The University of South Bohemia in České Budějovice
Faculty of Science

Activity of *Dnmt3l* gene in guinea pig oocytes

Bachelor thesis

Laura Andessner-Angleitner

Advisor: Mgr. Lenka Gahurová, Ph.D.

České Budějovice 2024

Andessner-Angleitner, L., 2024: Activity of *Dnmt3l* gene in guinea pig oocytes. Bc. Thesis, in English. – 48 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Annotation

This thesis investigates if the *Dnmt3l* gene is transcribed in guinea pig oocytes and if the *Dnmt3l* promoter is conserved between rodents belonging to the *Sciurognathi* and *Hystricomorpha* suborder.

Declaration

I declare that I am the author of this qualification thesis and that in writing it I have used the sources and literature displayed in the list of used sources only.

České Budějovice, 14.08.2024

Signature:

Table of Contents

Acknowledgements

At this point I want to thank everybody who supported me. First up, I have to thank my boyfriend Paul who is always cheering me on and giving me the best pep talks. I love you! I also want to thank my family, especially my dad, who is my biggest fan. I also have to thank, my second family, who embraced me with both arms and feet. Thank you all for loving and supporting me! We are always being spoiled when we drive home to Salzburg. I want thank my aunt Annerose who was always watching out for me and with whom I have the best phone calls. Last but not least, I want to thank my supervisor Mgr. Lenka Gahurová, Ph.D., for being such a supportive supervisor. She helped me with all my questions and I learned a lot working with her.

Abstract

DNA methylation is an epigenetic modification that transfers a methyl (-CH3) group to the 5-position of cytosine (5mC). The DNA methylation pattern is established by DNA methyltransferases (DNMTs). DNMT3A and DNMT3B are dependent on the cofactor DNMT3L. In general, the *Dnmt3l* gene is only expressed in oocytes of *Eumuroida* rodents. As the RNA-sequencing data presented a weak expression of the *Dnmt3l* gene in guinea pig oocytes, despite them not belonging to *Eumuroida* rodents, it was examined further. In mice, the annotated promoter for the *Dnmt3l* gene is located at the third intron of the *Aire* gene encoded on the complementary DNA strand. At the time of the project, the third intron of the *Aire* gene of guinea pigs, according to the Ensembl genome database, was not yet fully sequenced at the nucleotide level. For this purpose, primers were designed to amplify the third and fourth exons of the gene. The complete nucleotide composition of the third intron of the *Aire* gene was successfully identified using the Sanger sequencing procedure. For examining the *Dnmt3l* expression in guinea pig oocytes, suitable primer pairs were designed, targeting different parts of the gene. Analyzing the sequence of the third intron of the *Aire* gene demonstrated that guinea pig oocytes do not have a conserved promoter matching the rodents from the *Eumuroida* lineage. The weak expression readings resulted from a novel *Dnmt3l* promoter in guinea pig oocytes.

Chapter 1 Theoretical background

In this chapter, there is an introduction to epigenetic modifications, which includes the description of histone modifications, non-coding RNA, and DNA methylation. Due to its relevance for this project, the focus lies on DNA methylation.

1.1 Epigenetic modifications

Every multicellular organism is in need of cell-specific regulations. Although all cells of one individual have identical genetic material, the cells differ in morphological and physiological characteristics, depending on their location in the body and especially on their function. A representative among these cell type specific regulations are epigenetic modifications. These regulations control gene expression, without altering the genetic material in the process. Epigenetic modifications determine if specific genes are transcribed or transcriptionally silent (Stewart et al., 2016). Consequently, epigenetic modifications assist in defining the differentiation of a cell, for instance, whether it will develop into a heart or a nerve cell. In general, epigenetic modifications are not only influenced by inheritance but also by the behaviour, physical activities, eating habits, and age of an individual, as well as environmental factors. Environmental factors in the development phase are, for instance, stress or absence of nutrition, which can have long-lasting effects on the individual (Kanherkar et al., 2014). Epigenetic marks are mitotically stable and are maintained throughout the replication cycle of a cell (Skinner, 2011). There are three categories of epigenetic modifications: histone modification, non-coding RNA and DNA methylation, which are introduced in the next chapters 1.1.1- 1.1.3.

1.1.1 Histone modification

Histone modification is a post-translational epigenetic mechanism that affects the transcription process while leaving the DNA sequence unchanged. The DNA is stored in nucleosomes, which are the basic units of chromatin. In one nucleosome 147 base pairs of DNA are coiled around a core. The core consists of an octamer which contains four histones (H3, H4, H2A, and H2B) (Zhao & Shilatifard, 2019). Each core histone is composed of a globular region, which is the connecting site of a N-terminal tail. The N-terminal tail can undertake a variety of post-translational modifications, for instance, acetylation, methylation, and phosphorylation

(Inbar-Feigenberg et al., 2013). In histone acetylation, an acetyl group is transferred from acetyl-CoA to lysine in the N-terminal tail. This leads to changes in electrical charge, which weakens histone-histone interactions. Therefore, the chromatin is in a less condensed state that is more permissive for transcription. In histone methylation, lysine or arginine can be methylated with the assistance of specially assigned enzymes. This mechanism provides changes in the chromatin architecture by mono-, di- or trimethylating the amino acids. This offers diversity in the regulation of transcription, conferring either transcriptional activation or repression. Kinases regulates histone phosphorylation by adjoining phosphate groups. Known impacts of phosphorylation of histone proteins are DNA damage repair mechanism, regulation of chromatin state during mitosis as well as meiosis, and control over the transcription activity (Alaskhar Alhamwe et al., 2018).

1.1.2 Non-coding RNA

Non-coding RNA (ncRNA) describes RNA which is not coding for a protein. In eukaryotes, up to 90% of the genome is transcribed; however, only 1-2% is further translated into proteins. Therefore, the majority of transcribed RNA results in ncRNA. Increasing evidence indicates that the amount of ncRNA is proportional to the complexity of the organism. Non-coding RNA can be separated into two main groups: infrastructural ncRNAs and regulatory ncRNAs. Regulatory ncRNA is responsible for epigenetic modifications. For instance, miRNA is one of the major post-transcriptional gene expression regulators. It controls gene expression by affecting the translation or stability process of mRNAs (Kaikkonen et al., 2011). One miRNA can regulate multiple mRNAs and additionally, one mRNA can be influenced by more than one single miRNA. The miRNA is a perfect or almost perfect complementary to the 3´UTR of the targeted mRNA. The miRNA regulates the gene expression by base pairing to the sequence motif, through which the miRNA can inhibit translation or degrade the mRNA (Cai et al., 2009).

1.1.3 DNA Methylation

DNA methylation is one of the most extensively studied and mechanistically well-understood epigenetic modifications. In this mechanism, a methyl (-CH3) group is transferred from S-adenosyl-l-methionine to the 5-position of cytosine (5mC). This mechanism is presented in *Figure 1*.

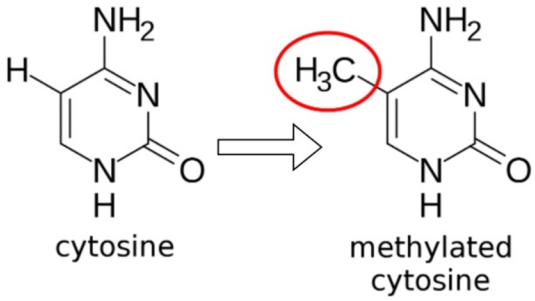

Figure 1: *DNA methylation of cytosine (5mC)*

If the promoter or first exon is methylated it results in gene silencing. However, if the gene body which is defined as the region of the gene beyond the first exon is methylated, the influence on the gene expression is not yet fully understood (Moore et al., 2012). DNA methyltransferases (DNMTs) are the enzymes that catalyze DNA methylation. *De novo* methylation is the process of attaching a methyl group to an unmethylated cytosine. This process involves the DNMT3 enzyme family, which includes DNMT3A, DNMT3B, DNMT3C and DNMT3L (Chen & Zhang, 2020). Rodents-specific (*Muroidea*) DNMT3C is considered to be limited to the male germline (Barau et al., 2016). DNMT3A as well as DNMT3B are the enzymes predominately responsible for the establishment of *de novo* DNA methylation (Unoki, 2019). *Dnmt3a* and *Dnmt3b* have a large N-terminal region containing Pro-Trp-Trp-Pro motif (PWWP) and ATRX–DNMT3–DNMT3L (ADD) domains, which can interact with different proteins. Their C-terminal regions possess a catalytic center (Zhang et al., 2010). DNMT3L is catalytically inactive but serves as a cofactor for both DNMT3A and DNMT3B (Chen & Zhang, 2020). DNMT3L works by uniting two duplicates of each enzyme (DNMT3A and DNMT3L) together to form a heterotetrameric complex (Jia et al., 2007). *Figure 2* displays a simple depiction of the DNMT3A-DNMT3L complex. The formation of this heterotetrameric complex leads to a positive stimulation of the DNMT3A activity. Moreover, the DNMT3L maintains DNMT3A stability.

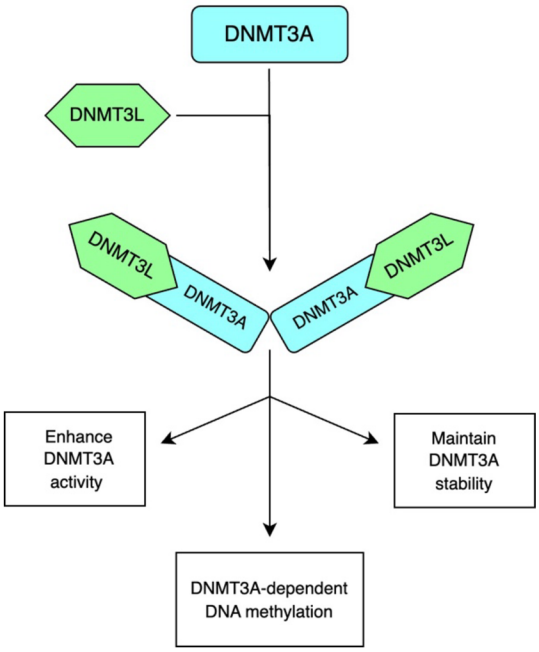

Figure 2: *Simplified graphical depiction of the DNMT3A-DNMT3L complex formation* (Veland et al., 2019a)

Once the DNA methylation is established, it is maintained through DNA replication by the enzyme DNMT1, which semi-conservatively copies the methylation pattern during cell division. This process effectively preserves the epigenetic memory. In mammals, most of the DNA methylation occurs at CpG dinucleotides (Chen & Zhang, 2020). CpG sites are specific regions within the DNA sequence, comprising a cytosine adjacent to a guanine. While DNA methylation primarily occurs at CpG sites, there is a considerable amount of non-CpG sites, which are also methylated (Moore et al., 2012). CpG islands (CGIs) are regions that contain a higher density of CpG dinucleotides and are, in general, hypomethylated (Inbar-Feigenberg et al., 2013). The CGIs seem to be mainly protected from methylation by transcription factor binding (Moore et al., 2012). The size of these regions is between 200 bp and several kilobases. The CGIs are commonly positioned on promoters of housekeeping genes and developmental genes (Inbar-Feigenberg et al., 2013). In mammals, more than two-thirds of the promoters are embedded in CGIs (Greenberg & Bourc'his, 2019). When CGIs are methylated, it leads to a stable silencing of the corresponding gene. However, normally CGIs promote the availability of a gene as well as its transcription factor binding sites and are rarely methylated (Moore et al., 2012). Instead, if silencing is required, it is achieved through silencing histone modifications.

DNA methylation is vital for the correct development of mammals and plays a key role in regulative processes, including X-chromosome inactivation as well as genomic imprinting.

In the early stage of embryogenesis, female embryos possessing two X-chromosomes experience an X-chromosome inactivation. In female mammals, one of the two X chromosomes in each somatic cell is transcriptionally inactive. The X-chromosome, which is inactivated, is randomly chosen (Inbar-Feigenberg et al., 2013). Genomic imprinting does not influence a whole chromosome but only specific genes. In genomic imprinting, only one parental allele (maternal or paternal) of a certain gene is transcriptionally active (Prawitt & Haaf, 2020). The other allele is strongly suppressed through the methylation of the imprinting control region (Altun et al., 2010).

1.1.4 Correlation of DNA methylation with histone modifications

Several research teams, for example, the one of Maxim V. C. Greenberg and Deborah Bourc'his, suggest multiple correlations and anticorrelations between DNA methylation and a variety of histone modifications (Greenberg & Bourc'his, 2019). Histone 3 lysine 4 trimethylation (H3K4me3) and histone 3 lysine 27 acetylation (H3K27ac), for example, present a negative relationship with DNA methylation and are associated with promoting gene expression (Hanna et al., 2018). H3K4me3 is a highly conserved epigenetic mark located at CpG islands as well as active promoters. In mammalian cells, H3K4me3 prevents *de novo* methylation at CGIs as well as gene promoters by inhibiting its interaction with DMNT3s (Hanna et al., 2022). In contrast to these histone modifications, unmethylated lysine 4 (Lys 4) of histone H3 (H3K4me0) was demonstrated to be favored by the DNMT3A–DNMT3L complex. The DNMT3A-DNMT3L heterotetrameric complex shares an ADD domain, which interacts with the N-terminal tail of histone 3 (H3) (Otani et al., 2009). H3K36me3 is a repressive modification linked to DNA methylation. A region marked by H3K36me3 is recognized by the PWWP domain of *Dnmt3a* (Dhayalan et al., 2010). H3K36me3 is attached to the DNA by SETD2, which is a conserved enzyme. The enzyme SET2 attaches the H3K36me3 over transcribed gene bodies and attracts DNA methylation. Typically, CpG islands are low in H3K36me3. However, as oocytes grow, intragenic CpG islands that will be methylated, gain H3K36me3 enrichment. The ablation of SETD2, specifically in oocytes, results in the loss of H3K36me3 and consequently eliminates gene-body methylation (Demond & Kelsey, 2020).

