University of South Bohemia in České Budějovice
Faculty of Science

&

Johannes Kepler University Linz
Faculty of Engineering and Natural Sciences

Dnmt3L in mammalian oocytes

Bioinformatics Project

Student:
Lirik Behluli

Supervisor:
Mgr. Lenka Gahurová, Ph.D.

České Budějovice, 2022

CONTENTS

1. Abstract

The gene Dnmt3L is a member of the Dnmt (DNA methyltransferases) genes family which also include Dnmt1, Dnmt3a, and Dnmt3b. All of them are needed for transferring a methyl group from universal methyl donor S-adenosyl-L-methionine (SAM) to the 5-position of cytosine residues in DNA, a process which is called DNA methylation and which is crucial for mammalian development. DNMT3L is a co-factor of DNMT3A and DNMT3B with similar structures to them, but misses catalytic protein domain and therefore is enzymatically inactive.

In literature, DNMT3L is presented as one of the key factors of de novo DNA methylation in mammalian oocytes. In this project, we analyzed the expression of Dnmt3l and other DNA methyltransferases in oocytes of multiple mammalian species and identified a specific sequence within the intron of the neighboring Aire gene that arose in a rodent lineage leading to mouse, rat and hamster, and serves as an oocyte-specific Dnmt3l promoter. Moreover, once the promoter sequence appeared, it became more conserved than other intronic sequences of the Aire gene. Our results, therefore, showed that Dnmt3l plays a role in oocyte de novo DNA methylation only in a specific rodent lineage and is not a universal mammalian oocyte factor.

2. Introduction

2.1. DNA Methylation

DNA methylation is an epigenetic mechanism controlling gene expression. DNA methylation is vital to various cellular processes comprising embryonic development, X-chromosome inactivation, genomic imprinting, carcinogenesis, and the stability of chromosomes. It also enables the suppression of repetitive elements including endogenous retroviral sequences. Overall, DNA methylation and its remodelling is crucial for correct development and differentiation, enabling a single cell to grow into a complex multi-cellular organism (Moore et al., 2013).

In the mammalian genome, DNA methylation involves the transfer of a methyl group onto the C5 position of the cytosine to form 5-methylcytosine. DNA methylation is one of the epigenetic mechanisms regulating gene expression by recruiting methyl-binding proteins that sterically interfere with transcription factor binding as well as affecting the binding of some transcription factors itself. Therefore, DNA methylation is considered a generally repressive epigenetic modification. During development, DNA methylation pattern in the genome is remodeled by both de novo DNA methylation and methyl group removal, and as a result, differentiated cells develop a stable and unique DNA methylation pattern that regulates tissue-specific transcription profile (Moore et al., 2013).

To facilitate DNA packaging into the cell nucleus, DNA is wrapped around nucleosomes, octamers that contain two histones H2a, H2B, H3, and H4. Histones have extended terminal tails that can acquire an excess of posttranslational modifications (PTMs) such as methylation or acetylation (Kouzarides et al., 2007). For example, histone 3 lysine 4 trimethylation (H3K4Me3) is localized at active promoters and transcription start sites (TSSs) and is associated with gene activation (Santos-Rosa et al., 2002), H3K36me3 marks gene bodies of active genes, while H3K9me2/3 and H3K27me3 are associated with transcriptionally silent heterochromatin. Altogether, DNA methylation and histone PTMs control the chromatin accessibility as well as packaging, and that allows gene activation or repression (Sendžikaitė et al., 2019).

In the mammalian somatic cells, DNA methylation is localized throughout the majority of the genome, excluding active gene regulatory regions such as promoters, enhancers, or CpG islands that are in general unmethylated. The oocyte is unique because both the genomic methylation pattern and its function are distinct from somatic cells. In oocytes, DNA methylation is mostly constrained to actively transcribed regions (gene bodies of active genes). This gives the oocyte genome the pattern of highly methylated gene bodies which are separated by intergenic or transcriptionally inactive regions with low methylation levels (Veselovska et al., 2015; Kobayashi et al., 2012).

DNA methylation is thought to be unessential for the development of oocytes as genetic ablation of oocyte methylation allows successful fertilization and embryonic development until the mid-gestation stage (Kaneda et al., 2004), but it is crucial for genomic imprinting (parent-of-

original allele specific gene expression, regulated by allele-specific DNA methylation). Failure to establish imprints leads to embryonic or postnatal lethality or congenital diseases in both mice and humans, and the observed pathologies are mostly linked to placental and fetal growth, as well as brain and metabolic functions (Sendžikaitė et al., 2019). Only three imprinted loci are regulated by DNA methylation in male gametes, meanwhile, there are at least 26 DNA methylation-dependent imprinted regions conversed in the oocyte (Tucci et al., 2019; Smallwood et al., 2011; Kobayashi et al., 2012).

Complete loss of DNA methylation in oocytes leads to embryonic lethality at E10.5, which was originally attributed to defects emerging from imprint loss (Kaneda et al., 2004; Bourc'his et al., 2001). However, recent evidence suggests that disruption of the oocyte methylome leads to developmental defects which are unrelated to effects of disrupted imprinting (Sendžikaitė et al., 2019). These include oocyte-ablation of *Kdm1a* or *Mll2*, which impair few or no imprints at all, but resulting minor global loss of gene-body methylation leads to defects of maternal-to-zygotic transcriptional transition or ovulation failure (Stewart et al., 2015). Nevertheless, the effects of *Kdm1a* and *Mll2* knockouts could also be mediated by chromatin alterations instead of direct effect of aberrant DNA methylation (Sendžikaitė et al., 2019). On the other hand, another example of *Stella* and *Uhrf1* knockout oocytes which also show limited alteration to imprints but a strong global effect on DNA methylation, arrest at the blastocyst stage (Bostick et al., 2007). These results lead to conclusion that although correct pattern of DNA methylation is not required for maturation or development of the oocyte and fertilization, it is crucial for embryonic development beyond imprinting, in a way that is not fully understood (Sendžikaitė et al., 2019).

As information at the epigenetic level is potentially transmitted to the next generation to influence gene expression, it is important to understand how the information is patterned during the development and growth of oocytes. It is known that all the cells within an organism contain the same genome, they acquire different functions, and different profiles of DNA methylation as histone modifications. Epigenetic modifications including DNA methylation play a key role during the cell lineage specification throughout development and maintenance during cell division by regulating chromatin function (Bernstein et al., 2007).

