



MONASH University

# Characterisation of testis- and ovary-specific long non-coding RNAs

Jennifer Xuelian Yang  
Master of Reproductive Sciences

A thesis submitted for the degree of Master at  
Monash University in 2015  
Department of Anatomy and Developmental Biology

# TABLE OF CONTENTS

COPYRIGHT NOTICE.....	1
ABSTRACT.....	2
DECLARATION.....	3
ACKNOWLEDGEMENTS.....	4
<b><u>INTRODUCTION</u></b>	
I. TESTIS DIFFERENTIATION.....	6
II. OVARY DIFFERENTIATION.....	8
III. MOLECULAR PATHWAYS OF SEX DEVELOPMENT.....	9
A. Testicular development	
B. Ovarian development	
C. Antagonisms	
IV. DISORDERS OF SEX DEVELOPMENT (DSD).....	14
V. NON-CODING RNAs.....	15
A. Overview	
B. Small non-coding RNAs	
VI. LONG NON-CODING RNAs.....	19
A. Overview	
B. Biogenesis	
C. Disease	
VII. FUNCTIONS OF LONG NON-CODING RNAs.....	23
A. Mechanism of function	
B. Gene regulation	

C. Imprinting	
VIII. LONG NON-CODING RNAs AND SEX DEVELOPMENT.....	25
A. Chromosome inactivation	
B. Gonad development	
X. SPECIFIC BACKGROUND FOR THE PROJECT.....	28
XI. HYPOTHESIS.....	29
XII. AIMS AND EXPECTED OUTCOMES.....	29
A. Aim 1	
B. Aim 2	

## **MATERIAL AND METHODS**

I. TISSUE COLLECTION AND PROCESSING.....	30
A. Embryos	
B. Sex genotyping	
C. Processing and embedding	
II. IN SITU HYBRIDISATION (ISH).....	31
A. Primers	
B. Cloning PCR of long non-coding RNA candidates	
C. DNA isolation	
D. DNA ligation	
E. Transformation	
F. Plasmid DNA	
G. M13 PCR	
H. <i>In vitro</i> transcription	

I. Paraffin section ISH	
III. QUANTITATIVE REVERSE TRANSCRIPTASE PCR (qRT-PCR).....	37
A. RNA extraction	
B. cDNA synthesis	
C. qRT-PCR and data analysis	
IV. EXPRESSION CONSTRUCTS.....	38
V. CELL CULTURE.....	40
A. Cell line transfection and cumate efficiency testing	
B. <i>In vitro</i> scratch assay	
C. MTS assay	
D. TUNEL assay	

## **RESULTS**

I. EXPRESSION ANALYSIS.....	46
A. Long non-coding RNAs are sexually dimorphically expressed in embryonic mouse gonads	
B. Quantification of long non-coding RNAs expression in embryonic mouse gonads showed sexually dimorphic pattern	
II. FUNCTIONAL ANALYSIS OF ONC3 OVEREXPRESSION.....	66
A. Relative efficacy of cumate activation of Onc3 expression was strongest at 24 hours for TM3 and GC2 cells	
B. Expression of Onc3 leads to reduced HEK293 and TM3 cell migration/proliferation and increased GC2 cell migration/proliferation	
C. Expression of Onc3 increases cell proliferation in TM3 cells and while no effect on proliferation was observed in GC2 cells	
D. Expression of Onc3 does not affect apoptosis in HEK293, GC2 and TM3 cells	

**DISCUSSION**

I. EXPRESSION ANALYSIS.....77

II. FUNCTIONAL ANALYSIS.....82

**FUTURE DIRECTIONS**

I. EXPRESSION ANALYSIS.....85

II. FUNCTIONAL ANALYSIS.....86

**CONCLUSION**.....88

**SUPPLEMENTARY DATA**.....89

**REFERENCES** .....100

**APPENDICES**.....102

## **Copyright notice**

© The author (2015). Except as provided in the Copyright Act 1968, this thesis may not be reproduced in any form without the written permission of the author.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

# ABSTRACT

Gonadogenesis is a highly regulated process in the mammalian fetus and disruption can lead to altered sexual development resulting in disorders of sex development (DSDs), which are congenital conditions in which the development of chromosomal, gonadal or anatomical sex is atypical. Many DSDs are currently unexplained at the molecular level, suggesting that genes and/or regulatory mechanisms that are important for gonad development are still unknown. There is emerging data that non-coding RNAs (ncRNA) play a role in many developmental processes. Long ncRNAs are ncRNAs defined as being longer than 200 nucleotides. Not much is known about the role of long ncRNA in regulation of testis and ovary development. Using microarrays techniques, we have identified a number of long ncRNAs, which displayed sexually dimorphic expression during mouse gonad development but no function has yet been assigned. In the current study I aimed to validate the expression of these long ncRNAs and elucidate their cellular function.

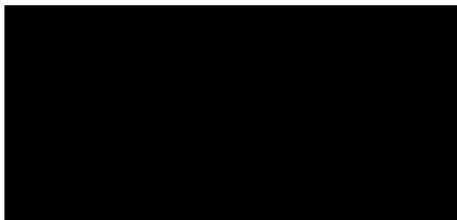
# DECLARATION

This thesis submitted to Monash University on the **29<sup>th</sup> of April 2015** is in fulfilment of Master of Reproductive Science and I, Jennifer Yang, confirm that it has not been submitted in whole or in part towards any other qualification.

I declare unless otherwise stated all work outlined herein has been completed solely by me. The supporting laboratory notebook for this project is located in the Wilhelm laboratory in the Department of Anatomy and Developmental Biology, Monash University.

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

**Signed:**

A large black rectangular box redacting the signature of the author.

# ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors for their guidance and support throughout the two years of my Masters course. To A/Prof Dagmar Wilhelm, you have always been an inspiration for me, both by the knowledge you hold and your passion for research. Thank you so much for your patience, encouragement and help you provided me, I really appreciate it.

To Dr Raphael Rastetter, thank you for always being there both in and out of the lab. The assistance you provided and the ideas you shared with me really helped me through my research. And also for willingness to listen and share experiences with me, you made my lab experience a very memorable and happy one.

To the Wilhelm Lab, the past members Kasia, who helped me heaps during the beginning of my Masters course, to Pascal, who was always very positive and solved many experimental difficulties I experienced, to Lisa, I really miss the days we worked together in the lab. To Simon and Sean, I want to thank you guys for making the lab a fun and lively environment. And to Michelle, your smile always brightens the day and I am sure you will achieve your goals.

To the people at Monash who helped me in various ways. To all the lovely people from department of Anatomy and Developmental biology, for your support and friendships throughout these two years. To Camilla, Steph and Jelena from the Monash histology platform, Judy from Monash imaging facility and Irene from Monash Biochemistry imaging facility, thank you all for the useful trainings and instructions you provided me in the lab.

And last but not least to my family, thank you so much for the support. It has been invaluable and without your support this two years would have been a lot more difficult.

# INTRODUCTION

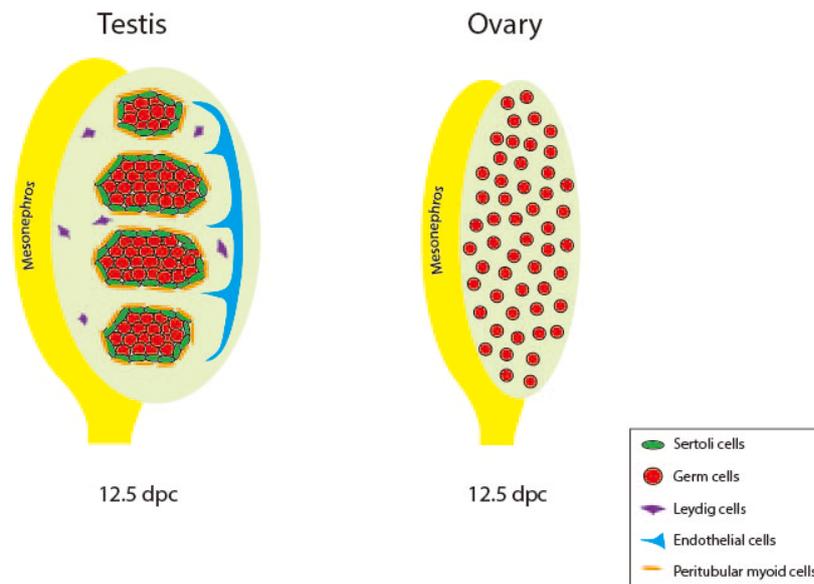
In 1947, French endocrinologist Alfred Jost defined sex development as a two-phase process: Sex determination and sex differentiation [1]. Jost performed surgery on rabbits *in utero* by either removing the ovary or the testis from embryos. The removal of the ovary did not affect female development, while the removal of the testis resulted in female instead of male development [1]. This suggested that the testis is dominant and determines male sex.

In mammals, the genetic sex is determined by the chromosomes, when either the X or the Y chromosome is inherited from the paternal side at the time of fertilisation. Several labs around the world tried to identify the factor on the Y chromosome that drives differentiation of the testis. To do so they examined three types of human patients: XX males with translocation from the Y chromosome, XY males with Y chromosome deletions and XY females. The groups first realised that the male determining factor is on the short arm of Y chromosome. After much effort, in 1990, Sinclair and colleagues showed that the gene *SRY* is always either mutated or deleted in XY females and that *SRY* is translocated onto an X chromosome in XX males, while it is intact in XY males that had other deletions on the Y chromosome [2], in combination with *Sry* transgenesis in mice, which was sufficient to cause XX sex reversal, ultimately showed that *SRY* is the male-determining gene on the Y chromosome in therian mammals [3].

The time point of sex determination is when the *Sry* gene is expressed at 11.5 days *post coitum* (dpc) in mice and between 41 and 44 days post ovulation in human [4, 5]. *Sry* acts as a switch that diverts the fate of the genital ridges towards testis development. If *Sry* is absent or non-functional, the genital ridges will develop into ovaries. Once testicular or ovarian differentiation has occurred, they produce the sex-specific hormones, which drive the differentiation of most, if not all, secondary sexual characteristics [6].

The precursors of the gonads are named the gonadal ridge. The bipotential precursors develop from the intermediate mesoderm which runs along either side of the midline structure of the embryo [7]. The pronephros, the mesonephros and the metanephros are located along the mesonephric duct. The mesonephros can function as a temporary embryonic kidney in mammals while the metanephros will give rise to the actual kidneys [8]. The genital ridges develop at the ventral surface of the mesonephros and are first visible at around 10 dpc in mouse. At around the same time, primordial germ cells (PGCs) arrive at the developing gonads. PGCs are not specified within the gonad, instead precursors are found in the epiblast near the extraembryonic ectoderm at around 6 dpc [9]. At 7 dpc, PGCs

are first seen in the region of the forming hindgut, they begin to migrate actively along the hindgut, dorsal mesentery and into the genital ridges by 10.5 dpc [10]. Males and females cannot be distinguished at this stage as the gonad is morphologically indifferent [11]. A few days later the testis cords will develop, which can be distinguished from an ovary through a microscope (Figure 1). Eventually the testis cords give rise to the seminiferous tubules in the adult testis. In comparison, the ovary appears to remain quiescent for the first few days.



**Figure 1. Murine testis (left) and ovary (right) morphology at 12.5 days post coitum (dpc).** Testis cords composed of primordial germ cells are surrounded by supporting Sertoli and peritubular myoid cells. Steroidogenic fetal Leydig cells are present in the interstitium. In contrast, fetal ovarian development starts at a more modest pace. At 12.5dpc, only germ cells and somatic precursor cell lineages are present.

## Testis differentiation

The bipotentiality that can be seen at the organ level is also true at the cellular level. Various testis-specific cell types are involved in testicular development. In the indifferent gonad, there are a number of cell precursors including supporting cells, steroidogenic cells and PGCs, which can enter either the male or the female pathway. In the male pathway, PGCs develop as prospermatogonia, which are the precursors of mature sperm [12]. The supporting Sertoli cells not only nurture and enclose PGCs, they are also the organisers of testis differentiation. They differentiate early in testicular development and drive the differentiation of other cell types [13]. Within testis cords, Sertoli cells are surrounded by a single layer of cells named peritubular myoid (PM) cells. PM cells function as muscles to

stabilise testis cord structure, and will later become contractile to support mature sperm movement through the seminiferous tubules into the epididymis in adult life [14]. Within the testicular interstitium, there are blood vessels, mesenchyme and steroidogenic cells. Sertoli cells express desert hedgehog (*Dhh*), which can program through its receptor Patched 1 steroidogenic precursor cells to differentiate into Leydig cells [15]. Foetal Leydig cells synthesise and secrete androgenic steroid hormones such as testosterone and insulin-like factor 3 (INSL3). Testosterone and 5-alpha-dihydroxytestosterone (DHT) are required for proper Wolffian duct development and external genitalia formation [16]. INSL3 is a member of the insulin peptide hormone family and is important for testicular descent [17].

In mammals, in order for a testis to develop, several essential testis-specific processes must occur before birth. In mouse, the gonads are hardly visible on top of the mesonephros where there are only a few cell layers at 10.5dpc. In the XY foetus at 11.5dpc, after *Sry* expression, cell migration occurs from the underlying mesonephros into the developing testis, followed by testis cord formation at 12.5dpc, steroidogenesis at 13.5dpc and lastly, differentiation of the Wolffian duct into male genital tract features such as epididymis, vas deferens and seminal vesicles by 14.5 to 15.5dpc [18].

Upon *Sry* expression, the Sertoli cells that are distributed evenly throughout the genital ridge secrete VEGFA (vascular endothelial growth factor A) to instruct the migration of endothelial cells from the mesonephric region into the differentiating testis, which takes place between 11.5 to 12.5dpc in mice [19]. The mesonephric cells do not migrate into the ovary hence cell migration is a testis-specific event. This process is required for the proper formation of the testis cords and the interstitial cell population [20]. Studies suggested that endothelial cells express cell surface markers and remodel the extracellular matrix during angiogenesis [14, 21]. When endothelial cell migration is inhibited, testis cord formation does not occur [22]. Meanwhile, *Sry* induces the differentiation of Sertoli cells and in its absence these cells become follicle cells. Sertoli cell differentiation and mesonephric cell migration together lead to the formation of cylindrical cord at 12.5 dpc [23]. Testis cord formation is the first morphological event of male sex determination [24]. Sertoli cells direct the encapsulated germ cells in the middle of the testis cords to undergo mitotic arrest at the  $G_0/G_1$  phase of mitosis from 13.5 dpc until after birth. After birth, germ cells resume proliferation. PGCs give rise to spermatogonia and then enter meiosis to form the primary spermatocytes, and later differentiate into sperm. The spermatogenesis process happens along the seminiferous tubule and mature sperm is then released and pumped into the epididymis [13].

## Ovary differentiation

In the absence of *Sry* expression the bipotential gonad develops into an ovary. In comparison to testis differentiation where massive morphological changes take place within a few days, the ovary appears to remain quiescent and looks almost the same for the first few days.

The mesonephros produces retinoic acid, which, at around 13.5dpc, induces the expression of retinoic acid gene 8 (*Stra8*), which is necessary to initiate germ cells to enter meiosis prophase I in an anterior to posterior wave [25-27]. The first morphological event of female development is the formation of germ cell cysts at around 14.5dpc in mouse [13]. Unlike testis, it is difficult to distinguish between different somatic precursor cell types in the ovary. Supporting somatic cells that express the transcription factor FOXL2 differentiate into pre-granulosa cells rather than Sertoli cells due to the absence of *Sry* expression [13]. Granulosa cells are the supporting cell lineage in the ovary. The steroidogenic precursor cells differentiate into theca cells after birth and produce testosterone, which is then converted to oestrogen by the granulosa cells [28]. In the absence of testosterone and the transforming growth factor- $\beta$  superfamily paracrine factor anti-Müllerian hormone (AMH), the Wolffian duct degenerates and the Müllerian duct gives rise to female genital tract features such as fallopian tubes, the uterus and the upper third of the vagina [29]. In addition, connective stromal cells and endothelial cells are also present in the foetal ovary [13, 30].

Oogenesis and folliculogenesis are intimately associated to produce the mature oocytes. Ovarian follicles differentiate only after birth. Shortly before birth, somatic cells break up the clusters of germ cells so that single oocytes are surrounded by a flat layer of granulosa cells to form primordial follicles [31]. Meanwhile, a large number of oocytes are lost through programmed cell death, which limits the number of primordial follicles [32]. During folliculogenesis, primordial follicles develop into primary follicles and are activated in preparation for oogenesis [13]. Until the preovulatory stage, the oocyte is arrested in prophase I of meiosis I and is not metabolically active. During late preovulatory stage, the oocyte starts growing and the proliferating granulosa cells form the secondary follicles with several layers of granulosa cells. In addition, stromal cells are recruited and undergoes differentiation into thecal externa and thecal interna. In tertiary stage follicles the antrum become visible and at the graffian follicles stage, cumulus cells wrap around the oocyte and the oocyte is ready for ovulation. After ovulation, collapsed follicles transform into the corpus luteum, made of theca and granulosa cells, which is important as it maintains pregnancy through progesterone production [33].

# **Molecular pathways of gonad development**

## **Testis development**

The most critical step during male sex determination in the majority of mammalian species is the expression of *Sry* gene. Three criteria must be fulfilled in order to be a testis-determining factor to switch on male development: the factor must be expressed in the bipotential gonad at the time when it starts to differentiate into a testis; XY reversal is expected when the factor is deleted or mutated and ectopic expression of the factor in a XX gonad will result in testis development. Gain and loss of function of *Sry* and the expression of *Sry* in mice and human were analysed [34, 35]. To prove that *SRY* is sufficient to drive male development, scientists mapped the mouse ortholog of *SRY* to generate transgenic mice by inserting the *Sry* gene with its own regulatory region into XX mice [3]. The XX mice developed externally as males, indistinguishable from a normal XY male. However, XX transgenic mice were infertile as other genes on the Y chromosome that are required for spermatogenesis were missing [36].

*SRY* stands for “sex determining region on the Y chromosome”. It is the founding member of the SOX transcription factor family. *SRY* is expressed in supporting cell lineages of human and mouse testis and contains a DNA binding domain named high mobility group (HMG) domain with two nuclear localisation signals (NLS). The HMG domain binds specifically to DNA primarily in the minor groove and thereby bends the DNA [37]. Binding and bending of DNA brings together other factors that activate the transcription of downstream genes. The majority of mutations found in *SRY*, which result in XY sex reversal, are within the HMG box [38]. *SRY* is not very well conserved between different species: the DNA binding domain from mouse and human is highly conserved while regions outside of this domain are not [39]. Hence the HMG domain is crucial and is the main domain in the *SRY* protein that is important.

The expression of *Sry* in the mouse is tightly regulated; it starts at 10.5dpc in the centre of the genital ridge and spreads out in a wave-like fashion towards the poles of the testis, then the expression reaches a peak at 11.5dpc, after which expression decreases in the same centre-to-pole manner. The last *SRY* positive cells are detected at 12.5dpc [4]. Expression of *Sry* must exceed a critical threshold level within 6 hours between 11.0 to 11.25dpc in order to activate the testis pathway or the genital ridges will develop as ovotestes or ovaries [40]. Within pre-Sertoli cells, Wilms tumor 1 (WT1+KTS), growth arrest DNA damage inducible 45 gamma (*GADD45 $\gamma$* ), GATA binding protein 4 (*GATA4*) /friend of GATA 2 (*FOG2*) and the insulin receptor family are examples of the factors that are expressed in the undifferentiated gonad prior to 11.5dpc and are involved in the regulation of *Sry* expression and [41-43].

During sex determination, SRY directly up-regulates the critical testis gene *Sox9*. Double immunofluorescence for SRY and SOX9 showed their expressions were almost completely overlapping [44]. Studies suggested that there are 3 stages of *Sox9* activation: initiation, up-regulation and maintenance. At the initiation stage, sex independently, orphan nuclear receptor steroidogenic factor 1 (SF1, also known as NR5A1) is expressed and binds to the promoter region of the *Sox9* gene, thereby leading to low levels of *Sox9* expression [45]. The initiation stage occurs in both XX and XY gonads. During the next stage: up-regulation, SRY in XY gonads, together with SF1, binds to the promoter and activates *Sox9* expression through the specific 3.2kb region named testis enhancer sequence core (TESCO), which is located 14kb upstream of the *Sox9* transcription start site [46]. Upon activation, SOX9 becomes rapidly upregulated and translocates to the nucleus. After SRY expression is down regulated, *Sox9* maintains its own expression through binding to its own promoter in Sertoli cells until after birth [46]. It has also been shown FGF9/FGFR2 and *Ptgds*/PGD<sub>2</sub> can maintain *Sox9* expression through positive feedback loops [47-49]. SOX9 acts as a transcriptional activator as it initiates and up-regulates the expression of a AMH and fibroblast growth factor 9 (FGF9) [47, 50]. The secreted signalling molecule FGF9 acts downstream of SOX9 and is essential for male sex development as it promotes cell proliferation [47]. XY mice mutant for FGF9 display sex reversal due to reduced *Sox9* levels, suggesting that FGF9 is involved in the up-regulation of *Sox9* expression [51]. Fibroblast growth factor receptor 2 (FGFR2) mediates FGF9 signals during male sex development to maintain *Sox9* expression [51]. Conditional inactivation of *Fgfr2* results in partial sex reversal [48, 52].

SOX9 is another SRY-like HMG domain transcription factor. Unlike SRY, SOX9 is highly conserved throughout vertebrates [53]. SOX9 is initially present in the indifferent gonad of both XX and XY fetuses. Shortly after the onset of *Sry* expression, *Sox9* only becomes up-regulated within pre-Sertoli cells in XY gonads. SOX9 also plays roles in other aspects of normal human and mouse development and disease. SOX9 mutation in human results in a disease called campomelic dysplasia, symptoms include bowing of the long bones and other skeletal defects, and hence SOX9 is important for bone formation [54]. Interestingly, 2 out of 3 XY babies who have mutations in SOX9 also show male-to-female sex reversal, which indicates that SOX9 is not only important for bone formation but also important for male development [55]. This observation in human was recapitulated in mouse models, where a study showed that inactivation of *Sox9* disrupts the testis pathway shown by loss of Sertoli cells, Leydig cells and lack of testis cords development in XY gonads, meanwhile, ovarian development was observed in *Sox9* mutant mice after the up-regulation of female markers forkhead box L2 (*Foxl2*) and wingless-related MMTV integration site 4 (*Wnt4*) [56]. In

addition, craniofacial skeletal defects were observed in heterozygous *Sox9* knockout mice [57]. Like *SRY*, *SOX9* is necessary and sufficient for testis development. Studies have shown that XX mice transgenic for *Sox9* are XX sex reversed and developed as infertile male [36, 58].

Apart from *Sox9* and *Sry*, ectopic expression of genes such as *Sox3* in ovary has also been shown to result in sex reversal [59, 60]. *Sox3* is almost identical to the *Sry* gene and it has been suggested that *Sry* and *Sox3* have a common ancestor [61]. *SOX3* is expressed in embryonic gonads as well as in the adult ovary and testis. Loss-of-function mutations of *SOX3* have shown that while it is not required for sex determination but for gonadal function [62]. In addition, duplications of *SOX3* in male patients result in pituitary defects [63, 64] and microarray data has shown duplications or deletions of *SOX3* in XX males lead to rearrangements, which cause *SOX3* to be expressed ectopically in the developing gonad so *SOX3* can act in the absence of *SRY* [65]. *Sox3* transgenic mice develop as males with small testes [66]. Moreover, another member of the *SOX* family, *SOX10*, which is closely related to *SOX9*, has been shown to be involved in neural crest and glial development. Transgenic mice overexpress *Sox10* in XX and XY gonads resulted in XX sex reversal, supporting the idea that human *SOX10* might play a role in 46, XX DSD [59, 67].

## Ovarian development

If *SRY* is not present in the supporting cell lineage during the brief crucial period, sustained *SOX9* expression is not established, and the female regulatory network predominates and drives ovarian development. Many pro-ovarian genes are involved in the regulation of the development of an ovary. R-spondin1 (*Rspo1*),  $\beta$ -catenin, *Wnt4*, *Foxl2*, follistatin and dosage sensitive sex reversal-adrenal hypoplasia congenital on X gene 1 (*Dax1* or *NR0B1*) have all been implicated in early development of the ovary, however none of these can be considered as *Sry*-equivalent ovarian determining gene. In contrast, there are at least two major and largely independent pathways. Disruptions to both pathways result in more severe phenotypes when compared with disruption of a single pathway [68-70].

*Rspo1*, *Wnt4* and  $\beta$ -catenin play a role in one of the two pathways that drive ovarian differentiation. The *Rspo1* gene encodes *RSPO1*, which activates the canonical WNT/ $\beta$ -catenin signalling pathway. XY mice with loss of *Rspo1* developed normally, while deletion of *Rspo1* in XX mice resulted in ovotestis formation [69, 71]. In contrast, mutations of *RSPO1* in human resulted in complete XX sex reversal [72].  $\beta$ -catenin is the key intracellular mediator of the canonical WNT pathway and acts antagonistically to testis differentiation.

Activation of  $\beta$ -catenin inhibits Sox9 transcription and reduces the binding of SF1 on the TESCO enhancer [73].

WNT4 is a secreted extracellular signalling protein that regulates embryonic ovarian development by binding to G-protein coupled receptors of the Frizzled family [74]. *Wnt4* is expressed in both XX and XY genital ridges before the start of testis and ovary differentiation at 9.5dpc, then becomes ovary-specifically expressed after 11.5dpc [75]. Like RSPO1, WNT4 also functions through the stabilisation of  $\beta$ -catenin, which allows it to be translocated into the nucleus, hence increases the transcription of target genes [76].  $\beta$ -catenin activation is important for the differentiation of the germ cells in the ovary to enter meiosis and also germ cell survival [69]. *Rspo1* acts upstream of *Wnt4* and regulates *Wnt4* expression via  $\beta$ -catenin, however the presence of *Rspo1* is not necessary for *Wnt4* function, as *Wnt4* expression is only reduced, but not completely gone in *Rspo1* knockout mice [69]. It is important to inhibit testis-specific events including the formation of the a testis-specific vasculature. Loss of *Wnt4* or *Rspo1* triggers cell migration and coelomic vessel formation in the ovary [77]. However, null mutation of *Wnt4* in XX mouse and overexpression of *Wnt4* in XY mouse only result in partial XX sex reversal, which indicates that although *Wnt4* is important it is not the main factor that induces ovarian differentiation [69, 71].

WNT4 upregulates the glycoprotein follistatin, a well-known activin B inhibitor. Follistatin is specifically expressed in ovaries; it encodes a TGF $\beta$  superfamily binding protein, which binds to activin B with high affinity [78]. Follistatin acts antagonistically to the testis pathway by inhibiting the formation of coelomic vessel, while maintaining germ cell survival in the ovary [79].