1.1.5 DNA- Methylation in the oocytes of mice

During mammalian development, DNA methylation undergoes two global events of reprogramming one in the germline and another one after fertilization. Reprogramming is necessary to eliminate established lineage patterns and to reinstate pluripotent potential in cells (Sendžikaitė & Kelsey, 2019a). This demethylation process is either passive or active. In passive demethylation the *Dnmt1* or its cofactors are absent, which leads to a dilution of methylation over successive rounds of DNA replication. In active demethylation, multiple chemical reactions are needed to remove the methyl group from cytosine. One active demethylation process involves the ten-eleven translocation (Tet) enzymes TET1, TET2, and TET3. These enzymes modify cytosine by attaching a hydroxyl group to its methyl group, resulting in the formation of 5-hydroxymethylcytosine (5hmC). The 5hmC can be modified further by two distinct mechanisms. Either by oxidation, in which 5hmC is first converted to 5-formyl-cytosine and then to 5-carboxy-cytosine. Or by AID/APOBEC, which transforms 5hmC into 5-hydroxymethyl-uracil. In the final step, the base excision repair (BER) recognizes the modified base and replaces it with plain cytosine (Moore et al., 2012).

In the early stages of the development of germ cells the DNA methylation is almost completely removed (Chen & Zhang, 2020). The demethylation procedure in germ cells is separated into two steps. The first step is the passive demethylation of most DNA sequences. In the second demethylation step, the DNA is actively demethylated by TET1 and TET2 enzymes which mainly affects germline-specific and imprinted genes (Greenberg & Bourc'his, 2019). During oocyte growth, DNA methylation is primarily established on an almost empty canvas. Only certain repetitive elements maintain their methylation pattern throughout oogenesis (Smallwood & Kelsey, 2012). Before the oocytes and sperm are fully mature the methylation pattern is re-established depending on the sex of the germline (Chen & Zhang, 2020). In mice, the sex-specific epigenome in the germline in males is established shortly after birth and in females after reaching adulthood (Sendžikaitė & Kelsey, 2019). Sperm methylome is revealed to have approximately 90% of the CpG sites methylated. In comparison, only around 40% of the CpG sites in oocytes are methylated. The oocyte methylome is assembled out of hyper- and hypomethylated segments (Stewart et al., 2016). In oocytes, methylation primarily occurs in actively transcribed regions, including gene bodies. This creates a bimodal pattern in the oocyte genome, characterized by hypermethylated gene bodies separated by intergenic or transcriptionally inactive regions with low methylation levels (Demond & Kelsey, 2020). In sperm, the DNA methylation is more evenly distributed (Stewart et al., 2016). Throughout

the growth and development of the oocytes, DNA methylation is ongoing. The non-CpG site methylation occurs more frequently in oocytes (Sendžikaitė & Kelsey, 2019b).

The two germ-cell genomes that unite during fertilization experience the second reprogramming wave. As the development proceeds towards pluripotency, the inherited gamete-specific DNA methylation patterns are demethylated (Greenberg & Bourc'his, 2019). How the second reprogramming process proceeds in detail is not yet fully understood and is being debated (Greenberg & Bourc'his, 2019). The pattern for the establishment of somatic cells is identical on both parental genomes and is re-established *de novo* (Prawitt & Haaf, 2020).

The expression of DNMT3A, DNMT3B, and DNMT3L in developing oocytes is synchronized. Their concentrations rise during oocyte maturation, reaching a peak at the GV stage when *de novo* methylation is fully accomplished, and then decline once the oocyte reaches the MII stage (Sendžikaitė & Kelsey, 2019a). The prevailing model recognizes DNMT3A as the primary enzyme responsible for *de novo* DNA methylation in oocytes (Kaneda et al., 2004). DMNT3A exists in two isoforms. *Dnmta1* is the full-length version of the protein, while *Dnmt3a2* lacks a segment at the N-terminus. Interestingly, *Dnmt3a2* is the predominant isoform found in oocytes (Kibe et al., 2021). In the case of DNMT1, oocytes produce a specific isoform known as *Dnmt1o*, which results from an alternative 5′ exon. It is 118 amino acids shorter than the full-length version of DNMT1 produced in somatic cells. The DNMT1O enzyme is mainly present in a high abundance in growing oocytes (Sendžikaitė & Kelsey, 2019a). The removal of either DNMT3A or DNMT3L results in a similar nearly complete loss of *de novo* methylation. Both proteins present limited DNA binding specificity. Their binding and activity are suggested to be influenced by the post-translational modification of histone proteins. The correlation between histone modifications and DNA methylation is described in more detail in chapter 1.1.4. It is suggested that the interaction of DNMT3A-DNTM3L complex with H3K4me3 and then H3K36me3, might be accountable for the gene-body methylation pattern in oocytes (Demond & Kelsey, 2020).

1.1.6 The *Dnmt3l* gene

DNMT3L is until now, only found in mammals with genomic imprinting (Sendžikaitė & Kelsey, 2019a). The DNMT3L-based model established in mice remains widely regarded as a general model for mammals (Chen et al., 2020). Unlike DNMT3A, DNMT3L lacks the PWWP domain and also misses a significant part of the catalytic methyltransferase domain (Dhayalan et al., 2010). In mouse oocytes, maternal DNMT3L is required to establish DNA methylation as well as genomic imprinting (Veland et al., 2019a). In somatic cells of mammals (Veland et al., 2019b) as well as human oocytes, the DNMT3L is transcriptionally silent (Petrussa et al., 2014). However, in mice, the protein is abundantly expressed in developing germ cells, embryos in early stages, and mESCs (Veland et al., 2019b). In mice, the promoters for the upstream autoimmune regulator (*Aire)* and *Dnmt3l* gene are closely positioned on the 10^{th} chromosome (Zamudio et al., 2011). The AIRE (Autoimmune Regulator gene) proteins function as a transcription factor that is responsible for establishing and preserving immune tolerance. Moreover, AIRE proteins are present in spermatogonia as well as spermatocytes. If a deficiency of this protein is detected, it results in infertility in males. In spermatids, evidence presents a co-regulatory interaction between the *Aire* and *Dnmt3l* gene (Zamudio et al., 2011).

In germ cells, sex-specific transcripts are produced by different *Dnmt3l* promoters. In the late pachytene spermatocytes, an internal promoter leads to the production of three short transcripts that lack significant open reading frames (ORFs). These three transcripts start from a novel exon which is positioned between exon 9 and 10. These short alternative transcripts were detected in the adult testis of mice. In growing oocytes, the full-length *Dnmt3l* protein is transcribed from the promoter which is within the third intron of the *Aire* gene and is encoded on the complementary DNA strand to the *Aire* gene. The specific functions of the different active *Dnmt3l* proteins are regulated by the distinct promoters (Shovlin et al., 2007).

The TATA-binding protein (TBP), a general transcription factor, is present exclusively in primordial follicular oocytes and is absent in later stages of oocyte development. As oocytes grow, TBP is replaced by its paralogue TBPL2 (also known as TBP2 or TRF3). As a result, TBPL2 is abundantly expressed in growing oocytes. Transcript analyses reveal that TBPL2 helps regulate oocyte-specific transcription. This indicates a TBPL2-dependent transcriptome in oocytes during growth. The TBPL2/TFIIA complex is commonly believed to influence *Dnmt3l* expression by attaching to the identified motif. However, how the TBPL2 recognizes and interacts with the corresponding gene is not fully uncovered (Yu et al., 2020).

1.2 Position of Guinea Pigs in the Rodent Clade

For numerous years, rodents, especially mice, have played a crucial role in scientific research. Mice are often presented as a general model for mammals; however, one model alone does not demonstrate the diversity of biological mechanisms among all mammalian species. The reasons for the popularity of mice for scientific purposes are the short generation intervals and the low expenses for feeding and housing (Maxeiner et al., 2021). In this project the *Eumuroida* lineage, containing the *Muridae* (mice, rats, gerbils) and *Cricetidae* (hamsters) families are compared to guinea pigs, damara mole/rats, and naked mole/rats. Mice (M*us musculus*), rats (*Rattus rattus*), and mongolian gerbil (*Meriones unguiculatus*) belong to the same family of *Muridae*, which is in the superfamily *Muroidea*. This superfamily is included in the infraorder *Myomorpha*, which is in the suborder *Sciurognathi*. Chinese hamster (*Cricetulus griseus*) and golden hamster (*Mesocricetus auratus*) for example belong to the *Cricetidae* family, which is also in the *Muroidea* superfamily. In contrast to that, guinea pigs (*Cavia porcellus*) are a part of the subfamily *Caviinae*, which belongs to the *Caviidae* family. This family is part of the infraorder *Caviomorpha*, which belongs to the suborder *Hystricomorpha*. Naked mole rats (*Heterocephalus glaber*) and damara mole rats (*Fukomys damarensis*) are in the *Bathyergidae* family. This family also belongs to the *Hystricomorpha* suborder (Cox & Hautier, 2015). *Figure 3* depicts the mouse-related clade, the squirrel-related clade, and the guinea pig-related clade. *Figure 3* also displays the Latin family, superfamily, and suborder name. As shown in this graph, the naked and damara mole-rats belong to the guinea pig-related clade. Moreover, the graph demonstrates how closely related mice and rats are. Hamsters on the other hand do not belong to *Muridae* but to the *Cricetidae* family.

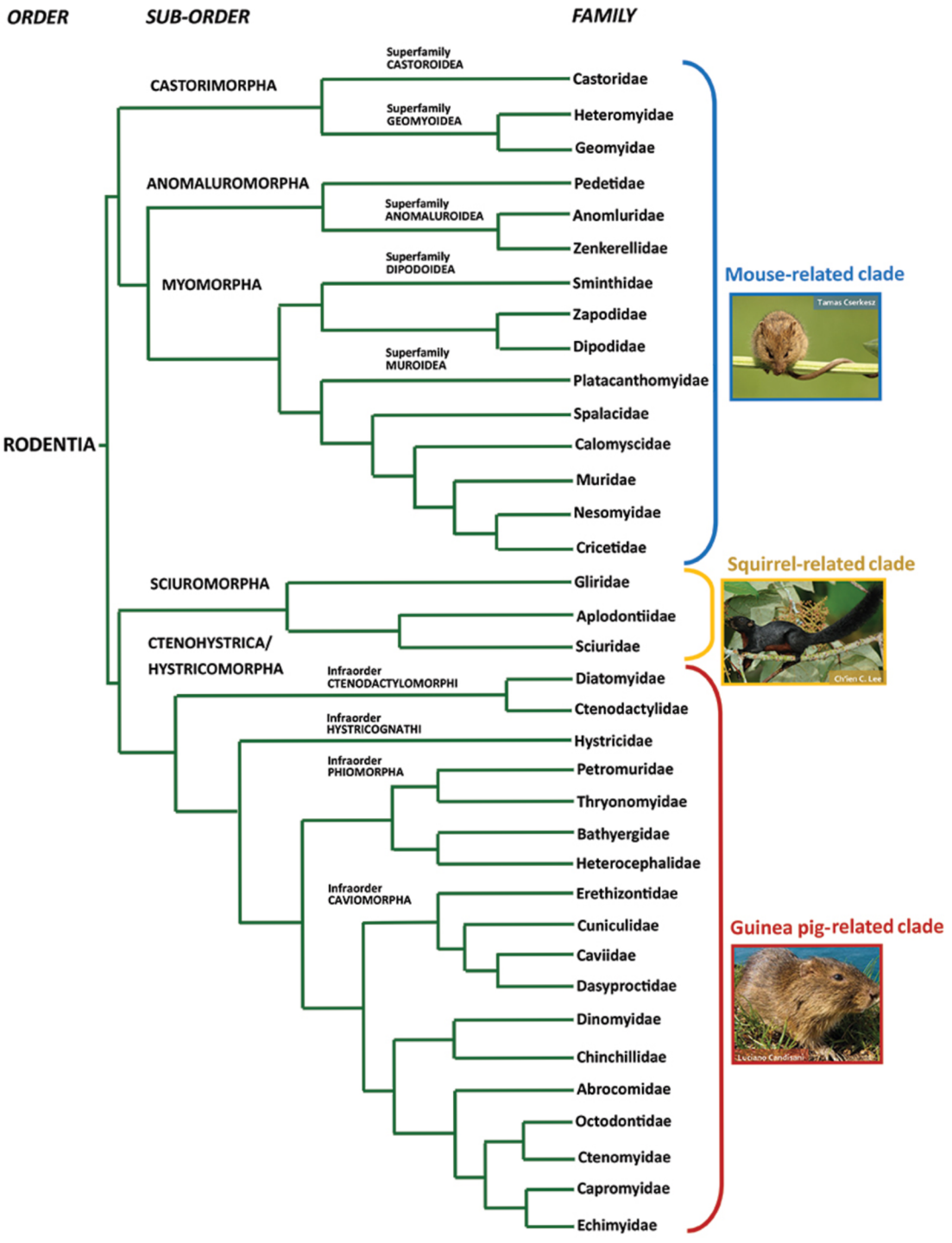

Figure 3: *Phylogenetic tree of rodents from: https://small-mammals.org/small-mammals/#rodents (last accessed on: 3.08.2024)*

Chapter 2 Aims

This bachelor thesis aims to advance the knowledge of the *Dnmt3l* expression in mammalian oocytes. DNMT3L helps establish *de novo* DNA methylation in oocytes and is generally only observed in *Eumuroida* rodents. In guinea pigs, a weak expression of the *Dnmt3l* gene was observed through RNA-sequencing, despite them not belonging to the *Eumuroida* lineage. This was investigated by further examining guinea pig oocytes. In mice, the annotated promoter for the *Dnmt3l* gene is located within the third intron of the *Aire* gene on the complementary DNA strand. This leads to the question of whether the annotated *Dnmt3l* promoter is conserved across rodents or if guinea pigs have a novel promoter sequence. This bachelor project aims to meet the following goals:

- Determine the sequence of the third intron of the *Aire* gene in the guinea pig genome.
- Identify the promoter of the *Dnmt3l* gene variant in guinea pig oocytes.
- Compare the obtained sequence of the third intron of the *Aire* gene of guinea pigs to other rodents to find out if the *Dnmt3l* promoter of the *Eumuroida* lineage is conserved in guinea pigs.