Throughout mammalian embryo development, DNA methylation is erased globally in the primordial germ cells, which specify from cells of the epiblast. Therefore, the primary oocytes are almost lacking methylation. The methylation is then established de novo in the later phases of oocyte growth, peaking in the oocyte-specific pattern. Oocyte represents a very interesting model to study the mechanisms of DNA methylation due to the fact that an entire methylation landscape is organized in a non-dividing cell (Demond et al., 2020).

Even though the oocyte has a unique DNA methylation pattern, it relies on classical DNA methylation machinery of DNA methyltransferases (DNMT) family. This DNMT family in mammals comprises five members: DNMT1, DNMT3A, DNMT3B, DNMT3L, and DNMT3C (which

has been only recently discovered), with one of them as maintenance, three as de novo methyltransferases and a cofactor (Sendžikaitė et al., 2019).

Classically, DNMT3A and DNMT3L are considered the key factors of de novo DNA methylation establishment in mammalian oocytes, as upon knockout of either gene, oocytes remain largely unmethylated (Demond et al., 2020). However, the research studying the oocyte role of DNMT3L was performed on the mouse, and later studies showed that *DNMT3L* is not expressed in human oocytes (Huntriss et al., 2004)

DNMT1, the maintenance DNMT, recognizes and methylates the unmethylated strand on hemimethylated DNA (Hermann et al., 2004). Homozygous deletion of DNMT1 results in embryonic lethality (Li et al., 1992), but does not appear to play a strong role in the oocyte DNA methylation (Shirane et al., 2013). DNMT3A together with DNMT3B and DNMT3C are de novo methyltransferases that use unmethylated DNA as a substrate. DNMT3A and DNMT3B show partial redundancy, and they are both needed for epigenetic preprogramming during embryogenesis; meanwhile DNMT3C, is a murine-specific de novo methyltransferase silences evolutionary young retrotransposons in prospermatogonia by methylating their promoters (Barau et al., 2016). DNMT3L is an odd member of the DNMT family, as it does not have an active catalytic domain (Aapola et al., 2000). It is also less conserved between the species and is only found in mammals with genomic imprinting. The C-terminal domain of DNMT3L can bind DNMT3A and DNMT3B C-terminal domains, enhancing their chromatin binding or catalytic activity by the formation of tetramers (Chen et al., 2005; Jia et al., 2019). The expression of *Dnmt3a*, *Dnmt3b*, and *Dnmt3l* increases during the oocyte growth, peaking towards the germinal vesicle (GV) stage when de novo methylation is complete (Sendžikaitė et al., 2019). DNMT3A is needed for the enzymatic reaction of methyl group transfer, although it relies on interaction with DNMT3L for genomic targeting (Smallwood et al., 2011). In mice, deletion of *Dnmt3a* or *Dnmt3l* results in failure to establish DNA methylation in the oocyte including the loss of maternal imprints in offspring (Sendžikaitė et al., 2019). On the other hand, DNMT3B, despite its expression, does not seem to play a role in DNA methylation in the oocyte. Nevertheless, in other biological contexts (Suetake et al., 2004), DNMT3B appears to be able to bind DNMT3L (Sendžikaitė et al., 2019).

Many researchers have performed studies to grasp how the oocyte-specific DNA methylation pattern is established. In general, DNMT3s show limited target sequence specificity (Smallwood et al. 2011). But, N-terminal regulatory domains of these proteins, the ADD (ATRX-DNMT3-DNMT3L) and PWWP (Pro-Trp-Trp-Pro motif) domains, have shown to interact with various histone PTMs and guide DNMT localization as well as enzymatic activity (Sendžikaitė et al., 2019). Newest advances in methods of low-input chromatin immunoprecipitation requiring just few hundred cells (Xu et al., 2018) have allowed the interrogation of the histone PTMs localization with respect to specific genomic features and DNA methylation status. In combination with knockouts of specific histone modifier enzymes, these advances are bringing

up light on interactions between DNA methylation, histone PTMs and transcriptional activity (Sendžikaitė et al., 2019).

ADD domains of DNMT3A and DNMT3L have a high affinity to the N-terminal region of histone 3, mainly when unmethylated at lysine 4 (H3K4), and this interaction promotes catalysis of DNA methylation (Ooi et al., 2007). Moreover, when DNMT3A is in complex with DNMT3L (which is supposedly the case in the oocyte), DNMT3L ADD domain recruitment is enough to engage the whole complex (Jia et al., 2007). Trimethylated lysine H3K4 (H3K4me3) is localized to active gene promoters in somatic cells and inhibits DNMT3A activity (Ooi et al., 2007). Structural studies identified that DNMT3A is intrinsically in an autoinhibitory allosteric conformation which is driven by the ADD domain. Therefore, the ADD domain masks the DNA binding site of the catalytic domain. Recognition of unmethylated H3K4 specifically allows a structural shift and disassembles ADD-catalytic domain interaction which allows activation of DNMT3A enzymatic function (Guo et al., 2015). This means that DNA methylation and H3K4me3 are mutually exclusive in the genome, and engineering of ADD domain to lose sensitivity to H3K4me3 leads to gaining methylation aberrantly over these domains (Noh et al., 2015). As the function of the ADD domain is relatively well studied in somatic cells with many insights into its mechanisms, recent evidence of methylation patterns in oocytes suggests that it plays a similar role in the oocyte, despite the fact that no studies have been conducted to study the precise mechanism in the oocytes (Sendžikaitė et al., 2019).

