CHARLES UNIVERSITY IN PRAGUE

Faculty of Science

Developmental and Cell Biology

Long Non-Coding RNAs in Oocyte-to-Embryo Transition

Ph.D. Thesis

Sravya Ganesh

Supervisor: Petr Svoboda, Ph.D.

Department of Epigenetics

Institute of Molecular Genetics of ASCR, v.v.i

Prague, 2018

Declaration

I declare that I wrote the thesis independently and that I cited all the information sources and literature. This work was not presented to obtain another academic degree or equivalent.

Prague, 30.9.2016 Sravya Ganesh

Acknowledgments

Petr, thank you for believing in me. Without your exceptional mentorship, patience and trust in me, this project wouldn't have been possible. Thank you.

All the members of Svoboda Lab (current and past), thanks for your help, support, and discussions. Especially, Radek Malik, for being an excellent mentor and a critical reviewer of this project since beginning. And, Jana Nejepinska, for always making me feel at home in the lab. I miss you.

All the members of Kristian Vlahovicek's lab, especially Kristian, Vedran, Filip and Rosa, Thank you. This project literally wouldn't have been possible without your help. I will miss our Split meetings.

RNA-Train, thanks for all the training you have provided. Especially, Anders and Anne, for all the efforts you have put in organizing RNA-Train. And, Elena and Hung for being great friends throughout.

All the members of IMG and collaborators, thank you for your constant cooperation and help throughout the project, especially the animal facility, CCP unit and microscopy facility.

Patrick, Dijana and Ines, thank you for always being there for me. Without you guys, these four years of my Ph.D. in Prague wouldn't be that fun. Utkarsh, thank you for all the discussions and suggestions. And all my friends in India and Prague, I appreciate all your support.

Martijn, you have been so patient, especially while writing this thesis. Thank you for all the support you have given me. And finally, this PhD wouldn't have been possible without the constant love and support from my family. Divya and Kavya thank you for being such great sisters. Mom and Dad, thank you for all the motivation and countless prayers for my success. I love you all.

Table of Contents

List of Abbreviations

cAMP	Cyclic Adenosine Monophosphate
cGMP	Cyclic Guanosine Monophosphate
circRNA	Circular RNA
CPAT	Coding-Potential Assessment Tool
CPE	Cytoplasmic Polyadenylation Element
CPEB	Cytoplasmic Polyadenylation Element Binding protein
CPSF	Cleavage and Polyadenylation Specificity Factor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
DMEM	Dulbecco Modified Eagle Medium
dsRNA	Double Stranded RNA
ENV	Envelope
ERV	Endogenous Retrovirus
FPKM	Fragments Per Kilo base per Million mapped reads
FSH	Follicle Stimulating Hormone
FSHD	Facioscapulohumeral Muscular Dystrophy
GFP	Green Fluorescent Protein
GR	Glucocorticoid Receptors
GRE	Glucocorticoid Response Elements
gRNA	Guide RNA
GV	Germinal Vesicle
GVBD	Germinal Vesicle Breakdown
HERV	Human Endogenous Retroviruses
IAP	Intracisternal A Particle
IBMX	3-isobutyl-1-methyl-xanthine
ICM	Inner Cell Mass
LH	Luteinizing Hormone
LHR	Luteinizing Hormone Receptor
LINE	Long Interspersed Nuclear Elements
LncRNA	Long Non-Coding RNA
LTR	Long Terminal Repeat
MaLR	Mammalian apparent LTR Retrotransposons
MII	Metaphase II

miRNA	MicroRNA
MT	Mouse Transcript
MuERV-L	Murine Endogenous Retroviruses
NAT	Natural Antisense Transcript
ncRNA	Non-Coding RNA
NSN	Non-Surrounded Nucleolus
OET	Oocyte-to-Embryo Transition
pancRNAs	Promoter Associated Non-Coding RNAs
PAP	Poly(A) Polymerase
PGC	Primordial Germ Cell
Pol II	RNA polymerase II
Pol III	RNA polymerase III
PRC2	Polycomb Repressive Complex 2
qPCR	Quantitative PCR
RISC	RNA-Induced Silencing Complex
RPM	Reads Per Million
RRM	RNA Recognition Motif
rRNA	Ribosomal RNA
SA	Splice Acceptor
SD	Splice Donor
SINE	Short Interspersed Nuclear Elements
siRNA	Small interfering RNA
SN	Surrounded Nucleolus
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
TEs	Transposable Elements
TE	Trophectoderm
TracrRNA	Trans-activating CRISPR RNA
tRNA	Transfer RNA
TSS	Transcription Start Site
UTR	Untranslated Region
ZGA	Zygotic Genome Activation

Abstract (English)

Oocyte-to-embryo transition (OET) is one of the most complex developmental events, during which a differentiated oocyte gives rise to a totipotent zygote. During OET a transcriptionally silent oocyte undergoes massive reprogramming of gene expression, which transforms it into a transcriptionally active zygote. Although numerous studies have contributed to understanding the mechanism of OET, many genes involved in OET are yet to be identified. A whole new level of possible regulation of OET came with the discovery of long non-coding RNAs (lncRNA). LncRNAs are pol II transcripts longer than 200 nucleotides, that are typically spliced and polyadenylated but do not encode proteins. While lncRNAs have been studied in many model systems including embryonic stem cells, their expression in oocytes and early embryos and contribution to OET were largely unexplored at the beginning of this project.

In my PhD project, I aimed to identify, annotate, and analyze lncRNAs expressed during OET. First, using RNA-Seq, 1600 highly reliable lncRNAs were identified and annotated in mouse oocytes and early embryos. Majority of lncRNAs were novel with expression exclusively at OET stages. A significant fraction of these lncRNAs was found associated with LTR retrotransposons, contributing to their novelty and evolution. Expression analysis of OET lncRNAs revealed, along with restricted maternal and zygotic expression profiles of OET lncRNAs, two unique classes of maternal lncRNAs. (I) A group of maternal lncRNAs were identified, which undergoes cytoplasmic polyadenylation, a mechanism which was previously associated with dormant maternal mRNAs, and (II) 100 lncRNAs with antisense pseudogene insertion were identified, which serve as substrates for endo-RNAi machinery in oocytes and give rise to endo-siRNAs. Finally, to study the role of lncRNAs during OET, loss of function mouse model of five selected lncRNAs were generated, of which three are reported here. Even though no phenotypic changes concerning fertility were observed, we validated cytoplasmic polyadenylation (i.e. lncRNA dormancy) and RNAi induction by maternal lncRNAs. Altogether, our study provides a comprehensive analysis of lncRNAs during OET, with the first look into contribution of LTR retrotransposons to lncRNA evolution in oocytes and zygotes, and the identification of two unique classes of maternal lncRNAs.

Abstrakt

Přechod z oocytů na embryo (OET) je jednou z nejkomplexnějších vývojových událostí, během níž se diferencovaný oocyt promění v totipotentní blastomery embrya. Během OET transkripčně neaktivní oocyt prochází masivním přeprogramováním genové exprese, které ho transformuje do transkripčně aktivní zygoty. Přestože četné studie přispěly k pochopení mechanismu OET, mnoho genů zapojených do OET je neznámých. Úplně nová úroveň možné regulace OET přišla s objevem dlouhých nekódujících RNA (lncRNA). LncRNAs jsou transkripty pol II delší než 200 nukleotidů, které jsou typicky sestřižené a polyadenylované, ale nekódují proteiny. Zatímco lncRNA byly studovány v mnoha modelových systémech včetně embryonálních kmenových buněk, jejich exprese v oocytech a časných embryích a příspěvek k OET byly na začátku tohoto projektu neznámé.

Ve svém doktorském projektu jsem se zaměřila na identifikaci, anotaci a analýzu lncRNA objevujících se během OET. Pomocí analýzy RNA-Seq bylo identifikováno 1600 lncRNA, které byly anotovány v myších oocytech a časných embryích. Většina lncRNA byla nová s expresí výlučně během OET. Významná část těchto lncRNA souvisela s LTR retrotransposony, což zřejmě souvisí s k jejich nedávným vznikem a evolucí. Expresní analýza OET lncRNA odhalila kromě výrazně odlišných maternálních a zygotických expresních profilů také dvě unikátní třídy maternálních lncRNA. (I) Byla identifikována skupina maternálních lncRNA, která zřejmě prochází cytoplazmatickou polyadenylací, která byla dříve spojena s dormantními maternálními mRNA. (II) Bylo identifikováno 100 lncRNA s antisense pseudogenní inzercí, které slouží jako substráty pro endo-RNAi mechanismus v oocytech a vznikají z nich endo-siRNA. Pro funkční analýzu role lncRNA během OET byly připraveny genomové delece pěti vybraných lncRNA; tři z nich jsou popsány v této práci. Ačkoliv u mutantů nebyly pozorovány žádné fenotypové změny týkajících se fertility, byla potvrzena cytoplazmatická polyadenylace (tzv. lncRNA dormance) a identifikovány maternální lncRNA sloužící jako substráty pro RNAi. Celkově tato práce poskytuje komplexní analýzu lncRNA během OET s originálním pohledem na příspěvek retrotransposonů LTR k vývoji lncRNA v oocytech a zygotách a přináší identifikaci dvou unikátních tříd mateřských lncRNA.

Introduction

In spite of enormous progress since the sequencing of several vertebrate genomes (Lander et al. 2001; Mouse Genome Sequencing et al. 2002; Consortium 2012; Howe et al. 2013), there is still much we do not understand about their evolution. For a long time we have known that, only 1.5% of the human genome contributes to protein coding genes, while the rest of the genome is occupied by non-coding DNA and transposable elements (TE) (Consortium 2012). TEs, which were previously presumed to be “junk” DNA, are now predicted to be one of the causes for substantial variation in human genome (Kapusta et al. 2017). Furthermore, TE insertions supply the genome with new functional sequences such as promoters, enhancers, and insulator elements. These functional sequences can establish novel species-specific gene regulatory networks and give rise to new genes (Gerdes et al. 2016, Franke, 2017).Simultaneously, advances in high throughput sequencing revolutionized our understanding of the non-coding regions of the genome. Several non-coding RNAs (ncRNA) including long non-coding RNAs (lncRNAs) are expressed from these regions, which not only participate in various biological processes, but also contribute to phenotypic and functional variance (Frias-Lasserre and Villagra 2017).

Altogether, investigation of activity and interplay between ncRNAs and transposable elements is key to understanding evolution. And oocytes and early embryonic stages provide the most suitable environment for study. Reprogramming of early embryonic genome is regulated by a set of genes, which are under constant evolutionary selection. And activity of TEs, which is seen exclusively in these stages, along with tissue specific expression of lncRNA are thought to be one of the key players in reprograming during oocyte-to-embryo transition (OET).

Therefore, before we explore the field of lncRNAs and Transposable elements, we need to understand the mechanisms involved in OET. The first part of the introduction introduces the oocyte biology, which is followed by our current knowledge of long non-coding RNA field and TEs.

Oocyte-to-Embryo Transition

Oocyte-to-Embryo transition (OET) is a transition where, an oocyte undergoes complex reprogramming to give rise to a totipotent embryo. The transition from oocyte to embryo involves many changes, including protein synthesis, protein and RNA degradation, and organelle remodeling (reviewed in Stitzel and Seydoux 2007, Schultz, 2018). Remarkably, this transition highly depends on messenger RNAs (mRNAs) and proteins accumulated during oogenesis. Therefore, understanding of the process of oogenesis is important to further study oocyte-to-embryo transition *(Fig 1)*.

Oogenesis

Formation of female gametes initiates before birth. During embryonic development day 7.5 (E7.5), small number of epiblast stem cells are induced to become the first precursors called Primordial Germ Cells (PGCs) (Ying et al. 2000, Sanchez, 2012). Signals from both extra embryonic ectoderm derived BMP4, BMP8B and extra embryonic endoderm derived BMP2 are required for PGC formation (Ying and Zhao 2001, Lawson, 1994). Migration, proliferation and colonization of PGCs tightly depends on a crosstalk between surrounding somatic cells and PGCs. Upon migration to the genital ridge, PGCs with XY chromosomes differentiate into male germ cells; under the influence of the Y-linked gene *Sry* and those with XX chromosomes, which lack *Sry*, develop into female germ cells (Bowles and Koopman 2010). PGCs then undergo rapid proliferation with incomplete cytokinesis and exist as “clusters” or “germ cell cyst” (Pepling 2006, Sanchez, 2012). At E13.5, PGCs within the clusters enter prophase and most of them arrest at diplotene stage of meiosis I. And by E18.5, PGC clusters breaks down to become primordial follicles (reviewed in Jagarlamudi and Rajkovic 2012).

Primordial follicles, which are surrounded by several flattened somatic cells called pre-granulosa cells have a diameter of 20μm *(Fig 1)*. Most of the primordial follicles are enclosed by pre-granulosa cells soon after birth (McGee and Hsueh 2000). Throughout germ cell cluster breakdown, a substantial number of oocytes are lost. From 6–7 million, the total number of oocytes falls drastically to less than 1 million at birth (Hansen et al. 2008). Oocytes that fail to enclose in pre-granulosa

cells are lost, presumably by apoptosis (De Pol et al. 1997). Primordial follicles formed either during mid-gestation in the human ovary, or shortly after birth in the mouse ovary provides the source of oocytes for the entire reproductive life (Green and Zuckerman 1951). Greenwald predicted, each primordial follicle may have three possible fates: to remain quiescent, to begin the development but undergo atresia later, or to develop, mature, and ovulate (Greenwald 1972). These primordial follicles have to become activated and initiate folliculogenesis to transform into primary follicles. Primordial follicle activation is a very dynamic and tightly controlled process and many molecular mechanisms regulating are still not fully understood.

During transition of primordial follicles to primary follicles, first comes changes in the pregranulosa cells followed by oocytes growth *(Fig 1)* (Lintern-Moore and Moore 1979). The growth of the oocyte during primary follicle activation is remarkable, for example, in mouse, the oocyte grows actively with an almost 300-fold increase in volume (Elvin and Matzuk 1998). Growth phase is accompanied by a 300-fold increase in RNA content (Sternlicht and Schultz 1981) and a 38-fold increase in the rate of protein synthesis (Schultz et al. 1979, Adhikari, 2009). Numerous studies have identified factors that regulate primordial follicle activation and initiation of their transition into primary follicles (Arraztoa et al. 2005, Park, 2005). For example, PI3K of PTEN/PI3K signaling pathway, is essential for primordial follicle activation (Reddy et al. 2005). Whereas, PTEN (phosphatase and tensin homolog) is a negative regulator of PI3K, and suppresses primordial follicle activation (Cantley 2002). Conditional deletion of PTEN from oocytes in primordial follicles resulted in premature activation of the entire primordial follicle pool (Reddy et al. 2008). Other factors responsible for maintenance and activation are, transcription factor FOXO3A, which helps in maintaining primordial follicles and suppresses their activation (Castrillon et al. 2003). Follicular activation is further supported by several growth factors, hormones, and cytokines (reviewed in Adhikari and Liu 2009).

When follicles reach the primary follicle or pre-antral stage, their development is considered to be gonadotropin-independent and progression to the early antral stage still requires locally released intraovarian factors. At this stage,

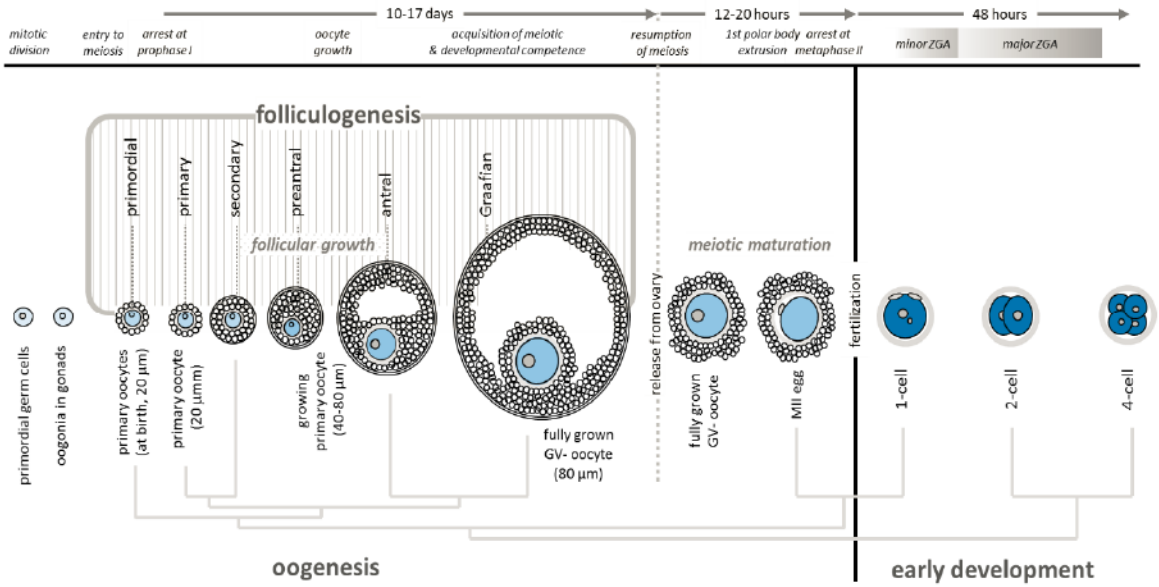


***Figure 1**: Schematic diagram of oogenesis, describing developmental stages occurring from primordial germ cell specification to formation of primordial follicles, and their growth to primary, secondary and antral follicles, to finally pre-implantation development. (Modified from Schultz et al. 2018).*

oocyte surrounding granulosa cells undergo rapid proliferation, giving rise to larger, multi-layered pre-antral follicles, followed by an antral cavity formation *(Fig 1)* (reviewed in Sanchez and Smitz 2012). Primary to secondary follicle transition of pre-antral follicles, depends on two members of the TGFβ superfamily proteins, GDF9 and BMP15. Their expression starts during early stages and continues throughout folliculogenesis and ovulation. GDF9 plays a crucial role as a potent granulosa cell mitogen (Orisaka et al. 2006). Mice ovaries lacking *Gdf9* were shown to form primordial follicles, but did not progress beyond the primary follicle stage (Hayashi et al. 1999, Nilsson, 2002). BMP15, on the other hand, induces proliferation of undifferentiated granulosa cells (Otsuka et al. 2000). However, mice lacking *Bmp15* gene display minor fertility problems unlike *Gdf9* knock outs (Yan et al. 2001). Communication among granulosa cells and between granulosa cells and oocyte is essential for successful folliculogenesis (Matzuk et al. 2002). Gap junction proteins also known as connexins 43 and 37 (CX43 and CX37), play an important role in the establishing this communication. *Cx43* is expressed in granulosa cells and forms gap junctions between granulosa cells, whereas, CX37, expressed in oocytes is crucial for formation of gap junction communications between oocyte and granulosa cell (Juneja et al. 1999, Gittens, 2003).

Antral development starts with differentiation of granulosa cells into the cumulus cells and formation of antrum (cavity). Germinal vesicle (GV) oocytes, now fully grown, within antral follicles are characterized by reduced or no transcription. Although GV oocytes in antral follicles are meiotically competent, the activation of meiosis occurs in later stages. The progression throughout the antral stages and ovulation is mainly dependent on gonadotropin hormones - FSH and LH (Dorrington and Hofmann 1973, Eppig, 1992). Follicle Stimulating Hormone (FSH) is one of the essential drivers of antral development. FSH induces luteinizing hormone receptor (LHR) expression in cumulus cells, which is required for follicles to respond to luteinizing hormone (LH) (Dorrington and Hofmann 1973, Findlay, 1999). Together they are crucial for activating the ovulatory process. FSH-LH stimulation induces a large-scale chromatin remodeling in GV oocytes (Debey et al. 1993, Sun, 2016). Chromatin in mouse oocytes is initially decondensed in a non-surrounded nucleolus (NSN) configuration, but is subsequently condensed, forming a surrounded nucleolus (SN) configuration with a heterochromatin rim around the nucleolus (Debey et al. 1993, Zuccotti, 1995). Transition from NSN to SN is essential for normal oocyte maturation.

Oocyte maturation

Upon FSH-LH stimulation, oocytes within the follicles are activated, triggering several changes leading to oocyte maturation (Albertini et al. 2001). After the LH surge, expansion of cumulus cells is observed, which is essential for the oocyte to resume meiosis (Fulop et al. 2003). Within the oocyte, LH stimulated meiotic resumption is initiated by a drastic drop in cAMP levels and a limited diffusion of cGMP and/or cAMP from the cumulus cells. Oocyte maturation requires both nuclear and cytoplasmic maturation. Nuclear maturation involves germinal vesicle breakdown (GVBD), extrusion of the first polar body, and resumption of first meiotic division. Whereas, cytoplasmic maturation involves acquisition of the competence to regulate nuclear maturation, fertilization and preimplantation development (Sorensen and Wassarman 1976, Eppig, 1989). Cytoplasmic maturation is reflected by changes in GV chromatin configuration (Bonnet-Garnier et al. 2012, Wang, 2009, De La Fuente, 2001). GV oocytes must shift from a NSN configuration to SN configuration to attain the full competence to complete meiosis

and development upon fertilization (Wang et al. 2009). Simultaneously, a selective degradation and accumulation of transcripts is observed upon/during oocyte maturation in mouse and humans, which is equally important for cytoplasmic maturation. Transcripts and/or proteins associated with meiotic arrest at the germinal vesicle (GV) stage are degraded or have decreased expression (Su et al. 2007). Simultaneously, transcripts required for signaling pathways essential for the regulation of oocyte meiosis and the maintenance of mitotic arrest upon maturation are translated (Su et al. 2007). Therefore, regulation of gene expression during oocyte maturation at transcriptional and post-transcriptional level is a crucial for the oocyte to attain full meiotic and developmental competence to ensure successful oocyte-to-embryo transition.

Regulation of gene expression and mRNA storage in oocytes

During the process of oogenesis, oocytes increase in size – from approximately 20 to 80 μm in mouse and from 35 to 120 μm in human and ~ 100-fold increase in volume (Eppig and O'Brien 1996, Picton, 1998). Fully-grown oocytes have been estimated to contain ~ 0.6 ng of total RNA (almost ~ 200 times the amount of RNA found in a typical somatic cell) (Sternlicht and Schultz 1981, Wassarman, 1992). Approximately 10–15% of all RNA in a fully-grown oocyte comprises of heterogeneous RNA, whereas ~ 65% comprises ribosomal RNA (Wassarman and Kinloch 1992, Abe, 2015). Increase in the RNA content and size is due to active transcription during the growth phase of oogenesis, which coincides with active proliferation of follicular cells (Moore and Lintern-Moore 1978, Bachvarova, 1985). However, by the end of the growth phase, fully grown oocyte enter a transcriptional quiescent state. Factors accumulated during oocyte growth phase are solely responsible for the control of oocyte maturation, fertilization and initiation development (Bachvarova et al. 1985, Su, 2007, De La Fuente, 2004). Accumulation of RNA in fully grown oocyte is subsequently reversed by RNA degradation. The fate of an oocyte RNA depends, to a large part, on its association with proteins that regulate its stability and/or stage-dependent translation (Eichenlaub-Ritter and Peschke 2002). One such mechanisms is selective polyadenylation, which regulates RNA stability and/or translation in the oocyte (Oh et al. 2000).

Regulation by cytoplasmic polyadenylation

Oocytes employ a class of mRNAs called dormant maternal mRNAs that are not translated (or poorly translated) during the growth phase and are recruited on translation machinery following initiation of maturation (Charlesworth et al. 2004). Translational potential of maternal mRNAs correlates with cytoplasmic changes in the length of the poly(A) tail. Increase in translation is associated with poly(A) tail elongation, whereas translational repression correlates with shortening of the poly(A) tail (Richter 1999). Dormant maternal mRNAs in oocyte's cytoplasm undergo deadenylation and are kept translationally silenced until oocyte maturation, where elongation of poly(A) tail by cytoplasmic polyadenylation activates them. Cytoplasmic polyadenylation requires two 3'UTR elements, a uridine rich sequence known as cytoplasmic polyadenylation element (CPE) and a canonical polyadenylation signal- AAUAAA. Three 3' UTR binding proteins are needed- a cytoplasmic polyadenylation element binding protein (CPEB), cleavage and polyadenylation-specificity factor (CPSF) and a poly(A) polymerase (PAP) (Mendez and Richter 2001). CPEB is a highly conserved zinc finger and RNA-recognition motif (RRM) RNA binding protein, which is regulated through phosphorylation. In immature oocytes, mRNAs with CPE are bound by CPEB and an inhibitory protein called Maskin, which interacts simultaneously with both CPEB and eIF4E repressing translation *(Fig 2)* (Stebbins-Boaz et al. 1999). Upon oocyte maturation, CPEB is phosphorylated by EG2, an enzyme activated upon oocyte activation by progesterone (Frank-Vaillant et al. 2000). Phosphorylated CPEB recruits CPSF, which in turn binds poly(A) signal and poly(A) polymerase, extending poly(A) length *.(Fig 2)* (Mendez et al. 2000). At the same time dislocation of Maskin, which also promotes translation, is also observed *(Fig 2)*.