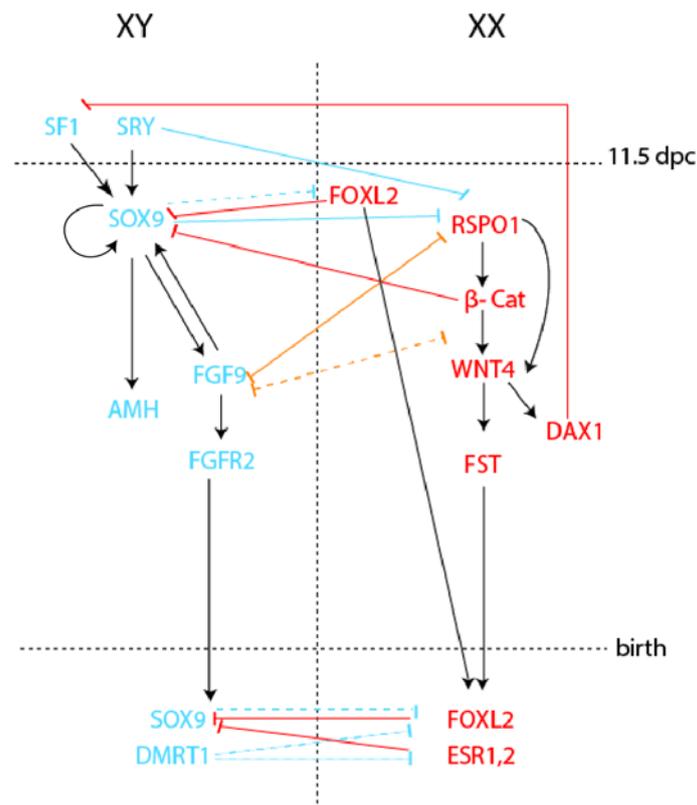
In parallel to *Wnt4/Rspo1* pathway, *Foxl2* is the key player of the second important pathway in ovarian differentiation. FOXL2 is a forkhead transcription factor involved in eyelid and ovarian development and function [80]. It is expressed in the somatic cells of the foetal ovary from 12.5dpc and at high levels in all granulosa cells in the adult ovary, but never in the testis [81]. Loss-of-function experiments showed species-specific requirements: in goat the loss of *Foxl2* leads to complete XX sex reversal, while in human loss of *FOXL2* causes blepharophimosis, ptosis and epicanthus inversus syndrome (BPES), associated with premature ovarian failure, which also can be observed in the *Foxl2*-null mouse [82, 83].

Lineage tracing experiment have shown FOXL2-positive cells become adult granulosa cells of the medullary follicles in the mouse ovary, which are activated shortly after birth and are then lost and only contribute to puberty and early fertility [84]. Cortical follicles are the ones

that determine the fertility for life and are activated throughout adulthood [85]. The adult stem cell marker LGR5 marks somatic cells in the foetal ovary that give rise to cortical granulosa cells of the adult ovary [86].

## Antagonisms

There are a number of interplays between testis and ovary both at the embryonic and the adult stages (Figure 2). In the embryonic testis pathway, SOX9 and FGF9 can suppress the key players of the ovarian pathway, while they are repressed in the ovary [47]. *Fgf9* is repressed by WNT4 in the female pathway to prevent testis differentiation, while endogenous FGF9 can suppress *Wnt4* to drive *Sox9* expression to reach threshold level in the male pathway [47].



**Figure 2. Genes involved in sex development during embryogenesis until after birth.** Testis-specific genes (blue) and ovary-specific genes (red) cross repress the pathways of the opposite sex to maintain gonadal phenotype. Antagonistic interactions between molecular players of the testis and ovary pathways are indicated by blue lines for the repression of female genes by male genes and red lines are vice versa. Orange lines show bidirectional influence from both XX and XY pathways. Dotted lines suggest possible antagonistic relationship between the genes.

In a fully differentiated mouse testis or ovary, the suppression is maintained throughout life. Doublesex and mab-3 related transcription factor 1 (DMRT1) is the protein essential to

maintain a testis. *Dmrt1* is expressed in the developing gonads with higher level within testes compared to ovaries [88]. Postnatally, DMRT1 represses the expression of FOXL2 and estrogen receptor 1 and 2 in the testis. In addition, it can also prevent the differentiation of granulosa cells by blocking retinoic acid signalling in Sertoli cells [89, 90]. In mice, deletion of *Dmrt1* during foetal development induces postnatal feminisation of the testis causing male-to-female sex reversal, while deletion of *Dmrt1* in the adult testis reprograms Sertoli cells into granulosa cells [90, 91].

On the other hand, a target gene of WNT4 signalling, *Dax1*, can act as an anti-testis gene by repressing SF1 and therefore preventing the up-regulation of *Sox9* [92]. It was initially thought that *Dax1* is a potential candidate ovary-determining gene, as duplication of human *DAX1* within the region Xp21 on the X chromosome causes male-to-female sex reversal [79, 93]. In addition, studies suggested *Dax1* also regulates testis cord organisation during testis development [94].

Moreover, *Foxl2* maintains mammalian ovary fate postnatally. In the ovary, FOXL2 normally binds to TESCO, repressing SOX9 and SF1 action. When *Foxl2* is deleted from a fully differentiated mouse ovary, TESCO is activated and therefore *Sox9* is upregulated, resulting in the transdifferentiation of granulosa cells and theca cells into Sertoli and Leydig cells respectively, as the result the ovary will acquire testis-like structures [95].

## **Disorders of Sex Development (DSD)**

Correct gonad development is the basis to fertility, which is determined early in fetal life. Factors that have a negative impact on early fetal gonadal development will lead to aberrant sexual development and therefore testicular and ovarian defects later in life. In the 19<sup>th</sup> century, people with aberrant sexual development were called hermaphrodites, however hermaphrodites only refer to people with both male and female organs [96]. Later on the name was changed to intersex and more recently, scientists have renamed it to disorders of sex development (DSDs). DSDs are defined as “a group of congenital conditions in which development of chromosomal, gonadal or anatomic sex is atypical” [97]. DSDs describes a range of conditions which affects gonadal development. Most of DSD patients are infertile and they are susceptible to ovarian or testicular cancer. In the general population, it is estimated that DSDs affects approximately 1 in 4500 to 5000 live births [97].

DSDs are very complex disorders, psychologists, surgeons and social workers work together to support the families affected by DSD. DSD can happen at fertilization, during sex determination and differentiation. Unfortunately, currently many types of DSDs are unexplained at the molecular level, therefore it is important to understand the factors and mechanisms that drive sexual development. For example, the cause of approximately 70% of 46,XY gonadal dysgenesis is still unknown, while 15% is due to mutation to *SRY* gene and another 15% have *SF1* mutation [100, 101]. Most of 46,XX testicular DSD can be explained by translocation of the *SRY* gene onto an X chromosome, very few cases are due to duplications of *SOX9* or mutations of *RSPO1*, and the remaining 10% of causes still remain unknown [102-104]. Klinefelter's syndrome, also known as 47,XXY, is another common DSD. Patients with this syndrome have at least one extra X chromosome in their karyotype [99]. Androgen insensitivity syndrome is a common type of DSD in boys, occurring in approximately 1 in 13,000 births. Although patients with this syndrome have external female genitalia, their testosterone level is often high and development of undescended testes is common [97]. In the case of persistent Müllerian duct syndrome, the Müllerian ducts remains in XY individuals due to the absence of anti-Müllerian hormone (AMH) expression or its receptor AMHRII [98]. Improved diagnosis of patients as well as the discovery of new genes that are linked to DSD is crucial. Novel DSD genes can be identified through copy number variation (CNV) arrays [105]. In order to understand the causes of DSD, it is also essential to know how sex is determined and how sex differentiation takes place. One class of regulatory mechanisms that has not been studied in great details are long non-coding RNAs.

## **Non-coding RNAs**

### **Overview**

Since the discovery of nucleic acids by Swiss scientist Friedrich Miescher in 1868, studies on ribonucleic acid (RNA) soon began in the early 1900s [106]. It took more than 50 years to identify the chemical and biological differences between DNA and RNA [107]. During the late 1950s, the concept of messenger RNA emerged, from which Francis Crick, the English scientist who discovered the structure of deoxyribonucleic acid (DNA), developed the "central dogma of molecular biology" to describe the flow of genetic information in all cells: DNA is transcribed into RNA and RNA is translated into protein [108].

For decades, the only known biological catalysts were proteins. RNA was assumed to be the essential bridge between DNA and protein, acting as a template for protein synthesis [109].

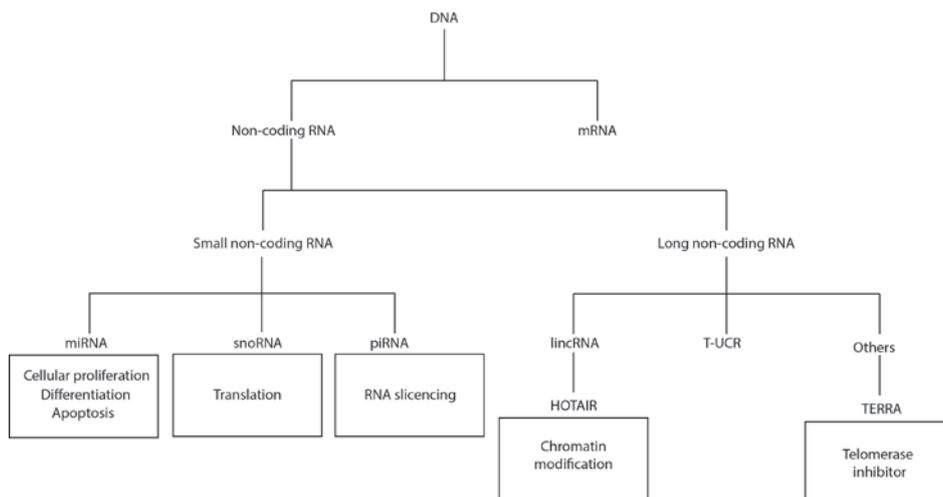
Proteins are the fundamental components of cells; they have a wide range of functions in almost every cellular process within a living organism. The protein centric view of “DNA makes proteins through temporary intermediate RNA” often leads to the assumption that most genetic information, including regulatory information is transacted by proteins. In the 1980s, Thomas Cech and Sidney Altman showed that certain RNAs function as enzymes, so-called ribozymes, which are involved in the processing of other RNA molecules such as the biogenesis of transfer RNA or splicing of nuclear pre-messenger RNA [110-112]. Nevertheless, the fascinating field of protein biology attracted molecular scientists around the globe to focus their research on the properties and biological functions of protein-coding RNAs or messenger RNAs (mRNAs), which acts as a template for protein synthesis through directing amino acid sequences of polypeptides in translation [113].

It is possible that the structure of genetic programming in humans and other complex organisms was fundamentally misunderstood for the last 50 years. The assumption that most genetic information is transacted by proteins may be untrue. Data of non-coding sequences emerged around 10 years ago. The Japanese Genome Network Project Core Group from RIKEN, Yokohama conducted global transcriptome sequencing in 2005, the surprising result was there were thousands of transcripts with little or no coding potential [114]. It turned out that only 1-2% of the human genome is protein-coding [115, 116], for some time the rest was considered to be “junk” DNA. Interestingly, it became apparent that this “junk” DNA is not just non-coding regions, but most of it is transcribed into RNA and possesses huge developmental and physiological importance [117-119]. Many of these RNAs appear to have no or little coding potential, raising the hypothesis that they are non-functional. These RNAs were given the name non-coding RNAs (ncRNAs) and are defined as RNAs that are not translated into a protein. It was originally hypothesised that ncRNAs do not participate in gene regulation. However, a rapidly increasing number of specific ncRNAs has been identified as the key regulators of many biological processes, including regulation of gene expression, cell cycle control, apoptosis, cell identity decisions, chromatin remodelling, and epigenetic modifications [120-122].

Additionally, the increasing number of functional ncRNAs in complex organisms may be the answer to the inconsistent correlation between the complexity of an organism, its cellular DNA content and the number of protein coding genes, respectively [122, 123]. The nematode *C. elegans* has around 1000 somatic cells, whereas humans have more than 100 trillion cells. Comparison between *C. elegans* and humans revealed a similar number of

protein-coding genes while there was an increase in the number of ncRNAs from worm to human. It appears that more complex organisms express more non-coding RNAs [123]. Thus, it is clear that the complexity of organisms cannot be explained based on the number of proteins but rather, at least partly, on the complexity of RNA.

The eukaryotic transcriptome is incredibly complex, it is composed of interlacing and overlapping of coding and non-coding transcripts on both strands, and the transcriptional landscape of any part of the genome is different in every cell [124]. ncRNAs include “house-keeping” RNAs such as ribosomal RNA (rRNA) and transfer RNA (tRNA), as well as regulatory RNAs. Regulatory RNAs are categorised, rather arbitrarily, according to their transcript length into small, shorter than 200 nucleotides (nt), and long ncRNAs (>200 nt). Members of small regulatory ncRNAs include microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs) and endogenous small interfering RNAs (endo-siRNAs) [125]. They are characterized by the biogenesis pathway, their length and their interaction with different proteins of the Argonaute family [126]. In contrast, long ncRNAs are a diverse class of mRNA-like non-coding transcripts that participate in a variety of biological processes through numerous mechanisms including chromatin modification, regulation of the activity or localization of proteins, organizational and structural frameworks, and as precursors for small ncRNAs.



**Figure 3. The family tree of regulatory non-coding RNAs.** Examples of small and long regulatory ncRNAs are listed with their functions.

## Small non-coding RNAs

Amongst different ncRNAs, the most studied small non-coding RNAs are miRNAs. miRNAs are a large family of single stranded ~21 nucleotide long RNA molecules [120]. Studies indicated that miRNAs act as negative gene regulators and are involved in almost every cellular process in both plants and animals, including cellular proliferation, differentiation and apoptosis [127]. miRNAs are estimated to control the activity of more than half of all mammalian protein-coding genes. miRNAs regulate gene expression by destabilising and repressing target RNAs at the post-transcriptional level [128, 129].

Another small non-coding RNA class consists of PIWI-interacting RNAs (piRNAs) [130]. piRNAs are 24 to 30 nucleotide long, single stranded small RNAs that mainly span along 50 to 100 kb in regions of genome that contain transcribed transposable elements and other repetitive elements [131]. piRNAs utilise the RNA interference (RNAi) machinery to repress transposon expression and mobilisation [132].

Like miRNA and piRNA, siRNA are small silencing RNAs that directs RNA interference (RNAi) to mediate post-transcriptional gene silencing in the cytoplasm [133]. In *Drosophila*, most siRNAs are initially derived from long exogenous double stranded RNAs, which is the main silencing trigger for siRNA biogenesis. siRNAs are encoded in many regions including introns of both coding and non-coding transcripts. In mammals, endogenous siRNA (endo-siRNA) have roles in oocyte maturation and transposon repression in the germ line [134].

Small nuclear RNAs (snoRNA) are intronic ncRNAs of 60 to 300 base pairs long. After transcription, pre-snoRNAs undergo splicing, debranching and exonucleolytic degradation to form mature snoRNAs. Together with small nucleolar ribonucleoprotein (snoRNP) core proteins, mature snoRNAs form part of snoRNP. snoRNP are thought to guide a multitude of RNA processing events, they can be either in the nucleus or in the nucleolus. snoRNP in the nucleus participate in alternative splicing of rRNA to facilitate rRNA folding and stability, while snoRNP in the nucleolus are involved in rRNA modification such as sequence-specific 2'-O-methylation and pseudouridylation of pre-rRNA [135-137].

# Long non-coding RNAs

## Overview

Long ncRNAs are a group of diverse heterogeneous ncRNAs. Within the mammalian genome, long ncRNAs are found in widespread loci with an expected number of tens of thousands, which likely make long ncRNAs the largest portion of the mammalian non-coding transcriptome [121]. Based on their position in the genome, long ncRNAs were categorized in three classes, intronic, natural antisense transcripts (NATs) and intergenic (large, intergenic ncRNAs or lincRNAs). However, this classification does not hold true when it comes to their functions. Long ncRNAs have a wide range of functions [121, 138], including the regulation of gene expression in *cis* and in *trans*, the regulation of epigenetic chromatin modification, post-transcriptional processes as well as structural functions. Well-known examples of long ncRNAs include *XIST*, *AIR* and *HOTAIR* [138, 139].

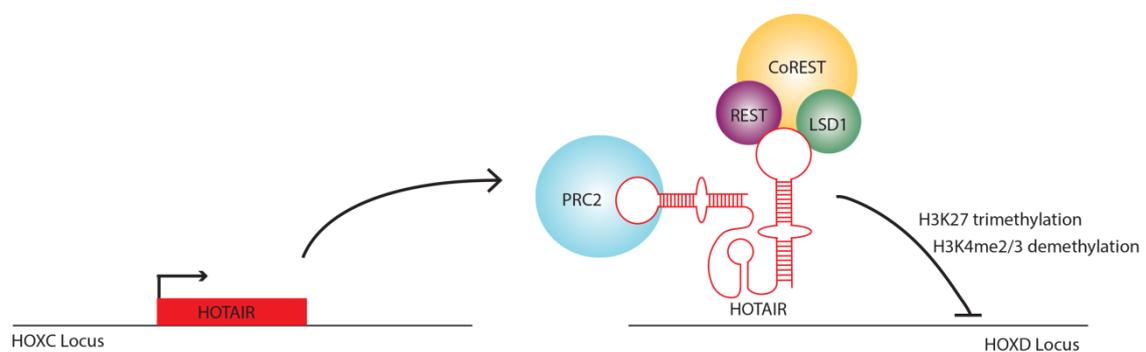
As a recently discovered class of ncRNA, long ncRNAs originate throughout the genome in regions such as promoters, enhancer sequences or 3'UTRs [140]. Their expression is very dynamic, for example long ncRNAs are found to participate in embryonic stem cell differentiation, T cell activation and muscle development [140-142]. Moreover, a number of long ncRNAs have been implicated in diseases such as breast and prostate cancer. Studies have shown aberrant transcription of long ncRNAs is a trigger of cancer development [143].

Most of the long ncRNAs can be effectively identified through strategies such as direct sequencing and microarray analysis [144]. The step after identification is the validation of the biological relevance of these long ncRNAs. Currently the functions for most long ncRNAs are still under research. Known long ncRNAs are involved in various important biological processes [121, 138]. Some long ncRNAs function in *cis* by regulating neighbouring protein-coding genes. Examples include *Xist* and *AIR* [139, 145]. While other long ncRNAs such as *HOTAIR* act in *trans* by regulating distant genes through epigenetic chromatin modification [146]. Furthermore, long ncRNAs are also involved in a range of transcriptional and post-transcriptional processes [147-149].

One example of long ncRNAs are large intergenic non-coding RNAs (lincRNAs). As their name suggests, lincRNAs are transcribed from intergenic regions. Currently, around 1,000 lincRNAs have been discovered [150]. LincRNAs are capped and polyadenylated, they are produced by the splicing of multi exon precursors that are transcribed by RNA polymerase II, and their genes carry chromatin markers commonly associated with protein-coding transcription units. One of the widely studied lincRNAs is trans-acting human homeobox

antisense intergenic RNA (*HOTAIR*). *HOTAIR* is transcribed from the *HOXC* locus and mediates epigenetic changes at the *HOXD* locus through recruitment of polycomb chromatin remodelling complex 2 (PRC2) (Figure 2). PRC2 then silences 40 kilobases of *HOXD* transcription [146]. This is a classic example of long ncRNA-mediated modification of chromatin [151]. Other than chromatin modification, lincRNAs also have roles in translational regulation. LincRNA p21 can act as a translation inhibitor for *JUNB* and *CTNNB1* mRNA. In response to DNA damage, p21 participate in the p53 mediated apoptosis pathway as a transcriptional repressor [152]. X chromosome inactivation mediators *Xist* and *Tsix* are also considered to be lincRNAs [150]. A recent knockout study suggested lincRNAs have important functions in not only embryonic development, but also in a range of tissues and organs. Some lincRNA knockout lines are lethal, while others have aberrant morphology in brain, lungs and skeleton [153].

HOTAIR in gene regulation



**Figure 4. Trans-acting ncRNA *HOTAIR* in gene regulation.** *HOTAIR* RNA is expressed at *HOXC* locus and acts as a molecular scaffold between PRC2 and LSD1-CoREST-REST complex. This in turn silences *HOTAIR* target genes on other locus such as *HOXD*.

Another type of long ncRNAs is called transcribed from ultraconserved regions (T-UCRS). UCRs are regions of DNA that are longer than 200bp and are conserved between human, rat and mouse [154]. More than half of UCRs are transcribed into T-UCRS. Currently more than 350 T-UCRS have been identified. The functions of T-UCRS are still under investigation. A recent study suggested T-UCRS might be involved in the regulation of miRNA and mRNA levels [155].

Many long ncRNAs do not fit into either lincRNAs or T-UCRS. One example is telomeric repeat containing RNA (*TERRA*), which act as a direct inhibitor of human telomerase [156-158].

## Biogenesis

In general, long ncRNAs are transcribed by RNA polymerase II and display the hallmarks of protein-coding genes. This includes the conservation of their chromatin structure, promoters, their regulation of expression by transcription factors and morphogens, their range of half-lives, tissue-specific expression, and splicing including alternative splice variants [125, 146, 159-161].

In addition, many long ncRNAs show the same characteristics as protein-coding mRNAs, they are spliced, 5'end capped and 3'end polyadenylated [162-164]. However, more recently through the development of more sophisticated approaches, alternative long ncRNA structures have been identified. Instead of endonucleolytic cleavage by CPSF73 and subsequent polyadenylation, which is necessary for transcript stability, some long ncRNAs possess a triple-helical structure at their 3'end, which protects them from degradation [165, 166]. This triple helical structure is formed by cleavage through RNaseP, which removes a conserved tRNA-structure from the 3'end, leaving a short A-rich tract that can form stable U-A.A triple helical structure [165, 166]. Some well-studied examples for long ncRNAs with a triple helical 3'end include *MALAT1* (metastasis-associated lung adenocarcinoma transcript 1), *NEAT1*, also called multiple endocrine neoplasia and RNAs from various viruses [167, 168].

A second alternative structure, which also results in stable transcripts, is utilized by some long ncRNAs derived from excised introns [169, 170]. As mentioned before, snoRNAs are approximately 70 to 200nt long, therefore are generally not classified as long ncRNAs. However, many introns contain two snoRNA-like sequences at their ends. Processing of these ends by the snoRNP machinery after splicing without removal of the sequences in between results in intronic long ncRNAs with snoRNA-like ends that are widely expressed in the human genome [171].

A third alternative mechanism of stabilisation is circularisation. Circular RNAs (circRNAs) can be formed by two different mechanisms. First, circRNAs are formed by head-to-tail joining through back-splicing, by which the 5'end acceptor site is spliced to a downstream 3'end donor site [172-175]. Secondly, circular intronic RNAs can derive from excised lariat introns [176]. Thousands of circRNAs have been identified in human cells [173, 176]. As already mentioned, some of these, especially those formed through back-splicing, display an enrichment for miRNA binding sites have been proposed to function as miRNAs sponges, indirectly regulating gene expression [174, 175]. In contrast, some intron-derived circRNAs accumulate at their site of transcription and positively regulate RNA polymerase II

transcription at the elongation step [176]. However, it remains to be seen, if all circRNA function through these mechanisms or if they can play a role through a plethora of different mechanism similar to linear long ncRNAs.

## Long ncRNAs and associated diseases

Long ncRNAs expression correlate to a number of diseases such as cancer. To date, lincRNAs, T-UCRs, psedogenes, enhancer RNAs and antisense RNAs have been demonstrated to have specific roles in carcinogenesis, many have been shown to have abrrated gene expression, translocation or deletion in cancer [155, 177-180]. Through high-throughput technologies such as next generation sequencing (RNA-Seq) [181], a significant number of long ncRNAs was found to be linked to lung, prostate, renal, breast and ovarian cancer [182, 183].

One of the best-studied long ncRNAs in cancer is *MALAT1*. This highly conserved nuclear long ncRNA act as a biomarker and oncogene for lung cancer through regulating expression of metastasis associated genes, cell migration and invasion of cancer cells [184-187].

*HOTAIR* is overexpressed in breast and hepatocellular carcinomas. This change of expression can predict subsequent metastasis of cancer [151, 177]. *HOTAIR* triggers epigenetic repression of PCR2 targets genes through increased recruitment of PRC2 [188]. Unlike *HOTAIR*, prostate cancer-associated transcript 1 (*PCAT1*) is repressed by PCR2 [189]. Studies have demonstrated *PCAT1* is selectively upregulated only in prostate cancer, it functions as a transcriptional repressor through trans-regulation of tumor suppressor genes such as *BRCA2* [190].

Similar to protein coding genes, long ncRNAs can function as tumor oncogenes and tumour suppressors. A well studied tumour suppressive long ncRNAs is *GAS5*. In breast cancer, *GAS5* binds to the glucocorticoid receptor to prevent gene expression, and at the same time induces apoptosis and growth arrest [191, 192].

*H19* is a long ncRNA involved in imprinting, interestingly it has been shown to be both oncogenic and tumor suppressive in breast and hepatocellular cancer. On the one hand, it promotes cell growth and proliferation, while on the other hand prolonged cell proliferation downregulates *H19* [193-195].

Apart from the well known long ncRNAs, in recent years many more long ncRNAs has been shown to participate in cancer-associated pathways. A T-UCR named *TUC338* has been

shown to act as an oncogene to regulate gene expression and promote cell proliferation and colony formation in hepatocellular carcinoma cell lines [196]. A member of *HOXA* genes, *HOXA9*, act as oncogene in human MLL-rearranged leukemia [197]. In 2010, Poliseno and colleagues found that the pseudogene *PTENP1* acts as a tumour suppressor through binding PTEN-suppressing miRNAs in prostate and colon cancer [198].

Long ncRNAs also plays a role in neurodegeneration. Long ncRNA *17A* and *BACE1-AS* are both related to Alzheimer's disease by regulating either the stability or inducing non-functional alternative splice isoforms of their target mRNA [199-201].

**Table 1. Examples of long ncRNAs implicated in oncogenesis.**

Long ncRNA	Function	Cancer type	Reference
<i>MALAT1</i>	Oncogenic	Lung, prostate, breast	185, 186, 187, 188
<i>HOTAIR</i>	Oncogenic	Breast, hepatocellular	152, 178, 189
<i>PCAT1</i>	Oncogenic	Prostate	190, 191
<i>GAS5</i>	Tumour suppressive	Breast	192, 193
<i>H19</i>	Oncogenic and tumour suppressive	Breast, hepatocellular	194,195,196

## **Functions of long non-coding RNAs**

### **Mechanisms of long ncRNA function**

Long ncRNAs play a role in many different processes, both at the molecular level such as splicing, transcriptional and epigenetic regulation of gene expression, and at the cellular level, for example in the regulation of proliferation, differentiation and apoptosis. Different groupings have been suggested for these functions, such as repressive, activating, and structural, or functioning in *cis*, in *trans* and as a decoy, or nuclear vs. cytoplasmic. However, the more we learn about the different mode of actions of specific long ncRNAs the more the boundaries of these categories do not hold true anymore. For example, some long ncRNAs function as activators as well as repressors depending on the cellular context such as the RNA component, *SRA*, of the steroid receptor coactivator complex acts as co-activator for the transcription factors MYOD and VDR [202, 203], but also as a repressor in association

with CTCF [204]. For some long ncRNAs the process of RNA transcription plays the important role instead of the RNA itself [205, 206].