Histone PTM H3K36me3 is a hallmark of active transcription and is localized predominantly to gene bodies of active genes. In *Saccharomyces cerevisiae*, H3K36 methylation is mediated solely by SET2 during transcription elongation (Huang et al., 2018). In metazoans nonetheless, multiple H3K36-specific methyltransferases exist (Huang et al., 2018). PWWP domain is found only in DNMT3A and DNMT3B (Qin et al., 2014). This domain contains a conserved aromatic cage that enables methylated lysines to bind, especially H3K36me3 (Wu et al., 2011; Rondelet et al., 2016). Biochemical studies suggest that this domain of DNMT3A interacts specifically and exclusively with H3K36me2/3 (Dhalayan et al., 2010), and they show that a point mutation (called D329A) within the aromatic cage of the DNMT3A PWWP domain disrupts binding of H3K36me3 in vitro. In mouse embryonic stem cells, it is DNMT3B, and not DNMT3A, catalyzing DNA methylation at H3K36me3-marked domains (Neri et al., 2007; Baubec et al., 2015). This was later supported by artificially expressing DNMT3B in yeast (Morselli et al., 2015). No DNA methylation defects within gene bodies or H3K36me3 domains were detected if mouse DNMT3A PWWP domain was disrupted in embryos, adult brains or oocytes (Sendžikaitė et al., 2019). Therefore, it remains a question whether DNMT3A in oocytes is recruited by transcription-dependent H3K36me3 mark over gene bodies (Stewart et al., 2015).

In this project, we would like to shed more light on the requirement for DNMT3L for DNA de novo methylation establishment in the oocytes across mammals. We would like to answer the question of whether the example of the mouse (with oocyte DNMT3L) or human (without oocyte DNMT3L) is more common in mammals and what are the molecular reasons for these

differences. Since de novo DNA methylation establishment in the oocytes is tightly associated with the profile of histone modifications, during this project we also dived into researching the expression of various epigenetic factors in oocytes, to see if there is a potential for differences in epigenomic profiles associated with the presence or absence of DNMT3L.

3. AIMS

- Analyzing DNMT3L expression in oocytes of multiple mammalian species

- Identify sequence-level changes which could lead to either activation or silencing the expression of Dnmt3l gene in the oocyte

- Predict oocyte expression of Dnmt3l across rodent species.

- Profile expression levels of genes encoding proteins with epigenetic functions in the oocytes of available mammalian species

4. Materials and methods

4.1. Analysis of gene expression

To quantify the expression of genes *Dnmt3l*, *Dnmt3a*, *Dnmt3b* and *Dnmt1*, as well as genes encoding proteins with epigenetic functions, in mammalian oocytes, we used publicly available oocyte RNA-seq datasets, as well as new datasets generated in our laboratory. All datasets were trimmed, quality controlled and mapped to the respective genomes previously in the laboratory. List of the datasets and genomes assembly versions to which the data were mapped can be found in table 1.

SeqMonk (https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) software was utilized to visualize and quantify RNA-seq datasets. SeqMonk is a bioinformatical program (software) for visualizing and analyzing mapped next-generation sequencing (NGS) data. SeqMonk has a visual interface incorporating genome browser that allows the user to navigate in a quick manner around annotated genomes, and moreover, it contains a set of tools that allows the user to quantify and filter data so that it is possible to find regions of interest in a systematic way. During our analyses, SeqMonk was used to quantify oocyte RNA-sequence datasets, using an RNA-sequence quantitation pipeline specifying strand specificity reads and whether the given data was the single-end or paired-end. The expression values are log2 transformed RPKM (reads per million of the reads in the dataset per 1 kilobase of the gene length) values for single end data or FPKM (fragments per million of the reads in the dataset per 1 kilobase of the gene length) values for paired end data.

Table1. The list of oocyte RNA-seq datasets and genome assemblies used in this project

Species	Accession code	Reference	Genome
mouse	GSE70116	Veselovska et al., 2015	GRCm38
rat	GSE112622	Brindamour et al. 2018	Rnor_6.0
golden hamster	GSE86470	Franke et al., 2017	MesAur1.0
guinea pig	n/a	Gahurova, unpublished data	Cavpor3.0
naked mole rat	n/a	Gahurova, unpublished data	hetGla2
cow	GSE61717	Reyes et al., 2015	UMD3.1
pig	GSE108900	Tsai et al., 2018	Sscrofa11.1
macaque rhesus	GSE103313, GSE86938	Chitwood et al., 2017, Wang et al., 2017	Mmul_10
human	GSE36552, GSE101571	Yan et al., 2013, Wu at al., 2018	GRCh38

4.2. *Dnmt3l* sequence in the naked mole rat genome

Resources listed below were used for the purpose of confirming that *Dnmt3l* is annotated at the correct position in the naked mole rat genome, and in case it is not; to identify where it is in the genome based on the sequence similarity with human Dnmt3l sequence. To blast Dnmt3l sequences to the naked mole rat genome, we used NCBI Genome Workbench (https://www.ncbi.nlm.nih.gov/tools/gbench/). It is a set of integrated tools to study and analyze genetic data where the user can explore and compare data from multiple sources including NCBI databases or own datasets. It is supported by alignment tools such as BLAST, Clustal, MAFFT.

Naked mole rat *Dnmt3a*, *Dnmt3b* and *Dnmt3l* sequences were used for a custom screening program (Appendix 1.) to identify whether in *Dnmt3l* there are identical sequences to *Dnmt3a* and *Dnmt3b* that can cause *Dnmt3l* reads mapping incorrectly to *Dnmt3a* and *Dnmt3b*.

Human Dnmt3l was retrieved from NCBI:

https://www.ncbi.nlm.nih.gov/gene/29947

Naked mole rat Dnmt3a transcript variants were retrieved from the NCBI:
https://www.ncbi.nlm.nih.gov/nuccore/XM_004839112.3,XM_004839113.3,XM_004839114.3,XM_004839115.3,XM_004839117.3,XM_021253567.1,XM_021253568.1,XM_021253569.1,XM_021253570.1,XM_021253571.1

Naked mole rat Dnmt3b variants were retrieved from the NCBI
https://www.ncbi.nlm.nih.gov/nuccore/XM_004867953.3,XM_004867957.3,XM_004867958.3,

XM_021246874.1,XM_021246875.1,XM_021246876.1,XM_021246877.1,XM_021246878.1,XM_021246879.1,XM_021246880.1,XM_021246881.1,XM_021246882.1

Naked mole rat Dnmt3l variants were retrieved from the NCBI: https://www.ncbi.nlm.nih.gov/nuccore/XM_004842591.2,XM_021256378.1

Naked mole rat genome was downloaded from https://hgdownload.soe.ucsc.edu/goldenPath/hetGla2/bigZips/

4.3. Multiple sequence alignment

Clustal programs were implemented heavily during our analyses. In general, Clustal is part of the family of sequence alignment programs, used to align related RNA, protein, or DNA sequences. Generally, Clustal starts with a set of sequences and performs a series of pairwise alignments to build a relationship tree that shows how similar given sequences are. Clustal program performs multiple sequence alignment, while other programs (generally) will do mostly just alignments of 2 sequences (pairwise sequence alignment). In our analysis, we compared *Aire* intron sequences using ClustalW (https://www.genome.jp/tools-bin/clustalw) with default parameters. We used ClustalW, as the idea was to perform multiple sequence alignment taking into consideration matches, substitutions and insertions/deletions with each of these having a score. ClustalW gives scores for each pairwise comparison.

