double helix conformation. In this everted state the cytosine base is readily accessible by methyltransferase enzymes. The different structures and functions of the DNA-methyltransferases (Dnmts) are described below but discussion of the conserved catalytic region is relevant when discussing the enzymatics of the reaction.

For the most part, the C-terminal end of the Dnmt contains the catalytic domain which is comprised of ten motifs (Dnmt3L, the exception, is outlined below) with the first (I) being the most proximal (closest to the N-terminal) and the tenth (X) being the most distal (closest to the C-terminal). Most of these motifs are detectable in the Dnmts of bacteria, fungi, plants and mammals while the six directly involved in the reaction are strongly conserved [reviewed by Bestor 2000]. Motifs VIII and IX flank the target recognition domain while IX plays a major role in maintaining the structure of that domain [reviewed by Goll and Bestor 2005]. As the Dnmt engages the cytosine base,

motif VI, which contains a glutamyl residue, protonates the nitrogen atom at position three of the cytosine pyrimidine ring (Figure 1). The double bond between the nitrogen atom and the carbon atom at the fourth position is shifted to between the carbon atoms at the fourth and fifth positions. At the same time motif IV (containing the prolylcysteiny dipeptide) binds the carbon atom at the sixth position of cytosine's pyrimidine structure breaking the 5-6 double bond allowing the carbon at the fifth position to maintain its four required covalent bonds. Most of the S-adenosyl L-methionine (AdoMet) binding site is formed by motifs I and X. AdoMet is the methyl-group donor. As the carbon atom at the fifth position binds the methyl-group the 4-5 double bond can no longer be maintained and shifts back to its original 3-4 position. In doing so the link between motif VI and the cytosine base is broken. However, the covalent link with motif IV remains. An unknown base or water molecule removes a proton from the carbon atom at the fifth position, freeing the cytosine base from motif IV and re-establishing the 5-6 double bond, completing the enzymatic reaction.

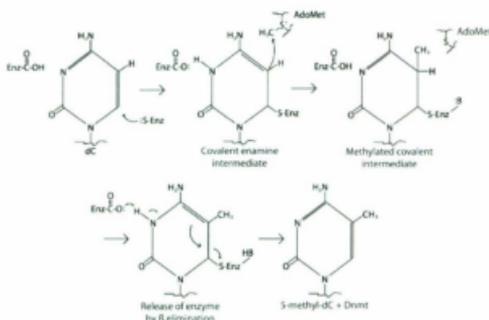


Figure 1: Catalytic mechanism of DNA (cytosine-5) methyltransferases. Adapted from Goll and Bestor 2005.

DNA-methyltransferase 1 (Dnmt1)

Dnmt1 has been shown to methylate hemimethylated DNA at CpG sites at a rate 5- to 30-fold greater than that of non-methylated DNA [Stein *et al.* 1982]. It is this preference for hemimethylated DNA that has resulted in Dnmt1 being termed the “maintenance” methyltransferase. Although there is variation in size the general structure of Dnmt1 remains consistent throughout vertebrate species. Dnmt1 from Friend murine erythroleukemia cells, which express high levels of Dnmt1 activity, was shown to encode a 1573 amino acid protein [Bestor *et al.* 1988]. The ~500 amino acid C-terminal possessed the characteristic ten catalytic motifs as described by Pósfai *et al.* [1989]. The remaining ~1000 amino acid N-terminal was connected via a region of alternating glycine and lysine residues.

The most distal domain present on the Dnmt1 N-terminal is the DNA methyltransferase 1-associated protein (DMAP1) binding domain [Marchler-Bauer *et al.* 2011]. DMAP1 has been implicated as a co-repressor supporting maintenance and activation of Dnmt1 preferentially at sites of homologous recombination repair [Lee *et al.* 2010]. Proximal to the DMAP1 binding domain is the replication foci domain (RFD) [Marchler-Bauer *et al.* 2011]. As the name implies it is responsible for targeting the Dnmt1 protein to replication foci during S phase DNA synthesis. This domain accounts for the dynamic localization documented by Leonhardt *et al.* [1992] in mammalian nuclei. Throughout G1 and G2 phases of the cell cycle Dnmt1 is diffuse throughout the nucleoplasm but associates with

replication foci during S phase. The targeting of Dnmt1 to replication foci adds further validity to the maintenance methyltransferase theory. Proximal still to the RFD is the zinc finger domain [Marchler-Bauer *et al.* 2011] containing eight conserved cysteine residues that bind two zinc ions. The zinc finger domain allows the Dnmt1 protein to bind nonmethyl-CpG dinucleotides and has been found in several mammalian proteins to be involved in chromatin and DNA modification [Frauer *et al.* 2011]. Between the zinc finger domain and the catalytic motifs of the C-terminus are two Bromo Adjacent Homology (BAH) domains [Marchler-Bauer *et al.* 2011]. BAH domains mediate protein-protein interactions and have been shown to be present in human origin recognition complex 1 (ORC1) where they promote chromatin association [Noguchi *et al.* 2006].

Mouse Dnmt1 isolated from somatic tissue is produced from a mammalian somatic cell-specific promoter which drives expression of a transcript including a somatic cell-specific exon (exon 1s) shortly after implantation [Tucker *et al.* 1996, Yoder *et al.* 1996]. Resulting protein products are only detected in nuclei. An additional promoter lies upstream to that utilized in somatic cells. This second promoter drives expression of an oocyte-specific exon (exon 1o) resulting in translation initiation in exon 4. The oocyte-specific variation of Dnmt1 lacks the N-terminal 118 amino acids of the somatic form but retains the functional nuclear localization signal (NLS) essential for nuclear proteins [Mertineit *et al.* 1998]. Though an NLS is present the Dnmt1o localizes to the cytoplasm of oocytes as well as early embryos and only moves into the nucleus during S phase of the 8 cell stage of development (discussed further below). A third Dnmt1 transcript

variant specific to pachytene spermatocytes was also detected although no protein product was found, possibly due to multiple upstream open reading frames (ORFs) strongly inhibiting translation.

DNA-methyltransferase 3 (Dnmt3) family

Mouse embryonic stem (ES) cells homozygous for null Dnmt1 mutation displayed stable, residual methyl cytosine levels and retained their ability to methylate integrated provirus DNA [Lei *et al.* 1996] providing the first evidence that a second DNA methyltransferase existed independently from Dnmt1. The new methyltransferase was dubbed Dnmt3 as the protein from mice showed little similarity to the previously described Dnmt1 or Dnmt2 (discussed below) [Okano *et al.* 1998]. There were two distinct variations of *Dnmt3* (3a and 3b) found in mice and humans by way of a tBLASTn search of the dbEST database using full-length bacterial type II cytosine-5 methyltransferase sequences as queries. These newly discovered methyltransferases were highly expressed in ES cells, early embryos and developing germ cells corresponding to observed *de novo* methylation patterns. Furthermore, these Dnmts showed no preference for hemi-methylated DNA over non-methylated DNA [Okano *et al.* 1998] lending credence to their role as *de novo* methyltransferases.

The mouse *Dnmt3a* gene encodes at least two protein products, 3a and 3a2, consisting of 908 and 689 amino acids respectively. Dnmt3a2 lacks the N-terminal region of Dnmt3a

as it is encoded by transcripts derived from a downstream promoter. Both variants are enzymatically active though they demonstrate different nuclear localizations. Dnmt3a2 is diffuse throughout nuclei whereas Dnmt3a concentrates at heterochromatin [Chen *et al.* 2002]. At least five human isoforms exist for *DNMT3b* by way of alternative splicing of exons 10, 21, and/or 22 (as reviewed by Chen and Li 2004). These isoforms exhibit different tissue distribution with *DNMT3b1* and *3b6* being the predominant forms in ES cells, *DNMT3b4* and *Dnmt3b5* predominantly expressed in testis, and *DNMT3b2* and *3b3* expressed at relatively high levels in testes, ovary, spleen, thymus and liver [Chen *et al.* 2002, Robertson *et al.* 1999, as reviewed by Chen and Li 2004]. Only DNMT3b1 and 3b2 display enzymatic activity [Aoki *et al.* 2001] as the rest lack the amino acid sequence present in motifs IX and X involved in maintaining the target recognition region between motifs VIII and IX.

The zebrafish (*Danio rerio*) shows evidence of six dnmt3s, twice as many as mammals. They were labelled as dnmts 3 through 8 by Shimoda *et al.* [2005] after using a partially determined peptide sequence of zebrafish dnmt3 as the query of a tBLASTx search of the zebrafish genome database. Smith *et al.* [2005] showed *dnmt5* to have at least three transcript variants with splicing variations occurring upstream to the translation start site suggesting a regulatory role in translation or localization.

Regardless of the model organism in question the general structure remains consistent for Dnmt3 as it does for Dnmt1. The C-terminal maintains the ten catalytic motifs required

for enzymatic activity as described by Pósfai *et al.* [1989]. The N-terminal harbours a PWWP domain named after a conserved Pro-Trp-Trp-Pro motif [Marchler-Bauer *et al.* 2011]. The PWWP domain has been demonstrated as being a recognition motif for methylated lysine 20 on histone 4 (H4K20me) [Wang *et al.* 2010]. The only deviations from this conserved structure are zebrafish dnmt3 and dnmt7 [Shimoda *et al.* 2005] which have a calponin-homology (CH) domain near their distal N-terminal capable of binding microtubules; this has not been seen in any other known methyltransferase.

DNA-methyltransferase 3-like (Dnmt3L) is structurally related to both Dnmt3a and Dnmt3b. Dnmt3L possesses a similar C-terminal catalytic-like region though it lacks the strong conservation of catalytic motifs and does not exhibit any enzymatic activity [Aapola *et al.* 2000]. Dnmt3L has been shown to play a major role in the maternal imprinting process of oocyte development and interacts with Dnmt3a and Dnmt3b in a manner that is not yet fully understood in female mammalian cells but appears to provide sequence specificity to DNA methylation [Hata *et al.* 2002]. In mammalian males Dnmt3L has been linked to the *de novo* methylation of dispersed repeated sequences in spermatogonia [Bourc'his and Bestor 2004]. Dnmt3L is catalytically inactive but remains functional in DNA methylation with an N-terminal region that recognizes histone H3 tails that are unmethylated at lysine 4 and acts to recruit/activate Dnmt3a [Ooi *et al.* 2007]. Additionally, Dnmt3L possess a plant homeodomain (PHD)-like region in its N-terminus which has been shown to recruit histone deacetylases (HDACs) [Deplus *et al.* 2002] which may also contribute to chromatin remodelling.

DNA-methyltransferase 2 (Dnmt2)

Dnmt2 is the most widely distributed methyltransferase. Organisms expressing both *Dnmt1* and *Dnmt3* always express *Dnmt2* and there are a number of species that solely express *Dnmt2* [reviewed by Goll and Bestor 2005]. Dnmt2 lacks the N-terminal extensions seen in other Dnmts [Yoder and Bestor 1998]. The catalytic motifs of the C-terminal end are well conserved, including the cytosine targeting region between motifs VIII and IX., yet Dnmt2 shows no activity towards either hemimethylated or unmethylated DNA [Okano *et al.* 1998]. Further analysis involving embryonic stem cell disruption of Dnmt2 expression revealed no methylation abnormalities and the development of normal phenotypes. While Dnmt2 is not involved in DNA methylation it was shown to methylate tRNA (Goll *et al.* 2006) in mice, *Arabidopsis thaliana* and *Drosophila melanogaster*. More recent experiments using *D. melanogaster* (Schaefer *et al.* 2010) have shown Dnmt2-mediated tRNA methylation to be associated with a reduction in stress-induced ribonuclease cleavage.

DNA methylation and development

The dynamic nature of chromatin modification is clearly demonstrated throughout early development of the mouse model. Initially the genomic DNA of the sperm is heavily methylated relative to the oocyte [Olek and Walter 1997]. Immediately following fertilization the sperm pronucleus is actively demethylated while protamines are replaced

by histones. In contrast, the maternal genome undergoes passive demethylation [Santos *et al.* 2002]. The oocyte specific variant Dnmt1 α is sequestered in the cytoplasm of embryonic cells during the first two cell divisions preventing the maintenance of DNA methyl patterns as DNA is semi-conservatively replicated. At the 8 cell stage Dnmt1 α moves into the nucleus to preserve imprints [Howell *et al.* 2001] and is then detected solely in the cytoplasm again at the 16 cell stage [Santos *et al.* 2002]. This depressed methylated state persists until the morula stage. *De novo* methylation begins by the blastocyst stage but only targets the genomes of cells making up the inner cell mass (ICM). Genomes of trophoctoderm cells remain relatively hypomethylated.

Initially, general trends of DNA methylation patterns observed in early mouse development were not detected in early zebrafish development. It was reported that methylation levels were not reduced in zebrafish pre- and early blastula stages compared with levels detected at the gastrula stage and in adult somatic tissue [Macleod *et al.* 1999]. The investigators postulated that the reduction in DNA methylation levels recorded in mammals may have been associated with the maintenance of imprinted genes during embryonic methylation programming. The lack of imprinted genes in the zebrafish [Streisinger *et al.* 1981, Corley-Smith *et al.* 1996] perhaps did not require a reduction and re-establishment of overall DNA methylation levels. Subsequent investigations have provided contradictory evidence to these initial reports. Relative changes in zebrafish *dnmt1* mRNA and enzyme activity during oogenesis and early development closely resemble changes observed in oogenesis and pre-implantation embryos of the mouse

[Mhanni and McGowan 2002]. Zebrafish sperm DNA is heavily methylated relative to that of the oocyte [Mhanni and McGowan 2004]. Following zebrafish fertilization substantial demethylation occurs and persists up to early blastula stages. *De novo* methylation initiates and remethylates the genome as the embryo enters later blastula stages coinciding with activation of the zygotic genome and differentiation of embryonic and extraembryonic cell lineages [MacKay *et al.* 2007]. The discrepancy between the Macleod *et al.* (1999) and following studies may in part be explained by timing. The earliest time point analysed by *HpaII* and *MspII* restriction digestion based experiments in the early report indicated a consistent level of DNA methylation in sperm, adult somatic tissue and early blastula embryos at 2.2 hours post fertilization (hpf). However the latter investigation claimed that methylation levels at the 2.2 hpf stage of development had already undergone the demethylation process and were undergoing the remethylation that would establish levels comparable with those seen in the more heavily methylated sperm and adult somatic tissues. Immunohistochemistry techniques utilized in the later experiments confirmed these findings revealing the early demethylation and subsequent remethylation of zebrafish embryos. The Macleod study also relied heavily on bisulphite sequencing data collected from small regions of only three genes randomly chosen from the GenBank database. Although data exists for and against the reduction of DNA methylation levels in early zebrafish development this present investigation operates on the assumption that three randomly chosen genes maintaining a consistent state of methylation throughout development does not reflect trends affecting the entire genome. The three genes reported by MacLeod *et al.* (1996) need not undergo

demethylation as a genome wide demethylation event is not necessary for the reduction in methylation levels observed by MacKay *et al.* (2007) or the reduction in methylation levels observed in the mouse model. Given that the mammalian model undergoes gene imprinting and the zebrafish does not suggests that a similarity in dynamic DNA methylation in both models is intrinsic for vertebrate development.

More specifically, each Dnmt plays a key role in development and differentiation. Mice homozygous for mutant *Dnmt1* failed to develop beyond the normal day 9.5 embryo and died prior to day 11. Brain, heart, and in some cases forelimb buds were all present but less developed than would be expected given the gestational age. Cell death was widespread throughout the embryo as was reduced cell proliferation and a 3-fold reduction in genomic cytosine methylation levels was documented [Li *et al.* 1992]. Morpholino disruption of *Dnmt1* transcript translation in zebrafish caused defects in terminal differentiation of intestine, retina and exocrine pancreas. The liver and endocrine pancreas developed normally but degenerated at 84 hours post fertilization (hpf). The same embryos exhibited a dramatic reduction in genomic cytosine methylation as well as genome-wide histone 3 lysine 9 (H3K9) trimethyl levels [Rai *et al.* 2006, Anderson *et al.* 2009]. Inactivation of *Dnmt3a* and *Dnmt3b* by gene targeting blocks *de novo* methylation in mouse ES cells and early embryos but has no effect on maintenance of imprinted methylation patterns [Okano *et al.* 1999]. It was determined that Dnmt3b was responsible for the methylation of centromeric minor satellite repeats and that mice deficient in Dnmt3b activity developed normally until embryonic day 9.5 but died before birth. Mice

deficient in Dnmt3a activity survived to term but became undersized and died by four weeks of age. Embryos lacking both Dnmt3a/3b activity showed a more severe phenotype similar to that observed for mutant Dnmt1 mice. Zebrafish development illustrates a tissue-specific function for Dnmt3 as it is required for proper differentiation of neurons, pharyngeal arches, the exocrine pancreas, and certain retinal tissues (Rai *et al.* 2010). It is not required for development of the jaw, intestine, endocrine pancreas or liver.

DNA methylation and human pathology

Immunodeficiency-centromeric instability-facial anomalies syndrome (ICF) is a rare autosomal recessive syndrome involving C-terminal mutations in the *DNMT3b* gene [Xu *et al.* 1999]. Distinctive signs of ICF include a severe immunodeficiency and instability of the pericentromeric heterochromatin at chromosomes 1 and 16 and less frequently at chromosome 9. Variable signs include hypertelorism, flat nasal bridge and macroglossia, psychomotor and mental retardation as well as intestinal dysfunction and developmental delays [as reviewed by Matarazzo *et al.* 2009]. The mutation within the catalytic region of the *DNMT3b* gene leads to complete demethylation of classical satellite sequences as well as demethylation of CpG islands on inactive X chromosomes in females [Xu *et al.* 1999]. Though silenced X chromosomes are affected there does not seem to be a difference in symptoms between male and female patients [Boure'his *et al.* 1999].

Rett syndrome is a neurological disorder almost exclusively affecting females and is linked to mutations in the *methylated DNA binding protein 2 (MeCP2)* gene [Amir *et al.* 1999]. It is thought that complete loss of the X-linked *MeCP2* gene in males would be lethal. Rett syndrome is characterized by an apparently normal pre- and perinatal period followed by a skill regression in early childhood (prior to age 2). There is loss of already acquired skills, regression of language, presence of repetitive motions (especially hand movements), seizures and mental retardation [reviewed by Matarazzo *et al.* 2009, Geiman and Muegge 2010]. MeCP2 is not a methyltransferase but a protein that binds methylated DNA and takes part in inducing nucleosome clustering and stabilization of large chromatin loops [Ghosh *et al.* 2010]. Although aberrations in m^5C are not the root of Rett syndrome the consequences of being unable to correctly respond to a particular epigenetic marker cannot be denied.

Aberrant DNA methylation is one of the most consistent epigenetic changes observed in human cancers [reviewed by Delcuve *et al.* 2009] and has the potential to influence tumourigenesis in three ways. First, methylated cytosine may alter coding regions of genes by inducing point mutations. Second, the overall depletion of m^5C in the genome may cause genome instability. Third, hypermethylation of gene promoters may inactivate gene transcription [reviewed by Gronbaek *et al.* 2007]. Hypermethylation of gene promoters sets the scene for potentially disastrous point mutations as methylated cytosines are deaminated to thymine. $G:C \rightarrow A:T$ has been shown to be the most common mutation to the *p53* gene in human colon and breast tumours as well as lung and

esophageal cancer [Hollstein *et al.* 1991]. p53 has been termed the guardian of the genome as it is one of the most important tumour suppressors mediating growth and apoptosis in response to oncogenic cellular stresses such as DNA damage [reviewed by Chen *et al.* 2011]. Hypomethylation of DNA causes chromatin decondensation and chromosomal rearrangements potentially resulting in chromosomal instability [reviewed by Kanai 2008]. Feinberg and Vogelstein [1983] showed a substantial degree of hypomethylation in cancer cell genes of a small group of patients with adenocarcinoma of the colon and small cell carcinoma of the lung when compared to the genes' normal counterparts in control cell types.

DNA methylation: the evolutionary query

DNA methylation as an epigenetic marker plays a major role in proper chromatin management throughout development. The integral role of DNA methylation in vertebrate development is illustrated by the severity of developmental and pathological aberrations documented in both the mammalian (mouse) and teleost (zebrafish) experimental models resulting from deviations in the normal methylation pattern. The mouse and zebrafish represent the tetrapod and teleost lineages respectively, two lineages that diverged from each other ~440 million years ago (mya) [Santini *et al.* 2009]. Though separated by millions of years of evolution these two groups use very similar proteins to establish very similar developmental programs in DNA methylation throughout early development. There are, however, lineage-specific differences in methyltransferase

numbers as well as how these methyltransferases are utilized with regard to each model's different reproductive strategy. The importance of DNA methylation cannot be denied but it is difficult to ascertain the targeting mechanism of DNA methyltransferases as well as how they are controlled when information bases we currently have are limited to so few species separated by over 400 million years of independent evolution.

In order to elucidate the core roles of vertebrate DNA methylation it would be beneficial to investigate an organism or group of organisms representative of the ancestral vertebrate condition. Class Chondrichthyes (the cartilaginous fish) diverged from bony fish ~528 +/- 56.4 mya [Kumar and Hedges 1998] and further split into two subclasses, Elasmobranchii (sharks, skates and rays) and Holocephali (chimeras) ~374 mya [Nelson 2006]. The Elasmobranchs exhibit fundamental vertebrate characteristics such as a developmental neural crest, jaws and teeth, a pressurized circulatory system and adaptive immune system [Mattingly *et al.* 2004]. Members of this subclass have previously been utilized as ancestral vertebrates before in numerous comparative physiological studies concerning endocrine systems and organ function [Bewley *et al.* 2006, Cai *et al.* 2001]. Furthermore, some species already being used in laboratories (such as *Leucoraja erinacea*) as representative ancestral vertebrate models have genome sizes comparable to our own [Gregory 2011]. This is a great advantage for investigating genome management while attempting to shed light on human-centric developmental abnormalities or pathologies. Subclass Elasmobranchii also displays a wide variety of reproductive strategies ranging from oviparity (single and multiple) to viviparity (yolk-sac, histotrophy

and oophagy) [as reviewed by Musick and Ellis 2005]. Studying genome management in this subclass could provide great insight into the key differences and similarities in DNA methylation observed in previously studied tetrapod and teleost models.