Many long ncRNAs function as link or scaffold to bring together specific regions in DNA and/or RNA and proteins. For example, the binding and recruitment of PRC2 to specific loci by long ncRNAs including the *Oct4*-pseudogene 5 [207], *HEIH* [208] and *HOTAIR* [151, 209, 210]. Long ncRNA *MALAT1* interact with SR splicing factors to regulate alternative splicing [165, 211] and the long ncRNA *Gadd7* associates with TDF43 to control mRNA decay [212]. Moreover, several long ncRNAs have been described to act as important structural component, especially within the nucleus, such as *Gomafu* [213], which constitutes a novel nuclear domain within certain neurons, and *NEAT1*, which is essential for the formation and maintenance of paraspeckles [214].

## Long ncRNAs and gene regulation

Long ncRNAs are regulators of transcription; they regulate gene transcription in either *cis* or *trans* [152, 215]. As *cis*-regulators, long ncRNAs displace DNA-binding proteins to activate or repress neighbouring genes. A knockdown study found that depletion of long ncRNAs result in a loss of expression in neighbouring genes [216]. Another study identified inhibitory effects of long ncRNA from the dihydrofolate reductase (*DHFR*) locus [217]. Here, a long ncRNA transcribed from the upstream minor promoter of *DHFR* represses the transcription of *DHFR* by the formation of a stable complex with the major promoter of *DHFR*, the interaction with the general transcription factor IIB and dissociation of the pre-initiation complex from the major promoter [217].

Long ncRNAs can regulate neighbouring genes through allosteric modulation. Following DNA damage, cell apoptosis requires the repression of cyclin D1 protein (CCD1), which is involved in cell cycle progression. Several long ncRNA (ncRNA<sub>CCND1</sub>A-D) derived upstream of the CCND1 promoter tether to the 5' regulatory regions of cyclin D1 and recruit the RNA binding protein TLS. In turn, TLS inhibits CREB binding protein and p300, which eventually results in cyclin D1 transcriptional repression [216].

Long ncRNAs also act as co-factors through recruitment of transcription factors to regulate target gene transcription. One example is *Evf2*, which is expressed in the mouse embryonic forebrain [218]. During the development of GABAergic neurons, *Evf2* mediates the recruitments of the homeobox protein DLX2 and the DNA methyl-binding protein MECP2 to

downregulate the expression of *Dlx5* and *Dlx6* and upregulate glutamic acid decarboxylase 1 (*Gad1*) [219].

*Trans*-acting long ncRNAs modulate chromatin through different ways. One example is that long ncRNAs can bind to complementary DNA to form DNA-RNA hybrid, which recruits chromatin modifier or transcriptional regulators to activate or repress target genes. DNA-binding proteins bind and use long ncRNAs as a platform for protein complex assembly. Alternatively, long ncRNAs induce conformational changes to DNA binding protein, so that DNA bound factors no longer inhibit or activate gene transcription [220].

### Long ncRNAs and imprinting

Genomic imprinting is a phenomenon by which an inherited functional gene from either maternal or paternal allele is not expressed. Studies have shown *cis*-acting long ncRNAs are involved in the methylation and inactivation of imprinted alleles [221]. Parental specific expression of long ncRNAs is controlled by imprinting control regions (ICEs), which are differentially methylated. For the imprinted allele, the unmethylated and active ICE induces a nearby long ncRNA to recruit repressive chromatin modifiers, which results in gene silencing. By contrast, methylated ICE in non-imprinted alleles prevents the expression of long ncRNA, allow gene transcription to proceed [222]. Examples for maternally imprinted gene clusters are the *Igf2r* and *Kcnq1*, where the corresponding long ncRNA *Airn* and *Kcnq1ot1* are expressed only on the paternal allele [223, 224].

## Long non-coding RNAs and sex development

### Xist and X chromosome inactivation

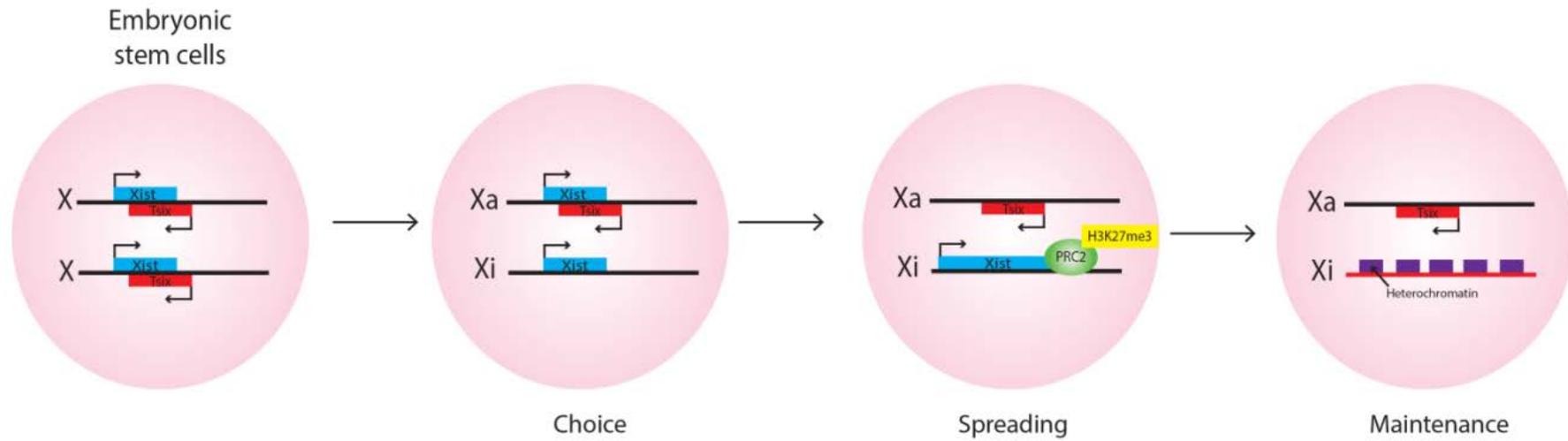
In mammals, female embryos have two X chromosomes, whereas males have one of each X and Y chromosome. To make sure females do not produce double the genes products from the X chromosomes, a process called X chromosome inactivation is required to silence one of the two X chromosomes in XX females. *Xist*, a 17kb ncRNA from the X-inactivation centre (*Xic*) locus, and the *Xist* antisense transcript ncRNA *Tsix*, are the main mediators of X inactivation [225, 226]. In embryonic stem cells, both X chromosomes weakly express *Xist* and *Tsix*. Upon differentiation, the X chromosomes undergo pairing, counting and choice.

One of the X chromosomes (Xi) is randomly selected to only express *Xist* as *Tsix* is antagonised by another long ncRNA named *Jpx*, while the other chromosome (Xa) continues to express *Tsix* [227]. An internal non-coding transcript named repeats A (*RepA*) within the *Xist* locus of Xi chromosome recruits PRC2 to deposit histone-modification marks (H3K27me3) at *Xist*, leading to the upregulation of *Xist* transcription [228, 229]. *Xist* then propagates repressive marks along of entire Xi chromosome through a single nucleation site. Heterochromatin is established through this process and will be maintained throughout life. On the other hand, the Xa chromosome is maintained due to silencing of *Xist* by *Tsix* via recruitment of DNMT3A to the *Xist* promoter [230] (Figure 5). Another study described a second mechanism for X inactivation: Small interfering RNAs (siRNAs) generated by Dicer from the *Xist/Tsix* duplex can also induce X inactivation [231].

### Long ncRNA and gonad development

Currently not much is known about the role of long ncRNAs in gonadal development as studies in this field only began in recent years. In 2011, Castillo and his group identified a novel ovarian long ncRNA named steroidogenic acute regulatory protein natural antisense transcripts (*Star NAT*). *Star NAT* is essential for the function and regulation of StAR, which takes part in hormone-induced cholesterol transport in steroidogenic cells. This long ncRNA is 3146bp long, polyadenylated and fully complementary to the 3.5 kb StAR *mRNA* sequence. It is expressed in MA-10 cells and steroidogenic tissues such as ovary and testis. After cyclic AMP stimulation, *Star NAT* is found to downregulate both StAR protein and progesterone production [232].

## Xist and Tsix in X inactivation



**Figure 5. Long ncRNA *Xist* (blue box) and *Tsix* (red box) in X chromosome inactivation.** In mammalian females, one of the two X chromosomes is silenced through the process of choice, spreading and maintenance. As the result, Xa (active) chromosome is maintained while Xi (inactive) chromosome becomes heterochromatin (purple boxes).

## **Specific background for the project**

A recent genomic study from this laboratory identified genes that were ovary-enriched expressed. The study used microarrays, which were custom-designed to detect long ncRNAs. Genes that were already known and characterised were excluded from this study. A new list of genes that could be implicated in ovary development was identified and expression profiles of ncRNAs during mouse foetal gonad development from 11.5 to 14.5dpc were determined. Microarray analysis identified 82 sexually dimorphic expressed long ncRNAs during mouse gonad development, 56 of them being ovary-enriched and 26 testis-enriched expressed [233]. The result showed that two out of three of the sexually dimorphic long ncRNAs were ovary-enriched expressed. This observation is in contrast to microarray data for protein-coding genes from 11.5 to 14.5dpc, for which more genes are expressed in the testis and only a few in the ovary. The data were further confirmed by section *in situ* hybridisation (ISH) of mouse embryos from 11.5 to 13.5dpc, which identified several ovary-enriched long ncRNAs (*Onc*) and testis-enriched long ncRNAs (*Tnc*). Of the ovary-enriched long ncRNAs, different expression patterns were observed. Some long ncRNAs were highly enriched in the ovary at all stages from 11.5 to 13.5dpc, while others were expressed at low levels in both testis and ovary at 11.5dpc, but were upregulated only in the ovary thereafter [233]. One example is *AK014986*, which was detected only in XX germ cells before entry into meiosis.

*AK144366* is an example of a *Tnc*, which was expressed in the nuclei of Sertoli cells. Data from the previous study suggested *AK144366* expression is regulated by pro-testis factor SOX9. Several SOX binding sites were identified on the putative promoter region of *AK144366*. *Ex vivo* analysis suggested the promoter of *AK144366* was more active in testes compared to ovaries. Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) using RNA from *Sox9* mutant mice did not show an upregulation of *AK144366* expression. Furthermore, functional analysis has shown that the expression of *AK144366* in ovaries resulted in down regulation of pro-ovarian factor FOXL2 (Wilhelm *et al*, unpublished data).

## **Hypothesis**

Although the functions of long ncRNAs identified in the previous study are currently unclear, their sexual dimorphic expression patterns suggest there may be potential roles for them in foetal gonad development. Long ncRNAs might be one of the missing links to understand sex development. They are likely to play functionally important roles similar to what has been observed in other tissues. A better understanding of the relationship between long ncRNAs and gonad development is required. Therefore I hypothesise long ncRNAs are expressed during gonad differentiation and play an important cellular roles such as migration, proliferation and apoptosis.

## **Aims and expected outcomes**

The mechanisms of gene regulation and the *in vivo* function of testis- and ovary-specific long ncRNAs in sex development are currently undetermined. This study aims to identify the role of long ncRNAs in gonad development, especially looking at cellular roles of long ncRNAs. Results from this study will verify the possible link between long ncRNAs and gonadogenesis which will potentially in the future allow a better understanding of the role of long ncRNAs in reproductive problems such as DSDs.

This study will validate the microarray results from the previous study using section ISH on paraffin embedded whole mouse embryos from 11.5 to 14.5dpc. Section ISH allows the identification of the cell type (ie. germ cell or somatic cell) that expresses the long ncRNA. In addition, quantitative PCR analysis will be carried out to quantify the expression of long ncRNA candidates. Lastly, cell culture experiments will be performed involving cloning and transfecting inducible expression constructs and monitoring of cellular events such as migration, proliferation and apoptosis. Mouse embryonic testicular cell line (TM3) and mouse germ cell cell line (GX2) will be used for the over-expression long ncRNAs of interest.

**Aim 1:** Detailed expression analysis of long ncRNAs in wildtype mouse models.

**Aim 2:** Functional analysis in cell culture.

# MATERIALS AND METHODS

## TISSUE COLLECTION AND PROCESSING

### **A. Mouse embryos**

All work conducted for this project has been approved by the Animal Ethics Committee of Monash University (MARF/2012/133). Mouse embryos were taken from timed matings with the presence of a vaginal plug being deemed 0.5dpc of either CD1 or C56BL/6 laboratory mouse strains that were obtained from the Monash Animal House facilities. Mice were sacrificed by cervical dislocation or CO<sub>2</sub> inhalation by a trained supervisor in accordance with the approved ethics. 70% ethanol (EtOH) [Merck Millipore] was applied to the abdominal region of the mother and the uterus was removed. Individual embryos of 11.5, 12.5, 13.5 or 14.5dpc were dissected from the uterus and washed in phosphate buffered saline (PBS). Where needed, gonads taken for RNA extraction were dissected from the embryos using a Stemi 2000 stereomicroscope [Zeiss] or a M295 stereozoom microscope [Leica], snap frozen on dry ice and transferred to the -80°C freezer [Thermo Scientific] for long term storage.

### **B. Sex genotyping**

Tails from wildtype embryos to be embedded were collected to extract genomic DNA for the identification of the genetic sex of individual embryos. 160µl of lysis buffer (Appendix) with 5% proteinase K (10µl/ml) was added to each tail sample and incubated at 55°C overnight in a heating block [Thermoline Scientific]. The next day temperature was increased to 95°C for 10 minutes. 160µl of isopropanol [Millipore] was added to each sample then incubated at room temperature for 1 hour. The samples were centrifuged [Eppendorf 5424] at 5000xg for 8 minutes, following the removal of the supernatant, washing of the pellet with 300µl of 70% EtOH [Merck Millipore] then centrifuged for 4 minutes at 5000xg. The genomic DNA pellet was air-dried at 37°C for 20 minutes and dissolved in 15µl of MilliQ water.

2µl of genomic DNA was added to 18µl of polymerase chain reaction (PCR) master mix containing 15.7µl MilliQ water + 2µl 10x ThermoPol Detergent-free Reaction buffer [Biolabs] + 0.4µl 10µM Sex reverse primer (Appendix) + 0.4µl 10µM Sex forward primer (Appendix) + 0.4µl 10mM deoxynucleotide triphosphates (dNTPs) + 0.1µl Taq DNA polymerase (5000U/ml) [Biolabs], with 20µl of master mix as a negative control. The sexing PCR

protocol in a BioRad T100 thermal cycler [BioRad] started with an initial denaturation 95°C for 3 minutes before undergoing 34 cycles of denaturation at 95°C for 30 seconds, annealing at 50°C for 30 seconds, extension at 72°C for 30 seconds, then final elongation at 72°C for 5 minutes before the temperature was reduced to 12°C indefinitely. 2µl of loading buffer (Appendix) was added to each PCR reaction, and 10µl was loaded alongside 8µl of 100bp DNA ladder [Fermantas] on a 1% agarose [Roche] / TAE buffer (Appendix) gel with 2µl Gelred nucleic acid stain (10000X) [Biotium], using either a BioRad PowerPac Basic [BioRad] or a Gibco BRL Electrophoresis Power Supply [Life Technologies] to run for 30 minutes at 125V. Resulting bands were visualised and analysed on an ImageStation 4000MMPro [BioRad] with the Carestream program [Kodak] or BioRad ChemiDoc MP imagine system with the Image Lab 5.0 program [Biorad]. Sex forward and reverse primers consist of a single primer specific to both *Xlr* (X-linked lymphocyte regulated complex) and *Sly* (*Sycp3*-like Y-linked). The genetic sex is distinguished by the length of the PCR fragments, 280 base pairs (bp) for XY and 685bp + 480bp for XX [234].

### **C. Processing and embedding**

Embryos to be embedded were fixed in 4% paraformaldehyde (PFA) [Sigma] in PBS at 4°C overnight. The next day, embryos were washed twice in PBS for 20 minutes. An ethanol series (25%, 50%, 70% EtOH [Merck Millipore] for 20 minutes each) was performed to dehydrate the embryos. Embryos were then placed in labelled cassettes in 70% EtOH [Simport tissue processing/embedding cassettes or Thermo Scientific tissue-loc histoscreen]. Monash Histology Platform completed alcohol processing of the embryos on a Peloris II machine [Leica]. Embryos were embedded in using a Shandon Histocentre 3 Embedding Centre [Thermo Scientific].

## **II. IN SITU HYBRIDISATION (ISH)**

### **A. PCR primers for cloning**

ISH probes of approximately 800 to 1000bp were designed from known cDNA sequences of long ncRNAs obtained from online database (UCSC genome Browser). ISH probes were generated using PCR primers [Integrated DNA technologies] that amplified either testis-enriched long ncRNAs (Tnc) or ovary-enriched long ncRNAs (Onc). A total of 11 long ncRNAs candidates were tested using ISH, 6 of them were Tnc: *AK034891* (Tnc3), *AK013819* (Tnc6), *AK013488* (Tnc7), *AK005877* (Tnc9), *AK045786* (Tnc10), *AK043086* (Tnc13) and 5 of them were Onc: *AK020106* (Onc3), *AK036014* (OncB), *AK182836* (OncF),

*AK015136* (OncK) and *AK044909* (OncL). To determine the expression pattern of long ncRNA candidates, section ISH was performed on sections of male (XY) and female (XX) embryos from wildtype (WT) mice at 11.5, 12.5, 13.5 and 14.5 dpc. For long ncRNA candidates with expression in gonads, paraffin-embedded embryos lacking germ cells (homozygote  $W^e$  mouse strain) were used to identify whether the expression of long ncRNA is dependent on germ cells. This mouse strain is characterised by a mutation in the stem cell growth factor receptor cKit. Germ cell migration to the gonad is disrupted in homozygous  $W^e$  mice, which results in the loss of germ cells for both the testis and the ovary [235]. An RNA probe detecting *Col1a1* was used as an overall positive control. *Col1a1* is expressed in most connective tissues including cartilage and bone [236]. Primers were dissolved in MilliQ water at a 100 $\mu$ M concentration and stored in -20°C, then used as a 1 to 10 dilution (10 $\mu$ M).

### **B. Cloning PCR of ISH probes for long ncRNA candidates**

Diluted primers were used in the cloning PCR protocol with genomic or cDNA from 13.5 dpc wildtype XX or XY gonads as templates. PCR mix were prepared in 0.2ml PCR tubes [Axygen] as follows: 41.3 $\mu$ l MilliQ water + 1 $\mu$ l 10mM dNTPs [BioLabs] + 0.2 $\mu$ l MyTaq DNA polymerase (5000U/ml) [BioLabs] + 1 $\mu$ l long ncRNA forward primer (10 $\mu$ M) + 1 $\mu$ l long ncRNA reverse primer (10 $\mu$ M) + 5 $\mu$ l 10x ThermoPol Detergent-free Reaction Buffer [BioLabs] + 0.5 $\mu$ l c/genomic DNA. The cloning PCR protocol on a BioRad T100 thermal cycler started with initial denaturation of 95°C for 3 minutes then 39 cycles of 95°C for 30 seconds to 50-60°C for 30 seconds to 72°C for 1 minute, final elongation at 72°C for 5 minutes and maintained at 12°C indefinitely. After cloning PCR, agarose gel electrophoresis was performed to check the length of the PCR fragments. 50 $\mu$ l of PCR reaction with 5 $\mu$ l of loading buffer was separated on a 1% agarose gel [Vivantis] at 125V for 30 minutes.

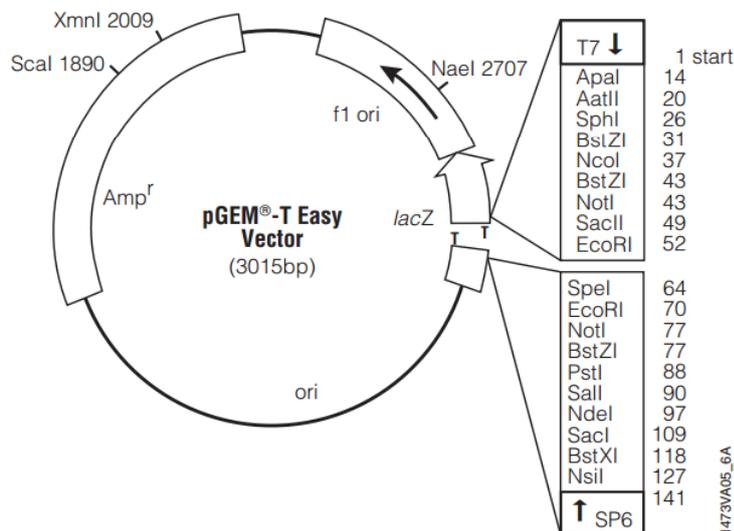
### **C. DNA isolation**

The correct sized DNA fragments were excised from the agarose gel with a razor blade [Swann-Moston] under UV light [Fotodyne] in the dark room. The protocol from Zymoclean Gel DNA Recovery Kit [Zymo Research] was followed to recover the DNA from the gel. The gel fragments were first weighed, then 3 volumes of Agarose Dissolving Buffer (ADB) [Zymo Research] were added to each volume of agarose gel. The mixture was incubated at 55°C for 10 minutes for the gel to completely dissolve. Dissolved solution was transferred to a Zymospin Column [Zymo Research] in a collection tube and centrifuged for 1 minute at 14680xg [Eppendorf 5424]. After discarding the flow through, 200 $\mu$ l of DNA Wash Buffer [Zymo Research] was added to the column and centrifuged for 30 seconds and this step was

repeated twice. Through the addition of 10µl DNA Elution Buffer [Zymo Research] to the column, ultra-pure DNA was eluted by centrifugation for 1 minute at 14680x g. DNA was stored at -20°C.

#### D. DNA ligation

The PCR fragments were ligated into the pGEM®-T Easy Vector (3015bp) [Promega] in the following reaction: 4µl isolated PCR fragment, 5µl 2x Rapid Ligation Buffer [Promega], 0.5µl T4 DNA ligase [Promega] and 0.5µl pGEM®-T Easy Vector [Promega], then incubated overnight at 4°C to generate the maximum number of plasmids. The pGEM®-T Easy Vector features a single T- overhang at the insertion site and numerous restriction sites within the multiple cloning region, providing easy and efficient ligation (Figure 6).



**Figure 6. pGEM®-T Easy Vector Map and Sequence Reference Points (Promega).** On the right is the list of sites for restriction enzymes, and the T7 and SP6 RNA polymerase promoters.

#### E. Transformation

5µl of ligation reaction was used to transform 50µl of MultiShot StripWell Top10 Chemically Competent Cells [Invitrogen]. Bacteria and ligation reaction were mixed and left on ice for 15 minutes, before incubated in a 37°C water bath for 3 minutes to heat-shock the bacteria, then left on ice for another 5 minutes. 500µl of room temperature Super Optimal Broth (SOC) medium [Invitrogen] was added to the samples and placed on a shaker for 1 hour at 37°C. The samples were centrifuged for 1 minute at 6000xg with an Eppendorf 5424 centrifuge, 400µl of supernatant was discarded and the remaining 150µl of bacteria culture was resuspended and plated onto 100µg/ml ampicillin [Roche] agar [Merck Millipore] plates

containing 30 $\mu$ l X-Gal (20mg/ml) [Thermo Scientific] and 10 $\mu$ l of 100 $\mu$ M Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) [Thermo Scientific]. The plates were incubated at 37°C overnight. Colonies on the plates were checked the next day, positive colonies appear white due to the disruption of LacZ activity by the insert in the vector, while bacteria transformed with vectors without an insert resulted in blue colonies. Single white colonies were picked from the plates and placed in 5ml of 100 $\mu$ g/ml ampicillin in LB medium (Appendix) on a shaker overnight at 37°C.

## **F. Plasmid DNA**

5ml of each bacteria culture was centrifuged at room temperature for 1 minute at 14680x g and the supernatant discarded. The ZR Plasmid Miniprep-Classical kit protocol was followed [Zymo Research] to isolate the plasmid DNA. To check the size of the insert, a restriction enzyme digest was performed using the following mix: 3 $\mu$ l purified plasmid DNA + 3 $\mu$ l 10x NEB buffer [BioLabs] + 0.3 $\mu$ l restriction enzymes [BioLabs] + 23.7 $\mu$ l MilliQ water. One most frequently used restriction enzyme was EcoRI, which cuts at both sites of the insert and thereby provides single enzyme digestion for the release of the insert. The enzyme digestion reaction was left at 37°C for 2 hours. 3 $\mu$ l of DNA loading buffer was added to each sample and the whole 33 $\mu$ l alongside 8 $\mu$ l of 1kb DNA ladder was separated on a 1% agarose/TAE buffer gel containing 2 $\mu$ l Gelred for 30 minutes at 125V. Identities of the sequence of the PCR fragments were further confirmed via DNA sequencing through the Micromon Sequencing Facility using either an SP6 or T7 primer. If necessary, the remainder of the transformed samples were used to inoculate a larger volume of LB medium to purify plasmid DNA on a large scale following the Plasmid Maxi kit [Qiagen] or the Zyppy Plasmid Maxiprep Kit [Zymogen] protocol. Concentration of the purified DNA was measured using the Nanodrop 2000 spectrophotometer system [Thermo Scientific].

## **G. M13 PCR**

The PCR fragments cloned into the pGEM®-T Easy Vector [Promega] were amplified in a M13 PCR reaction, the master mix contained: 0.5 $\mu$ l plasmid DNA + 1 $\mu$ l 10 $\mu$ M M13 forward primer + 1 $\mu$ l 10 $\mu$ M M13 reverse primer + 5 $\mu$ l 10x ThermoPol Detergent-free Reaction buffer [BioLabs]+ 1 $\mu$ l 10mM dNTPs [BioLabs]+ 0.2 $\mu$ l Taq DNA Polymerase [BioLabs]+ 41.3 $\mu$ l MilliQ water. The reaction was subjected to the M13 PCR protocol, which is identical to the sex genotyping PCR protocol described earlier. 2 $\mu$ l of loading buffer was added to each sample and 10 $\mu$ l was loaded alongside 8 $\mu$ l of 100bp DNA ladder on a 1% agarose [Roche] /

TAE buffer gel containing 2 $\mu$ l Gelred nucleic acid stain [Biotium] and electrophoresis [Thermo Scientific] was separated for 30 minutes to 1 hour at 125V. Resulting bands were visualised on an ImageStation 4000MMPro using the Carestream program [Kodak].

## **H. *In vitro* transcription**

The M13 PCR product was used as a template for *in vitro* transcription to produce the antisense RNA probe for ISH. Each reaction required: 13.5 $\mu$ l MilliQ water + 0.5 $\mu$ l M13 template DNA + 2 $\mu$ l DIG-labelling mix [Roche] + 2 $\mu$ l 10x transcription buffer [Roche] + 1 $\mu$ l T7 or SP6 RNA polymerase [Roche] + 0.5 $\mu$ l RNase inhibitor [Roche]. The reaction with a total of 20 $\mu$ l was placed in the 37°C room for 2 to 4 hours. DIG allows the qualitative detection of specific binding of ISH probes to long ncRNA candidates. To ensure the RNA probe had been successfully synthesised, 1 $\mu$ l of the solution was added to 10 $\mu$ l of MilliQ water and 2 $\mu$ l loading buffer then separated on a 1% agarose/TAE gel alongside 10 $\mu$ l of 1kb ladder with electrophoresis running at 125V for 30 minutes. 30 $\mu$ l of MilliQ water + 8 $\mu$ l 2M sodium acetate (pH 4.2) + 1 $\mu$ l glycogen (20mg/ml) [Roche] + 150 $\mu$ l of 100% EtOH [Merck Millipore] was added to precipitate the RNA at -20°C overnight. The next day, precipitated RNA probes were spun down at 18,000xg for 15 minutes at 4°C. After removal of the supernatant, the RNA pellets were washed twice with 400 $\mu$ l of 75% EtOH [Merck Millipore] and air dried at room temperature for 15 to 30 minutes. The transparent pellet was dissolved in 50 $\mu$ l of MilliQ water. 5 $\mu$ l of the probe was used to check on a 1% agarose/TAE gel and the remaining 45 $\mu$ l was stored at -20°C for future use.