4.4. Usage of Python and its libraries

Programming language Python was used to develop scripts for specified analyses. Python is an interpreted high-level general-purpose language and is widely used among natural scientists due to its design philosophy emphasizing code readability. The language constructs as well as its object-oriented approach aim to help to write clear, logical code, either for small or for large projects. Two python scripts were utilized during our analyses. The first script (Appendix 1) takes two sequence files as input, compares between them and checks if given number of base pairs (in this case 5, 10, 11, 12, 13, 15) is the same in Dnmt3l, Dnmt3a and Dnmt3b. The second script takes the intronic sequences as input and graphically plots the promoter regions which gives us insights whether they are more conserved than normally (Appendix 2). The parameters used are explained in Appendix 2.

Biopython, Numpy & Matplotlib

Python usually has built-in libraries, or libraries that have to be downloaded manually. Libraries are a collection of related modules which contain bundles of code that can be used repeatedly in different programs and makes it simpler for the programmer to not write the same code many times.

Biopython is a library containing sets and available tools for biological computation.

NumPy is a library for linear algebra. At its core, there is NumPy array, a multi-dimensional data structure that can be used to represent vectors and matrices.

Matplotlib is a library used for data visualization, where you can create bar-plots, scatter-plots, histograms, and so on.

4.5. Usage of Notepad++ software

Notepad++ was used for searching keywords in txt/fasta files containing a list of thousands of results, to select our desired sequences as needed, as they were marked with special keywords or Ensembl codes. Notepad++ is a text and source code editor which works for the operating system Windows. It supports tabbed editing and allows working with multiple open files in a single window.

Notepad++ software allows us to search for particular keywords thus making it easier to find our targets (Figure 1.). At particular organisms, they were mostly annotated with Ensembl codes or numbers, and they were searched using these keywords.

4.6. Usage of R and its package ggplot2

R is a sophisticated computer programming language and environment for statistical computing and graphics. It is an open- source statistical environment modeled after S and S-plus (which was developed back then in the 1980s at AT&T labs). R project was started by Robert Gentleman and Ross Ihaka, at the Statistics Department of the University of Auckland in 1995 (Gardener et al., 2012). It has a widespread audience, especially in today's time when science generates a lot of data to be analyzed, and Biology is one of them, especially Bioinformatics.

For generation of boxplots of expression levels of Dnmts in selected mammalian species (section 5.1.), R ggplot2 package was used. ggplot2 is a R package for producing statistical, or data graphics. These plots can be built up iteratively from the user using programming commands (Wickham et al., 2016). During our analysis, ggplot2 was used to make boxplots with specifications to show average expression values +/- standard deviation as lower and upper border of the box, and +/- 3x standard deviation as whiskers (Appendix 7.).

5. Results

5.1. Quantification of expression of Dnmt1, Dnmt3a, Dnmt3b and Dnmt3l in oocytes from different mammalian species

First, we aimed to analyse the oocyte expression of Dnmt1, Dnmt3a, Dnmt3b and Dnmt3l in multiple mammalian species (mouse, rat, golden hamster, guinea pig, naked mole rat, cow, pig, macaque rhesus and human). We used publicly available RNA-seq datasets and datasets generated in our laboratory (see Table 1 in Methods section). All these datasets were

previously processed and mapped to respective genomes. The quantification of the expression of all genes was performed using SeqMonk.

After exporting the expression values from SeqMonk, we used Notepad++ to search through the whole list to find the expression values of Dntm3l, Dnmt3a, Dnmt3b, and Dnmt1 in each species. A table with the expression values was created (Appendix 4).
Figures 1-9 show the average values of Dnmt3L, Dnmt3A, Dnmt3B, and Dnmt1 expression in the individual species (besides hamster where Dnmt3B is not present). Guinea pig (figure 8) and hamster (figure 9) could not be visualized as boxplots as there was only one replicate of RNA-seq data.

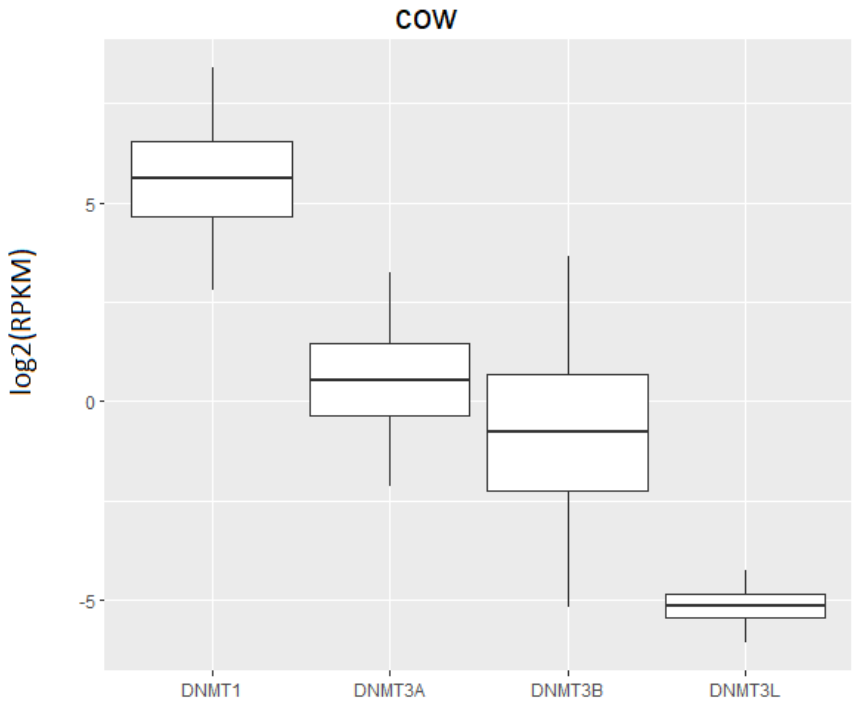

Figure 1. Boxplot showing average expression values (boxplot mid-line), standard deviation (lower and upper box boundary) and 3x standard deviation as whiskers, of each DNMT gene in the oocytes of cow

Figure 15. Bar plots: Pairwise alignment scores of multiple sequence alignments for each Intron (1 to 5) sequences of golden hamster, chinese hamster, northern american deer mouse, mongolian gerbil, steppe mouse, mouse, algerian mouse, ryukyu mouse, rat. The columns are in order from comparison of sequences 1:2 to 8:9 and the numbers correspond to species as seen in Figure 23, Appendix 5.