Objectives of Thesis

The animal model presented in this work is the Winter Skate (*Leucoraja ocellata*) belonging to Class Chondrichthyes, Order Rajiformes and Family Rajidae. *L. ocellata* is an oviparous skate found in the western Atlantic Ocean off the coasts of Newfoundland and Labrador and as far south as North Carolina [Bester 2011]. It belongs to the same genus as the currently studied Little Skate (*L. erinacea*). Though the Winter Skate was chosen for this study due to its convenient habitat location it also serves as a further expansion of the *Leucoraja* genus as a scientific model. Previously in my undergraduate program I established that *L. ocellata* methylates its genome. I isolated a 274 bp *L. ocellata Dnmt3* 3' sequence by reverse transcription – polymerase chain reaction (RT-PCR) using degenerate primers based on ClustalW [Larkin *et al.* 2007] alignment of *M. musculus*, *H. sapiens* and *D. rerio de novo* DNA methyltransferases. The present study expands upon that work by investigating DNA methylation levels in testis and ovary tissues of *L. ocellata*. I also report and characterize the first full length Chondrichthyes DNA methyltransferase cDNA sequences corresponding to the *L. ocellata Dnmt1* and *Dnmt3*.

Materials and Methods

DNA and RNA samples

Ovary, testis and muscle tissue from *L. ocellata* specimens were previously collected by Dr. Ross McGowan from animals maintained at the Ocean Sciences Centre (Logy Bay, NL, Canada) and stored at -80°C.

Genomic DNA was extracted from tissue samples using chloroform/ phenol methodology [Sambrook *et al.* 1989]. Total RNA had been previously extracted from tissue samples by Dr. Ross McGowan using an acid guanidinium thiocyanate-phenol-chloroform extraction [Chomezynski and Sacchi 1987]. Additional testes, brain and gut total RNA from *L. ocellata* specimens maintained at the Ocean Sciences Centre were provided by Dr. H el ene Volkoff.

Comparison of testis, ovary and somatic tissue methylation levels

Digestion of gDNA from *L. ocellata* testis, ovary and somatic tissue was performed using the restriction enzyme *HpaII* (Promega, Madison, WI, USA). *HpaII* is a 4 bp cutter targeting the sequence 5' - CCGG - 3' but is inhibited from cutting the target sequence when the interior cytosine is methylated.

500 ng of *L. ocellata* gDNA from each tissue type were separated by gel electrophoresis in 0.8% agarose providing a means of assessing the initial integrity of each gDNA sample. 2.5 µg of gDNA from each tissue type were then digested by *HpaII* as per the manufacturer's protocol (Promega). Comparison of untreated gDNA samples to samples digested by *HpaII* (Promega) provided a means of assessing the methylation level specific to the endonuclease target sequence of each gDNA sample. The contents of the restriction digestions were separated by gel electrophoresis in 0.8% agarose, stained with ethidium bromide and visualized under ultraviolet light. Complete digestion was verified by the addition of control plasmid DNA to each sample of the experimental digests. Digestion was considered to be complete when the internal control gave a digestion pattern identical to that obtained with test plasmid and *HpaII* alone.

Primer design and synthesis

Primer design was carried out using the on-line Primer3 software (<http://frodo.wi.mit.edu/primer3/>) [Rozen and Skaletsky 2000]. All primers were synthesized by Invitrogen Inc. (Carlsbad, CA, USA) and used at an experimental concentration of 10 µM unless otherwise stated (see Table 1 for primer list and sequences).

Table 1: Primers used throughout the isolation and characterization of *Leucoraja ocellata Dnmt3* and *Dnmt1*. Refer to cDNA schematics for primer locations and orientations. The degenerate sequence code used by Invitrogen Inc. is as follows. *B*: T/C/G, *D*: A/T/G, *H*: A/T/C, *K*: T/G, *M*: A/C, *N*: A/C/G/T, *R*: A/G, *S*: C/G, *V*: A/C/G and *Y*: C/T

Leucoraja ocellata Dnmt3

Primer	Purpose	Sequence
1	3' RACE RT-PCR	GCAGCCCCGTGAACGATCT
2	RT First Strand Synthesis	ACCCATGGCAACAACATTCT
3	3' Splice Variant Investigation	TGAGGGCACAGGAAGACTTT
4	3' Splice Variant Investigation	GCATGCAGCCTTCACAATAA
5	5' RACE RT-PCR	AGGCCTGTCTCCCACTCCTTTGGT
6	Isolation of Long 3' Splice Variant	ATCCTTGCCCTGCTTGATGGA
7	Isolation of Short 3' Splice Variant	GCAAAGTAGAGAGATTCTGACA
8	Isolation of Long 5' Splice Variant	AGCCAAAGATCCTGTTCAAT
9	Isolation of Short 5' Splice Variant	AGCCAAAGATCCTGTTCAAT
10	RT First Strand Synthesis	CTCCCCATGTTTCGACACGCTCTGTGT
11 (M13F)	Sequencing	GTA AAAACGACGGCCAG
12	Sequencing	TACATGTAGGAAGTCTCCAG
13	Sequencing	AGAAATCAGCTTGTAAATG
14	Sequencing	CTGGAAAATTATACGTGGT
15	Sequencing	TGCAGCAGTTTGCAATCCCA
3' Degenerate (not shown in Figure 2)	Initial Isolation of <i>Dnmt3</i> cDNA	SATTGGDDGMAGYCCHTYAA
5' Degenerate (not shown in Figure 2)	Initial Isolation of <i>Dnmt3</i> cDNA	CCARAARTANCKDGCYCTGTG

Leucoraja ocellata Dnmt1

Primer	Purpose	Sequence
A	5' RACE RT-PCR	TCCTTTAGCTCCAGTGGCTCCAGCA
B	Sequencing	ACCCAAGCCCCAAACTTCA
C	Sequencing	GCTGACTGCCACACAATCAT
D	Sequencing	TAGTGGTAGTGGCTTTAGTG
E	Sequencing	TGGGTTATCATCATAGATTG
3' Degenerate (not shown in Figure 4)	Initial Isolation of <i>Dnmt1</i> cDNA	TGGGCHATHGARATGTGGGA
5' Degenerate (not shown in Figure 4)	Initial Isolation of <i>Dnmt1</i> cDNA	GGHACRCRTTNCMACCTG

Sequencing reactions and data analysis

Sequencing of all clones obtained throughout this investigation was carried out at the Genomics and Proteomics facility of Memorial University of Newfoundland following laboratory protocols and using the Applied Biosystems 3130 four capillary Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Sequencing data was assessed using the free on-line software CHROMAS (<http://www.softpedia.com/get/Science-CAD/Chromas-Lite.shtml>) (School of Health Sciences, Southport, Queensland, Australia). Reliable sequence data was further trimmed and analyzed using the free on-line software Gene Runner version 3.05 (<http://www.generunner.net/>).

*Isolation of the *L. ocellata Dnmt1* cDNA 5' end*

Using the *L. ocellata Dnmt1* 3' cDNA previously isolated by the McGowan Laboratory, a primer was designed (Primer A, see Figure 2 for primer location and orientation) having a compatible melting temperature with the Universal Primer Mix (UPM) used by the BD SMARTer RACE cDNA Amplification Kit (Clontech, Palo Alto, CA, USA). The PCR program recommended by the manufacturer was used; it started with a 2 min 94°C denaturation followed by thirty cycles consisting of 5 cycles with a 30 sec 94°C denaturation and a 3 min 72°C elongation step, 5 cycles with a 30 sec 94°C denaturation, a 30 sec 70°C annealing step and a 3 min 72°C elongation step, and 20 cycles with a 30 sec 94°C denaturation, a 30 sec 68°C annealing step and a 3 min 72°C elongation step. A



Figure 2: Schematic of *L. ocellata Dnmt1* cDNA. (A) 3' sequence acquired by the McGowan laboratory is indicated in orange. Untranslated regions are shown in grey. Primer locations and orientations are indicated by arrows. (B) Locations of primers corresponding to the labeled schematic. See Table 1 for primer sequences and purposes.

final 1 min elongation step at 72°C plus a holding step at 10°C followed. All PCR amplification reactions performed throughout this investigation were carried out using an Eppendorf Mastercycler ep Gradient (Eppendorf Inc., Mississauga, ON, Canada). Resulting amplicons were separated by gel electrophoresis in 0.8% agarose. All electrophoresis gels ran 1 Kb Plus DNA Ladders (Invitrogen Inc.) as marker lanes. Gels were stained with ethidium bromide and visualized under ultraviolet light. PCR products were ligated into the TOPO TA Cloning Vector (Invitrogen Inc.) as per the kit protocol. Vectors containing PCR amplicons were used to transform Mach T1 chemically competent *E. coli* cells (Invitrogen Inc.). Transformed cells were incubated overnight at 37°C on L-broth agar containing ampicilin. Resulting colonies were used to inoculate L-broth liquid cultures and incubated overnight in a 37°C water bath set to shake at 200 rpm. Plasmid DNA was recovered from cells using the Wizard[®] Plus Minipreps DNA Purification System (Promega) and the presence of desired PCR products was confirmed by *Eco*RI (Promega) digestion and gel electrophoresis in 0.8% agarose.

Sequencing of cDNA was carried out as outlined above. Further characterization of cDNA utilized the following on-line databases: Conserved Domain Database (CDD) [Marchler-Bauer *et al.* 2011], nucleotide and protein sequence databases (National Center for Biotechnology Information) and skate EST database (http://decypher.mdibl.org/decypher/algo-tera-blast/tera-blastn_nn.shtml). Additional characterization took advantage of the following free on-line software: ClustalW [Larkin *et al.* 2007], Mobyle@Pasteur v1.0 (<http://mobyle.pasteur.fr/cgi-bin/portal.py>), SWISS-

MODEL Workspace [Arnold *et al.* 2006], WoLF PSORT [Horton *et al.* 2007], and in the case of *L. ocellata Dnmt3* (discussed below), Scion Image software (National Institute of Health, Bethesda, MD, USA).

Preliminary research setting the stage for this Dnmt3 investigation

Previously, I obtained 274 bps of *Leucoraja ocellata*'s *Dnmt3* transcript within the 3' catalytic region (Figure 3) [Lake 2008]. The sequence was isolated using RT-PCR with degenerate primers designed based on a ClustalW [Larkin *et al.* 2007] alignment of *Homo sapiens DNMT3A* (GenBank number [AB208833](#)) and *3B* (GenBank number [NM_006892.3](#)), *Mus musculus Dnmt3A* (GenBank number [AF068625](#)) and *3B* (GenBank number [AF068626](#)) as well as *Danio rerio dnmts 3* through *8* (GenBank numbers [AB196914](#), [AB196915](#), [AB196916](#), [AB196917](#), [AB196918](#) and [AB196919](#) respectively). It was from this 274 bp sequence that species specific primers were designed for the purpose of determining the remaining sequence of both the 3' and 5' ends of the *L. ocellata Dnmt3* cDNA.

Isolation of the complete 3' end of L. ocellata Dnmt3 cDNA

First strand cDNA was synthesized using the BD SMART RACE cDNA Amplification Kit and protocol (Clontech). Template RNA was derived from pooled *L. ocellata* oocytes (0.75 – 1.0 cm in diameter). The initial reverse transcription reaction consisting of a 1.5

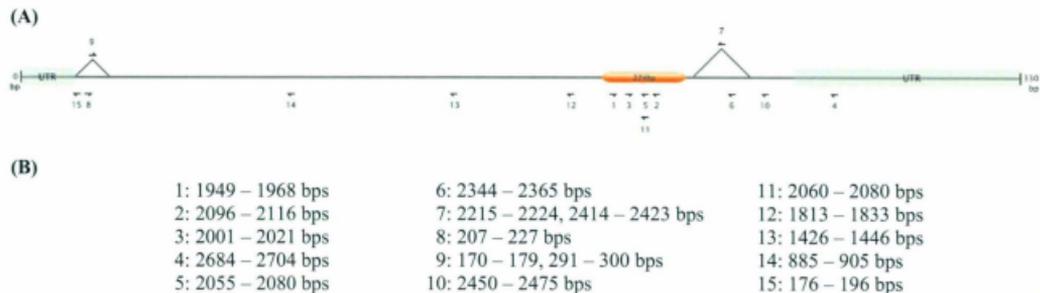


Figure 3: Schematic of *L. ocellata Dnmt3* cDNA. (A) The 274 bps previously isolated [Lake 2008] are indicated in orange. Untranslated regions are shown in grey. Areas of sequence subjected to splicing are demarcated by triangles. Primer locations and orientations are indicated by arrows. (B) Locations of primers corresponding to the labeled schematic. See Table 1 for primer sequences.

hour 42°C incubation followed by 7 minutes at 72°C was performed using an Eppendorf Mastercycler ep Gradient (Eppendorf Inc.).

PCR primer #1 was designed to amplify the remaining 3' end of the *L. ocellata Dnmt3* transcript. The position of primer #1 within the original 274 bp sequence is shown in Figure 3. The PCR program started with a 2 min 94°C denaturation followed by thirty cycles consisting of a 30 sec 94°C denaturation, a 30 sec annealing step using a block gradient (55°C – 60°C) and a 30 sec elongation step at 72°C. A final 1 min elongation step at 72°C plus a holding step at 10°C followed. PCR products were cloned and sequenced as previously described.

Sequencing analysis and tissue distribution of the L. ocellata Dnmt3 3' variants

Gel electrophoresis of the *Dnmt3* 3' clones recovered via the above mentioned Miniprep system indicated two variants (Figure 4A). ClustalW [Larkin *et al.* 2007] alignment of the two variant sequences indicated that one lacked an internal region comprising 189 bps. Primers #3 and #4 were designed to flank the variable 3' region (Figure 3) and were utilized in investigating the tissue distribution of the two *Dnmt3s* containing variable 3' catalytic regions. Tissue types tested in this manner were *L. ocellata* ovary, testis, brain and gut. First strand cDNA synthesis reactions were carried out using the AncT primer (Invitrogen Inc.). PCR amplifications utilizing primers #3 and #4 started with a 2 min 94°C denaturation followed by thirty cycles consisting of a 30 sec 94°C denaturation, a

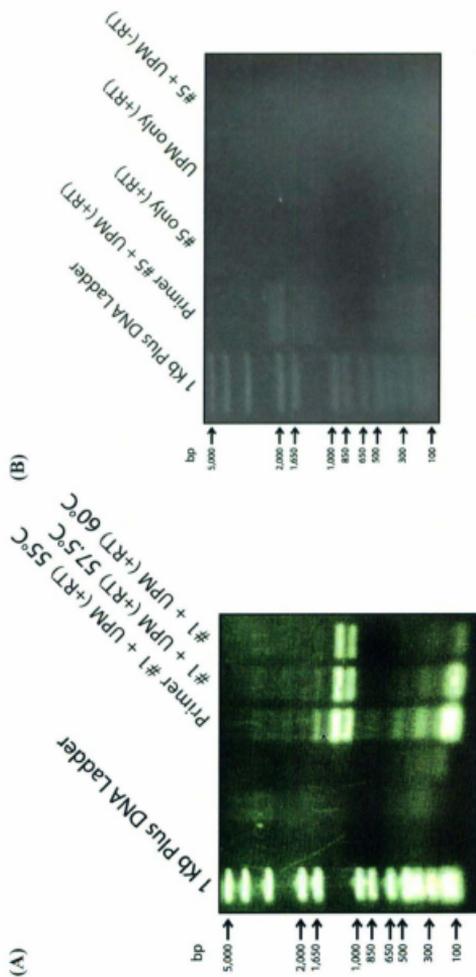


Figure 4: SMARTer RACE RT-PCR products of the 3' and 5' ends of the *L. ocellata Dmmt5*. (A) Separation by gel electrophoresis in 0.8% agarose of the ~1000 bp and ~1200 bp products of the 3' reaction. Amplicons were generated in three different reactions using different primer annealing temperatures. (B) Gel electrophoresis separation in 0.8% agarose of the ~2000 bp and ~2200 bp products of the 5' reaction. 1Kb Plus DNA ladder was used in both gels and is labeled to the left of each image (Invitrogen Inc.).

30 sec 65°C annealing step and a 1 min elongation step at 72°C. A final 1 min elongation step at 72°C plus a holding step at 10°C followed. PCR products were separated via gel electrophoresis. Comparison of transcript levels of each of the two *Dnmt3* 3' variant amplicons was conducted within each lane of the electrophoresis gel using Scion Image software (National Institute of Health).

Isolation of the L. ocellata Dnmt3 cDNA 5' end

Primers #2 and #5 were designed to isolate the 5' end of the Winter Skate *Dnmt3* transcript. Primer #2 was used with BD Bioscience's SMARTer RACE cDNA Amplification Kit (Clontech) to generate first strand cDNA. Subsequent PCR was performed using Primer #5 and the UPM provided by the SMARTer RACE kit (Clontech). The PCR program recommended by the manufacturer was used; it started with a 2 min 94°C denaturation followed by thirty cycles consisting of 5 cycles with a 30 sec 94°C denaturation and a 3 min 72°C elongation step, 5 cycles with a 30 sec 94°C denaturation, a 30 sec 70°C annealing step and a 3 min 72°C elongation step, and 20 cycles with a 30 sec 94°C denaturation, a 30 sec 68°C annealing step and a 3 min 72°C elongation step. A final 1 min elongation step at 72°C plus a holding step at 10°C followed.

PCR amplicons resulting from the above reaction were cloned, isolated and sequenced as previously described. Gel electrophoresis of the *Dnmt3* 5' clones recovered via the

Miniprep system indicated two variants (Figure 4B). ClustalW [Larkin *et al.* 2007] alignment of the two variant sequences indicated one lacked an internal region comprising 111 bps.

Matching Dnmt3 5' variants to 3' variants

Two sets of discriminating primers were designed in order to match the variable 5' ends to the variable 3' ends (Primer #s 6, 7/8, 9). Primer #10 was designed to be used in conjunction with BD Bioscience's RT kit (Clontech) for first strand cDNA synthesis, the product of which was used as template in amplification reactions pairing primer #6 with primers #8 and #9 as well as reactions pairing primer #7 with primers #8 and #9 (see Figure 3 for positions and orientations). The parameters for all six amplification reactions were the same. The PCR program started with a 2 min 94°C denaturation followed by thirty cycles consisting of a 30 sec 94°C denaturation, a 30 sec 50°C annealing step and a 3 min elongation step at 72°C. A final 1 min elongation step at 72°C plus a holding step at 10°C followed.

Investigation of potential Dnmt3 pseudogenes

While investigating the 3' variants of the *L. ocellata Dnmt3* cDNA by way of RT-PCR using primers #3 and #4 results similar to those found in the experimental reactions were observed in the control PCR reactions using the "no-RT" reaction (reverse transcription

reaction lacking a reverse transcriptase) as template. Fresh primer aliquots were prepared and fresh enzymes were used in an effort to eliminate any potential contamination to reaction reagents. The “no-RT” PCR results resembling the experimental results persisted. A new experiment was designed to investigate the possibility that the results were due to genomic DNA (gDNA) carryover in the RNA extraction. 1 µg of gDNA extracted as outlined above from *L. ocellata* testis underwent RNase A (Qiagen Inc., Mississauga, ON, Canada) treatment as per product protocol. One half of the 20 µl RNase treated gDNA volume was then removed and subjected to DNA Wipeout (Qiagen Inc.) as per the product protocol using a 9 min 42°C incubation period. Both the RNase treated gDNA and RNase/DNA Wipeout treated gDNA samples were used as template for PCR using primers #3 and #4. The PCR program started with a 2 min 94°C denaturation followed by thirty cycles consisting of a 30 sec 94°C denaturation, a 30 sec 65°C annealing step and a 1 min elongation step at 72°C. A final 1 min elongation step at 72°C plus a holding step at 10°C followed.

Results

Higher levels of DNA methylation in testis relative to ovary and somatic tissue

Research conducted during my B.Sc. (Honours) dissertation project [Lake 2008] using restriction enzymes demonstrated that the *L. ocellata* genome was methylated. However, only somatic tissues were examined in that study. Since differences in relative DNA methylation levels exist between sperm and oocytes in both mammals and teleosts an investigation into whether or not the same trend was evident in *L. ocellata* was undertaken. It should be noted that Winter Skate DNA methylation levels were assessed using DNA isolated from testis and ovary tissues rather than sperm or oocyte populations. As a result methylation levels of reproductive cells may have been partially masked by methylation levels of surrounding reproductive tissue cells. Figure 5 shows the results of a *HpaII* restriction enzyme digestion. Although a direct quantitative comparison cannot reliably be made between lanes of the electrophoresis gel it is evident that, of the 2.5 µg of DNA digested in each case, the majority of DNA in the testis samples persisted as high molecular weight fragments indicating that *HpaII* digestion was inhibited. *HpaII* inhibition is indicative of a methylated cytosine residue in the restriction enzyme target sequence. Therefore *L. ocellata* testis tissues appear to exhibit higher levels of cytosine methylation relative to somatic and ovary tissues.

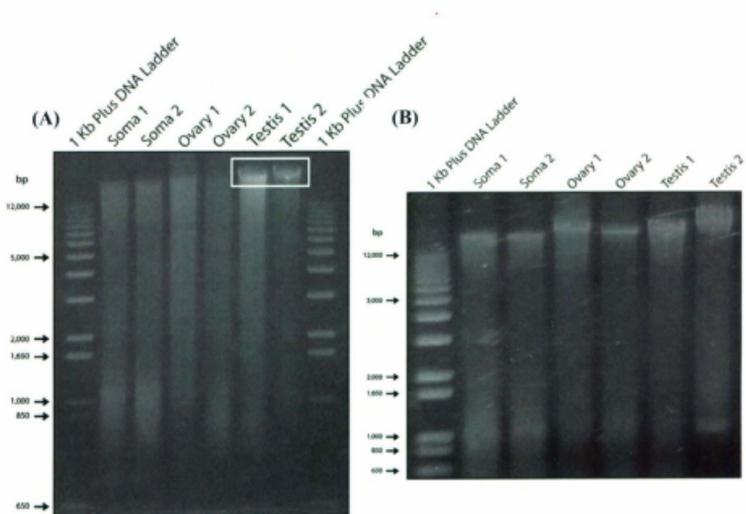


Figure 5: Methylated cytosine is elevated in the testis of *L. ocellata* when compared to ovary and somatic tissues. (A) Methylated cytosine in the target sequence 5'-CCGG-3' prevents *HpaII* from digesting genomic DNA resulting in a greater proportion of high molecular weight fragments (outlined in white) upon separation via gel electrophoresis in 0.8% agarose. (B) Control electrophoresis gel. Initial integrity of gDNA samples is confirmed by large proportion of high molecular weight fragments present in all samples prior to *HpaII* digestion.