By addressing these points, this thesis explores the field of oocyte *de novo* DNA methylation in rodents outside the *Eumuroida* lineage.

Chapter 3 Materials and Methods

This chapter presents the methods and materials employed in this study. The methods are explained to a level of detail which allows reproduction of the results but not further since this would be beyond the scope of this thesis. All experiments were conducted in the laboratory of the University of South Bohemia in Budweis. To avoid contamination, the workplace and used tools were always cleaned at the beginning of every experiment with 70% ethanol. If not differently stated, the water used was HPLC water.

3.1 Determination of the third intron of the *Aire* gene sequence

Ensembl genome database (ensembl.org) provides a substantial part of the domesticated guinea pig genome. However, there is a fraction of the third intron of the *Aire* gene that was not determined at nucleotide level. This experiment series aimed to amplify the third intron of the *Aire* gene and sequence it by using Sanger sequencing.

3.1.1 DNA extraction

Two samples of fresh guinea pig liver tissue were acquired by dissecting two female individuals. The collected tissue was immediately used to perform a DNA extraction. Therefore, the tissue samples were placed in 2 mL Eppendorf tubes. After that, 1200 µL of STE buffer (1 M NaCl, 200 mM Tris-HCl, 100 mM EDTA) and 30 µL of proteinase K (10 mg/mL, Merck) were added to each sample. The test tubes were incubated at a heating block at 55 °C with 900 rpm for 4 hours. The dissolution process was enhanced by manually stirring the sample every 30 minutes. After 4 hours, the test tubes were heated for another 5 minutes at 70 °C to inactivate the proteinase K. Then, the samples were placed on ice for 5 minutes. Afterwards, the tubes were positioned in the centrifuge for 10 minutes at 10°C at full speed (14000 rpm). The supernatant was transferred into a new 2 mL test tube, containing 700 µL 100% isopropanol. Due to the isopropanol, the DNA precipitated and became visible. To obtain the DNA, the sample was centrifuged at 4 °C for 5 minutes at full speed. Consequently, the DNA formed a pellet at the bottom of the test tube. This way, the isopropanol could easily be removed as supernatant. The DNA pellet was washed with 500 µL of 70% ethanol. The test tube was again placed in the centrifuge for 5 minutes at 4 °C at full speed. The supernatant was discarded.

The DNA was dried for a few minutes by opening the test tube lid to allow the remaining ethanol to vaporize. Afterwards, 300 µL of water was inserted into the test tube to dissolve the DNA. In the last step, the sample was placed on the heating block at 65°C for 5 minutes to assist the dissolution of DNA. The genomic DNA was then stored at -18°C.

3.1.2 Polymerase chain reaction (PCR) from extracted genomic DNA

This project aimed to amplify the unknown sequence of the third intron of the *Aire* gene. Suitable primers targeting the required region were designed by using the program Primer3web v4.1.0 (https://primer3.ut.ee/). These primers are presented in *Table 1*.

Table 1: *Primers designed for the third intron of the Aire gene. The cen presents the forward and reverse primers as well as their oligonucleotide sequence The size of the PCR product was expected to be 517 base pairs (bp).*

Primers	Forward(F)/ Reverse(R)	DNA/RNA	Primer name	Sequence	Size (bp)
Aire intron3	F	DNA	Aire3_gp_F	AAGACTCCTGTACTGCCACC	517
	R	DNA	Aire3_gp_R	GCTGCAACTCCGAATTACCC	

For each experiment, a negative control was prepared. The negative control contained water instead of the genomic DNA but was otherwise treated the same as the samples. The genomic DNA was diluted with water in a 1:20 ratio to reach a concentration of approximately 50 ng μL^{-1}. The samples were always prepared on ice. The PCR reaction was prepared with a high-fidelity DNA polymerase within the Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB). *Table 2* displays the PCR reaction mixture.

Table 2: *PCR reaction mixture was prepared in 0.2 mL PCR tubes.*

Volume / µL	Reagent	Original concentration
10	Mastermix	2X
1	Forward primer	10 µM
1	Reverse primer	10 µM
7	Water	-
1	DNA	50 ng μL^{-1}

After the preparation of the PCR reaction mixture, it was shortly vortexed and centrifuged. The PCR program, as displayed in *Table 3*, was chosen due to its suitability for the high-

fidelity DNA polymerase. This program was then modified in further experiments to obtain purer products.

Table 3: *Basic PCR program suitable for the high-fidelity DNA polymerase. The temperatures used are listed with the corresponding duration time. The red marked area demonstrates the three steps that are repeated 30 times.*

Steps of PCR	Temperature / °C	Time / s
Initial denaturation	98	30
Denaturation	98	10
Annealing	59	15
Extension	72	30
Final extension	72	600

The annealing temperature (°C) was modified during the experiment series: 59, 61, 60, 62, 63, 65, 68, 69, 70. The experiments conducted at 61°C and 63°C were performed by using two different DNA dilution ratios: 1:20 and 1:100. At the other annealing temperatures, the only DNA concentration ratio used was 1:20. After several experiments the final annealing temperature, which yielded the wanted single DNA fragment, was at 70 °C. After successfully modifying the PCR program, which is presented in *Table 4*, the procedure was repeated with an increased reaction volume, as displayed in *Table 5*.

Table 4: *Modified PCR program suitable for the third intron of the Aire gene. The red square highlights the three steps which were repeated 30 times.*

Steps of PCR	Temperature / °C	Time / s
Initial denaturation	98	30
Denaturation	98	10
Annealing	70	15
Extension	72	30
Final extension	72	600

Table 4 displays the optimized PCR program which yielded the wanted single DNA fragment. As visible in *Table 4*, the final annealing temperature was at 70 °C. The rest of the original

PCR program (*Table 3*) remained the same. The individual components of the increased reaction volume are demonstrated in *Table 5*.

Table 5: *The PCR reaction mixture with an increased reaction volume was prepared in a 0.2 mL PCR tube. To obtain a higher DNA concentration for sequencing, undiluted DNA was added to the test tube.*

Volume / µL	Reagent	Concentration
25	Mastermix	2X
2.5	Forward primer	10 µM
2.5	Reverse primer	10 µM
17.5	Water	-
2.5	DNA	undiluted

After preparing the PCR reaction mixture the test tube was shortly vortexed and centrifuged. The higher reaction volume was beneficial because one part of the reaction was used for gel electrophoresis to confirm the presence of the product. The other part was used for purification before sequencing.

3.1.3 Gel electrophoresis, purification, and sequencing of the PCR product

The agarose gel was prepared during the last 30 minutes of the running PCR procedure. First, the gel tray was prepared by closing both open ends with casting dams. Then the comb was inserted into the assembly. The amount of agarose powder and solvent was calculated according to the size of the used electrophoresis instrument. In this experiment series, the quantity of agarose powder was 0.8 g which equals 1% of the total volume (80 mL). The amount was weighed and put into an Erlenmeyer flask. Then 80 mL of 1X TAE (Tris-acetate-EDTA) was poured into the flask. After that, the Erlenmeyer flask was heated in a microwave to dissolve the agarose powder. During the heating process, the microwave was frequently stopped to check on the solvent's activity and to mix it for better dissolution. Once, the agarose powder was fully dissolved the flask was cooled, by placing it under cold running water, to reduce the evaporation process. This was done to subsequently prevent the evaporation of Ethidium Bromide (EtBr), which was added in the next step. 1.5 µL of EtBr was pipetted into the Erlenmeyer flask. Since EtBr effectively binds between the base pairs of the alpha-helix, it is commonly used to visualize DNA fragments in the electrophoresis gel. These DNA fragments

are visualized under UV-light by illumination (Sigmon & Larcom, 1996). The final solution was gently mixed and poured into the gel tray. During the polymerization process of the gel, the test tubes were taken out of the PCR machine. The sample and negative control were vortexed and centrifuged. 5 µL out of the 20 µL PCR reaction mixture was transferred into a new 0.2 mL PCR tube. In case of the increased reaction volume of 50 µL, used for the final product purification, 10 µL was transferred to a new 0.2 mL PCR tube. Then 1-2 µL of 6X purple gel loading dye was added to each solution. The sample and negative control were vortexed and centrifuged. 4 µL of a pre-prepared New England Biolabs Quick-Load 100 bp DNA ladder was transferred to a new 0.2 mL test tube. The test tubes were then taken to the electrophoresis device. The comb was removed from the gel and the solidified gel was placed into the buffer tank. The buffer tank was filled with the 1X TAE buffer to the maximal line. Subsequently, the samples were pipetted into the wells of the gel. The electrophoresis was working at 80 V, for approximately 30-50 minutes. The gel was analyzed by a Gel Imager 200 from Azure Biosystems. To determine whether the DNA concentration was sufficient for the Sanger sequencing procedure, 1 µL of the remaining sample was measured using the NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer. The PCR product (of the PCR reaction mixture with an increased volume of 50 µL) was purified using the Monarch Nucleic Acid Purification Kit (BioLabs) according to the manufacturer's instructions. After that, the product was sequenced using the Sanger sequencing procedure with PCR primers employed as sequencing primers.

3.2 *Dnmt3l* expression in guinea pig oocytes

The expression of *Dnmt3l* in guinea pig oocytes was questioned due to the RNA-sequencing data demonstrating a low level of the transcribed protein. Therefore, the project focused on finding out if *Dnmt3l* is expressed in guinea pig oocytes and whether the promoter is conserved within rodent species. Or if *Dnmt3l* isoforms are transcribed using a novel promoter.

3.2.1 RNA extraction from oocytes

For the total RNA isolation of guinea pig oocytes, the PicoPure RNA isolation Kit (Life Technologies) was used. Before starting the procedure, the heating block was warmed up to 42 °C. After 10 µL of extraction buffer (XB) had been inserted into an empty 0.5 mL Rnase & Dnase free tube (provided in the kit), the oocytes, which were obtained by an experienced member of the laboratory, were transferred to the test tube. This was followed by three cycles of freezing the sample tube in liquid nitrogen and then melting it. Then, 10 µL of XB was inserted and mixed by pipetting. The test tube was placed in the heating block at 42°C for 30 minutes. Three minutes before the end of the incubation time a column was prepared. In the course of this, 250 µL of CB buffer was added into the column and the column was left for five minutes at room temperature. After that, the column was centrifuged for one minute at maximum velocity (13000 g). The test tube was then taken out of the heating block and placed in a centrifuge at 16°C at a velocity of 2200 g for two minutes. 20 µL of 70 % ethanol was inserted into the tube and mixed by pipetting. Then, the mixture was transferred to the previously prepared column and centrifuged for two minutes at the lowest possible velocity (100 g). Directly afterwards, the column was centrifuged at 1300 g for 30 seconds. 100 µL of WB1 buffer was pipetted into the column and subsequently, centrifuged at 6900 g for one minute. 100 µL of WB2 buffer was added and centrifuged at 8000 g for one minute. At last, 100 µL of WB2 buffer was inserted and centrifuged at maximum speed (13000 g) for two minutes. The flowthrough was discarded by pipetting. The sample was centrifuged at 13000 g for one minute. Then the column was transferred to a new 0.5 mL collection tube. 9.5 µL of NFW was inserted and the column was incubated at room temperature for one minute. The sample was centrifuged at 2450 g for one minute and subsequently, at 13000 g for one minute. The RNA sample was then stored at -80°C.