5.3. Local conservation

We then analysed if the region of oocyte-specific promoter sequence is more conserved than the rest of the intron 3 sequence.

A python script was written by the author and utilized to perform the analysis allr all alignments of the introns 1 to 5 of the species golden hamster, chinese hamster, northern american deer mouse, mongolian gerbil, steppe mouse, mouse, algerian mouse, ryukyu mouse, rat to show that the promoter regions are more conserved than other sequences in these intronic sequences (Appendix 2).

Upon implementation of the script, the results (plots) showed that the promoter sequence within intron 3 show high local conservation (Figure 18.), compared to the other parts of the same intron or other intronic sequences of Aire gene (the only other similar region is around base 50 in intron 4).

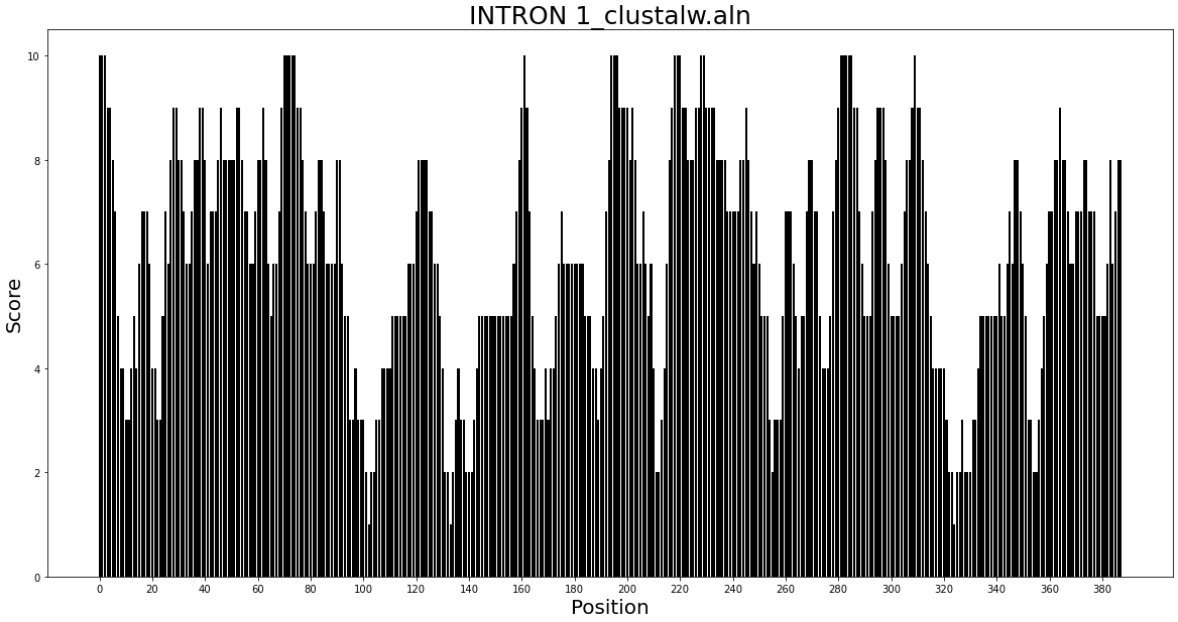

Figure 16. Graphical plot (Intron 1) using Matplotlib library of the Python language.

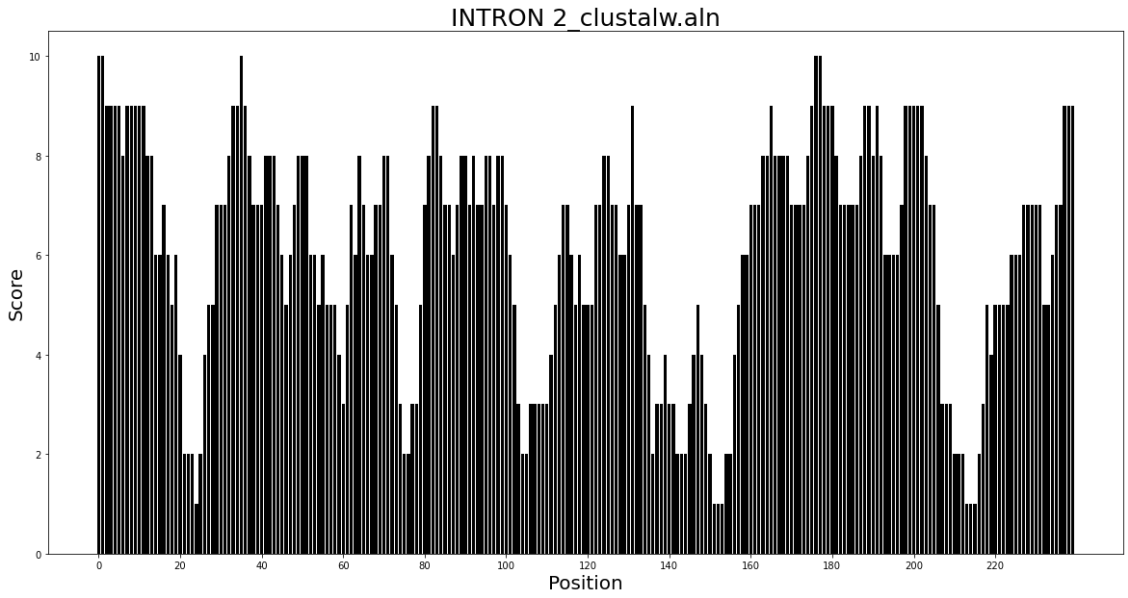

Figure 17. Graphical plot (Intron 2) using Matplotlib library of the Python language.