## **I. Paraffin section ISH**

Paraffin embedded whole mouse embryos were sectioned at 7 $\mu$ m using a CUT 4060 microtome [microTec]. The sections were mounted on SuperFrost Plus microscope slides [Menzel-Glaser] and air dried overnight at room temperature.

The first day of ISH should be completely RNase-free to avoid degradation of RNA. The samples were dewaxed twice in xylene [Sigma Aldrich] for 10 minutes each then rehydrated through an ethanol [Merck Millipore] series (2 mins each: 2x 100%, 95%, 90%, 80%, 50%, 30% EtOH). This is followed by 2x RNase-free PBS washes for 5 minutes each. The slides were placed in 10 $\mu$ g/ml solution of proteinase K in PBS for 20 minutes to permeabilise the cell membranes, washed twice in RNase-free PBS for 5 minutes each, refixed in freshly

prepared 4% PFA [Sigma] in PBS for 10 minutes, washed again twice in RNase-free PBS, then placed in the acetylation solution for 10 minutes. The acetylation solution was made with 196ml MilliQ water + 2.6ml triethanolamine [Sigma] + 350 $\mu$ l hydrochloric acid [Sigma], mixed and poured into slide container, added 500 $\mu$ l acetic anhydride [Sigma] then mixed again by dipping the slides. Whatman filter paper was placed into an RNase-free humidified plastic slide tray which was soaked with 2x saline-sodium citrate (SSC)/50% formamide. Subsequently, the slides were washed in RNase-free PBS for 3x 5 minutes before laid out in the slide tray. 300 $\mu$ l of prehybridisation solution (Appendix) was added onto each slide and incubated for 2 to 3 hours at room temperature. Hybridisation solutions containing 1.5 $\mu$ l of probe with 150 $\mu$ l of prehybridisation solution was subsequently added to each slide and slides were covered with parafilm [Bemis]. The tray was sealed with cling wrap for a formamide-saturated atmosphere for hybridisation then incubated overnight at 55°C to 65°C in an incubator [S.E.M].

Post hybridisation, the slides were placed in 5x SSC (Appendix) for 10 minutes at hybridisation temperature to remove the parafilm and then in 0.5x SSC for 1 hour at 60°C followed by 0.5x SSC for 10 minutes at room temperature. The slides were washed 2x 10 minutes in NT buffer (Appendix), then incubated in a blocking solution (10% HIHS (heat-inactivated horse serum) in NT buffer) for 1 hour at room temperature. A mixture of 1:1000 anti-digoxigenin antibody coupled to alkaline phosphatase ( $\alpha$ DIG-AP) [Roche] in 1% HIHS/NT was prepared, 500 $\mu$ l of the mixture was added to each slide and incubated at 4°C overnight.

Slides were washed in 3x 5 minutes with NT buffer at room temperature to ensure no unbound antibody remained and then washed twice in NTM (Appendix) for 10 minutes each. A colour reaction solution with 3.5 $\mu$ l nitro blue tetrazolium chloride (NBT) [Roche] and 3.5 $\mu$ l 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) [Roche] per ml NTM was prepared and passed through a 0.45  $\mu$ m filter to remove any precipitates. 500 $\mu$ l of the filtered solution was added onto each slide and incubated in a slide box at room temperature until the colour was deemed sufficient by looking under a dissecting microscope [Zeiss]. The reaction was stopped by briefly washing the slides in PBS and fixing in 4% PFA [Sigma] in PBS for a minimum of 10 minutes at room temperature. The samples were then washed again in PBS before being mounted with VectaMount permanent mounting media [Vector] and covered with 24x60mm coverslip [Menzel-Glaser]. Imaging was done on an Olympus BX60 bright field microscope using an attached camera [Olympus DP70] and DP controller software [Olympus]. Figures were prepared using Adobe Photoshop and Illustrator [Adobe].

### **III. QUANTITATIVE REVERSE TRANSCRIPTASE PCR (qRT-PCR)**

#### **A. RNA extraction**

Foetal mouse gonads dissected from the mesonephroi or cell lines (TM3 and GC2) stored at -80°C were thawed on ice. RNA was extracted from gonads or cells by following the RNeasy Micro and Mini kit protocol respectively [Qiagen]. Zymo Direct-Zol RNA MiniPrep kit [Zymo Research] was also used for RNA extraction. After adding the lysis buffer from the kit, samples were homogenised with needles of 0.4mm diameter [Terumo]. Purified RNA concentration was measured using the Nanodrop 2000 Spectrophotometer system [Thermo Scientific].

#### **B. cDNA synthesis**

The Superscript III First-Strand Synthesis System [Invitrogen] for RT-PCR was used for the synthesis of cDNA from testis and ovary RNA (11.5, 12.5, 13.5 and 14.5dpc) as well as RNA from cells (TM3 and GC2). The manufacturer's protocol was followed. RNA/primer mixture was prepared using approximately 100ng of total RNA + 1µl 50ng random hexamers [Invitrogen] + 1µl 10mM dNTP [Invitrogen] + sterile distilled water to a volume of 13µl. The mixture was incubated at 65°C for 5 min to denature the RNA and then placed on ice for at least 1 minute. cDNA Synthesis Mix containing 1µl 0.1M DTT [Invitrogen]+ 1µl RnaseOUT Recombinant RNase Inhibitor (40U/µl) [Invitrogen] + 4µl 5X First Strand Buffer [Invitrogen] + 1µl Superscript III RT (200U/µl) [Invitrogen] was added to each RNA/primer mixture, mixed gently and incubated at 25°C for 5 minutes to anneal primers to the RNA, at 50°C for 60 minutes for cDNA synthesis before terminating the reaction at 70°C for 15 minutes. After brief centrifugation, 1µl of *E.Coli* RNase H [Invitrogen] was added to each sample to remove RNA before the mixture was incubated at 37°C for 20 minutes. The first-strand cDNA obtained in the synthesis reaction can be used directly as template in PCR reactions or stored at -20°C for future experiments.

#### **C. qRT-PCR and data analysis**

Forward and reverse primers have been designed for all genes of interest: Tnc9, OncL, Tnc7, Tnc3, Tnc6, Onc3, AK019493 (OncP), OncF, AK015184 (OncQ), OncB, *Foxl2*, *Amh* and the house keeping gene *Sdha* [Integrated DNA Technologies] (Appendix) [237]. A qRT-PCR

master mix was prepared using cDNAs from gonads at different developmental stages (11.5 to 14.5 dpc) and sexes, from which 10 $\mu$ l was added to each replicate in a 96-well clear non-skirted PCR plate with 8-strip flat caps [4titude]. Each well consists of: 3 $\mu$ l MilliQ water + 1 $\mu$ l cDNA (25ng/ $\mu$ l) + 1 $\mu$ l 10 $\mu$ M primer pair + 5 $\mu$ l 2x Power SYBR Green PCR Master Mix [Applied Biosciences]. Each developmental stage was measured in at least three independent biological triplicates for statistical calculation, and within each qRT-PCR experiment three technical triplicates were performed for each condition. Quantification was conducted by an Mx3000P real-time PCR machine [Stratagene] through MxPro [Stratagene] program of: initial denaturation at 95°C for 10 minutes, 40 cycles of 95°C for 30 seconds and 60°C for 30 seconds, a final elongation cycle of 95°C for 1 minute to 55°C for 30 seconds then to 95°C for 30 seconds. After PCR amplification, expression levels of long ncRNA candidates as well as testis-specific *Amh* and ovary-specific *Foxl2* mRNA as controls were measured and normalised to the signal intensity of the housekeeping gene *Sdha* then converted all Ct threshold readings to  $2^{-Ct}$ , using three independent biological and the average of three technical replicates for both sexes and four different developmental stages. As the current study tried to compare the difference between XX and XY expression levels at different time points for each gene, two-tailed unpaired t-test was used to analyse each data sets with a p value of less than 0.05 deemed statistically significant. Data analyses and graphical representation of results were performed using Microsoft Excel [Microsoft] and Prism 6 [Graphpad].

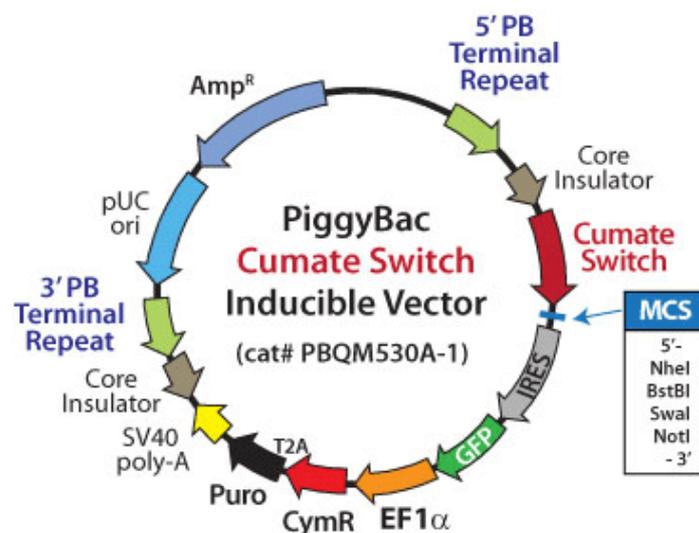
#### **IV. Expression constructs**

To generate expression constructs of the long ncRNAs the gene of interest has to be amplified from cDNA by PCR before cloning it into an expression vector. After designing the necessary primers, Expand High Fidelity PCR System [Roche] protocol was followed. Genomic DNA from the R1 mouse ES cell line, 13.5dpc testis and ovary cDNA were used. PCR mix 1 contained 20 $\mu$ l MilliQ water + 1 $\mu$ l 10mM dNTPs [BioLabs]+ 1.5 $\mu$ l 10 $\mu$ M forward primer + 1.5 $\mu$ l 10 $\mu$ M reverse primer + 1 $\mu$ l cDNA. PCR mix 2 contained 19.25 $\mu$ l MilliQ water + 5 $\mu$ l 10x EHF Buffer [Roche] + 0.75 $\mu$ l EHF enzyme mix [Roche]. Mix 1 and 2 were combined and subjected to the HF cloning PCR reaction: initial denaturation at 94°C for 2 minutes, followed by 10 cycles of 94°C for 15 seconds to 45-65°C for 30 seconds to 68/72°C for 45 seconds (for 0.75kb fragments), 29 cycles of 94°C for 15 seconds to 45-65°C for 30 seconds to 72°C for 45 seconds (for 0.75kb fragments) with an extra 5 seconds added to each cycle,

final elongation at 72°C for 7 minutes and being held at 12°C indefinitely. Gel electrophoresis was performed after the cloning PCR to check the length of PCR fragments.

Amplified sequences were first ligated into pGEM®-T Easy Vector [Promega] and transformed into competent *Escherichia coli* bacteria, before plated onto 1:1000 ampicillin [Roche] agar [Merck Millipore] plates. Single colonies were picked and plasmid DNA isolated using the ZR Plasmid Miniprep kit [Zymo Research]. The restriction enzyme EcoRI [BioLabs] was used to check the sizes of the inserts. Using EcoRI digestion, the inserts were released from the vector. Gel electrophoresis was performed to confirm the successful ligation into the pGEM®-T Easy Vector. DNA sequencing through Monash Micromon Facility was performed to check the identities of the inserts and to identify any mutations in the cloned sequence.

Inserts in pGEM®-T Easy were digested with specific restriction enzymes, with the aim to ligate the inserts into the PB-CuO-MCS-IRES-GFP-EF1-CymR-Puro cDNA Cloning and Inducible Expression Vector (PiggyBac vector) of approximately 9.5kb [System Biosciences] (Figure 7). By adding the small molecule cumate [System Biosciences], the expression of the cloned long ncRNA is induced, which allows for superior control of expression induction. The PiggyBac cumate switch can also be turned off via removal of cumate. The advantages of this vector are that it allows stable transfection into cells due to a puromycin selection gene, a GFP expression cassette downstream of the multiple cloning site that allows for fluorescence visualisation of the expression, and the expression of the gene of interest can be switched on and off by the addition and removal of cumate. There are four restriction sites within the multiple cloning site of the PiggyBac vector (NheI, BstBI, SwaI and NotI).



**Figure 7. PiggyBac Vector Map and Sequence Reference Points (System Biosciences).** Cumate switch can turn on the expression of the gene cloned into the multiple cloning sites as well as GFP through an internal ribosomal entry site (IRES). Other features of this vector include puromycin selection and transposon-specific inverted terminal repeat sequences located on both ends of the transposon vector.

Both the empty PiggyBac vector and pGEM®-T Easy vector with insert were digested with two restriction enzymes. In the cases where there were differences in the properties of restriction enzymes (ie. incubation temperature and choice of reaction buffers), the DNA had to be precipitated at -80°C for 20 minutes in between two digestions. Precipitation mixture included 50µl of digested plasmid + 5µl Sodium Acetate (NaAc) at pH5.2 [Sigma Aldrich] + 125µl 100% EtOH [Merck Millipore] + 1µl glycogen [Roche]. The PiggyBac vector was dephosphorylated to prevent religation from incomplete enzyme digestion. MilliQ water was added to 100ng of vector DNA + 2µl of 10x rAPid Alkaline Phosphatase Buffer [Roche] + 1µl rAPid Alkaline Phosphatase [Roche] to a final volume of 20µl. This reaction was incubated at 37°C for 10 minutes and then rAPid Alkaline Phosphatase was inactivated for 2 minutes at 75°C. Ligation between digested PiggyBac vector and insert DNA was performed using: 50ng of vector DNA + 150ng of insert DNA + 2µl 5x DNA Dilution Buffer [Roche] + MilliQ water to 10µl, then 10µl of 2x T4 DNA Ligation Buffer [Roche] + 1µl T4 DNA Ligase [Roche] was added to the mixture and incubated for 30 minutes at room temperature. The ligation reaction was transformed into competent bacteria and plasmid DNA was isolated from bacteria as described earlier. PiggyBac vector with insert was digested with restriction enzyme to check the size of the insert. Gel electrophoresis was performed to confirm the successful ligation of the insert into the PiggyBac vector.

Plasmid DNA was isolated from the agarose gel with Zymoclean Gel DNA Recovery Kit [Zymo Research]. The concentration of DNA was measured by Nanodrop2000 system [Integrated technologies]. Plasmid DNA with the correct sequence was used in cell culture analysis.

## **IV. Cell culture**

All tissue culture techniques were performed in laminar flow tissue culture hood [Bio-Cabinets] using sterilised glassware. The human embryonic kidney 293 (HEK293), mouse embryonic testicular TM3 (immature Leydig cell line) and mouse embryonic germ cell line (GC2) were grown in High Glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum (FBS)[Invitrogen]. Tissue culture flasks were incubated at 37°C in 5% CO<sub>2</sub> and 95% air in a humidifying incubator [Sanyo CO2 incubator or Thermo Scientific Hera cell 150].

Cells were passaged once every 2-3 days. After removing the DMEM+FBS solution from cells in a tissue culture dish, cells were washed with 3ml PBS to remove dead cells. PBS

was removed and cells were incubated in the 37°C incubator for 5 minutes with 2ml PBS + EDTA. Subsequently the solution was pipetted up and down several times to detach cells from the bottom of the flask and an aliquot of around 1/10 of the single cell suspension was transferred to a new flask with 10ml fresh DMEM + FBS. In cases where a certain number of cells needed, cell counts were determined with a Neubauer cell counting chamber.

### **A. Cell line transfection and cumate sufficiency testing**

Cells were grown to 60 to 80% confluence before splitting into 6-well plates [Falcon] with each well containing 200,000 cells in 2ml of culture medium. 16µl Lipofectamine 2000 Reagent [Invitrogen] in 300µl serum-free DMEM were added to two 1.5ml Eppendorf safe-lock microcentrifuge tubes [Eppendorf] containing either 2.5µg/ml expression construct plasmid DNA with transposase or 2.5µg/ml empty PiggyBac vector with transposase (as control) in 150µl of serum-free DMEM. Tubes were incubated for 20 minutes at room temperature then added to the 6-well plates with pre-seeded cells. After transfection, cells were cultured on a single 10cm plate [Falcon]. As the PiggyBac vector harbours a puromycin resistance gene, 1µg/ml puromycin in DMEM was used to select for cells that had integrated the plasmid into their genome. Puromycin selection was maintained in cell passages through the change of medium. Cumate induction solution (10,000x) [System Biosciences] was titrated to 1x, 5x or 10x concentration to test the optimal induction concentration for each cell types. Appropriate level of cumate was maintained in the media after passages to keep the cumate switch on. The PiggyBac vector also harbours a GFP gene to visualised transfected cells using a fluorescence microscope.

The peak time of the activation by cumate [System Biosciences] for specific cell types was determined first. Different cell lines respond differently to cumate activation. For this part of the experiment, all three cell lines, TM3 and GC2 cells were used to test cumate efficiency. The cell lines were not expected to express Onc3 without the Onc3 expression construct. TM3 cells were chosen as they are mouse testicular Leydig cell line and GC2 cells are of mouse testicular germ cell origin. TM3 and GC2 cells were seeded separately in 6-well plates with 100,000 cells in each well. 5x cumate were added to 4 wells with the 5<sup>th</sup> well without cumate as control. cDNA was prepared from cells at 4, 11, 24, and 48 hours after the addition of the cumate. To extract RNA from cells, 1ml of DMEM + FBS was removed from each well of the 6 well plates then cells were pipetted up and down to detach and remove cell clumps. After transfer cells into a 1.5ml Eppendorf tube, the tubes were centrifuged at 400xg in Eppendorf 5424 for 2 minutes and the supernatant was discarded. The cells were

re-suspended in 500µl RNase free PBS then centrifuged again for 2 minutes at 400xg. RNA and cDNA was synthesized as described earlier.

Both conventional RT-PCR and qRT-PCR were used to check the expression level of the Onc3 at different time points.

For conventional RT-PCR, 2µl of cDNA was added to 18µl of PCR master mix containing 15.7µl MilliQ water + 2µl 10x ThermoPol Detergent-free Reaction buffer [Biolabs] + 0.4µl 10µM expression construct forward primer + 0.4µl 10µM expression construct reverse primer + 0.4µl 10mM dNTPs + 0.1µl Taq DNA polymerase (5000U/ml) [Biolabs], with Onc3 cDNA from cloning PCR as positive control and water as a substitute for cDNA as negative control. The reaction was subjected to the sex genotyping PCR protocol described earlier.

On the other hand, each well of qRT-PCR consists of: 3µl MilliQ water + 1µl cDNA (25ng/µl) + 0.5µl 10µM expression construct forward primer + 0.5µl 10µM expression construct reverse primer + 5µl 2x Power SYBR Green PCR Master Mix [Applied Biosciences] with three technical triplicates performed for each condition. The reaction was subjected to the qRT-PCR protocol described earlier and each cDNA sample was normalised to housekeeping gene *Sdha*.

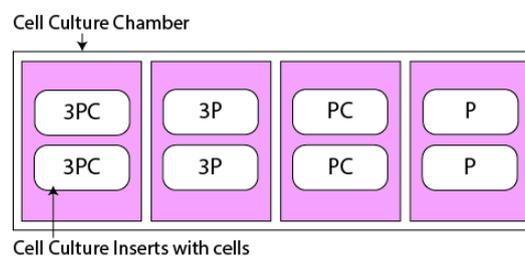
## **B. *In vitro* scratch assay (Ibidi insert system)**

The effect of overexpression of long ncRNA *AK020106* (Onc3) on cell migration/proliferation was tested using an *in vitro* scratch assay with stably transfected cells. Instead of the conventional scratch assay, the ibidi insert system was used. This system provides a uniform gap of 500µm between two inserts containing cells whereas conventional scratch assay uses pipette tips to scratch a line between the cells. The advantages of the insert system are that cells are protected from potential damage from the pipette tip and the size of the gap is uniform, leading to more consistent results.

The rate of cell migration/proliferation was measured by quantifying the distance of cells moved from the edge to the centre of the scratch. 40,000 cells were seeded in each culture insert [Ibidi], with 4 culture inserts in each chamber of the 4 well Cell Culture Chamber [Life Technologies]. Each Culture Insert can hold up to a maximum of 100µl of solution, 50µl was filled with DMEM + FBS with 10x cumate and the other 50µl contained cells with a concentration of 800,000 per ml (Figure 8). After 24 hours incubation at 37°C and 5% CO<sub>2</sub>, Culture Inserts were removed in a laminar flow tissue culture hood with sterilised tweezers

and the chambers filled with 1ml DMEM+FBS media with or without 5x cumate. The Cell Culture Chambers were imaged in a time-lapse microscope Widefield LX [Leica] with constant CO<sub>2</sub> supply from the Monash Micro Imaging Facility. Leica imaging software was used to create markings as references points close to the edges of the scratch to make sure images were acquired from the same field automatically. Three test conditions were used for HEK293 and TM3 cells, one was the test group consistant of cells that express Onc3, activated with cumate (3PC), and the other two conditions act as controls. These were cells stably transfected with Onc3 in the PiggyBac vector without cumate (3P) and the empty PiggyBac vector with cumate (PC). For GC2 cells, one more control condition of cells transfected with the empty PiggyBac vector without cumate (P) was used. The control group P can determine whether the changes in migration/proliferation of the test cell line are due to the expression of Onc3 or direct effects of the cumate treatment.

The incubation time was determined at 40 hours as this is the time when the faster moving cells were about to close the scratch. Images of the scratches were captured at the interval of once every 15 minutes for 40 hours. Live cell imaging started at 24 hours after the addition of cumate and images were taken once every 15 minutes for 24 hours. In addition to the migration/proliferation rate, the changes in cell morphology and the way cells migrate within 48 hours after the addition of cumate can also be analysed through images and videos captured by the microscope.



**Figure 8. *In vitro* scratch assay (Ibidi insert system) setup.** Migration/proliferation of Onc3 in PiggyBac activated with cumate (3PC), Onc3 in PiggyBac without cumate (3P), empty PiggyBac vector with cumate (PC) and empty PiggyBac vector only (P) were tested for HEK293, TM3 and GC2 cells. Each chamber contains two cell culture inserts where cells were seeded. Pink areas represent DMEM media. After 24 hours incubation, inserts were removed to create a uniform gap between the two cell groups.

Images were processed with Adobe Illustrator [Adobe]. Distance between one side of scratch and the other was measured in pixels. Through comparing the images at the beginning (0 hours) and at 24 hours incubation, the average micrometres per hour was calculated.

### C. MTT assay

MTT assay is a useful tool to study cell proliferation. In the current study, the assay was used to measure changes in cell proliferation in TM3 and GC2 cells with or without the expression of Onc3. CellTiter 96 AQueous One Solution Cell Proliferation Assay [Promega] contains a MTT tetrazolium compound with an electron-coupling reagent PES, together they form a stable yellow solution that can be bio-reduced into purple formazan in the mitochondria of metabolically active cells. In this assay, the intensity of absorption measured with spectrophotometer reflects the number of live cell.

TM3 and GC2 cells were first activated with 5x cumate 24 hours before the MTT experiment. Four conditions were tested: Onc3 in PiggyBac activated with cumate (3PC), Onc3 in PiggyBac without cumate (3P), empty PiggyBac vector with cumate (PC) and empty PiggyBac vector only (P). TM3 and GC2 cells were seeded ~2,000 per well in DMEM+FBS media with/without cumate in Falcon 96-well plates, they were first diluted to 200,000 per ml then 10 $\mu$ l of the diluted cells were added to 90 $\mu$ l of DMEM+FBS media to make up a total of 100 $\mu$ l per well. For each plate, all conditions were seeded in triplicates, with 2 sets for MTT induction and 1 set without MTT as control (Figure 9). A total of 5 96-well plates were prepared, one for each MTT induction point at 24, 28, 32, 36 and 48 hours after cumate activation.

24, 28, 32, 36 and 48 hours after cumate activation, 20 $\mu$ l of MTT solution was added to each test well in the corresponding plate then immediately incubated in the 37°C, 5% CO<sub>2</sub> incubator [Thermo Scientific]. 3 hours after each MTT induction point, the absorbance of the colour solution was quantified by measuring at 490nm with PHERAstar plate reader from the Biochemistry imaging facility. The experiment was repeated 3 times and the statistical analysis was calculated through Microsoft Excel [Microsoft] and graphs plotted with Prism 6 [Graphpad].

	1	2	3	4	5	6	7	8	9	10	11	12
A	M	M	M	M	M	M	M	M	M	M	M	M
B	M	M	M	M	M	M	M	M	M	M	M	M
C	M	M	TM3 3PC	TM3 PC	TM3 3P	TM3 P	M	GC2 3PC	GC2 PC	GC2 3P	GC2 P	M
D	M	M	TM3 3PC	TM3 PC	TM3 3P	TM3 P	M	GC2 3PC	GC2 PC	GC2 3P	GC2 P	M
E	M	M	TM3 3PC	TM3 PC	TM3 3P	TM3 P	M	GC2 3PC	GC2 PC	GC2 3P	GC2 P	M
F	M	M	M	M	M	M	M	M	M	M	M	M
G	M	M	M	M	M	M	M	M	M	M	M	M
H	M	M	M	M	M	M	M	M	M	M	M	M

**Figure 9. MTT plate setup.** Proliferation of Onc3 in PiggyBac activated with cumate (3PC), Onc3 in PiggyBac without cumate (3P), empty PiggyBac vector with cumate (PC) and empty PiggyBac vector only (P) were tested for both TM3 and GC2 cells. Shaded boxes were activated with MTT. M for DMEM media, they prevent the evaporation of test wells from incubation.

#### **D.TUNEL assay**

TUNEL assay is used to detect DNA fragmentation from apoptosis. The assay relies on the presence of nicks in the DNA that can be identified by TdT, an enzyme that catalyses the addition of dUTPs that are secondarily labelled with a marker.