3.2.2 Reverse transcription

The RNA extracted from guinea pig oocytes was reverse-transcribed into cDNA using Superscript III Reverse Transcriptase (ThermoFisher Scientific) according to manufacturer's instructions. 10 µL DNA-free RNA (approximately 100 ng), 1.5 µL oligo $d(T)_{16}$ (50 µM, ThermoFisher Scientific), 4 µL dNTPs (2.5 mM each, ThermoFisher Scientific) and 4 µL water were mixed into a 0.2 mL test tube. The test tube was then placed into the PCR machine at 65 °C for five minutes. Directly afterwards, the tube was cooled on ice for two minutes. Then,

6 µL of the 5X First Strand Buffer was added. 1.5 µL of DTT (0.1M) as well as 1.5 µL of Rnase OUT were inserted into the test tube. The solution was incubated in the PCR machine at 50 °C for two minutes. Next, 1.5 µL SuperScript III Reverse Transcriptase (ThermoFisher Scientific) was added and the reaction was incubated at 50 °C for 60 minutes. The reaction was stopped by heating the tube at 70 °C for 15 minutes. The resulting cDNA was stored at - 20 °C until further use.

3.2.3 PCR with oocyte cDNA

The cDNA, which was obtained as described in chapter 3.2.1 and 3.2.2, was used for the experiments below. To confirm that the *Dnmt3l* gene is expressed and that it starts from the annotated promoter, a set of six primer pairs aligning to multiple locations across the *Dnmt3l* transcript was applied (ENSCPOT00000004115.3 in Ensembl genome database). The primer pairs were designed using Primer3web v4.1.0 (https://primer3.ut.ee/). Two primer pairs targeted the beginning of the transcript, amplifying the region between exon 1 and a downstream exon. The other four primer pairs focused on different downstream regions, amplifying areas between various downstream exons. The specific regions targeted by these primer pairs are displayed in *Table 6*.

Table 6: *The table presents the required primers designed for the Dnmt3l gene. The forward (F) and reverse (R) primers as well as their oligonucleotide sequence are listed. Additionally, the exact size of each PCR product is presented. The numbers in the "Primer name" column refer to the corresponding exon where the primers align.*

Primers	Forward(F)/ Reverse(R)	DNA/RNA	Primer name	Sequence	Size (bp)
Dnmt3L_ex_1_2	F	RNA	Dnmt3l_gp_ex_1_2_F	CGGTCCAGGAAAGAGCAAAG	133
	R	RNA	Dnmt3l_gp_ex_1_2_R	GGTCTTGAGGATGGAGCAGG	
Dnmt3L_ex_4_6	F	RNA	Dnmt3l_gp_ex_4_6_F	GAAGTCTCAGTGAACCAGCG	144
	R	RNA	Dnmt3l_gp_ex_4_6_R	ATCATAAAGGAACAGGGCGC	
Dnmt3L_ex_6_8	F	RNA	Dnmt3l_gp_ex_6_8_F	GCGCCCTGTTCCTTTATGAT	325
	R	RNA	Dnmt3l_gp_ex_6_8_R	ATCGGCTGTCTCTTCCACAC	
Dnmt3L_ex_8_10	F	RNA	Dnmt3l_gp_ex_8_10_F	AGAGCTGATGAGTTTGGGCT	127
	R	RNA	Dnmt3l_gp_ex_8_10_R	GCCGTACACCAGGTCAAATG	
Dnmt3L_ex_11_12	F	RNA	Dnmt3l_gp_ex_11_12_F	CCCTCAGCCTTTCTTCTGGA	179
	R	RNA	Dnmt3l_gp_ex_11_12_R	GTCCCCAGTACCTCCTCAC	
Dnmt3L_ex_1_4	F	RNA	Dnmt3l_gp_ex_1_4_F	GCGCCCACTGTTATACTGC	335
	R	RNA	Dnmt3l_gp_ex_1_4_R	ATGTTCCGCTGGTTCACTGA	

For these experiments, the Enzo Life Sciences Ampigene Taq Mix (ThermoFisher Scientific) was used. The cDNA was diluted 1:100 with water to obtain a concentration of approximately

100 ng μL^{-1}. The preparation of the PCR reaction mixtures was always done on ice and is demonstrated in *Table 7*.

Table 7: *The PCR reaction mixture was prepared in 0.2 mL PCR tubes. After the addition of the Ampigene Taq mix, the primer pairs, and water, the test tubes were shortly vortexed and centrifuged. 4 µL of the mixture was extracted and transferred to a new sample tube. Then 1 µL cDNA was added. For the negative control again 4 µL of the mixture was extracted and 1 µL water was added.*

Volume / µL	Reagent	Original concentration
12.5	Ampigene Taq Mix	2X
1	Forward primer	10 µM
1	Reverse primer	10 µM
10.5	Water	-
1	cDNA	100 ng μL^{-1}

After the preparation of the PCR reaction mixtures, the test tubes were shortly vortexed and centrifuged. The samples were then positioned in the PCR machine. The PCR program, displayed in *Table 8*, was selected due to its suitability for the Ampigene Taq polymerase. Subsequent experiments involved modifying this program to enhance the purity of the resulting products.

Table 8: *Basic PCR program suitable for the Ampigene Taq polymerase. In the table, the temperature and its corresponding duration time are presented. The red square highlights the three steps that were repeated 35 times.*

Steps of PCR	Temperature / °C	Time / s
Initial denaturation	95	60
Denaturation	95	15
Annealing	58	15
Extension	72	20
Final extension	72	600

The annealing temperature (°C) was modified during the experiments: 58, 56, 54, 56. After conducting the experiment several times, the best annealing temperature was established to be

56 °C. The duration time of the annealing step was increased to 25 s. Furthermore, the elongation step was raised to 40 s. The number of cycle repetition was adjusted to 40. All relevant modifications done on the PCR program to enhance the results are summarized in *Table 9*.

Table 9: *Modified PCR program. The temperature and its corresponding duration time are demonstrated below. The red square highlights the three steps that were modified to repeat for 40 times.*

Steps of PCR	Temperature / °C	Time / s
Initial denaturation	98	30
Denaturation	98	10
Annealing	70	25
Extension	72	40
Final extension	72	600

3.2.4 Gel electrophoresis of PCR product

The gel electrophoresis procedure was described in detail in chapter 3.1.3. In this chapter, only the differences between these methods are stated. During the last 30 minutes of the PCR procedure, the 1.5 % agarose gel was prepared. Therefore, 1.2 g of the agarose powder was dissolved in 80 mL of 1X TAE (Tris-acetate-EDTA). The more concentrated agarose gel (in contrast to 3.1.3) was prepared to improve the separation of individual cDNA bands. Once the PCR procedure was finished the samples were placed back on ice. Then 1 µL of the purple gel loading dye was directly added to the sample tubes. Additionally, 4 µL of a pre-prepared New England Biolabs Quick-Load 100 bp DNA ladder was transferred to a new test tube. The ladder and the samples were loaded into the wells of the gel. The electrophoresis was working at 90 V for 30-50 minutes. The gel was again analyzed by a Gel Imager 200 from Azure Biosystems.

3.3 Comparison of the third intron of the *Aire* gene of guinea pigs with other rodents

A selected part of the third intron of the *Aire* gene, containing the oocyte transcription factor binding site and the transitional start site of mice, was downloaded from Ensembl. These annotated sites of the third intron of the *Aire* gene on the complementary DNA strand in mice are compared to other rodents belonging to the *Sciurognathi* and *Hystricomorpha* suborder. Only this specific segment of the third intron of the *Aire* gene is depicted due to better visualization of the important transcription sites. The sequences of the other rodents were downloaded from either Ensembl or the NCBI database. These two websites retrieve publicly available sequencing data from the International Nucleotide Sequence Database Collaboration. *Sequence 1- Sequence 8* are presented below, with the corresponding accession code and source of downloading.

The following rodents belong to the suborder *Sciurognathi* and the *Muridae* family:

Mouse (*Mus musculus*):

Accession code: ENSMUSG00000000730 (Ensembl)

```
GTACACTCAAGAGGAGCTAGCCAGGGTTGCTGGGCCCTCCCCAACCGGCTCTTAGGAGCTTCTGTCT
TACTGACACCACCCCAGGGCCAGCCTGCCAGGGTCACAGAGTCACCTCTGAGCCCTCAGACCTGAGC
ATTGGAGGAGGCCCACAGCCTCTCAGCGTCTTACTGTCCCAAAGGCTGAGTTTCTGGGCGGTGAGGC
AGGCAGGTGGTTTTGATTTCCTTTCTGTTGAAGAAGGAAACAGCCCATCACAGCTTAAGAACCGTCG
ATCTGACCCTTACCAGCTGCTCTCTCTCCCATCCTCACTTTCTACCCTGGATCCGTCAACATGACCC
CAGCCCAGAAAAGTGGGCCCAGGCTGCCTCTACCTCCCCTTCGCAG
```

Sequence 1: *Segment of the third intron of the Aire gene of mice*

Rat (*Rattus rattus*):

Accession code: ENSRNOG00060022269 (Ensembl)

```
GTACTCAAGAGGAACTACCCAGGGTTGCTGGGCCCTCCCCGACCGGCTCTCAGGAGCTTCTGTCTTA
CTGACACCACCCCAGGGCCAGCCTGCCAGAGCTGGAGTCACCTTTGAGCGCTCAGACCTGAGCATTG
GAGGAAGAGGGCACAGCCTCTCAGCGTCTTACTGTCCTGAAGGCTGGGTTCCAGGGTGGTGGGTGAG
GCAGGCAGGTGGTTTTGATTTCCTTTTCTGTTGAAGAAGGAAACACAGCCCAGACCAGCTTAAGAAC
CCTCAGGGATCCAGGGAACCCTTGATATGACCCTTACCAGGTGCTTTCTCCCCAACCTTACTTGACA
TCCTGGACCTGTCAACACCCCCTAACCCAACCTGTACCTCCCCTTCACAG
```

Sequence 2: *Segment of the third intron of the Aire gene of rats*

Mongolian gerbil (*Meriones unguiculatus*):

Accession code: ENSMUGG00000017526 (Ensembl)

```
GTACTCAGGGAGAGGAGCTGGGTCCTCCCCAAATGACTCTCAGGAGCTTACTGATACCACCCCAAGG
AGGGCCAGAGCCGCAGAGTCACGTCTGAGTGCTCAGACCTGAGCATCGGTGGAAGCTCACGGCCTCT
CAGCTTCTCACTGGTCTGAAGGCCGAGTTCCTTCGTGGTGAGGCAGGCAGGTGCTTTTGATTTCCTT
TCTGTTGAAGAAGGAAACAGCCTGTGCCAGCTTAAGAACTCTCAGGGAGCCAGGGAACCCATGATCT
GACCCTGCCAGGTGCTTTCTCCCACCCCCCCACCTCACTTTATACCCTGGGCCCACCAGCACCACCC
CAGCCCAGAAAGGAGGCCTAGCCTGCCTCCACCTCCCTTTCCCAG
```

Sequence 3: *Segment of the third intron of the Aire gene of mongolian gerbils*

The following rodents belong to the suborder *Sciurognathi* and the *Cricetidae* family:

Chinese hamster (*Cricetulus griseus*):

Accession code: ENSCGRG00001016841 (Ensembl)

```
GTACTCAGGGAGAGGAGCTAGCCAGGGTTTCAGGGCCCTCTCCAACTGGCTCGCAGGAGCTTCTGTC
CCAATGATGCCTGCCCAGGGCCAGCCTGCTAGAGCCGTAGTCACCTCTGAGCACTGGTGGAGGCTCA
GTACCTCTCAGCTTCTTCCTGCTCTGAAGGCTGAGTTCCTGGCTGGTGAGGCAGGCAGGTGGTTTTG
ATTTCATTTCCGTTGAAGAAGGAAACGGCCTAGCCTGGCTTAAGAACACTCAGGGGGTCAGGAAATC
CCTTGGTCTGGCCCTACCAGGTGCTCTCCCCCGCCTCACTTTATACCCTGGACCCACCAACACCACC
CCAGCCCAGAAAGGGCATTCATTGCCTTCATGCTGCTTCCAACTCCCCTCAGCAG
```

Sequence 4: *Segment of the third intron of the Aire gene of chinese hamsters*

Golden hamster (*Mesocricetus auratus*):

Accession code: ENSMAUG00000019680 (Ensembl)

```
GTACTCAGGGAGAGGAGCTAGCCAGGATGTCGGGGCCCTCTCCAACTGGCTCTCGGGAGCTTCTGTC
CCAGGGATGCCTGCCCAGGGCCAGCCTGATAGAGCCGCAGTCACCTCTGAGCCCTGGTGGAGGCTCA
CCGCCTCGCAGCTTCTTCCTGTTCTGAAGGCTGAGTTCCTCCTGGCTGGTGAGGCAGGCAGGTGGTT
TTGATTTCGTTTCTGTTGAAGAAGGAAATGGCCTAACACGGCTTAAGAACACTCAGGGGGTCAGGAA
ACCCTTGATCTGCCCCTACCAGGTGCTCTCCCCCACCCAGGACCGCCCCGGCCCAGAAAGGGCTTTA
CCTTCAGGCTGCCTCCACCTCTCCTTGGCAG
```

Sequence 5: *Segment of the third intron of the Aire gene of golden hamster*

The following rodents belong to the suborder *Hystricomorpha* and the *Bathyergidae* family:

Naked mole rat (*Heterocephalus glaber*):