Isolation and characterization of the L. ocellata Dnmt1

The 3' end of the Winter Skate *Dnmt1* was previously acquired by the McGowan laboratory (unpublished data). The cDNA cloning was completed by amplifying an ~2600 bp 5' sequence using RACE RT-PCR (Figure 6). The full length *L. ocellata Dnmt1* cDNA sequence consists of 4990 bp (Figure 7). Based on the longest open reading frame (ORF) identified in the cloned sequence (4531 bp) the first 73 bp and final 386 bp are untranslated (see Figure 4 for schematic of cDNA clone).

The translated ORF of the *L. ocellata Dnmt1* yields a 1509 amino acid sequence (Figure 8). The sequence surrounding the translation start codon (AUG) was compared to the Kozak consensus sequence [Kozak 1987] (Table 2) which acts to increase recognition efficiency of the start codon by the 43S translation pre-initiation complex [as reviewed by Kozak 2005]. *L. ocellata Dnmt1* displayed a weak resemblance to the Kozak consensus sequence as only the -3 located nucleotide (three nucleotides upstream from the adenine of the AUG start codon) and not the +4 located nucleotide (four nucleotides downstream from the adenine of the AUG start codon) was conserved. Regardless of the weak nature of the *L. ocellata Dnmt1* Kozak consensus sequence, a complete ORF is present with both a start and stop site yielding a protein sequence possessing the expected domains for the enzyme type in question. Three nuclear localization signals (NLS) were predicted using the on-line WoLF PSORT software (<http://wolffpsort.org/>) [Horton *et al.* 2007]

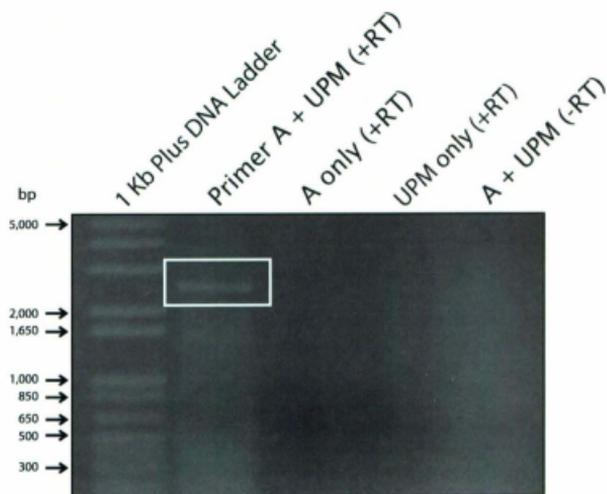


Figure 6: The 5' end of *L. ocellata* Dnmt1 generated by 5' RACE RT-PCR. ~2600 bp (outlined in white) amplified with a primer designed from 3' sequence previously acquired by the McGowan laboratory.

TTTATTTACTCATTATAGGGCAAGCAGTGGTATCAACGCAGAGTACATGGGGACCGACAACAAT
ACAATACAAGCGCGGCACTACCCGCTGCTCGTGCCTGCCGACGAGCTCAAGAAGCGGCTTCAAGCA
CTTGAGGATAAAGCGGCTCATCAGAGAAGGAATGTGTGAAAGAAAAGCTGAGTTTGGTGCATGG
CTTCTTCAAGCAGATGCACAAAATGACCTCAGCGAGCTGGAACAAAATTTAAAAAGAAAGAGC
TCTCGGAGGAGGTTTATCTGGCCAAAAGTGAAGCATTGCTCAGCCAGGAGCTTCCGTAAGGAAAT
GGAAATCTGAGCTTGGTGTGAATCAAATGGTTGCACAGAAAATGGAGCCACTGACAGTGTATAA
AAGTGACACCAATGGTGATCTGATGAGCCAAAATGGCCAAAGAGGAGCAAAATATGGAAAAGAAAG
CTGTGGTTTTCCCAACAGCAAAGAGTCTGGGAAGCGACTAAGTAAATCCAATGGAGATAACAAA
AGACTATCAGGAAGTCCAAGGTTCCACAGGAATCGGCAAAAGCAGCAAGCCACAATACATCCAT
GTTTCTAAAGTAGCAAAACAACGCAAGTCAGATGACCTAAATGAACCAAGTCAAGAAAGAAATG
AGACCGATGAATGTAATGCTGAACAGGAGCAAGATGAAAAGAAAATCAAATGTAATCTGAACAA
ACGCTTAGTGGAGCCGAAGTACCTCTAAGTGTAAACCCAAAAGTGAACAACTCTTAAGACGCC
ACCTCCAAAGTGACAGATTGCAAGCAGTCTTGGATGACCCAGATCTCAAATACTCCAGGGCG
ATCCTGATAATGCATTAGAAGAGCCTGAGATGTTGACTAATGAGCGGTTGCTACTCTTTGAAGGA
ACGAACGATGAAGGTTTTGAAAGTTATGATGACTTGCCCTCAACATAAAGTTACATCTTTCAGTAT
TTATGATCGGAAGGCACTTGTGTGCTTTTGATACCTGGCCCTCATCGAAAAGAAATGGAAATTTG
GCTTCAGTGGTGTGGTGAAGCCAATCTATGATGATAACCAAGCCCTGGATGGTGGAGTCAGAGCT
AAAAACTGGGACCAATAAATGCCTGGTGGATAACTGGTTTTGATGGTGGAGAAAAGCTTTGAT
TGGATTTCAACAGCGTTTGGCTGATTATATATATGAGGATCCAGTGAACAGCAATCAAGAACCT
TTGCTGTGATGCAAGAGAAAATCCACATAAGCAAAAATGTGATGAAATTTCTACAAAATAATCTT
GATTCACCTTATGAAGACTTGCTAAACAAGATAGAGACTACAGTACCTCTGCTGGACTGAGTTTT
CTCTCGCTTACGGAGGATTCGCTGTGAGACACGCCCAAGTTTTGCTTGAACAAGTGGAGATTT
ATGATGAAGCTGGTGTGATGATGATGAACAGCCTATCATTATAACTCCCTGTATGAGGGATCTGATC
AAGCTGGCTGGTGTCACTCTTGGGAAAAGACGAGCTGCAAGGCGACAAGCAATACGTCACCCCTAC
CAAGATCGATAAAGGACAAGGGACCCACTAAAGCCACTACCCTAAGTTGGTATATCAGATTTTTG
ATACTTTCTTCTGAGGAAATGATCAGAATGACAAGAGAATGGATCAAAGCGTAGGCGCTGT
GGAGTGTGTAGGTTTTGTCAACAGCCTGACTGTGGGAAGTGAATGCTTGAAGGACATGGTGA
GTTTTGGGGCTTGGGTCGTGAACAAAACAAGCCTGCTTACAGAGAAGTGTCCAATCTGGCTGTGA
AGGAGGCTGATGACGATGAGAATGATGAGGATGACTCTGATCTAATTGACAAGACTCTCCTAAA
AGAAATGTTGCAGGCTCGGAAAAAAAAGCAATCGAAGAGCCGCATCTCTGGATGGAGGCTAT
GAAGTCTGACGGACGGAAGACCTTATACCAGAAAGTGTCTTGGACGATGAAATTTGCGCATCA
ATGATTTGTGGCAGTCAGCCCGGATGACCCCAACAAAGCGTTTACTTAGCAAGAAATCACATCA
ATGTGGGAAGAAGTCGGTGGGAAGATGTTCCATGCAAAATGGTTCTGCGGTGGCACAGACACTGT
ACTGGGTGAAACCTCTGATCCACTTGAAGCTCTTTCTAGTGGATGAGATGGAAGACTGCAAGTAT
CTTATGTAGACAGCAAAGTAAAGGTTATTCACAAGGCTGCCTCGGAGAGCTGGGCAGTGGAGGGT
GGAATGGATGATGAGTTGAGTTAAAAATGGTTGAAGATGGAAGACACTACTCTACCCAGAT
GTGGTATGATCCAGAGTATTCCCGATTCCAATTCCTCACCATGTGAATCAACGAGGAGGAACA
AGCACAAGTCTGTGACAGTTGCACTCGGTTGGCTGAAATCAGACAGCGAGAGATGCCACGGGTTG
CTGGAGCCACTGGAGCTAAAGGATGATTCAAAGGTTTTCTACGCATGAGACTAAGAATGGAAAC
GCAGTATAAAGTTGGAGATGGCATTACCTCCTGCAAGATGCATTCTCATTACGCGTTAAACCAT
CTAGTCTGGCAAGCGCAGGATGAAGAAAGATGACGTGGATGAAGCACTGATCCCGAGTACTAT
CGAACTCATCTGATTACATCAAAGGGAGCAACTTGATATCTCGAACCAATCCGAATTTGGCGG
CATTTCATGAAATCTTCTGTCAACGCTAGCAATGGCAAACCAATGAAGCAGATATTAAGCTAC
GAATCAATAAATCTCAGAGCTGAGAATCGCACAAAAGGCTGAAAGGCGAGTTACCACACTGAC
ATCAATTTGCTATATTTGAGCGATGAAGAAGTGCACGCTGGACTCAAAGGACATCCAAGGCGCTG
TACTGTGGAATATGGTGAAGATCTGACTGTATGCGTTACGAAATACTGTGCTGGAAGCCAGACA

GATTCTATTTCCTAGAGGCCACAAATGCAAAAGACAAAATCCTTTGAAGACCCTCCCAACCCGTGCA
 CGCAGTTCGGCCAACAAGGCAAGGGTAAAGGGAAAGGCAAAAGGCAAAAGGAAAGGGGAAGTCTGC
 TACTGAACCTGAACAACAGAGCCAGAGGTTTCGGACAAATTCGAGAAACTTCGCAGCCTTGATG
 TATTTTCAGGCTGTGGGGGTCTGTCTGAAGGATTCCATCAGGCAGATCTTTTCGAAAACCATGTGG
 GCCATTGAGATGTGGATCCAGCAGCCAGGCGTTCGGTTTGAATAATCCCGGTGCCACCGTCTT
 CACTGAGGACTGCAACGTTCTCTTGAAACTGTGCATGGCTGGAGAGAAGCAATTCACATTTGGAC
 AGAAACTGCCACAGAAGGGTGTGTGGAAATGCCTGTGGTGGCCCTCCGTGCCAGGGCTTTAGC
 GGGATGAATCGGTTCAATTCGCCACCTACTCCAATTCAGAAGTCCCTTTGGTCTCTTATCT
 CAGCTATTGTGACTATTACCGGCCCGGTTTTCCTCCTGGAAAACGTGAGGAACTTTGTCTCAT
 TCAAGCGCTCCATGGTCTGAAGCTAACTCTTCGCTGCCTTGTTCGAATGGGCTACCAGTGCACA
 TTTGGTGTCTTCAGGCTGGTCACTACGGAGTTGCCAGACCCGCCGGAGAGCATCGTTCTGGC
 CGCAGCCCTGGAGAGAAGCTGCCGCTCTCCAGAGCCCTGCACGTCTTCGCACCCAGAGCCT
 GCCAGCTCAGTGTGGCAGTGGATGACAAGAAGTTTGTCAAGTATGTCACCAGGACAAAAGTCTGTCT
 CCGTACAGAACCATCACTGTAAGAGATACCATGTCCGATCTGCCTGAAATTCGCAATGGAGCAT
 GGGCGTGGAAATCTCGTACAATGGTGAACCTCAGTCTGGTTTCAGAGGCAGATTCGAGGCACAC
 AGTACCAGCCTATCTGCGGGATCATGTCTGTAAGGATATGAGTGCCTTTGGTGGCAGGCAGAAATG
 CGCCACATTCCTCCGCCCGGGGTCTGACTGGCGTGATCTTCCAATAATTGAGGTGCCCTTTC
 CGATGGCACCATGACCAAGAAGCTGAGGTACACACACCAGATAAAAAAGAACGGGCATAGCAGCA
 CTGGTGGCTACGTGGAGTGTGTTCTTGCCGACAGCTGAAGCAGTGTGAGCCGGCTGACAGACAG
 CAGGGGTTCCAGACACCTACAGGCTCTTTGGGAACGTTTTGGACAAGCACAGGCAGGTGGCAA
 GTATGGCAGGCTGGAATGGATGGATTCTCAGCACCACCGTCACCAATCTTGAGCCGATGGGAA
 AACAGGTCGTGTTCTACATCTGAGCAGCACCCTGTGGTGTGAGTGTGCGAGAGTGTGCACGATCC
 CAGGGGTTCCAGACACCTACAGGCTCTTTGGGAACGTTTTGGACAAGCACAGGCAGGTGGCAA
 TGCAGTACCACCTCCTCTGGCTAAGCTATCGGCACGGAAATCAAACCTTTGCGCACTTGACAGGA
 AGAAAGGGAATACAGAGCACATCAAACCTAGAGACAATGGACACAAGCGCCCTGATTCATCTCTTC
 AGCTTGATGTTCCACCACATCTTCATGCAGTGCATTCGAAGGCAGGAGGAAAGAAAATACGATG
 GGATTCCTGTGACTGTTCCGGTAAATAATGTTTTTTTTTAAACTGGGTGATGTGAGGCAGCCAAT
 GAAGATGTAACATTGTTTTAGTTATGAATGAACATTTTTTTTTGATTGTGCAGTGTCTGTCAATG
 CATTGTGGATTTTAAATGTGGTTTTAAAAATGCACAGTATTTGAATAAATGCCCACTTTTTGCAGTG
 GTAATTCATATGTTTAAATAAATGTAGTTTTTATATGTTGTAATATTTCAATAAATATTTTAAAGT
 GGAATGCTATTATGCCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Figure 7: *Leucoraja ocellata* DNA-methyltransferase 1 cDNA sequence. 4990 bp
Dnmt1 clone with the largest ORF identified consisting of 4531 bp. Start
 (ATG) and Stop (TGA) codons are indicated in green and red text
 respectively.

MPGTTTRCSLPDDVKKRLQALEDKAGLSEKECVKEKLSLLHGFLQADAQNDLSELETKFK
 KEELSEEGYLAKVKALLSQELSVGNNGNSELGVNSNGCTENGATDSDKSDTNGVSDPEPNG
 QEESNMETEAVVFTAKSRGKRLSKSNGDNKRLSGSPRFTRNSAKQATITSMFAKVAN
 KRKSDDLNEPESKKNETDECNAEQEQDEKKIKIESEQTLSGAEATSNCCKPKSEQTPKTP
 PPKCTDCKQFLDDPDLKYFQGDPDNALEEEPEMLTNERLSLFEGTNDEGFESYDDLPOHK
 VTFFSIYDRKGHLCAFDGTGLIEKNVELCFSGVVKPIYDDNPSLDGGVRAKLLGPIINAW
 ITGFDGGEKALIGPTTAFADYILMDPSDEYATIFAVMQEKIHSKIVIEFLQNNLDSTY
 EDLLNKIETTVPAGLSFSRFTEDSLRHAQFVLEQVESYDEAGDVDEQPIIITPCMRD
 LIKLAGVTLGKRRRAARRQAIRHPTKIDKDKGPTKATTTKLVYQIFDTFFSEEIDQNDKE
 NGSKRRCGVCVEVCQPPDCGCKNACKDMVKFGGLGRNKQAQLQRRCPNLAVKEADDDEN
 DEDDSLIDKASPKRMLQGRKKKQSKSRISWIGEAMKSDGRKTYQKQVSVDDDEILATND
 CVAVSPDDPTKPLYLARITSMWEEVGGKMFHANWFCRGTDTVLGETSDPLEFLVDECE
 DMQLSYVDISKVKVIHKAASESWALEGGMDDEFELKMWEDDGKTYFYQMWYDPEYSRFQI
 PHPCESTEENKHKFCDSCTRLAEIRQREMPRVLEPLELKDDSKVYALATKNGTQYKVG
 DGIYLLQDAFVSFKVSPSGPKRPVKKDDVDEDLYPEYRKSDDYIKGSNLDIPEPFRIG
 RIHEIFCHKRSNGKPNEADIKLRINKFYRAENTHKGLKGSYHTDINLLYWSDEEVTVD
 KDIQGRCTVEYGEDLTVCVQEYCAQSPDRFYFLEAYNAKTKSFEDPPNRRARSANKGKG
 KGKGGKGGKGSATEPEQTEPEVSDKFEKLRSLDVFSGCGGLSEGFHQADLSETMWAIE
 MWDPAQAQFRLNPGATVFTEDCNVLLKLVMAGETKNSLGQKLPQKGDVEMLCGGPPCQ
 GFGSMNRFNSRTYSKFKNSLVVSYLSYCDYRPRFFLENVNRFVSFKRSMVLKTLTRC
 LVRMGYQCTFGVLQAGQYQVQTRRRRAIVLAAAPGEKLPFPEPLHVAFPRACQLSVAV
 DDKKFSVSNVTRTKSAPYRTITVRDMSDLPEIRNGASALEISYNGEPQSWFQRQIRGTQ
 YQPILRDHVCKDMSALVAGMRHIPLAPGSDWRDLPNIEVRLSDGTMTKKLRVTHHDKK
 NGHSSGTALRGVCSADVKQCEPADRQFNLTLPWCLPHTGNRHHWAGLYGRLEWDGFF
 STTVTNPEPMGKQGRVLHPEQHRVSVVRECARSQGFPDITYRLFNGVLDKHRQVGNVPP
 PLAKAIGTEIKLALDRKKGNTEHIKLETMDTSA

Figure 8: *Leucoraja ocellata* DNA-methyltransferase 1 amino acid sequence. Translation of the largest identified ORF of the *Dnmt1* clone yields a 1509 aa sequence. N-terminus domains are indicated in orange and the C-terminus C-5 DNA methylase domain is indicated in green (refer to Figure 9 for N-terminus domain identities).

Table 2: Comparison of *Dnmt1* translation start sites in vertebrates. Comparison to the Kozak consensus sequence [Kozak 1987] focuses on nucleotides in the -3 location (3 nucleotides upstream from the adenine nucleotide beginning the translation start codon) and +4 location (4 nucleotides downstream from the adenine nucleotide beginning the translation start codon) proven to be the most influential in targeting the 43S pre-initiation complex to the translation start codon (green text) [Nakagawa *et al.* 2007]. Similarity to the Kozak consensus sequence is categorized as **Strong** (both the -3R and +4G nucleotides are conserved), **Weak** (one of the nucleotides in either the -3R or the +4G location is conserved), or **Poor** (neither of the nucleotides in the -3R or the +4G locations are conserved). R: A or G

Nucleotide Position	-6	-5	-4	-3	-2	-1	1	2	3	4	
Kozak Seq	G	C	C	R	C	C	A	T	G	G	
Organism											Similarity to Kozak consensus sequence
<i>L. ocellata</i>	A	A	T	A	C	A	A	T	G	C	Weak
<i>X. laevis</i>	A	T	A	G	C	C	A	T	G	C	Weak
<i>G. gallus</i>	C	C	C	C	C	G	A	T	G	C	Poor
<i>M. musculus</i>	T	G	C	A	A	G	A	T	G	C	Weak
<i>H. sapiens</i>	T	C	C	G	A	G	A	T	G	C	Weak
<i>D. rerio</i>	C	T	T	G	A	A	A	T	G	C	Weak

(NLS Score of 0.41): PDDVKKR, KRRR and RKKK. Their locations begin at amino acids 10, 535 and 610 respectively. Both the Conserved Domain Database (CDD) [Marchler-Bauer *et al.* 2011] and SWISS-MODEL Workspace [Arnold *et al.* 2006] were used to identify the remaining protein domains. Their identities and amino acid positions are as follows (Figure 9): DNA Methyltransferase 1-Associated Protein binding domain (DMAP1) 5-96, Replication Foci Domain (RFD) 288-423, Zinc Finger domain (Zf) 532-578, Bromo Adjacent Homology domains (BAH) 643-769, 857-991 and C-5 cytosine-specific DNA methyltransferase domain 1033-1487. Figure 10 outlines the ten catalytic motifs of the *L. ocellata* Dnmt1 methyltransferase domain. Conserved sequences determined by Pósfai *et al.* [1989] were used to probe the subject *L. ocellata* sequence for similarity. In addition the CDD was utilized in outlining specific features of the methyltransferase domain. As a result motifs I, II, III, IV, V, VI and VIII superimpose very well over the CDD data while the remaining motif locations do not align with amino acids indicated to be involved in specific features.

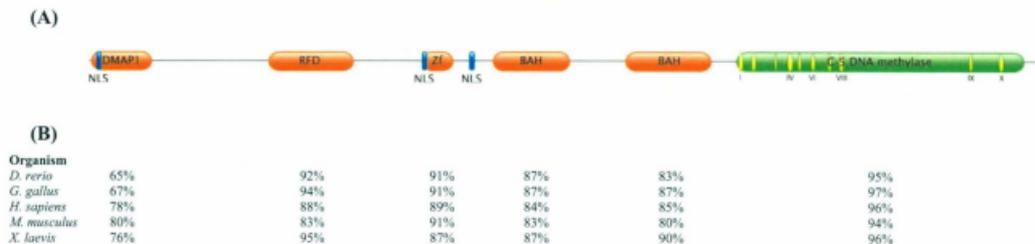


Figure 9: Dnmt1 functional domains in representative vertebrate models. (A) Schematic of *L. ocellata* Dnmt1 protein. Functional domains identified in the N-terminus are indicated in orange (DMP1: DNA methyltransferase 1-associated protein binding domain, RFD: replication foci domain, ZF: zinc-finger domain, BAH: bromo adjacent homology domain). Predicted nuclear localization signals (NLS) are shown in blue. The C-terminus catalytic region is shown in green with the predicted 10 catalytic motifs shown in yellow. (B) Percentage similarity of the subject *L. ocellata* Dnmt1 functional domains with other vertebrate Dnmt1 proteins (*H. sapiens* GenBank [X63692](#), *M. musculus* GenBank [X14805](#), *D. rerio* GenBank [NM_131189](#), *G. gallus* GenBank [D43920](#), *X. laevis* GenBank [D78638](#)).

```

      #####
LRSLDVFSGCGGLSEGFHQADLSETMWAIEMWDPAAQAFRLNNPGATVFTEDCNVLLKL
      #####
VMAGEKTNSLGQKLPQKGDVEMLCGGPPCQGFSGMNRFNSRTYSKFNSLVVSYLSYCD
      #
      #
YRPRFFLLENVRNFVSVFKRSMVLKLTLRCLVRMGYQCTFGVLQAGQYGVVAQTRRRRAIV
      #
      #
LAAAPGEKLPFPPEPLHVFAPRACQLSVAVDDKKFVSNVTRTKSAPYRTITVRDTMSDL
PEIRNGASALEISYNGEPQSWFQQRQIRGTQYQPILRDHVCKDMSALVAGRMRHIPLAPG
SDWRDLPNIEVRLSDGTMTKKLRYTHHDKKNHSSSTGALRGVCSCADVKQCEPADRQFN
TLIPWCLPHTGNRRHNHAGLYGRLEWDGFFSTVTNPEPMGKQGRVLHPEQHRVVSRE
      #
CARSQGFDPTYRLFGNVLDKHRQVGNVPPPLAKAIGTEIKLCA

```

Figure 10: Catalytic motifs present in the *L. ocellata* Dnm1 C-5 cytosine-specific DNA methyltransferase domain. Underlined regions correspond to the 10 motifs predicted using 13 bacterial DNA methyltransferases [Pósfai *et al.* [1989]. Pound symbols located above amino acid residues indicate their incorporation into specific features as reported by the Conserved Domain Database [Marchler-Bauer A *et al.* 2011]. Pound symbol colour code is as follows. **Red:** Feature 1 (Cofactor Binding). **Green:** Feature 3 (DNA Binding). **Blue:** Component of Feature 1 and 3. **Orange:** Component of Feature 2 (Substrate Interaction) and 3.