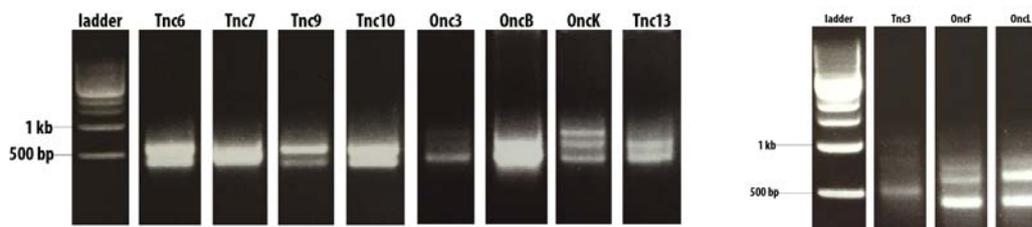
Cells were seeded in 100,000 per ml DMEM with FBS on coverslips in 6-well plates and incubated at 37°C, 5% CO<sub>2</sub> overnight. The next day, the coverslips with the cells were removed from the 6-well plates and fixed on a SuperFrost Plus microscope slide [Menzel-Glaser]. All of the TUNEL assay experiments were carried out in a slide tray. Cells were fixed in 1% PFA [Sigma]/PBS at pH7.4 for 10 minutes at room temperature before being washed with 2 changes of PBS for 5 minutes each and postfixed in precooled ethanol: acetic acid (2:1 ratio) for 5 minutes at -20°C, before excess liquid was drained. Cells were washed twice in PBS for 2 minutes each, then in blocking solution made of 3% H<sub>2</sub>O<sub>2</sub> in PBS for 5 minutes to block endogenous peroxidase activity, followed by two washes in PBS for 5 minutes each. Equilibrium Buffer was applied to the specimen (75µl/5cm<sup>2</sup>) and incubated for at least 10 seconds at room temperature. TdT enzyme was diluted in Reaction Buffer and applied to the slides (55µl/5cm<sup>2</sup>) then incubated in a humidified chamber at 37°C for 1 hour. The reaction was stopped with Stop/Wash Buffer, slides were agitated for 15 seconds and left at room temperature for another 10 minutes before the buffer was washed off with PBS 3 times for 1 minute each. Meanwhile, anti-digoxigenin peroxidase conjugate was warmed to room temperature and applied to slides (65µl/5cm<sup>2</sup>), again incubated in a humidified chamber at room temperature for 30 minutes. The buffer was washed off with 4 rounds of PBS for 2 minutes each before the DAB peroxidase substrate [Vector] was applied to the slides (75µl/5cm<sup>2</sup>) and incubated for 6 minutes at room temperature. Slides were washed 4 times with distilled water with 1 minute each for the first 3 washes and 5 minutes for the last wash. Coverslips with cells were mounted with VectaMount Aqueous medium [Vector]. Imaging was performed using an Olympus BX60 bright field microscope with an attached camera [Olympus DP70] and associated software. Live and apoptotic cells were counted to calculate the percentage of apoptotic cells.

# RESULTS 1

## EXPRESSION ANALYSIS

### 1. Long ncRNAs are sexually dimorphically expressed in embryonic mouse gonads

In previous work from the Wilhelm lab identified a number of sexually dimorphically expressed long ncRNAs. In order to validate the microarray data from the previous study and to establish a detailed temporal and spatial expression pattern for testis and ovary enriched long ncRNA candidates, section *in situ* hybridization (ISH) was performed. Gel electrophoresis of the *in vitro* transcribed products confirmed the generation of a total of 11 long ncRNA probes: AK034891 (Tnc3), AK013819 (Tnc6), AK013488 (Tnc7), AK005877 (Tnc9), AK045786 (Tnc10), AK043086 (Tnc13), AK020106 (Onc3), AK036014 (OncB), AK182836 (OncF), AK015136 (OncK), AK044909 (OncL) and sizes of the products were approximately the expected 800-1000 bp in length (Figure 10). In cases where multiple bands were present, the upper bands correspond to the predicted size.



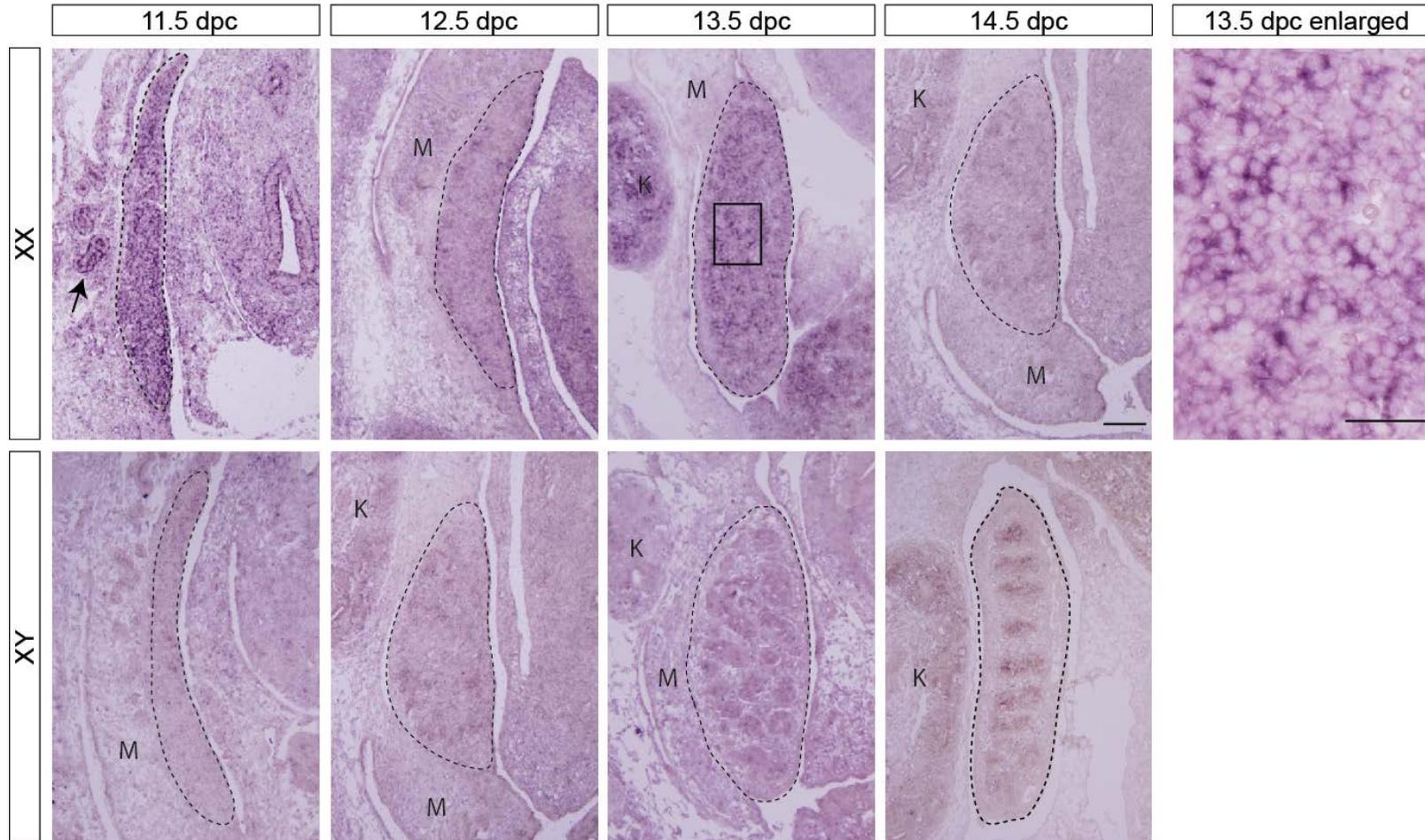
**Figure 10. Long ncRNA candidates *in vitro* transcription.** Gel electrophoresis of *in vitro* transcription products for long ncRNA ISH probes alongside a 1kb DNA ladder. The resulting fragments were of the expected size (approximately 800 to 1000 bp depending on the specific probe).

Different expression patterns of ovary-enriched long ncRNAs (Onc) were observed. Using ISH, expression in somatic cells can be distinguished from expression in germ cells. In high magnification images, germ cells appear as big, round cells whereas somatic cells appear to have a more irregular shape.

The expression of *AK020106* (*Onc3*) was specific in ovarian germ cells at 11.5 and 13.5 dpc (Figure 11). In comparison, no expression was detectable in XY gonads at all stages examined. At 14.5 dpc, no expression of *Onc3* could be detected in ovaries and testes. More broadly, *Onc3* expression was also detected in mesonephric tubules at 11.5 dpc and in kidneys at 13.5 dpc.

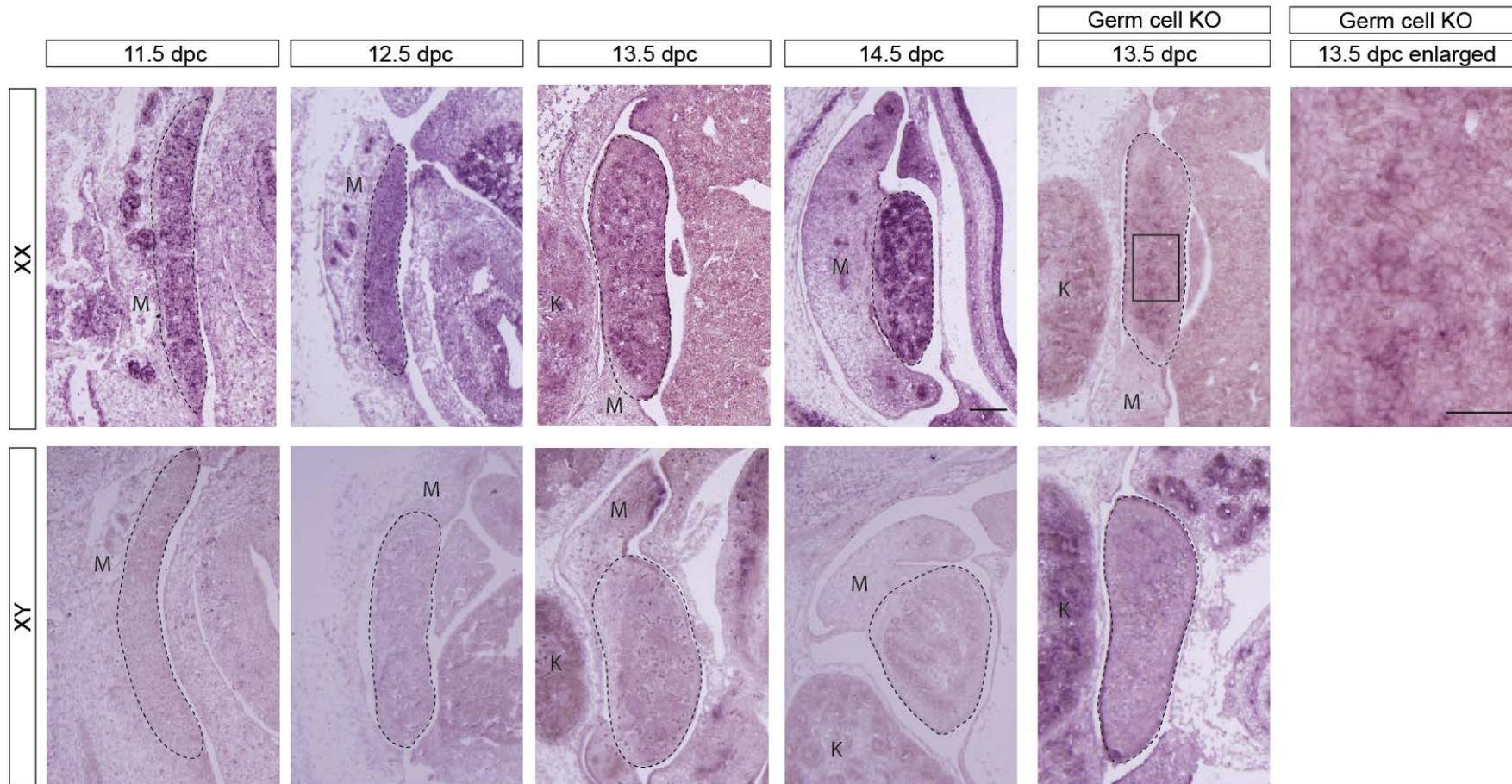
For *AK036014* (*OncB*), ovarian germ cell expression was evident from 11.5 to 14.5 dpc. *OncB* was also expressed at the anterior pole of the mesonephros at 11.5, 12.5 and 14.5 dpc in XX mouse embryos. Section ISH performed on the XX and XY  $W^e$  mouse strain at 13.5 dpc showed reduced *OncB* expression compared to WT embryos suggesting that *OncB* may be expressed in germ cells and/or *OncB* expression is dependent on germ cells. In contrast, *OncB* expression was not detected in XY gonads at the stages tested, while expression was detected in the kidney of 13.5 dpc WT and  $W^e/W^e$  XY gonads (Figure 12).

# Onc3



**Figure 11. Pattern of Onc3 expression in the developing gonads by *in situ* hybridisation.** Onc3 expression (purple staining) is evident in germ cells of XX gonads (outlined by black dashed line) at 11.5 and 13.5 dpc. No expression is detectable in XY gonads from 11.5 to 14.5 dpc. Onc3 expression is also detectable in the kidney (K) at 13.5 dpc and in the mesonephric tubules at 11.5 dpc (indicated by black arrow). Scale bars, 100  $\mu$ m (left panels), 50  $\mu$ m (right panel).

# OncB



**Figure 12: Pattern of OncB expression in the developing gonads by *in situ* hybridisation.** OncB expression (purple staining) is detectable in XX gonads (outlined by dotted lines) ranging from 11.5 to 14.5 dpc, while no expression is visible in XY gonads from 11.5 to 14.5 dpc. Little expression of OncB was detected in  $W^e/W^e$  XX gonads at 13.5 dpc. No expression was detectable in  $W^e/W^e$  XY gonads at 13.5 dpc. Expression in the kidney (K) was apparent in 13.5 dpc XX and XY WT and  $W^e/W^e$  embryos. Expression was detectable in the anterior pole of the mesonephric tissue (M) in XX gonads at 11.5, 12.5 and 14.5 dpc. Scale bar, 100  $\mu$ m (left panels), 50  $\mu$ m (right panel).

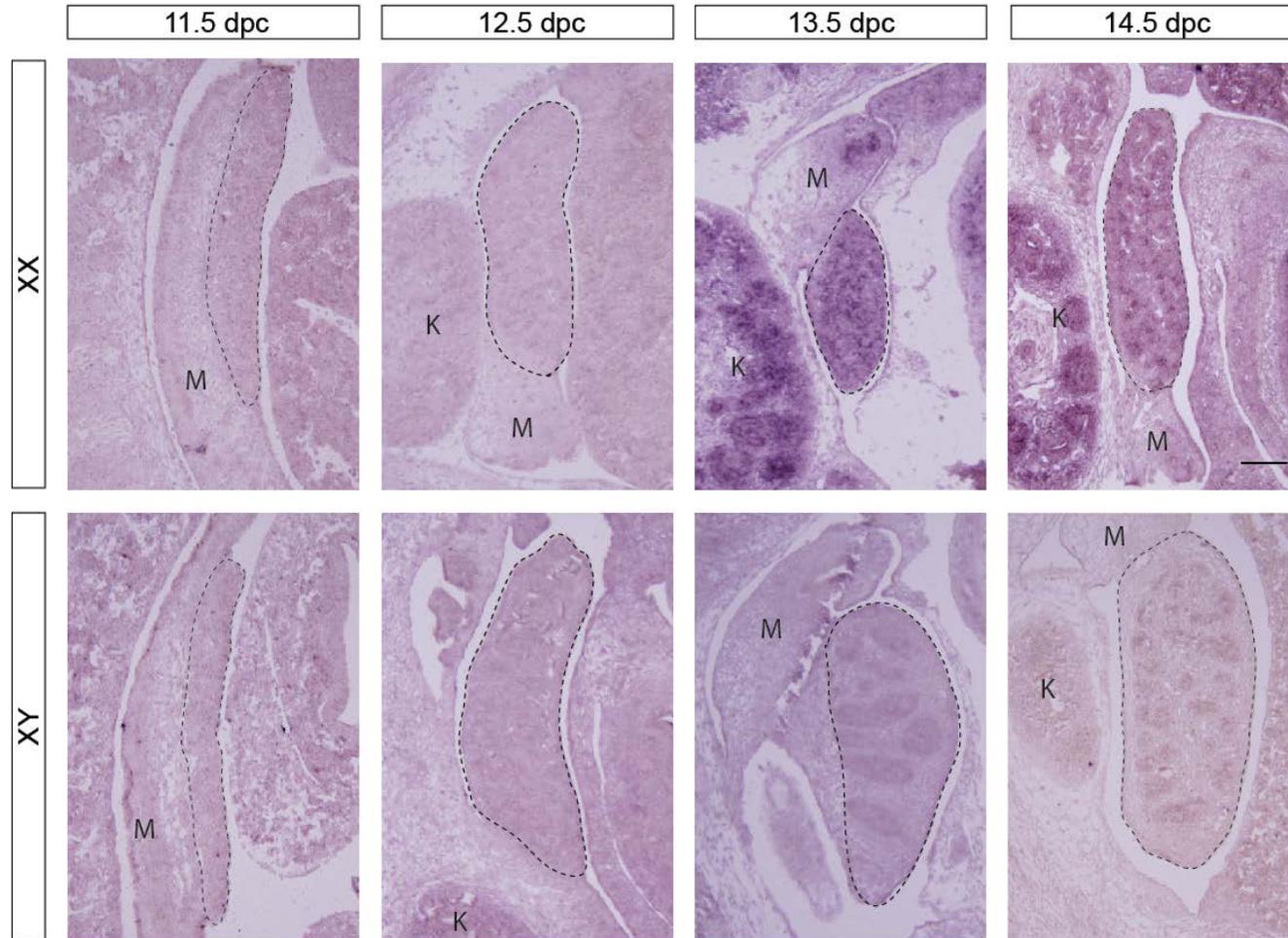
No *AK182836* (OncF) expression was detected at 11.5 and 12.5 dpc XX gonads but expression was subsequently up-regulated in ovarian germ cells from 13.5 dpc to 14.5 dpc. Similar to most Oncs, no expression of OncF was detected in the XY embryos at all stages. OncF expression was also detected in the developing ribs (data not shown) and 13.5 and 14.5 dpc XX kidney (Figure 13).

*AK015136* (OncK) expression was detected in 13.5 to 14.5 dpc XX gonads and the expression pattern suggested that OncK was expressed in ovarian germ cells. Conversely, no expression was detected in testes at all stages investigated. The only other OncK expression detected was in the liver of 14.5 dpc XX embryos, expression was not observed in kidney or mesonephric tissue (Figure 14).

No specific expression was detected for *AK044909* (OncL) in both XX gonads at 12.5 and 13.5 dpc, however expression was detected in 11.5 and 14.5 dpc XX gonads. Expression of OncL was also evident in the liver of 11.5 dpc XX embryo, as well as in the trigeminal ganglia (data not shown) at 13.5 dpc. Staining in testis cords was observed in 13.5 dpc testis, however this was considered to be unspecific trapping of the colour solution. Overall no specific OncL expression was detected in XY gonads at all stages (Figure 15).

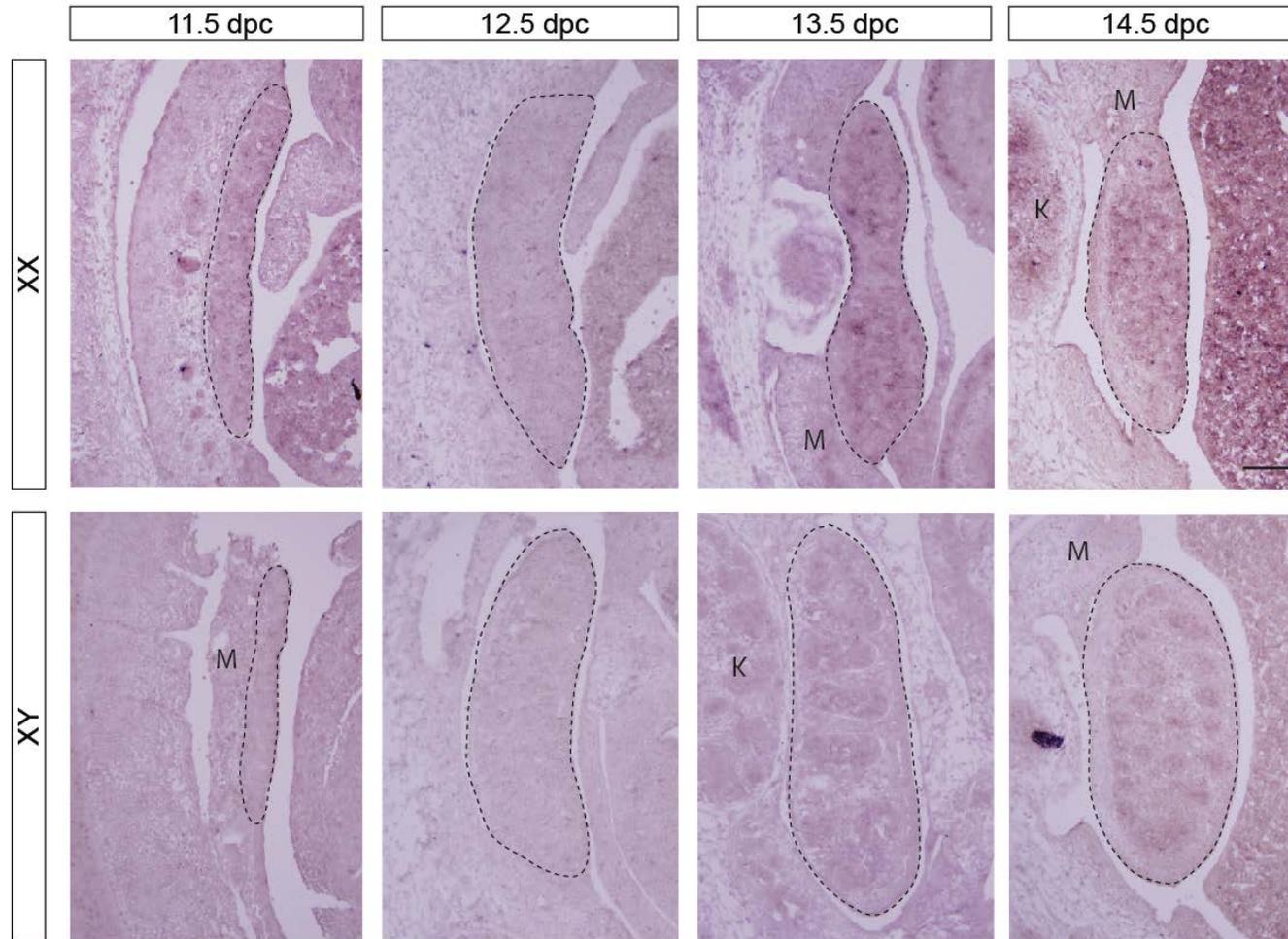
Overall, the expression patterns suggested that most of the ovary-enriched long ncRNAs investigated were expressed in germ cells. Some of the Oncs had expression throughout the developmental stages investigated while others were only expressed at certain stages. In addition, expression was only enriched in XX gonads but not XY gonads, which confirmed the sexually dimorphic expression of Oncs. Expression in kidney and the mesonephros was detected in some but not all embryos.

# OncF



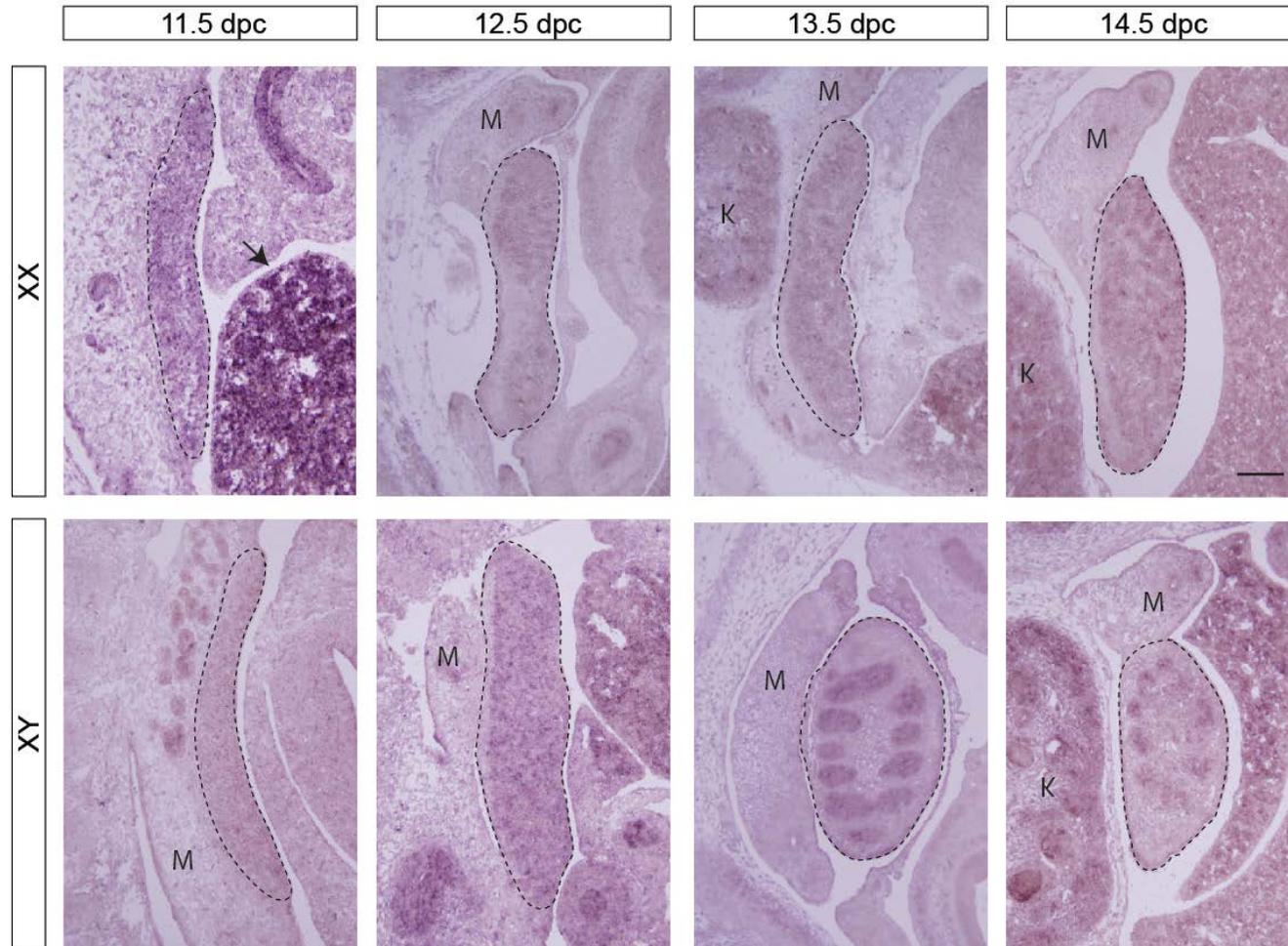
**Figure 13. Pattern of *OncF* expression in the developing gonad by *in situ* hybridisation.** *OncF* expression (purple staining) is evident in XX gonads (outlined by black dashed lines) at 13.5 and 14.5 dpc but no expression is detectable at 11.5 and 12.5 dpc. No expression is detectable in XY gonads from 11.5 to 14.5 dpc. Expression in the kidney (K) is observed in 13.5 and 14.5 dpc XX embryos. Scale bar=100  $\mu$ m.