Gene ID: 101719207 (NCBI)

```
GTACCAGCAGGAGGAGCCCGCCAGGTCACAGCCCCCCAACACCCCTCGATGCTGAGGCCCACGGGCT
GGGCCAGGCTCCTGCCTCTCCCGCCCAGCTGCCCAGCGCCCAGGCGGCCCAGGAGCCCAGGTGGCCG
GGGTTCCAGATCCCAGCGGCAAAGGGATCCACTGGGACCCCCAGGGAGCCCAGAGCCCCAGCACCCA
CCGCACCAGTGCTCTCCAGCCAGATTACCCCAGCACCCCTTTGCAG
```

Sequence 6: *Segment of the third intron of the Aire gene of naked mole rats*

Damara mole rat (*Fukomys damarensis*):

Accession code: ENSFDAG00000015738 (Ensembl)

```
GTACCAGAGGAGAAGCCTGCCAGGGTCATGCCCCCACCCTCCCTGTGGTGCGTAGGCCCCCAGCTGG
GCCCACTCCTGCCTCTTCTGCCCAAGGCCAGGATCCCAGGGCTTGGGGTTTCCACAGGGGGCCCCCA
GGGATGCCGGGAGACCCAGCACCTGCCCTACCCAGTGCTCCCCAGCCAGCTGACCCCAGTACCCCTT
TGCAG
```

Sequence 7: *Segment of the third intron of the Aire gene of damara mole rats*

Guinea pigs belong to the suborder *Hystricomorpha* and the *Caviidae* family:

Guinea pig (*Cavia porcellus*):

The missing part of the third intron of the *Aire* gene was successfully sequenced in this project.

```
AAGACTCCTGTACTGCCACCCAGAAACCCCACCAAGAGGAAGGCTGTGGAGGAGCCCCGAGCTACCC
CACCAGCAGCCCTGACCTCGAAGGGCACCCCGAGTCCAGGTACCAGCAGGAGGAGCCTGCCTGGTCA
CAGACCTCCTACTCTTTCTATGGGGCCAGCACACTCCCCACAGGTACCTAACCCACTGCCAGGCTCC
TAGGTGCCAGGCAGTGCCCTGGGTGGGGCTTCCTGTTCCTGACAGCAGAGGAGTCCACATGGAGCCC
GGGGACCCTAGCACCCACCCCCTGGGTGCTGCCCAGCCAGCTGCCCCCTGTATCCCCTGCAGGATCC
CAGTGAAGTCAAGCCCCCCAAGAAGCCAGAGGGTAATTCGGAGTTGCAGC
```

Sequence 8: *Segment of the third intron of the Aire gene of guinea pigs*

The determination of gene similarity across diverse rodent species and guinea pigs was accomplished by using the multiple sequence alignment program Clustal Omega accessible

through https://www.ebi.ac.uk/Tools/msa/clustalo/. This is displayed in chapter 4.3, in which *Sequence 1- Sequence 8* are compared with each other by multiple alignment.

In Appendix A, the whole third intron of the *Aire* gene of guinea pigs is compared with the other rodents.

Chapter 4 Results

4.1 Determination of sequence of Aire intron 3

In the first part of this project, the missing sequence of the third intron of the *Aire* gene of guinea pigs, highlighted in red in *Figure 4*, was determined. At the time of the project, the red highlighted part was, according to the Ensembl genome database, not deciphered at the nucleotide level. In mice, the *Dnmt3l* oocyte promoter is located in the third intron of the *Aire* gene on the complementary DNA strand. Obtaining the full sequence of this region would allow us to assess whether the mouse oocyte *Dnmt3l* promoter is conserved in guinea pigs and therefore could drive oocyte *Dnmt3l* expression.

>ENSCPOG00000019868:ENSCPOT00000026960.2 intron 3:protein_coding
GGTACCAGCAGGAGGAGCCTGCCTGGTCACAGACCTCCTACTCTTTCTATGGGGCCAGCA
CACTCCCCACAGGTACCTAACCCACTGCCAGGCTCCTAGGTGCCAGGCAGTGCCCTGGGT
GGGGCTTCCTGTTCCTGACAGCAGAGGAGTCCACATGGAGCCCGGGGACCCTAGCACCCA
CCCCCTGGGTGCTGCCCAGCCAGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNGTC

Figure 4: *Third Intron of Aire gene in guinea pigs. The red-colored cycle highlights the region without sequence determination.*

The DNA extracted from the liver tissue of two guinea pig specimens was amplified by designing a customized primer pair inclosing the missing part of the sequence. For this purpose, the program Primer3web v4.1.0 (https://primer3.ut.ee/) was used. *Sequence 9* displays the forward (green) and reverse (blue) primer amplifying the third intron of the *Aire* gene.

ATGTGGACCTCAGCCAGCCCCGGAAGGGAAGGAAGCCCCCCAGCGGTCCCAAGACTCCTGTACTGCC
ACCCAGAAACCCCACCAAGAGGAAGGCTGTGGAGGAGCCCCGAGCTACCCCACCAGCAGCCCTGACC
TCGAAGGGCACCCCGAGTCCAGGTACCAGCAGGAGGAGCCTGCCTGGTCACAGACCTCCTACTCTTT
CTATGGGGCCAGCACACTCCCCACAGGTACCTAACCCACTGCCAGGCTCCTAGGTGCCAGGCAGTGC
CCTGGGTGGGGCTTCCTGTTCCTGACAGCAGAGGAGTCCACATGGAGCCCGGGGACCCTAGCACCCA
CCCCCTGGGTGCTGCCCAGCCAGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTCAAGCCCCCC
AAGAAGCCAGAGGGTAATTCGGAGTTGCAGCGCCTTCCCCTGGGGAACG

Sequence 9: *The colored areas of the third intron of the Aire gene present the designed primer pair. The forward primer is depicted in green and the reverse primer is colored in blue.*

The PCR procedure was done using the Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB). The experiment was conducted multiple times with a total sample volume of 20 µL. The PCR program was optimized and improved while the experiments were performed. For the gel electrophoresis, 5 µL sample + 1 µL 6X purple gel loading dye were loaded into the gel well. The gel was analyzed by a Gel Imager 200 from Azure Biosystems. Due to the final purification and Sanger sequencing procedures, the total sample volume was increased from 20 µL to 50 µL. To test if the DNA concentration is sufficient for the Sanger sequencing procedure, 1 µL of the sample was measured by the NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer. The expected band size of the PCR product was 517 bp, while the product on the gel appeared a little shorter, approximately 350-450 bp. *Figure 5* depicts the result of the electrophoresis.

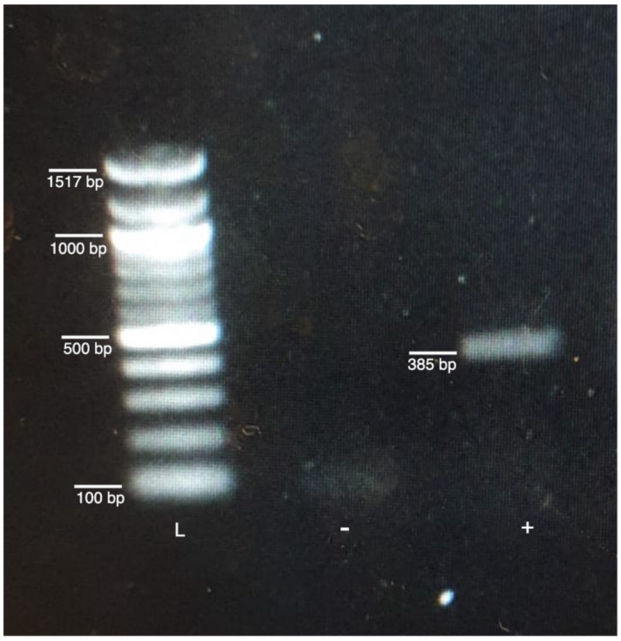

Figure 5: *PCR result of third intron in the Aire gene of guinea pigs*

In *Figure 5*, the New England Biolabs Quick-Load 100 bp DNA ladder is marked with a **L**. The negative control is presented with **-** and the PCR product is presented with a +. The sequencing process was successful. The sequenced segment of the third intron of the *Aire* gene, beginning at the forward primer (green) and ending at the reverse primer (blue), is exactly 385 pb. *Sequence 10* presents the fully sequenced part of the third intron of the *Aire* gene of guinea pigs.

AAGACTCCTGTACTGCCACCCAGAAACCCCACCAAGAGGAAGGCTGTGGAGGAGCCCCGAGCTACCC
CACCAGCAGCCCTGACCTCGAAGGGCACCCCGAGTCCAGGTACCAGCAGGAGGAGCCTGCCTGGTCA
CAGACCTCCTACTCTTTCTATGGGGCCAGCACACTCCCCACAGGTACCTAACCCACTGCCAGGCTCC
TAGGTGCCAGGCAGTGCCCTGGGTGGGGCTTCCTGTTCCTGACAGCAGAGGAGTCCACATGGAGCCC
GGGGACCCTAGCACCCACCCCCTGGGTGCTGCCCAGCCAGCTGCCCCCTGTATCCCCTGCAGGATCC
CAGTGAAGTCAAGCCCCCCAAGAAGCCAGAGGGTAATTCGGAGTTGCAGC

Sequence 10: *Fully sequence segment of the third intron of the Aire gene of guinea pigs. The sequence marked in green is the designed forward primer and the part colored in blue is the reverse primer.*

The relevant segments of the third intron of the *Aire* gene of guinea pigs are compared with other rodents in chapter 4.3.

4.2 Determination of the structure of guinea pig oocyte *Dnmt3l* transcript variant

RNA-sequencing data generated previously in our laboratory demonstrated low expression of *Dnmt3l* in guinea pig oocytes. We wanted to confirm by an alternative method, quantitative real-time PCR, whether *Dnmt3l* is truly expressed, and whether its oocyte promoter is annotated and matches the mouse genome. In mice, the *Dnmt3l* transcription starts from an oocyte-specific promoter localized in the third intron in the *Aire* gene on the opposite DNA strand. In contrast, the only annotated promoter in guinea pigs is located closely upstream of *Aire* first exon on the complementary DNA strand. The *Dnmt3l* gene of guinea pigs visualized in *Figure 6*, is obtained from the Ensembl genome browser (https://www.ensembl.org/index.html).

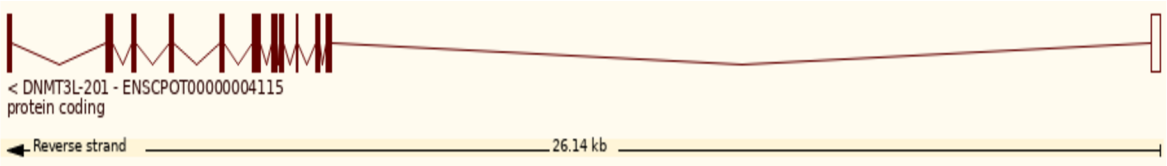

Figure 6: *Dnmt3l gene of guinea pigs*

As visible in *Figure 6*, the *Dnmt3l* gene in guinea pigs has 12 exons and 11 introns. To confirm other possible oocyte-specific novel *Dnmt3l* promoters we designed primers amplifying across the whole *Dnmt3l* gene. *Sequence 11-Sequence 16* display the exon sequences and the designed primer pairs aligning at the exons. The introns were excluded for better visualization of the customized primer pairs. The forward primers are depicted in green, and the reverse primers are coloured in blue.

Exon 1-2:

ATGGCGGTGTCACCCGGTCCCCGCGCGGCACAGCTGCGCCCCCTGCCGGCCGCGGCGTGGCGTGCA
CACCTGGCGGGCACCTGCGCGCCCACTGTTATACTGCCCGGTCCAGGAAAGAGCAAAGGCCGCTGG
CGCGGGAGCCAATCAGGGCCGGAGCTCCCACGGACCGAGTGGAGCCACCGGTGGCTCTGGAAGGCC
TCCTCTCCGGACTCCATGGCCTGCTCCATCCTCAAGACCCGGAGCCCAAAGGCCCTGGACAGGCGG
GACCCTGGCCCTGCCGGCTGCCCAG

Sequence 11: *Primer pair aligning from exon 1 to 2*

Exon 1-4:

ATGGCGGTGTCACCCGGTCCCCGCGCGGCACAGCTGCGCCCCCTGCCGGCCGCGGCGTGGCGTGCA
CACCTGGCGGGCACCTGCGCGCCCACTGTTATACTGCCCGGTCCAGGAAAGAGCAAAGGCCGCTGG
CGCGGGAGCCAATCAGGGCCGGAGCTCCCACGGACCGAGTGGAGCCACCGGTGGCTCTGGAAGGCC
TCCTCTCCGGACTCCATGGCCTGCTCCATCCTCAAGACCCGGAGCCCAAAGGCCCTGGACAGGCGG
GACCCTGGCCCTGCCGGCTGCCCAGCCCTGGCGCTGGAGGCTGAGCGCAGTGTAGACGTGCTCCTG
CTGGACTCCATTGAGCCGCCACAGGCCCCTGCACGCAGGCGGGGCAGAGAGCGCATCACTTACGAA
GTCTCAGTGAACCAGCGGAACATAGAAG

Sequence 12: *Primer pair aligning from exon 1 to 4*

Exon 4-6:

AGCGCATCACTTACGAAGTCTCAGTGAACCAGCGGAACATAGAAGCCATCTGCCTCTGCTGCGGGA
GTTTCCAGGTGCACATACGGCATCCCCTGTTCGAAGGAGGAATGTGTGCCCCCTGCAAGGACAAGT
TCCTGGGCGCCCTGTTCCTTTATGATGAGGATGGGTACCAGTCCTGCTGCTCCATCTGTGCCTCGG
GGAGTGTTCTGCTCATCTGCGAGAACCCCCAGTGCACAAG

Sequence 13: *Primer pair aligning from exon 4 to 6*

Exon 6-8:

GACAAGTTCCTGGGCGCCCTGTTCCTTTATGATGAGGATGGGTACCAGTCCTGCTGCTCCATCTGT
GCCTCGGGGAGTGTTCTGCTCATCTGCGAGAACCCCCAGTGCACAAGGTGCTACTGCTTCGAGTGC
CTGGACGTGTTGGTGGGCCCCGGGACCTCAGACAAGGTGCACAGCATGTCCAGCTGGGTGTGCTTC
CTGTGCCTGCCCTTCCCATGCAGCGGGCTGCTGCGGCGGCGGCCCAAGTGGCGCGGCCAGCTGAAG
GCCTTCCACGACCTCGAGGCGGGAAGTCCCCTGGAGATGTACAAAACAGTGCCTGTGTGGAAGAGA
CAGCCGATGCGGGTGCTATCCCTGTTTGGGGACATTAAGCAAG

Sequence 14: *Primer pair aligning from exon 6 to 8*

Exon 8-10:

GGAAGTCCCCTGGAGATGTACAAAACAGTGCCTGTGTGGAAGAGACAGCCGATGCGGGTGCTATCC
CTGTTTGGGGACATTAAGCAAGAGCTGATGAGTTTGGGCTTTTTGGAAGCTGGGTCCCAGGCAGGA
AGGCTGCGGCACTTGGAGGATGTCACAGATGTCGTGCGGAGAGACGTGGAGGAGTGGGGCCCATTT
GACCTGGTGTACGGCTCAACACCCACTCTGGGCCACGCCTGTGACCGTTCTCCCG

Sequence 15: *Primer pair aligning from exon 8 to 10*

Exon 11-12:

CTTGGTTCCTGTTCCAGTTCCACCGGCTGCTGCAGTACTCGAGGCCCCCGCCTGGCAGCCCTCAGC
CTTTCTTCTGGATGTTTGTGGACCACCTGCTGCTGACCAGTGATGACCAGGCCATTGCCACCCGCT
TCCTTGAGATAGAGCCGGCCACCCTCCAGGATGTCCATGGCCGGGTCCTGCAGGGCGCTGCCAGAG
TGTGGAGCAACATCCCCGCCGTGAGGAGGTACTGGGGACCAAGGGGACACTGA

Sequence 16: *Primer pair aligning from exon 11 to 12*

The RNA was isolated from guinea pig oocytes, using the PicoPure RNA isolation Kit (Life Technologies). In the next step, the extracted RNA was reverse-transcribed by using the Superscript III Reverse Transcriptase (ThermoFisher Scientific). The corresponding cDNA was diluted in a ratio of 1:100 with water prior to the PCR procedure. The PCR program was modified and improved during the experiments. The final modified PCR program is summarized in *Table 9* (chapter 3.2.3). The PCR product was then analyzed by the gel electrophoresis. *Figure 7* displays the results of an electrophoresis experiment done with the primer pairs targeting exons 1+2, 1+4, 4+6, 6+8, 8+10 and 11+12.

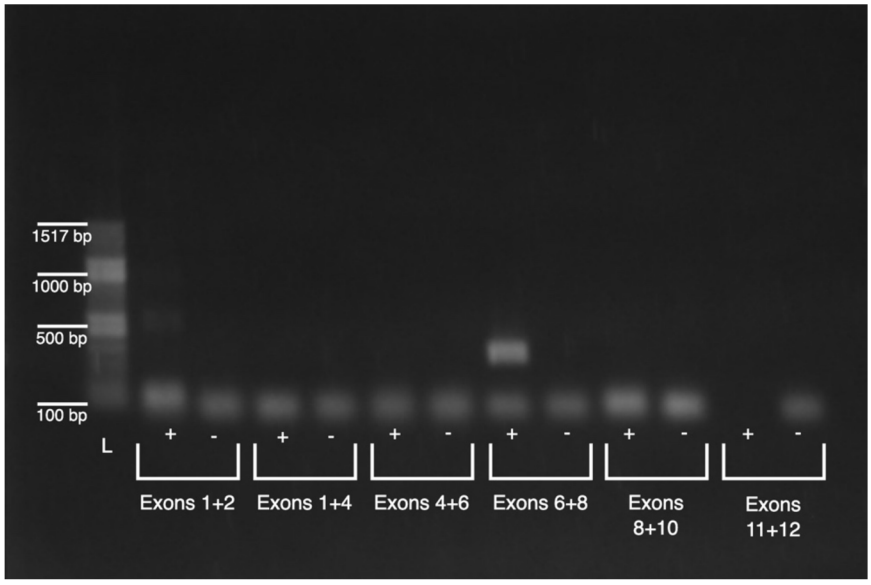

Figure 7: *Experiment using all designed primers pairs*

In *Figure 7*, the New England Biolabs Quick-Load 100 bp DNA ladder is marked with a **L**. The negative control is presented with **-** and the sample with the PCR product is marked with a **+**. *Figure 7* is labeled with the primer pairs that were used in the experiment. *Figure 7* also depicts a strong visible PCR product, corresponding to the exons 6+8 (325 bp). As conclusion,

there is a novel promoter located in the 6th exon in guinea pig oocytes leading to a weak expression of a *Dnmt3l* isoform. As visible in *Figure 7*, there are also very slight PCR product bands observable at exon 1+2. To confirm this annotated promoter located closely upstream of *Aire* first exon on the complementary DNA strand, more experiments were conducted. All lower product bands represent the remaining primer pairs. However, at exon 11+12 (+) there is no visible primer pair band at the bottom of the graph, which leads to the conclusion that the primer pairs were forgotten to be added to this specific sample.

The experiments were conducted multiple times, using the designed primer pairs listed in *Sequence 11-Sequence 16*. *Figure 8* demonstrates visible PCR products at exon 1+2 (133 pb) and exon 8+10 (127 pb). It is unclear why these two products were not observable in the experiments conducted before. However, *Figure 8* demonstrates two short alternative transcripts of the *Dnmt3l* gene. This results in two short isoforms, which could be transcribed in the guinea pig oocytes. This confirms the weak expression of one annotated promoter positioned closely upstream of *Aire* first exon on the opposite DNA strand. The oocyte-specific full DNMT3L protein with the promoter starting in the third intron of the *Aire* gene of the complementary DNA strand is not conserved in guinea pig oocytes. In guinea pig oocytes, no observable product was verified to be transcribed beginning at this annotated promoter visible in the *Eumuroida* lineage.

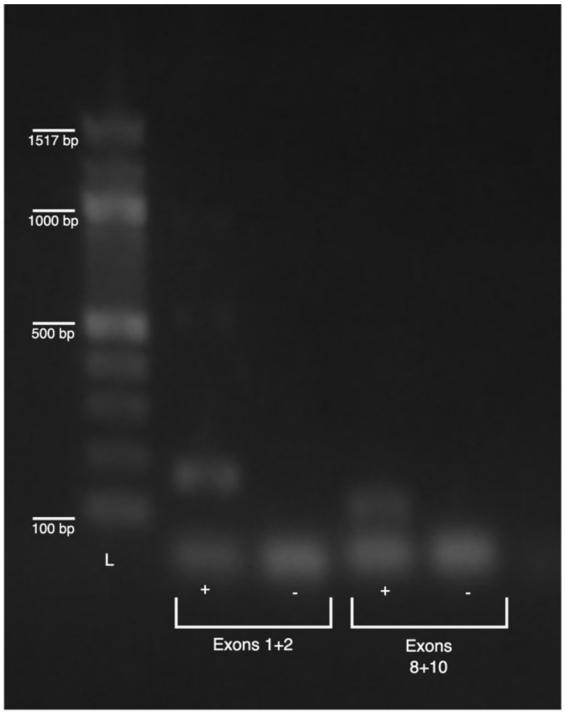

Figure 8: *Visible PCR products for exon 1+2 and 8+10*

In *Figure 8*, the New England Biolabs Quick-Load 100 bp DNA ladder is marked with a **L**. The negative control is displayed with a – and the sample is presented with a +. The product size of exon 1+2 is 133 pb. The product size of exon 8+10 is 127 bp.

4.3 Comparison of guinea pigs to other mammals

The *Dnmt3l* gene was shown to be transcribed in oocytes of the *Eumuroida* lineage, comprising families *Muridae* (mice, rats, gerbils) and *Cricetidae* (hamsters). As mentioned before, the situation in guinea pig oocytes was not clear, as RNA-sequencing data showed a weak expression of the *Dnmt3l* gene (despite not being classified as an *Eumuroida* rodent). Therefore, guinea pig oocytes were examined in this experiment. As noted before, the transcriptional start site (transcriptional start site) in mice oocytes is located at the third intron of the *Aire* gene, encoded on the opposite DNA strand. The sequencing procedure of the third intron of the *Aire* gene of guinea pigs (chapter 4.1) presented that the oocyte promoter sequence is not conserved. This led to the question if an evolutionarily novel *Dnmt3l* promoter, unique for guinea pigs, could be responsible for the low level of expression instead of the annotated guinea pig promoter. The experiments conducted revealed that the *Dnmt3l* isoforms expressed in guinea pigs are stretching from exon 1+2, 6+8 and 8+10. The short isoforms are likely neither functional nor helping to establish the *de novo* DNA methylation. A longer isoform which is transcribed beginning at exon 6 and stretching to exon 10 was not confirmed.

Both the naked mole-rat and guinea pigs belong to the same suborder, *Hystricomorpha*. Therefore, the base pairs around the annotated guinea pig transcriptional start site (± 250 bp) were compared to the ones of naked mole-rats. As visible in *Figure 9*, the sequence is largely conserved between the two species of the same suborder. The guinea pig *Dnmt3l* transcriptional start site is colored in yellow. The green highlighted bases demonstrate the not conserved part of this sequence comparison.

Figure 9: *Comparison of guinea pigs and naked mole rats. The guinea pig Dnmt3l transcriptional start site is highlighted in yellow. A region colored in green presents the low conserved sequence part.*

In *Figure 10*, a selected part of the third intron of the *Aire* gene, containing the oocyte transcription factor binding site and the transitional start site of *Eumuroida* rodents is presented. The multiple alignment was performed between the following rodents: mouse (*Mus musculus*), rat (*Rattus rattus*), mongolian gerbil (*Meriones unguiculatus*), chinese hamster (*Cricetulus griseus*), golden hamster (*Mesocricetus auratus*), naked mole rat (*Heterocephalus glaber*), damara mole rat (*Fukomys damarensis*) and guinea pig (*Cavia porcellus*).

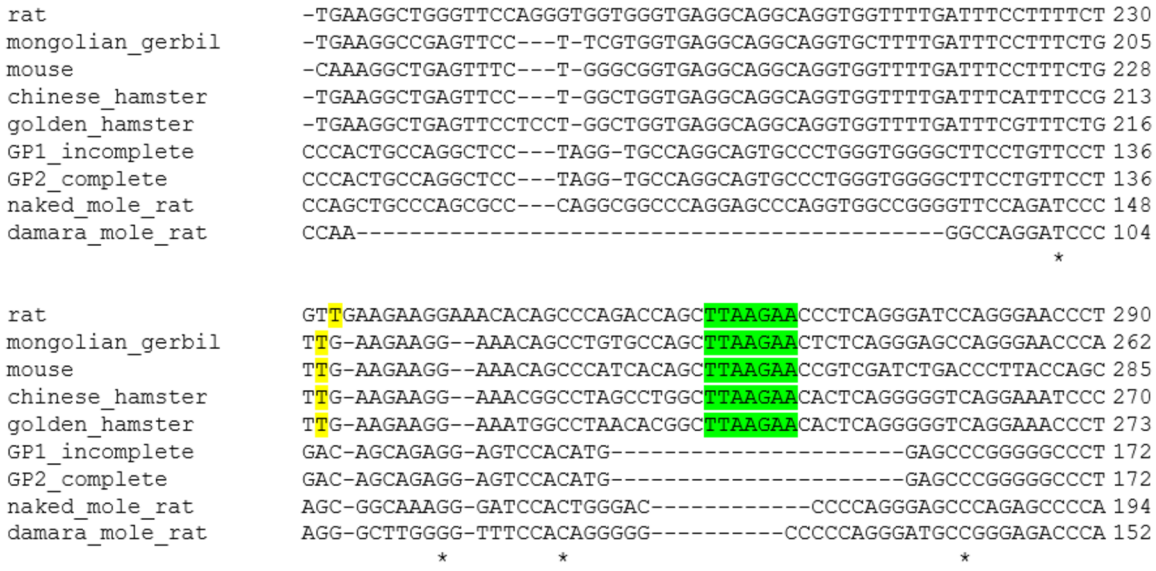

Figure 10: *Third intron of the Aire gene of guinea pigs compared to other rodents. The yellow highlighted base is the transcriptional start site. The sequence displayed in green is the oocyte transcription factor binding site.*

In *Figure 10*, the yellow highlighted part marks the beginning of the transcriptional start site. The sequence highlighted in green displays the oocyte transcription factor binding site. *Figure 10* displays, that the oocyte transcription factor binding site and the beginning of the transcriptional start site are conserved within the *Eumuroida* lineage. However, guinea pigs, naked mole rats, and damara mole rats, which all belong to the *Hystricomorpha* suborder, lack a conserved transcription factor binding and a beginning of the transcriptional start site. This results in guinea pigs, naked mole rats, and damara mole rats not transcribing the full DNMT3L protein in their oocytes.