Regulation by small RNAs

Post-transcriptional gene regulation can also be mediated by small RNAs. Small RNAs are 20–30 nt long regulatory RNAs, which are typically involved in silencing target mRNAs. Three major classes of small non-coding RNAs are present in mouse oocytes: endogenous small-interfering RNAs (siRNAs), microRNAs (miRNAs) and Piwi-interacting RNAs (piRNAs) (reviewed in

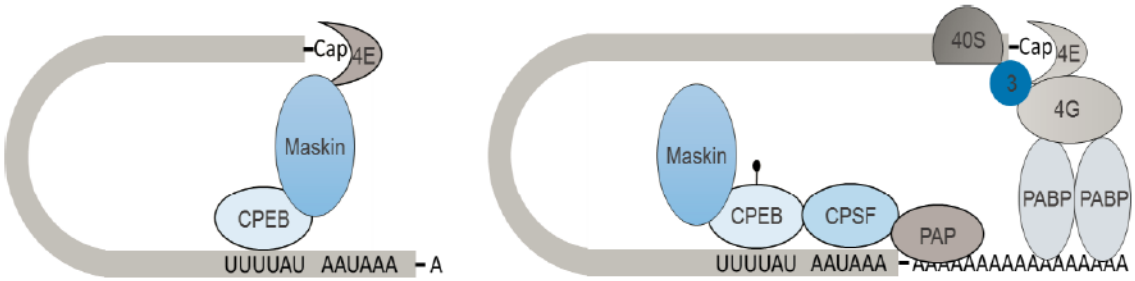


Figure 2: *CPEB mediated translational control. Messenger RNAs containing a CPE are masked and reside in a complex containing CPEB, Maskin and eIF4E. Upon maturation, newly phosphorylated CPEB recruits cleavage and CPSF and PAP, which elongates the poly(A) tail. And dissociating Maskin from eIF4E. (Adapted from Mendez, 2000)*

Svoboda 2017, Suh, 2011). siRNAs are derived from double strand dsRNAs, which are directly cut by RNase II Dicer, into a short 21–23 nt RNAs *(Fig 3)*. miRNAs are transcribed as long primary-miRNAs, which are first cleaved by the microprocessor complex, composed of Drosha, an RNase III enzyme and dimer of RNA-binding protein DGCR8 (DiGeorge syndrome critical region gene 8) into a precursor miRNAs (Han et al. 2004), and later cleaved by Dicer into ~ 22-nt miRNAs *(Fig 3)* (Kim 2005). piRNA biogenesis is Dicer independent and their function in oogenesis is still poorly understood (Suh and Blelloch 2011).

Both miRNAs and siRNAs get incorporated into the RNA-induced silencing complex (RISC) as single stranded small RNAs and repress target mRNAs *(Fig 3)*. The Argonaute 2 (AGO2), a component of RISC possesses endonuclease activity and is responsible for direct endonucleolytic cleavage of target mRNAs (Denli et al. 2004, Gregory, 2004, Song, 2004).

With regards to miRNA function during OET, studies suggest that beside their expression, miRNAs function is observed during early oogenesis and in late pre-implantation embryos but are non-functional in oocytes and during oocyte-to-embryo transition (Suh et al. 2010, Tang, 2007). In contrast, importance of siRNA in oocyte maturation was shown in Dicer null oocytes. Loss of Dicer in oocytes leads to loss of miRNAs, siRNAs and dysregulation of thousands of mRNAs; furthermore, mutant oocytes displayed aberrant spindle organization and chromosomal segregation, resulting in meiotic arrest. On the other hand, *Dgcr8* null oocytes, an essential miRNA biogenesis factor, did not exhibit any effect on mRNA levels or on

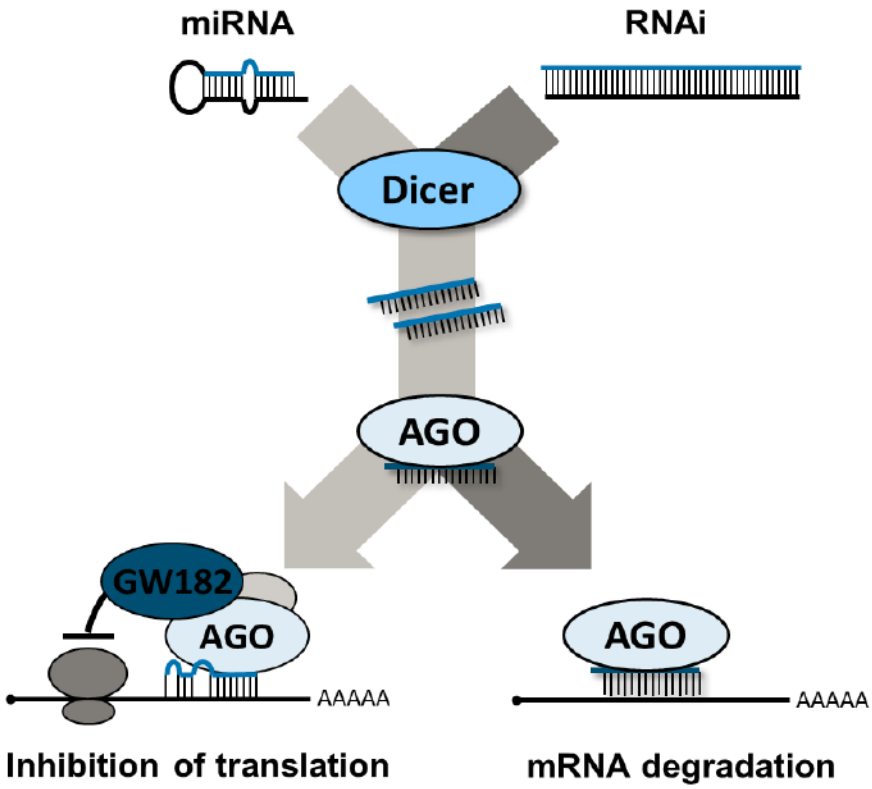


Figure 3: *mRNA regulation by small RNAs. Both miRNAs and siRNAs are processed by dicer into 20-30nt small RNAs, which are bound by Ago2 and together influence gene expression in oocytes.*

oocyte maturation. Therefore, siRNAs mediated mRNA regulation is critical for oocyte maturation and early development and not miRNA mediated regulation (Suh et al. 2010, Tang, 2007, Murchison, 2007).

Oocyte-to-embryo transition

Following fertilization, maternal factors initiate parental gene expression reprograming by maternal mRNA degradation and zygotic genome activation (ZGA), which is essential for a successful oocyte-to-embryo transition. Mammalian ZGA takes place in “waves”, where, in mouse, cohorts of genes are transcribed at specific time point during OET.

The first zygotic transcription begins during S-phase of one-cell stage, this first wave of transcription is also known as the minor ZGA (Aoki et al. 1997, Nothias, 1996, Abe, 2018). The second wave of transcription, also known as major ZGA, occurs at the 2-cell stage and is required for the next cell divisions (Golbus et al.

1973, Hamatani, 2004, Li, 2010). The next phase of transcription initiates at the 4- to 8-cell transition and marks the beginning of dynamic morphological changes that leads to formation of the blastocyst (Hamatani et al. 2004, Jukam, 2017).

At the 8-cell stage, blastomeres increase the surface area of their cell-cell contacts in a process known as compaction. Around this time, the first cell fates are specified into inner cell mass (ICM) and trophectoderm (TE), which give rise to the embryonic and extraembryonic (placental) tissues of the blastocysts, respectively. ZGA in humans differs from mice with respect to timing; major ZGA in humans takes place at 4-to 8-cell stage rather than 2-cell stage (Xue et al. 2013, Yan, 2013).

Two key features distinguish transcription during OET to other cellular transitions. First, during OET, zygotes with a little or no transcription state are taken to a state where up to thousands of genes are transcribed, whereas during other cellular transitions, the global transcription profile remains largely undisturbed. Second, during OET, zygote undergoes divisions with no significant growth (a process known as cleavage), this results in changes in the ratio between maternally deposited RNAs/proteins and genomic content within each blastomere (reviewed in Jukam et al. 2017).

How this unique transition of non-transcribing, differentiated oocyte to an embryo made possible? In John Gurdon's classic experiment, a nucleus of a frog somatic cell transferred into an enucleated frog oocyte “reprograms” the somatic nucleus, in a way that supports the development of an adult frog (Gurdon 1962). Therefore, maternal cytoplasm has a remarkable ability to reset the chromatin state of a terminally differentiated nucleus. Although the importance of maternal contribution in OET is well accepted, the factors involved in are yet to be thoroughly studied.

Long Non-Coding RNAs

LncRNA discovery

Only, ~1.5% of our genome encodes for proteins (Consortium 2012), which has been historically considered to be the functional part of the genome. For the long time, remaining part of the genome (the non-coding region) was considered to be largely "junk". This notion has been debunked thanks to the large scale genome-wide transcriptome and chromatin analysis. While the existence of non-coding RNAs was known for decades; genomic studies opened up a whole new era of non-coding RNAs, where countless studies have emphasized the importance of non-coding RNAs from DNA replication to protein synthesis *(Fig.4)* (reviewed in Jarroux et al. 2017).

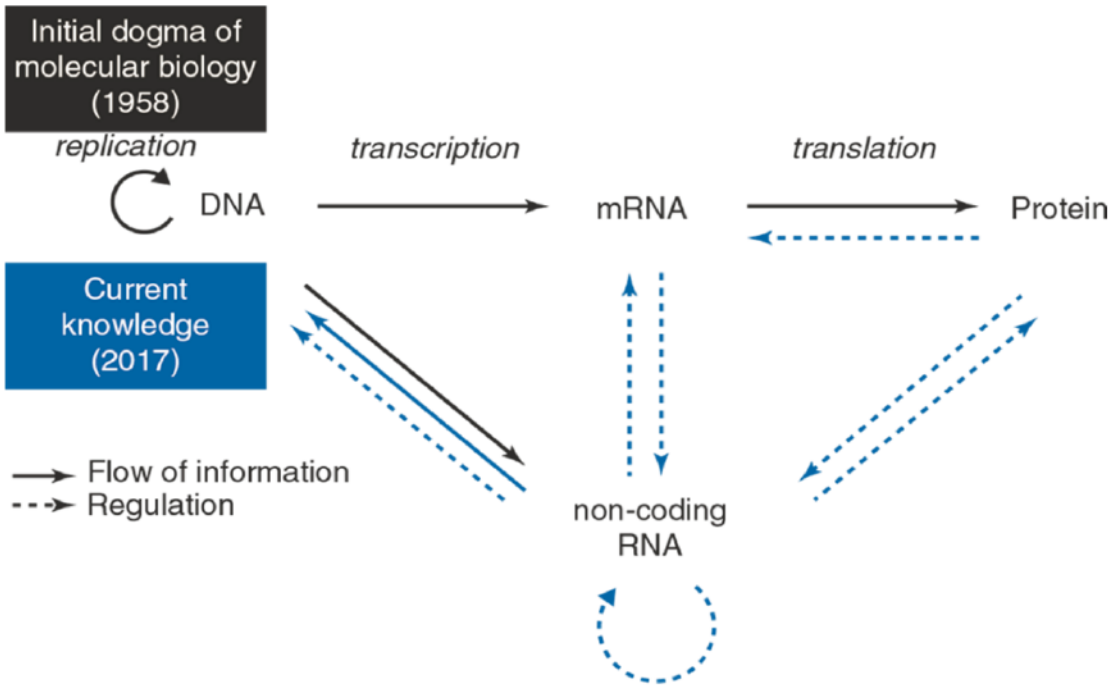


Figure 4: *Modified central dogma emphasizing the role of non-coding RNAs in previously known central dogma. Non-coding RNAs support every stage of gene flow from DNA replication to protein synthesis. Adapted from (Jarroux et al. 2017).*

There are many types of non-coding RNAs; small RNAs (which includes miRNAs, siRNAs, piRNA, described earlier), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), small nucleolar RNA (snoRNA), circular RNAs (circRNAs) and

long non-coding RNAs (lncRNAs). LncRNAs have proved themselves to be one of the most versatile class of non-coding RNAs. A few lncRNAs such as "*H19*" (Bartolomei et al. 1991) and *Xist* (Brown et al. 1991) were already characterized before the genomic era but remained an exception until the early Encode genome project (Consortium 2012). *H19* is the very first lncRNA gene to be identified, which is a part of the imprinted *H19/Igf2* locus. (Bartolomei et al. 1991) At that time *H19* was considered to be a unique RNA as it had all features of an mRNA except for its lack of translation (Brannan et al. 1990). Activation of *H19* expression in transgenic mice was found to be lethal in prenatal embryos, suggesting regulated *H19* expression was required for proper development. LncRNA *H19* remained a controversy until the discovery of lncRNA *Xist*, another non-protein coding gene (Brown et al. 1991). *Xist* was essential for X-chromosome inactivation in humans even though it did not encode a protein. Studies on *H19* and *Xist* changed our view of non-coding genes and on their biological relevance in general. In 2000's, larger-scale genomic studies conducted by the ENCODE project in human and mouse (Lander et al. 2001, Mouse Genome Sequencing, 2002) revealed 54% and 46% of the transcriptome respectively consisted of non-protein coding transcripts. This study was later supported by Guttman *et.al,* where they identified thousands of long non-coding RNAs in four mouse cell types using chromatin-state maps (Guttman et al. 2009). Since then lncRNAs have been identified and studied in several model systems.

LncRNA characteristics and features

What makes lncRNAs special? Let's start with their definition, which is still a controversial topic. Most accepted definition states: LncRNAs are RNAs, which do not encode for proteins and are longer than 200 nt (Moran et al. 2012, Wilusz, 2009, Rinn, 2012, Ulitsky, 2013, Guttman, 2009, Derrien, 2012, Cabili, 2011). Although the upper limit of lncRNAs varies, the lower limit is set to 200 nt. Although, the reason for set limit is arbitrary and subject to discussion. One of the important features of lncRNAs is their lack of ability to encode for a protein, which distinguishes them from mRNAs. Although, lncRNAs may encode small peptides (van Heesch et al. 2014).

In many aspects, lncRNAs resemble mRNAs. LncRNAs are transcribed by RNA polymerase II (Pol II), 5' capped and usually 3' polyadenylated. However there are few exceptions, for example, lncRNA *BC2oo* (Mus et al. 2007) and *asOct4-pg5* (Hawkins and Morris 2010) are non-polyadenylated and transcribed by Polymerase III. Some lncRNAs originate from unusual processing of RNA transcripts. Recent studies have identified circular RNAs (circRNAs), originating from back-spliced exons (Memczak et al. 2013, Liang, 2014) and circular intronic long ncRNAs originating from lariat introns that escape from debranching (Zhang et al. 2013, Sun, 2015).

LncRNAs, like to protein coding genes, are multi-exonic, although, they have fewer exons, with slightly longer exon length (Derrien et al. 2012). LncRNAs often undergo alternate splicing compared to mRNAs, producing tissue specific isoforms (Derrien et al. 2012). LncRNAs also have generally lower expression compared to mRNAs, which is due to lower stability and lower rate of transcription (Derrien et al. 2012). Although, chromatin marks on lncRNAs resemble those of any other Pol II transcribed mRNAs, they have reduced levels of H3K4me3 marks, which explains their lower transcription (Guttman et al. 2009). LncRNAs also display high tissue-specific expression, unlike most of the protein-coding genes.

Different types of classifications exist based on the lncRNA length, location, properties, and function (St Laurent et al. 2015). Classification based on location are broadly divided into intragenic and intergenic lncRNAs (Ulitsky and Bartel 2013, Guttman, 2009, Derrien, 2012). Intragenic lncRNAs are of two types, (1) lncRNAs that overlap protein coding genes in antisense orientation, referred to as antisense lncRNAs, and (2) lncRNAs that are expressed from within an intronic region. Antisense RNAs, or so called natural antisense transcripts (NATs) are further classified as *cis*-NATs and *trans*-NATs (Su et al. 2010). Intergenic lncRNAs are those, which do not overlap with any protein coding gene *(Fig.5).*

LncRNA conservation and evolution

Another recognized feature of lncRNAs is their poor conservation compared of protein coding genes (Ulitsky et al. 2011, Washietl, 2014). LncRNA conservation can be classified into four categories: conservation based on the sequence,

structure, function, and syntenic transcription (Diederichs 2014). First category is based on sequence conservation. Although, there are few examples of sequence conserved lncRNAs in vertebrates, most lncRNAs lack conservation (Ulitsky et al. 2011, Derrien, 2012). LncRNAs are thought to be products of spurious transcription; some are positively selected and retained, while the others are lost during evolution (Necsulea et al. 2014). Some lncRNAs, even though do not share a perfect nucleotide conservation, can retain the secondary structure there by retaining the

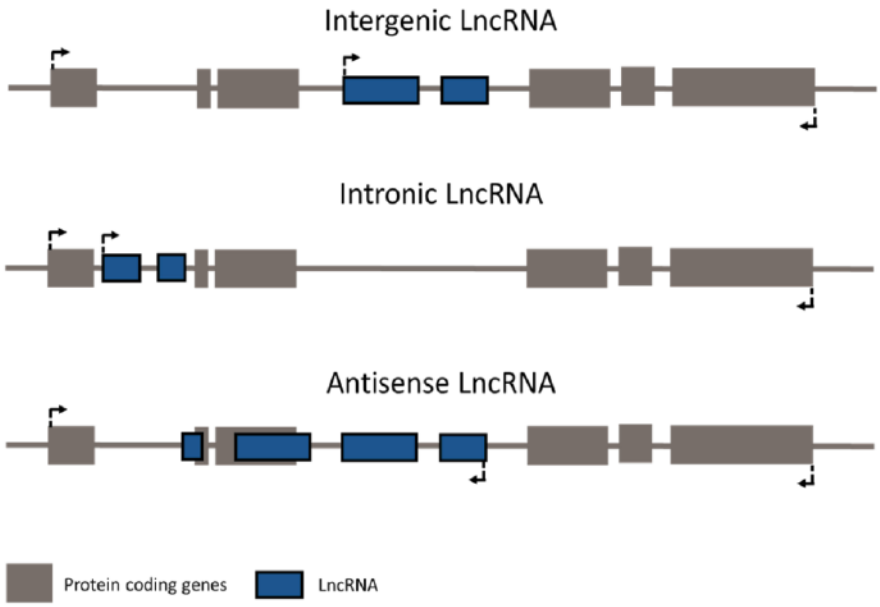


***Figure 5**: LncRNA classification based on their genomic location. Classification of lncRNAs (blue) based on their genomic position in a relation to neighboring protein coding genes (grey).*

function (Derrien et al. 2012). Finally, some lncRNAs share syntenic genomic locus but lack sequence conservation, for example lncRNAs, which can influence neighboring gene expression (Ulitsky et al. 2011).

LncRNA functional characterization

One of the biggest challenges in lncRNA field is to assign functional significance to lncRNAs. As mentioned above, most lncRNAs are in general products of spurious transcription and perhaps do not have any function, while a

small portion of lncRNAs may play important roles. Functionally, lncRNAs can be classified in many ways. Based on their site of action, lncRNAs can be classified as *cis* or *trans* acting lncRNAs. For example, *linc-HOXA1* located 50kb from the *HoxA* gene cluster in mouse ESCs represses *Hoxa1* gene in *cis* by recruiting purine-rich element-binding protein b (Purb) (Maamar et al. 2013). LncRNA *HOTAIR*, expressed within *HoxC* cluster, interacts with polycomb repressive complex 2 (PRC2) and repress *HoxD* gene is *trans* (Rinn et al. 2007).

LncRNAs can be classified into five categories based on their mode of action *(Fig 6)* (Wang and Chang 2011, Guttman, 2012): 1. Guide lncRNAs, 2. LncRNA scaffolds, 3. LncRNA decoys, 4. Allosteric lncRNAs and 5. Small-RNA substrates. One of the most studied group of lncRNAs are guide lncRNAs. Guide lncRNAs interact with chromatin modifying complexes and recruit these complexes on to genes influencing their expression. Guide lncRNAs are found to form a triple helix structure with target genomic regions (Mondal et al. 2015). For example, lncRNA *MEG3* guides PRC2 complex to TGFβ-regulated genes (Mondal et al. 2015). Some lncRNAs facilitate formation of large multiprotein complexes and localize these complexes to specific genomic locations. For example, in the nucleus, lncRNA *HOTAIR* adopts a four-module structure and interacts with PRC2/LSD1/REST complexes to silence specific genes (Tsai et al. 2010). LncRNAs also act as decoys, preventing certain proteins binding their target genes. For example, lncRNA *GAS5* binds glucocorticoid receptors (GR) by mimicking glucocorticoid response elements (GRE) and represses GR-regulated genes (Kino et al. 2010). Recently, some lncRNA have been shown to change activity of target protein allosterically by directly interacting with proteins, for example, binding of *CCND1* lncRNA to TLS protein, can change its activity and activate its target genes by changing the modification states (Wang et al. 2008b). Finally, some lncRNAs host small RNAs or even serve as substrates for small RNAs. *H19* is host to miR-675 (Cai and Cullen 2007); *Gas5* hosts 10 highly conserved snoRNAs (Smith and Steitz 1998); and the imprinted lncRNAs *Gtl2* and *Mirg* hosts almost 50 miRNAs and 40 snoRNAs (da Rocha et al. 2008). Furthermore, lncRNAs carrying pseudogene sequences are proposed to produce endo-siRNAs found in mouse oocytes by complementary

base-pairing between mRNAs and expressed pseudogenes (Tam et al. 2008, Watanabe, 2008).

LncRNAs in disease

Several lncRNAs were shown to be dysregulated in human diseases. For example, *DBE-T* lncRNA is involved in facioscapulohumeral muscular dystrophy (FSHD) (Cabianca et al. 2012), a disease caused due to reduced D4Z4 repeats size. Decreased D4Z4 repeats size activates *DBE-T* lncRNA expression, which in turn deposits activating H3 lysine-36 demethylation (H3K36me2) at neighboring genes, activating genes responsible for FSHD phenotype. FSHD provided the first proof of lncRNA-mediated gene activation in human disease. (Cabianca et al. 2012). Similarly, Angelman Syndrome (AS) is caused by the loss of expression of imprinted *UBE3A* gene. (Meng et al. 2013). In normal conditions, *UBE3A* gene is repressed on the paternal chromosome by antisense lncRNA *UBE3A-ATS*, while the maternal allele is active. Based on mouse studies, mutations in *UBE3A* gene can either activate lncRNA expression on both chromosomes or just distort the protein itself, leading to loss of *UBE3A* protein expression (Meng et al. 2013).

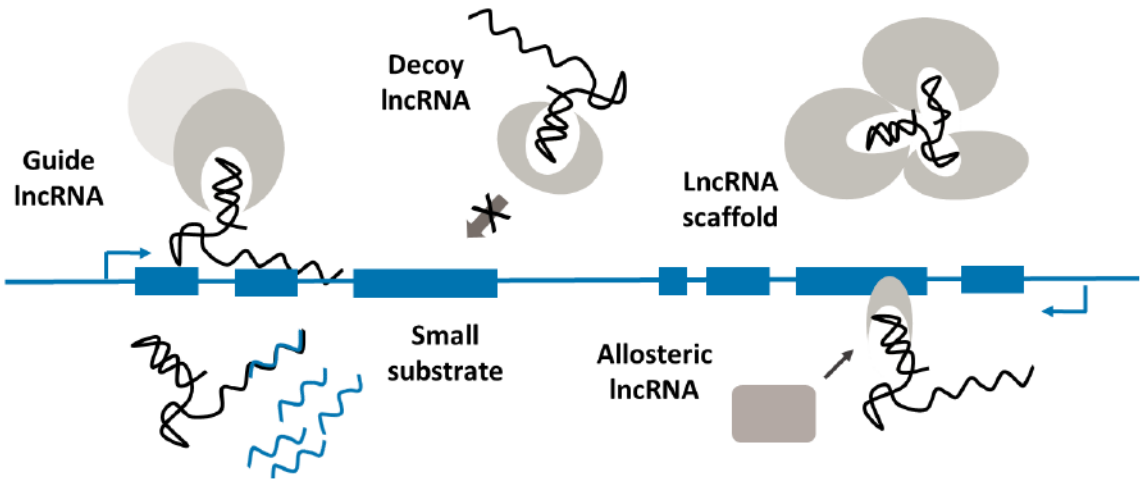


***Figure 6**: Schematic diagram of lncRNA mechanisms of action (left to right): (1) LncRNAs can guide chromatin modifiers and transcription factors to DNA to both repress and activate gene expression; (2) LncRNAs can act as molecular decoys to move proteins away from specific DNA locations; (3) LncRNA can serve as molecular scaffolds to bring proteins into stable complexes that can modulate gene expression; (4) LncRNAs can modify target proteins allosterically and influence their activity and (5) Finally, lncRNAs can also host and/or produce small RNAs.*

LncRNAs have also been shown to have diverse role in lipid metabolism. LncRNAs function in regulation of sterol regulatory element binding protein (SREBP) family of transcription factors (Li et al. 2018, Yan, 2016, Walker, 2011). SREBP family genes are transcription factors, which regulate lipid homeostasis by orchestrating expression of genes required for cholesterol, fatty acid, and phospholipid synthesis and uptake (reviewed in Horton et al. 2002). Two lncRNAs *MALAT 1* and *HP14* were shown to bind SREBP protein and promote lipogenesis (Li et al. 2018, Yan, 2016). Another set of antisense lncRNAs were shown to regulate Apolipoproteins (APOs), which regulate the plasma lipoproteins (reviewed in Mahley et al. 1984) (van Biervliet et al. 1986). Experimental studies on obese mice showed loss of lncRNA *APOA4-AS* in liver reduces levels of plasma cholesterol and triglycerides (Halley et al. 2014). An elegant recent study emphasizes the role of micro peptides expressed by lncRNA in muscle disorders. There are increasing studies emphasizing role of lncRNA in cancer. For example, lncRNA *HOTAIR* promotes metastasis in breast cancer, dysregulation of lncRNA *PCGEM1* and *PCA3* is observed in prostate cancer, where it is being used as a biomarker. LncRNA *Malat1*, one of most highly expressed and studied lncRNA has been implicated to have role in lung cancer and the list goes on (reviewed in Huarte 2015). LncRNAs are also being used as non-invasive biomarkers for early detection of cancer. Recent studies have emphasized the role of several exosomal enriched lncRNAs in tumorigenesis by regulating several factors like angiogenesis, apoptosis and metastasis. Certain exosomal lncRNAs are dysregulated in certain cancer types and can be used as diagnostic biomarkers. For example, lncRNA-*UCA1* enriched exosomes are released in hypoxic bladder cancer cells, which facilitates their growth and development. The detection of lncRNA-*UCA1* level in human serum could be used as diagnostic biomarker for bladder cancer (reviewed in Sun et al. 2018). Taking into account all the studies it is apparent that lncRNAs are emerging new therapeutic targets for the treatment of several diseases.

Retrotransposons

Transposable elements (TEs) make up almost half of the mammalian genome and retrotransposons are predominant among them (Babushok et al. 2007). Based on the mode of transposition, TEs fall into two major classes: Class I includes “copy and paste” retrotransposons and Class II harbors “cut and paste” DNA transposons *(Fig.7)* (reviewed in Havecker et al. 2004). DNA transposons are characterized by their ability to mobilize themselves within a genome with the help of terminal inverted repeats and transposase (TPase), by a cut and paste mechanism. However, mammalian genomes currently do not seem to have any active DNA transposons and their remnants are so-called fossils (Babushok et al. 2007). Class I transposable elements, also known as retrotransposons, transpose through an RNA intermediate, which is reverse transcribed and integrated, hence “copy and paste”. Retrotransposons are divided into four subclasses based on the presence/absence of long terminal repeats (LTRs) at their 5′ and 3′ ends (LTR vs. non-LTR elements) and on their retrotransposition ability (autonomous vs. non-autonomous).