# Onck



**Figure 14. Pattern of Onck expression in the developing gonad by *in situ* hybridisation.** Onck expression (purple staining) is detectable in XX gonads (outlined by black dashed lines) at 13.5 and 14.5 dpc but no expression is detectable at 11.5 and 12.5 dpc. No expression is detectable in XY gonads from 11.5 to 14.5 dpc. Expression in the kidney (K) is observed in 14.5 dpc XX embryos. No expression is detectable in the mesonephric tissue (M). Scale bar=100  $\mu$ m.

# OncL



**Figure 15. Pattern of OncL expression in the developing gonad by *in situ* hybridisation.** OncL expression (purple staining) is evident in XX gonads (outlined by black dashed lines) at 11.5 and 14.5 dpc but no expression is visible at 12.5 and 13.5 dpc. No expression is visible in XY gonads from 11.5 to 14.5 dpc. Expression in the liver (indicated by black arrow) is observed in 11.5 dpc XX embryos. Scale bar=100  $\mu$ m.

A second group of lncRNAs, which are enriched in the testis but not the ovary, were called Tncs in this study, which stands for testis-enriched long ncRNAs. Expression of Tncs within testis cords could indicate Sertoli and/or germ cell expression. It is difficult to distinguish the cell types even in high magnification images. Unlike mRNA, which is localized to the cytoplasm, ncRNAs can also be localized to the nucleus [238]. The cytoplasm of Sertoli cells wrap around the germ cells, therefore expression of Tncs within testis cords could be from germ cells, Sertoli cells or from both.

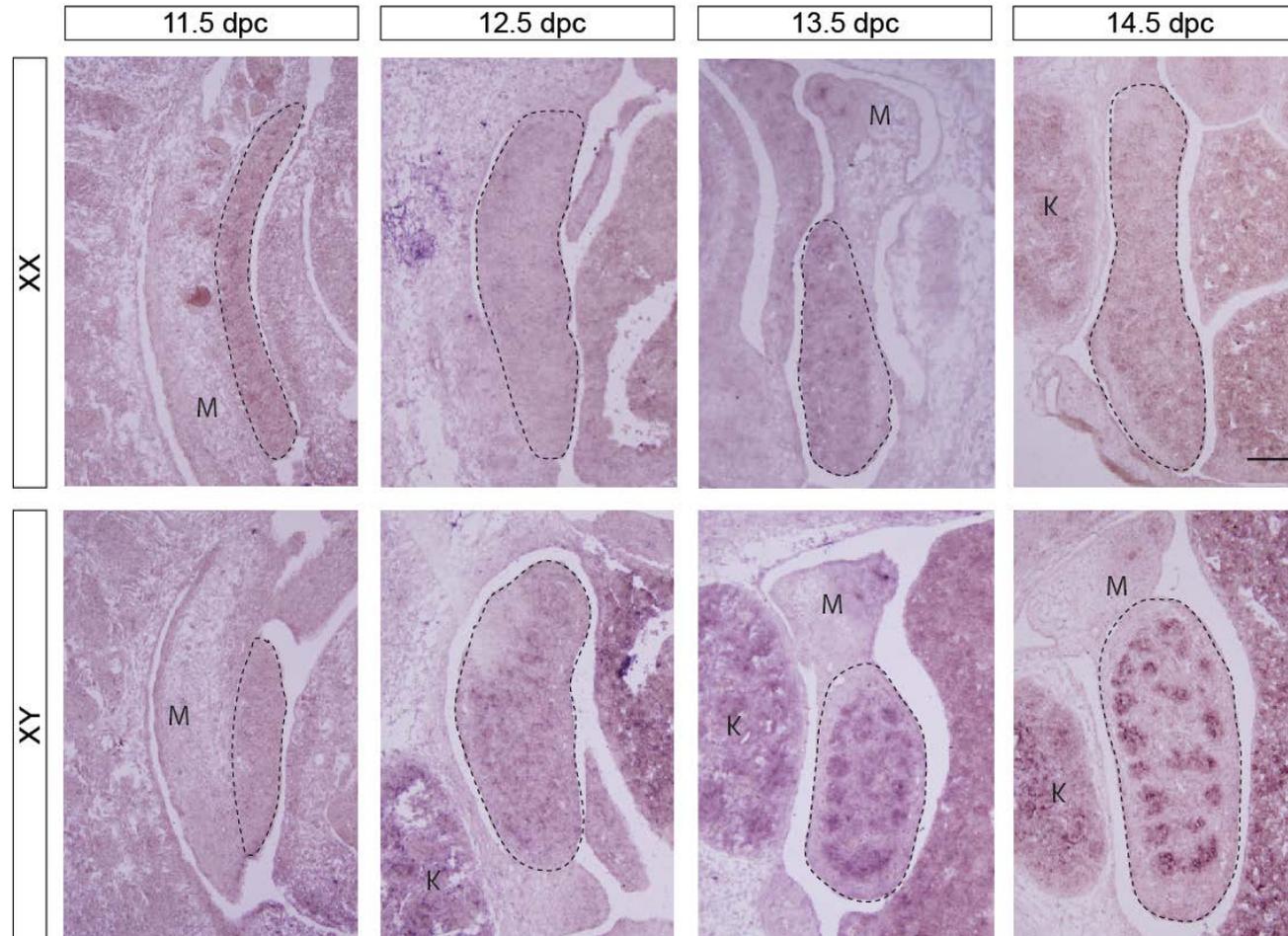
In XY gonads expression of *AK045786* (Tnc10) was detected at 12.5 dpc, by 13.5 dpc expression was detected within the testis cords, which was still evident at 14.5 dpc. No specific expression was observed in 11.5 to 14.5 dpc ovary. In addition, expression in the kidney was detected in 12.5 to 14.5 dpc XY embryos (Figure 16).

In testis cords of XY embryos *AK013819* (Tnc6) expression was detected at 13.5 and 14.5 dpc, a similar expression pattern that was observed for *AK013488* (Tnc7). For both Tnc6 and Tnc7, expression in the kidney and liver was evident in 13.5 and 14.5 dpc XY embryos. No expression can be detected in XX embryos at all stages investigated (Figure 17A and B).

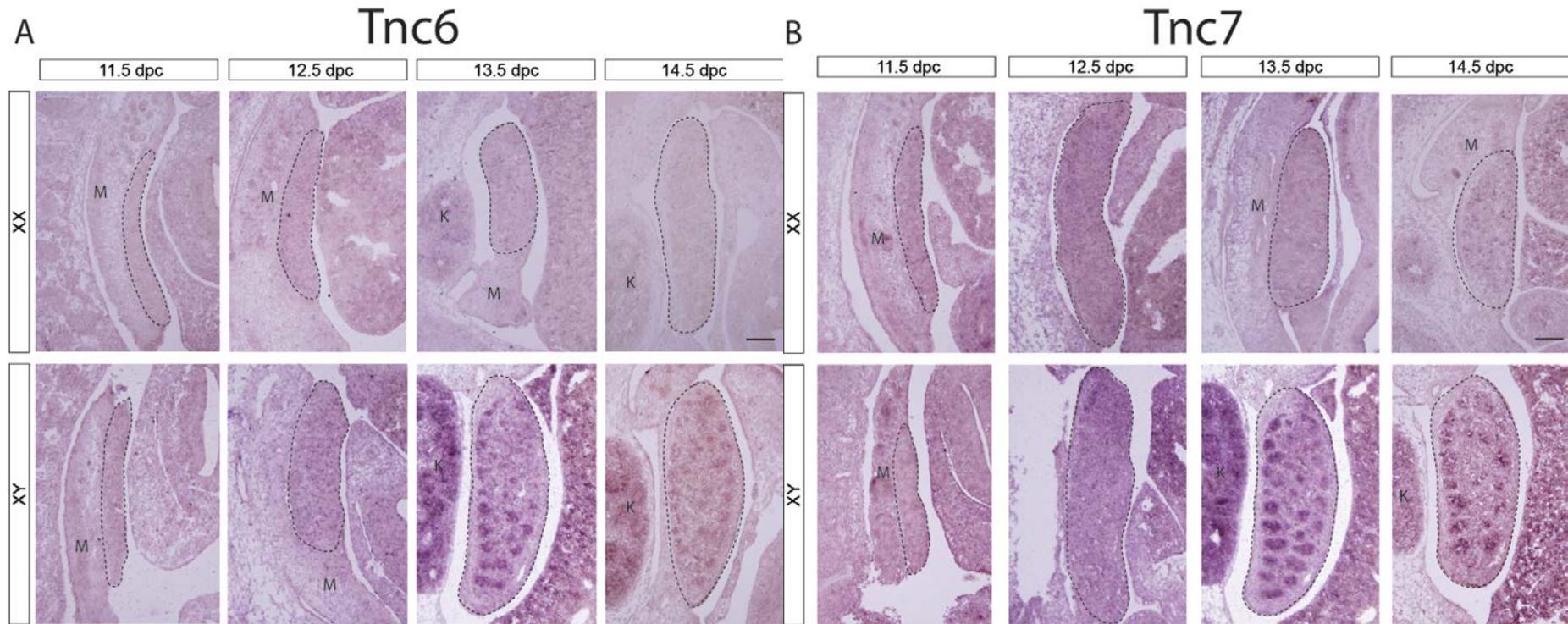
Similar to Tnc6 and 7, *AK005877* (Tnc9) expression was detectable in the testis cords of 13.5 and 14.5 dpc XY gonads. Additionally, Tnc9 was also expressed in 11.5 dpc XY gonad, 11.5 dpc XX and XY mesonephric tubules, and the kidney of 13.5 dpc XX and XY embryos (Figure 18).

No specific *AK034891* (Tnc3) expression was detected in XX gonads at all stages investigated. Conversely, Tnc3 expression was detected in 13.5 dpc XY gonad and kidney, however not at other stages (Figure 19).

# Tnc10

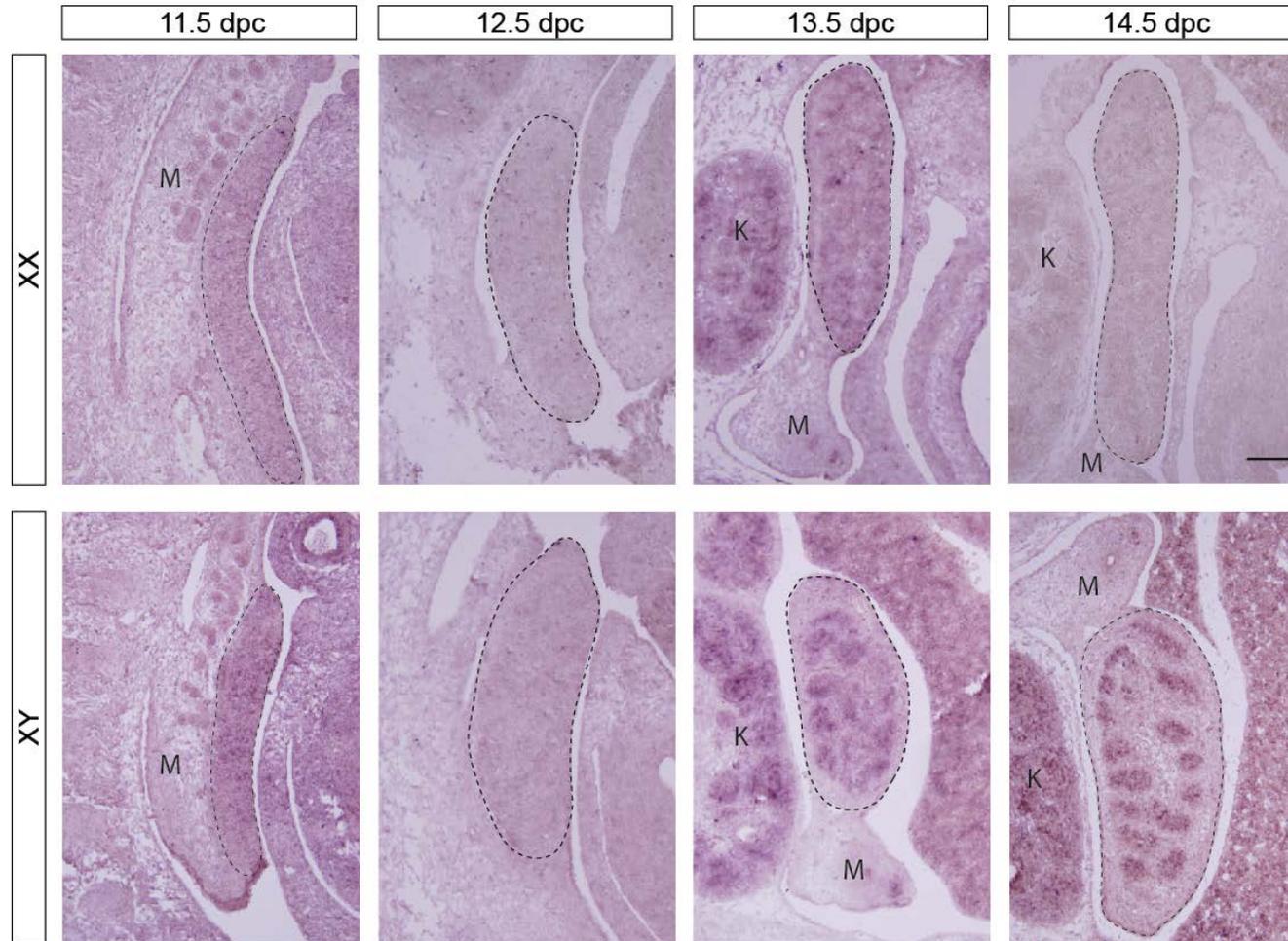


**Figure 16. Pattern of Tnc10 expression in the developing gonad by *in situ* hybridisation.** Tnc10 expression (purple staining) is evident in XY gonads (outlined by black dashed lines) at 12.5 to 14.5 dpc but no expression is visible in the testis at 11.5 dpc. No expression is detected in XX gonads from 11.5 to 14.5 dpc. Expression in the kidney (K) is apparent in 12.5 to 14.5 dpc XY embryos. Scale bar, 100 $\mu$ m.



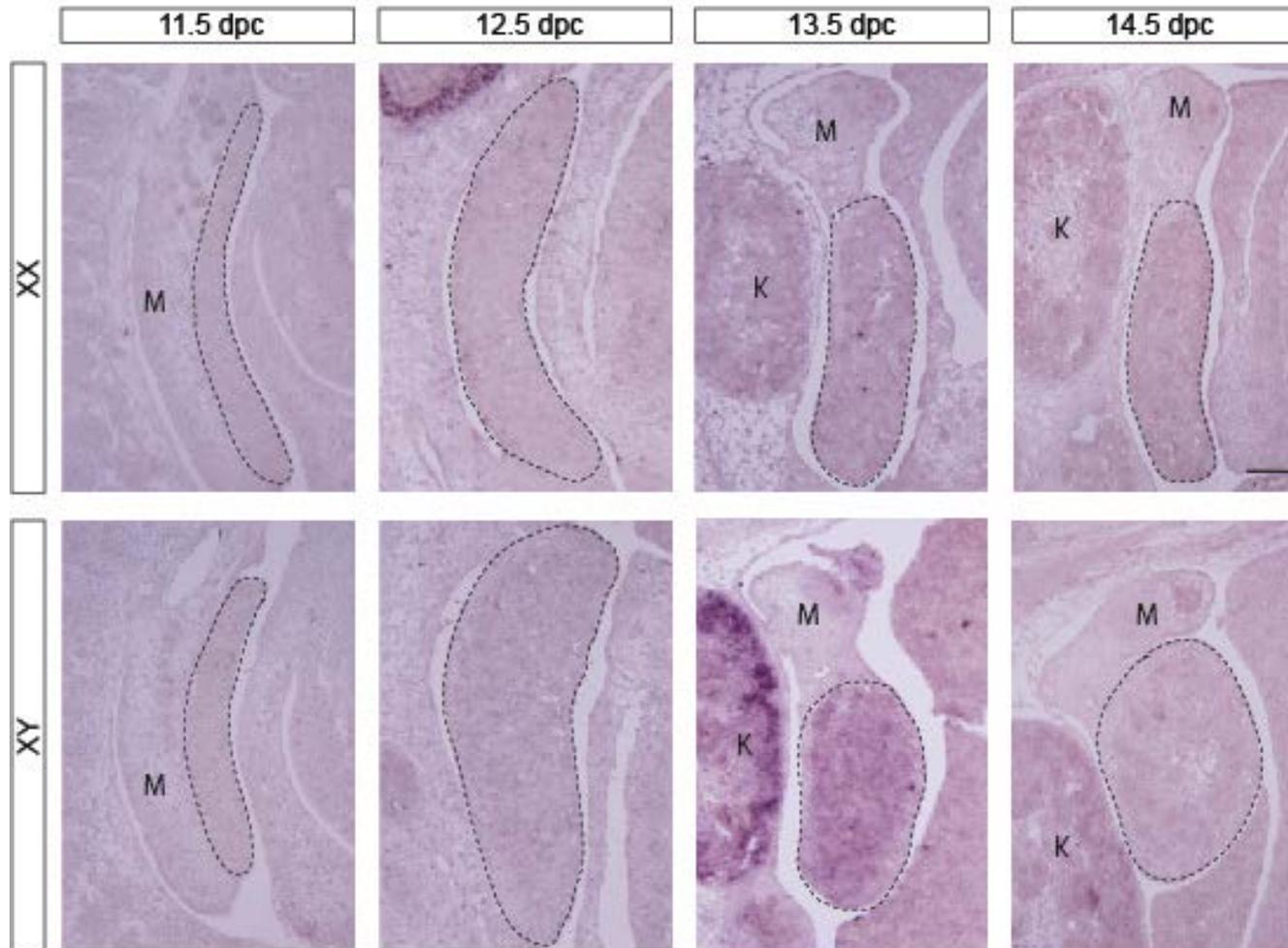
**Figure 17. Pattern of Tnc6 (A) and Tnc7 (B) expression in the developing gonad by *in situ* hybridisation.** Tnc6 and 7 expression (purple staining) are evident in XY gonads (outlined by black dashed lines) at 13.5 and 14.5 dpc but no expression is visible in the XY gonad at 11.5 and 12.5 dpc. No expression is detectable in XX gonads at all stages. Expression in the kidney (K) is apparent in 13.5 and 14.5 dpc XY embryos, but no expression is evident in mesonephric tissue (M). Scale bar = 100 $\mu$ m.

# Tnc9



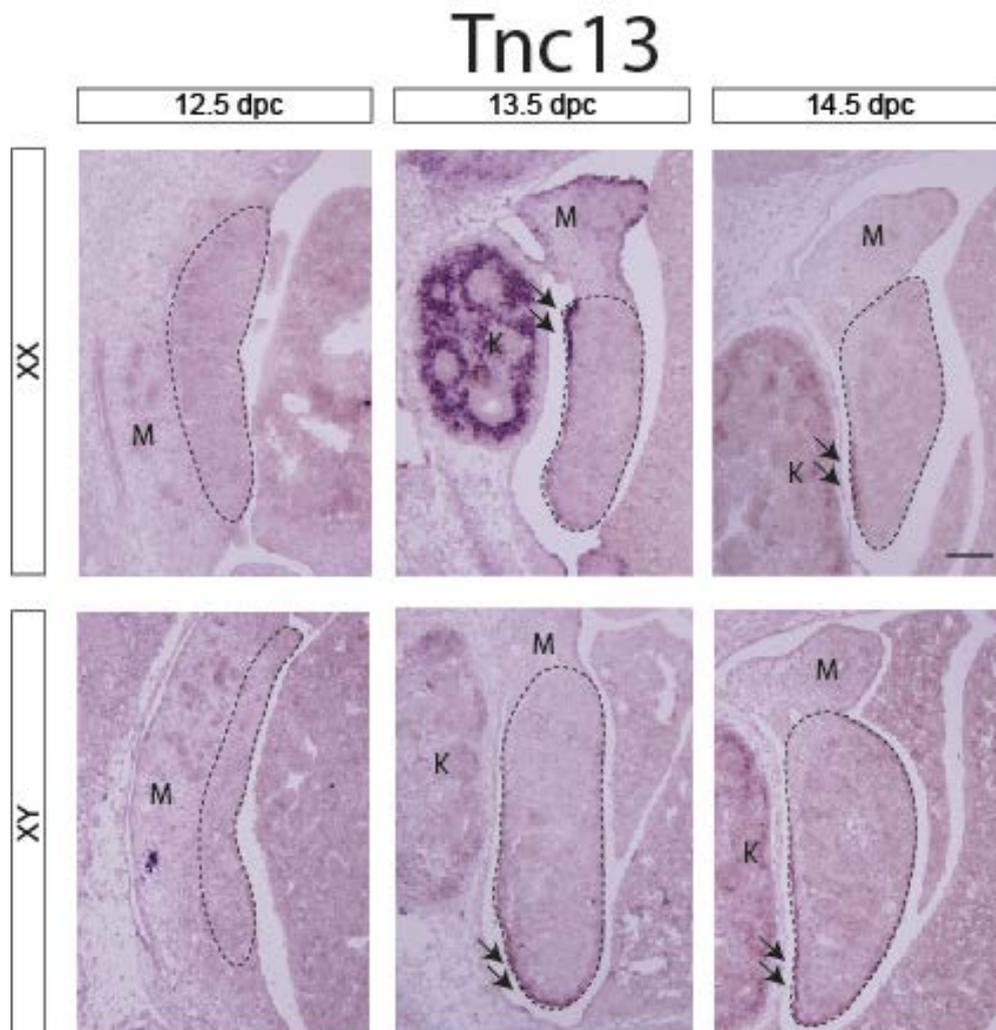
**Figure 18. Pattern of Tnc9 expression in the developing gonad by *in situ* hybridisation.** Tnc9 expression (purple staining) is evident in XY gonads (outlined by black dashed lines) at 11.5, 13.5 and 14.5 dpc but no expression is detectable in the XY gonad at 12.5 dpc. No expression is detectable in XX gonads at 11.5 to 14.5 dpc. Expression in the kidney (K) is apparent in 13.5 dpc XX embryos, as well as 13.5 and 14.5 dpc XY embryos. Scale bar = 100µm.

# Tnc3



**Figure 19. Pattern of Tnc3 expression in the developing gonad by *in situ* hybridisation.** Tnc3 expression was visible in XY gonads at 13.5 dpc. Expression in the kidney (K) was observed in 13.5 dpc XY embryos. No other expression was detected. Scale bar = 100µm.

According to a previous study, *AK043086* (*Tnc13*) is expressed in the coelomic epithelium at the anterior pole of the testis and ovary (Wilhelm *et al*, unpublished data). Figure 20 shows specific expression in the coelomic epithelium of 13.5 and 14.5 dpc XY and XY embryos, while no expression was detected in 12.5 dpc XX and XY embryos. This observation further confirmed the previous finding. In addition, expression in the kidney was also observed in 13.5 dpc XX and 14.5 dpc XX and XY embryos (Figure 20).

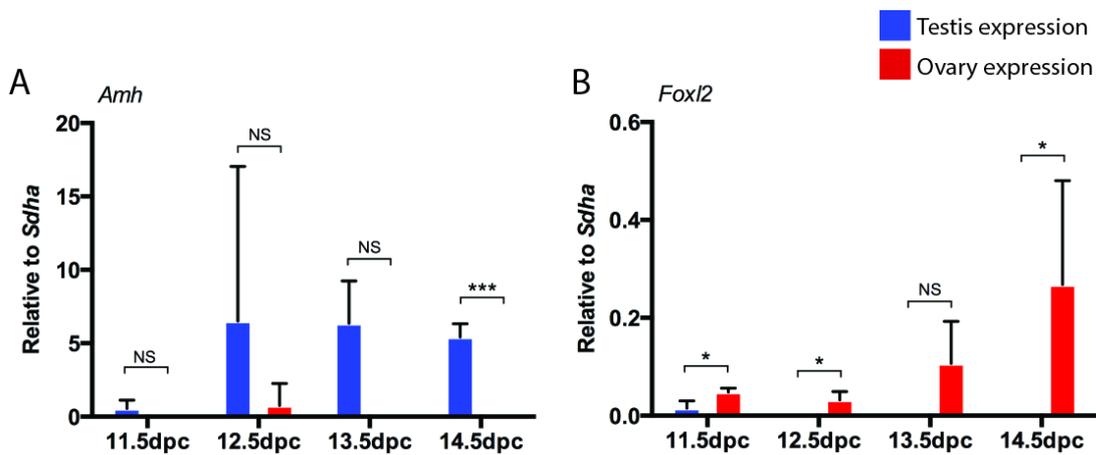


**Figure 20. Pattern of *Tnc13* expression in the developing gonad by *in situ* hybridisation.** *Tnc13* expression (purple staining) is evident in coelomic epithelium (indicated by black arrows) of XY and XX gonads (outlined by black dashed lines) at 13.5 and 14.5 dpc. Expression in the kidney (K) is apparent in 13.5 dpc XX and 14.5 dpc XX and XY embryos. No expression was evident in mesonephric tissue (M). Scale bar = 100µm.

Analysis of *Tncs* revealed an overall trend that most *Tncs* were expressed from 13.5 dpc onwards. For some of the *Tncs*, polarised expression at the anterior pole of the mesonephros was observed. Moreover, expression in the kidney was detected in some but not all embryos.

## 2. Quantification of long ncRNAs expression in embryonic mouse gonads showed sexual dimorphic pattern.

After showing that long ncRNAs are expressed sexually dimorphically in the developing gonad through ISH, the next step was to quantify long ncRNA expression levels in the gonad at different developmental stages. In order to do this, quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed using RNA from embryonic WT XX and XY mouse gonads of 11.5 to 14.5 dpc.



**Figure 21. qRT-PCR quantification of *Amh* and *Foxl2* mRNA from 11.5 to 14.5 dpc.** qRT-PCR analysis of isolated XX and XY gonads from 11.5, 12.5, 13.5 and 14.5 dpc mouse embryos using gene specific primers for *Amh* (A) and *Foxl2* (B) relative to *Sdha*. Error bars= ± SEM; n=3; P-values = NS (not significant) >0.05, \* <0.05, \*\*\* <0.001. Blue and red bars represent testis and ovary expression respectively.