Isolation and characterization of L. ocellata Dnmt3

RACE RT-PCR was performed using Primer3 [Rozen and Skaletsky 2000] designed primers based on the previously acquired 274 bp region of the Winter Skate *Dnmt3*. These reactions yielded two amplicons for the 3' end as well as two amplicons for the 5' end sized at ~1000/1200 bp and ~2000/2200 bp respectively (Figure 4). The full length *L. ocellata Dnmt3* cDNA sequence incorporating the longest amplicon of each end consists of 3301 bp (Figure 11). The largest ORF identified (2340bp) results in a 5' UTR of 214 bp and a 3' UTR of 747 bp (see Figure 3 for schematic of cDNA clone).

The longest identified ORF of the *L. ocellata Dnmt3* yields a 779 amino acid sequence (Figure 12). WoLF PSORT [Horton *et al.* 2007] predicted an NLS of six amino acids (RHKKKK, NLS score of 0.82) starting at position 61. The SWISS-MODEL Workspace [Arnold *et al.* 2006] identified the remaining protein domains. Their identities and amino acid positions are as follows (see Figure 13): Pro-Trp-Trp-Pro domain (PWWP) 141-214, Plant Homeo Domain (PHD) 367-444 and C-5 cytosine-specific DNA methyltransferase domain 473-746. Figure 14 outlines the ten catalytic motifs of the *L. ocellata Dnmt3* methyltransferase domain. Motif assessment was carried out as outlined above. Again, motifs I, II, III, IV, V, VI and VIII superimpose very well over the CDD data. Motif X also shows some similarity with the CDD feature description. The remaining motif

ACATGGGGCAGTGAAGTGAAGTGGAGCCGGAGTGAAGTCGGAGCGGGCGATCGCGGACCCAGCGCAATATCCCCCT
 CCTCACCTTCCCTCGAGCCGAGATCTTAAACAGAAAGGCCAGCAATGAAGTC~~TAAGCATTCGAACCTCGAAACCTGTCT~~
 GTCGACATCAGCAAACTAGAGCTTCCATGAACGGGAAACTCACCACCTGGATCA~~ATGGGGAGCAAGCAACAACATCT~~
 AGTTCCCAAACCCAGCCCAAATCCCACCTCAAAGTGGAACTGATAGCACCCGATCTTGACATGGAGATCTATCGAGT
 GAAAGCAGTCTAGGGAGGAAGCTTGAATGATGGGAGGCCAAAGGGTAAATCGAGATGCCAAGAAGAAAGCGAAGAGGCCAA
 AGCGCGAGGAGATACAGTGTGGGAGGACGCTCTCAGACAAAGGCCAGCACAGCTATAAATTTCCAGGCAAGGATTTCT
 CTTTCGCCACAAAAAAGAAAGAGATAAATAATGAAACTAAAAAGGCATGAACAATTCATTTGTCTCAGAAAGCAG
 AGGACATCTCTACCAAGGCTTTTATCGCTGAAAAATCTGTCTATAGACCTCACTTTGGAGCCTCTGGATAGCAGCAAGA
 AACCTCCAGTTCACGACAGAAAGCTTTGTCTGCGAAAAATAGCATTGTGAAGGGTACAGATGGAATCCCTCAGTACOCAGGA
 TGCCAAAGGTTACGGTATTTGGGGAGTTGGTGTGGGAAAGATAAAGGGTTCCTCGTGGTGGCCGCCCATCGTGGTGTG
 TGCGCACACCCGGAGCGAGGCAAGCGCGCTCGGGGATCGGTTGGCTCAATGGTTCGGAGACGGCAAGTCTCAGAGG
 TTTCTGCCGCAAACTATGCTCTTAACTGCGATTGGTCAATATTTCCACACATCTGCATTAAACAAGCTGATTTCTTA
 CAAAGAGCTGTGATCAGGCTTTGGAGATGCGAGCAGCAGATCTGGGATTCATCCCTCCCAATGACCGGGGCACT
 TTGGAGGAACAGATAAAGCCCATCTAGACTGGGCATTTGCTGGTTTCCAACCGAAAGGCTACGAAAGAAATTAACCCCA
 AACAAAAACAGAGAACTACTCAGATGGCACTCAAGTCGAGGCTGTGTGCCGAGTACTACTCGCCAAACAAAGAA
 ACAGAAGACTAGTCTTATAAAGCAAGAAGGACCTGAGGAGAACCCCTGGCAGAGAAAAAATCTTCTTCAAGTT
 ACATCGAATAATAAAGCATTGAAGAATTTTGTCTTGCTTGGCGCAGCATTAGAATCGGCACCTTTTCATCCACTGTTTG
 AAGTGGTCTATGTTCCATTTGCAAGGATATCTACTGGAGACTTCTACATGTATGACGATGATGGCTACACGATCCTTA
 CTGTACCGTGTGCTGTGGTGGGGGGAGGTTCTGCTGTGGGAATGCAAACTGCTGCAAGTGTTTTGTGGAGCTG
 ATAGATATTTTGGTGGTCCGGGTGCACTGAAGAAGCCAAAGTCTCGGACCTTGGCGATGTACATGTGCTCGCCG
 ACGAAAGCTACGGAGTGTGGAGGCGCAGGAGACTGGACCATGAAACTCAAGAGTTTTTCGACGGCAATGGGCA
 GGAATATGATCGCCTAAAATTTACCAGCCGCTACTGCGCAGAAACAGAAAGCCAAATCAAAGTCTCTGTGTTGTTGAT
 GGGATAGCAACGGGATATTTAGTTTAAAGGATTTAGGATTTAAGTTGAGAGGTATGTGCTCTGAAATATGTGAAG
 ACTCGATCGCAGTGGGAGCGGTGCGCATGAAGGAAGAAATCACATACGTACATGATGTCAGGAACATCAGCCGGCAGAA
 CATTATGAGTGGGGTCCGTTTGTATCTGGTATTGGAGGACGCCCTGTAACGATCTTTCTATTTGAAATCTCAGAA
 AAGGGTTATATGAGGGCCAGGAAAGACTTTTCTTGAATTTATCGACTGCTTCATGACACAGCAAAAGAGGTGGG
 AAGCAGGCTTTCTCTGGCTATTGAGAATGTTGGTCCATGGGTGCAATGATAAAGAGGACATCTCACAGTTTCT
 GAGTGAACCCAGTATGATCGATGCAATCGACGTGTCTGTCTGCCACCAGGCTCGCTACTTTTGGGAAACTTACCA
 GGAATGAACAGGCCCTGGTGTCTTCTCAGCAGATAAACTGGAACCTGCAGCACTGCTTGACATGGTATGGATAGCAA
 AGTTTAGCAAGTTAGGACCATACACAAAGGTCGAATCCATCAAGCAGGGCAAGGATCAACATTTCCAGTTATCAT
 GAATGGGAAGGAGACATCTGTGGTGTACAGAACTGGAGGAGATCTTTGGCTTTCCAGTGCATACACAGACGCTGTG
 AACATGGGGAGGAGCTCGCCAGAAACTCTTGGAAAGTCTTGGAGTGTCCCTGTCACTCCGCAACTATTTCAGCCTC
 TGAAGGATATTTTGTCTGTGAATAACACAGCATATAATAGTCTTTCAGAAACTTACGGTGCTCTCTAGAAACTAAC
 AAGTAGTGAACAAATCCAGACTGGTTTTCAAGTACTGTGACACAAAAATTTTCCCTGGCATTATTGTGAAGGCTGCA
 TGCTGTACATTTACGATACTGTGCCATTATTCACAGTCAACAACCTAGGACAGATAGGTAGCTAGACTGCCTACTATA
 GTTGTTTTAGATTTTGAATTTTCATCTTTTATATGGGAAGAAACAGAAATTCCTACTTTTAGTTATAGTTCTTACAA
 TAGTGTGAGGATTACTTCACAGTTTTTTAGTACTTTGTAGCCAGTTTTATATGGGATTAAGGGCTTTTTATTT
 CTGTAGTGTCTATTGTCGCCCTTTAGATTTTCACTATTTTTAAATTTCCAAACATAGATTTTATTGTTACTATA
 CATACAGTATCAGCAGCTGGATTTGGTAGACAGATAAATAGTCTTGACCTATTTAACAAAGATCAACCTAGCAGAGT
 TAAATCATGTGTTCAAATGTTGAAGTGTGTTCTGTTTAAACAGTATCCAAGTATTTTAACTTTTCCAGAAAGT
 GTAATTTTTCATGTAACCTGTGTATGTGATTTGGGAAATAATCATGATACATTTGTTAAATTTGTTATGTTTGTGA
 TTTTAGCAAGGTTGCAGATGATAAATGAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Figure 11: *Leucoraja ocellata* DNA-methyltransferase 3 cDNA sequence. 3301 bp
Dnmt3 clone with the largest ORF identified consisting of 2340 bp. Underlined sequence indicates regions of transcript subjected to splicing. Start (ATG) and Stop (TAA) codons of ORF are indicated in green and red text respectively.

MGSKQQHPVPKPPAQIPLKVELIAPHSDMEILSSESSLGRKLD DGRPKGKSRCEESE
 AKRRGVSGWESSLRQPAPRIIFQAGISFRHKKKKEDNNMMLKDKDMNNSLSQKTRTSAT
 KAFIAENSVIDLLEPLDSSKKPPVPAESFVCENSIVKGT DGIPOYQDGRGYGIGELVW
 GKIKGFSWWPAIVVSWRTPGRRQAASGMRWLQWFGDCKFSEVSADKLMPLTAIGQYFHT
 SAFNKLISYKRAVYQALEIASSRSGIPFPSNDRGTLEEQIKPMLDWAFAGFQPKGYEGI
 KPKQNTENDTADGTPVEVCVPEYYPTKKQKTSLYKSKEGPEEEHRGREKIFFQVTSNN
 KSIEEFCLACGSIRIGTFHPLFEGGLCSICKDIYLET SYMYDDDGYSYCTVCCGGREV
 LLCGNANCCRCFCVDCIDILVGGPASEEAKVLDLPWRCYMCLPHESYGVLRRRGDWTMKL
 QEFFASDNGQEYDPPKIYPVPAENRKPIKVLSLFDGIATGYLVLRDLGFKVERYVASE
 ICEDSIAVGTVRHEGRITYVHDVRNISRQNIHEWGPFDLVIGGSPCNDLSIVNPARKGL
 YEGTGRLFFEFYRLLHDTRPKEWEDRPFVWLFENVVAMGVNDRDISRFLECNPMVIDA
 IDVSAHRARYFWGNLPGMNRPLVASSADKLELQHCLEHGRIAKFSKVRTITTRSNSIK
 QGKDQHFVIMNGKEDILWCTELERIFGFPVHYTDVSNMGRGARQKLLGRSWSVPVIRH
 LFAPLKDYFACE

Figure 12: *Leucoraja ocellata* DNA-methyltransferase 3 amino acid sequence. Translation of the largest identified ORF of the *Dnmt3* clone yields a 779 aa sequence. N-terminus domains are indicated in orange and the C-terminus C-5 DNA methylase domain is indicated in green (refer to Figure 13 for N-terminus domain identities). Underlined sequence indicates regions of protein affected by transcript splicing.



(B)

Organism	PMSBP	PHD	Long	Short
<i>D. rerio</i> dnm13	66%	74%	85%	88%
<i>D. rerio</i> dnm14	87%	88%	95%	95%
<i>D. rerio</i> dnm15	62%	76%	85%	86%
<i>D. rerio</i> dnm16	71%	87%	92%	70%
<i>D. rerio</i> dnm17	67%	82%	88%	91%
<i>D. rerio</i> dnm18	69%	84%	92%	94%
<i>G. gallus</i> dnm13a	74%	89%	93%	91%
<i>G. gallus</i> dnm13b	89%	86%	91%	84%
<i>H. sapiens</i> dnm13a	75%	89%	94%	94%
<i>H. sapiens</i> dnm13b	85%	85%	93%	93%
<i>M. musculus</i> dnm13a	75%	89%	93%	91%
<i>M. musculus</i> dnm13b	83%	86%	92%	93%

Figure 13: Dnmt3 functional domains in representative vertebrate models. (A) Schematic of *L. ocellata* Dnmt3 protein. Functional domains identified in the N-terminus are indicated in orange (PWWP: Pro-Trp-Trp-Pro motif, PHD: Plant Homeo Domain). The C-terminus catalytic region is shown in green with the predicted 10 catalytic motifs shown in yellow. The predicted nuclear localization signal (NLS) is shown in blue. Regions showing evidence of transcript splicing are indicated by triangles. (B) Percentage similarity of the subject *L. ocellata* Dnmt3 functional domains with other vertebrate Dnmt3 proteins (*H. sapiens* 3a GenBank [AB208833](#), *H. sapiens* 3b GenBank [AB208880](#), *M. musculus* 3a GenBank [AF068625](#), *M. musculus* 3b GenBank [AF068626](#), *D. rerio* 3 GenBank [AB196914](#), *D. rerio* 4 GenBank [AB196915](#), *D. rerio* 5 GenBank [AB196916](#), *D. rerio* 6 GenBank [AB196917](#), *D. rerio* 7 GenBank [AB196918](#), *D. rerio* 8 GenBank [AB196919](#), *G. gallus* 3a GenBank [NP_001020003.1](#) and *G. gallus* 3b GenBank [NP_001019999.1](#)). Columns titled Long & Short indicate alignments carried out using the 274 amino acid and 211 amino acid *L. ocellata* domains respectively.

DGTPVEVCVPEYYPPTKKQKTSLYKSKEGPEEEHRGREKIFFQVTSNNKSIEEFCLACGSIRIGTFHFLF
 EGGLECSICKDIYLEASYMYDDDDGYQSYCTVCCGGREVLVLCGNANCCRCPCVDCIDILVGPBASEAKVLD
 PWRCYMCLPHESYGVLRRRGDWTMKLQEFFASDNGQEYDPPKIYPVPAENRKPICKVLSLFDGIATGYLV
 LRD LGFKVERYVASEICEDSIAVGTVRHEGRITYVHDVNRNISRQNIHEWGPFDLVIGGSPCNDLSIVNPA
 RKGLYEGTGRLLFFEFYRLLHDTRPKEWEDRPFWLFENVVAMGVNDKRDISRFLCNPVMIDAI DVSAAH
 RARYFWGNLPGMNRPLVASSADKLELQHCHLEHGRIAKFSKVRTITTRSNSIKQGKDQHFVIMNGKEDIL
 WCTELERIFGFPVHYT DVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE

Figure 14: Catalytic motifs in the long 3' variant *L. ocellata* Dnmt3 C-terminus. Underlined regions correspond to the 10 motifs predicted using 13 bacterial DNA methyltransferases [Pósfai *et al.* 1989]. Pound symbols located above amino acid residues indicate their incorporation into specific features as reported by the Conserved Domain Database [Marchler-Bauer *et al.* 2011]. Pound symbol colour code is as follows. **Red:** Feature 1 (Cofactor Binding). **Green:** Feature 3 (DNA Binding). **Orange:** Component of Feature 1 & 2 (Substrate Interaction). **Purple:** Component of Feature 2 & 3. **Black:** Component of all three Features. **Pink** text highlights target recognition domain affected by 3' transcript splicing.

locations as predicted by Pósfai *et al.* [1989] do not align themselves with amino acids indicated to be involved in specific features by the CDD.

Variation in L. ocellata Dnmt3 transcripts

The different lengths of 5' cDNA detected in this study (Figure 3B) represent alternate transcript splices (Figure 11) revealing multiple potential translation start codons (Figure 12). The upstream start site results in the translation of an additional 28 amino acids within the same reading frame as the downstream start site. The sequences surrounding the two translation start codons were compared to the Kozak consensus sequence [Kozak 1987] (see Table 3) as previously performed for *L. ocellata Dnmt1*. The upstream translation start site yielding the longer N-terminus bears a weak resemblance to the Kozak consensus sequence showing only the nucleotide at the +4 location to be conserved. The downstream translation start site yielding the truncated N-terminus shows a strong resemblance to the Kozak consensus sequence with nucleotides in both the -3 and +4 locations conserved.

Two different 3' cDNA sequences were also detected in this study (see Figure 4A). The shorter of the two sequences lacks an internal 189 bp region coding for 63 amino acids within the reading frame of the longer 3' sequence. Consequently, removal of these 189 bp does not create a reading frame shift (see Figure 12). The *Dnmt3* region affected by

this splicing event lies between motifs VIII and IX of the C-5 cytosine-specific DNA methyltransferase domain (see Figure 14).

Table 3: Comparison of *Dnmt3* translation start sites in vertebrates. Comparison to the Kozak consensus sequence [Kozak 1987] focuses on nucleotides in the -3 location (3 nucleotides upstream from the adenine nucleotide beginning the translation start codon) and +4 location (4 nucleotides downstream from the adenine nucleotide beginning the translation start codon) proven to be the most influential in targeting the 43S pre-initiation complex to the translation start codon (outlined in green text) [Nakagawa *et al.* 2007]. Similarity to the Kozak consensus sequence is categorized as **Strong** (both the -3R and +4G nucleotides are conserved), **Weak** (one of the nucleotides in either the -3R or the +4G location is conserved), or **Poor** (neither of the nucleotides in the -3R or the +4G locations are conserved). R: A or G

Nucleotide Position	-6	-5	-4	-3	-2	-1	1	2	3	4	
Kozak Seq	G	C	C	R	C	C	A	T	G	G	
Organism											Similarity to Kozak consensus sequence
<i>L. ocellata</i> Upstream	G	G	A	T	C	A	A	T	G	G	Weak
<i>L. ocellata</i> Downstream	T	C	T	G	A	C	A	T	G	G	Strong
<i>H. sapiens 3a</i>	G	C	C	C	A	G	A	T	G	C	Poor
<i>H. sapiens 3b</i>	G	A	A	A	G	C	A	T	G	A	Weak
<i>M. musculus 3a</i>	C	C	A	G	C	A	A	T	G	C	Weak
<i>M. musculus 3b</i>	G	A	A	A	C	A	A	T	G	A	Weak
<i>D. rerio 3</i>	G	G	A	A	A	A	A	T	G	G	Strong
<i>D. rerio 4</i>	G	A	C	A	G	G	A	T	G	A	Weak
<i>D. rerio 5</i>	C	T	T	G	A	A	A	T	G	C	Weak
<i>D. rerio 6</i>	T	C	T	G	T	G	A	T	G	A	Weak
<i>D. rerio 7</i>	C	A	G	A	A	G	A	T	G	G	Strong
<i>D. rerio 8</i>	A	C	A	C	A	C	A	T	G	C	Poor
<i>G. gallus 3a</i>	T	G	C	G	C	C	A	T	G	G	Strong
<i>G. gallus 3b</i>	A	C	C	G	C	G	A	T	G	A	Weak

Since a splicing event within the catalytic region of the *L. ocellata* Dnmt3 may be relevant to tissue specificity, a tissue comparison of 3' variants was undertaken. Figure 15 shows gel electrophoresis separation of RT-PCR products from *L. ocellata* ovary, testis, brain and gut tissues amplified using primers flanking the variable region (#3 and #4). Both 3' variants are present at the transcript level in all four tissue types. Due to the fact that an appropriate control gene sequence has not been identified in skates, quantitative comparisons between gel electrophoresis lanes could not be made. However, SenImage software (National Institute of Health) was used to evaluate relative band intensities of the 3' variant PCR products within each tissue type (Table 4). Of the four tissue types tested, only gut tissue showed a majority of the 514 bp amplicon over the 703 bp amplicon based on the latter's 93.2% relative intensity to the former. Ovary, testis and brain tissues all showed majorities of the 703 bp amplicon over the 514 bp amplicon with relative intensities of 123.8%, 150.0% and 160.0% respectively.

In an effort to determine which variable 5' end paired with which variable 3' end new primers capable of isolating the longer and shorter variants in each scenario were designed (primers #s 6, 7, 8 and 9; see Figure 3 for positions). Pairing the long 5' region to the long 3' region for PCR amplification was expected to yield a 2159 bp fragment (primers #8 and #6 in Figure 16). Two fragments amplified with sizes of ~2200 bp and ~2000 bp. The larger of the two fragments was closest to the expected size but was the less prominent of the two. Pairing the long 5' region to the short 3' region for PCR amplification was expected to yield a 2028 bp fragment (primers #8 and #7 in Figure 16).

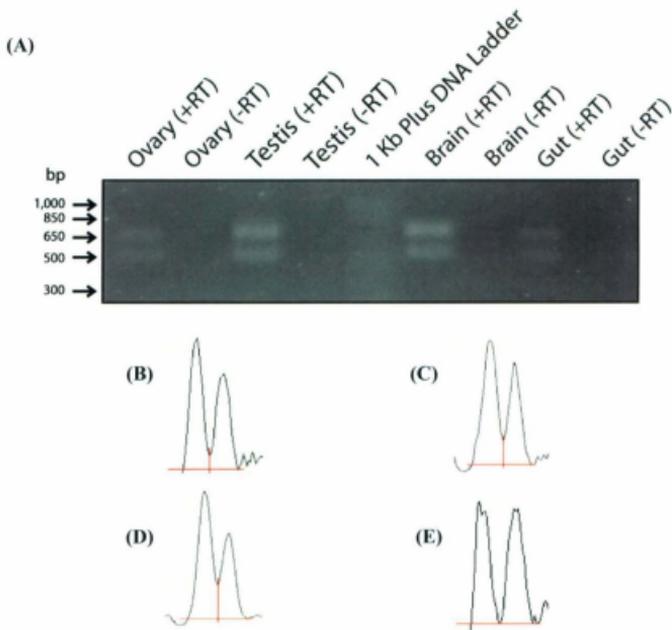


Figure 15: Both the 703 bp and 514 bp 3' transcript variants of the *L. ocellata Dnmt3* are present in four different tissue types. (A) RT-PCR products were amplified using primers #3 and #4 working from template synthesized using Invitrogen's AncT primer and 1 μ g of total RNA. Absence of PCR amplification in the -RT reactions indicates no genomic DNA contamination. (B) Graphical representation of band intensities from Lane 1 (ovary +RT) generated by Scion Image software. (C) Band intensities from Lane 3 (testis +RT). (D) Band intensities from Lane 6 (brain +RT). (E) Band intensities from Lane 8 (gut +RT). Red lines indicate the boundaries set allowing pixel counts beneath each curve to be obtained separate from pixel counts of adjacent curves. Peaks on left side of each individual image correspond to the 703 bp band while peaks on the right correspond to the 514 bp band. See Table 3 for square pixel counts and calculated ratios.