LTR retrotransposons

LTR retrotransposons are of retroviral origin, with their life cycle confined within a host cell (Tarlinton et al. 2006, Kaneko-Ishino, 2012). LTR retrotransposons are further divided into autonomous and non-autonomous based on their ability to autonomously retrotranspose. Autonomous LTR retrotransposons have two long terminal repeats (LTRs) flanking a protein-encoding region. The protein coding region encodes an RNA-dependent DNA polymerase (POL, reverse transcriptase) but often lacks gene encoding envelope protein (ENV) *(Fig 7)* (reviewed in Havecker et al. 2004). In the mouse genome, there is currently only one highly active LTR retrotransposon group (Intracisternal A Particle (IAP)) and several presumable LTR retrotransposon fossils, including Mouse Endogenous Retrovirus type-L (MuERV-L) insertions. MuERV-L are highly transcribed during mouse early development (Svoboda et al. 2004). In humans, there is family of actively retrotransposing Human Endogenous Retroviruses (HERVs) (reviewed in Khodosevich et al. 2002). Autonomous LTR-retrotransposons usually reach

hundreds to several thousands of insertions before they die out due to accumulation of mutations abolishing their coding capacity and *trans* complementation (Sanchez et al. 2017). Non-autonomous LTR elements possess flanking LTRs but do not contain open-reading frames encoding factors that would mobilize them; their retrotransposition is dependent on autonomous LTRs *(Fig.7)*. They are significantly smaller than autonomous LTR elements, ranging usually between 1-1.5 kb. An example of non-autonomous LTR elements are Mammalian apparent LTR Retrotransposons (MaLR) (reviewed in Smit 1993).

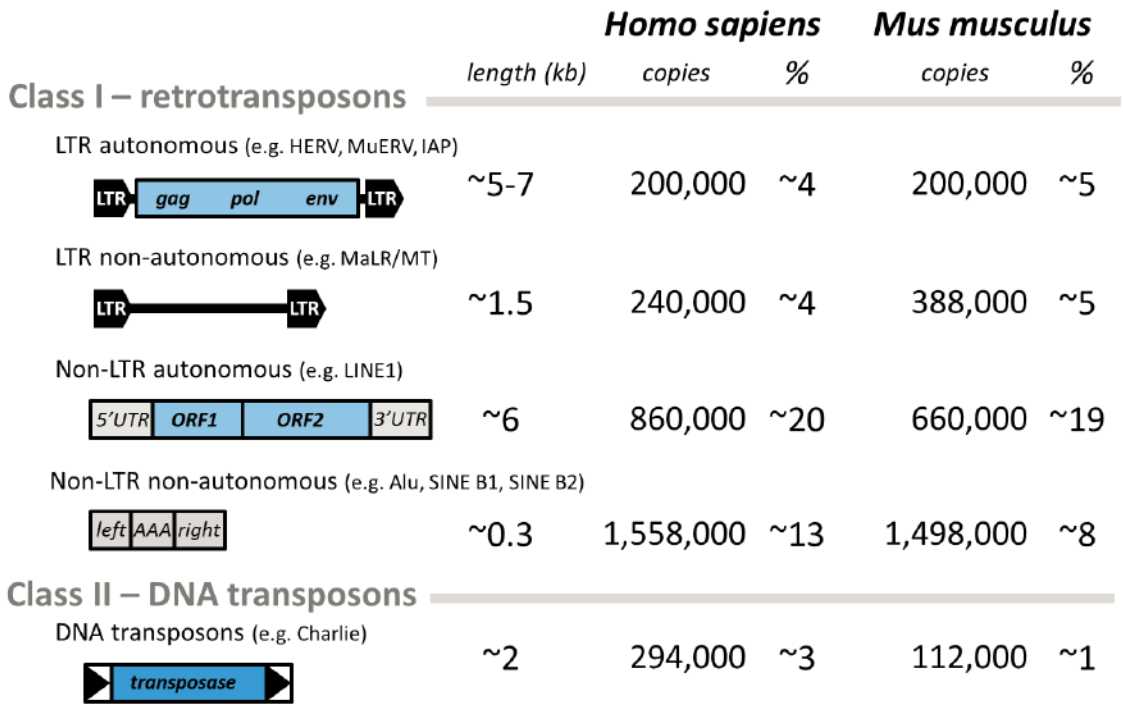


Figure 7: *Overview of two major classes of transposable elements. Copy numbers of transposable elements per haploid genome (copies) and fraction of the genome occupied by each transposable element type (%) were obtained from the literature (Mouse Genome Sequencing et al. 2002, Lander, 2001).*

Non-LTR Retrotransposons

Non-LTR elements are represented by long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) *(Fig.7)*. Non-LTR retrotransposons are also divided into autonomous and non-autonomous types. The autonomous long interspersed nuclear elements (LINEs) are among the most abundant retrotransposons in mammalian genomes (20% in human genome and

19% mouse genome) (Mouse Genome Sequencing et al. 2002, Lander, 2001). LINE elements are 6–7 kb long and carry two open reading frames. Most of the genome insertions, however are truncated at the 5′ end (Beck et al. 2011). LINE elements have a strong *cis*-preference for the retrotransposition machinery; proteins translated from a LINE RNA preferentially associate with and retrotranspose the same RNA molecule, from which they were translated (Wei et al. 2001). Non-autonomous short interspersed nuclear elements (SINEs) are relatively short sequences (<0.5 kb) related to RNA Polymerase III (Pol III)-transcribed small RNAs and do not encode proteins (Singer 1982). The most studied mammalian SINEs are human Alu elements derived from the small cytoplasmic 7SL RNA and are the most abundant transposable elements in the human genome (~1 million insertions) (Lander et al. 2001).

Retrotransposon in gene evolution

Although mutations caused by retrotransposon insertions can be harmful to the host genome (Hancks and Kazazian 2012), they are also proposed to be one of the major contributors to genome plasticity and evolution. LTR retrotransposons provide distinct paths for genome remodeling by providing promoters, enhancers, splice sites, or polyadenylation sites *(Fig. 8)* (Babushok et al. 2007, Deininger, 2003, Faulkner, 2009, Fort, 2014, Hasler, 2006, Peaston, 2004, Kapusta, 2013). Estimated 16% of eutherian-specific conserved non-coding elements are derived from mobile elements, implicating their major contribution to mammalian evolution (Feschotte 2008). Furthermore, retrotransposons are evolutionarily among the most lineage-specific sequence elements, especially in mammals (Marino-Ramirez et al. 2005, Ganesh, 2016).

Retrotransposon mediated gene remodeling in oocytes and zygotes

Peatson *et al* revealed a high contribution of retrotransposons to the maternal and pre-implantation embryo transcriptome (Peaston et al. 2004). Oocytes and early embryos provide an environment suitable for transcription of the MaLR and ERV-L class LTR retrotransposons (Peaston et al. 2004, Evsikov, 2004). These LTR retrotransposons provide alternative promoters to several maternal and zygotic genes, for example, *Zbed3*, the most abundant maternal gene is transcribed

from the mouse transcript (MT) MaLR retrotransposon family (Peaston et al. 2004). A recent study from our lab, partially represented here, provides new insight into mechanism behind retrotransposon mediated gene remodeling in oocytes and zygotes (Franke et al. 2017). We observed a high number of insertions of MaLR retrotransposons in protein-coding genes as well as non-coding genes (refer to results and discussion for more information).

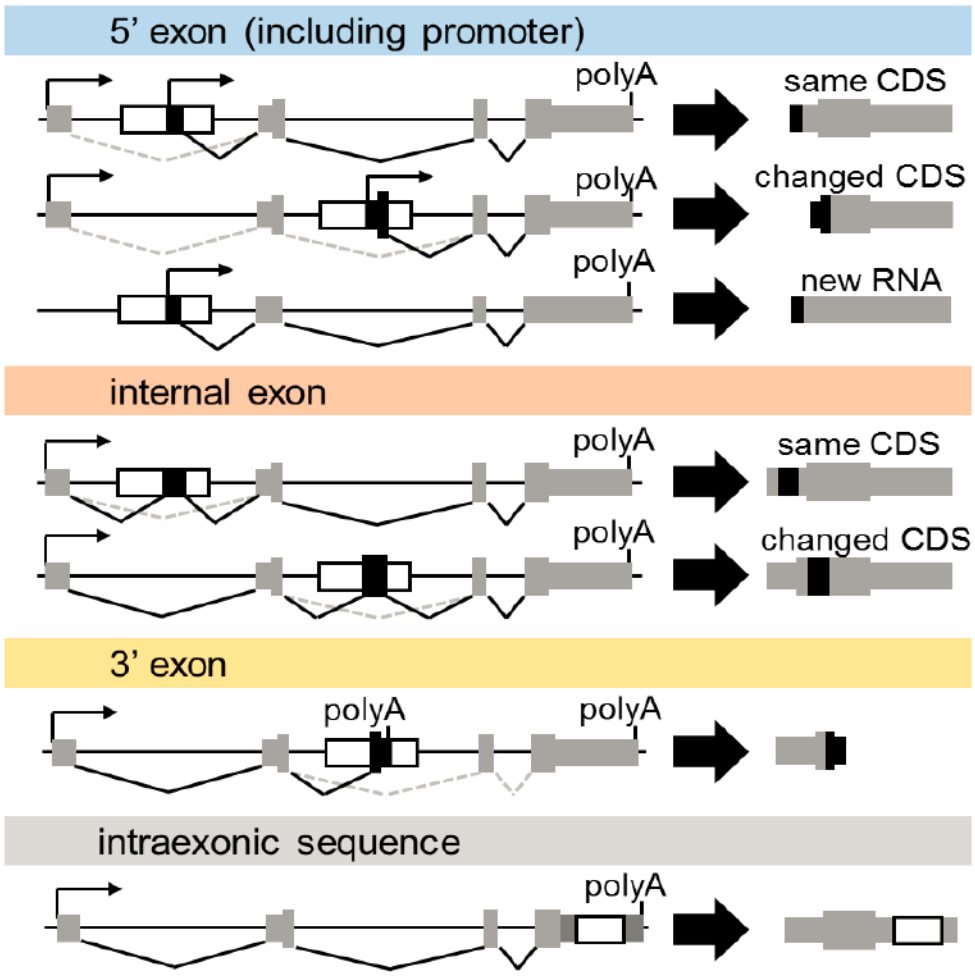


Figure 8: *Schematic representation of different categories by which retrotransposons may contributes to gene remodeling. (A).They may contribute to 5' end of the genes either as a promoter and/or first exon. (B).Retrotransposons may contribute to internal sequence of a gene. (C)& (D) or retrotransposon sequence may contribute to 3' UTRs by providing polyadenylation signals.*

Aim of the study

There is a plethora of literature describing physiological significance of long non-coding RNAs in various tissues and model organisms. However, the expression and role of lncRNAs in mammalian oocyte development and oocyte-to-embryo transition (OET) is not yet well studied. Although there have been few preliminary efforts to annotate and study lncRNAs in preimplantation embryos, there still is a requirement for a comprehensive study focused on maternal and zygotic lncRNAs and their role in OET. At the beginning of the study, nothing was known about lncRNAs in oocytes and zygotes. Aim of this study was to provide a comprehensive analysis of lncRNAs in mouse oocytes and during OET. Project was further divided into four specific aims:

1. Identification and annotation of OET lncRNAs: Design bioinformatics pipeline to identify and annotate lncRNAs expressed in mouse oocytes and embryos from total non-amplified RNA-Seq data.

2. Characterization of OET lncRNAs*:* Characterize OET lncRNAs and compare with lncRNAs expressed in somatic tissues to reveal, if any, difference in the structure and expression pattern.

3. Evolution of OET lncRNAs: Study the evolution of OET lncRNAs, especially their association with retrotransposons.

4. Functional analysis of OET lncRNAs*:* And finally and most importantly, examine functional significance of OET lncRNAs in oocyte development and in oocyte-to-embryo transition, using CRISPR-cas9 mediated loss of function mouse models.

Materials and Methods

Oocyte and Embryo Collection

Oocytes and early embryos were obtained from superovulated C57Bl6/J or C57BL/6NCrl mice as described previously (Nagy 2003). Resumption of meiosis during culture of germinal vesicle (GV) oocytes was prevented with 0.2 mM 3-isobutyl-1-methyl-xanthine (IBMX; Sigma). Hamster and rat full-grown GV oocytes were collected as mouse oocytes without superovulation. Animal experiments were approved by the Institutional Animal Use and Care Committees and were carried out in accordance with the European Union regulations.

RNA Isolation, Reverse Transcription, PCR, and Real-Time PCR

Total RNA from mouse oocytes and embryos (20-30 oocytes/embryo) were released by incubating oocytes in water with RNase inhibitor for 5 min at 85°C prior to reverse transcription. Reverse transcription was performed using Premium RevertAid First Strand cDNA Synthesis Kit (Fermentas). Quantitative PCR was performed using Maxima SYBR Green qPCR Master Mix (Fermentas) on LC480 (Roche) systems. Total RNA from mouse tissues and cultured cells was isolated using RNeasy mini kit (Qiagen) according to the manufacturer's instructions. One microgram of total RNA was reverse transcribed using RevertAid First Strand cDNA Synthesis Kit (Fermentas) and subjected to qPCR using Maxima SYBR Green qPCR Master Mix (Fermentas) on LC480 (Roche) systems. qPCR data was normalized to Hprt1 and Alas expression, by the ΔΔCt approach using an in-house software *(List of primers in Table 1).*

Next Generation Sequencing

For earlier sequencing data, total RNA was extracted from 3000 fully-grown germinal vesicle (GV)-intact oocytes obtained from C57BL6/J mice, respectively, using Isogen (Nippon Gene, Tokyo, Japan), according to the manufacturer's instructions. Poly(A) RNA was isolated by using mRNA purification kit (Invitrogen, Carlsbad, CA; cat# 610.06). High throughput sequencing of size-selected RNA (>200 nt) was performed using Genome Analyzer IIx (Illumina) and 76-nt paired-end-sequencing reads as described previously (Abe et al. 2015). The complete set

of NGS data is available in the Array Express database under accession IDs E-MTAB-2950 and E-MTAB-4775.

For RNA-Seq libraries form lncRNA knock-outs, total RNA was extracted from 25 oocytes using PicoPure RNA Isolation Kit with on-column genomic DNA digestion according to the manufacturer's instruction (Thermo Fisher Scientific). Each sample was spiked in with 0.2 pg synthesized *Renilla* luciferase mRNA before extraction as a normalization control. RNA-Seq libraries were constructed using the Ovation RNA-Seq system V2 (NuGEN) followed by Ovation Ultralow Library system (DR Multiplex System, NuGEN). RNA-Seq libraries were pooled and sequenced by 125 bp paired-end reading using Illumina HiSeq.

Production of LncRNA Knock-Out Models

LncRNA deletion mutant models were produced in the Transgenic Unit of the Institute of Molecular Genetics ASCR, Czech Centre for Phenogenomics using Cas9-mediated deletion of lncRNA promoters (Cong et al. 2013). All animal experiments were approved by the Institutional Animal Use and Care Committees (project number 58-2015) and were carried out in accordance with the law. Sequences of guide RNAs are listed in the *Table 1*. To produce guide RNAs, synthetic 128 nt guide RNA templates including T7 promoter, 18nt sgRNA and tracrRNA sequences were amplified using T7 and TracrRNA primers. Guide RNAs were produced *in vitro* using the Ambion mMESSAGE mMACHINE T7 Transcription Kit, and purified using the mirPremier™ microRNA Isolation Kit (Sigma). The Cas9 mRNA was synthesized from pSp Cas9-puro plasmid using Ambion mMESSAGE mMACHINE T7 Transcription Kit, and purified using the Qiagen RNasy mini kit. A sample for microinjection was prepared by mixing two guide RNAs in ultra-pure water at a concentration of 25ng/μl for each one together with Cas9 RNA (100 ng/μl). Five picoliters of the microinjection mixture were injected into male pronuclei of C57Bl/6 zygotes and transferred into pseudo pregnant recipient mice. PCR genotyping was performed on tail biopsies from four weeks-old animals. *(Primers are listed in Table 1)*

Luciferase Assay

Oocytes (5 oocytes per sample) were lysed in 5 µl of 1× Passive Lysis Buffer (Promega) and the lysate was transferred into 96-well plates (Fisher Scientific). Luciferase activity was measured on the Modulus Microplate Reader (Turner Biosystems) luminometer using Dual Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol. The obtained *Renilla* luciferase signal (experimental reporter) was normalized to firefly luciferase (reference reporter).

Cell Culture and Transfection

Mouse fibroblasts NIH3T3 and adenocarcinoma HeLa cells were cultured at 37°C in 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (Sigma), and penicillin/streptomycin (100 units/ml; Invitrogen). Lipofectamine 3000 Reagent (Invitrogen) was used for cell transfection using manufacturer's instructions.

Results

Part 1: Identification and Annotation of OET lncRNAs

Over the past decade, advances in high throughput sequencing have revolutionized our understanding of lncRNAs; form identifying novel lncRNAs to deciphering their possible functions. As described in the aim of this study, the first step towards studying role of lncRNAs in OET, requires their identification and annotation. At the beginning of this project, three RNA-Seq based studies provided initial but incomplete insights into some aspects of lncRNA biology during OET in mammals. In single-cell RNA profiling data of human preimplantation embryos, Yan *et al*. (Yan et al. 2013) identified 2,733 novel expressed lncRNAs. In 2014, Zhang *et al* (Zhang et al. 2014) used single-cell SOLiD NGS data from mouse OET stages and reported 5,563 novel lncRNAs. However, this lncRNA assembly is of low quality and does not represent a reliable OET lncRNA dataset. Finally, Hamazaki *et al*. (Hamazaki et al. 2015) focused their study on a specific class of lncRNAs termed promoter associated non-coding RNAs (pancRNAs) in ovulated oocytes and two-cell zygotes. All the above mentioned studies accompany a large portion of artifacts and unreliable lncRNA annotation. Therefore, this study identified and annotated a highly reliable set of *de novo* assembled lncRNAs expressed during oocyte-to-embryo transition (referred to as OET lncRNAs hereafter) and performed its characterization in terms of structure and expression (Karlic et al. 2017).

Mapping and transcript assembly

RNA-Seq data used in this study were based on a previously published work from a collaborating lab of Fugaku Aoki (except for the fully grown GV-intact oocytes data) (Abe et al. 2015). LncRNA transcript model assembly and annotation pipeline were designed exclusively for this study. 76-nt paired end (76PE) non-directional RNA-Seq data were obtained from total non-amplified RNA from oocytes and early embryos. RNA-Seq data have depths of 33–58 × 10^6 (~4-9 x 10^6 mapped non-rRNA reads) sequence reads per sample. Due to the lower depth, only highly expressed lncRNAs (>1 FPKM –FPKM stands for Fragments Per Kilobase of transcript per

Million mapped reads) were identified and annotated. This would lower possibility of artifacts.

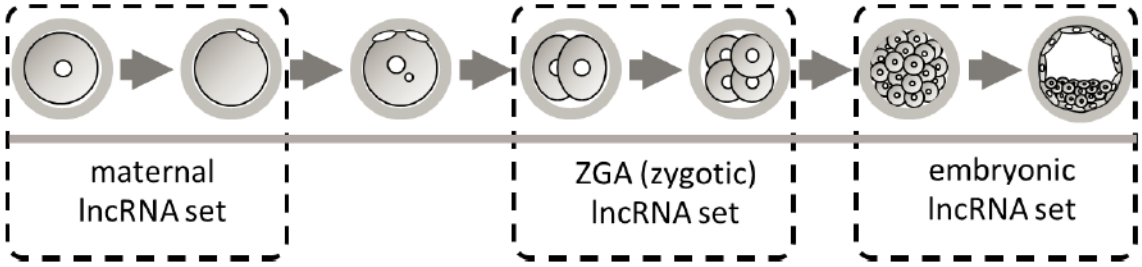


***Figure 9**: Overview of RNA-Seq data sets used for lncRNA annotation and their grouping for assembling transcript models.*

Total RNA-Seq samples were organized into three sets (*Fig 9*).:

(I) the **maternal** set with fully-grown GV oocyte and MII egg RNA-Seq data.

(II) the ZGA or **zygotic** set with 2-cell and 4-cell RNA-Seq data.

(III) the late preimplantation (**embryonic)** set with morula and blastocyst stage data.

Transcript model assembly was performed for each set separately to reduce artifacts from degraded maternal RNAs and to achieve accurate assembly of differentially expressed overlapping sense and antisense transcripts.

To build transcript models from RNA-Seq data, “Scripture” was used as a transcript assembly tool (Guttman et al. 2010). Scripture performed better than Cufflinks or String tie on a diagnostic set of 20 lncRNAs. Transcript models generated by Scripture were further refined to eliminate artifacts of mapping and transcript assembly. Introns with length <20 nt and transcript models containing introns >250 kb, which were typically repeat-derived artifacts, were removed. We also removed all single-exon transcripts to eliminate any artifacts derived from repeat derived locus and also due to their unpredictable orientation. Only spliced transcripts were annotated as introns within transcripts could be used to reveal their directionality based on the presence of GU/AG splice sites. And of course, transcripts shorter than 200 nt were eliminated from this annotation (*Fig 10*).

Refined transcript models from each developmental set were then clustered to eliminate redundancy in annotation. A cluster encompasses all transcript models with the same orientation, which share at least one splice site. Transcript models in

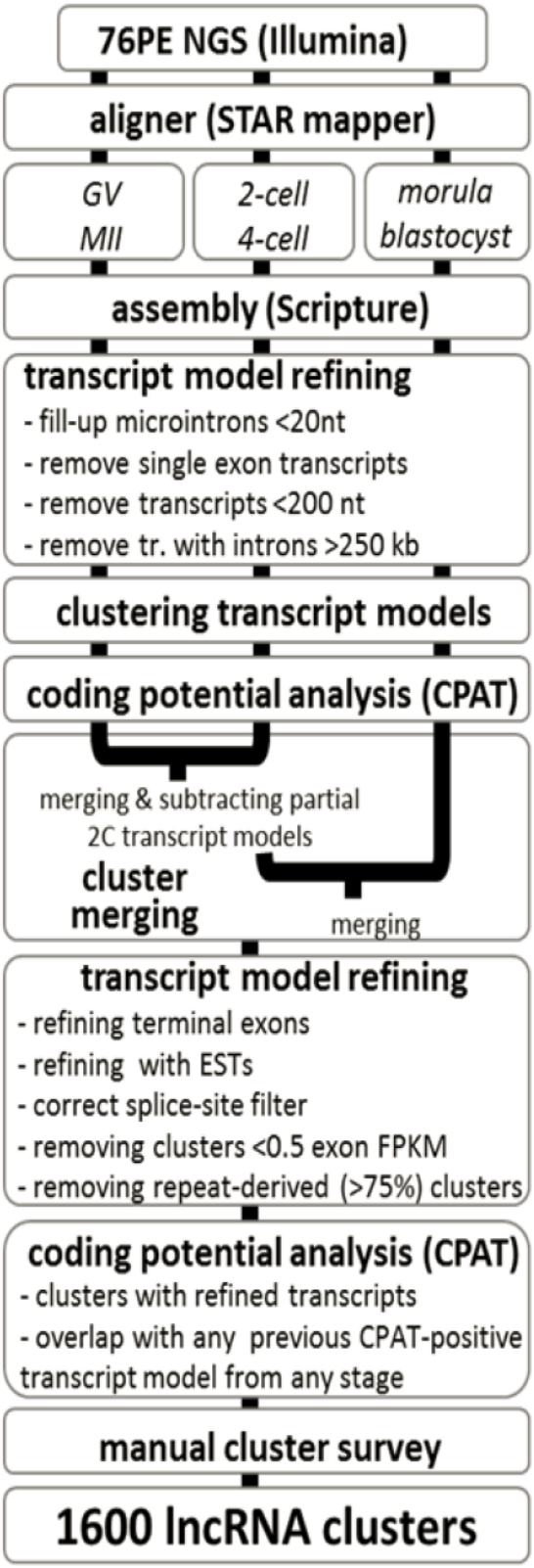


***Figure 10**: Workflow of OET lncRNA identification and clustering from oocytes and early embryos RNA-Seq data.*

clusters were further filtered for coding potential by using CPAT (Wang et al. 2013). CPAT is a Coding-Potential Assessment Tool, which filters out transcripts with coding potential or a potential open reading frame. Clusters containing any transcript model with coding potential were removed, regardless of how many transcript models were CPAT-negative. This step eliminated false lncRNA clusters made of partially assembled transcripts of protein coding genes, although, several loci producing both, a protein-coding transcript and a lncRNA were observed. Next, we merged clusters containing the same transcript models from the three RNA-Seq sets, to eliminate any transcript model redundancy and as an extra step to ensure reliable transcript assembly (*Fig 10*). Transcript models in each cluster were refined further by revising terminal exon predictions, since Scripture tend to produce truncated terminal exon variants. At the end, ~1,200 clusters were manually curated, with a particular focus on low RPM (reads per million) clusters. Manual refining helped to eliminate artifacts neglected by bioinformatic refining. During manual annotation, a lncRNA expression was visualized in UCSC browser along with the transcript model

and examined for their over-lap with expression data of the transcript model. At the end, a highly reliable set of **1,600 non-redundant lncRNA clusters** were obtained (*Fig 10*) referred to as OET lncRNAs hereafter. These lncRNA clusters are used for further analysis. .