In Appendix A, the whole third intron of the *Aire* gene of guinea pigs is compared with the other rodents.

Chapter 5 Discussion

This project demonstrates that not only naked mole rats and damara mole rats, but also guinea pigs, do not rely on the transcription of the complete DNMT3L protein in their oocytes. This raises the question of whether DNMT3L is truly necessary for *de novo* DNA methylation in mammalian oocytes. In contrast to *Dnmt3a* and *Dnmt3b*, *Dnmt3l* presents a unique evolutionary development in mammals. In 2013, a research team comprising M. Guenatri, R. Duffié, J. Iranzo, et al., examined whether DNMT3L is essential for the *de novo* methylation in mice. In the germ cells of mice, they confirmed the requirement of the DNMT3L protein for DNA methylation. However, they argue that it might not be essential in embryo development. They suspect that DNMT3L only accelerates the *de novo* methylation in the mouse embryos (Guenatri et al., 2013). In 2019, another research team comprising N. Veland, Y. Lu, S. Hardikar, et al., supports the perspective stated above. Their own results demonstrate that DNMT3L is required for the *de novo* methylation in the oocytes of mice. Zygotic DNMT3L, however, is not necessary for embryonic development (Veland et al., 2019c).

The expression of the *Dnmt3l* in oocytes is regulated by a conserved promoter found in mice, rats, gerbils, and hamsters. In contrast to that, rodents not belonging to the *Eumuroida* lineage generally show an absence of *Dnmt3l* expression in their oocytes. This was demonstrated by recent experiments previously conducted in our laboratory. Naked mole rats or damara mole rats, which belong to the suborder *Hystricomorpha*, were confirmed to lack *Dnmt3l* expression in their oocytes. This indicates that the unique oocyte promoter of *Dnmt3l* may have originated within the *Eumuroida* rodents. Consequently, *Dnmt3l* might only be expressed as a functional cofactor in oocytes of species belonging to the mouse-related rodent clade. However, to definitively determine the function of DNMT3L in mammals beyond the *Eumuroida* lineage, it is necessary to generate female germ line-specific DNMT3L-knockout breeds. Without such individuals, the function of DNMT3L in these mammals cannot be conclusively confirmed or rejected.

The transcription of *Dnmt3l* in guinea pig oocytes (despite them not belonging to the *Eumuroida* lineage) was uncertain, as RNA-sequencing data indicated low levels of expression. The short isoforms of *Dnmt3l* which might be expressed in low levels comprise the exons 1+2, 6+8 and 8+10. However, these alternative transcriptions are unlikely to be functional in oocytes.

Moreover, given the extremely low expression levels, the *Dnmt3l* isoforms are not expected to either form a tetramer complex with DNMT3A or help establish the *de novo* DNA methylation in oocytes.

Mice serve as a general model for studying biological mechanisms; however, obtained knowledge cannot represent all mammals, nor does it apply to every rodent species. Therefore, examining non-model species helps demonstrate the diversity among mammals and provides a better understanding of universal biological processes. For example, *Dnmt3l* might only be expressed in *Eumuroida* rodent lineage, which would lead to the conclusion that there is another biological mechanism for DNA methylation in oocytes. This variety in genetics and biological processes reveals evolutionary relations and individual traits of these species.

There are limitations to this project. First, it might be possible that *Dnmt3l* is expressed prior to the oocyte growth in the primordial germ cells. The DNMT3L protein could remain folded until oocyte maturation and support establishing the *de novo* DNA methylation. Also, the *Dnmt3l* may be transcribed earlier during oogenesis. In mature oocytes, that could make the protein undetectable at protein or mRNA levels. This way, DNMT3L would again participate in the *de novo* DNA methylation while being not detectable in the experiment.

The results are mainly supported by the determination sequence motif of the *Dnmt3l* gene in mice. Even though it might be possible that *Dnmt3l* is expressed by other sequence motifs identified by different oocyte TFs, the RNA-seq data of oocytes available online does not indicate this. The data does not support the presence of a universal DNMT3L expression pattern across mammals. The TBPL2/TFIIA complex is generally assumed to regulate the *Dnmt3l* expression by binding to the identified motif. Moreover, chromatin immunoprecipitation could most likely confirm the TBPL2 presence. It is believed that the transcription procedure is mostly regulated by the same transcription complexes, as stated in the paper "TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage". In oocyte maturation, the TBP is replaced by TBPL2 which plays a crucial role in DNA methylation. Knockout experiments correlated to the absence of H3K4me3 methylation signals at the primary as well as secondary follicle stage of the oocytes (Yu et al., 2020). Therefore, results were leaned on the sequence motif as a factor of a functional DNA methylation process.

Chapter 6 Conclusion

This project investigated the *Dntm3l* expression in guinea pig oocytes. The analyzed *Dnmt3l* sequences of mammalian oocytes presented a specific upstream promoter which is found in mice, rats, gerbils, and hamsters. In contrast to that, rodents outside the *Eumuroida* lineage generally lack DNMT3L in their oocytes. For example, naked mole rats and damara mole rats, which belong to the suborder *Hystricomorpha*, lack the expression of *Dnmt3l* in their oocytes. At the beginning of the project, the transcription of *Dnmt3l* in guinea pig oocytes was uncertain. Although guinea pigs also belong to the *Hystricomorpha* suborder and not to the *Eumuroida* rodents, RNA-sequencing data indicated low levels of *Dnmt3l* expression. Experiments were conducted to examine the guinea pig oocytes further. The *Dnmt3l* promoter in mice oocytes is located at the third intron of the *Aire* gene, encoded on the opposite DNA strand. At the time of this project, the third intron of the *Aire* gene in guinea pigs had not yet been entirely sequenced. However, the missing part of the third intron of the *Aire* gene was successfully sequenced using the Sanger sequencing procedure. The analyses presented that the annotated *Dnmt3l* promoter of *Eumuroida* rodents is not conserved in guinea pig oocytes. Also, guinea pigs lack a conserved factor binding site and beginning of the transcriptional start site. Further experiments resulted in identifying isoforms of *Dnmt3l*, expressed from novel promoters in guinea pig oocytes, stretching from exon 1+2, 6+8 and 8+10. These alternative transcriptions are unlikely to be functional in the oocytes. Given the extremely low expression levels, these *Dnmt3l* isoforms are not expected to form a tetramer complex with DNMT3A or aid in establishing *de novo* DNA methylation in oocytes. Examining non-model species reveals the diversity among rodents (as well as mammals) and enhances the understanding of universal biological processes. If *Dnmt3l* is only expressed in the *Eumuroida* rodent lineage, it indicates that there is another mechanism for DNA methylation that exists in oocytes of other species. There are limitations to this project, for example, *Dnmt3l* might be expressed earlier in the primordial germ cells and remain as folded DNMT3L until oocytes growth to support the *de novo* DNA methylation. To definitively establish the function of DNMT3L in mammals outside the *Eumuroida* lineage, it would be necessary to produce female germ line-specific DNMT3L-knockout breeds. Without these, the function of DNMT3L in these rodents cannot be decisively confirmed or rejected.

Appendix A

```
rat                 ---GTACTCAAGAGGAACTACCCAGGGTTGCTGGGCCCTCCCCGACCGGCTCTCAGGAGC      57
mongolian_gerbil    ------------GTACTCAGGGAGAGGAGCTGGGTCCTCCCCAAATGACTCTCAGGAGC      47
mouse               -GTACACTCAAGAGGAGCTAGCCAGGGTTGCTGGGCCCTCCCCAACCGGCTCTTAGGAGC      59
chinese_hamster     GTACTCAGGGAGAGGAGCTAGCCAGGGTTTCAGGGCCCTCTCCAACTGGCTCGCAGGAGC      60
golden_hamster      GTACTCAGGGAGAGGAGCTAGCCAGGATGTCGGGGCCCTCTCCAACTGGCTCTCGGGAGC      60
GP1_incomplete      GGTACC------------------------------------------------------      6
GP2_complete        GGTACC------------------------------------------------------      6
naked_mole_rat      -GTACC------------------------------------------------------      5
damara_mole_rat     -GTA-C------------------------------------------------------      4

rat                 TTCTGTCTTACTGACACCACCCCAGGGCCAGCCTGCCAGAGCTGGAGTCACCTTTGAGCG      117
mongolian_gerbil    T-----------TACTGATACCACCCCAAGGAGGGCCAGAGCCGCAGAGTCACGTCTGAG      96
mouse               TTCTGTCTTACTGACACCACCCCAGGGCCAGCCTGCCAGGGTCACAGAGTCACCTCTGAG      119
chinese_hamster     TTCTGTCCCAATGATGCCTGCCCAGGGCCAGCCTGCTAGAGCCGTA--------------      106
golden_hamster      TTCTGTCCCAGGGATGCCTGCCCAGGGCCAGCCTGATAGAGCCGCA--------------      106
GP1_incomplete      --------------------AGCAGGAGGAGCCTGCCTGGTCACAG--------------      32
GP2_complete        --------------------AGCAGGAGGAGCCTGCCTGGTCACAG--------------      32
naked_mole_rat      --------------------AGCAGGAGGAGCCCGCCAGGTCACAG--------------      31
damara_mole_rat     --------------------CAGAGGAGAAGCCTGCCAGGGTCATG--------------      30
                                        *   *   *

rat                 CTCAGACCTGAGCATTGGAGGAAGAGGGCACAGCCTCTC---AGCGTCTTACTGTCC---      171
mongolian_gerbil    TGCTCAGACCTGAGCATCGGTGGAAGCTCACGGCCTCTC---AGCTTCTCACTGGTC---      150
mouse               CCCTCAGACCTGAGCATTGGAGGAGGCCCACAGCCTCTC---AGCGTCTTACTGTCC---      173
chinese_hamster     --GTCACCTCTGAGCACTGGTGGAGGCTCAGTACCTCTC---AGCTTCTTCCTGCTC---      158
golden_hamster      --GTCACCTCTGAGCCCTGGTGGAGGCTCACCGCCTCGC---AGCTTCTTCCTGTTC---      158
GP1_incomplete      ACCTCCTACTCTTTCTATGGGGCCAGCACACTCCC-------CACAGGT-----ACCTAA      80
GP2_complete        ACCTCCTACTCTTTCTATGGGGCCAGCACACTCCC-------CACAGGT-----ACCTAA      80
naked_mole_rat      CCCCCCAACACCCCTCGATGCTGAGGCCCACGGGCTGGGCCAGGCTCCTGCCTCTCCCGC      91
damara_mole_rat     CCCCCACCCTCCCTGTGGTGCGTAGGCCCCCAGCTGG--GCCCACTCCTGCCTCTTCTGC      88
                              *     *  *              *   *       *

rat                 -TGAAGGCTGGGTTCCAGGGTGGTGGGTGAGGCAGGCAGGTGGTTTTGATTTCCTTTTCT      230
mongolian_gerbil    -TGAAGGCCGAGTTCC---T-TCGTGGTGAGGCAGGCAGGTGCTTTTGATTTCCTTTCTG      205
mouse               -CAAAGGCTGAGTTTC---T-GGGCGGTGAGGCAGGCAGGTGGTTTTGATTTCCTTTCTG      228
chinese_hamster     -TGAAGGCTGAGTTCC---T-GGCTGGTGAGGCAGGCAGGTGGTTTTGATTTCATTTCCG      213
golden_hamster      -TGAAGGCTGAGTTCCTCCT-GGCTGGTGAGGCAGGCAGGTGGTTTTGATTTCGTTTCTG      216
GP1_incomplete      CCCACTGCCAGGCTCC---TAGG-TGCCAGGCAGTGCCCTGGGTGGGGCTTCCTGTTCCT      136
GP2_complete        CCCACTGCCAGGCTCC---TAGG-TGCCAGGCAGTGCCCTGGGTGGGGCTTCCTGTTCCT      136
naked_mole_rat      CCAGCTGCCCAGCGCC---CAGGCGGCCCAGGAGCCCAGGTGGCCGGGGTTCCAGATCCC      148
damara_mole_rat     CCAA--------------------------------------------GGCCAGGATCCC      104
                                                                       *
```

```
rat                 GTTGAAGAAGGAAACACAGCCCAGACCAGCTTAAGAACCCTCAGGGATCCAGGGAACCCT      290
mongolian_gerbil    TTG-AAGAAGG--AAACAGCCTGTGCCAGCTTAAGAACTCTCAGGGAGCCAGGGAACCCA      262
mouse               TTG-AAGAAGG--AAACAGCCCATCACAGCTTAAGAACCGTCGATCTGACCCTTACCAGC       285
chinese_hamster     TTG-AAGAAGG--AAACGGCCTAGCCTGGCTTAAGAACACTCAGGGGGTCAGGAAATCCC      270
golden_hamster      TTG-AAGAAGG--AAATGGCCTAACACGGCTTAAGAACACTCAGGGGGTCAGGAAACCCT      273
GP1_incomplete      GAC-AGCAGAGG-AGTCCACATG----------------------GAGCCCGGGGGCCCT      172
GP2_complete        GAC-AGCAGAGG-AGTCCACATG----------------------GAGCCCGGGGGCCCT      172
naked_mole_rat      AGC-GGCAAAGG-GATCCACTGGGAC------------CCCCAGGGAGCCCAGAGCCCCA      194
damara_mole_rat     AGG-GCTTGGGG-TTTCCACAGGGGG----------CCCCCAGGGATGCCGGGAGACCCA      152
                              *       *                           *

rat                 TGATATGACCCTTACCAGGTGCTTTCTCCCCAACC------TTACTTGACATCCTGGACC      344
mongolian_gerbil    TGATCTG-ACCCTGCCAGGTGCTTTCTCCCACCCCCCCACCTCACTTTATACCCTGGGCC      321
mouse               TGCTCTCTCT----------------CCCA--------TCCTCACTTTCTACCCTGGATC      321
chinese_hamster     TTGGTCTGGCCCTACCAGGTGCTCTCCCCC--------GCCTCACTTTATACCCTGGACC      322
golden_hamster      T-GATCTGCCCCTACCAGGTGCTCTCCCCC--------ACCCAGGACCGCC---------      315
GP1_incomplete      A-G-CACCCACCCCCTGGGTGCTGCCCAGCCAGCT----GCCCCCTGTA-----------      215
GP2_complete        A-G-CACCCACCCCCTGGGTGCTGCCCAGCCAGCT----GCCCCCTGTATCCCCTGCA--      224
naked_mole_rat      G---CACCCACCGCACCAGTGCTCTCCAGCCAGAT----TACCCCAGCACCCCTTTGC--      245
damara_mole_rat     G-C-ACCTGCCCTACCCAGTGCTCCCCAGCCAGCT----GACCCCAGTACCCCTTTGC--      204

rat                 TGTCAACACCCCCTAACCCAACCTGTACCTC---------------------------CC      377
mongolian_gerbil    CACCAGCACCACCCCAGCCCAGAAAGGAGGC---------CTAGCCTGCCTCCACCTCCC      372
mouse               CGTCAACATGACCCCAGCCCAGAAAAGTGGG--------CCCAGGCTGCCTCTACCTCCC      373
chinese_hamster     CACCAACACCACCCCAGCCCAGAAAGGGCATTCATTGCCTTCATGCTGCTTCCAACTCCC      382
golden_hamster      -------------CCGGCCCAGAAAGGGCT----TTACCTTCAGGCTGCCTCCACCTCTC      358
GP1_incomplete      ------------------------------------------------------------      215
GP2_complete        -------------GGATCCCAGTGAAGTC------------------------------      240
naked_mole_rat      -------------AG--------------------------------------------      247
damara_mole_rat     -------------AG--------------------------------------------      206

rat                 CTTCACAG  385
mongolian_gerbil    TTTCCCAG  380
mouse               CTTCGCAG  381
chinese_hamster     CTCAGCAG  390
golden_hamster      CTTGGCAG  366
GP1_incomplete      --------  215
GP2_complete        --------  240
naked_mole_rat      --------  247
damara_mole_rat     --------  206
```