Table 4: Tissue specific differential expression of the *L. ocellata Dnmt3 3'* transcript variants. Square pixel counts were determined using Scion Image software to analyze a tiff image of RT-PCR products separated by gel electrophoresis (see Figure 10). Ratios were determined by calculating the band intensity of the 703 bp product relative to the 514 bp product. PCR amplification was carried out using primers #3 and #4. Pixel counts were taken from the centers of the electrophoresis bands.

Tissue	Ovary	Testes	Brain	Gut
703 bp Square Pixels	2513	4400	4245	1804
514 bp Square Pixels	2030	2934	2667	1935
Ratio	1.238	1.500	1.600	0.9323

Sample Calculation

$$\text{Ovary: } 703\text{bp pixel count} / 514\text{bp pixel count} = 2513 / 2030 = 1.238$$

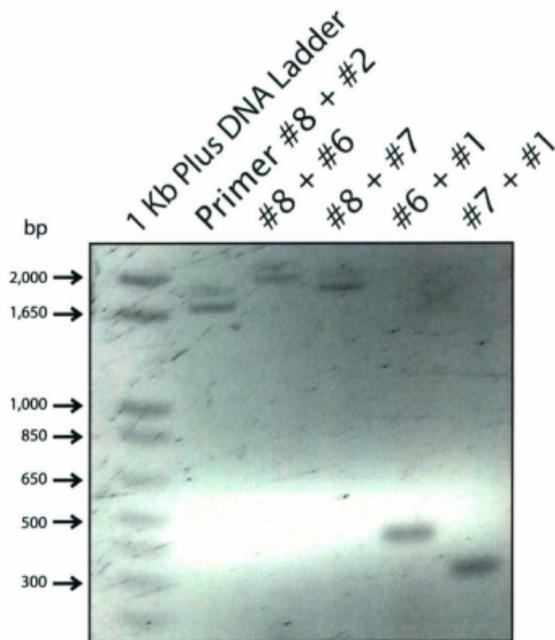


Figure 16: The long 5' end of the *L. ocellata Dnmt3* transcript matches with both the long and short 3' ends. Primer #8 isolates transcripts with extended 5' ends while primer #2 serves as a landmark to ensure #8 is annealing where it was designed to. Primers #6 and #7 isolate the long and short 3' ends respectively while primer #1 acts as a landmark ensuring both #6 and #7 anneal where they were designed to.

Again, two fragments resulted from amplification only this time the fragment sizes were ~1900 bp and ~2000 bp. The largest fragment was closest to the expected size and also appeared far less prominent than the smaller fragment. PCR amplification using primer #8 paired with #2 was performed to act as a landmark ensuring #8 was annealing where it was designed to (primer #2 had been previously shown to anneal to its designed target sequence; data not shown). This reaction was expected to amplify a single 1910 bp fragment. As seen in Figure 16 two fragments amplified with the larger, less prominent one being ~1900 bp and the smaller of the two being ~1700 bp.

Pairing the short 5' region to the long 3' region for PCR amplification was expected to yield a 2085 bp fragment (primers #9 and #6 in Figure 17). A single fragment amplified at ~1900 bp. Pairing the short 5' region to the short 3' region for PCR amplification was expected to yield a 1955 bp fragment (primers #9 and #7 in Figure 13). A single fragment amplified at ~1800 bp. PCR amplification using primer #9 paired with #2 was also performed to act as a landmark ensuring #9 was annealing where it was designed to. This reaction was expected to amplify a 1836 bp fragment but yielded a fragment of ~1650 bp (as seen in Figure 13).

*Evidence of *L. ocellata* Dnmt3 pseudogenes*

Initial PCR amplifications, prior to the use of DNA Wipeout treatments for total RNA samples, in the “no-RT” controls amplified cDNA fragments matching the 3' end

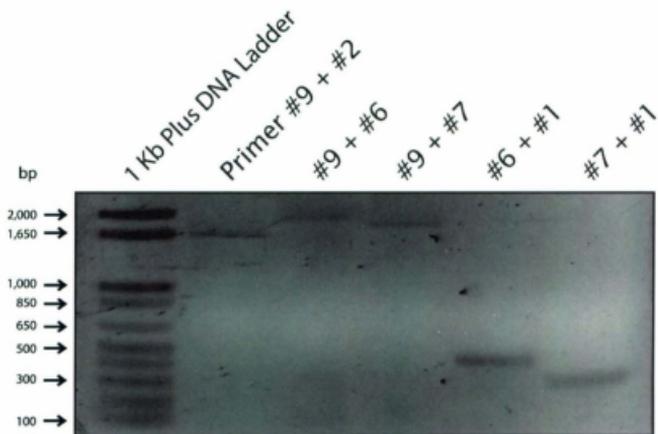


Figure 17: The short 5' end of the *L. ocellata Dnmt3* transcript matches with both the long and short 3' ends. Primer #9 isolates transcripts with spliced 5' ends while primer #2 serves as a landmark to ensure #9 is annealing where it was designed to. Primers #6 and #7 isolate the long and short 3' ends respectively while primer #1 acts as a landmark ensuring both #6 and #7 anneal where they were designed to.

doublet seen in Figure 15. This result suggested DNA contamination. Using gDNA treated with RNases as the PCR template resulted in 705 bp and 514 bp amplicons consistent with the previous reverse transcribed cDNA amplification results (Figure 18). Using the RNase treated/DNA Wipeout treated gDNA as PCR template yielded no amplicons indicating that the 705 bp and 514 bp doublet was indeed amplifying from the intact gDNA template. The 705 bp and 514 bp gDNA PCR products were cloned and sequenced (Figures 19 and 20 respectively). BLASTn alignment against the corresponding 703 bp and 514 bp cDNA PCR products revealed a 99% identity match in each instance (Figures 21 and 22 respectively).

Having shown that both the 703 bp and 514 bp *L. ocellata* cDNA amplicons match two same sized amplicons derived from gDNA I wanted to investigate this region in representative mammalian and teleost *Dnmt3* sequences (*H. sapiens*, *M. musculus* and *D. rerio*; refer to Materials and Methods for GenBank numbers). A ClustalW [Larkin *et al.* 2007] alignment of the single *L. ocellata*, two *H. sapiens*, two *M. musculus* and six *D. rerio de novo* DNA methyltransferases was performed. The entire *L. ocellata* cDNA sequence was used to preserve the structure of the alignment since the region in question

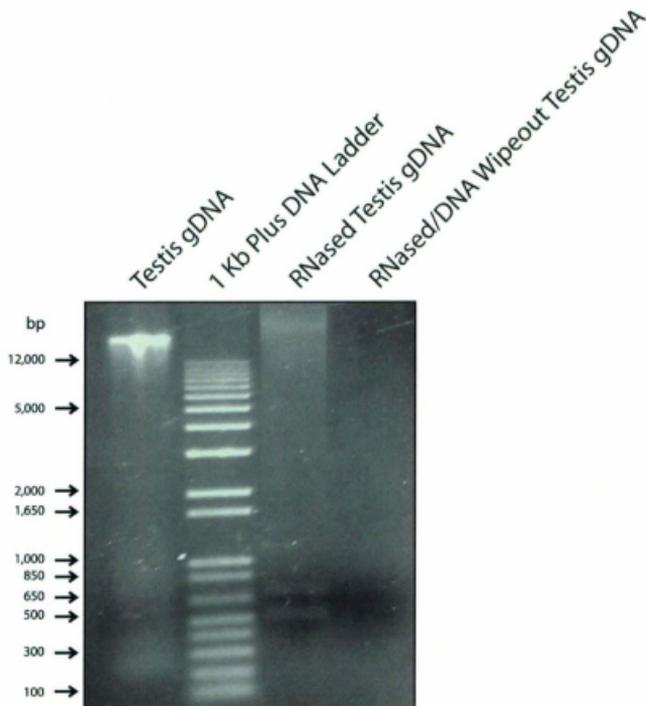


Figure 18: Both the long and short 3' ends detected in the *L. ocellata Dnmt3* transcript are present at the genomic level. PCR amplification was carried out using primers #3 and #4. RNase treatment degrades RNA ensuring that any amplification that does occur is due to a genomic DNA (gDNA) template. DNA wipeout removes the gDNA template revealing that amplification was due to a gDNA template and not contaminating mRNA. The first lane on the left reveals the integrity of the gDNA sample prior to RNase or DNA wipeout treatments.

TGAGGGCACAGGAAGACTTTTCTTTGAATTTTATCGACTGCTTCATGACACCAGACCAA
AGGAGTGGGAAGACAGGCCTTCTTCTGGTTATTTGAGAATGTTGTTGCCATGGGTGTC
AATGATAAAAGGGACATCTCACGTTTCTTGGAGTGTAACCCAGTTATGGTCGATGCAAT
CGACGTGTCTGCTGCCACCGGCTCGCTGCTTTTGGGGAAACTTACCAGGAATGAACA
GGCCCCTGGTTGCTTCTTCAGCAGATAAACTGGAAGTGCAGCACTGCCTTGAGCATGGT
AGGATAGCAAAGTTTAGCAAAGTTAGGACCATAACAACAAGGTCGAACTCCATCAAGCA
GGGCAAGGATCAACATTTCCAGTTATCATGAATGGGAAGGAAGACATTCTGTGGTGTA
CAGAACTGGAGAGGATCTTTGGCTTCCAGTGCCTACACAGACGTGTCGAACATGGGG
AGAGGAGCTCGGCAGAACTTCTTGGAAAGTCTTGGAGTGTGCCTGTCAATTCGCCACCT
ATTTGCACCTCTGAAGGATTATTTTGGCTTGTGAATAACACAGCATATAATAGGTCTTTC
AGAAACTTACGGTGCTCTCTTAGAAAATAACAAGTAGTGAACAATAATCCAGACTGGTT
TTCAGTATACTGTGACACAAAATCATTTGCTTGGCATTATTGTGAAGGCTGCATGC

Figure 19: 705 bp of *L. ocellata* gDNA amplified using primers #3 and #4. Primers were originally designed to amplify sequence in the catalytic region of *L. ocellata Dnmt3* reverse transcribed mRNA. The gDNA amplicon is very similar in size to the cDNA amplicon generated under the exact same PCR conditions.

TGAGGGCACAGGAAGACTTTTCTTTGAATTTTATCGACTGCTTCATGACACCAGACCAA
AGGAGTGGGAAGACAGGCCTTCTTCTGGTTATTTGAGAATGTTGTTGCCATGGGTGTC
AATGATAAAAGGGACATCTCACGTTTCTTGGAGTGTAACCCAGTTATGATCGATGCAAT
CGACGTGTCTGCTGCCACCGGCTCGCTGCTTTGGGGAAACTTACCAGGAATGAACA
GGATCTTTGGCTTTCAGTGCACTACACAGACGTGTCGAACATGGGGAGAGGGGCTCGG
CAGAAACTTCTTGAAGGTCTTGGAGTGTCCCTGTCATTGCCACCTATTTGCACCTCT
GAAGGATTATTTGCTTGTGAATAACACAGCATATAATAGTCTTTTCAGAAACTTACGG
TGCTCTCTTAGAACTAACAAGTAGTGAACAATCCAGACTGGTTTTTCAGTATACTGT
GACACAAAATATTTGCTTGGCATTATTGTGAAGGCTGCATGC

Figure 20: 514 bp of *L. ocellata* gDNA amplified using primers #3 and #4. Primers were originally designed to amplify sequence in the catalytic region of *L. ocellata* *Dnmt3* reverse transcribed mRNA. The gDNA amplicon is very similar in size to the cDNA amplicon generated under the exact same PCR conditions.

(A)

```

cDNA 1   TGAGGGCACAGGAAGACTTTTCTTTGAATTTTATCGACTGCTTCATGACACCAGACCAAA 60
      |||
gDNA 1   TGAGGGCACAGGAAGACTTTTCTTTGAATTTTATCGACTGCTTCATGACACCAGACCAAA 60
      |||

cDNA 61   GGAGTGGGAAGACAGGCCCTTCTCTCGGTTATTTGAGAATGTTGTGCCATGGGTGTCAA 120
      |||
gDNA 61   GGAGTGGGAAGACAGGCCCTTCTCTCGGTTATTTGAGAATGTTGTGCCATGGGTGTCAA 120
      |||

cDNA 121  TGATAAAAGGGACATCTCACGTTTCTTGGAGTGTAAACCAGTTATGATCGATGCAATCGA 180
      |||
gDNA 121  TGATAAAAGGGACATCTCACGTTTCTTGGAGTGTAAACCAGTTATGATCGATGCAATCGA 180
      |||

cDNA 181  CGTGTCTGCTGCCACCGGGCTCGCTGCTTTTGGGAAACTTACCAGGAATGAACAGGCC 240
      |||
gDNA 181  CGTGTCTGCTGCCACCGGGCTCGCTGCTTTTGGGAAACTTACCAGGAATGAACAGGCC 240
      |||

cDNA 241  CCTGGTTGCTTCTTCAGCAGATAAACTGGAACCTGCAGCACTGCCTTGAGCATGGTAGGAT 300
      |||
gDNA 241  CCTGGTTGCTTCTTCAGCAGATAAACTGGAACCTGCAGCACTGCCTTGAGCATGGTAGGAT 300
      |||

cDNA 301  AGCAAAAGTTTAGCAAAGTTAGGACCATAACAACAAGSTCGAACTCCATCAAGCAGGGCAA 360
      |||
gDNA 301  AGCAAAAGTTTAGCAAAGTTAGGACCATAACAACAAGSTCGAACTCCATCAAGCAGGGCAA 360
      |||

cDNA 361  GGATCAACATTTCCAGGTTATCATGAATGGGAAGGAAGACATTCGTGGTGTACAGAACT 420
      |||
gDNA 361  GGATCAACATTTCCAGGTTATCATGAATGGGAAGGAAGACATTCGTGGTGTACAGAACT 420
      |||

cDNA 421  GGAAGGATCTTTGGCTTTCCAGTGCCTACACAGACGTGTGGAACATGGGAGAGGAGC 480
      |||
gDNA 421  GGAAGGATCTTTGGCTTTCCAGTGCCTACACAGACGTGTGGAACATGGGAGAGGAGC 480
      |||

cDNA 481  TCGGCAGAAACTTCTTGGAAAGTCTTGGAGTGTGCTGTCAATTCGCCACCTATTTGCACC 540
      |||
gDNA 481  TCGGCAGAAACTTCTTGGAAAGTCTTGGAGTGTGCTGTCAATTCGCCACCTATTTGCACC 540
      |||

cDNA 541  TCTGAAGGATTATTTGCTTGTGAATAACACAGCATATAATAGGCTTTTCAGAACTTAC 600
      |||
gDNA 541  TCTGAAGGATTATTTGCTTGTGAATAACACAGCATATAATAGGCTTTTCAGAACTTAC 600
      |||

cDNA 601  GGTGCTCTCTTAGAAA-CTAACAAAGTAGTGAACAAATCCAGACTGGTTTTTCAGTATACT 659
      |||
gDNA 601  GGTGCTCTCTTAGAAAATAACAAGTAGTGAACAAATCCAGACTGGTTTTTCAGTATACT 660
      |||

cDNA 660  GTGACACAAAAT-ATTGCTTGGCATTATTGTGAAAGCTGCATGC 703
      |||
gDNA 661  GTGACACAAAATCATTGCTTGGCATTATTGTGAAAGCTGCATGC 705
      |||

```

(B) Score = 1279 bits (692), Expect = 0.0
Identities = 701/705 (99%), Gaps = 2/705 (0%)
Strand=Plus/Plus

Figure 21: Alignment of *L. ocellata* 3' region of cDNA and gDNA. **(A)** BLASTn alignment showing two mismatches and two additional nucleotides in the gDNA sequence. **(B)** Data readout from the BLASTn alignment indicating 99% identity matches between the cDNA and gDNA sequences.

(A)

```

cDNA 1   TGAGGGCACAGGAAGACTTTCTTTGAAATTTATCGACTGCTTCATGACACCAGACCAAA 60
      |||
gDNA 1   TGAGGGCACAGGAAGACTTTCTTTGAAATTTATCGACTGCTTCATGACACCAGACCAAA 60

cDNA 61   GGAGTGGGAAGACAGGCCCTTCTCTCGGTATTTGAGAATGTTGTGCCATGGGTGTCAA 120
      |||
gDNA 61   GGAGTGGGAAGACAGGCCCTTCTCTCGGTATTTGAGAATGTTGTGCCATGGGTGTCAA 120

cDNA 121  TGATAAAAGGGACATCTCACGTTTCTTGGAGTGTAAACCAGTTATGATCGATGCAATCGA 180
      |||
gDNA 121  TGATAAAAGGGACATCTCACGTTTCTTGGAGTGTAAACCAGTTATGATCGATGCAATCGA 180

cDNA 181  CGTGTCTGCTGCCACCAGGCTCGCTACTTTGGGAAACTTACCAGGAATGAACAGGAT 240
      |||
gDNA 181  CGTGTCTGCTGCCACCAGGCTCGCTACTTTGGGAAACTTACCAGGAATGAACAGGAT 240

cDNA 241  CTTTGGCTTCCAGTGCACTACACAGACGTGTGAAACATGGGGAGAGGAGCTCGGCAGAA 300
      |||
gDNA 241  CTTTGGCTTCCAGTGCACTACACAGACGTGTGAAACATGGGGAGAGGAGCTCGGCAGAA 300

cDNA 301  ACTTCTTGAAGGCTCTGGAGTGTGCCTGTCAATCGCCACCTATTTCACCTCTGAAGGA 360
      |||
gDNA 301  ACTTCTTGAAGGCTCTGGAGTGTGCCTGTCAATCGCCACCTATTTCACCTCTGAAGGA 360

cDNA 361  TTATTTTGCTTGTGAATAACACAGCATATAATAGGTCCTTTCAGAAACTTACGGTGCTCTC 420
      |||
gDNA 361  TTATTTTGCTTGTGAATAACACAGCATATAATAGGTCCTTTCAGAAACTTACGGTGCTCTC 420

cDNA 421  TTAGAAACTAACAGTAGTGAACAATAACAGACTGGTTFTCAGTATACTGTGACACAAA 480
      |||
gDNA 421  TTAGAAACTAACAGTAGTGAACAATAACAGACTGGTTFTCAGTATACTGTGACACAAA 480

cDNA 481  ATATTTGCTTGGCATTATTGTGAAGGCTGCATGC 514
      |||
gDNA 481  ATATTTGCTTGGCATTATTGTGAAGGCTGCATGC 514

```

(B) Score = 939 bits (508), Expect = 0.0
 Identities = 512/514 (99%), Gaps = 0/514 (0%)
 Strand=Plus/Plus

Figure 22: Alignment of *L. ocellata* spliced 3' region of cDNA and gDNA. **(A)** BLASTn alignment showing two mismatches between the cDNA and gDNA sequences. **(B)** Data readout from the BLASTn alignment indicating 99% identity matches between the cDNA and gDNA sequences.

did not differ between cDNA and gDNA sequences. Figure 23 highlights the 703 bp region of *L. ocellata* cDNA that corresponds to the equivalent 705 bp gDNA region. In this region that is free of introns in the *L. ocellata* genome there is a minimum of one (*D. rerio dnmt8*) and a maximum of five (*M. musculus Dnmt3a*) post-transcriptional splice sites present in all the cDNA sequences aligned (Table 5).

Table 5: Tally of post transcriptional splice sites. Splice sites detected by comparison of genomic and cDNA sequences found in GenBank (NCBI). Tallied splice sites of the aligned vertebrates occurred within the 703 bp region of the *L. ocellata* intron-less *Dnmt3* isolated from gDNA.