OET lncRNA classification

The OET lncRNA clusters were classified into four categories based on their genomic location relative to protein-coding genes. Clusters with no overlap to protein coding genes were classified as **intergenic lncRNAs** (1,142 intergenic lncRNA clusters). Clusters with an overlap with protein coding genes, were further divided based on their orientation and location. LncRNA clusters overlapping intronic regions of protein coding genes, were classified as **intronic lncRNAs** (193 intronic lncRNA clusters) and those overlapping protein coding genes in antisense were classified as **antisense lncRNAs** (235 antisense lncRNA clusters). Finally, we found 30 lncRNA clusters, which overlapped with more than one protein coding gene, which were classified as **complex**. Clusters, which overlapped protein coding genes in sense were omitted from further analysis, as we could not rule out their coding potential (*Fig 11A*).Among annotated lncRNA clusters, 973 clusters were assembled exclusively from the maternal set, while 197 clusters overlapped with other two sample sets. Around 600 clusters were assembled from the zygotic set (362 clusters exclusively from the zygotic set), and only 93 lncRNA clusters were assembled from the embryonic set (only 24 of those clusters were exclusively from the embryonic set) (*Fig 11C*). Low abundance of lncRNAs from embryonic and extraembryonic lineages in morulae and blastocysts could be due to the presence of surrounding non-expressing cells (discussed in discussion). Notably, majority of the 1,600 lncRNA clusters annotated were novel. Around 1200 lncRNA clusters were annotated just in OET stages, while as few as 350 lncRNAs clusters were previously annotated in other tissues (*Fig 11B*).

Characteristic feature of OET lncRNAs

OET lncRNA loci were randomly distributed across the genome *(Fig.12A).* The highest density of lncRNA loci was observed on chromosome 10, while the lowest density was on chromosome 17, which harbored just 17 lncRNAs *(Fig.12A).*

As mentioned in the introduction, lncRNAs have several characteristic features, which distinguish them from mRNAs, like low expression, shorter length/less exons, tissue specific conservation and low conservation. OET lncRNAs are no different in this respect from previously studied lncRNA population in other tissues. When compared with maternal mRNAs, lncRNA loci were shorter and produced shorter transcripts *(Fig.12.B,C,D)*. This difference could be due to a higher number of exons in mRNAs compared to lncRNAs. OET lncRNAs had markedly shorter 5′ exons compared to mRNAs while lengths of internal and 3′ exons of mRNAs and lncRNAs were pretty similar *(Fig.12E)*. In several cases, shorter 5′ exons came from long

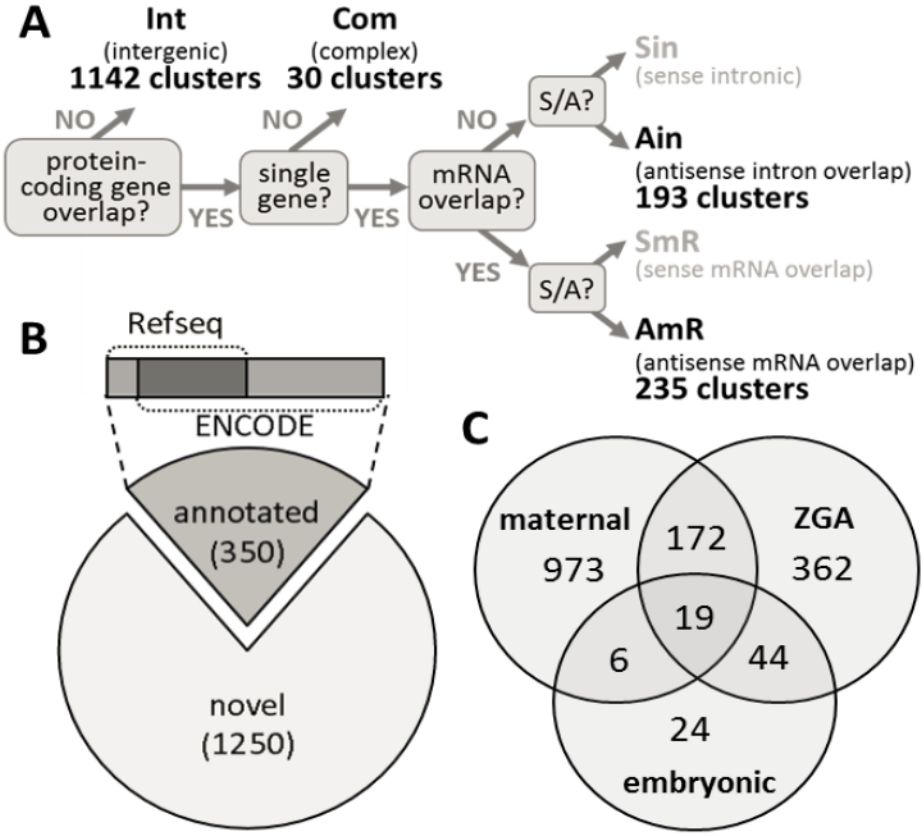


***Figure 11**: OET lncRNA classification (A) Classification system used for annotating lncRNA transcript models and clusters. (B) Novel and annotated lncRNA loci identified in this study. Exons from transcript models from each clusters were compared with data in ENCODE and Refseq databases. Clusters, in which none of the exon-exon junctions from transcript models matched an exon-exon junction annotated in these databases were considered novel. (C) Origin of transcript models in lncRNA clusters. The Venn diagram depicts data set to which lncRNA clusters belong. For example, 19 lncRNA clusters contain transcript models assembled in all three developmental sets while 973 clusters comprise of exclusively oocyte-derived transcript models.*

terminal repeat (LTR) retrotransposons, which gave rise to one third of the 5′ exons. SINE and LINE elements contributed to mature lncRNA sequences more often than to lncRNA promoters and transcription start sites. Detailed results on retrotransposon associated lncRNAs will be described later.

Most of the 1,600 OET lncRNAs were more than an order of magnitude less expressed than the 1,600 most expressed OET mRNAs *(Fig.13),* which is consistent with low lncRNA expression reported in several studies before (Guttman et al. 2010, Guttman, 2009, Derrien, 2012, Ulitsky, 2011). Lower lncRNA expression could reflect a minimal selective pressure compared to high expression levels of evolving lncRNAs lacking a function. However, this comparison only demonstrates

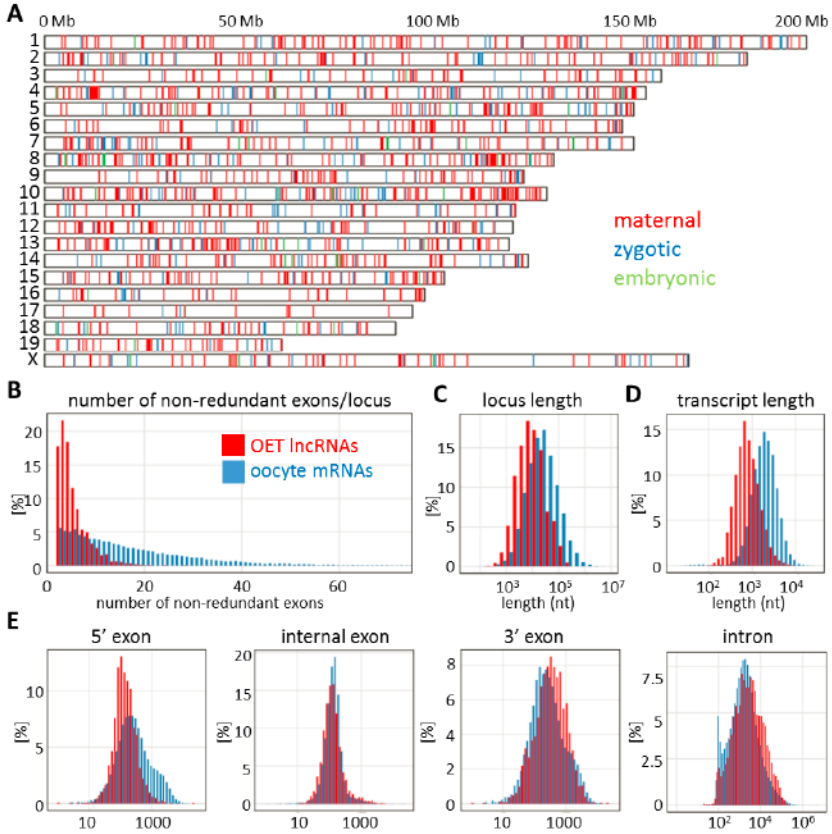


***Figure 12**: Structural features of OET lncRNAs. (A) Genomic distribution of 1,600 OET lncRNA loci across the mouse genome. The color-coding indicates the highest expression (maternal, GV oocyte or MII egg; zygotic, two- or four-cell stage; late preimplantation, morula or blastocyst). (B) Number of exons in OET lncRNAs; (C) OET lncRNA locus lengths; (D) Median transcript length produced from an OET lncRNA locus; (D) Length distribution of OET lncRNA exons and introns; (F) Distribution of LTR-derived first exon sequences. (B–F) All features of OET lncRNAs (depicted in red) are compared with oocyte mRNA data (depicted in blue).*

the expression difference between lncRNAs and most expressed maternal mRNAs while it does not provide an accurate comparison of expression of mRNAs and lncRNAs. Because most expressed maternal mRNAs may carry specific adaptations associated with their high expression Thus, expression levels of 1600 OET lncRNAs were compared with random selection of 1600 OET mRNAs. The idea was that, OET lncRNAs and mRNAs are pol II transcripts whose expression is regulated by the same set of transcription factors. Thus, if there would be no difference in control of expression of lncRNAs and mRNAs, a random set of 1600 mRNA should have expression levels very similar to those of 1600 OET lncRNAs. To make the analysis more robust, the expression of 1600 OET lncRNAs was compared with one thousand random selections of 1600 OET mRNAs. The obtained distribution of mRNA expression levels indicated that on average mRNAs generally retain a higher expression than lncRNAs *(Fig.13)*. However, the same analysis also revealed that within the least expressed quartile of lncRNAs and mRNAs, expression levels of both types of RNAs are essentially the same (Kumar and Hedges 1998) *(Fig.13)*. Whether the differences in expression levels between lncRNAs and mRNAs, reflect this lncRNA-specific feature in transcriptional or post-transcriptional regulations remains unclear. Most probable explanation for lower expression of lncRNAs could be lack of functionality, where non-functional lncRNAs would not be maintained and their expression would presumably decline over time due to mutations affecting transcriptional control elements.

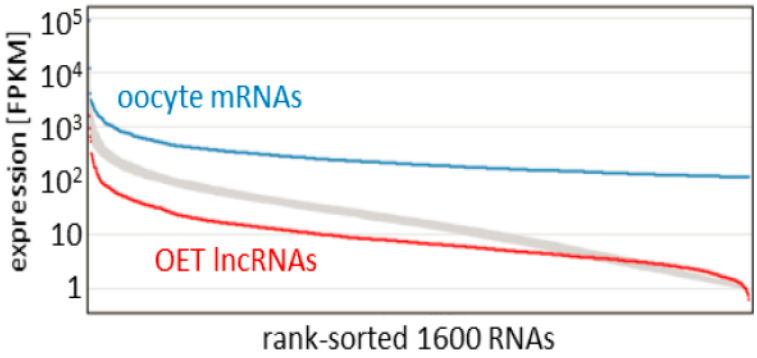


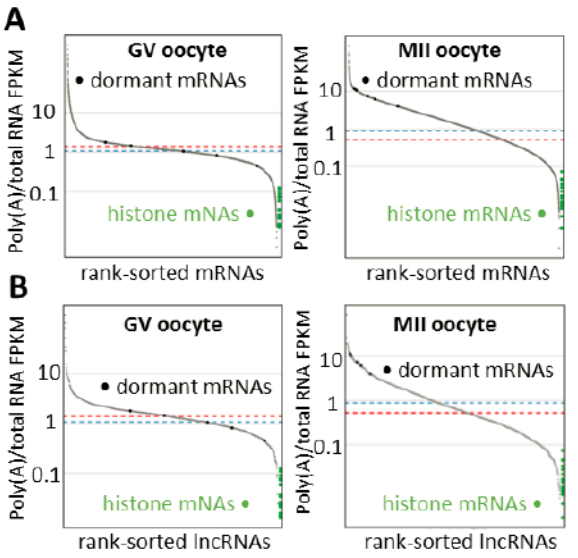
Figure13: *Comparison of OET lncRNA expression levels with OET mRNAs. The Y axis shows log10 FPKM expression, the Y axis are rank-sorted RNAs as follows: blue, 1,600 most expressed maternal RNAs; red, 1,600 OET lncRNAs. The broad grey curve represents values for thousand random selections of 1,600 RNAs*

Polyadenylation status of OET lncRNAs

Biogenesis of lncRNAs and mRNAs is common in most aspects, both type of RNAs are spliced, pol II transcripts whose transcription would utilize essentially the same set of transcription factors. One of the features, by which lncRNAs could differ, is in their 3′ end processing. Therefore, total RNA and poly(A) RNA-Seq expression data was used to study polyadenylation status of lncRNAs. Ratio of poly(A) FPKM/total RNA FPKM was used, which, for simplicity, is referred to as poly(A) score. GV oocytes and MII eggs are an excellent model system for testing this idea because of two possible internal controls: (i) Replication dependent histone mRNAs carrying unique stem loop structures at 3′ ends instead of poly(A) tails (Marzluff et al. 2002). Presence of these transcripts within poly(A)-selected RNA would therefore indicate the extent of contamination with non-polyadenylated mRNA. (ii) Dormant maternal mRNAs (Richter 1999). Translationally repressed mRNAs with short poly(A) tails stored in the GV oocytes, which are readenylated and translated during meiotic maturation. Thus, 20 highly expressed replication-dependent histone genes lacking alternative polyadenylated transcript isoforms were selected and five dormant maternal mRNAs were selected, for which the cytoplasmic polyadenylation during meiosis was previously demonstrated: *Mos, Plat, Cyclin B1, Orc6l,* and *Dcp1a*.

The distribution of the poly(A) score for mRNAs and lncRNAs from GV and MII stages generated sigmoidal curves with slightly different slopes *(Fig.14A,B)*. Although, the difference in slopes might reflect intrinsic differences of GV and MII transcriptomes, poly(A) scores accurately reflected the lack of polyadenylation of histone mRNAs and cytoplasmic polyadenylation of dormant maternal mRNAs during meiotic maturation *(Fig.14A,B)*. The dynamics of poly(A) scores of dormant maternal mRNAs raised a question, whether similar behavior could also be found among maternal lncRNAs. Remarkably, 91 maternal lncRNAs with expression >1 FPKM whose poly(A) scores increased more than 5-fold during meiosis were identified. Similar to dormant maternal mRNAs, these 91 maternal lncRNAs carried putative cytoplasmic polyadenylation elements (CPEs) and at least 14 lncRNAs carried a combination of a canonical poly(A) signal and a CPE-like motif at their 3′ ends *(Fig.14C)*. LncRNAs resembling dormant maternal RNAs are remarkable

because they suggest that cytoplasmic polyadenylation and dormancy could play a more general role in RNA regulation, i.e. a role that goes beyond translational control of maternal mRNAs. We hypothesize that controlled cytoplasmic polyadenylation during transcriptional quiescence could 'activate' stored maternal lncRNAs. Thus, putative dormant maternal lncRNAs represent excellent candidates for further functional studies of lncRNAs functioning between ovulation and ZGA. (For experimental validation refer to functional analysis chapter)



C

Gene	Score	CPE-like sequence
lncRNA-OET-07-157	36.9	5'-UUUUAAU-117nt-AAUAAA-3'
lncRNA-OET-14-235	16.1	5'-UUUUAU-12nt-AAUAAA-3'
lncRNA-OET-16-037	13.4	5'-UUUUUAU-32nt-AAUAAA-3'
lncRNA-OET-03-030	10.7	5'-UUUUAAU-16nt-UUUUAU-4nt-AAUAAA-3'
lncRNA-OET-14-153	9.2	5'-UUUUAAU-70nt-AAUAAA-3'
lncRNA-OET-01-013	9.2	5'-UUUUAU-20nt-AAUAAA-3'
lncRNA-OET-09-055	8.4	5'-UUUUAAU-13nt-AAUAAA-3'
lncRNA-OET-13-145	8.1	5'-UUUUAU-20nt-UUUUAAUAAA-3'
lncRNA-OET-02-259	7.4	5'-UUUUUAU-7nt-AAUAAA-3'
lncRNA-OET-12-064	7.2	5'-UUUUAU-29nt-AAUAAA-3' (*Snhg10*)
lncRNA-OET-05-363	6.8	5'-UUUUUAU-12nt-AAUAAA-3'
lncRNA-OET-03-178	5.5	5'-UUUUUAU-29nt-AAUAAA-3'
lncRNA-OET-07-478	5.5	5'-UUUUAU-88nt-AAUAAA-3'
lncRNA-OET-08-341	6.0	5'-UUUUUAU-58nt-AAUAAA-3'
Orc6l	12.2	5'-UUUUAUU-20nt-UUUUAU-27nt-AAUAAA-3'
Dcp1a	10.7	5'-UUUUAU-22nt-AAUAAA-3'
Plat	6.1	5'-UUUUAAU-17nt-AAUAAA-3'
Mos	5.0	5'-UUUAUUUUUUUAAA-38nt-AAUAAA-3'
Ccnb1	4.2	5'-UUUUAAU-22nt-AAUAAA-3'

Figure 14: *Polyadenylation analysis. (A) Distribution of poly(A) scores for mRNAs in GV and MII oocytes. (B) Distribution of poly(A) scores for lncRNAs in GV and MII oocytes. The Y axis depicts the poly(A) score calculated as the ratio of poly(A) RNA-Seq FPKM/total RNA FPKM. The X axis represents rank-sorted OET lncRNAs that had FPKM >0. Dashed red and blue lines represent median poly(A) score values for lncRNAs and mRNAs, respectively. Green points on the right site indicate poly(A) scores of histone mRNAs, black points on the curve indicate the rank of poly(A) scores of dormant maternal mRNAs. (B) Examples of putative CPE elements found among*

Major changes in cytoplasmic RNA polyadenylation during OET occur also post-fertilization. (Sakurai et al. 2005). A scatter plot of relative changes of lncRNAs in poly(A) and total RNA RNA-Seq sets in MII eggs and one-cell embryos showed a relative enrichment in lncRNA polyadenylation upon fertilization; similar to changes observed for mRNAs *(Fig.15)*. In contrast to mRNAs, the number of lncRNAs showing stronger decrease in poly(A) RNA upon fertilization was minimal. Taken together, our data show that the bulk of OET lncRNAs are polyadenylated at

their 3′ end and that maternal lncRNAs utilize the same cytoplasmic polyadenylation mechanisms as mRNAs. Therefore, cytoplasmic polyadenylation of OET lncRNAs most likely regulates their availability and stability, similar to mRNAs.

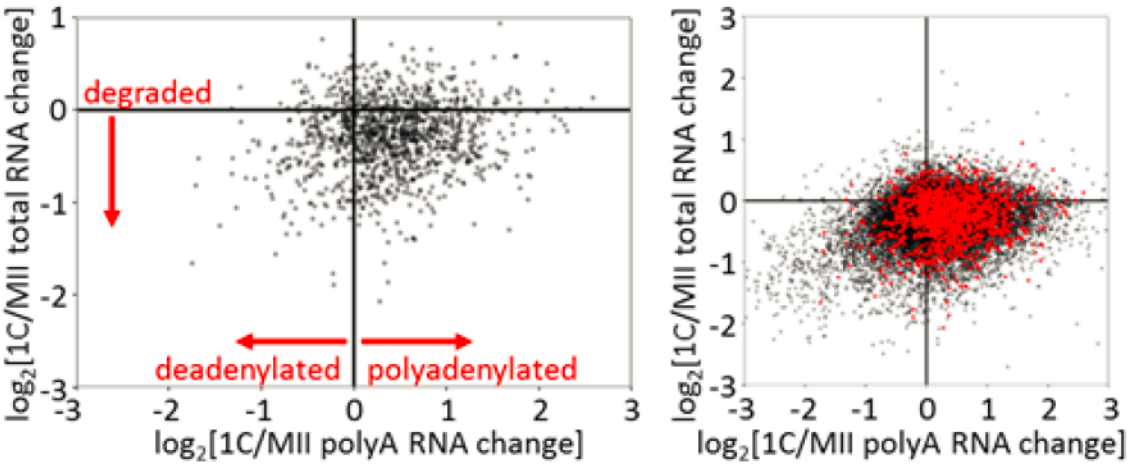


Figure 15: *polyadenylation changes upon fertilization. The Y axis depicts relative changes of gene expression in total RNA upon fertilization (log_2[1-cell/MII total RNA FPKM]), the X axis shows relative changes in poly(A) RNA (log_2[1-cell/MII poly(A) FPKM]). The left scatterplot displays only OET lncRNAs, the right plot shows OET lncRNAs in red superimposed onto mRNAs (black). Each point represents expression of one gene.*

Part 2: Expression Dynamics of LncRNA During OET

LncRNA expression display a strong cell/tissue specific expression compared to protein coding genes, which is one of their key characteristic features. To study dynamics of lncRNA expression during OET, RNA-Seq based transcriptome analysis was employed.

Expression of OET lncRNAs in other tissues

Since the majority of the OET lncRNAs appeared polyadenylated, their expression was inspected across 22 tissues selected from the ENCODE poly(A) RNA RNA-Seq mouse tissue panel (GSE4941744). To increase the specificity of the expression analysis, only clusters with four or more spliced reads and expression of >4 FPKM in at least one of the tissues in the tissue panel were included. The cut-off 4 FPKM for poly(A) RNA was used because it is a rough equivalent of 1 FPKM in total RNA RNA-Seq from mouse oocytes. Under these filtering conditions, expression values for 356 clusters were obtained (*Fig.16*). Analysis revealed a small population of ubiquitously expressed lncRNA clusters (28 cluster with expression >4 FPKM in all tissues), consistent with the notion that mammalian lncRNAs typically have a cell type-restricted expression (Derrien et al. 2012). Of the 28 ubiquitously expressed lncRNA clusters, 26 were annotated. These clusters were from small nucleolar RNA host genes and other annotated lncRNAs, such as *Malat 1, Firre*, or *Rian*. Remarkably, OET lncRNAs were mostly also expressed in the testis. Within the tissue panel (which included the ovary and the placenta), testis stood out as the tissue that had the highest number of expressed clusters across tissues (121 clusters) (*Fig.16*). Ovary ranked second after the testis, having the maximum expression of 19 clusters.

LncRNAs expressed in testis and during OET stages represent an interesting group of germ-line specific lncRNAs. Testis expressed 121 lncRNA clusters were further analyzed to determine 1) if they are associated with maternal or zygotic expression, and 2) whether they share the same promoters in the testis and OET stages. Most of the 121 lncRNA clusters during OET were highly expressed maternally (93 clusters), while 25 clusters had the highest expression in the zygotic stage and 3 clusters in the embryonic stage. 58 clusters of the 121

clusters shared promoters in testis and OET stages. In the case of 37 clusters, a unique non-repetitive lncRNA promoter functioned in testes, while oocytes or early embryos employed a different unique promoter (19/37) or a retrotransposon-derived promoter such as MaLR class LTR promoter (12/37). In any case, the 93 lncRNA loci expressed in oocytes and testes are the prime candidates for analysis of lncRNAs with germline-specific functions.

LncRNA dynamics during OET

Next, expression dynamics of OET lncRNAs during OET was analyzed. Gene expression during OET can be divided into three basic classes based on dynamics of maternal RNAs with zygotic transcripts:

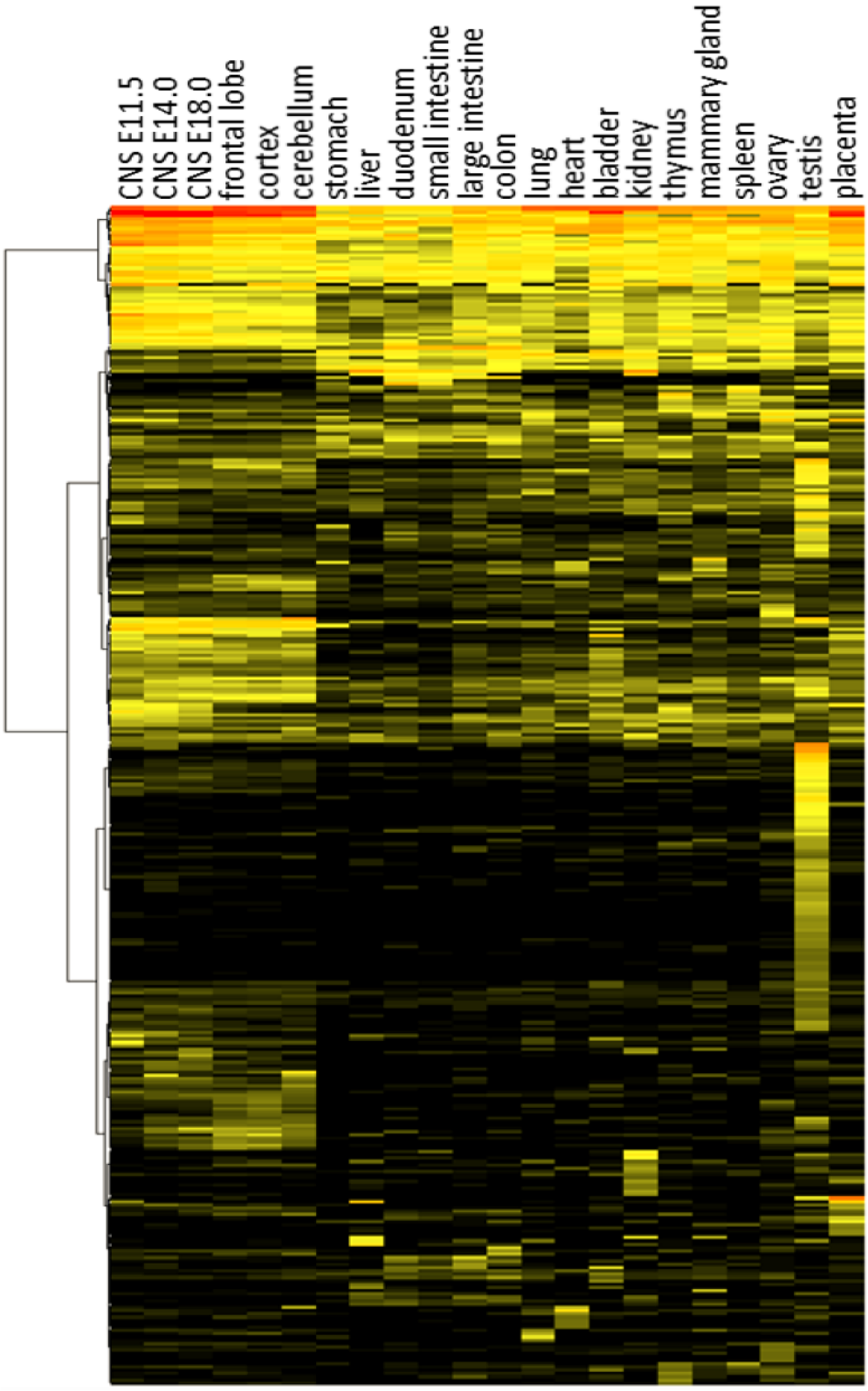
(i) **Maternal transcripts**—are transcripts, which are dominantly expressed only in oocytes. These transcripts survive until certain time-points/stages of OET and are not replaced with zygotic/embryonic transcripts. Based on their expression, maternal transcripts may be important just for oocyte development or they may function during meiotic maturation or soon after fertilization, where they can contribute to zygotic genome activation (ZGA) and initiation of development.