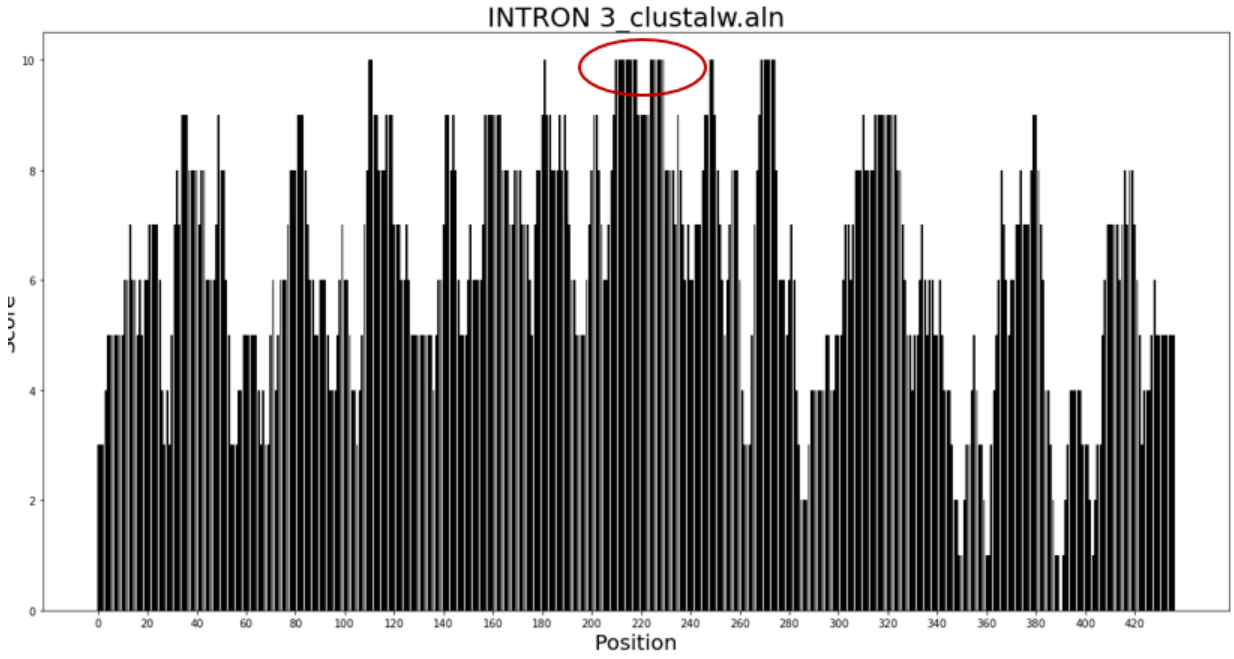

Figure 18. Graphical plot (Intron 3) using Matplotlib library of the Python language.

Highlighted area where it is more conserved shown in red

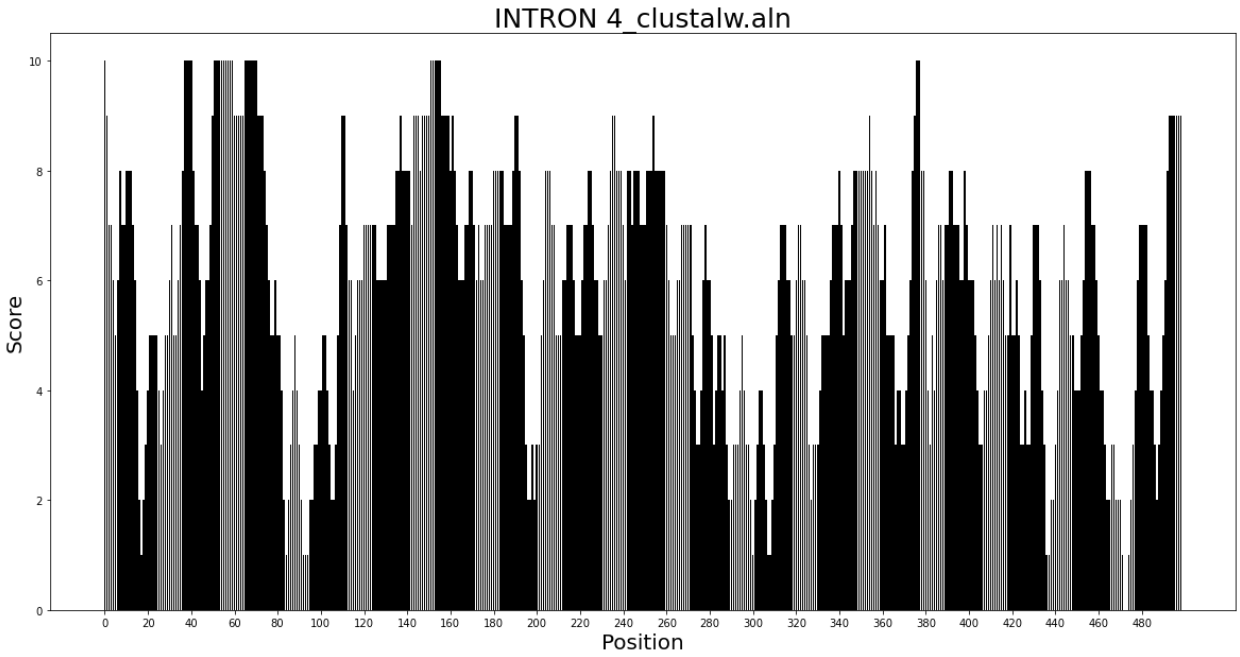

Figure 19. *Graphical plot (Intron 4) using Matplotlib library of the Python language.*