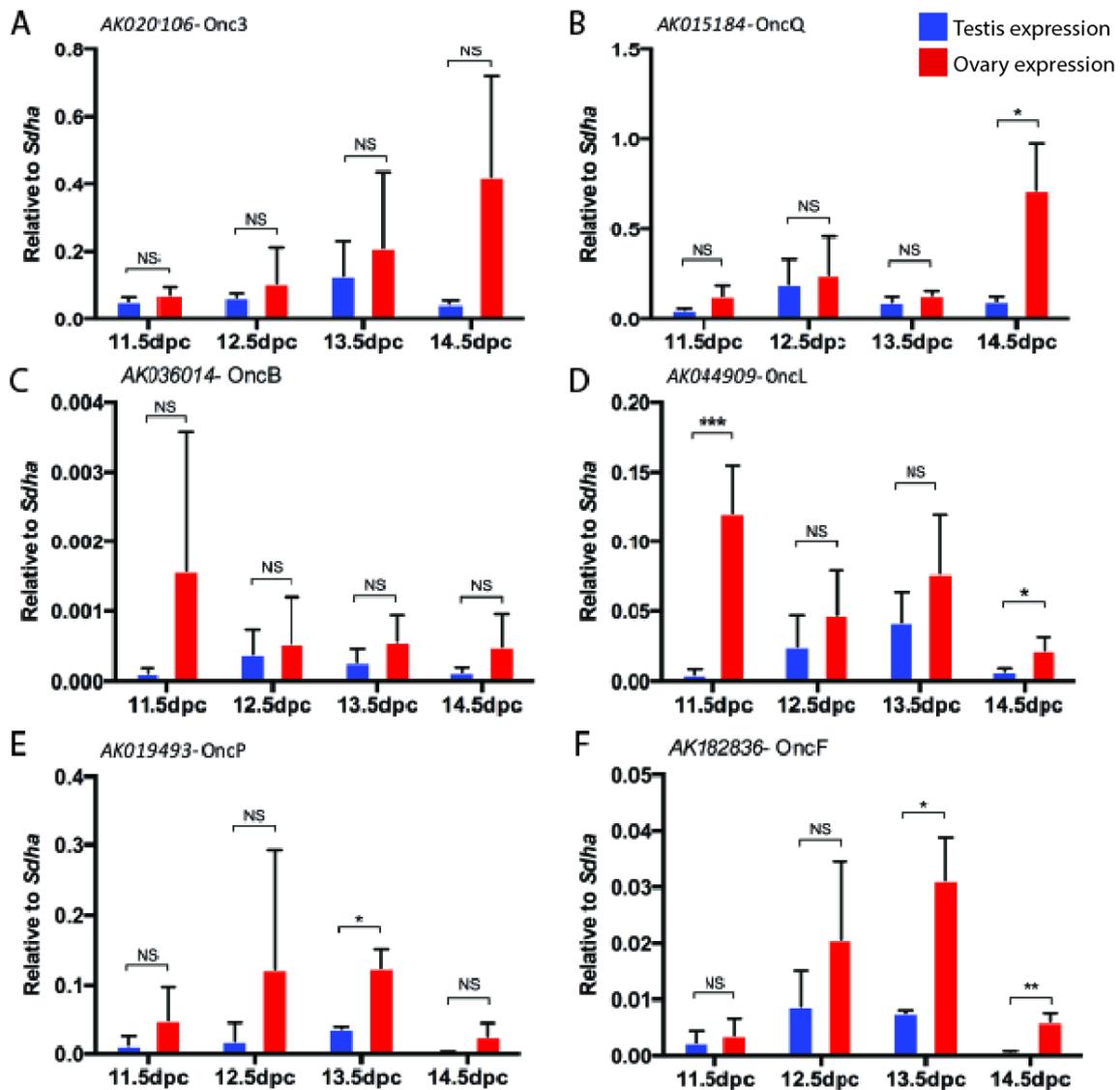
The expression level of the positive control for testis expression *Amh* was as expected higher in XY gonads compared to XX gonads at 14.5 dpc (P=0.000161). However from 11.5 to 13.5 dpc, *Amh* expression was not statistically significantly higher in XY gonads than in XX gonads due to the large variation between the three independent biological replicates (Figure 21A). The expression levels of *Foxl2* were significantly higher in XX gonads than in XY gonads at 11.5, 12.5 and 14.5 dpc (p=0.045672, 0.017006 and 0.013867 respectively) (Figure 21B).

As described in the introduction, testis-specific *Amh* and ovary-specific *Foxl2* expression increases in levels throughout early gonad development. As expected, the current results showed an increase in expression of *Foxl2* specifically in ovaries as development progresses. Although not statistically significant, the data for *Amh* showed the expected

trend towards a higher expression in the XY gonad compared with the XX gonad from 12.5 to 14.5 dpc.

As expected, expression levels of all ovary-enriched long ncRNAs (Oncs) appeared consistently higher in XX gonads compared to XY gonads at all developmental stages, however many differences were not statistically significant. This is due to the large variation between the three independent biological replicates. An increase in biological replicates may correct for this natural variation. As long ncRNA expression is often relatively low, variation in cDNA input can greatly affect the qRT-PCR results.

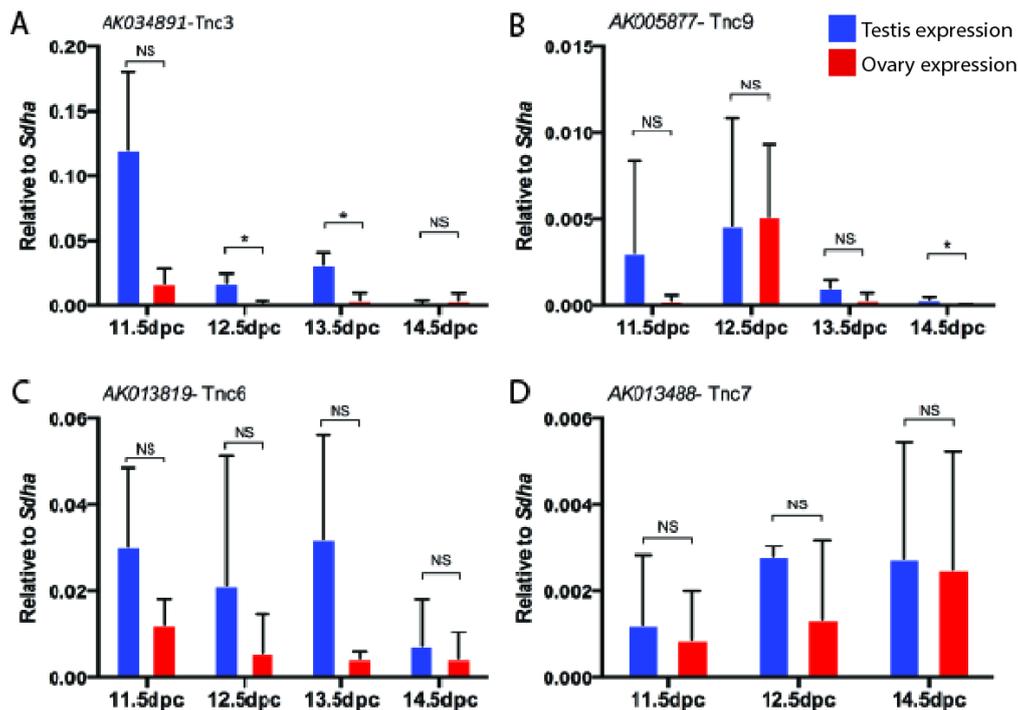
Expression of Onc can be classified into 3 groups. In group A, Onc3 (Figure 22A) and OncQ (Figure 22B) both had gradual increase in expression levels in XX gonads from 11.5 dpc to 14.5 dpc. For Onc3, expression increased at a constant rate from 11.5 to 13.5 dpc in both XX and XY gonads, but at higher levels in the XX compared to the XY gonad. Onc3 expression reached a peak at 14.5 dpc in XX gonad, whereas expression was reduced at 14.5 dpc in the XY gonad, although not statistically significant ( $P= 0.212477$ ). This expression was not consistent with the ISH data, suggesting that Onc3 ISH for 14.5 dpc XX embryos may not have worked. Similarly, OncQ expression also peaked at 14.5 dpc in the XX gonad, and the expression was statistically significant higher than in testes ( $P=0.014892$ ).



**Figure 22. qRT-PCR of selected ovary-enriched long ncRNAs in mouse embryonic gonads from 11.5 to 14.5dpc.** qRT-PCR analysis of isolated XX and XY gonads from 11.5, 12.5, 13.5 and 14.5 dpc mouse embryos using gene specific primers for *AK020106*, *Onc3* (A) *AK015184*, *OncQ* (B) *AK036014*, *OncB* (C) *AK044909*, *OncL* (D) *AK019493*, *OncP* (E) and *AK182836*, *OncF* (F) relative to *Sdha*. Error bars =  $\pm$  SEM; n=3; P-values = NS (not significant) >0.05, \* <0.05, \*\* <0.01, \*\*\* <0.001. Blue and red bars represent testis and ovary expression respectively.

Group B consists of *OncB* (Figure 22C) and *OncL* (Figure 22D). In contrast to group A, expression of *OncB* and *OncL* peaked at 11.5 dpc before gradually decreasing until 14.5 dpc. For *OncL*, significant difference in expression was observed between 11.5 dpc testes and ovary (P=0.00045), the expression in the XX gonad was decreased at 12.5 dpc gonad before increasing slightly at 13.5 dpc, then lowered again at 14.5 dpc (P=0.017401).

Finally, *OncP* (Figure 22E) and *OncF* (Figure 22F) were classified into group C. Both had a gradual increase of expression from 11.5 to 13.5 dpc before expression dropped down at 14.5 dpc. Expression of *OncF* and *OncP* was statistically significant higher in 13.5 dpc XX than in XY gonad with P values of 0.006195 and 0.041607 respectively, and *OncF* expression was also statistically significant higher in 14.5 dpc ovary than in testes (P=0.006741).



**Figure 23.** qRT-PCR quantification of selected testis enriched long ncRNAs expression in mouse embryonic gonads from 11.5 to 14.5dpc. qRT-PCR analysis of isolated XX and XY gonads from 11.5, 12.5, 13.5 and 14.5 dpc mouse embryos using gene specific primers for *AK034891*, *Tnc3* (A) *AK005877*, *Tnc9* (B) *AK013819*, *Tnc6* (C) and *AK013488*, *Tnc7* (D) relative to *Sdha*. Error bars=  $\pm$  SEM; n=3; P-values = NS (not significant) >0.05, \* <0.05. Blue and red bars represent testis and ovary expression respectively.

On the other hand, expression patterns varied between the *Tncs* investigated by qRT-PCR. *Tnc3* expression peaked at 11.5 dpc in XX gonads then declined afterwards (Figure 23A). *Tnc3* appeared to be expressed at higher levels in the XY gonads when compared with the expressions in the XX gonads, however statistically significance was only found at 12.5 and 13.5 dpc (P=0.020796 and 0.020434 respectively).

Relative expression levels of Tnc9 showed that this long ncRNA was expressed at higher levels in 11.5 dpc XY gonads compared to 11.5 dpc XX gonads (Figure 23B). Interestingly, by 12.5 dpc expression levels were similar between XX and XY gonads. Tnc9 expression became sexually dimorphic again at 13.5 and 14.5 dpc with higher expression in the XY gonad compared to the XX gonads. At 14.5 dpc, although the difference of expression in XX and XY gonad was found to be statistically significant ( $P=0.023567$ ), the overall expression of Tnc9 at this time point was very low.

Relative expression levels of Tnc6 showed that this long ncRNA was expressed at consistent higher levels in XY compared to XX gonads at all stages from 11.5 to 13.5 dpc before declining at 14.5 dpc in the XY gonad. However, at none of the stages was the expression found to be statistically significantly different between testes and ovaries (Figure 23C).

Lastly for Tnc7, high expression was observed in both XX and XY gonads at 11.5, 12.5 and 14.5 dpc (qRT-PCR failed to work for Tnc7 at 13.5 dpc), with expression slightly higher in XY than in XX gonads, which indicates that Tnc7 may not be sexually dimorphically expressed. However, this is in contrast to the result of the ISH in which Tnc7 was found to be enriched in the testis cords at 13.5 and 14.5 dpc (Figure 23D).

Although the large error bars and variations between the results made it difficult to draw any conclusions from the qRT-PCR data, the overall trend of high expression of Tncs and Oncs in the testis and ovary respectively at 11.5 to 14.5 dpc was mostly consistent with ISH data. The enrichment of most of the long ncRNA candidates in gonads in these embryonic stages may be consistent with their role in mammalian gonadal development.

**Table 2. Summary of long ncRNA candidates validated in the current study.**

<b>Long ncRNA</b>	<b>Microarray [233]</b>	<b>ISH</b>	<b>qPCR</b>
Onc3 (AK020106)	Preferentially expressed in ovary from 12.5 to 14.5 dpc.	Enriched ovarian germ cells expression at 11.5 and 13.5 dpc.	Increase in expression in XX gonads from 11.5 dpc to 14.5 dpc, peaks at 14.5 dpc.
OncQ (AK015184)	Highly enriched in ovary at 12.5 and 13.5 dpc.	Not validated in the current study.	Increase in expression in XX gonads from 11.5 dpc to 14.5 dpc, sexually dimorphic at 14.5 dpc.
OncB (AK036014)	Highly enriched in ovary at 12.5 and 13.5 dpc.	Enriched ovarian germ cell expression from 11.5 to 14.5 dpc.	Decrease in expression in XX gonads from 11.5 dpc to 14.5 dpc, peaks at 11.5 dpc.
OncL (AK044909)	Preferentially expressed in ovary from 12.5 to 14.5 dpc.	Ovarian germ cell expression at 11.5 and 14.5 dpc.	Decrease in expression in XX gonads from 11.5 dpc to 14.5 dpc, sexually dimorphic at 11.5 dpc.
OncP (AK019493)	Highly enriched in ovary from 11.5 to 13.5 dpc.	Not validated in the current study.	Increase in expression in XX gonads from 11.5 dpc to 13.5 dpc, sexually dimorphic at 13.5 dpc.
OncF (AK182836)	N/A	Enriched ovarian germ cell expression from 13.5 to 14.5 dpc.	Increase in expression in XX gonads from 11.5 dpc to 13.5 dpc, sexually dimorphic at 13.5 and 14.5 dpc.
OncK (AK015136)	Highly enriched in ovary at 13.5 dpc.	Ovarian germ cell expression from 13.5 to 14.5 dpc.	Not validated in the current study.
Tnc3 (AK034891)	N/A	Expression in the testis cords at 13.5 dpc.	Decrease in expression in XX gonads from 11.5 dpc to 14.5 dpc, sexually dimorphic at 12.5 dpc and 13.5 dpc.
Tnc9 (AK005877)	N/A	Enriched in the testis cords at 11.5, 13.5 and 14.5 dpc.	Higher expressed in XY than XX at 11.5 dpc. Sexually dimorphic at 13.5 dpc and 14.5 dpc.
Tnc6 (AK013819)	N/A	Enriched in the testis cords at 13.5 and 14.5 dpc.	Expressed at higher levels in XY gonads from 11.5 dpc to 13.5 dpc before expression declines at 14.5 dpc in the XY gonad.
Tnc7 (AK013488)	N/A	Enriched in the testis cords at 13.5 and 14.5 dpc.	Expression slightly higher in XY than in XX gonads at 11.5, 12.5 and 14.5 dpc.
Tnc10 (AK045786)	N/A	Testis cord expression at 13.5 and 14.5 dpc.	Not validated in the current study.
Tnc13 (AK043086)	N/A	Expressed in the coelomic epithelium at the anterior pole of the testis and ovary at 13.5 and 14.5 dpc.	Not validated in the current study.

# RESULTS 2

## FUNCTIONAL ANALYSIS OF ONC3 OVEREXPRESSION

In order to functionally analyse the impact of long ncRNA expression, inducible expression constructs were used to monitor cellular events such as proliferation, migration and apoptosis.

Onc3 was amplified from cDNA of 13.5 dpc WT XX mouse gonads by PCR. Gel electrophoresis performed after cloning PCR confirmed the expected size of the fragment, which was approximately the expected 821bp.

Amplified Onc3 sequences were first ligated into pGEM®-T Easy vector. Restriction enzyme EcoRI was used to check the sizes of the insert as there is a SacI digestion site within Onc3. EcoRI digestion releases the insert from the vector. Gel electrophoresis performed after restriction enzyme digestion showed two fragments, a 3kb fragment for pGEM®-T Easy vector and a 821bp fragment for the insert . One stable integration clones were isolated and carried forward to the next section.

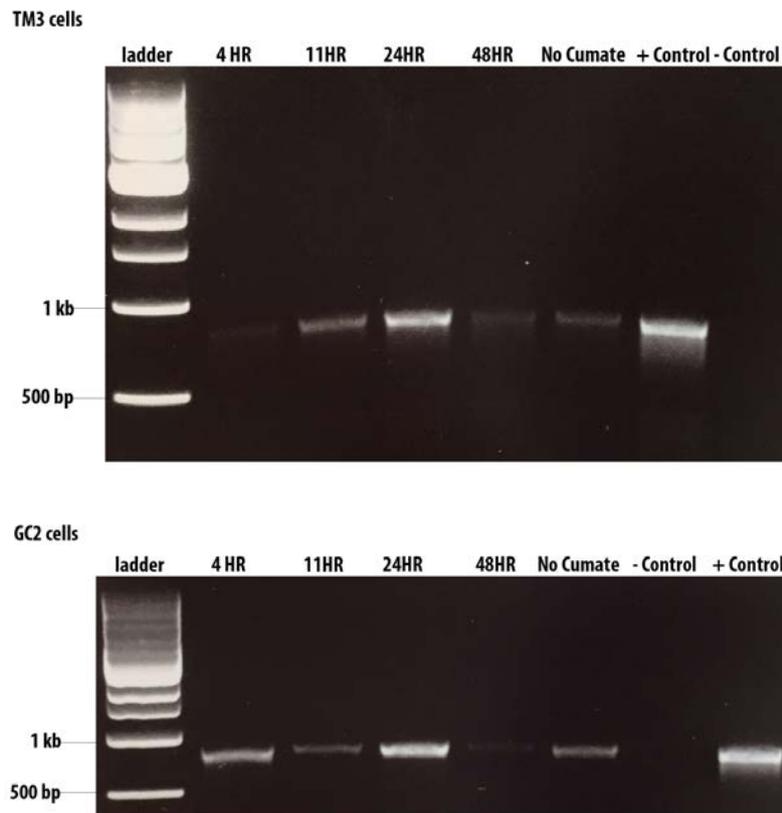
Following this Onc3 was cloned into the PiggyBac vector. Onc3 in pGEM®-T Easy and the empty PiggyBac vector were digested with specific restriction enzymes NotI and NheI before ligation. Gel electrophoresis performed after enzyme digestion showed the release of the Onc3 insert from the pGEM®-T Easy vector and the linearization of the PiggyBac vector.

The PiggyBac vector was dephosphorylated and then ligated with the Onc3 insert. Restriction enzyme SacI was used to check the size of the insert. SacI digested the plasmid twice which results in two fragments of 4820bp and 5455bp. Gel electrophoresis confirmed the successful ligation of the Onc3 with PiggyBac vector.

### **Relative efficacy of cumate activation of Onc3 expression**

First, the relative efficacy of cumate was tested to determine the peak of the long ncRNA expression after activation by cumate.

Both conventional RT-PCR and qRT-PCR were used to assess Onc3 expression activated by cumate at different time points. Gel electrophoresis of the RT-PCR showed that in both GC2 and TM3 cells the strongest expression was at 24 hours after the addition of cumate (Figure 24).



**Figure 24. Cumate efficiency test for TM3 and GC2 cells.** Gel electrophoresis of PCR product for cDNA of Onc3 activated with cumate at 4, 11, 24 and 48 hours, along with 1 kb DNA ladder. Cloned Onc3 cDNA was used as positive control and water as a substitute for cDNA as a negative control. The resulting bands were of the correct sizes (approximately 821 bp).

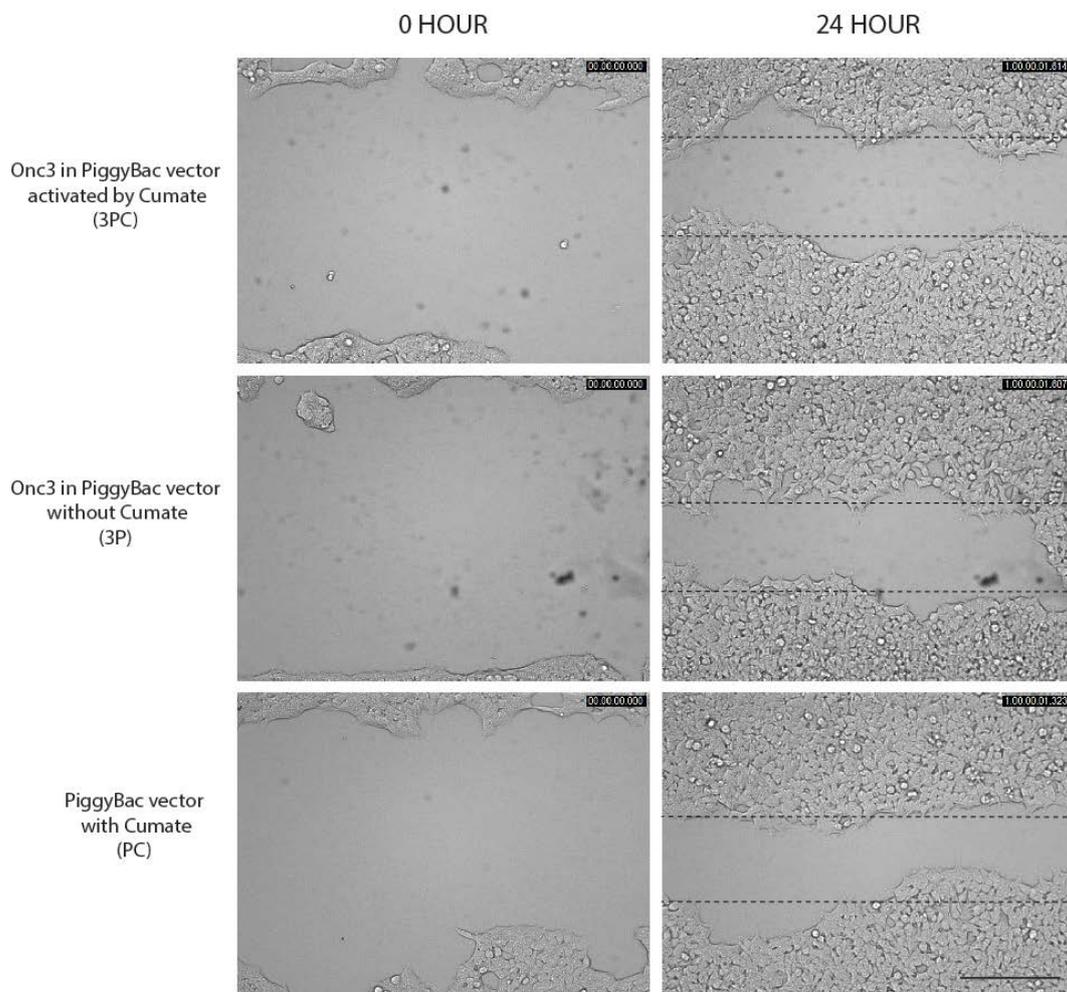
The result was further analysed with qRT-PCR (Supplementary data). The expression levels were normalised to the housekeeping gene *Sdha*. In contrast to the RT-PCR results, the expression of Onc3 at 24 hours for GC2 and TM3 cells was very low. qRT-PCR was only performed once therefore the data collected could not be statistically analysed.

## Expression of Onc3 leads to reduced HEK293 and TM3 cell migration/ proliferation and increased GC2 cell migration/proliferation

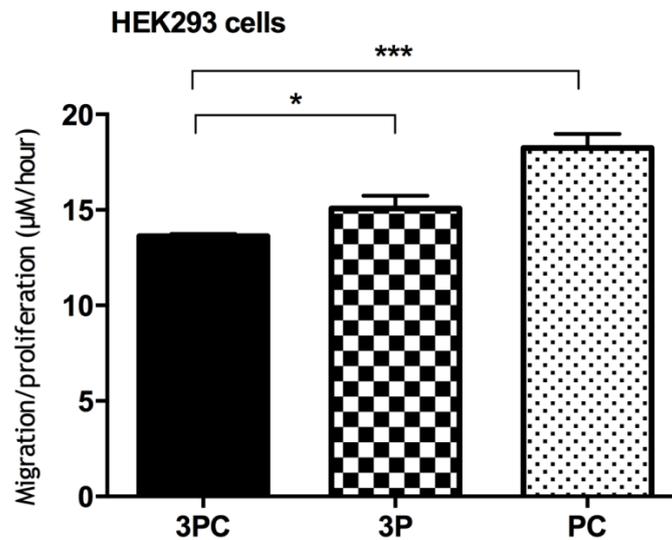
To study the effect of Onc3 on migration of HEK293, TM3 and GC2 cells, live cell imaging analysis was performed.

In general, HEK293 cells migrate together towards the middle of the gap as one front. Comparison of cell morphology showed no differences between 3PC, 3P and PC cells (Figure 25). Analysis of HEK293 cells images at 24 hours showed significant differences in migration/proliferation between 3PC and 3P cells as well as between 3PC and PC cells (Figure 26).

### HEK293 cells



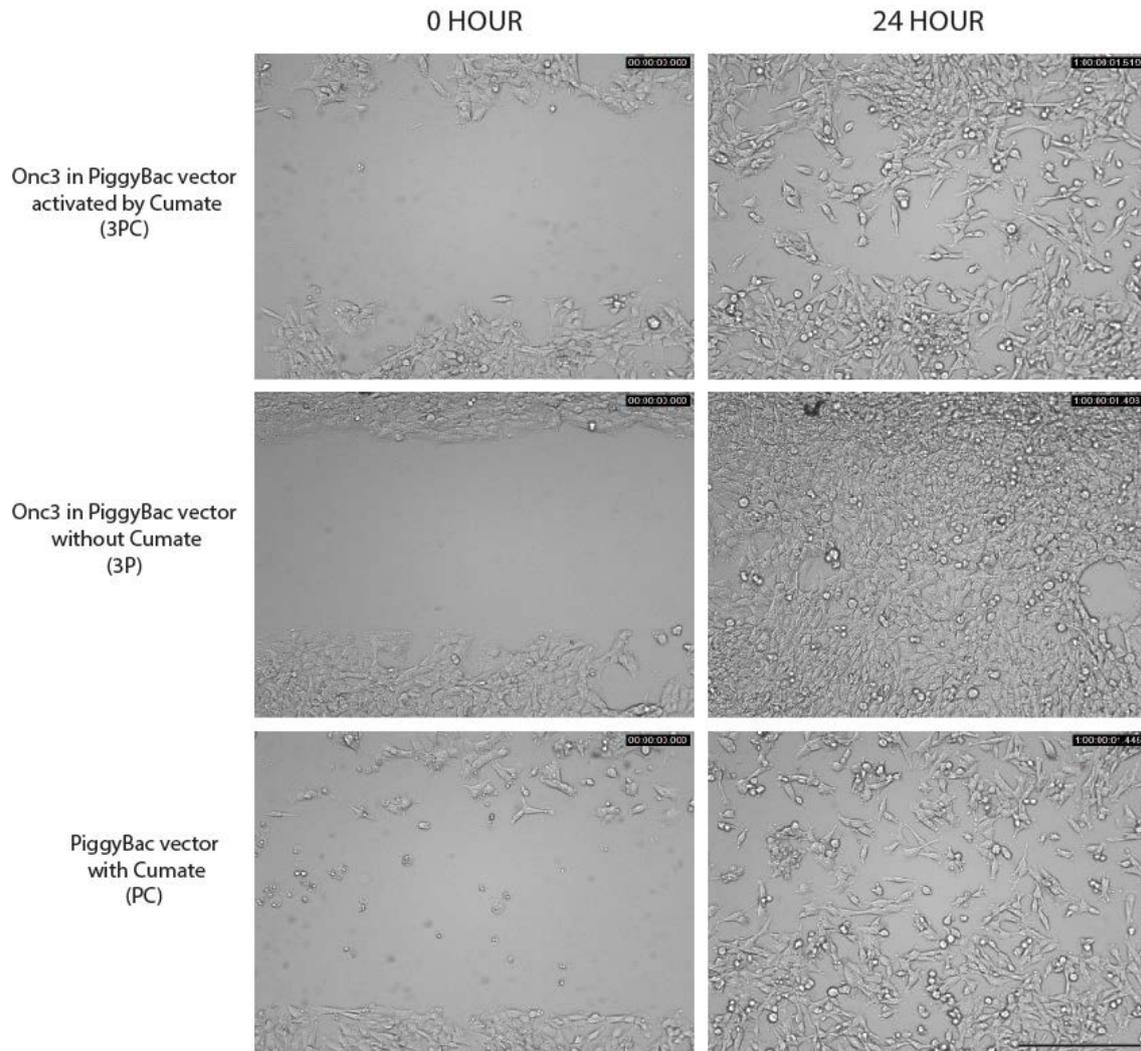
**Figure 25. Analysis of HEK293 cell migration/proliferation by *in vitro* live cell imaging.** Images were acquired once every 15 minutes and images from 0 and 24 hours were used for comparison between 3PC, 3P and PC cells. The dotted lines define the areas lacking cells. Scale bar= 250 $\mu$ m.