(ii) **Zygotic transcripts**—are expressed during ZGA and are absent in oocyte. Zygotic transcripts may be transcribed just transiently during ZGA or could remain expressed during early development. Zygotic transcripts, for example, could include genes involved in the establishment and maintenance of totipotency.

(iii) **Maternal-zygotic transcripts** —are transcripts found in both oocytes and early embryonic stages. This is due to expression of these genes in oocytes as well as during ZGA. This category can be exemplified by housekeeping genes. Maternal-zygotic transcripts, may have higher expression in oocytes than in early embryos or vice versa.

First, to simplify, we grouped all OET stages (as it was done for transcript model assembly) into three basic expression states: maternal (M), zygotic (Z), and embryonic (E). The maternal expression level was calculated as an average lncRNA expression level in GV and MII oocytes; it represented lncRNA expression before fertilization. The zygotic expression level was calculated as an average lncRNA level in two- and four-cell stages; it represented the transition period of

Figure 16: *OET lncRNA expression in different tissues. The heatmap displays expression of 356 lncRNA clusters with expression values >4 FPKM in at least one of 22 tissues selected from the ENCODE poly(A) RNA NGS mouse tissue panel (GSE49417 (Yue et al. 2014)).*



gene expression during ZGA. The embryonic expression level was calculated as an average lncRNA level in morulae and blastocysts stages; it represented gene expression at a later embryonic stage, during which maternal RNAs were mostly degraded (and so was ZGA-specific expression). Next, Maternal, Zygotic, and Embryonic lncRNA values were adjusted, so that the highest value observed in each state was set to one, and we displayed values for all lncRNA clusters in a single plot, where there were defined into M, Z, and E groups *(Fig.17A).* Using this expression plot, OET lncRNA clusters were further classified into six groups representing the maternal, zygotic, and maternal-zygotic expression types *(Fig.17B)*.

To obtain a comprehensive view of temporal dynamics of OET lncRNAs, clusters were organized into a heatmap based on six basic patterns *(Fig.17B),* while displaying expression in all stages *(Fig.17C)*. Most lncRNA clusters (1,166) were categorized in M group. Of these 993 of M clusters declined rapidly during ZGA, reaching low levels in the blastocyst stage. These represent maternal lncRNA clusters. This class is the most abundant in our dataset. In total 393 and 131 lncRNA clusters had maximum expression in Z and E stages, respectively. While, 251 of those had minimal maternal expression, thus representing zygotic lncRNA clusters *(Fig.17B)*. Among these zygotic clusters, 107 lncRNAs were transiently expressed during ZGA. Finally, 446 lncRNA clusters were considered maternal-zygotic transcripts. Maternal-zygotic lncRNAs could be further divided into two categories: (i) those transcripts constantly present during OET, i.e. zygotic transcripts, which are expressed even before maternal ones were eliminated, and (ii) those whose maternal transcripts were strongly eliminated before zygotic/embryonic transcripts were activated—there was a distinct minor group of 59 maternal-zygotic clusters whose expression reached the minimum at the two- and four-cell stages. The dynamics of OET lncRNA expression highly differed from mRNAs (*Fig.17C* bottom), mainly in the proportion of maternal and maternal-zygotic expression. While, 62% of the 1,600 OET lncRNA loci were maternal, maternally expressed mRNAs made only 20% of all OET mRNAs. Furthermore, maternal-zygotic lncRNAs made up a minor fraction of OET lncRNAs (28%), while this class was highly abundant (68%)

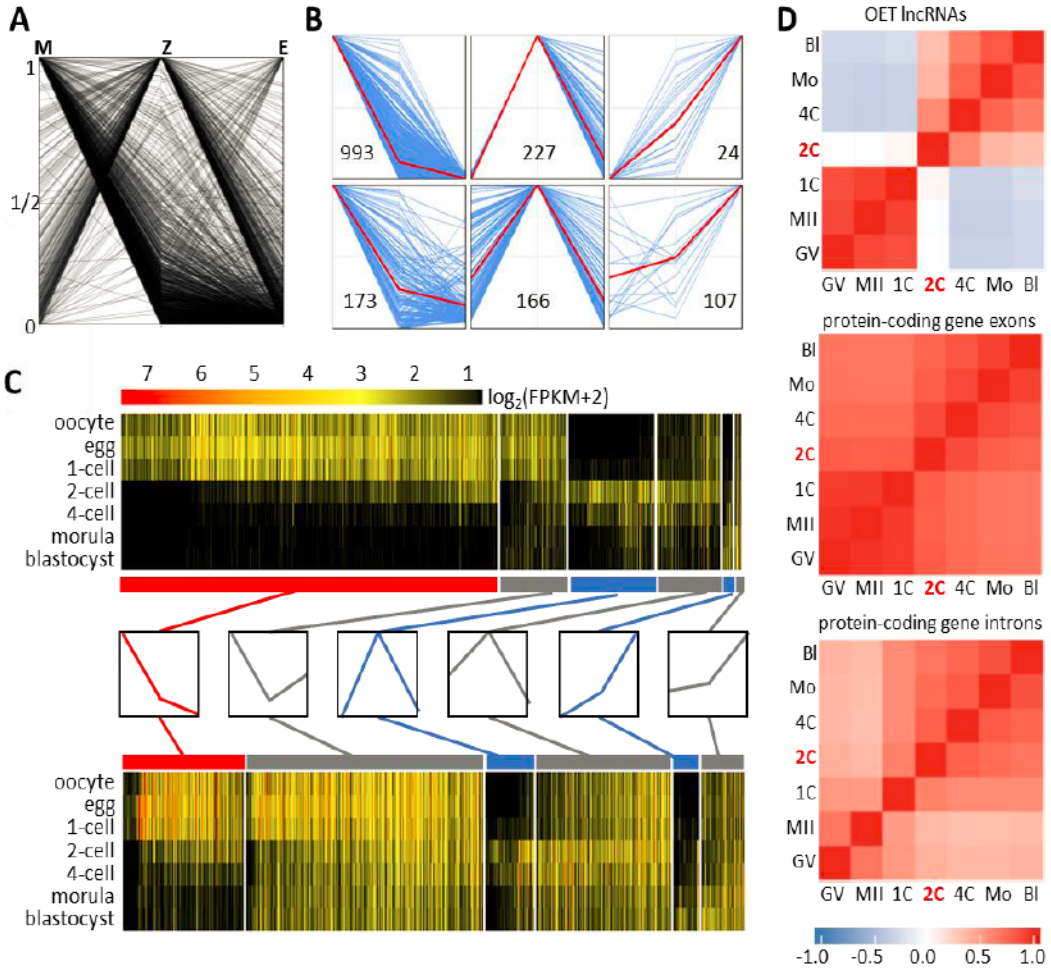


Figure 15 *OET lncRNA population dynamics during early development. (A) Overview of expression patterns of OET lncRNAs. To simplify expression pattern classification, average FPKM values were used: M, maternal (GV and MII oocytes); Z, zygotic (two- and four-cell stages); and E, late preimplantation embryo (morula and blastocyst). The plot shows dynamics of all clusters where the maximum average FPKM value of each cluster in M, Z, E was set to 1. (B) Main expression patterns of OET lncRNAs. The six panels display six basic patterns separating maternal (top left panel), zygotic (top middle and top right panels), and maternal-zygotic lncRNA (bottom panels) expression. The red lines represent the average values per each panel. (C) Expression patterns of 1,600 OET lncRNAs and 19741 mRNAs. The heatmap for lncRNAs and mRNAs was assembled from the six basic patterns (shown in (B)) (D) Expression correlations estimated from reads matching different types of sequences—exons of 1,600 lncRNA and exons and introns of protein-coding genes. The color scale on the left indicates the correlation coefficient for the analyzed features. Note the negative correlation for lncRNA expression between maternal and zygotic/embryonic stages reflects the apparently mutually exclusive expression patterns observed in the upper heatmap in (C)*

among mRNAs, which was not surprising considering the multitude of housekeeping roles of encoded proteins. Interesting observations were made from analysis of RNA expression correlations between individual stages (*Fig.17D*). Interestingly, when OET lncRNA and mRNA exons and introns were compared, lncRNAs showed positive correlations among maternal stages (GV oocyte and unfertilized/fertilized eggs) and among zygotic stages (two-, four-cell, morula, and blastocyst), but not between two stages from the two groups. Although two-cell stage lncRNAs showed no correlation with between stages, the later stages even had negative correlations (*Fig.17D*). This apparently reflected the extensive mutually exclusive nature of maternal and zygotic lncRNA transcriptions, which was also apparent from the expression heatmap (*Fig.17C*). In contrast, analysis of exons of protein coding genes revealed good correlations between two of the analyzed stages. This was due to dominant maternal-zygotic expression of protein-coding genes (68% of mRNA transcriptome). These results are consistent with a previous observation that lncRNA expression drastically varies at different stages of cleavage stage embryos, suggesting stage-specific expression (Zhang et al. 2014). This study identified two main expression patterns of OET lncRNAs—maternal expression, which diminished by early development, and zygotic expression dominated by a transient major ZGA expression pattern.

.

Part 3: Evolution of OET LncRNAs

Retrotransposon contribution to lncRNA evolution

Among 1600 OET lncRNAs identified, ~1200 lncRNA clusters were novel. Interestingly, retrotransposons made a strong contribution to novel OET lncRNA transcripts; one third of 5′ exons of novel lncRNAs came from long terminal repeat (LTR) retrotransposons. Whereas, SINE and LINE elements contributed to mature lncRNA sequences (*Fig.18A,B*). LTR retrotransposons, especially the MaLR class LTRs, made a strong contribution to lncRNA promoters and transcription start sites (*Fig.18A,B*). LTRs in 5' exons have the ability to splice in with novel splice acceptors downstream to give rise to new genes, which was the case in several novel OET lncRNAs (*Fig.18D*). This study mainly focused on contribution of MaLR and MT2 LTRs (fully or partially) to promoters and first exons of OET lncRNAs. Among the annotated lncRNAs, 333 LTRs were shown to make full 5' exon contributions to the lncRNA genes. These 333 OET lncRNAs thus represent cases of complete "plug-and-play" gene remodeling by MaLR and MT2 LTRs in mouse. LTRs retrotransposons not only provide promoter and first exon to lncRNAs, but also contribute to their tissue specific expression (*Fig.18C*). MaLR MT LTR driven lncRNAs were expressed exclusively in the oocytes, while MT2 LTRs driven lncRNA transcripts were expressed in zygotes, and ORR1 LTRs driven transcripts were expressed in both oocytes and zygotes. This stage specific lncRNA expression is probably due to the adaptation of LTR promoters to oocyte specific and zygote specific transcription control. Taken together, MaLR and MT2 LTRs are important contributors to lncRNA evolution, either by giving birth to new lncRNA or by recycling existing genes.

Maternal lncRNA function in RNAi

As described in the introduction, gene regulation by endogenous-siRNAs is essential for oocytes development and OET. Endogenous RNAi is a mechanism where mRNAs in the oocyte are targeted and degraded by small perfectly complementary 20-23 nt small interfering RNAs (siRNAs) Endogenous-siRNAs in mouse oocytes come from three distinct sources: (i) a transcription of an inverted repeat, (ii) a convergent transcription, and (iii) base pairing of mRNA and antisense

RNA originating, e.g. from a processed pseudogene (Tam et al. 2008; Watanabe et al. 2008). Even though

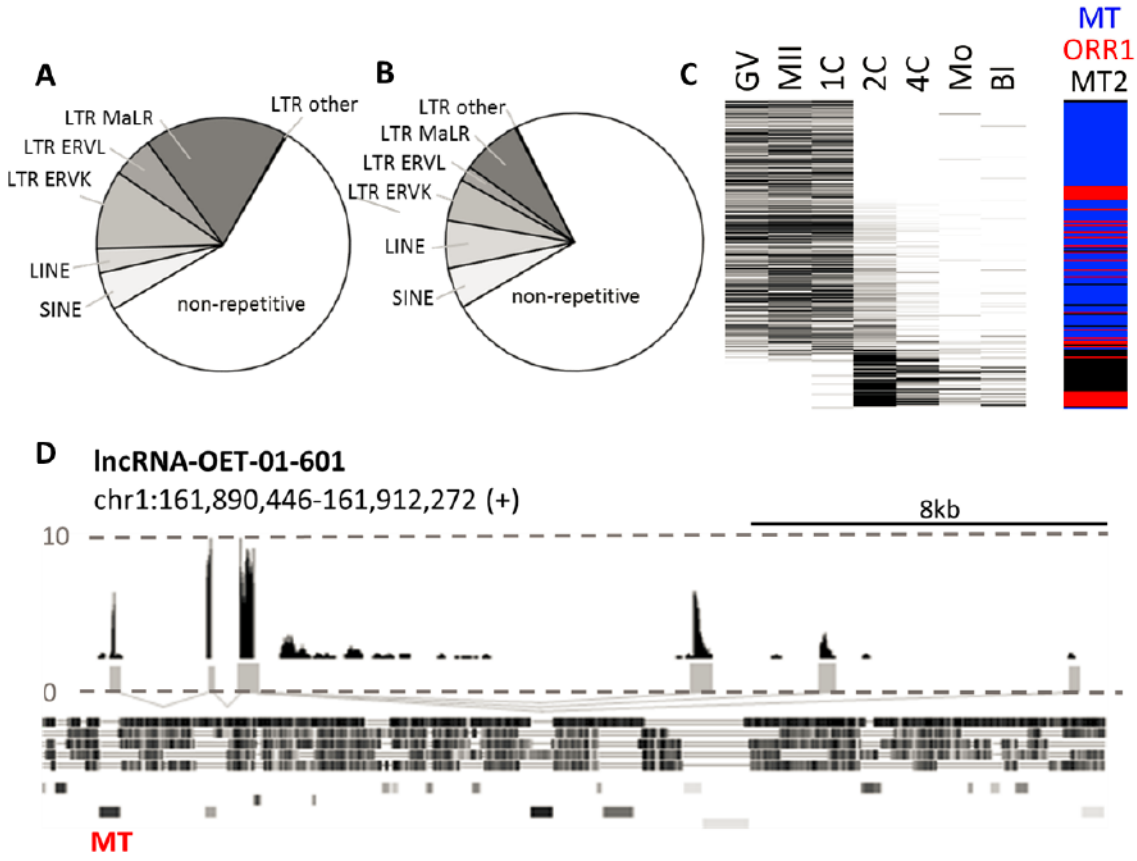


Figure 18: *Retrotransposon contribution to evolution of lncRNAs. (A) Contribution of retrotransposons to OET lncRNA transcriptional regulation. The graph depicts fractions of 5' OET lncRNA exons, which contain a given type of a repetitive sequence over the putative transcription start site and 50 bp upstream. (B) Contribution of repetitive sequences to mature (spliced) OET lncRNA sequences.(C) Expression of MaLR and MT2 LTR-derived 5' exons from lncRNAs ordered by the maternal/ZGA expression ratio (GV+MII)/(2C+4C). The heatmap shows log_2 FPKM values of the annotated LTR 5' exons (full contribution) with FPKM >0.1 in at least one sample. The colored bar indicates the LTR family. (D) UCSC browser snapshot of novel lncRNA locus showing lncRNA expression from a new MT insertion in mouse.*

antisense transcripts from pseudogene were considered as a source for siRNAs for a long time, the nature of these transcripts was unknown.

This study identified antisense sequences from processed pseudogenes among maternal lncRNAs. Antisense pseudogene carrying transcripts were typically multi-exon lncRNAs with one or more pseudogenes within them. LncRNAs were either seen to originate at the site of pseudogene insertion or in some cases,

pseudogene sequence was inserted into a pre-existing lncRNA transcript (*Fig 19*). Over 100 lncRNAs with processed pseudogene sequences were identified and 30 pseudogene sequences were 'recycled' by MaLR LTR-derived promoters into transcripts serving as endo-siRNA substrates in mouse oocytes. This lncRNAs serve as a unique class of maternal lncRNAs, which serve as substrates for the endo-siRNA machinery.

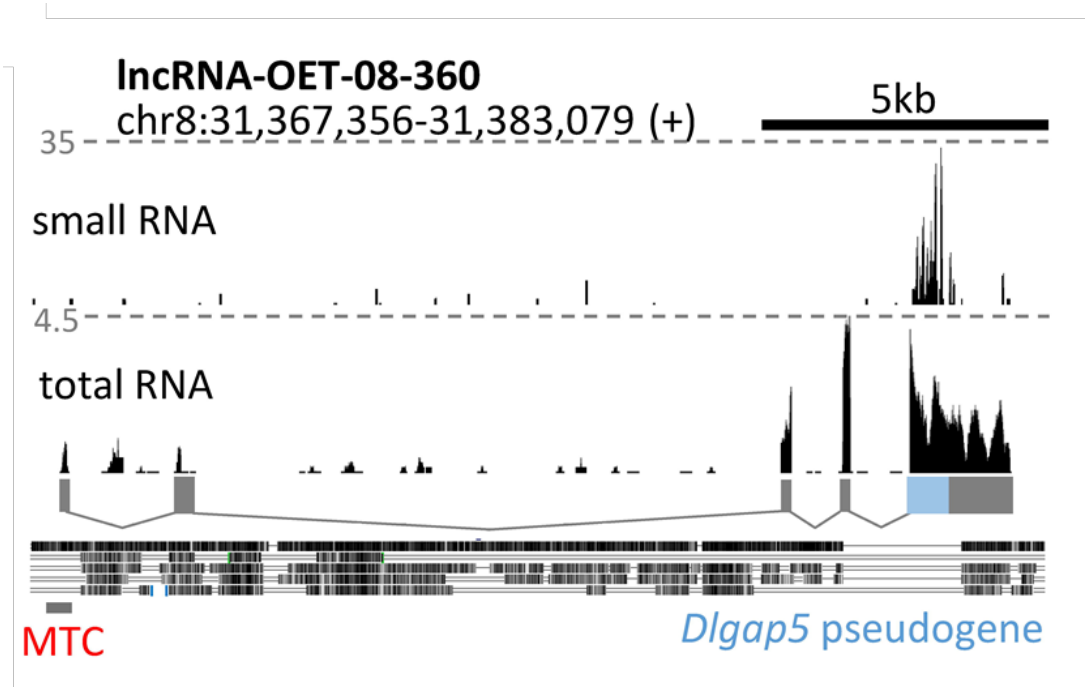


Figure 19: *An example of antisense pseudogene sequence rewiring yielding a lncRNA substrate for endo-siRNAs where a pseudogene (Dlgap5) was inserted into a locus already containing an MTA solo LTR.*

Part 4: Functional Analysis of LncRNAs During OET

Function of most of the 1,600 annotated OET lncRNAs is unknown and it is highly likely that majority of them are non-functional. However, based on their features, functions of few OET lncRNAs could be predicted. For example, we found two novel maternally expressed lncRNA clusters located in the imprinted loci, whose expression correlates with the maternal pattern of expression. LncRNA-OET-17-106, overlaps antisense with 3' end of *Airn* lncRNA, which is maternally silenced *(Fig 20A)*. LncRNA *Arin*, located on mouse chromosome 17, regulates imprinting of *Igf2r, Slc22a2* and *Slc22a3* genes in *cis*. And, LncRNA-OET-12-253, is expressed just downstream of an imprinted microRNA (miRNA) cluster, which is expressed from the maternal allele *(Fig 20A)*. We also observed interesting expression patterns of several known lncRNA loci, such as *Malat1*, *Neat1* or *Cyrano (Fig 20B)*. Metastasis-associated lung adenocarcinoma transcript 1 (*Malat1*) (Gutschner et al. 2013) is among the most studied lncRNA after *Xist*. It is a conserved extremely abundant lncRNA, non-essential for normal development (Eissmann et al. 2012, Zhang, 2012) However, in oocytes, *Malat1* transcript levels were minimal relative to later preimplantation stages *(Fig 20B).* Similarly, expression of *Neat1*, lncRNA encoded adjacent to *Malat1*, was seen only after ZGA and a shorter *Neat1* transcript isoform accumulates from the 4-cell stage on *(Fig 20B).* Thus, the *Malat1/Neat1* locus transcription is a zygotic component of OET. In contrast, *Cyrano* lncRNA, which has been implicated in embryonic development in zebrafish (Ulitsky et al. 2011), exhibited both maternal and zygotic expression pattern, while zygotic expression in the locus extends into a conserved 3' end region *(Fig 20B).*

Functional analysis of candidate OET lncRNAs

To gain insights into functional importance of OET lncRNAs, five maternal lncRNAs were selected based on their expression, conservation, promoter sequence and synteny. To study these lncRNAs, mouse knock-out models were created using RNA guided CRISPR-Cas9 system. We decided to delete promoter along with first exon, in order to suppress transcription at the locus and not just the accumulation of mature lncRNA. (Chen et al. 2013). For the first CRISPR deletion

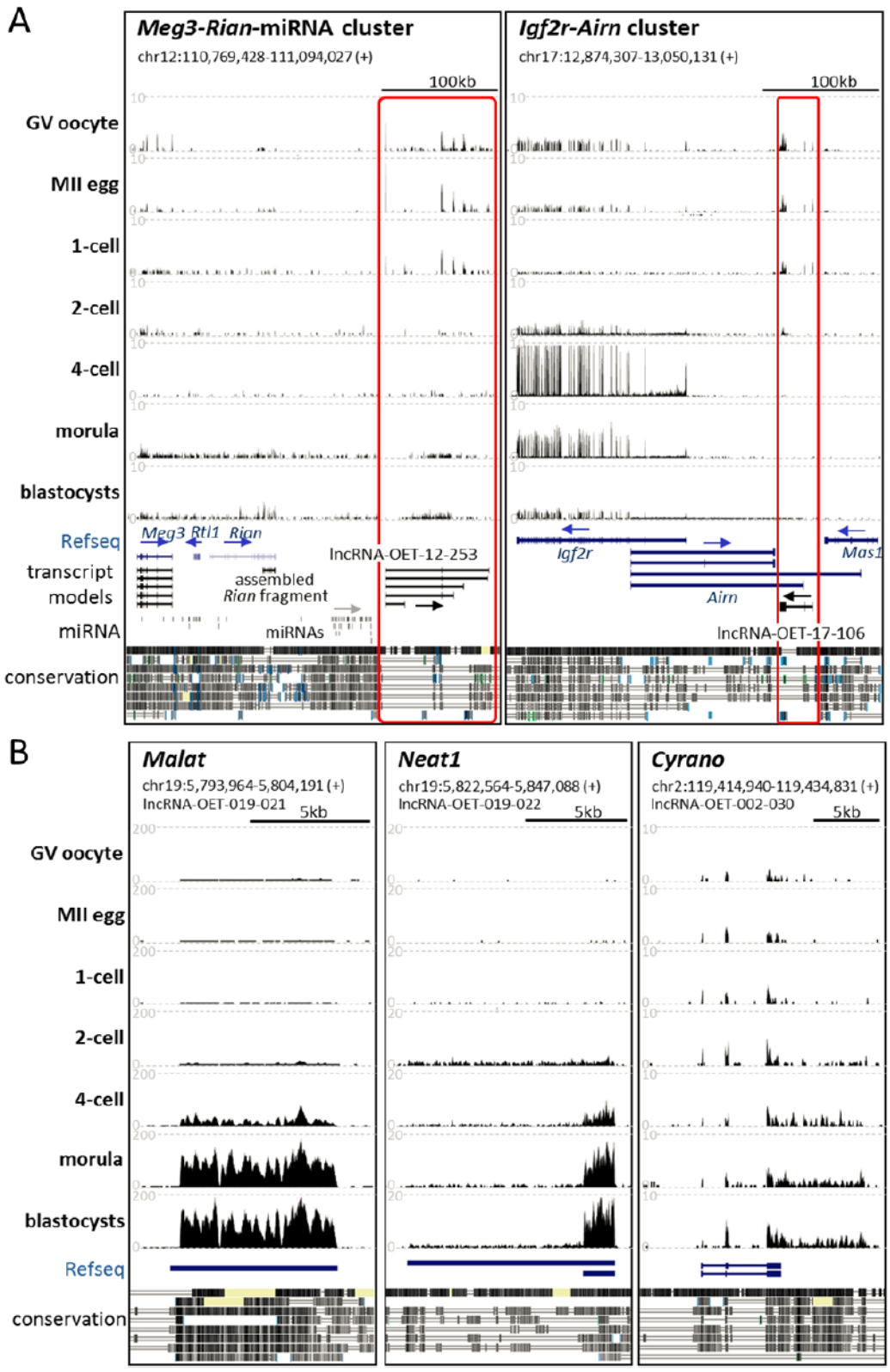


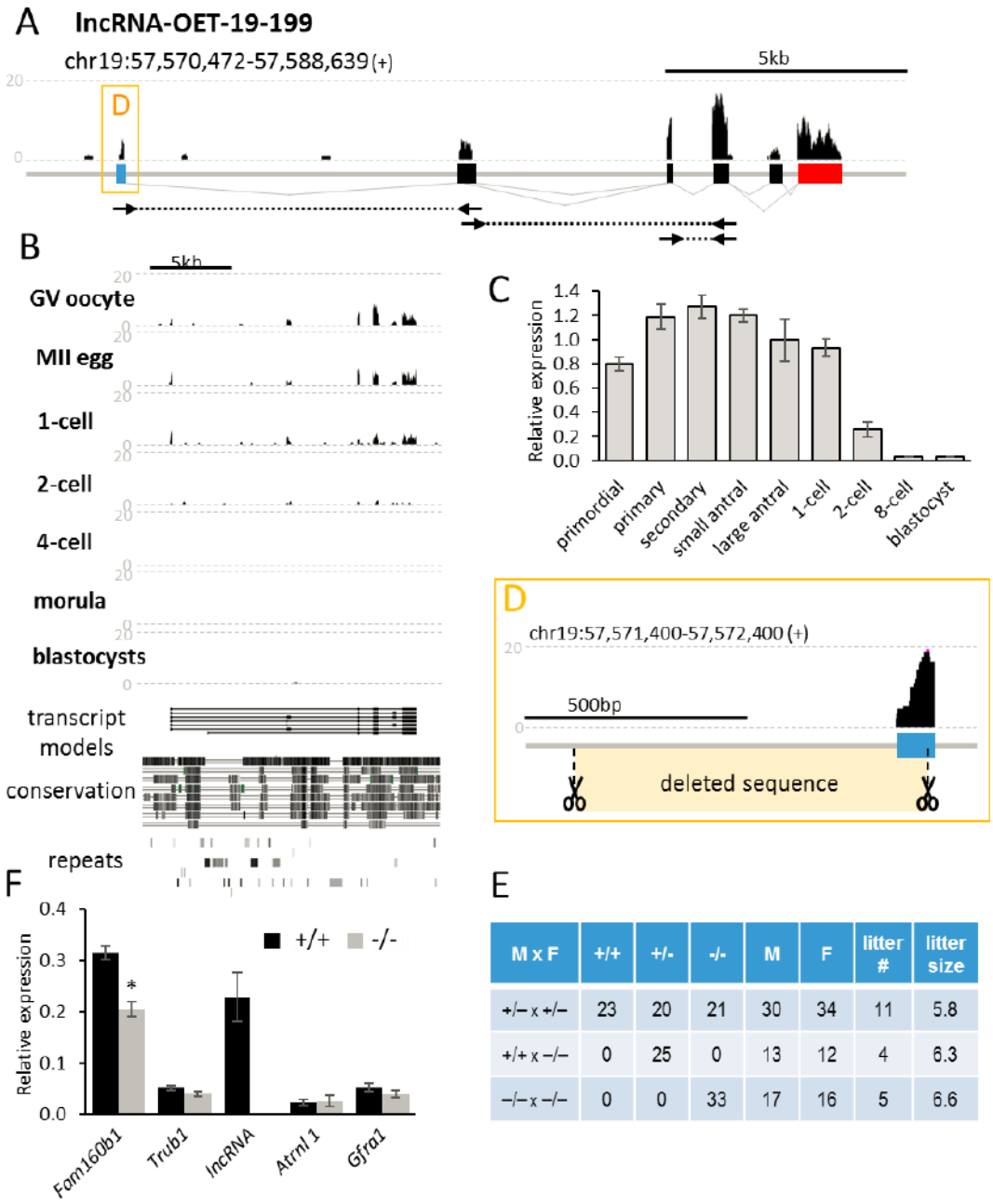
Figure 20: *Annotated USCS snapshots displaying expression of selected lncRNAs during oocyte-to-embryo transition. Grey dashed horizontal lines indicate expression maximum (FPKMs). The conservation track displays data eight genomes (top to bottom): Conservation track displays data from the following eight genomes rat, rabbit, human, tree shrew, dog, horse, cow, and platypus.*