Dnmt	Human		Mouse		Zebrafish					
	3a	3b	3a	3b	3	4	5	6	7	8
# of intron/exon splice sites present in 703bp region	4	2	5	4	4	4	4	4	4	1

Zf7 -----
 Sd3 -----
 mouse3b -----
 Human3b -----
 Zf4 -----
 mouse3a -----
 Human3a TCTAGAACGTTCCAGTGCAGCTGTGGAAATGTGGCTAGGTGTAATTCAGCTTCTCTGG 2717
 Zf6 -----
 Zf3 -----
 Zf5 -----
 Zf8 -----

Zf7 -----TGC AATCTGTGATGATTG 3945
 Sd3 -----TGT AACCCAGTTATGATCG 2156
 mouse3b -----TGT AACCCAGTGAATCG 2450
 Human3b -----TGT AATCCAGTGAATG 2649
 Zf4 -----TGT AATCCAGTGAATG 2121
 mouse3a -----TCT AACCCCTGTGATGATTG 2545
 Human3a GAGGCTGCAGGCTAGCCAGTGTGTGGCTCCTGAGAGAGAATCC AACCCCTGTGATGATTG 2777
 Zf6 -----TGC AACCCAGTGAATG 2114
 Zf3 -----TGT AACCCCTGTGCTGTGG 4130
 Zf5 -----TGT AACCCCTGTGCTAATTG 3626
 Zf8 -----TGT AATCCAGTCAATGATCG 2198
 * * * * *

Zf7 ATGCTGTAAAAAGTAAGCCCA--GCCACAGGGCCAGTACTTCTGGGGGAATTTACCTGG 4003
 Sd3 ATGC AATCGAGCTGTCTGCT--GCCACCGGGCTCGCTACTTTGGGGGAATTTACCAAG 2214
 mouse3b ATGCCATCAAGGTTCTGCT--GCTCACAGGGCCCGGTACTTCTGGGGTAACCTACCCGG 2508
 Human3b ATGCCATCAAGTTTCTGCT--GCTCACAGGGCCCGATACTTCTGGGGCACTACCCGG 2707
 Zf4 AGCCCATAGAGGTTCTGGCT--GCTCACAGGGCCAGCTATTCTGGGGCAACTGCCAGG 2179
 mouse3a ACGCCAAAGAAAGTGTCTGCT--GCACACAGGGCCCGTACTTCTGGGGTAACCTTCTGCG 2603
 Human3a ATGCCAAAGAAAGTGTCACT--GCACACAGGGCCCGTACTTCTGGGGTAACCTTCCCGG 2835
 Zf6 ATGCCAAAGAAAGTGTCTGCA--GCCACAGAGCCAGCTATTTTGGGGGAACCTGCTCG 2172
 Zf3 ATGCCGTGAAAGTG--AGTCCAGCTCACAGAGCAAGATACTTCTGGGGGAACATACTCG 4188
 Zf5 AGCCTGTGAAAGTG--AGTCCAGCTCACAGAGCAAGATACTTCTGGGGGAACATACTCG 3684
 Zf8 ATGCTAAGGAGGTTCTGACCC--GCACACAGAGCTCGATACTTCTGGGGGAACCTTCTCG 2256
 * * * * *

Zf7 AATGAACCGACCTGTGTGGACTTCTCTCAC-TGACAATGTAGATCTGCAGGACTGCCTGG 4062
 Sd3 AATGAACAGGCCCTGTGTGCTTCTTCAGC-AGATAAATGGAGCTGCAGCAGCTGCCTGG 2273
 mouse3b AATGAACAGGCCCTGTGTGCTTCTTCAGC-AGATAAATGGAGCTGCAGCAGCTGCCTGG 2567
 Human3b GATGAACAGGC----- 2117
 Zf4 AATGAAGAGGCCTCTGTGTGCTTCTGGGAT--GGATAAATYAGAGCTTCAGGACTGTTGG 2238
 mouse3a CATGAACAGGCCCTTGTGGCATCCACTGTGAA-TGATAAGCTGGAGCTGCAGGAGTGTCTGG 2662
 Human3a TATGAACAGGCCCTTGTGGCATCCACTGTGAA-TGATAAGCTGGAGCTGCAGGAGTGTCTGG 2894
 Zf6 CATGAATAGACCAATGTCTGCCA-TGTGCCACTGATAAAGTGTGATCTTCAAGACTGTTTGG 2231
 Zf3 CATGAACAGACC AATCATAGCATCTCAGAA-TGATAAAGCTGTGCTTCAAGAAATGTCTGG 4247
 Zf5 CATGAACAGACC AATCATAGCATCTCAGAA-AGATAAAGTCAAGTCTTCAAGACTGTTTGG 3743
 Zf8 GATGAATCGGCCACTGACTGCTATGTGTGAA-TGACAAACTCGATCTGCAGGACTGTCTGG 2315

Zf7 AATCTGGACGAACCGCAATGTT CAGCAAAAGTTCGCACCATCACCACCAAGTCCAATTC 4122
 Sd3 AGCATGCTAGGATAGCAAAGTTAGCAAAGTTAGGACCAATAACAAACAGGTCGAATCCA 2333
 mouse3b AGTTCAGTAGGACAGCAAAGTTAAGCAAAGTGCAGACAATAACCCAAAGTCCAATCCA 2627
 Human3b -----

Zf4 AACATGGCCGCTTGGCAAGTTCCGGGAAGGTACGTACCATCACAAACAGCTCCAATTC 2298
 mouse3a AAGCAGCGCAGAAATAGCCAAAGTTCAGCAAAGTGGAGGACCAATACCACAGGTCAAACTCTA 2722
 Human3a AGCATGGCAGGAATGCCAAGTTCAGCAAAGTGGAGGACCAATCACTACTGGGTGAAATCCA 2954
 Zf6 AAGCATGGCAGGACAGCTAAGTTGGTAAAGTGGGACCAATCACTACTGGGTGAAATCCA 2291
 Zf3 AGCTTGGCCGCTGCAAAAGTATGAAAAAGTTCGCACTATCACTACACGGGCAAACTCCC 4307
 Zf5 ATGGAGGCCGCACTGCCAAGTATGAAAAAGTACGCACTTATCAACACGGGCAAACTCAC 3803
 Zf8 AACATGGCAGCAGCTAAGTTCATAAAGTGGGACCAATCACAACTGGTCAAATCCA 2375

Zf7	-----TGAAATTGATGTTATTTTAAAGCTTCTGCAAAATTCACITTCACAGAATATTTTAAACA	4459
Sd3	----- CTTACGGTGTCTCTCT -----TAGAAA----- CTAACCA	2608
mouse3b	-----CACCCCTCCCTGAAGGCACCTCACCTGTGCCCTTTTAGCT	2936
Human3b	-----GTGTGATTCCTGAAGGCATCC--CCAGGCCGTGTCTCTPCT	2951
Zf4	-----CAAATGCTCTCAAAAA-----GGAGCTCTTAGAGACC	2601
mouse3a	-----CAAAAAC--AAAAACAAA-----CAATAAAAACCAAGA	3032
Human3a	-----AAAAAC--AAAAACA-----TAATAAAAACCAAGA	3255
Zf6	-----AAAAATC--GAAATA-----TGATTTAAACCAAGA	2564
Zf3	-----GAATTC-----CCTGTG	4540
Zf5	-----GAAATC-----CCTGCA	4036
Zf8	-----GAAGCC-----CCGCC	2606
Zf7	TTTCATTAGTTTCCAATATGAGAAGCTGTCTCAAAATCCTGGTCAGTGGTCTT-----	4514
Sd3	----- AGTAGTGA-AAC ----- AAATCCAGAC ----- TGGTTTTCT -----	2637
mouse3b	-----CACCTGTGTGGGG-----TCACATCACCTGTACCTC-----	2968
Human3b	-----CAGCTGTGTGGG-----TCATACCG-TGTACCTC-----	2979
Zf4	-----CACAAATTTGA-----CAAGACC-----ATCTC-----	2624
mouse3a	A-----CGAGGACGGAGA-----AAAGT-TCAGCACCCAGAAGA	3067
Human3a	A-----CATGAGATGGAGA-----GAGGTATCAGCACCCAGAAGA	3291
Zf6	G-----CATCAAAAC--RAGC-TCT-TATCT-----	2587
Zf3	-----A--AATGA-----AGA-----GAGGCTACG--	4558
Zf5	-----A--AATGA-----AGA-----GAAATTT--	4052
Zf8	-----GCCAGTGC-----ACC-----ATGACATC-----	2625
Zf7	-----GGAATGCACAGACACTTTTGGATCTGATTTTTAAAGTCATAAGTACTGTG--	4565
Sd3	----- AGTATACTGTGACAC ----- AAAAAT-AT ----- TTG -----	2662
mouse3b	-----AGCTTTCCTCTGCTC-----AGT-GGGAGCAGAGCCT--	2999
Human3b	-----AGTTCCTCTCTGCTC-----AGT-GGGGGCAGAGCCA--	3010
Zf4	-----AGCTTTGGAAAAACA-----AGC-ATGAG--AAGCTG--	2653
mouse3a	GAAAAAGGAATTTAAAGCAAA-----CCACAGAGGAGGAAAACGGCCGAGG	3113
Human3a	GAAAAAGGAATTTAAUACAAAA-----CCACAGAGGCGGAAA-YACCAGGAGG	3338
Zf6	-----GTTATTTGCAACA-----	2600
Zf3	-----	
Zf5	-----	
Zf8	-----	
Zf7	-TTTTTAACTGTAACGTTGGAGAAAAACACAGATTCTCCACCCCGCTAECTCAGTGCCA	4624
Sd3	- CTTGGCATTAT ----- TGTGAAGG ----- CTGCATGCTG-TACATT--TACGA	2702
mouse3b	-CCTGGCCCTTG-----CAGGGG--AGCCCC-GGTGCTCCCTCGGTGTCACAGCTCAGA	3050
Human3b	-CCTGACTCTTG-----CAGGGGT-AGCTGAGGTGCTCCGCTCTTGTGC-ACAATCAGA	3063
Zf4	----TGACTTGAAG--ACAGGGT-AGTGTGAGCAACAGGCC--AAACTCAAGCTAGA	2703
mouse3a	GCTTGGCCTTGC--AAAGGG--TTGGACATCATCTCCTGATTTTCAATGTTA	3163
Human3a	GCTTGGCCTTGG--AAAGGG--TTGGACATCATCTCCTGATTTTCAATGTTA	3389
Zf6	-----TCT-----AAAACGG-----T--ATTTCATGT--TGGTTTCTATTTTTCA	2636
Zf3	-----ATGAGGA-----TTTTAATCTAAT--TTGGCTCAATGTAAAAA	4594
Zf5	-----ATGAAGC-----ATTTCGTTTATTC--TTCCTCTTGTGAAAAA	4089
Zf8	-----ATCAA-----CTCCAACCTTCTGTGTCTTCAATGTGTG--	2659
Zf7	TTCTGTGATTGGTCTTGTCTGGAGCAATGGCTATTTATAATTGAATGAATAGTTAAC	4684
Sd3	TACTGTGGCCA -----TTAATTCACAG-----TCACACTCA--GGACAGAGTAGG	2745
mouse3b	C-CTG-G-----CTGCTTAGAG-----TAGCCCG--GCATGGTGCT	3082
Human3b	C-CTG-G-----CTGCTTAGAG-----CAGCCCTA-----ACACGGTGCT	3095
Zf4	CACCTGTG-----ATGCTAAGAG-----CGA-AA--ACACAGT--	2732
mouse3a	ACCTTCAGTC-----CTATCTAAAA-----AGCAA-----AATAGCC-CC	3197
Human3a	TTCTTCAGTC-----CTAATTTAAAA-----ACAAA-----ACCAGCTCC	3424
Zf6	TCATTTCTGT-----TTTGGAA-----T--A--ATTTG--CT	2661
Zf3	G-----ACTCAGGG-----ACACACAAGT	4613
Zf5	G-----ACTCAGGG-----CCACAGTAGT	4108
Zf8	-----CTTGTGG-----CCTCTGAAC--	2675

Figure 23: Alignment of *L. ocellata Dnmt3* 3' region with *Dnmt3* cDNA of other vertebrate models. (A) Section of ClustalW alignment using the full length *L. ocellata Dnmt3* cDNA (Sd3) with *H. sapiens DNMT3a* GenBank [AB208833](#), *H. sapiens DNMT3b* GenBank [AB208880](#), *M. musculus Dnmt3a* GenBank [AF068625](#), *M. musculus Dnmt3b* GenBank [AF068626](#), *D. rerio dnmt3* GenBank [AB196914](#), *D. rerio dnmt4* GenBank [AB196915](#), *D. rerio dnmt5* GenBank [AB196916](#), *D. rerio dnmt6* GenBank [AB196917](#), *D. rerio dnmt7* GenBank [AB196918](#), *D. rerio dnmt8* GenBank [AB196919](#). Region of *L. ocellata* sequence corresponding to the 703 bp in question is highlighted in green text. Post transcriptional splice sites of remaining sequences are preceded by blue text and followed by red text.

Discussion

Study of the class Chondrichthyes, subclass Elasmobranchii, provides an excellent opportunity to understand DNA methylation and its role in managing the vertebrate genome. The position of subclass Elasmobranchii in the evolutionary record makes it an ideal choice to act as a representative of the ancestral vertebrate condition linking the teleost and tetrapod lineages. Extensive studies using both mammalian and teleost models have shown how integral proper chromatin management is to the complete and successful development of an organism. Specifically, DNA methylation of the genome and controlled dynamic methylation adjustments have been shown to be vital in maintaining normal, wild type development. Both the teleost and mammalian lineages appear to use very similar proteins to establish and maintain this epigenetic marker. In addition both lineages display similar strategies in the hypermethylation of paternal germ cell genomes relative to those of maternal germ cells and, perhaps more importantly, both show a genome-wide demethylation/remethylation cycle following fertilization. These similarities are important for understanding the evolution of this process but the role played by variations in the numbers of DNA methyltransferase genes remains unclear. Although both lineages demonstrate similar relative germ cell DNA methylation levels and post-fertilization dynamics, they utilize two very different reproductive strategies (external ovipary versus internal vivipary). Expanding the investigation of vertebrate chromatin management into the subclass Elasmobranchii may help shed light on the significance of these differences and similarities.

This study investigates DNA methylation in *Leucoraja ocellata* (the Winter Skate) found off the shores of Newfoundland and Labrador. *L. ocellata* is from the subclass Elasmobranchii, order Rajiformes and family Rajidae. As it is a member of the family Rajidae its reproductive strategy is single oviparity, that is, eggs are deposited one at a time on the rocky substrate and embryonic development occurs externally from the mother [Bester 2011]. Having previously shown that *L. ocellata* methylates its genome [Lake 2008] I proceeded with an investigation into relative germ cell DNA methylation levels to determine if the trends observed in zebrafish and mammals were conserved. This would add further validity to the use of *L. ocellata* as a representative ancestral vertebrate while studying genomic methylation. In order to test this genomic DNA from ovary, soma and testis tissues were subjected to *HpaII* restriction digestions. *HpaII* targets 5'-CCGG-3' sequences and creates a single strand nick between the cytosine nucleotides of each strand of DNA resulting in two fragments, each ending with a 3' GC overhang. In the presence of a methyl-group covalently bonded to the cytosine of the internal CpG dinucleotide *HpaII* is prevented from cutting at that specific target sequence. Undigested genomic DNA runs as a single large sized band (>12,000 bp) when separated by gel electrophoresis. Both ovary and soma DNA were digested by the *HpaII* endonuclease; hence displaying the continuous gradient from very large to relatively small DNA fragments separated via gel electrophoresis indicative of large scale restriction digestion. Although both samples were extensively digested, there was the persistence of very large DNA fragments indicating that not all potential *HpaII* target sequences had been cleaved. This suggested that there was some level of DNA

methylation present in these two tissue types. Comparatively, the majority of the *HpaII*-digested DNA from the testis samples migrated as a single band of a large fragment size. This drastic reduction in digestion can be attributed to the prevention of *HpaII* gaining access to its target sequence via extensive genomic CpG methylation. There was also a continuous smear of DNA fragments in the digested testis samples indicating that at least some of the *HpaII* target sequences were unmethylated, but at lower levels than those seen in the other two tissue types.

It should be noted that these results are applicable to DNA methylation levels of reproductive tissues and not pure germ cell populations. As there are a number of contributing cell types to each tissue it can not be clearly stated that *L. ocellata* sperm are hypermethylated relative to oocytes. However, these results do reflect a trend in paternal hypermethylation relative to maternal methyl levels regarding reproductive structures and provides a jumping off point for the remainder of this study which focuses primarily on the DNA-methyltransferases of *L. ocellata*.

L. ocellata Dnmt1: structure

This study reports the isolation and identification of a 4990 nucleotide (nt) *L. ocellata* DNA-methyltransferase 1 cDNA sequence (see Figure 7) building upon the previous 2669 base pairs (bp) acquired by the McGowan laboratory. Comparisons of the sequence to those present in the nucleotide sequence database (by National Center for

Biotechnology Information BLASTn) using the full length *L. ocellata* cDNA as the query returned no significant similarities. Performing a DeCypher Tera-BLASTN search using the full length *L. ocellata Dnmt1* cDNA sequence to probe the skate EST database (http://decypher.mdibl.org/decypher/algo-tera-blast/tera-blastn_nn.shtml) resulted in five hits all originating from *Leucoraja erinacea* (the Little Skate) and all scoring 99% similarity (data not shown). The longest EST returned in the search was 831 nt in length and spanned both bromo-adjacent homology domain coding regions (protein domains are discussed below). The remaining four ESTs overlapped with the first to varying degrees and spanned regions of the *L. ocellata Dnmt1* cDNA coding for portions of the methyltransferase catalytic region as well as the Replication Foci domain (data not shown). However, the *L. ocellata* clone presented in this study is the first instance of a complete *Dnmt1* cDNA being reported for a member of the subclass Elasmobranchii.

The nucleotides directly adjacent to the translation initiation codon have been implicated in the 43S pre-initiation complex's (40S ribosomal subunit + eIF1, eIF1 α and eIF3) efficient recognition of the AUG start codon [as reviewed by Kozak 2005]. This sequence consisting of 5'-GCCRCCaugG-3' (*R: A or G*), termed the Kozak consensus sequence [Kozak 1987], is >10-fold more efficient than a simple AUG start codon at binding the 43S pre-initiation complex as it scans the mRNA transcript in the 5' to 3' direction. Of the entire Kozak sequence it has been determined that the third nucleotide upstream (-3*R*), as well as the fourth nucleotide downstream (+4*G*) from the adenine of the start codon are the most important nucleotides involved in translation initiation aside from the

start codon itself [Nakagawa *et al.* 2007]. Although both the -3R and +4G nucleotides play important roles in translation initiation the effect of +4G alone is minor compared to that of -3R alone. The sequence surrounding the start codon of the *L. ocellata Dnmt1* was compared to the Kozak consensus sequence revealing a weak similarity consisting only of the -3R nucleotide (see Table 1). For the purposes of this study all *Dnmt1* sequence comparisons were carried out using well known scientific models as representatives for the teleost (*Danio rerio*) and tetrapod (*Xenopus laevis*, *Gallus gallus*, *Mus musculus* and *Homo sapiens*) lineages. Of the sequences examined all but *G. gallus* showed weak similarity to the Kozak sequence with only the -3R nucleotide conserved. *G. gallus* showed a poor degree of similarity with neither the -3R nor the +4G nucleotides conserved. The degree of conservation of the Kozak sequence in the *L. ocellata Dnmt1* may only be partial but it is consistent with other major vertebrate models and does preserve the -3R nucleotide known to be the more instrumental of the two key nucleotide locations.

The open reading frame (ORF) of *L. ocellata Dnmt1*, as reported by Gene Runner software, was 1509 amino acids (aa) in length (see Figure 8), well within the expected size range of this protein type (*X. laevis Dnmt1*: 1490 aa – *M. musculus Dnmt1*: 1620 aa). Comparisons to on-line protein databases (National Center for Biotechnology Information BLASTp) reported an average of 85% similarity when compared to the above mentioned vertebrate model organism *Dnmt1* proteins. The highest degree of similarity was seen in *H. sapiens* (88%) although the percentage of the *L. ocellata Dnmt1*

that aligned with the *H. sapiens* DNMT1 was the lowest (84%) (Table 6). The lowest degree of similarity was found in *D. rerio* (83%), however this comparison utilized 99% of the *L. ocellata* Dnmt1 amino acid sequence. The free on-line software MobyE@Pasteur v1.0 (<http://mobyE.pasteur.fr/cgi-bin/portal.py>) was used to perform a multiple protein alignment and unrooted tree analysis of the above mentioned vertebrate model organism Dnmt1s with the *L. ocellata* Dnmt1. This revealed that *L. ocellata* Dnmt1 protein was most similar to *D. rerio* dnmt1 when considering the entire amino acid sequence (Figure 24). Furthermore, there is a pattern indicating a higher conservation of amino acid types (hydrophilic, hydrophobic, large or small residues) over actual amino acid identities in all of the compared Dnmt1 proteins. This conservation of amino acid type would be more important in maintaining protein structure and function than specific identities as multiple amino acids with similar hydrophobic/hydrophilic nature may be of similar sizes allowing for proper protein folding.

The order and identities of conserved domains within the *L. ocellata* Dnmt1 protein, as reported by the Conserved Domain Database (CDD) [Marchler-Bauer *et al.* 2011], were consistent with other vertebrate maintenance Dnmts. No single species held a monopoly on domain similarities as top hits were found in all species except *H. sapiens*. The range of percentage similarity for each domain remained fairly narrow amongst the vertebrate species with the widest spread being only 15%. The *L. ocellata* Dnmt1 was predicted to have three nuclear localization signals (NLS) by WoLF PSORT [Horton *et al.* 2007], a number consistent with that found in mice [Cardoso and Leonhardt 1999] and zebrafish

[Mhanni *et al.* 2001]. There was little conservation over all of these signals between *L. ocellata* and the representative vertebrate models (data not shown). Though position and degree of similarity varies between the species, the NLS do seem to be heavily biased towards positively charged, hydrophilic lysine and arginine residues, a bias preserved in the *L. ocellata* Dnmt1.

Table 6: Comparison of *L. ocellata* Dnmt1 to representative vertebrate Dnmt1 proteins. *L. ocellata* translated open reading frame (as reported by GeneRunner software) was aligned to other vertebrate amino acid sequences using the National Center for Biotechnology Information BLASTp search tool. Vertebrate amino acids sequences used were as follows: *D. rerio* GenBank [NM_131189](#), *G. gallus* GenBank [D43920](#), *H. sapiens* GenBank [X63692](#), *M. musculus* GenBank [X14805](#) and *X. laevis* GenBank [D78638](#).

	% Identity Match	% Similarity	% <i>L. ocellata</i> protein covered in alignment
<i>D. rerio</i>	72	83	99
<i>G. gallus</i>	74	85	97
<i>H. sapiens</i>	78	88	84
<i>M. musculus</i>	73	84	85
<i>X. laevis</i>	74	85	99

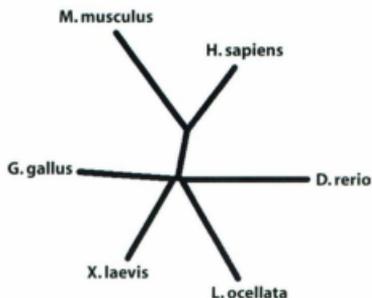


Figure 24: *L. ocellata* Dnmt1 amino acid sequence is more similar to that of *D. rerio* Dnmt1 than other representative vertebrate maintenance methyltransferases. Multiple protein alignment and unrooted tree analysis was performed using Mobyle@Pasteur v1.0 (<http://mobyle.pasteur.fr/cgi-bin/portal.py>) and the following protein sequences: *D. rerio* GenBank NM_131189, *G. gallus* GenBank D43920, *H. sapiens* GenBank X63692, *M. musculus* GenBank X14805, and *X. laevis* GenBank number D78638.

The least conserved *L. ocellata* Dnmt1 domain present in all the chosen representative vertebrate species was the DNA methyltransferase 1-associated protein (DMAP1) binding domain. DMAP1 co-localizes with recombinant chromatin following homologous recombination and appears to have a strong binding preference for hemimethylated DNA. DMAP1 stimulates DNA methylation mediated by Dnmt1 affecting epigenetic alterations associated with the repair of double stranded DNA breaks during homologous recombination [Lee *et al.* 2010]. Additionally, DMAP1's affinity for hemimethylated CpG dinucleotides promotes the recruitment of Dnmt1 to semi-conservatively synthesized DNA during S phase [Takebayashi *et al.* 2007].