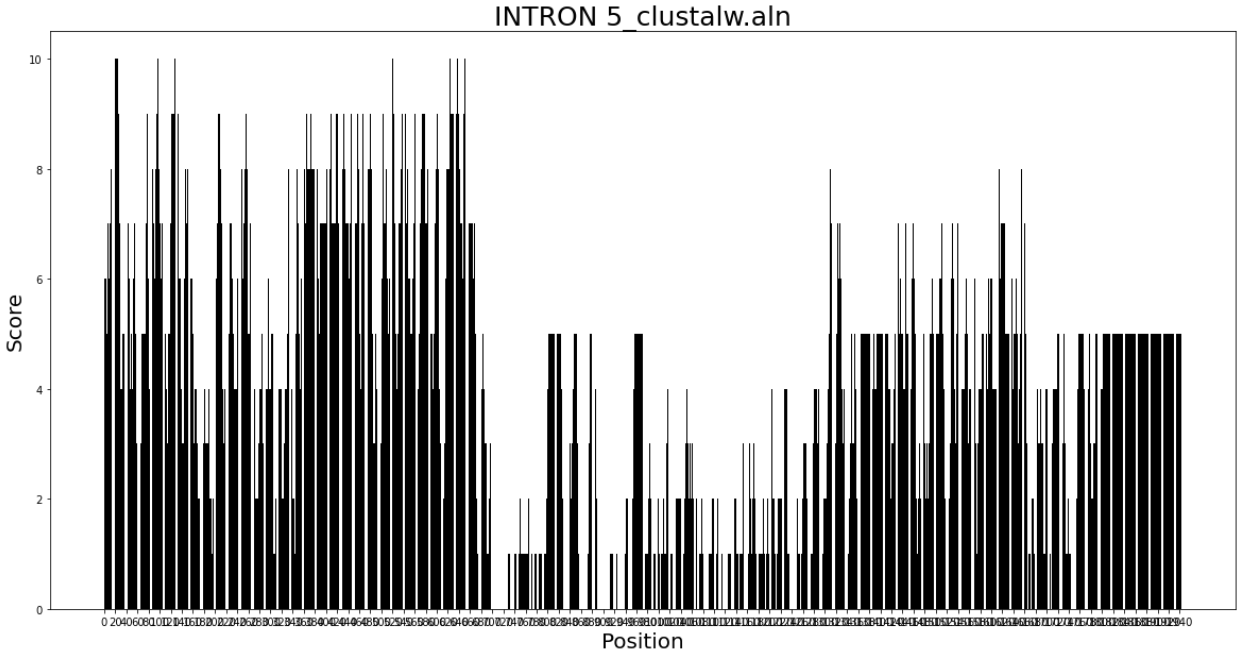

Figure 20. Graphical plot (Intron 5) using Matplotlib library of the Python language.

7. References

1. Aapola U., Kawasaki K., Scott H.S., Ollila J., Vihinen M., Heino M. et al. . (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 10.1006/geno.2000.6168
2. Andrews, S. (2007). SeqMonk: A tool to visualise and analyse high throughput mapped sequence data.
3. Barau J., Teissandier A., Zamudio N., Roy S., Nalesso V., Hérault Y. et al. . (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354, 909–912 10.1126/science.aah5143
4. Baubec T., Colombo D.F., Wirbelauer C., Schmidt J., Burger L., Krebs A.R. et al. . (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243–247 10.1038/nature14176
5. Bell, O., Wirbelauer, C., Hild, M., Scharf, A. N., Schwaiger, M., MacAlpine, D. M., ... & Schübeler, D. (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. The EMBO journal, 26(24), 4974-4984.
6. Bernstein B.E., Meissner A. and Lander E.S. (2007) The Mammalian Epigenome. Cell 128, 669–681 10.1016/j.cell.2007.01.033
7. Bostick M., Kim J.K., Estève P.-O., Clark A., Pradhan S. and Jacobsen S.E. (2007) UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science 317, 1760–1764 10.1126/science.1147939
8. Bourc'his D., Xu G.-L., Lin C.-S., Bollman B. and Bestor T.H. (2001) Dnmt3L and the Establishment of Maternal Genomic Imprints. Science 294, 2536–2539 10.1126/science.1065848
9. Brind'Amour, J., Kobayashi, H., Richard Albert, J., Shirane, K., Sakashita, A., Kamio, A., ... & Lorincz, M. C. (2018). LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. Nature communications, 9(1), 1-14.
10. Chedin, F., Lieber, M. R., & Hsieh, C. L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proceedings of the National Academy of Sciences, 99(26), 16916-16921.
11. Chen Z.-X., Mann J.R., Hsieh C.-L., Riggs A.D. and Chédin F. (2005) Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J. Cell. Biochem. 95, 902–917 10.1002/jcb.20447
12. Chitwood, J. L., Burruel, V. R., Halstead, M. M., Meyers, S. A., & Ross, P. J. (2017). Transcriptome profiling of individual rhesus macaque oocytes and preimplantation embryos. Biology of reproduction, 97(3), 353-364.
13. Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... & De Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25(11), 1422-1423.
14. Demond, H., & Kelsey, G. (2020). The enigma of DNA methylation in the mammalian oocyte. F1000Research, 9.

15. Dhayalan A., Rajavelu A., Rathert P., Tamas R., Jurkowska R.Z., Ragozin S. et al. . (2010) The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. J. Biol. Chem. 285, 26114–26120 10.1074/jbc.M109.089433
16. Duymich, C. E., Charlet, J., Yang, X., Jones, P. A., & Liang, G. (2016). DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nature communications, 7(1), 1-9.
17. Edmunds, J. W., Mahadevan, L. C., & Clayton, A. L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. The EMBO journal, 27(2), 406-420.
18. Franke, V., Ganesh, S., Karlic, R., Malik, R., Pasulka, J., Horvat, F., ... & Svoboda, P. (2017). Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome research, 27(8), 1384-1394.
19. Gardener, M. (2012). Beginning R: the statistical programming language. John Wiley & Sons.
20. Guo X., Wang L., Li J., Ding Z., Xiao J., Yin X. et al. . (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640–644 10.1038/nature13899
21. Hanna, C. W., Huang, J., Belton, C., Reinhardt, S., Dahl, A., Andrews, S., ... & 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.
22. Hanna, C. W., Taudt, A., Huang, J., Gahurova, L., Kranz, A., Andrews, S., ... & Kelsey, G. (2018). MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. Nature structural & molecular biology, 25(1), 73-82.
23. Hermann A., Goyal R. and Jeltsch A. (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. J. Biol. Chem. 279, 48350–48359 10.1074/jbc.M40342720
24. Huang, C., & Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. Biophysics reports, 4(4), 170-177.
25. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03), 90-95.
26. Huntriss, J., Hinkins, M., Oliver, B., Harris, S. E., Beazley, J. C., Rutherford, A. J., ... & Picton, H. M. (2004). Expression of mRNAs for DNA methyltransferases and methyl-CpG-binding proteins in the human female germ line, preimplantation embryos, and embryonic stem cells. Molecular Reproduction and Development: Incorporating Gamete Research, 67(3), 323-336.
27. Jiang, Z., Sun, J., Dong, H., Luo, O., Zheng, X., Obergfell, C., ... & Tian, X. C. (2014). Transcriptional profiles of bovine in vivo pre-implantation development. BMC genomics, 15(1), 1-15.
28. 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, 449(7159), 248-251.