experiment, mouse zygotes were injected with GFP tagged Cas9 protein + gRNA complexes, which were incubated together for 30min at 37°C prior to injection. Recombinant Cas9 was a gift from Martin Jinek lab, UZH. Inspection of zygotes two hours post injection did not reveal any GFP expression in zygotes, which we presumed was due to Cas9 degradation. Cas9 protein degradation was probably due to their low stability in zygotes or the used buffer was not compatible with the injection in mouse zygotes. After several trials, pronuclear injection of 100ng/ul Cas9 mRNA: 25ng/ul of gRNA in zygotes gave us successful mutant mice. We obtained 1 in 25 mice with positive deletion, which of course varied from gRNA to gRNA. Using this protocol we generated five lncRNA mutant mice; functional characterization of three select lncRNAs is explained below (lncRNA-OET-19-199, lncRNA-OET-06-154 and lncRNA-OET-07-157), whereas other two lncRNA mutant mice are still under investigation.

LncRNA-OET-19-199

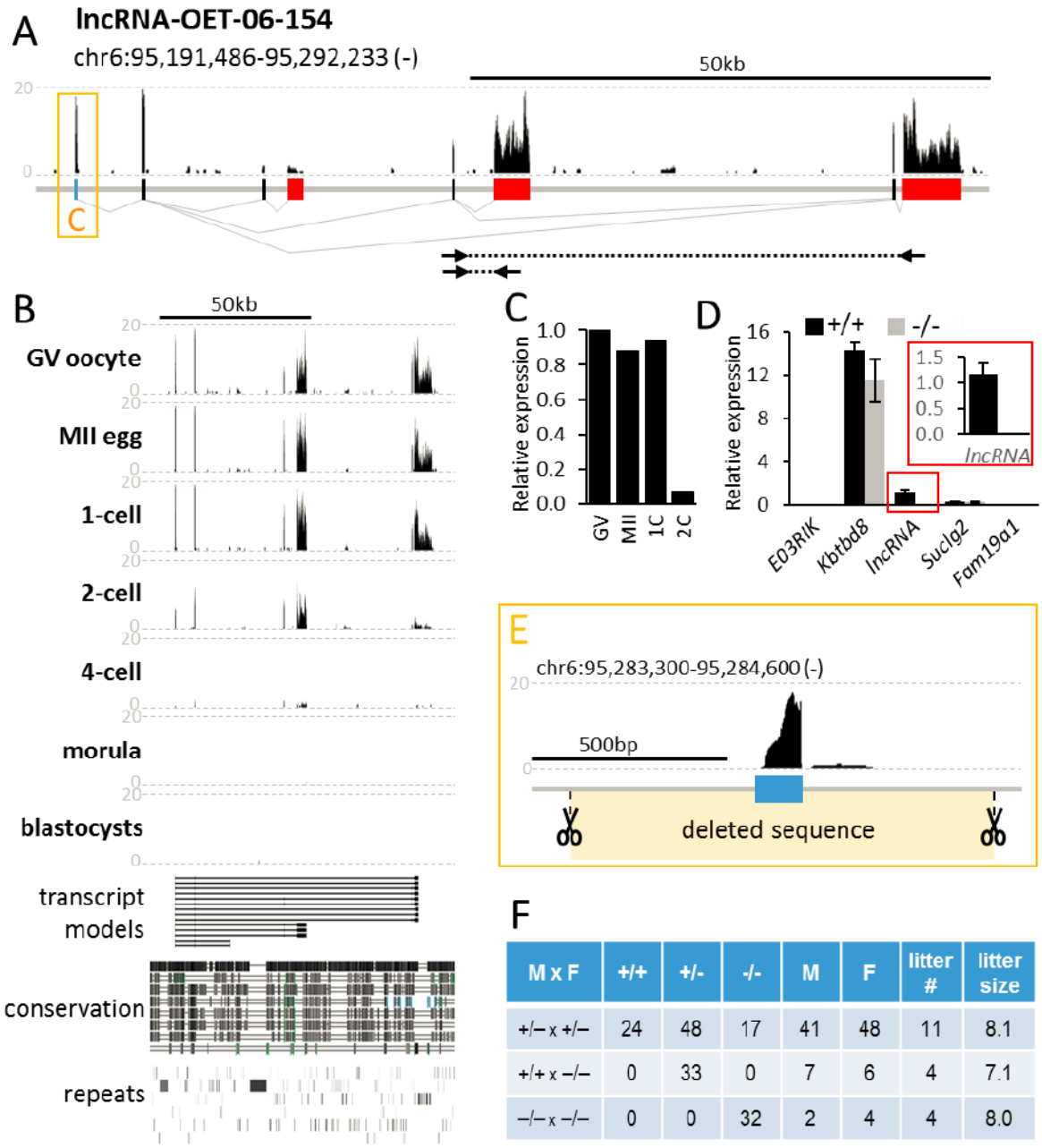
Lnc-OET-19-199 is a partially conserved maternal intergenic lncRNA, expressed in a relatively conserved region of chromosome 19 *(Fig 21A)*. The promoter sequence and first exon of lncRNA-OET-19-199 are partially conserved, whereas the last two exons are conserved in all mammals. Lnc-OET-19-199 consists of 6 exons, which our transcript model predicted to splice into six isoforms; two of which are unique to oocytes, whereas rest are also expressed in testis from an alternative promoter (Vasiliauskaite et al. 2018). Expression of Lnc-OET-19-199 begins in the primordial follicles and persists until fertilization, which triggers its degradation *(Fig 21B)*. To determine lncRNA-OET-19-199 function during OET, I generated a knock-out mouse model by CRISPR-cas9 mediated 500nt deletion of promoter along with first exon. Loss of lncRNA expression was confirmed by qPCR analysis *(Fig 21C, F).*

Breeding of lncRNA-OET-19-199 mutant mice did not reveal any effects on viability and fertility of homozygotes, although breeding of lncRNA-OET-19-199, heterozygotes produced heterozygotes with a lower frequency than expected *(Fig 21E)*. The basis of this effect is unclear. Nonetheless, homozygous null females were fertile and breeding of null animals produced viable offspring. Ovarian histology of knock-out animals appeared normal and normal amounts of fully-grown



M x F	+/+	+/-	-/-	M	F	litter #	litter size
+/– x +/–	23	20	21	30	34	11	5.8
+/+ x –/–	0	25	0	13	12	4	6.3
–/– x –/–	0	0	33	17	16	5	6.6

Figure 21: *LncRNA-OET-19-199 (A) Exon structure of lncRNA-OET-19-199 and its expression in the fully-grown GV oocyte. The grey dashed horizontal line indicates expression value 20 FPKMs. Terminal exons are displayed as red rectangles, the 5' exon in blue. (B) Annotated UCSC snapshot displaying expression of lncRNA-OET-019-199 during oocyte-to-embryo transition. Grey dashed horizontal lines indicate expression maximum of 20 FPKMs. The conservation track displays data from the following eight genomes (top to bottom): rat, rabbit, human, tree shrew, dog, horse, cow, and platypus. (C) Microarray profiling of lncRNA-OET-06-154 during oocyte development and in early embryos confirms temporal pattern of lncRNA-OET-06-154 expression Error bar = SEM. (D) Detailed view of positions of gRNAs for CRISPR-mediated deletion. (E) Breeding performance of mutant mice. (F) qPCR expression analysis of LncRNA-OET-19-199 and neighboring genes in WT and null oocytes.*



M x F	+/+	+/-	-/-	M	F	litter #	litter size
+/– x +/–	24	48	17	41	48	11	8.1
+/+ x –/–	0	33	0	7	6	4	7.1
–/– x –/–	0	0	32	2	4	4	8.0

Figure 22: *LncRNA-OET-06-154.(A) Exon structure of lncRNA-OET-06-154 and its expression in the fully-grown GV oocyte. The grey dashed horizontal line indicates expression value 20 FPKMs. Terminal exons are displayed as red rectangles, the 5' exon as a blue one. Positions of primers are indicated below the gene scheme. (B) Annotated UCSC snapshot showing expression of lncRNA-OET-06-154 during oocyte-to-embryo transition. Grey dashed horizontal lines indicate expression maximum of 20 FPKMs. The conservation track displays data from the following eight genomes (top to bottom): rat, rabbit, human, tree shrew, dog, horse, cow, and platypus. (C) qPCR expression analysis of lncRNA-OET-06-154 during oocyte-to-embryos transition. (D) qPCR expression analysis of LncRNA-OET-06-154 neighboring genes in WT and null oocytes. (E) Detailed view of positions of gRNAs for CRISPR-mediated deletion. (D) Breeding performance of lncRNA-OET-06-154 mutant mice.*

oocytes were recovered from null ovaries. Expression of neighboring genes was examined, as lncRNA transcription has been shown to regulate nearby genes, either positively (upregulation) or negatively (downregulation) (Kapranov et al. 2007). We observed a significant downregulation of neighboring gene Fam160b1 *(Fig 21F),* however, whether this reflected the biological role of lncRNA-OET-19-199 or whether it was a consequence of the introduced DNA deletion remains unknown. Altogether, even though lncRNA-OET-19-199 is a well expressed, conserved maternal lncRNA, it is not required for the normal oocyte development or during OET in mice.

LncRNA-OET-06-154

LncRNA-OET-06-154, is another partially conserved maternal lncRNA, located on chromosome 6 *(Fig 22A)*. Even though lncRNA-OET-06-154 has a highly conserved promoter and first two exons, conservation of the rest of the exons varies. Interestingly, the very last exon of lncRNA-OET-06-154 exist only in mouse and gives rise a splice isoform unique to mouse oocytes. LncRNA-OET-06-154 expression is predominantly maternal and the transcripts are degraded by 4-cell stage *(Fig 22B, C)*. To study LncRNA-OET-06-154 function, I produced knock-out mouse model by CRISPR-cas9 mediated 500nt deletion of promoter along with first exon, similar to previous lncRNA *(Fig 22E)*. Loss of lncRNA expression was confirmed by qPCR analysis *(Fig 22D)*. Breeding of LncRNA-OET-06-154 mutant mice did not reveal any effects on viability and fertility of homozygous mice and normal litter size was recorded *(Fig 22F).* Similarly, ovarian histology of knock-out animals appeared normal and normal amounts of fully-grown oocytes were recovered from mutant ovaries. Expression of neighboring genes was not influenced by the loss of LncRNA-OET-06-154 *(Fig 22D).*

Even though the lncRNA-OET-06-154 mutant mice lacked phenotypic function, a biological function to one of its transcript isoforms could be predicted. The most downstream terminal exon of lncRNA-OET-06-154 carried an antisense sequences from the *Eef1g* pseudogene *(Fig 23A)*. The pseudogene insertion took place in the common ancestor of mice and rats, which was subsequently disrupted by several LTR insertions in the mouse lineage. The transcript isoform with the last exon, thus carries antisense pseudogene sequence, which could base pair with

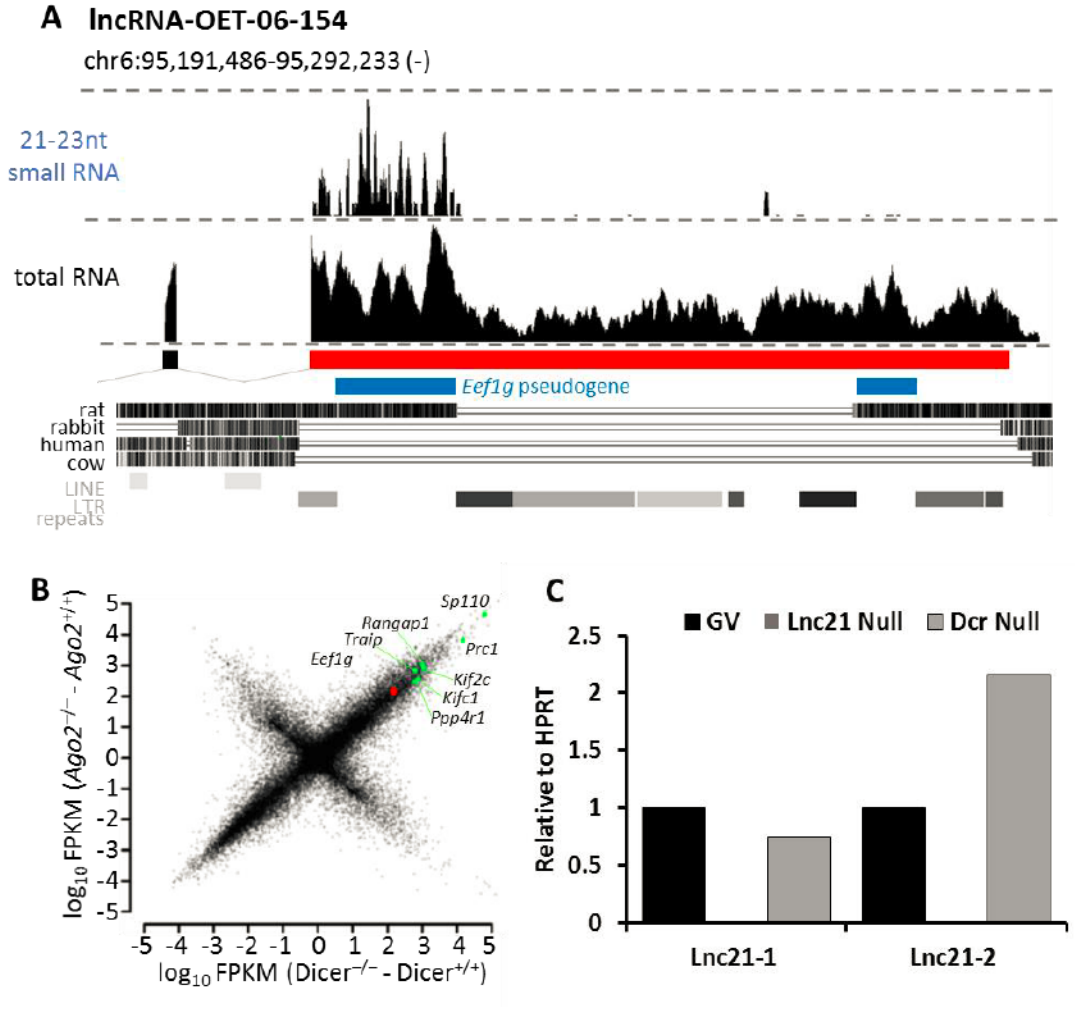
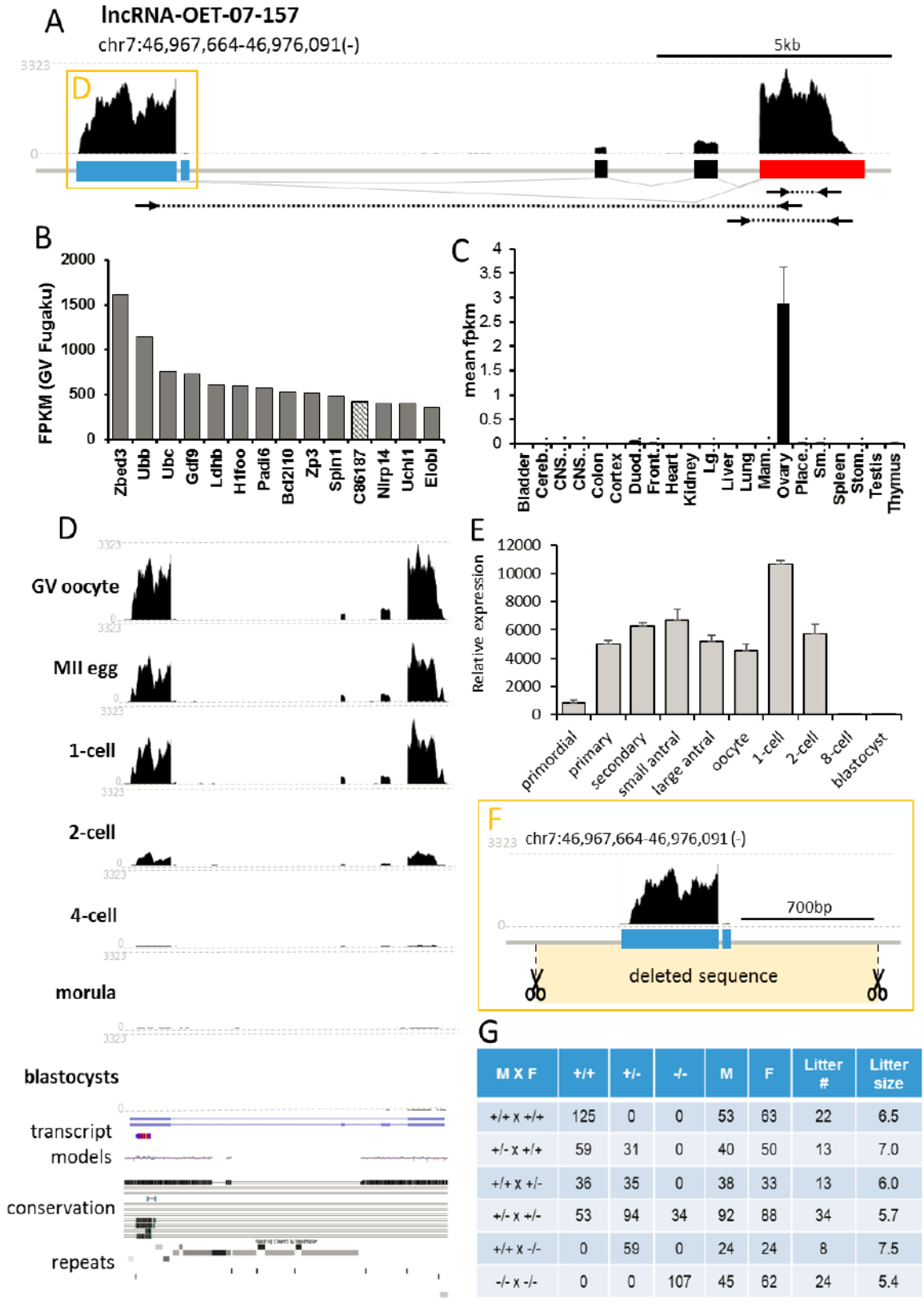


Figure 23: *LncRNA-OET-06-154 endo-siRNA activity (A) LncRNA-OET-06-154 terminal 3' exon region displaying lncRNA and small RNA expression in GV oocytes (B) Analysis of RNA-Seq data from Dicer and Ago2 knock-out oocytes. The Y-scale depicts the FPKM difference in Ago2 knock-outs (catalytically dead mutant), the X-scale in Dicer knock-outs. The FPKM difference was used because it better reflects the effect of suppressed RNAi on the transcriptome than the ratio, which is distorted by varying expression levels. (C) qPCR analysis of lncRNA-OET-06-154 expression in WT, Null and Dcr Null oocytes.*

Eef1g mRNA to give rise to dsRNA. This dsRNA is recognized by Dicer to generate 21-23 nt endo-siRNAs, shown by mapping small RNAs from small RNA-Seq data to the lncRNA-OET-06-154 locus *(Fig 23A)* (Tam et al. 2008). Small RNAs targeting *Eef1g* are biologically active as evidenced by *Eef1g* mRNA upregulation in both *Dicer* and *Ago2* knock-out oocytes *(Fig 23B)*. Simultaneously, overexpression of LncRNA-OET-06-154 overexpression in the absence of Dicer,



M X F	+/+	+/-	-/-	M	F	Litter #	Litter size
+/+ x +/+	125	0	0	53	63	22	6.5
+/- x +/+	59	31	0	40	50	13	7.0
+/+ x +/-	36	35	0	38	33	13	6.0
+/- x +/-	53	94	34	92	88	34	5.7
+/+ x -/-	0	59	0	24	24	8	7.5
-/- x -/-	0	0	107	45	62	24	5.4

Figure 24: *LncRNA-OET-07-157. (A) Exon structure of lncRNA-OET-07-157 and its expression in the fully-grown GV oocyte. The grey dashed horizontal line indicates expression value 3323 FPKMs. Terminal exons are displayed as red rectangles, the 5' exon as a blue one. Positions of qPCR primers are indicated below the gene scheme. (B) LncRNA-OET-07-157 expression compared to highly expressed oocyte genes. (C) lncRNA-OET-07-157 expression in ENCODE tissue RNA-Seq set Error bar = SEM (LncRNA-OET-07-157=Lnc1) (D). Annotated UCSC snapshot showing expression of lncRNA-OET-07-157 during oocyte-to-embryo transition. Grey dashed horizontal lines indicate expression maximum of 3323 FPKMs. (E) Microarray profiling of lncRNA-OET-07-157 during oocyte development and in early embryos confirms temporal pattern of lncRNA-OET-06-154 expression Error bar = SEM. (F) Detailed view of positions of gRNAs for CRISPR-mediated deletion. (G) Breeding performance of mutant mice*

as confirmed by qPCR expression analysis in RNAi-deficient (Dcr null) oocytes *(Fig 23C).* Taken together, lncRNA-OET-06-154 represents an example of a locus expressing multiple transcript isoforms with probably different functions: the most downstream 3' terminal exon is shown to be engaged in RNAi-mediated repression of *Eef1g* in the oocyte, while other transcript isoforms may have another function (or no function at all).