References

Alaskhar Alhamwe, B., Khalaila, R., Wolf, J., Bülow, V., Harb, H., Alhamdan, F., Hii, C. S., Prescott, S. L., Ferrante, A., Renz, H., Garn, H., & Potaczek, D. P. (2018). Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy, Asthma & Clinical Immunology 2018 14:1*, *14*(1), 1–16. https://doi.org/10.1186/S13223-018-0259-4

Altun, G., Loring, J. F., & Laurent, L. C. (2010). DNA methylation in embryonic stem cells. *Journal of Cellular Biochemistry*, *109*(1), 1. https://doi.org/10.1002/JCB.22374

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., & Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, *354*(6314), 909–912. https://doi.org/10.1126/SCIENCE.AAH5143/SUPPL_FILE/BARAU.SM.PDF

Cai, Y., Yu, X., Hu, S., & Yu, J. (2009). A Brief Review on the Mechanisms of miRNA Regulation. *Genomics, Proteomics & Bioinformatics*, *7*(4), 147–154. https://doi.org/10.1016/S1672-0229(08)60044-3

Chen, Z., Chen, Z., Chen, Z., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., & Zhang, Y. (2020). Role of Mammalian DNA Methyltransferases in Development. *Annual Review of Biochemistry*, *89*(Volume 89, 2020), 135–158. https://doi.org/10.1146/ANNUREV-BIOCHEM-103019-102815/CITE/REFWORKS

Chen, Z., & Zhang, Y. (2020). *Annual Review of Biochemistry Role of Mammalian DNA Methyltransferases in Development*. https://doi.org/10.1146/annurev-biochem-103019

Cox, P. G., & Hautier, L. (2015). Evolution of the rodents: advances in phylogeny, functional morphology, and development. In *Evolution of the rodents : advances in phylogeny, functional morphology, and development*. Cambridge University Press. 10.1017/CBO9781107360150

Demond, H., & Kelsey, G. (2020). The enigma of DNA methylation in the mammalian oocyte. In *F1000Research* (Vol. 9). F1000 Research Ltd. https://doi.org/10.12688/f1000research.21513.1

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *Journal of Biological Chemistry*, *285*(34), 26114–26120. https://doi.org/10.1074/jbc.M109.089433

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology 2019 20:10*, *20*(10), 590–607. https://doi.org/10.1038/s41580-019-0159-6

Guenatri, M., Duffié, R., Iranzo, J., Fauque, P., & Bourc'his, D. (2013). Plasticity in Dnmt3L-dependent and -independent modes of de novo methylation in the developing mouse embryo. *Development*, *140*(3), 562–572. https://doi.org/10.1242/DEV.089268

Hanna, C. W., Demond, H., & Kelsey, G. (2018). Epigenetic regulation in development: is the mouse a good model for the human? *Human Reproduction Update*, *24*(5), 556–576. https://doi.org/10.1093/HUMUPD/DMY021

Hanna, C. W., Huang, J., Belton, C., Reinhardt, S., Dahl, A., Andrews, S., Francis Stewart, A., Kranz, A., & Kelsey, G. (2022). Loss of histone methyltransferase SETD1B in oogenesis results in the redistribution of genomic histone 3 lysine 4 trimethylation. *Nucleic Acids Research*, *50*(4), 1993–2004. https://doi.org/10.1093/NAR/GKAC051

Inbar-Feigenberg, M., Choufani, S., Butcher, D. T., Roifman, M., & Weksberg, R. (2013). Basic concepts of epigenetics. *Fertility and Sterility*, *99*(3), 607–615. https://doi.org/10.1016/J.FERTNSTERT.2013.01.117

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature 2007 449:7159*, *449*(7159), 248–251. https://doi.org/10.1038/nature06146

Kaikkonen, M. U., Lam, M. T. Y., & Glass, C. K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research*, *90*(3), 430. https://doi.org/10.1093/CVR/CVR097

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, H., Li, E., & Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature 2004 429:6994*, *429*(6994), 900–903. https://doi.org/10.1038/nature02633

Kanherkar, R. R., Bhatia-Dey, N., & Csoka, A. B. (2014). Epigenetics across the human lifespan. *Frontiers in Cell and Developmental Biology*, *2*(SEP). https://doi.org/10.3389/FCELL.2014.00049

Kibe, K., Shirane, K., Ohishi, H., Uemura, S., Toh, H., & Sasaki, H. (2021). The DNMT3A PWWP domain is essential for the normal DNA methylation landscape in mouse somatic cells and oocytes. *PLOS Genetics*, *17*(5), e1009570. https://doi.org/10.1371/JOURNAL.PGEN.1009570

Maxeiner, S., Gebhardt, S., Schweizer, F., Venghaus, A. E., & Krasteva-Christ, G. (2021). Of mice and men – and guinea pigs? *Annals of Anatomy - Anatomischer Anzeiger*, *238*, 151765. https://doi.org/10.1016/J.AANAT.2021.151765

Moore, L. D., Le, T., & Fan, G. (2012). DNA Methylation and Its Basic Function. *Neuropsychopharmacology 2013 38:1*, *38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., & Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Reports*, *10*(11), 1235–1241. https://doi.org/10.1038/EMBOR.2009.218/SUPPL_FILE/EMBR2009218-SUP-0001.PDF

Petrussa, L., Van de Velde, H., & De Rycke, M. (2014). Dynamic regulation of DNA methyltransferases in human oocytes and preimplantation embryos after assisted reproductive technologies. *Molecular Human Reproduction*, *20*(9), 861–874. https://doi.org/10.1093/MOLEHR/GAU049

Prawitt, D., & Haaf, T. (2020). Basics and disturbances of genomic imprinting. *Medizinische Genetik*, *32*(4), 297–304. https://doi.org/10.1515/MEDGEN-2020-2042/ASSET/GRAPHIC/J_MEDGEN-2020-2042_FIG_002.JPG

Sendžikaitė, G., & Kelsey, G. (2019a). The role and mechanisms of DNA methylation in the oocyte. *Essays in Biochemistry*, *63*(6), 691–705. https://doi.org/10.1042/EBC20190043

Sendžikaitė, G., & Kelsey, G. (2019b). The role and mechanisms of DNA methylation in the oocyte. In *Essays in Biochemistry* (Vol. 63, Issue 6, pp. 691–705). Portland Press Ltd. https://doi.org/10.1042/EBC20190043

Shovlin, T. C., Bourc'his, D., La Salle, S., O'Doherty, A., Trasler, J. M., Bestor, T. H., & Walsh, C. P. (2007). Sex-specific promoters regulate Dnmt3L expression in mouse germ cells. *Human Reproduction*, *22*(2), 457–467. https://doi.org/10.1093/HUMREP/DEL379

Sigmon, J., & Larcom, L. L. (1996). The effect of ethidium bromide on mobility of DNA fragments in agarose gel electrophoresis. *Electrophoresis*, *17*(10), 1524–1527. https://doi.org/10.1002/ELPS.1150171003

Skinner, M. K. (2011). Role of epigenetics in developmental biology and transgenerational inheritance. *Birth Defects Research Part C: Embryo Today: Reviews*, *93*(1), 51–55. https://doi.org/10.1002/BDRC.20199

Smallwood, S. A., & Kelsey, G. (2012). De novo DNA methylation: a germ cell perspective. *Trends in Genetics*, *28*(1), 33–42. https://doi.org/10.1016/J.TIG.2011.09.004

Stewart, K. R., Veselovska, L., & Kelsey, G. (2016). Establishment and functions of DNA methylation in the germline. In *Epigenomics* (Vol. 8, Issue 10, pp. 1399–1413). Future Medicine Ltd. https://doi.org/10.2217/epi-2016-0056

Unoki, M. (2019). Recent Insights into the Mechanisms of De Novo and Maintenance of DNA Methylation in Mammals. *DNA Methylation Mechanism*. https://doi.org/10.5772/INTECHOPEN.89238

Veland, N., Lu, Y., Hardikar, S., Gaddis, S., Zeng, Y., Liu, B., Estecio, M. R., Takata, Y., Lin, K., Tomida, M. W., Shen, J., Saha, D., Gowher, H., Zhao, H., & Chen, T. (2019a). DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Research*, *47*(1), 152–167. https://doi.org/10.1093/NAR/GKY947

Veland, N., Lu, Y., Hardikar, S., Gaddis, S., Zeng, Y., Liu, B., Estecio, M. R., Takata, Y., Lin, K., Tomida, M. W., Shen, J., Saha, D., Gowher, H., Zhao, H., & Chen, T. (2019b). DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Research*, *47*(1), 152–167. https://doi.org/10.1093/NAR/GKY947

Veland, N., Lu, Y., Hardikar, S., Gaddis, S., Zeng, Y., Liu, B., Estecio, M. R., Takata, Y., Lin, K., Tomida, M. W., Shen, J., Saha, D., Gowher, H., Zhao, H., & Chen, T. (2019c). DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Research*, *47*(1), 152–167. https://doi.org/10.1093/NAR/GKY947

Yu, C., Cvetesic, N., Hisler, V., Gupta, K., Ye, T., Gazdag, E., Negroni, L., Hajkova, P., Berger, I., Lenhard, B., Müller, F., Vincent, S. D., & Tora, L. (2020). TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage. *Nature Communications 2020 11:1*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-20239-4

Zamudio, N. M., Scott, H. S., Wolski, K., Lo, C. Y., Law, C., Leong, D., Kinkel, S. A., Chong, S., Jolley, D., Smyth, G. K., de Kretser, D., Whitelaw, E., & O'Bryan, M. K. (2011). DNMT3L Is a Regulator of X Chromosome Compaction and Post-Meiotic Gene Transcription. *PLoS ONE*, *6*(3), 18276. https://doi.org/10.1371/JOURNAL.PONE.0018276

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., & Jeltsch, A. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the

histone H3 tail. *Nucleic Acids Research*, *38*(13), 4246–4253. https://doi.org/10.1093/NAR/GKQ147

Zhao, Z., & Shilatifard, A. (2019). Epigenetic modifications of histones in cancer. *Genome Biology 2019 20:1*, *20*(1), 1–16. https://doi.org/10.1186/S13059-019-1870-5