Targeting Dnmt1 to newly synthesized DNA where it can fulfill its role as a maintenance DNA methyltransferase is the replication foci domain (RFD). The RFD binds the proliferating cell nuclear antigen (PCNA) [Chuang *et al.* 1997]. PCNA assists in DNA polymerase delta processivity [Langston and O'Donnell 2008]. While this association with the replication machinery is not strictly necessary for Dnmt1 maintenance activity it does improve its efficiency [Spada *et al.* 2007].

Dnmt1's targeting of CpG dinucleotides is aided by the presence of a zinc finger domain containing a cluster of eight cysteinyl residues in the form CX₂CX₂CX₄CX₂CX₂CX₁₅CX₄C [Bestor 1992]. This cluster forms two short helical segments embedded with Zn²⁺ cations that interact with both the major and minor grooves of DNA as nonmethylated CpG dinucleotides are bound [Song *et al.* 2011]. Interestingly the study by Song *et al.* [2011] suggests that methylation of either cytosine of a CpG dinucleotide would create steric clashes with peptide atoms preventing Dnmt1 from binding.

Adjacent to the C-terminal catalytic domain are two Bromo Adjacent Homology (BAH) domains. The core of a BAH domain is an open, distorted β -barrel comprised of six strands interrupted by a helix positioned between the fourth and fifth strands [Oliver *et al.* 2005]. BAH domains mediate protein-protein interactions and target origin recognition complexes (ORC) to chromatin in humans [Noguchi *et al.* 2006] potentially through an

interaction with heterochromatin-associated protein 1 (HP1) as seen in *X. laevis* [Pak *et al.* 1997].

Of the analyzed representative vertebrate models the C-terminal catalytic domain is by far the most highly conserved domain. *L. ocellata* Dnmt1 showed a minimum of 94% similarity with the *M. musculus* and a high of 97% identity matches with the *G. gallus* dnmt1. Both CDD features one and two show an excellent correlation with the Pósfai *et al.* [1989] predictions (see Figure 10). Feature one describes the co-factor binding region of the domain. The CDD source for feature one, O'Gara *et al.* [1999], analyzed the crystal structure of the bacterial *Haemophilus haemolyticus* methyltransferase focusing on the region involved in binding the methyl donor AdoMet. It is not surprising that this feature would align well with the Pósfai *et al.* [1989] predicted motifs as the *HhaI* methyltransferase was included in the thirteen bacterial enzymes used in making their predictions. The CDD's second feature describes the substrate interaction site which, to a slightly lesser degree than feature one, also aligns very well with the Pósfai *et al.* [1989] predictions. The third CDD feature is involved in DNA binding as well as target recognition. The target recognition domain (TRD) lies between predicted motifs VIII and IX [as reviewed by Cheng 1995]. Amino acid residues determined to be involved by the CDD do not align perfectly with the predicted motifs. There is overlap with features two and three corresponding to the binding of DNA (see Figure 10) but the means of recognizing an everted cytosine nucleotide may be less conserved between species due to

issues involved with approaching a distorted DNA complex and avoiding steric interactions.

L. ocellata Dnmt3: structure

This study reports the isolation and identification of a 3301 nt *L. ocellata DNA-methyltransferase 3* cDNA sequence (see Figure 11) building upon the previous 274 nt acquired during my Honours research [Lake 2008]. Using the full length *L. ocellata* cDNA sequence in a National Center for Biotechnology Information BLASTn search, as well as a DeCypher Tera-BLASTN search (http://decypher.mdibl.org/decypher/algo-terablast/tera-blastn_nn.shtml), reported no significant similarity found. The *L. ocellata* Dnmt3 ORF, as reported by Gene Runner software, was 779 aa at its longest (see Figure12). There was evidence of 5' and 3' transcript variants. The 3' variants will be discussed further below while discussion of the 5' variants is relevant when analyzing the translation start sites. The *L. ocellata Dnmt3* cDNA presented here shows two potential translation start codons in the same reading frame separated by 27 aa. The sequence surrounding the upstream start codon was compared to the Kozak consensus sequence revealing a weak similarity of only the +4G nucleotide (see Table 3). Comparison of the downstream start codon to the Kozak consensus sequence revealed a strong similarity with conservation of both the -3R and +4G nucleotides. Based on these data it appears that the downstream start codon would be preferred for translation initiation. However, when the sequences surrounding the Dnmt3 start codons in other vertebrate models were

assessed, the majority showed only a weak similarity to the Kozak consensus sequence and there was no preference for conservation of the -3R or +4G nucleotides. Having a weak translation start site adjacent to a stronger translation start site in *L. ocellata* may not necessarily mean it would be passed over by scanning pre-initiation complexes as other vertebrate *Dnmt3* transcripts possess similar weak, and even poor, similarity to the Kozak consensus sequence.

The 779 aa Dnmt3 of *L. ocellata* falls within the lower size limits of the *de novo* methyltransferases of the representative vertebrate models (*D. rerio* dnmt6: 731 aa – *D. rerio* dnmt3: 1448 aa). Comparisons to protein databases by a National Center for Biotechnology Information BLASTp search reported a top hit from *D. rerio* dnmt4 registering 80% similarity. Mobylye@Pateur v1.0 (<http://mobylye.pasteur.fr/cgi-bin/portal.py>) protein alignment and unrooted tree analysis of the *L. ocellata* Dnmt3 with the representative vertebrate models also reported a higher degree of similarity with *D. rerio* dnmt4 than with all others aligned (Figure 25). The next BLASTp result which was not a predicted protein product was *H. sapiens* DNMT3b isoform 2 with 75% similarity. Considering the unrooted tree analysis, the top two BLAST hits and the fact that *D. rerio* dnmt4 is very similar to the mammalian *DNMT3b* in both sequence and expression patterns [Shimoda *et al.* 2005, Smith *et al.* 2011], it seems that the transcript isolated in this study codes for an Elasmobranchii *DNMT3b* equivalent.

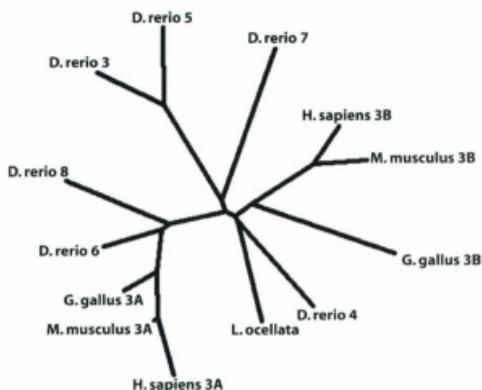


Figure 25: *L. ocellata* Dnmt3 amino acid sequence is more similar to that of *D. rerio* dnmt4 than other representative vertebrate *de novo* methyltransferases. Protein alignment and unrooted tree analysis was carried out using the full length *L. ocellata* Dnmt3 amino acid sequence and the primary isoform of each of the representative vertebrate *de novo* methyltransferases. Multiple protein alignment and unrooted tree analysis was performed using Mobyle@Pasteur v1.0 (<http://mobyle.pasteur.fr/cgi-bin/portal.py>) and the following protein sequences: *D. rerio* dnmt3 through 8 GenBank [AB196914](#), [AB196915](#), [AB196916](#), [AB196917](#), [AB196918](#) and [AB196919](#) respectively, *G. gallus* dnmt3a and 3b GenBank [NP_001020003.1](#) and [NP_001019999.1](#), *H. sapiens* DNMT3a and 3b GenBank [AB208833](#) and [NP_008823.1](#) as well as *M. musculus* Dnmt3a and 3b GenBank [AF068625](#) and [AF068626](#).

The 5' variant shortens the potential ORF by 28 aa which does not affect the NLS predicted by WoLF PSORT [Horton *et al.* 2007] located at amino acid 61 (see Figure 13). Regulatory roles have been suggested for 5' splice variants in zebrafish *de novo* methyltransferases, however those splice junctions involve untranslated regions, not the coding sequence [Smith *et al.* 2005]. Even with the slightly truncated N-terminus the order and identities of conserved domains within the *L. ocellata* Dnmt3, as reported by the CDD [Marchler-Bauer *et al.* 2011], were consistent with other vertebrate *de novo* Dnmts.

Similar to *L. ocellata* Dnmt1, there was no trend in domain similarities across the representative species although *D. rerio* dnmt4 did hold the highest degree of similarity in both splice variants of the C-terminus catalytic domain (see Figure 13). The most distal N-terminus domain is the PWWP characterized by a moderately conserved 100-150 aa region with the highly conserved Proline-Tryptophan-Tryptophan-Proline motif. The N-terminus half forms a five stranded β -barrel while the C-terminus half forms a helical bundle [Qiu *et al.* 2002]. PWWP domains in proteins other than methyltransferases have been implicated in binding methylated histones [Wang *et al.* 2010]; however the same domains of mammalian Dnmt3a and Dnmt3b have been shown to be involved in their functional specialization. Mammalian Dnmt3a and Dnmt3b concentrate in the major satellite repeats at pericentric heterochromatin. Disrupting the PWWP domain of each enzyme prevents this association and abolishes their ability to methylate major satellite repeats. It was found that the PWWP domain of Dnmt3a had little DNA binding ability

while the Dnmt3b equivalent had a nonspecific DNA binding capacity suggesting the targeting of Dnmt3 enzymes to their specific DNA regions is mediated by a mechanism other than direct protein-DNA contact [Chen *et al.* 2004].

Between the PWWP domain and the C-terminus catalytic region lies the Plant Homeo Domain (PHD). The PHD is a metal-dependent folding motif comprising ~50 amino acid residues featuring a conserved Cysteine₄-Histidine-Cysteine₃ zinc binding motif [Pascual *et al.* 2000]. The PHD has been shown to mediate interactions between mammalian Dnmt3a and histone deacetylase 1 (HDAC1) contributing to the active repression of transcription [Fuks *et al.* 2003].

The most highly conserved region between the *L. ocellata* Dnmt3 and the representative vertebrate *de novo* methyltransferases was the C-terminal catalytic region. The Dnmt3 presented here shows evidence of at least one transcript variant involving the 3' end. This transcript splice results in the removal of 64 aa corresponding to the target recognition domain situated between motifs VIII and IX (see Figure 14). As was the case for *L. ocellata* Dnmt1 there was a high degree of overlap between CDD features one and two (co-factor binding and substrate interaction respectively) and the Pósfai *et al.* [1989] predicted motifs. Feature three (DNA binding) was independent of the other features with the exception of a single amino acid. A large component of the DNA binding feature was removed from the 3' splice of the *L. ocellata* Dnmt3 transcript. A National Center for Biotechnology Information BLASTp search using the *L. ocellata* 3' spliced protein as the

query returned a top hit indicating 73% similarity to *H. sapiens* DNMT3b isoform 3. Mammalian *Dnmt3b3* is a ubiquitously expressed catalytically inactive isoform that is co-expressed in cells with other active Dnmt3 isoforms. Dnmt3b3 is capable of interacting with Dnmt3b1 and Dnmt3a as well as modulating the activity of Dnmt3a-Dnmt3L complexes *in vitro* by an as yet unknown mechanism [Van Emburgh and Robertson 2011]. The presence of a Dnmt3 transcript lacking its target recognition domain and having a high degree of similarity to the mammalian Dnmt3b3 isoform indicates that *L. ocellata* could potentially possess a similar mechanism as that seen in mammals for modulating Dnmt3 complex activities.

Tissue distribution of Dnmt3 3' splice variants

With *L. ocellata* showing evidence of a potentially catalytically inactive Dnmt3 isoform I wanted to investigate tissue distribution of the 3' splice variants as well as in-tissue relative abundances. Both 3' transcript variants cloned in this investigation were present in each of the four tissue types tested. If the shorter transcript lacking the TRD does indeed represent an equivalent of the mammalian *Dnmt3b3* isoform it would appear that the mammalian trend of *Dnmt3b3* being co-expressed with other isoforms is present in *L. ocellata* as well. All RT-PCR reactants/template concentrations and conditions were identical in all four scenarios; however no internal standard to control for RNA quantity is available for the skate so quantitative comparisons between lanes are not reliable. Nevertheless, the general band intensity differences seen between ovary and testis tissue

(lanes 1 and 3 respectively; see Figure 15) correlate well with the comparative methylation level data.

A higher level of DNA methylation in testis over ovary tissue (see Figure 5) implies a higher concentration of *de novo* methyltransferases. Mammalian *de novo* methyltransferases re-program methylation patterns in pre-implantation embryos and remain active in germ cell lines while becoming down-regulated in adult differentiated somatic tissues [as reviewed by Hermann *et al.* 2004]. Similar dynamic *de novo* methyltransferase expression is seen in *D. rerio* [Smith *et al.* 2011]. The relatively low levels of *Dnmt3* 3' isoforms seen in *L. ocellata* gut tissue seem to mirror this reduction in somatic *Dnmt3* levels. *L. ocellata* brain tissue, however, does not appear to down-regulate *Dnmt3* transcript expression as Figure 15A indicates levels of the 3' isoforms comparable with those seen in testis. Studies in mice have shown that *Dnmt3b* expression is detected in early developmental stages of the central nervous system (CNS) within a narrow window corresponding to embryonic day 11-15. *Dnmt3a* expression is high in neural precursor cells but is subsequently maintained at only slightly lower levels in postmitotic CNS neurons [Feng *et al.* 2005]. *M. musculus* *Dnmt3a* shows evidence of playing a complementary role with *Dnmt1* in altering methylation patterns in response to behavioural changes or external signals involved in synaptic plasticity, learning and associative memory formation [Yu *et al.* 2011]. The persistence of *Dnmt3* transcript expression in *L. ocellata* brain tissue may reflect a similar means of brain function regarding dynamic DNA methylation patterns.

BLASTp results indicated a higher degree of similarity of the *L. ocellata* Dnmt3 to other vertebrate Dnmt3b equivalents. It may be that Elasmobranchii possess fewer *de novo* methyltransferases than either teleosts or mammals and therefore rely on variable transcript splicing to produce specific *Dnmt3* forms with diversified functions, as opposed to having separate *Dnmt3a* and *Dnmt3b* genes. This would account for an apparent *L. ocellata* *Dnmt3b* equivalent reflecting tissue specific expression patterns characteristic of mammalian *Dnmt3a*. Of course an exhaustive characterization of the full *L. ocellata* Dnmt complement is required before such a claim can be made.

The relative abundance of each 3' transcript isoform within a single tissue sample was investigated using square pixel counts of areas beneath band intensity peaks determined by ScnImage software (see Figure 15B-E). DNA band intensity is a result of ethidium bromide (EtBr) binding the DNA and subsequent ultraviolet (UV) light illumination. On average one EtBr molecule intercalates into native-DNA every four nucleotides at physiological pH [Chitre and Korgaonkar 1979]. Therefore the longer a DNA fragment is the more EtBr it will bind and the more intense it will appear under UV light. A 703 bp amplicon would be expected to bind ~1.4 times as many EtBr molecules as a 514 bp amplicon resulting in a greater band intensity based on EtBr binding capacity and not abundance of amplicons.

The ratios of 3' transcript isoforms from ovary, testis and brain tissue all indicate majorities of 703 bp amplicons (see Table 4). Taking into account the baseline intensity

determined by size discrepancy alone (1.4) it seemed that ovary tissue had the 514 bp amplicon present slightly more than the 703 bp amplicon. This bias of a potentially catalytically inactive form over a potentially active form may reflect the down regulation of *de novo Dnmts* seen in the ovary tissue of other organisms. Testis and brain tissue maintained their 703 bp bias after accounting for their baseline intensities however the ratios were slightly diminished. The higher expression level of a potentially active isoform relative to a potentially inactive isoform is consistent with the higher *de novo* expression levels seen in these tissue types in mice and zebrafish.

Gut tissue appeared to be the outlier of the four tissue types as the 514 bp amplicon was the dominant *Dnmt3* 3' isoform present. The apparent low levels of the potentially catalytically active isoform relative to the shorter, potentially inactive isoform are again consistent with trends seen in other animals where *de novo* expression is down regulated in somatic tissues. It should be noted that, of the primers used to amplify the 3' variable region, the most 3' primer (#4) oriented in the 5' direction (see Figure 3) was designed to anneal in the 3' untranslated region (UTR) of the sequenced cDNA. Primer design was carried out previous to sequence translation so it was not known until after that this was the case. Having a primer anneal to the 3' UTR may have biased the observed isoform types as the PCR amplification would have only occurred from mRNA templates having that same 3'UTR. Variable UTRs are often related to transcript stability and translation rates [Moucadel *et al.* 2007]. It is possible these *L. ocellata* tissue types had additional

Dnmt3 mRNA with different 3' UTRs which would not have been detected in this PCR experiment using primer #4.

Which end goes where?

New primers were designed capable of discriminating against each of the 5' and 3' splice variants previously described. Their placement can be seen in Figure 3. Primer #6 annealed to a sequence present only in the TRD of the 3' isoform corresponding to mammalian *Dnmt3b2*. In the case of the *Dnmt3b3* equivalent with the TRD removed primer #6 would not anneal. This removal of sequence juxtaposed two sections of sequence previously separated by 189 nt creating the target sequence to which primer #7 annealed. The same strategy was employed on the 5' end. Primer #8 annealed to sequence within the spliced region whereas primer #9 could not recognize its target sequence unless the internal 84 nt were absent. RT-PCR was then carried out pairing each 5' primer with each of the 3' primers. Primer #2 was paired with both #8 and #9 for parallel RT-PCR reactions to ensure that the 5' discriminators were annealing where they were designed to. Similarly, primer #1 was paired with each of the 3' discriminating primers. Both primer #1 and #2 had been shown in previous experiments to reliably anneal where they were designed to (data not shown).

Whereas PCRs pairing primer #1 with primers #6 and #7 produced single DNA fragments of expected sizes, the pairing of primer #2 with #8 did not. In addition to the

expected amplicon of ~1900 bp, a second, far more prominent amplicon of ~1700 bp was generated (see Figure 16). This indicated that there were at least two different *Dnmt3* transcript isoforms present having the 5' insert but differing by ~200 bp at some location between primers #2 and #8. The PCR reaction matching primer #8 to #6 returned a similar result. The expected amplicon of 2159 bp was amplified but to a much lesser degree than a second amplicon approximately 200 bp smaller. The same result was seen from PCR using primers #8 and #7. The expected 2028 bp amplified but to a much lesser degree than a second amplicon ~200 bp smaller. Taken together these results suggest that four different *L. ocellata Dnmt3* transcript isoforms exist with the extended 5' sequence: two isoforms with and without the 3' TRD as well as two isoforms, with and without the TRD and an additional 200 bp located between nucleotides 207 (primer #8) and 1949 (primer #1) (Figure 26). Since the PCR amplification between primers #1 and #6 or #7 did not generate multiple bands, the additional ~200 bp missing from the two additional isoforms present in much higher concentrations must lie upstream of primer #1's location. A ClustalW [Larkin *et al.* 2007] alignment of *H. sapiens* DNMT3b1 (GenBank [NP_008823.1](#)), DNMT3b2 (GenBank [NP_787044.1](#)) and DNMT3b3 (GenBank [NP_787045.1](#)) proteins showed a 20 aa region present between the PWWP and PHD domains of Dnmt3b1 missing from both DNMT3b2 and DNMT3b3. The addition of 198 nt (~200 nt) would code for 66 aa which is approximately three times the insert seen in *H. sapiens* DNMT3b1. Adding *H. sapiens* DNMT3a (GenBank [AB208833](#)) to the alignment shows a number of small protein regions upstream from the 20 aa insert present in *H. sapiens* DNMT3b1 unique to the DNMT3a. All of these regions together account for 68



Figure 26: Schematic of *L. ocellata Dnmt3* transcripts illustrating the combinations of 5' and 3' splice variants. Regions of transcripts determined experimentally to be subjected to splicing are indicated in orange. Dashed lines indicate regions of transcripts suspected of harbouring additional splice sites. Asterisks indicate transcripts corresponding to PCR products that were of expected sizes (see Figures 16 and 17) but were present in much lower concentrations than PCR products ~200 bp smaller having the same 5' and 3' splice variant combinations.

additional amino acids present upstream in *H. sapiens* DNMT3a from the 20 aa distinguishing DNMT3b1 from DNMT3b2/3. Although *H. sapiens* DNMT3a lacks 16 of the 20 aa between the PWWP and PHD domain corresponding to the DNMT3b2/3 region mentioned above there is still a 68 aa (204 nt) difference in N-terminus regions between the *H. sapiens* DNMT3a and DNMT3b species. Further study is required to determine the exact nature of the full complement of *L. ocellata* Dnmts. It may be that the clone obtained in this investigation was in fact an isoform transcribed at lower levels than the primary *L. ocellata* *Dnmt3* gene product and may be more similar to *H. sapiens* DNMT3a in sequence than originally thought, even though it appears to produce an equally abundant transcript with its TRD spliced out. This would partially account for the DNMT3a-like higher levels of transcription observed in brain tissue over other adult somatic tissue.

Similar unexpected results were obtained from PCR using the above strategy substituting primer #9 for #8. Both primers #6 and #7 were shown to anneal to their expected target sequences. The PCR reaction using primers #2 and #9 was expected to yield a 1836 bp amplicon but instead amplified a DNA fragment ~1650 bp in size. There did not appear to be a second, less prominent band present in this reaction as there had been using primer #8 (Figure 17). Judging from this reaction alone it would seem that the *L. ocellata* *Dnmt3* isoform utilizing the shorter 5' sequence is lacking the ~200 bp seen to separate *H. sapiens* DNMT3a from DNMT3b. This absence of ~200 bp is consistent in the matching of primer #9 with #6 and #7. PCR was expected to return amplicons of 2085 bp

and 1955 bp using #6 and #7 respectively. The DNA fragments that amplified, however, were ~1900 bp and ~1800 bp. These sizes are consistent with the results from the primer #9 landmarking PCR amplification. It would appear that at least two *L. ocellata Dnmt3* transcript isoforms exist where the shorter 5' sequence is utilized with both the long 3' sequence and the short, potentially catalytically inactive, 3' sequence (see Figure 26).