LncRNA-OET-07-157

LncRNA-OET-07-157 is a ~2kb long maternal intergenic lncRNA *(Fig 24A)*. It likely exists only in rodents, as this particular gene locus is disrupted by gaps in other mammalian genomes. Among 1600 OET lncRNAs annotated, lncRNA-OET-07-157 was found to one of the most abundant maternal lncRNA expressed. Interestingly, for a lncRNA (considering they have lower expression compared to mRNAs), lncRNA-OET-07-157 was found among top fifteen highly expressed genes in oocytes *(Fig 24B)*. High expression of lncRNA-OET-07-157 is probably due to its adaption to key maternal transcription factors in oocytes. Lhx8 and Nobox are two important maternal transcription factors required for oocyte development (Choi et al. 2008). Interestingly, *Lhx8* and *Nobox* null oocytes microarray data, revealed loss of lncRNA-OET-07-157 expression in *Lhx8* and *Nobox* null oocytes (Choi et al. 2008). This also explains the lack of lncRNA-OET-07-157 expression in other tissues *(Fig 24C)*. In oocytes, lncRNA-OET-07-157 expression initiates at the primary follicle stage and is seen until the 2 cell stage *(Fig 24D,E)*. LncRNA-OET-07-157 has an interesting gene structure. The first exon of the lncRNA is entirely

composed of retrotransposon sequences and a pseudogene; it consists of two MTC-int sequences and a SINE B3 element. Whereas, the last exon is free of retrotransposon sequences. The only intron of lncRNA-OET-07-157 consist of several LINE insertions, part of it also contributes to its splice isoforms *(Fig 24D)*. Knock-out mouse model was generated for LncRNA-OET-06-154 by CRISPR-Cas9 mediated 700nt deletion of promoter along with the first exon *(Fig 24E)*. Loss of lncRNA expression was confirmed by qPCR analysis. Long term breeding of LncRNA-OET-06-154 mutant mice did not reveal any effects on viability and fertility

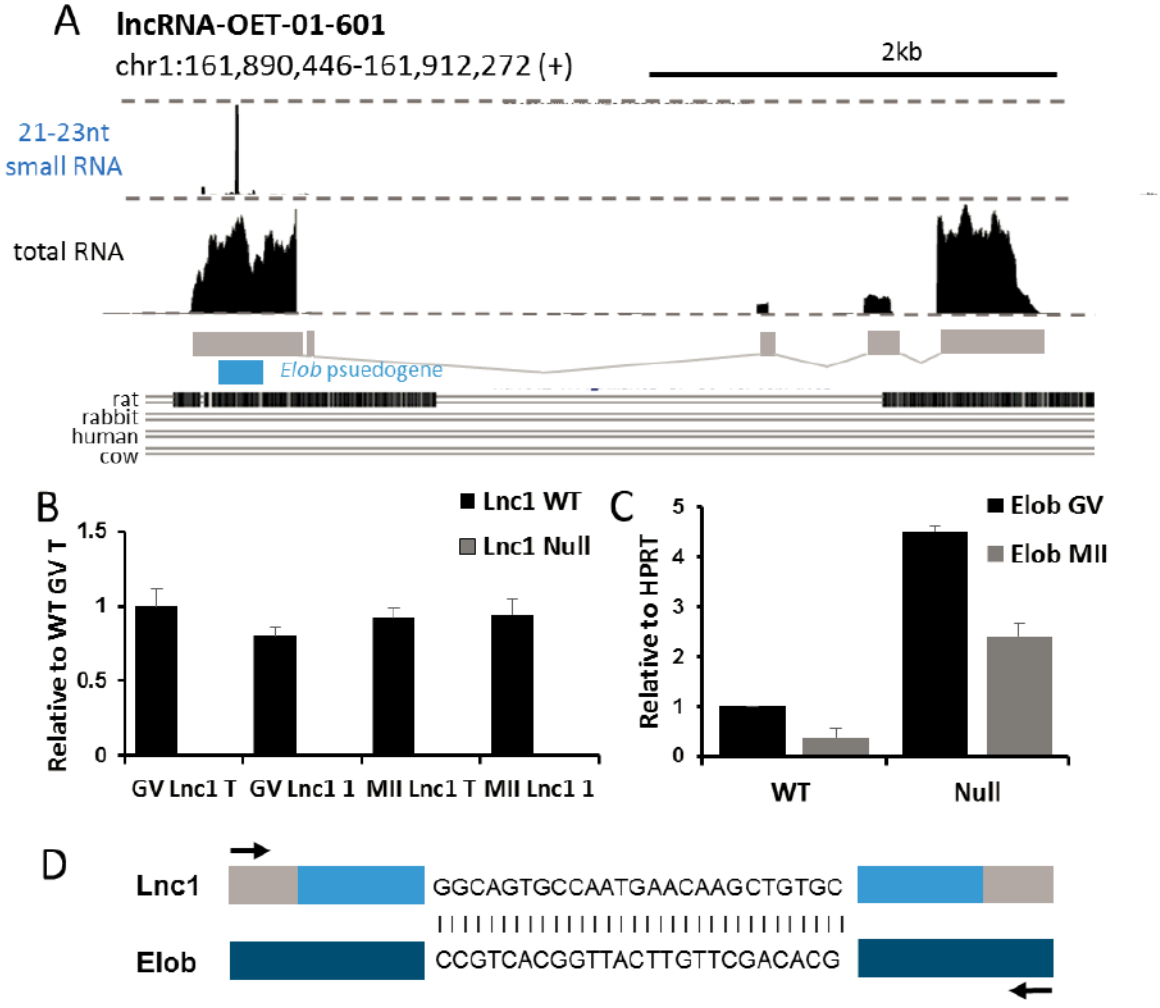


Figure 25: *Endo-siRNA activity of LncRNA-OET-07-157 (Lnc1)(A) LncRNA-OET-07-157 5' first exon region displaying lncRNA and small RNA expression in GV oocytes (B) qPCR expression analysis of total (GV Lnc1 T) and isoform containing the first and last exon (GV Lnc1 1) of LncRNA-OET-07-157 in WT and null, GV and MII oocytes. (C) qPCR expression analysis of Elob mRNA in LncRNA-OET-07-157 WT and null, GV and MII oocytes. (D) Schematic representation of the 25nt perfect match between LncRNA-OET-07-157 and Elob mRNA*

of homozygous mice, and produced normal litter *(Fig 24F)*. To examine the transcriptome dynamics of null LncRNA-OET-07-157 oocytes, I performed total RNA-Seq from GV oocytes. Analysis of null transcriptome revealed no significant difference compared to WT oocytes. However, two interesting functional properties of LncRNA-OET-07-157 were discovered LncRNA-OET-07-157 harbors an *Elob* pseudogene in antisense orientation in its first exon. The pseudogene insertion apparently occurred in the common ancestor of mice and rats *(Fig 25A)*. LncRNA-OET-07-157 transcript has 25nt perfect match with *Elob* mRNA, thus base pairing of transcripts can form dsRNA and generate endo-siRNAs, which can be seen by mapping small RNAs to this locus *(Fig 25D)*.Noticeably, upregulation of *Elob* mRNA is observed in LncRNA-OET-07-157 null oocytes *(Fig 25C)*. Therefore, LncRNA-OET-07-157 is another example of lncRNA class, which serves as small RNA substrates. Even though a significant increase in *Elob* mRNA was seen, no interference in normal fertility of LncRNA-OET-07-157 null mice was observed, suggesting the *Elob* regulation by RNAi is not critical for oocyte function.

While characterizing the polyadenylation status of OET lncRNAs we stumbled upon a set of lncRNAs, which displayed an increase in polyadenylation during oocyte maturation. Upon closer inspection, putative cytoplasmic polyadenylation elements (CPE) were found in their 3'UTR; which, as described before, were found only in dormant maternal mRNAs, where they are involved in selective translation of dormant maternal transcripts upon oocyte maturation. Notably, LncRNA-OET-07-157 was one of the lncRNAs, which displayed increased polyadenylation in MII oocytes compared to GV oocytes, and carried four canonical cytoplasmic polyadenylation elements in its 3'UTR *(Fig 26A)*. To examine LncRNA-OET-07-157 CPE functionality, a luciferase reporter assay was designed, where deletion mutant reporters were constructed for putative CPE sequences *(Fig 26B)*. Well characterized dormant mRNA *CcnB1* was used as a positive control (Yang et al. 2017). An increase in translation of LncRNA-OET-07-157 and *CcnB1* 3'UTRs was observed during oocyte meiotic maturation, while mutant reporters of LncRNA-OET-07-157 showed minimal translation, validating CPE functionality. Activity in *CcnB1* mutant reporters could be detected presumably due to deletion of only one of four CPEs.

In rats, LncRNA-OET-07-157, lacks two of the four CPE elements in their 3'UTR. Furthermore, *Elob* pseudogene insertion adjacent to the first exon, does not seem to be a part of the LncRNA-OET-07-157 transcript. Therefore, rats lack the endo-siRNA function in oocytes and probably has lower cytoplasmic polyadenylation activity compared to mouse. Taken together, in mouse, LncRNA-OET-07-157 is an example of a lncRNA with dual functionality, serving as both a small RNA substrate and a dormant maternal lncRNA whose function is currently under further investigation.

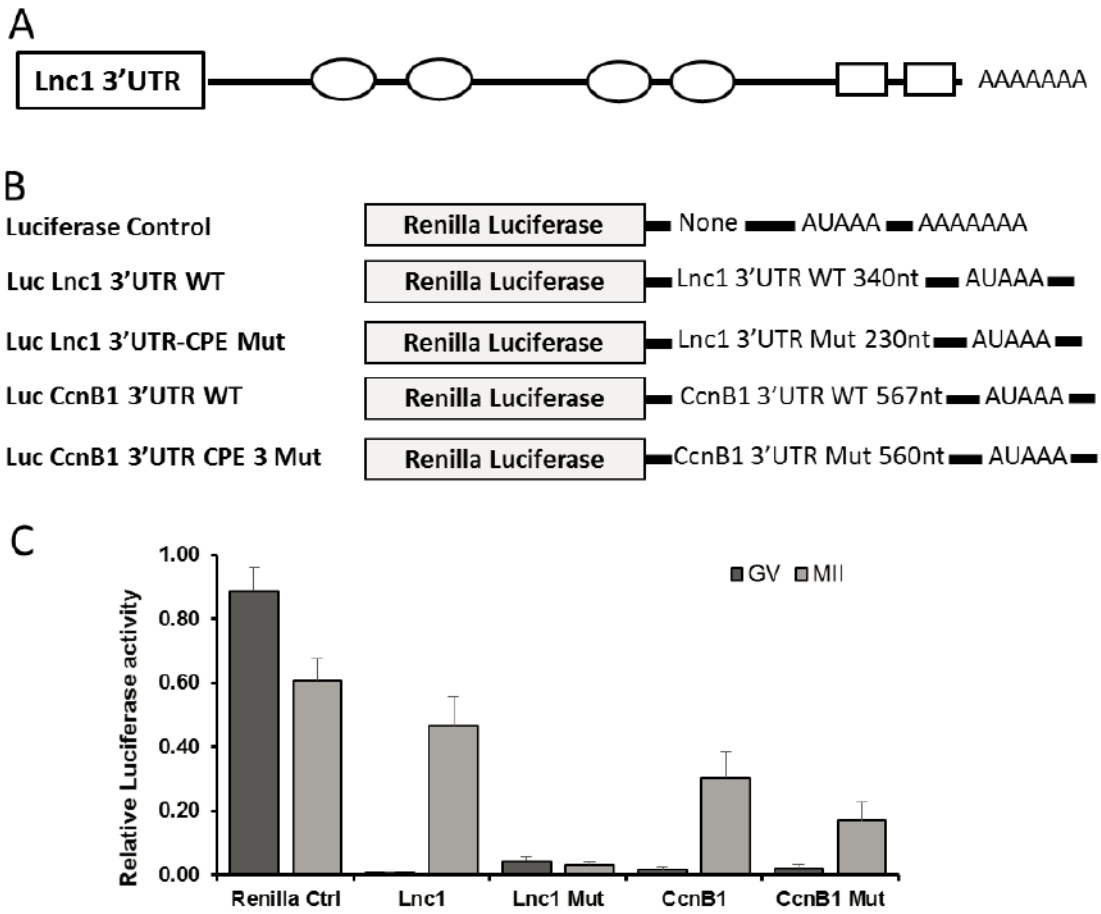


Figure 25: *LncRNA-OET-07-157 cytoplasmic polyadenylation: (A) schematic depiction of LncRNA-OET-07-157 3'UTR and the four canonical CPEs (in ovals) along with two hexanucleotide (poly(A) signal) sequence (in squares). (B) Luciferase reporters designed and injected for the luciferase reporter assay. (C) Relative luciferase activity of injected Renilla reporters. For these experiments, Renilla luciferase reporter activities were normalized to a coinjected firefly luciferase control. The experiment was conducted five times, and data are presented as the mean ± SEM.*

Discussion

At the beginning of this study absolutely nothing was known about expression or function of lncRNAs during mouse OET, except for two preliminary studies by Zhang et al (Zhang et al. 2014) and Hamazaki *et al*. (Hamazaki et al. 2015). Hamazaki *et al*. focused on promoter associated non-coding RNAs (pancRNAs), whereas Zhang et al identified 5,563 novel lncRNAs using single-cell SOLiD RNA-Seq data from mouse OET stages. However, analysis of their data revealed low quality lncRNA assembly and annotation, which did not represent a reliable dataset. LncRNAs predicted by Zhang et al. had less exons per transcript (typically only two) and less exons per lncRNA loci. Therefore, a robust pipeline for de novo transcript model assembly was required for identification and annotation of OET lncRNAs. We then use these lncRNA to explore their expression and potential contribution to OET.

OET lncRNA annotation

OET lncRNA study begins with their identification and annotation, which requires a robust pipeline and RNA expression data (Abe et al. 2015) For this study, 76PE non-amplified total RNA-Seq data from oocyte to blastocyst stages were used. Analysis of total RNA allowed us to explore the entire lncRNA population including non-polyadenylated lncRNAs. Additionally, considering lower depth of RNA-Seq data, focus was on annotating well-expressed (>1 FPKM) lncRNAs rather than annotating as many loci as possible. This step was included in the pipeline to prevent accumulation of artifacts from spurious transcription and from partial assembly of low-expressed lncRNAs.

Furthermore, to avoid miss annotation during mapping and clustering, 1200 lncRNA clusters were manually curated to identify specific sources of artifacts, which could be eliminated by optimizing the pipeline. The manual annotation was assisted by two other colleagues and made this annotation more reliable than other approaches, which relied solely on bioinformatics based filtering. For example, any transcripts with introns <20 nt and >250 kb were removed from the annotation. These introns were products of transcripts originating form repeat elements. This

study also excluded all single exon transcripts in order to eliminate transcripts originating from non-specific mapping. For example, transcripts form pseudogenes or reads originating from repeat elements, would map to multiple loci but only loci showing splicing could be positively identified as transcribed ones. This refining step drastically reduced the number of lncRNA transcript models in our annotation, which also explained the high number of transcripts and shorter length in Zhang *et al*'s annotation. Finally, CPAT tool was used to eliminate any transcripts with protein coding potential.

With the help all refining steps, we obtained 1600 non-redundant highly reliable set of OET lncRNA clusters. Majority of the clusters were expressed in maternal stages while the least number of clusters were found in morula and blastocysts stages. Low abundance of lncRNAs from these stages was a bit surprising as several thousands of lncRNAs have been annotated in mouse embryonic stem cells (Guttman et al. 2009). The embryonic set had had a relatively smaller depth compared to maternal set even though blastocyst has ~ three times more total RNA (~1.5 ng) than an oocyte (~0.5 ng) (Piko and Clegg 1982). The low abundance in these embryonic stages could be due to the presence of surrounding non-expressing cells from extraembryonic lineages, which could dilutes the transcriptome. Therefore, the same FPKM cut-off value would be more stringent on the blastocyst-expressed genes than the maternal genes. While lower counts of expressed lncRNAs in later preimplantation stages could be explained by the different depth, it does not explain the difference between the numbers of maternally and zygotically expressed lncRNAs. A 2-cell zygote contains approximately four times more total mRNA than a GV oocyte (Piko and Clegg 1982), thus, RNA-Seq data from 2-cell zygotes should have a larger relative depth. Yet, there are much less 2-cell specific lncRNAs than maternal lncRNAs. A partial explanation of higher count of maternal lncRNA can be provided by an order of magnitude higher number of transcriptionally active LTR promoters discovered in GV oocytes (Franke et al. 2017).

While this project was in progress, Gavin Kelsey's lab also published lncRNA annotation from oocytes and zygotes (Veselovska et al. 2015). This work provided a much more comprehensive analysis of OET lncRNAs compared to Zhang et al

and hamazaki et al. They had a much deeper maternal RNA-Seq data, which gave them longer transcript models with more splicing variants. Comparison of our 1600 OET clusters, revealed 70% overlap of our lncRNAs with their data, although we had much lower depth compared to their RNA-Seq data. Taken together, we successfully developed a bioinformatic pipeline for de novo assembly and annotation of lncRNAs and we produced a list of 1600 high-quality reliable maternal and zygotic lncRNAs expressed during OET.

Evolutionary origin of OET lncRNAs

LncRNAs exhibit characteristic features that distinguish them from protein coding genes. LncRNAs transcripts are generally shorter length, which is due to lower number of exons in their transcript compared to mRNAs. (Kapusta et al. 2013), which was also observed in OET lncRNAs. They also have low and very tissue specific expression. The expression analysis of OET lncRNAs revealed a very stage restricted expression pattern compared to protein coding genes. Most of the OET lncRNAs exhibited either maternal expression or zygotic expression pattern, whereas, protein coding genes predominantly exhibited a maternal-zygotic pattern of expression. These results are consistent with a recent study in human MII stage and cleavage stages embryos, where similar expression dynamics between stages was reported (Bouckenheimer et al. 2018). This implies that expression of the bulk of OET lncRNA clusters is driven by either oocyte- or ZGA-specific transcription factors, because only a minority of OET lncRNA clusters is ubiquitously expressed, i.e. presumably controlled by ubiquitous transcription factors. But why would lncRNAs adapt their expression to tissue or stage-specific transcription factors? A possible explanations could be that the local transcription factors provide less selective pressure on these random lncRNA transcription compared to ubiquitous transcription factor regulated genes, which are under stronger selective pressure.

LncRNAs are also less conserved compared to mRNAs and are among one of the most actively evolving genes. A study showed only ~92% of human intergenic lncRNAs were also detected as expressed in chimpanzee or bonobo and only ~72% were expressed in macaque, whereas more than 98% of conservation in expression

was observed for protein-coding genes, for all primates (Necsulea et al. 2014). LncRNA evolution studies in tetrapods shows that testis has a stronger specificity for young lncRNAs (55%) than for old lncRNAs, which is probably due to permissive testis chromatin, which favors new gene origination (Soumillon et al. 2013; Necsulea et al. 2014). Similar permissive chromatin might exist in the female germline.

The poor conservation of lncRNAs is probably because of emergence of lncRNAs from random transcription with low selective pressure. Such lncRNA transcripts, most of the time fail to acquire a function, and would therefore disappear from the genome at a much rapid pace compared to protein coding genes. This process may be enhanced in germline due to the permissive chromatin, which activates retrotransposons and retrotransposon driven specific and nonspecific gene transcription (Franke et al. 2017, Vasiliauskaite, 2018).

Interestingly, most of the novel OET lncRNA clusters were found associated with retrotransposons especially to their 5'exon. Retrotransposons have been previously shown to make a strong contribution to lncRNA sequences. Over two thirds of lncRNA sequences (75% and 68% of human and mouse, respectively) have at least a partial retrotransposon insertion in their sequence (Kapusta et al. 2013). This is more than other type of RNA sequences, such as protein coding sequences, small RNAs or untranslated regions (Kapusta et al. 2013) The high content of retrotransposon in lncRNA sequences is one of the likely contributing factor to their sequence diversification and high complexity.

Retrotransposons could contribute to lncRNAs by providing (as explained in the introduction)– internal part of an exon, transcription start sites (TSS), polyadenylation (poly(A)) site, splice donor or acceptor.(Kapusta et al. 2013). There are lncRNAs, which are almost completely made of retrotransposon sequences. For example, *lncRNA UCA1*, is enriched in bladder carcinomas and is conserved only in a few primate species (Wang et al. 2008a). In addition, many annotated lncRNAs share a significant proportion of their sequence with retrotransposons, for example *XIST* (Elisaphenko et al. 2008), *lincRNA-RoR* (Loewer et al. 2010) ,lncRNA *BORG*, etc.

Although all four major retrotransposon types contribute to lncRNA exons, ERV LTR retrotransposon family, were found to be the most enriched retrotransposon families in mouse and human lncRNAs (Kapusta et al. 2013). In mouse oocyte and zygotes, ERVL and ERVL/MaLR LTR retrotransposon families significantly contribute to TSS of several mRNAs and lncRNAs (Peaston et al. 2004, Franke, 2017). Interestingly, most of the ERVL/MaLR insertions in the mouse are solo LTR insertions (Franke et al. 2017). Solo ERVL/MaLR LTRs typically provided lncRNAs with retrotransposon-derived promoter, TSS, and often also with splice donors (SD). Thus a solo LTR insertion can yield a novel lncRNA where the LTR splicing donor will splice with a downstream splicing acceptor(s), hence will facilitate lncRNA genesis. Remarkably, 333 lncRNA clusters in mouse oocytes and zygotes were products of solo LTR insertions, exclusively in their 5'exons.

Altogether, almost half of the novel lncRNAs in oocyte and zygotes are products of retrotransposon activation by permissive chromatin, yielding poorly conserved and rapidly evolving lncRNA genes. At the same time, this raise the question of what fraction of lncRNAs would be functional and what fraction would represent result of opportunistic transcription from cryptic promoters (Kim et al. 2016).

Function(s) of OET lncRNAs

There has been a long standing debate regarding functionality of lncRNAs. Although thousands of lncRNAs have been discovered in vertebrates, very few have their biological function documented. One possible reason for why it is difficult to identify functionally important lncRNAs, could be their unique features, such as low conservation, lack of functional domains, low expression, which makes it technically difficult to investigate lncRNA functions. Especially in model systems where *in vitro* studies are not feasible. Alternatively it is also possible that most of the lncRNAs expressed are products of cryptic promoters, which have not acquired a function.

To study OET lncRNA functions, five OET lncRNAs were selected. The selection was solely based on different features of lncRNAs; such as, their expression levels in oocytes and zygotes, their expression profile during OET, their

genomic conservation and the promoter sequence. We aimed at conserved or partially conserved intergenic lncRNAs, which would be well expressed and would have a single annotated promoter in the oocyte.

There are few experimental strategies that have been frequently employed to gain insights into lncRNA mechanisms and functions. For example, fluorescence in situ hybridization technique, is used to detect lncRNA localization within a cell, which could distinguish lncRNA function from nuclear chromatin regulation from cytoplasmic post transcriptional regulation. Likewise, RNA-immunoprecipitation or RNA affinity pulldown techniques can provide insights into possible lncRNA interaction with proteins, RNA and/or DNA. However, several of these *in vitro* techniques are inaccessible for model systems like oocytes and zygotes, where it is difficult if not impossible, to acquire enough starting material for biochemical analyses and their architecture complicates the use of FISH for detecting specific localization of a chosen RNA. Unlike cell culture and *in vitro* model systems, oocytes and zygotes severely restrict options for a preliminary functional analysis. Considering the model system and the funding window, we thus opted to directly produce loss of function mouse models of selected lncRNAs using CRISPR-Cas9 mediated deletions of promoters and first exons. The idea was to eliminate transcription at the lncRNA locus rather than eliminating mature RNA levels. The intrinsic issue of the CRISPR-Cas9 system is off-targeting (Cho et al. 2014). However, this issue is not a major problem in mice because off-targets effects of CRISPR-Cas9 could be eliminated by several back crossing events. A more problematic issue was the presence cryptic promoters in the locus. Deletion of just promoters and first exon was in some cases not sufficient to knock-out the lncRNA transcription completely. In two of the mouse models, alternative transcription start sites were observed. In fact, in one of the cases lncRNA was overexpressed and reached two times higher levels than normal. Therefore, an optimal strategy for analyzing a lncRNA in vivo would be deletion of the entire lncRNA locus instead of just the annotated transcription start site, which would be combined with suitable rescue experiments.

All mutant mice models were subjected to breeding analysis, histology, expression analysis by qPCR and RNA-Seq, to examine possible lncRNA functions.

All five lncRNA mutant mice model were viable and fertile, suggesting that selected lncRNAs were functionally insignificant. However, a few interesting observations were made. The most significant concerned two mouse models, which revealed two oocyte-specific lncRNA features.

During expression analysis we identified 100 maternal OET lncRNAs with antisense pseudogene insertions. Interestingly, many of these lncRNAs could base pair with their respective mRNAs and form dsRNAs. Mapping of small RNA-Seq data revealed 21-23 nt small RNAs produced from these lncRNAs. LncRNA-OET-06-154 and lncRNA-OET-07-157, represent two validated examples of endo-siRNA producing lncRNAs. LncRNA-OET-06-154 carries *Eef1g* antisense pseudogene sequence and LncRNA-OET-07-157, carries antisense pseudogene insertion for *Elob* gene. Even though, upregulation of target mRNAs could be observed, mouse loss of function models of these lncRNAs were viable and fertile. This is probably because, RNAi mediated *Eef1g* and *Elob* expression is not critical for the oocyte development. None the less, these lncRNAs represent a unique class of maternal lncRNAs, which provide substrate to the endo-RNAi machinery in mouse oocytes.

Another interesting class of lncRNAs was observed while analyzing polyadenylation status of OET lncRNAs. 91 maternal lncRNAs displayed an increased level during meiotic maturation. Many of these lncRNAs carried putative cytoplasmic polyadenylation elements (CPEs) in their 3'UTRs. CPEs have been shown to suppress mRNA translation in growing oocytes and mediate translational activation during meiotic maturation through cytoplasmic polyadenylation (Charlesworth et al. 2004). Cytoplasmic polyadenylation activity in oocytes was always thought to be coupled with protein coding genes. Surprisingly, our study for the first time identified lncRNAs exhibiting the same cytoplasmic polyadenylation as mRNAs. LncRNA-OET-07-157 is one of the most expressed RNAs in the oocyte, which carries an *Elob* antisense pseudogene in its 5'exon also contains four CPEs in its 3'UTR. LncRNA-OET-07-157 shows increased levels in poly(A)-selected RNA during meiotic maturation. To validate the CPE functionality, we designed a luciferase assay with mutant lncRNA-OET-07-157 3' end, and used *CcnB1* 3' UTR as a positive control. Interestingly, loss of CPEs in mutant reporters correlated with reduced luciferase activity in meiotically matured eggs, while reporters with wild

type (WT) lncRNA-OET-07-157 3' end showed an increased luciferase activity. This analysis demonstrated that CPE sequences in lncRNAs are functional. As described earlier, loss of function mouse model of lncRNA-OET-07-157 was viable and fertile. Nonetheless, a more detailed examination of oocyte maturation in lncRNA-OET-07-157 mutant mice is required to completely understand its CPE functionality. In any case, these results define a novel class of lncRNAs: dormant maternal lncRNAs, which undergo cytoplasmic polyadenylation during meiotic maturation. We predict that, instead of translation, CPEs in lncRNAs might regulate their stability or localization.