29. Kaneda M., Okano M., Hata K., Sado T., Tsujimoto N., Li E. et al. . (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429, 900–903 10.1038/nature02633
30. Kobayashi H., Sakurai T., Imai M., Takahashi N., Fukuda A., Yayoi O. et al. . (2012) Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. PLos Genet. 8, e1002440 10.1371/journal.pgen.1002440
31. Kouzarides T. (2007) Chromatin Modifications and Their Function. Cell 128, 693–705 10.1016/j.cell.2007.02.005
32. Kuznetsov, A., & Bollin, C. J. (2021). NCBI Genome Workbench: desktop software for comparative genomics, visualization, and GenBank data submission. In Multiple Sequence Alignment (pp. 261-295). Humana, New York, NY.
33. Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. Neuropsychopharmacology, 38(1), 23-38.
34. Morselli M., Pastor W.A., Montanini B., Nee K., Ferrari R., Fu K. et al. . (2015) In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. eLife. 4, e06205 10.7554/eLife.06205
35. Neri F., Rapelli S., Krepelova A., Incarnato D., Parlato C., Basile G. et al. . (2017) Intragenic DNA methylation prevents spurious transcription initiation. Nature 543, 72–77 10.1038/nature21373
36. Noh K.-M., Wang H., Kim H.R., Wenderski W., Fang F., Li C.H. et al. . (2015) Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. Mol. Cell 59, 89–103 10.1016/j.molcel.2015.05.017
37. Nowialis, P., Lopusna, K., Opavska, J., Haney, S. L., Abraham, A., Sheng, P., ... & Opavsky, R. (2019). Catalytically inactive Dnmt3b rescues mouse embryonic development by accessory and repressive functions. Nature communications, 10(1), 1-16.
38. Ooi S.K.T., Qiu C., Bernstein E., Li K., Jia D., Yang Z. et al. . (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714–717 10.1038/nature05987
39. Python, R. (2019). Python. Python Releases for Windows, 24.
40. Qin S. and Min J. (2014) Structure and function of the nucleosome-binding PWWP domain. Trends Biochem. Sci. 39, 536–547 10.1016/j.tibs.2014.09.001
41. Rondelet G., Dal Maso T., Willems L. and Wouters J. (2016) Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J. Struct. Biol. 194, 357–367 10.1016/j.jsb.2016.03.013
42. Santos-Rosa H., Schneider R., Bannister A.J., Sherriff J., Bernstein B.E., Emre N.C.T. et al. . (2002) Active genes are tri-methylated at K4 of histone H3. Nature 419, 407–411 10.1038/nature01080
43. Sendžikaitė, G., & Kelsey, G. (2019). The role and mechanisms of DNA methylation in the oocyte. Essays in biochemistry, 63(6), 691-705.
44. Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., ... & Sasaki, H. (2013). Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of

non-CpG methylation and role of DNA methyltransferases. PLoS genetics, 9(4), e1003439.
45. Shum, J. (2013). Jerry Blogger: Notepad++ v6. 5.1-Current Version with DIRECT DOWNLOAD link.
46. Smallwood S.A., Tomizawa S., Krueger F., Ruf N., Carli N., Segonds-Pichon A. et al. . (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat. Genet. 43, 811–814 10.1038/ng.864
47. Stewart, K. R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S. I., ... & Kelsey, G. (2015). Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. Genes & development, 29(23), 2449-2462.
48. Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinformatics, (1), 2-3.
49. Tsai, T. S., Tyagi, S., & St. John, J. C. (2018). The molecular characterisation of mitochondrial DNA deficient oocytes using a pig model. Human Reproduction, 33(5), 942-953.
50. Tucci V., Isles A.R., Kelsey G., Ferguson-Smith A.C., Tucci V., Bartolomei M.S. et al. . (2019) Genomic Imprinting and Physiological Processes in Mammals. Cell 176, 952–965 10.1016/j.cell.2019.01.043
51. Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. Computing in science & engineering, 13(2), 22-30.
52. Van Emburgh, B. O., & Robertson, K. D. (2011). Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. Nucleic acids research, 39(12), 4984-5002.
53. Veselovska, L., Smallwood, S. A., Saadeh, H., Stewart, K. R., Krueger, F., Maupetit-Méhouas, S., ... & Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome biology, 16(1), 1-17.
54. Wagner, E. J., & Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. Nature reviews Molecular cell biology, 13(2), 115-126.
55. Wang, X., Liu, D., He, D., Suo, S., Xia, X., He, X., ... & Zheng, P. (2017). Transcriptome analyses of rhesus monkey preimplantation embryos reveal a reduced capacity for DNA double-strand break repair in primate oocytes and early embryos. Genome research, 27(4), 567-579.
56. Wickham, H. (2016). ggplot2: elegant graphics for data analysis. springer.
57. Wu H., Zeng H., Lam R., Tempel W., Amaya M.F., Xu C. et al. . (2011) Structural and Histone Binding Ability Characterizations of Human PWWP Domains. PLoS One 6, e18919 10.1371/journal.pone.0018919
58. Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z., ... & Sun, Y. (2018). Chromatin analysis in human early development reveals epigenetic transition during ZGA. Nature, 557(7704), 256-260.
59. Xu Q. and Xie W. (2018) Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. Trends Cell Biol. 28, 237–253 10.1016/j.tcb.2017.10.008

60. Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., ... & Tang, F. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature structural & molecular biology, 20(9), 1131-1139.
61. Yokomine, T., Hata, K., Tsudzuki, M., & Sasaki, H. (2006). Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. Cytogenetic and genome research, 113(1-4), 75-80.
62. Yu, C., Cvetesic, N., Hisler, V., Gupta, K., Ye, T., Gazdag, E., ... & Tora, L. (2020). TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage. Nature communications, 11(1), 1-13.