Evidence of L. ocellata Dnmt3 pseudogenes

Pseudogenes are sequences present in a population's genome that are characterized by close similarities to one or more paralogous genes and are typically non-functioning. Pseudogenes can be found in bacteria, plants, insects and vertebrates. They can arise due to the duplication of genes occurring through unequal crossing over or through retrotransposition where a single stranded RNA molecule is reverse transcribed and the resulting double stranded molecule is inserted back into the genome. Processed pseudogenes resulting from retrotransposition typically lack 5' promoter sequences and introns, flanking repeats and 3' polyadenylation tracts (as reviewed by Mighell *et al.* 2000). Though they are present in the genome with sequences similar to those of actively transcribed genes, retrotransposed pseudogenes are not likely to be co-expressed with their gene of origin as they more often than not are removed from the promoters and enhancers governing expression of said gene. In the process of investigating the *L. ocellata Dnmt3* 3' transcript isoforms several "no-RT" PCR controls yielded the same results as the experimental reactions. The 703 bp and 514 bp amplicons were amplified

under reaction conditions lacking a cDNA template. This suggested the RNA sample was contaminated with genomic DNA capable of acting as template using primers #3 and #4. Figure 18 shows the initial integrity of the *L. ocellata* testis genomic DNA sample. The sample's high molecular weight and minimal smearing following separation by electrophoresis indicates no evidence of degradation. Both the 703 bp and 514 bp amplicons were generated from the genomic DNA following RNase A treatment indicating RNA was not acting as template. Treatment with DNA wipeout subsequently degraded the genomic DNA template. Removal of template in this manner resulted in no amplification of the 703 bp and 514 bp fragments. PCR products from the first reaction were cloned and sequenced to confirm that *L. ocellata* genomic DNA contained the same sequences as the *Dnmt3* 3' transcript isoforms (see Figures 19 and 20). Alignments of the 703 bp and 514 bp fragments amplified from mRNA templates and genomic DNA templates show 99% similarity in both cases (see Figure 21 and Figure 22). The 1% discrepancy may be accounted for by sequencing errors.

Figure 23 shows a ClustalW [Larkin *et al.* 2007] alignment of the *D. rerio*, *M. musculus* and *H. sapiens de novo* methyltransferases along side the *L. ocellata* 703 bp amplicon. The region corresponding to *L. ocellata* clone spanned multiple intron/exon splice sites in all the aligned teleost and mammalian sequences, many of which were highly conserved in their position across species. The parsimonious explanation is that these splice sites were inherited from an evolutionary ancestor and not spontaneously incorporated into multiple species' genomes at homologous locations. Therefore it appears that *L. ocellata*

has at least two retrotransposed pseudogenes, one with the TRD intact and one with the TRD removed, each consisting only of exonic sequence. The conventional *L. ocellata* *Dnmt3* gene consisting of both introns and exons would not have yielded a PCR product under the experimental conditions used as the extension time was sixty seconds, time enough to amplify only a few thousand base pairs. If primers #3 and #4 annealed to exonic sequences flanking one or more introns, full extension would likely not occur in such a short period of time. *Dnmt3* pseudogenes are not unique to *L. ocellata* and have been documented in both *M. musculus* and *H. sapiens* [Lees-Murdock *et al.* 2004]. *H. sapiens* carry a *DNMT3a* pseudogene that has remained functional while *M. musculus* carry at least one for each member of the *Dnmt3* family, all of which have resulted from retrotransposition and not genome duplication.

Summary and Conclusion

Development of multicellular organisms requires an elegant and sophisticated means of dynamically controlling chromatin structure. DNA methylation has been shown to play an integral role in vertebrate genomic management as aberrations in methylation patterns have been linked to phenotypic abnormalities, cancer and even embryonic death. Much of our understanding of the enzymes responsible for establishing and maintaining these methylation patterns has originated from studies using teleosts and mammals. There are similarities in enzyme structure and basic developmental expression patterns between these two evolutionary groups, but the significance of functional roles related to different gene numbers and reproductive strategies is still not known. Investigating DNA methylation in subclass Elasmobranchii, the evolutionary outgroup of teleosts and mammals, provides an opportunity to explore the core roles of vertebrate DNA methylation and how it is established and maintained. Here I have presented the first full length cDNA sequences of the maintenance methyltransferase *Dnmt1*, and *de novo* methyltransferase *Dnmt3* of *Leucoraja ocellata*. At the nucleotide level they show no significant similarities to known *Dnmts*. At the protein level they are very similar possessing all the domains we recognize to be essential for their function assembled in the order consistent with all other documented maintenance and *de novo* methyltransferases. I have provided evidence of multiple *Dnmt3* splice variants involving alternate translation start codons as well as a 3' splice removing a catalytically important region of the methyltransferase domain. Although the *L. ocellata* amino acid sequence

shows a high degree of similarity to *H. sapiens DNMT3b* isoform2 there appears to be a trend in tissue distribution similar to that of *H. sapiens DNMT3a* in the central nervous system. It may be that *L. ocellata* possesses a single *Dnmt3*, or a reduced *Dnmt3* family, relying on multiple transcript splices to perform the functions carried out by various *Dnmt3* genes in the more derived mammalian and teleost lineages. There are also indications of at least two retrotransposed *Dnmt3* pseudogenes with their origins linked to the potentially catalytically active and inactive methyltransferase splice variants. A more rigorous investigation of the total number of *de novo* methyltransferases of the *L. ocellata* genome is currently being conducted. New degenerate primers flanking the most common intron/exon splice sites shared by all zebrafish and mammalian *Dnmt3* genomic sequences have been designed. These primers are being used in PCR amplification with gDNA as template in an effort to detect *L. ocellata Dnmt3* intronic sequences. Multiple intronic sequences confirmed within the *L. ocellata* genome would imply the existence of multiple *Dnmt3* genes. Cloning of the exonic sequences flanking these multiple intronic sequences would provide a means of creating primers specific to each of the individual *L. ocellata Dnmt3s*. Further probing of the transcriptome with these new primers would yield full clones of all the potential *Dnmt3* family members. Developmental expression analysis by way of immunoprecipitation could then be conducted to elucidate the pattern of *de novo* methyltransferase expression in the developing *L. ocellata*.

Preliminary work has shown the trend of hypermethylated paternal germ cells relative to maternal germ cells seen in both teleosts and mammals is present in *L. ocellata*. While

the exact number of methyltransferases is still being investigated it appears that early developmental methylation dynamics can be traced back to *L. ocellata*. This preservation of unbalanced paternal and maternal methylation contributions to the Elasmobranchii embryo indicates a dynamic epigenetic reprogramming event is essential, not only in Osteichthyes, but also in the sister group Chondrichthyes.

This investigation has been undertaken in the hopes of assisting the exploration of DNA methylation in subclass Elasmobranchii, a group of organisms employing a wide range of reproductive strategies and developmental programs capable of providing insight into intrinsic vertebrate chromatin management as well as the derived forms studied in both teleosts and tetrapods.

References

- Aapola, U., Shibuya, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Shinsei, M., Krohn, K., Antonarakis, S., Shimizu, N., Kudoh, J. and 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.
- Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoghbi, H.Y. (1999) Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2.
- Anderson, R.M., Bosch, J.A., Goll, M.G., Hesselson, D., Dong, P.D., Shin, D., Chi, N.C., Schlegel, A., Halpern, M. and Stainier, D.Y. (2009) Loss of Dnmt1 catalytic activity reveals multiple roles for DNA methylation during pancreas development and regeneration. *Developmental Biology* 334: 213-223.
- Aoki, A., Suetake, I., Miyagawa, J., Fujio, T., Chijiwa, T., Sasaki, H. and Tajima, S. (2001) Enzymatic properties of *de novo*-type mouse DNA (cytosine-5) methyltransferase. *Nucleic Acids Research* 29: 3506-3512.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics* 22: 195-201.

Bestor, C. (2011) Winter Skate Biological Profile. Florida Museum of Natural History. Retrieved from <http://www.flmnh.ufl.edu/fish/Gallery/Descript/WinterSkate/WinterSkate.html>.

Bestor, T.H. (2000) The DNA methyltransferases of mammals. *Human Molecular Genetics* 9: 2395-2402.

Bestor, T.H. (1992) Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *European Molecular Biology Organization Journal* 11: 2611-2617.

Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. *Journal of Molecular Biology* 203: 971-983.

Bewley, M.S., Pena, J.T.G., Plesch, F.N., Decker, S.E., Weber, G.J. and Forrest, J.N.Jr. (2006) *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 219: R1157-R1164.

Bourc'his, D. and Bestor, T.H. (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431: 96-99.

Bourc'his, D., Miniou, P., Jeanpierre, M., Gomes, D.M., Dupont, J.M., De Saint-Basile, G., Maraschio, P., Tiepolo, L. and Viegas-Péquignot, E. (1999) Abnormal methylation does not prevent X inactivation in ICF patients. *Cytogenetics and Cell Genetics* 84: 245-252.

Cai, S., Wang, L., Ballatori, N. and Boyer, J.L. (2001) Bile salt export pump is highly conserved during vertebrate evolution and its expression is inhibited by PFIC type II mutations. *American journal of Physiology. Gastrointestinal and Liver Physiology* 281: G316-G322.

Cardoso, M.C. and Leonhardt, H. (1999) DNA methyltransferase is actively retained in the cytoplasm during early development. *The Journal of Cell Biology* 147: 25-32.

Chen, H., Chen, H., Khan, M.A., Rao, Z., Wan, X., Tan, B. and Zhang, D. (2011) Genetic mutations of p53 and k-ras in gastric carcinoma patients from Hunan, China. *Tumor Biology* 32: 367-373.

Chen, T. and Li, E. (2004) Structure and function of eukaryotic DNA methyltransferases. *Current Topics in Developmental Biology* 60: 55-89.

Chen, T., Tsujimoto, N. and Li, E. (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Molecular and Cellular Biology* 24: 9048-9058.

Chen, T., Ueda, Y., Xie, S. and Li, E. (2002) A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *The Journal of Biological Chemistry* 277: 38746-38754.

Cheng, X. (1995) Structure and function of DNA methyltransferases. *Annual Review of Biophysics and Biomolecular Structure* 24: 293-318.

Chitre, A.V. and Korgaonkar, K.S. (1979) Binding of ethidium bromide and quinacrine hydrochloride to nucleic acids and reconstituted nucleohistones. *Biochemical Journal* 179: 213-219.

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* 162: 156-159.

Chuang, L.S.H., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G. and Li, B.F.L. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21^{WAF1}. *Science* 277: 1996-2000.

Corley-Smith, G.E., Lim, C.J. and Brandhorst, B.P. (1996) Production of androgenetic zebrafish (*Danio rerio*). *Genetics* 142: 1265-1276.

Delcuve, G.P., Rastegar, M. and Davie, J.R. (2009) Epigenetic Control. *Journal of Cellular Physiology* 219: 243-250.

Deplus, R., Brenner, C., Burgers, W.A., Putmans, P., Kouzarides, T., de Launoit, Y. and Fuks, F. (2002) Dnmt3L is a transcriptional repressor that recruits histone deacetylase. *Nucleic Acids Research* 30: 3831-3838.

Feinberg, A.P. and Vogelstein, B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301: 89-92.

Feng, J., Chang, H., Li, E. and Fan, G. (2005) Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. *Journal of Neuroscience Research* 79: 734-746.

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenöder, S., Wang, M., Qin, W., Söding, J., Spada, F. and Leonhardt, H. (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *Public Library of Science One* 6: e16627.

Fuks, F., Hurd, P.J., Deplus, R. and Kouzarides, T. (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferases. *Nucleic Acids Research* 31: 2305-2312.

Geiman, T.M. and Meugge, K. (2010) DNA methylation in early development. *Molecular Reproduction and Development* 77: 105-113.

Ghosh, R.P., Nikiitina, T., Horowitz-Scherer, R. A., Gierasch, L.M., Uversky, V.N., Hite, K., Hansen, J.C. and Woodcock, C.L. (2010) Unique physical properties and interactions of the domains of methylated DNA binding protein 2. *Biochemistry* 49: 4395-4410.

Goll, M.G. and Bestor, T.H. (2005) Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry* 74: 481-514.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C., Zhang, X., Golic, K.G., Jacobsen, S.E. and Bestor, T.H. (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase Homolog Dnmt2. *Science* 311: 395-398.

Gregory, T.R. (2011). Animal Genome Size Database. <http://www.genomesize.com>

Gronbaek, K., Hother, C. and Jones, P.A. (2007) Epigenetic changes in cancer. *ACTA Pathologica, Microbiologica et immunologica Scandinavica* 115: 1039-1059.

Hata, K., Okano, M., Lei, H. and Li, E. (2002) Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice. *Development* 129: 1983-1993.

Hermann, A., Gowher, H. and Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cellular and Molecular Life Science* 61: 2571-2587.

Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C.C. (1991) p53 mutations in human cancers. *Science* 253: 49-53.

Horton, P., Park, K., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007) WoLF PSORT: Protein localization predictor. *Nucleic Acids Research* 35: W585-W587.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M. and Chaillet, J.R. (2001) Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene. *Cell* 104: 829-838.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Dass, S.U., Landsberger, N., Strouboulis, J. and Wolfe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics* 19: 187-191.

Kanai, Y. (2008) Alterations of DNA methylation and clinicopathological diversity of human cancers. *Pathology International* 58: 544-558.

Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate RNAs. *Nucleic Acids Research* 15: 8125-8148.

Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361: 13-37.

Kumar, S. and Hedges, S.B. (1998) A molecular timescale for vertebrate evolution. *Nature* 392: 917-920.

Lake, G.D. (2008) Isolation and characterization of the *Leucoraja ocellata* DNA-methyltransferase 3. Memorial University of Newfoundland, NL, Canada.

Langston, L.D. and O'Donnell, M. (2008) DNA polymerase delta is highly processive with proliferating cell nuclear antigen and undergoes collision release upon completing DNA. *Journal of Biological Chemistry* 283: 29522-29531.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.

Lee, G.E., Kim, J.H., Taylor, M. and Muller, M.T. (2010) DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair. *Journal of Biological Chemistry* 285: 37630-37640.

Lees-Murdock, D.J., McLoughlin, G.A., McDaid, J.R., Quinn, L.M., O'Doherty, A., Hiripi, L., Hack, C.J. and Walsh, C.P. (2004) Identification of 11 pseudogenes in the DNA methyltransferase gene family in rodents and humans and implications for the functional loci. *Genomics* 84: 193-204.

Lci, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K.A., Jaenisch, R. and Li, E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122: 3195-3205.

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

Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics* 3: 662-673.

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

Ma, Y., Jacobs, S.B., Jackson-Grusby, L., Mastrangelo, M., Torres-Betancourt, J.A., Jaenisch, R. and Rasmussen, T.P. (2005) DNA CpG hypomethylation induces heterochromatin reorganization involving the histone variant macroH2A. *Journal of Cell Science* 118: 1607-1616.

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.

MacKay, A.B., Mhanni, A.A., McGowan, R.A. and Krone, P.H. (2007) Immunological detection of changes in genomic DNA methylation during early zebrafish development. *Genome* 50: 778-785.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., Deweese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.L., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C. and Bryant, S.H. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* 39(Database issue):D225-9. Epub 2010 Nov 24.

Matarazzo, M.R., De Bonis, M.L., Vacca, M., Della Ragione, F. and D'Esposito, M. (2009) Lessons from two human chromatin diseases, ICF syndrome and Rett syndrome. *The International Journal of Biochemistry and Cell Biology* 41: 117-126.

Mattingly, C., Parton, A., Dowell, L., Rafferty, J. and Barnes, D. (2004) Cell and Molecular Biology of Marine Elasmobranchs: *Squalus acanthias* and *Raja erinacea*. *zebrafish* 1: 111-120.

Mertineit, C., Yoder, J.A., Taketo, T., Laird, D.W., Trasler, J.M. and Bestor, T.H. (1998) Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Development* 125: 889-897.

Mighell, A.J., Smith, N.R., Robinson, P.A. and Markham, A.F. (2000) Vertebrate Pseudogenes. *Federation of European Biochemical Societies Letters* 468: 109-114.

Mhanni, A.A. and McGowan, R.A. (2004) Global changes in genomic methylation levels during early development of the zebrafish embryo. *Development Genes and Evolution* 214: 412-417.

Mhanni, A.A. and McGowan, R.A. (2002) Variations in DNA (cytosine-5)-methyltransferase-1 expression during oogenesis and early development of the zebrafish. *Development Genes and Evolution* 212: 530-533.

Mhanni, A.A., Yoder, J.A., Dubesky, C. and McGowan, R.A. (2001) Cloning and sequence analysis of a zebrafish cDNA encoding DNA (cytosine-5)-methyltransferase-1. *Genesis* 30: 213-219.

Moucadel, V., Lopez, F., Ara, T., Benech, P. and Gautheret, D. (2007) Beyond the 3' end: experimental validation of extended transcript isoforms. *Nucleic Acids Research* 35: 1947-1957.

Musick, J.A. and Ellis, J.K. (2005) Reproductive Evolution of Chondrichthyans. Reproductive biology and phylogeny of Chondrichthyes: Sharks, Batoids and Chimaeras in Hamlett, W.C. (Ed) *Reproductive biology and phylogeny Vol.3*: 45-71. Enfield, New Hampshire, USA.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H. and Miura, K. (2007) Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Research* 36: 861-871.

Nelson, J.S. (2006) *Fishes of the world*. New York. Wiley: 624

Noguchi, K., Vassilev, A., Ghosh, S., Yates, J.L. and DePamphilis, M.L. (2006) The BAH domain facilitates the ability of human Orc1 protein to activate replication origins *in vivo*. *European Molecular Biology Organization Journal* 25: 5372-82.

O'Gara, M., Zhang, X., Roberts, R.J. and Cheng, X. (1999) Structure of a binary complex of *HhaI* methyltransferase with S-adenosyl-L-methionine formed in the presence of a short non-specific DNA oligonucleotide. *Journal of Molecular Biology* 287: 201-209.

Okano, M., Xie, S. and Li, E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics* 19: 219-20.

Okano, M., Bell, D.W., Haber, D.A. and Li, E., (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* 99: 247-257.

Olek, A. and Walter, J. (1997) The pre-implantation ontogeny of the *H19* methylation imprint. *Nature Genetics* 17: 275-276.

Oliver, A.W., Jones, S.A., Roe, S.M., Matthews, S., Goodwin, G.H. and Pearl, L.H. (2005) Crystal structure of the proximal BAH domain of the polybromo protein. *Biochemical Journal* 389: 657-664.

Ooi, S.K.T., Qui, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S., Allis, C.D., Cheng, X. and Bestor, T.H. (2007) Dnmt3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* 448: 714-717.

Pak, D.T.S., Pflumm, M., Chesnokov, I., Huang, D.W., Kellum, R., Marr, J., Romanowski, P. and Botchan, M.R. (1997) Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell* 91: 311-323.

Pascual, J., Martinez-Yamout, M., Dyson, H.J. and Wright, P.E. (2000) Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *Journal of Molecular Biology* 304: 723-729.

Pósfai, J., Bhagwat, A.S., Pósfai, G. and Roberts, R.J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Research* 17: 2421-2435.

Qiu, C., Sawada, K., Zhang, X. and Cheng, X. (2002) The PWWP domain of mammalian DNA methyltransferases Dnmt3b defines a new family of DNA-binding folds. *Nature* 9: 217-224.

Rai, K., Jafri, I.F., Chidester, S., James, S.R., Karpf, A.R., Cairns, B.R. and Jones, D.A. (2010) Dnmt3 and G9a cooperate for tissue-specific development in zebrafish. *Journal of Biological Chemistry* 285: 4110-4121.

Rai, K., Nadauld, L.D., Chidester, S., Manos, E.J., James, S.R., Karpf, A.R., Cairns, B.R. and Jones, D.A. (2006) Zebrafish Dnmt1 and Suv39H1 regulate organ-specific terminal differentiation during development. *Molecular and Cellular Biology* 26: 7077-85.

Robertson, K.D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F.A. and Jones, P.A. (1999) The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Research* 27: 2291-2298.

Rozen, S. and Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: A laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Santini, F., Harmon, L.J., Carnevale, G. and Alfaro, M.E. (2009) Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BioMed Central Evolutionary Biology* 9: 194-209.

Santos, F., Hendrich, B., Reik, W. and Dean, W. (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. *Developmental Biology* 241: 172-182.

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusbürger, M., Helm, M. and Lyko, F. (2010) RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes and Development* 24: 1590-1595.

Shimoda, N., Yamakoshi, K., Miyake, A. and Takeda, H. (2005) Identification of a gene required for de novo DNA methylation of the zebrafish *no tail* gene. *Developmental Dynamics* 233: 1509-1516.

Smith, T.H.L., Collins, T.M. and McGowan, R.A. (2011) Expression of the *dnmt3* genes in zebrafish development: similarity to Dnmt3a and Dnmt3b. *Development Genes and Evolution* 220: 347-353.

Smith, T.H.L., Dueck, C.C., Mhanni, A.A. and McGowan, R.A. (2005) Novel splice variants associated with one of the zebrafish *dnmt3* genes. *BioMed Central Developmental Biology* 5:23

Song, J., Rechkoblit, O., Bestor, T.H. and Patel, D.J. (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* 331: 1036-1040.

Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Längst, G. and Leonhardt, H. (2007) DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *Journal of Cell Biology* 176: 565-571.

Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A. and Cedar, H. (1982) Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proceedings of the National Academy of Sciences of the United States of America* 79: 61-65.

Streisinger, G., Walker, C., Dower, N., Knauber, D. and Singer, F. (1981) Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* 291: 293-296.

Takebayashi, S., Tamura, T., Matsuoka, C. and Okano, M. (2007) Major and essential role for the DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions. *Molecular and Cellular Biology* 27: 8243-8258.

Tucker, K.L., Talbot, D., Lee, M.A., Leonhardt, H. and Jaenisch R. (1996) Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. *Proceedings of the National Academy of Sciences of the USA* 93:12920-12925.

Van Emburgh, B.O. and Robertson K.D. (2011) Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. *Nucleic Acids Research* 39: 4984-5002.

Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J.R. III and Jia S. (2010) Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Molecular Cell* 33: 428-437.

Xu, G., Bestor, T.H., Bourc'his, D., Hsieh, C., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J. and Viegas-Péquignot, E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402: 187-190.

Yoder, J. A., Yen, R. C., Vertino, P. M., Bestor, T. H. and Baylin, S. B. (1996) New 5' regions of the murine and human genes for DNA (cytosine-5)-methyltransferase. *The Journal of Biological Chemistry* 271: 31092-31097.

Yoder, J.A. and Bestor, T.H. (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Human Molecular Genetics* 7: 279-284.

Yu, N., Baek, S.H. and Kaang, B. (2011) DNA methylation-mediated control of learning and memory. *Molecular Brain* 4: 5.