Conclusions

My PhD project “Long non-coding RNAs in Oocyte-to-Embryo Transition”, yields a comprehensive analysis of OET lncRNAs in during OET in mice. The key conclusions are:

- 1600 highly reliable non-redundant lncRNA loci were identified and annotated; majority of the annotated lncRNAs were novel.
- Expression analysis revealed shorter length, low conservation, and restricted expression of OET lncRNA.
- A half of novel OET lncRNA loci in mice oocytes and zygotes utilizes LTR promoters.
- Two oocyte specific lncRNA features were identified:
  LncRNAs serving as endo-siRNA substrates.
  Dormant maternal lncRNAs, which undergo cytoplasmic polyadenylation.
- Five CRISPR mediated lncRNA loss of function mouse models were produced with no phenotype concerning fertility.
- LncRNA endo-siRNA activity and lncRNA cytoplasmic polyadenylation were validated in two of the loss-of-function mouse models

LncRNAs identified by our annotation provides an overview of maternal and zygotic lncRNA expression, regulation and evolution in oocytes and zygotes. Our data highlights the contribution of retrotransposons to the evolution of lncRNAs. And, two unique lncRNAs types were identified and characterized, which are associated with molecular mechanisms essential for oocyte development and oocyte-to-embryo transition.

References

Abe K, Yamamoto R, Franke V, Cao M, Suzuki Y, Suzuki MG, Vlahovicek K, Svoboda P, Schultz RM, Aoki F. 2015. The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3' processing. *EMBO J* **34**: 1523-1537.

Adhikari D, Liu K. 2009. Molecular mechanisms underlying the activation of mammalian primordial follicles. *Endocr Rev* **30**: 438-464.

Albertini DF, Combelles CM, Benecchi E, Carabatsos MJ. 2001. Cellular basis for paracrine regulation of ovarian follicle development. *Reproduction* **121**: 647-653.

Aoki F, Worrad DM, Schultz RM. 1997. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev Biol* **181**: 296-307.

Arraztoa JA, Zhou J, Marcu D, Cheng C, Bonner R, Chen M, Xiang C, Brownstein M, Maisey K, Imarai M et al. 2005. Identification of genes expressed in primate primordial oocytes. *Hum Reprod* **20**: 476-483.

Babushok DV, Ostertag EM, Kazazian HH, Jr. 2007. Current topics in genome evolution: molecular mechanisms of new gene formation. *Cell Mol Life Sci* **64**: 542-554.

Bachvarova R, De Leon V, Johnson A, Kaplan G, Paynton BV. 1985. Changes in total RNA, polyadenylated RNA, and actin mRNA during meiotic maturation of mouse oocytes. *Dev Biol* **108**: 325-331.

Bartolomei MS, Zemel S, Tilghman SM. 1991. Parental imprinting of the mouse H19 gene. *Nature* **351**: 153-155.

Beck CR, Garcia-Perez JL, Badge RM, Moran JV. 2011. LINE-1 elements in structural variation and disease. *Annual review of genomics and human genetics* **12**: 187-215.

Bonnet-Garnier A, Feuerstein P, Chebrout M, Fleurot R, Jan HU, Debey P, Beaujean N. 2012. Genome organization and epigenetic marks in mouse germinal vesicle oocytes. *Int J Dev Biol* **56**: 877-887.

Bouckenheimer J, Fauque P, Lecellier CH, Bruno C, Commes T, Lemaitre JM, De Vos J, Assou S. 2018. Differential long non-coding RNA expression profiles in human oocytes and cumulus cells. *Sci Rep* **8**: 2202.

Bowles J, Koopman P. 2010. Sex determination in mammalian germ cells: extrinsic versus intrinsic factors. *Reproduction* **139**: 943-958.

Brannan CI, Dees EC, Ingram RS, Tilghman SM. 1990. The product of the H19 gene may function as an RNA. *Mol Cell Biol* **10**: 28-36.

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**: 38-44.

Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D. 2012. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**: 819-831.

Cai X, Cullen BR. 2007. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* **13**: 313-316.

Cantley LC. 2002. The phosphoinositide 3-kinase pathway. *Science* **296**: 1655-1657.

Castrillon DH, Miao L, Kollipara R, Horner JW, DePinho RA. 2003. Suppression of ovarian follicle activation in mice by the transcription factor Foxo3a. *Science* **301**: 215-218.

Charlesworth A, Cox LL, MacNicol AM. 2004. Cytoplasmic polyadenylation element (CPE)- and CPE-binding protein (CPEB)-independent mechanisms regulate early class maternal mRNA translational activation in Xenopus oocytes. *J Biol Chem* **279**: 17650-17659.

Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS et al. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**: 1479-1491.

Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**: 132-141.

Choi Y, Ballow DJ, Xin Y, Rajkovic A. 2008. Lim homeobox gene, lhx8, is essential for mouse oocyte differentiation and survival. *Biol Reprod* **79**: 442-449.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

De Pol A, Vaccina F, Forabosco A, Cavazzuti E, Marzona L. 1997. Apoptosis of germ cells during human prenatal oogenesis. *Hum Reprod* **12**: 2235-2241.

Debey P, Szollosi MS, Szollosi D, Vautier D, Girousse A, Besombes D. 1993. Competent mouse oocytes isolated from antral follicles exhibit different chromatin organization and follow different maturation dynamics. *Mol Reprod Dev* **36**: 59-74.

Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231-235.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**: 1775-1789.

Diederichs S. 2014. The four dimensions of noncoding RNA conservation. *Trends Genet* **30**: 121-123.

Dorrington KJ, Hofmann T. 1973. Effects of ionic strength on the circular dichroism spectrum of chymotrypsin. *Can J Biochem* **51**: 1059-1065.

Eichenlaub-Ritter U, Peschke M. 2002. Expression in in-vivo and in-vitro growing and maturing oocytes: focus on regulation of expression at the translational level. *Hum Reprod Update* **8**: 21-41.

Eissmann M, Gutschner T, Hammerle M, Gunther S, Caudron-Herger M, Gross M, Schirmacher P, Rippe K, Braun T, Zornig M et al. 2012. Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA biology* **9**: 1076-1087.

Elisaphenko EA, Kolesnikov NN, Shevchenko AI, Rogozin IB, Nesterova TB, Brockdorff N, Zakian SM. 2008. A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PloS one* **3**: e2521.

Elvin JA, Matzuk MM. 1998. Mouse models of ovarian failure. *Rev Reprod* **3**: 183-195.

Eppig JJ, O'Brien MJ. 1996. Development in vitro of mouse oocytes from primordial follicles. *Biol Reprod* **54**: 197-207.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9**: 397-405.

Frank-Vaillant M, Haccard O, Thibier C, Ozon R, Arlot-Bonnemains Y, Prigent C, Jessus C. 2000. Progesterone regulates the accumulation and the activation of Eg2 kinase in Xenopus oocytes. *Journal of cell science* **113 (Pt 7)**: 1127-1138.

Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J et al. 2017. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res* **27**: 1384-1394.

Frias-Lasserre D, Villagra CA. 2017. The Importance of ncRNAs as Epigenetic Mechanisms in Phenotypic Variation and Organic Evolution. *Front Microbiol* **8**: 2483.

Fulop C, Szanto S, Mukhopadhyay D, Bardos T, Kamath RV, Rugg MS, Day AJ, Salustri A, Hascall VC, Glant TT et al. 2003. Impaired cumulus mucification and female sterility in tumor necrosis factor-induced protein-6 deficient mice. *Development* **130**: 2253-2261.

Gerdes P, Richardson SR, Mager DL, Faulkner GJ. 2016. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol* **17**: 100.

Golbus MS, Calarco PG, Epstein CJ. 1973. The effects of inhibitors of RNA synthesis (alpha-amanitin and actinomycin D) on preimplantation mouse embryogenesis. *J Exp Zool* **186**: 207-216.

Green SH, Zuckerman S. 1951. The number of oocytes in the mature rhesus monkey (Macaca mulatta). *J Endocrinol* **7**: 194-202.

Greenwald GS. 1972. Of eggs and follicles. *Am J Anat* **135**: 1-3.

Gurdon JB. 1962. Adult frogs derived from the nuclei of single somatic cells. *Dev Biol* **4**: 256-273.

Gutschner T, Hammerle M, Diederichs S. 2013. MALAT1 -- a paradigm for long noncoding RNA function in cancer. *Journal of molecular medicine* **91**: 791-801.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**: 223-227.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C et al. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* **28**: 503-510.

Halley P, Kadakkuzha BM, Faghihi MA, Magistri M, Zeier Z, Khorkova O, Coito C, Hsiao J, Lawrence M, Wahlestedt C. 2014. Regulation of the apolipoprotein gene cluster by a long noncoding RNA. *Cell reports* **6**: 222-230.

Hamatani T, Carter MG, Sharov AA, Ko MS. 2004. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell* **6**: 117-131.

Hamazaki N, Uesaka M, Nakashima K, Agata K, Imamura T. 2015. Gene activation-associated long noncoding RNAs function in mouse preimplantation development. *Development* **142**: 910-920.

Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**: 3016-3027.

Hancks DC, Kazazian HH, Jr. 2012. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* **22**: 191-203.

Hansen KR, Knowlton NS, Thyer AC, Charleston JS, Soules MR, Klein NA. 2008. A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. *Hum Reprod* **23**: 699-708.

Havecker ER, Gao X, Voytas DF. 2004. The diversity of LTR retrotransposons. *Genome Biol* **5**: 225.

Hawkins PG, Morris KV. 2010. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* **1**: 165-175.

Hayashi M, McGee EA, Min G, Klein C, Rose UM, van Duin M, Hsueh AJ. 1999. Recombinant growth differentiation factor-9 (GDF-9) enhances growth and differentiation of cultured early ovarian follicles. *Endocrinology* **140**: 1236-1244.

Horton JD, Goldstein JL, Brown MS. 2002. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J Clin Invest* **109**: 1125-1131.

Howe K Clark MD Torroja CF Torrance J Berthelot C Muffato M Collins JE Humphray S McLaren K Matthews L et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**: 498-503.

Huarte M. 2015. The emerging role of lncRNAs in cancer. *Nat Med* **21**: 1253-1261.

Jagarlamudi K, Rajkovic A. 2012. Oogenesis: transcriptional regulators and mouse models. *Mol Cell Endocrinol* **356**: 31-39.

Jarroux J, Morillon A, Pinskaya M. 2017. History, Discovery, and Classification of lncRNAs. *Adv Exp Med Biol* **1008**: 1-46.

Jukam D, Shariati SAM, Skotheim JM. 2017. Zygotic Genome Activation in Vertebrates. *Dev Cell* **42**: 316-332.

Juneja SC, Barr KJ, Enders GC, Kidder GM. 1999. Defects in the germ line and gonads of mice lacking connexin43. *Biol Reprod* **60**: 1263-1270.

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL et al. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**: 1484-1488.

Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* **9**: e1003470.

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci U S A* **114**: E1460-E1469.

Karlic R, Ganesh S, Franke V, Svobodova E, Urbanova J, Suzuki Y, Aoki F, Vlahovicek K, Svoboda P. 2017. Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. *DNA Res* **24**: 129-141.

Khodosevich K, Lebedev Y, Sverdlov E. 2002. Endogenous retroviruses and human evolution. *Comp Funct Genomics* **3**: 494-498.

Kim JH, Lee BB, Oh YM, Zhu C, Steinmetz LM, Lee Y, Kim WK, Lee SB, Buratowski S, Kim T. 2016. Modulation of mRNA and lncRNA expression dynamics by the Set2-Rpd3S pathway. *Nat Commun* **7**: 13534.

Kim VN. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nature reviews Molecular cell biology* **6**: 376-385.

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. 2010. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* **3**: ra8.

Kumar S, Hedges SB. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**: 917-920.

Lander ES Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Li D, Guo L, Deng B, Li M, Yang T, Yang F, Yang Z. 2018. Long noncoding RNA HR1 participates in the expression of SREBP1c through phosphorylation of the PDK1/AKT/FoxO1 pathway. *Mol Med Rep* **18**: 2850-2856.

Lintern-Moore S, Moore GP. 1979. The initiation of follicle and oocyte growth in the mouse ovary. *Biol Reprod* **20**: 773-778.

Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S et al. 2010. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* **42**: 1113-1117.

Maamar H, Cabili MN, Rinn J, Raj A. 2013. linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes Dev* **27**: 1260-1271.

Mahley RW, Innerarity TL, Rall SC, Jr., Weisgraber KH. 1984. Plasma lipoproteins: apolipoprotein structure and function. *J Lipid Res* **25**: 1277-1294.

Marino-Ramirez L, Lewis KC, Landsman D, Jordan IK. 2005. Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet Genome Res* **110**: 333-341.

Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. 2002. The human and mouse replication-dependent histone genes. *Genomics* **80**: 487-498.

Matzuk MM, Burns KH, Viveiros MM, Eppig JJ. 2002. Intercellular communication in the mammalian ovary: oocytes carry the conversation. *Science* **296**: 2178-2180.

McGee EA, Hsueh AJ. 2000. Initial and cyclic recruitment of ovarian follicles. *Endocr Rev* **21**: 200-214.

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**: 333-338.

Mendez R, Murthy KG, Ryan K, Manley JL, Richter JD. 2000. Phosphorylation of CPEB by Eg2 mediates the recruitment of CPSF into an active cytoplasmic polyadenylation complex. *Mol Cell* **6**: 1253-1259.

Mendez R, Richter JD. 2001. Translational control by CPEB: a means to the end. *Nature reviews Molecular cell biology* **2**: 521-529.

Meng L, Person RE, Huang W, Zhu PJ, Costa-Mattioli M, Beaudet AL. 2013. Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. *PLoS Genet* **9**: e1004039.

Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, Mitra S, Mohammed A, James AR, Hoberg E et al. 2015. MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. *Nat Commun* **6**: 7743.

Moore GP, Lintern-Moore S. 1978. Transcription of the mouse oocyte genome. *Biol Reprod* **18**: 865-870.

Moran VA, Perera RJ, Khalil AM. 2012. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res* **40**: 6391-6400.

Mouse Genome Sequencing C Waterston RH Lindblad-Toh K Birney E Rogers J Abril JF Agarwal P Agarwala R Ainscough R Alexandersson M et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520-562.

Mus E, Hof PR, Tiedge H. 2007. Dendritic BC200 RNA in aging and in Alzheimer's disease. *Proc Natl Acad Sci U S A* **104**: 10679-10684.

Nagy A. 2003. *Manipulating the mouse embryo : a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**: 635-640.

Oh B, Hwang S, McLaughlin J, Solter D, Knowles BB. 2000. Timely translation during the mouse oocyte-to-embryo transition. *Development* **127**: 3795-3803.

Orisaka M, Orisaka S, Jiang JY, Craig J, Wang Y, Kotsuji F, Tsang BK. 2006. Growth differentiation factor 9 is antiapoptotic during follicular development from preantral to early antral stage. *Mol Endocrinol* **20**: 2456-2468.

Otsuka F, Yao Z, Lee T, Yamamoto S, Erickson GF, Shimasaki S. 2000. Bone morphogenetic protein-15. Identification of target cells and biological functions. *J Biol Chem* **275**: 39523-39528.

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* **7**: 597-606.

Pepling ME. 2006. From primordial germ cell to primordial follicle: mammalian female germ cell development. *Genesis* **44**: 622-632.

Piko L, Clegg KB. 1982. Quantitative changes in total RNA, total poly(A), and ribosomes in early mouse embryos. *Dev Biol* **89**: 362-378.

Reddy P, Liu L, Adhikari D, Jagarlamudi K, Rajareddy S, Shen Y, Du C, Tang W, Hamalainen T, Peng SL et al. 2008. Oocyte-specific deletion of Pten causes premature activation of the primordial follicle pool. *Science* **319**: 611-613.

Reddy P, Shen L, Ren C, Boman K, Lundin E, Ottander U, Lindgren P, Liu YX, Sun QY, Liu K. 2005. Activation of Akt (PKB) and suppression of FKHRL1 in mouse and rat oocytes by stem cell factor during follicular activation and development. *Dev Biol* **281**: 160-170.

Richter JD. 1999. Cytoplasmic polyadenylation in development and beyond. *Microbiology and molecular biology reviews : MMBR* **63**: 446-456.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E et al. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**: 1311-1323.

Sakurai T, Sato M, Kimura M. 2005. Diverse patterns of poly(A) tail elongation and shortening of murine maternal mRNAs from fully grown oocyte to 2-cell embryo stages. *Biochem Biophys Res Commun* **336**: 1181-1189.

Sanchez DH, Gaubert H, Drost HG, Zabet NR, Paszkowski J. 2017. High-frequency recombination between members of an LTR retrotransposon family during transposition bursts. *Nat Commun* **8**: 1283.

Sanchez F, Smitz J. 2012. Molecular control of oogenesis. *Biochim Biophys Acta* **1822**: 1896-1912.

Schultz RM, Letourneau GE, Wassarman PM. 1979. Program of early development in the mammal: changes in the patterns and absolute rates of tubulin and total protein synthesis during oocyte growth in the mouse. *Dev Biol* **73**: 120-133.

Schultz RM, Stein P, Svoboda P. 2018. The oocyte-to-embryo transition in mouse: past, present, and future. *Biol Reprod* **99**: 160-174.

Singer MF. 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* **28**: 433-434.

Smit AF. 1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res* **21**: 1863-1872.

Smith CM, Steitz JA. 1998. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol Cell Biol* **18**: 6897-6909.

Sorensen RA, Wassarman PM. 1976. Relationship between growth and meiotic maturation of the mouse oocyte. *Dev Biol* **50**: 531-536.

Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthes P, Kokkinaki M, Nef S, Gnirke A et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell reports* **3**: 2179-2190.

St Laurent G, Wahlestedt C, Kapranov P. 2015. The Landscape of long noncoding RNA classification. *Trends Genet* **31**: 239-251.

Stebbins-Boaz B, Cao Q, de Moor CH, Mendez R, Richter JD. 1999. Maskin is a CPEB-associated factor that transiently interacts with eIF-4E. *Mol Cell* **4**: 1017-1027.

Sternlicht AL, Schultz RM. 1981. Biochemical studies of mammalian oogenesis: kinetics of accumulation of total and poly(A)-containing RNA during growth of the mouse oocyte. *J Exp Zool* **215**: 191-200.

Stitzel ML, Seydoux G. 2007. Regulation of the oocyte-to-zygote transition. *Science* **316**: 407-408.

Su WY, Xiong H, Fang JY. 2010. Natural antisense transcripts regulate gene expression in an epigenetic manner. *Biochem Biophys Res Commun* **396**: 177-181.

Su YQ, Sugiura K, Woo Y, Wigglesworth K, Kamdar S, Affourtit J, Eppig JJ. 2007. Selective degradation of transcripts during meiotic maturation of mouse oocytes. *Dev Biol* **302**: 104-117.

Suh N, Baehner L, Moltzahn F, Melton C, Shenoy A, Chen J, Blelloch R. 2010. MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Current biology : CB* **20**: 271-277.

Suh N, Blelloch R. 2011. Small RNAs in early mammalian development: from gametes to gastrulation. *Development* **138**: 1653-1661.

Sun Z, Yang S, Zhou Q, Wang G, Song J, Li Z, Zhang Z, Xu J, Xia K, Chang Y et al. 2018. Emerging role of exosome-derived long non-coding RNAs in tumor microenvironment. *Mol Cancer* **17**: 82.

Svoboda P. 2017. Long and small noncoding RNAs during oocyte-to-embryo transition in mammals. *Biochemical Society transactions* **45**: 1117-1124.

Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM. 2004. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol* **269**: 276-285.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM et al. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**: 534-538.

Tarlinton RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. *Nature* **442**: 79-81.

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**: 689-693.

Ulitsky I, Bartel DP. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**: 26-46.

Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. 2011. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**: 1537-1550.

van Biervliet JP, Rosseneu M, Bury J, Caster H, Stul MS, Lamote R. 1986. Apolipoprotein and lipid composition of plasma lipoproteins in neonates during the first month of life. *Pediatr Res* **20**: 324-328.

van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M. 2014. Extensive localization of long

noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* **15**: R6.

Vasiliauskaite L, Berrens RV, Ivanova I, Carrieri C, Reik W, Enright AJ, O'Carroll D. 2018. Defective germline reprogramming rewires the spermatogonial transcriptome. *Nat Struct Mol Biol* **25**: 394-404.

Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit-Mehouas S, Arnaud P, Tomizawa S, Andrews 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 Biol* **16**: 209.

Wang F, Li X, Xie X, Zhao L, Chen W. 2008a. UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. *FEBS letters* **582**: 1919-1927.

Wang HL, Sui HS, Liu Y, Miao DQ, Lu JH, Liang B, Tan JH. 2009. Dynamic changes of germinal vesicle chromatin configuration and transcriptional activity during maturation of rabbit follicles. *Fertil Steril* **91**: 1589-1594.

Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43**: 904-914.

Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* **41**: e74.

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R. 2008b. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**: 126-130.

Wassarman PM, Kinloch RA. 1992. Gene expression during oogenesis in mice. *Mutat Res* **296**: 3-15.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**: 539-543.

Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. 2001. Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol* **21**: 1429-1439.

Xue L, Cai JY, Ma J, Huang Z, Guo MX, Fu LZ, Shi YB, Li WX. 2013. Global expression profiling reveals genetic programs underlying the developmental divergence between mouse and human embryogenesis. *BMC Genomics* **14**: 568.

Yan C, Wang P, DeMayo J, DeMayo FJ, Elvin JA, Carino C, Prasad SV, Skinner SS, Dunbar BS, Dube JL et al. 2001. Synergistic roles of bone morphogenetic protein 15 and growth differentiation factor 9 in ovarian function. *Mol Endocrinol* **15**: 854-866.

Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J et al. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* **20**: 1131-1139.

Yang Y, Yang CR, Han SJ, Daldello EM, Cho A, Martins JPS, Xia G, Conti M. 2017. Maternal mRNAs with distinct 3' UTRs define the temporal pattern of Ccnb1 synthesis during mouse oocyte meiotic maturation. *Genes Dev* **31**: 1302-1307.

Ying Y, Liu XM, Marble A, Lawson KA, Zhao GQ. 2000. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol Endocrinol* **14**: 1053-1063.

Ying Y, Zhao GQ. 2001. Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Dev Biol* **232**: 484-492.

Yue F Cheng Y Breschi A Vierstra J Wu W Ryba T Sandstrom R Ma Z Davis C Pope BD et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**: 355-364.

Zhang K, Huang K, Luo Y, Li S. 2014. Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. *BMC Genomics* **15**: 845.

Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL. 2013. Circular intronic long noncoding RNAs. *Mol Cell* **51**: 792-806.

Supplemental Material

Table 1

PCR primers

name	sequence	note
006-154-GP1.fwd	GGCCACCATTTCTTCATTCAGA	lncRNA-OET-006-154 knock-out genotyping
006-154-GP1.rev	ACAAGGTCAAGGAGTTTTGTGG	
006-154-GP1M.fwd	TAGAAGCTTCCTGTGTTCCCTT	
006-154-2.fwd	AAACTGACCAGATTCACGAAGATC	lncRNA-OET-006-154 expression analysis
006-154-2.rev	TGTGAGTCTATTGATCTCTCCTCC	
006-154-7.fwd	GGACATGAAACAGTTCATGACTTC	
006-154-7.rev	GTGTTAAAGATGCGATCTCTCAGG	
019-199-GP1.fwd	CGCTGAGAGGCTAGCTATCA	lncRNA-OET-019-199 knock-out genotyping
019-199-GP1.rev	AGGGCTTTGTGGTGAAGAGT	
019-199-GP1M.fwd	TAGTCTGACGGGGAAAAGGC	
019-199-1.fwd	GTCTATACTTACGGCTTCCCAACT	lncRNA-OET-019-199 expression analysis
019-199-1.rev	CTTTCCTGGGCTTCTAGTCAGTTT	
019-199-2.fwd	CCTCAACTTAAGACTTGTGGACTC	
019-199-2.rev	CTGGAACACTGTCTCTAGCTGA	
019-199-4.fwd	GTGTGCAGGGGTTTGTTGAC	
019-199-4.rev	CGAATGGCCTGTTCTCACTTG	
007-157-GP1.fwd	CTGCACTTTCATCAATAACCTCAC	lncRNA-OET-007-157 knock-out genotyping
007-157-GP1.rev	TAGGTCCATCTCATGCATACTTTC	
007-157-GP1M.fwd	CTTGATCACTGAAAGTTTCTCCG	
007-157-1. fwd	CTCCAGTGACCAGCATCAAGTA	lncRNA-OET-007-157 expression analysis
007-157-1. rev	GTCTGTAGGCCTATCTTGAGTCC	
007-157-2. fwd	CAGCTTATGCAATAGCTACACACC	
007-157-2. rev	CTTCTAGCTACCTTCACAGCTACC	
T7.fwd	CTTTATAATACGACTCACTAT	Guide RNA template amplification
sg.rev	AAAAAGCACCGACTCGGTGCC	

guide RNAs for CRISPR-mediated deletion

name	sequence
lncRNA-OET-006-154 CR1 3'	TAATACGACTCACTATAGG**GCACAGAAATACATCTGG**GTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGCTTTTT
lncRNA-OET-006-154 CR3 5'	TAATACGACTCACTATAGG**GTTTCCCATGTGACATGC**GTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGCTTTTT
lncRNA-OET-019-199 CR1 3'	TAATACGACTCACTATAGG**GCAAGTGACTCCCCTCTG**GTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGCTTTTT
lncRNA-OET-019-199 CR4 5'	TAATACGACTCACTATAGG**GACCTTAGTCTATACTTA**GTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGCTTTTT
lncRNA-OET-007-157 CR1 3'	TAATACGACTCACTATAGG**TCTCAGCACAGCATCTTC**GTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGCTTTTT
lncRNA-OET-007-157 CR4 5'	TAATACGACTCACTATAGG**CGTCTTGAAGATCAGACT**GTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGCTTTTT

Publications

Ganesh S, Svoboda P. Retrotransposon-associated long non-coding RNAs in mice and men. *Pflugers Arch*. 2016; 468(6):1049-60. doi: 10.1007/s00424-016-1818-5. Review. PubMed PMID: 27044413.

Karlic R*, **Ganesh S***, Franke V, Svobodova E, Urbanova J, Suzuki Y, Aoki F, Vlahovicek K, Svoboda P. Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. *DNA Res*. 2017; 24(2):129-141. doi: 10.1093/dnares/dsw058. PubMed PMID: 28087610

* these two authors contributed equally

Franke V, **Ganesh S**, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J, Svobodova E, Ma J, Suzuki Y, Aoki F, Schultz RM, Vlahovicek K, Svoboda P. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res*. 2017 Aug;27(8):1384-1394. doi: 10.1101/gr.216150.116. PubMed PMID: 28522611.