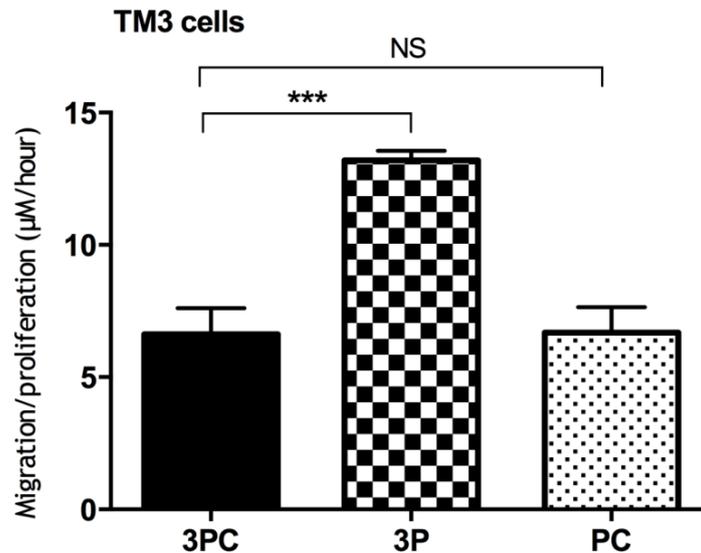
**Figure 26. Migration/proliferation of HEK293 cells.** Areas between the cells were measured at 0 hours and 24 hours then migration/proliferation rate in  $\mu\text{M}$  per hour was calculated and compared between the 3 groups. Statistical analysis showed that 3PC cells migrate at a statistically significant slower rate compared to 3P and PC cells.

Unlike HEK293 cells, TM3 cells migrate more individually in a random fashion. This way of migration made it harder to measure the area of the gap, therefore in future experiments techniques such as Boyden chamber assay may be more suitable for TM3 cells. On the other hand, morphology between the three test groups resembles one another (Figure 27). At 24 hours, statistically significant difference was found between the migration/proliferation rate of 3PC and 3P cells in TM3 cells, with 3PC cells migrate/proliferate at a slower rate. In addition, TM3 cells transfected with vector alone and activated with cumate (PC) also showed slower rate of cell migration/proliferation when compared with 3P cells (Figure 28). As the control of PiggyBac vector only (P) was not included in this experiment, the difference in migration/proliferation rates between the three groups of cells can either be due to an inhibitory effect from Onc3 or due to the cumate treatment.

# TM3 cells



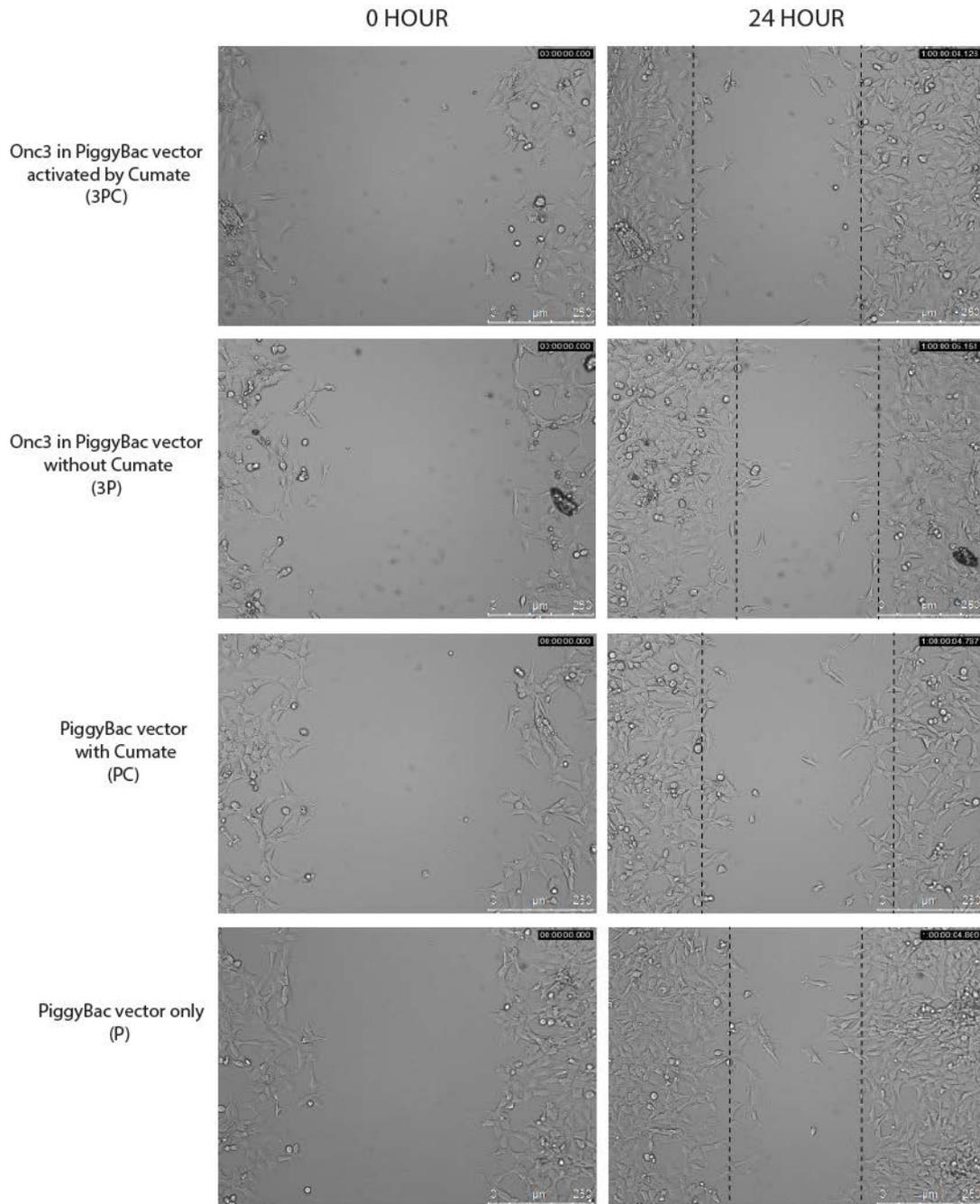
**Figure 27. Analysis of TM3 cell migration/proliferation by *in vitro* live cell imaging.** Images from 0 and 24 hours were used for comparison of cell migration/proliferation and morphology between 3PC, 3P and PC cells. Scale bar= 250 $\mu$ m.



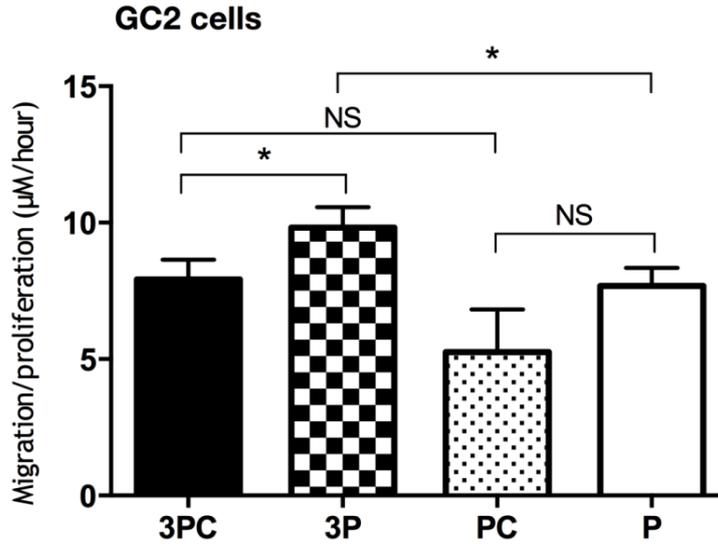
**Figure 28. Migration/proliferation of TM3 cells.** Areas between the cells were measured at 0 hours and 24 hours then migration/proliferation rate in  $\mu\text{M}$  per hour was calculated and compared between the 3 groups. 3PC cells migrate/proliferate at a significantly slower rate compared to 3P cells and migration/proliferation rate were similar between 3PC and PC cells.

Lastly the migration/proliferation rates for GC2 cells were analysed. GC2 cells migrate in a similar way as HEK293 cells. Interestingly, the morphology of GC2 cells activated with cumate (3PC and PC) appeared smaller and the dendrites of cells extended longer compared to cells without cumate (3P and P) (Figure 29). Overall, GC2 cells transfected with Onc3 expression construct (3PC and 3P) migrated/proliferated faster compared to cells transfected with PiggyBac vector alone (PC and P). 3PC cells migrated/proliferated significantly slower compared to 3P cells and at the same time, PC cells migrated/proliferated slower than P cells, although not statistically significant. If the effect of cumate was not considered, 3P cells migrate/proliferate significantly faster than P cells, which indicate that the cumate system may be leaky. Interestingly, no significant difference was found between 3PC and PC cells, which may be due to the large variation using PC cells (Figure 30).

# GC2 cells

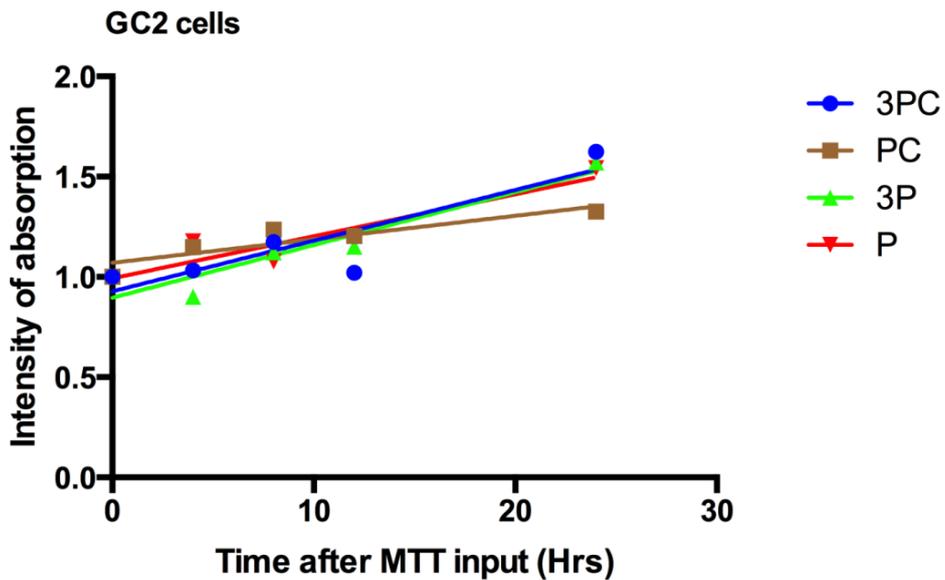


**Figure 29. Analysis of GC2 cell migration/proliferation by *in vitro* live cell imaging.** Images from 0 and 24 hours were used for comparison of cell migration/proliferation and morphology between 3PC, 3P PC and P cells. The dotted lines define the areas lacking cells. Scale bars= 250μm.



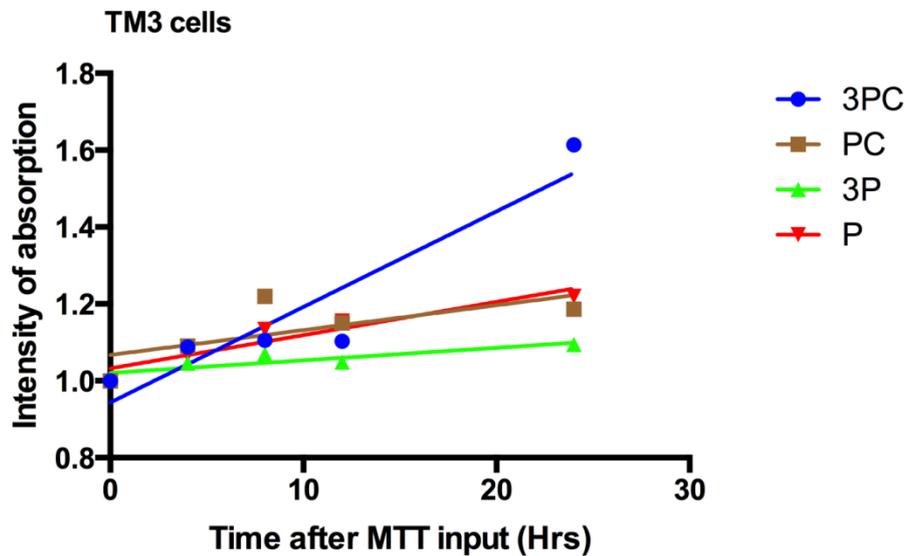
**Figure 30. Migration/proliferation of GC2 cells .** Areas between the cells were measured at 0 hours and 24 hours then migration/proliferation rate in  $\mu\text{M}$  per hour was calculated and compared between the 4 groups. 3PC cells migrate/proliferate at a significantly slower rate compared to 3P cells and 3P cells migrate/proliferate significantly faster than P cells. Overall expression of Onc3 increases the migration/proliferation rate of GC2 cells.

**Expression of Onc3 increases cell proliferation in TM3 cells while no effect on proliferation was observed in GC2 cells**



**Figure 31. MTT analysis of proliferation of GC2 cells.** Intensity of absorption of MTT was measured by a plate reader for 3PC, 3P, PC and P cells at 0, 4, 8,12, 24 hours after MTT input. Proliferation rate between 3PC, 3P and P cells were similar while PC cells proliferated at a slightly slower rate compared to other groups.

No differences were found between the proliferation rate of Onc3 expressed in GC2 cells compared to 3P and P cells. Interestingly, proliferation rate of PC cells was slightly slower compared to other cell groups (Figure 31).

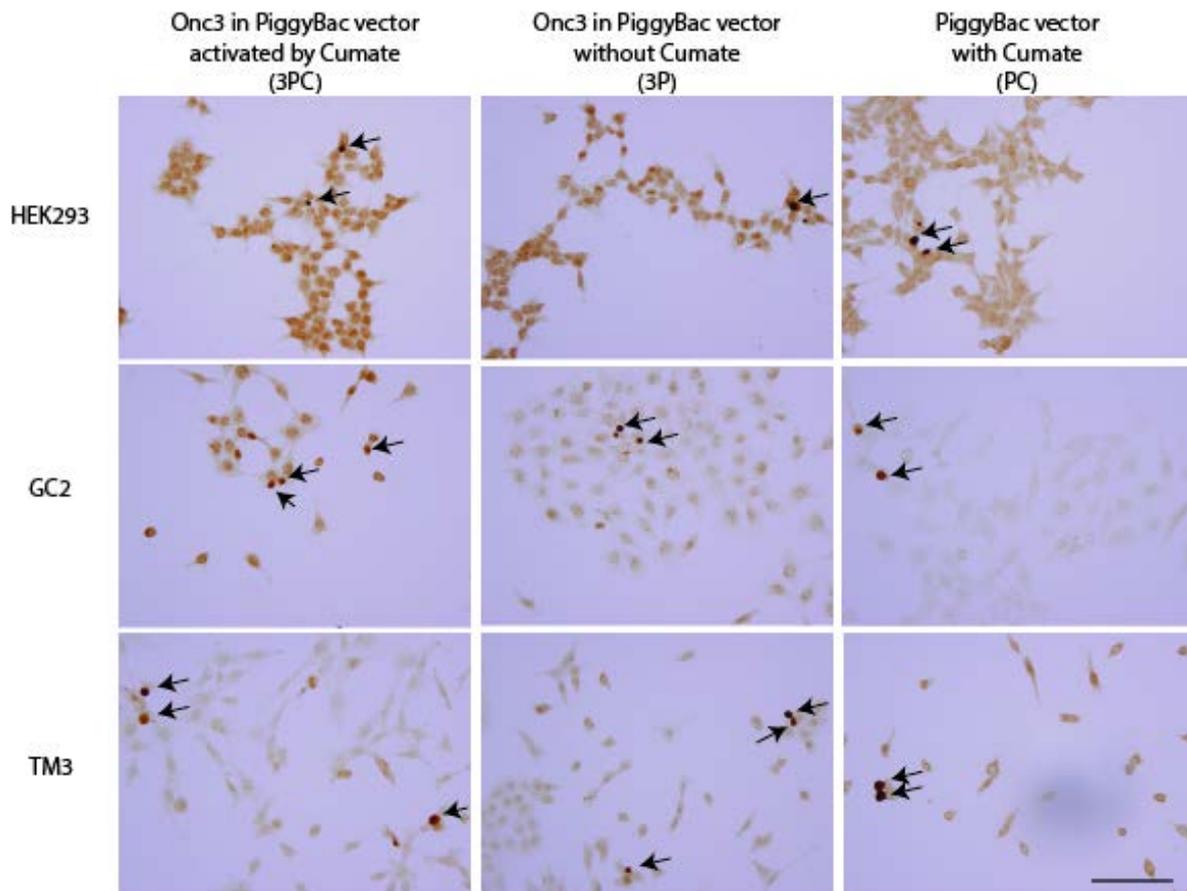


**Figure 32. MTT analysis of proliferation of TM3 cells.** Intensity of absorption of MTT was measured by a plate reader for 3PC, 3P, PC and P cells at 0, 4, 8, 12, 24 hours after MTT input. Proliferation rate between 3PC, 3P and P cells were similar while PC cells proliferated at a slightly slower rate compared to other groups.

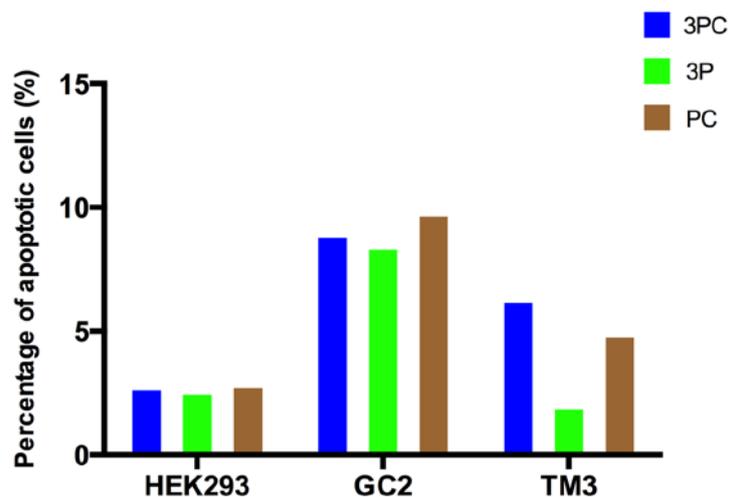
The MTT results suggested that the expression of Onc3 activated by cumate in TM3 cells causes higher proliferation rate compared to the three control cell groups. In addition, PC and P cells proliferate in a similar rate while 3P cells have the slowest proliferation rate among all cell groups (Figure 32).

### Expression of Onc3 does not affect apoptosis in HEK293, GC2 and TM3 cells

In order to determine if Onc3 expression caused an increase in cell death, TUNEL assay was performed on HEK293, TM3 and GC2 cells to observe changes in apoptosis in cells with or without Onc3 expression.



**Figure 33. The apoptotic effect of Onc3 expression on HEK293, GC2 and TM3 cells.** Cells marked by black arrow heads are deemed to be apoptotic cells from their morphology. Magnification: 40X. Scale bar= 100  $\mu$ m.



**Figure 34. The percentage of apoptotic cells in HEK293, GC2 and TM3 cells.** TUNEL assay was performed to observe apoptotic activity in HEK293, GC2 and TM3 cells. The number of live and apoptotic cells were counted for 3PC, 3P and PC cells, the percentage of apoptotic cells was calculated. Comparisons were made between test (3PC) and control (3P and PC) cells. Expression of Onc3 does not affect the apoptotic activity in HEK293 and GC2 cells, while in TM3 cells with cumate had increased apoptotic activity compared to cells without cumate.

From preliminary results it appears that Onc3 has no effect on the apoptotic activity of HEK293 and GC2 cells. For HEK293 cells, the apoptotic activity was slightly higher in 3PC and PC cells compared to 3P cells. Similar pattern was observed for GC2 cells. The slight increase in apoptotic cells may be due to a toxic effect of the cumate. 3PC and PC cells appear to have greater apoptotic activity in TM3 cells when compared to the 3P cells, this may be the result of cumate toxicity and it is likely that TM3 cells are more susceptible to the toxic effect of the cumate (Figure 36). It is difficult to conclude whether expression of Onc3 increases apoptosis of TM3 cells as PC cells showed similar percentage of apoptotic cells compared to 3PC cells.

# DISCUSSION

Long ncRNAs have a wide range of functions including the regulation of epigenetic chromatin modification and X-chromosome inactivation. However, the role of long ncRNA in early mouse gonad development is largely unexplored. Recently a microarray analysis found long ncRNAs are expressed sexually dimorphic in foetal gonads. The hypothesis for my research was that long ncRNAs are expressed during gonad differentiation and play important cellular roles such as proliferation, apoptosis and migration. In this project I analysed the expression pattern of a number of long ncRNA candidates throughout gonad development and cloned expression construct for one of the candidates, the ovary-enriched long ncRNA *AK020106* (*Onc3*), to perform further functional analysis in cell culture.

## I. EXPRESSION ANALYSIS

### **Sexually dimorphic expression of long ncRNAs suggest that they may have roles in gonad development**

In order for long ncRNAs to play a role during the crucial stages of sex differentiation they must be expressed during these stages, beginning at 11.5 dpc when *Sry* is expressed, until 14.5 dpc when gonad differentiation is well underway.

First, I characterised the expression of long ncRNAs in embryonic mouse gonads of both sexes. Through the use of ISH the expression of most long ncRNA candidates were shown to be sexual dimorphic in WT testis and ovary from 11.5 to 14.5 dpc, and that they display a range of different expression patterns within this time period. In addition, some long ncRNA expression was also seen in the liver, kidney and mesonephric tubules at specific time points in either XX or XY embryos, suggesting that these long ncRNAs also play roles within these tissues. Subsequently, the expression of long ncRNA candidates was quantitatively validated using qRT-PCR on 11.5 to 14.5 dpc WT mouse testis and ovary RNA.

## 1. Ovary-enriched long ncRNAs (Oncs)

All ovary-enriched long ncRNA candidates analysed by qRT-PCR showed higher expression levels in the ovary compared to the testis, which supports the microarray data that these long ncRNAs are ovary-enriched expressed.

It is known that long ncRNAs may directly or indirectly regulate genes located nearby [239]. Biological processes such as nuclear import, alternative splicing and epigenetic changes are mediated by long ncRNAs for gene regulation [240, 241]. In general, long ncRNAs can either regulate genes in *trans* through inactivation of transcription factor or inhibition of RNAPII binding to protein coding gene, or they can regulate in *cis*. One example is *XIST* in X chromosome inactivation as described earlier. Moreover, long ncRNAs can regulate locus-specifically in *trans* (*HOTAIR*) or in *cis* through recruitment of genetic modifiers to the promoters of protein-coding gene [239].

According to my qRT-PCR results, *AK044909* (OncL) and *AK036014* (OncB) expression peaks at 11.5 dpc in the ovary and is then gradually down-regulated over the next three days of development. The WNT4/RSPO1 signalling pathway is one of the major pathways that drive ovarian differentiation. As *Wnt4* expression becomes ovary-specific from 11.5 dpc [75]. It is possible that *Wnt4* is involved in the up-regulation OncL and OncB expression in the ovary. My results contradict the finding from the previous study [233], which showed that OncB expression was low in both ovary and testis at 11.5 dpc then up-regulated in the ovary thereafter. It is unclear what caused this discrepancy, however Chen et al. used the outbred CD1 mouse strain, whereas I quantified the expression of these long ncRNAs using the inbred C57BL/6 mouse strain. It is known that there are differences in gene expression within the developing gonads of these two mouse strains [242], however, more experiments would be necessary to show that this dramatic difference in OncB expression is due to the mouse background.

OncB may regulate the overlapping gene ionotropic glutamate receptor kainite 3 (*GRIK3*), which was found to associate with psychiatric illness such as schizophrenia. However, its expression or function within the developing ovary has not been investigated to date.

*AK019493* (OncP) qRT-PCR was performed in the previous study [233], which showed that the expression of OncP peaks at 11.5 dpc then gradually decreases until 13.5 dpc. According to the previous study, OncP appears to be co-expressed with *FOXL2* and has the highest expression within the medulla of the ovary, close to the mesonephros, while little or no expression was detected within the coelomic epithelium [233]. In my study, OncP

expression was low at 11.5 dpc, then peaked at 12.5 dpc followed by a reduction of expression by 14.5 dpc. The previous study has identified OncP to be expressed predominantly by ovarian somatic cells and may regulate the gene encoding pantothenate kinase 1 (*Pank1*), which is involved in the regulation of intracellular CoA concentration [243]. However *Pank1* expression is not sexually dimorphic and this gene appears to have no function during gonad development [244]. Therefore, the role of OncP in gonad development requires further functional analysis.

*AK182836* (OncF) expression peaked at 13.5 dpc before being down-regulated at 14.5 dpc. As described previously, the ovary appears quiescent until 13.5 dpc, when germ cells enter meiosis in an anterior-to-posterior wave [26, 245] and meiosis-related genes such as *Stra8* are up-regulated [25]. According to my ISH data, OncF is highly expressed in germ cells and hence it may participate in events leading to entry into meiosis or the up-regulation of genes such as *Stra8*. OncF overlaps with transmembrane 4 superfamily member tetraspanin 3 (*Tspan3*). A recent study suggested *Tspan3* regulates cell migration during *Xenopus laevis* embryonic development [246]. Since cell migration is a testis-specific event, it is possible that OncF down-regulates *Tspan3* in the developing ovary.

A previous study suggested that *AK015184* (OncQ) was expressed within XX but not XY germ cells from 11.5 to 13.5 dpc [233]. This was further extended in the current study, showing that OncQ expression further increases at 14.5 dpc. Based on this expression it is possible that OncQ is involved in germ cell entry into meiosis and germ cell cyst formation, which takes place from 13.5 to 14.5 dpc [13].

Similarly, the expression pattern of *AK020106* (Onc3) was lowest at 11.5 dpc then gradually peaked at 14.5 dpc in the XX gonads, hence the functions of Onc3 may be similar to OncQ. Onc3 is located upstream of *Fgf11*, a member of the FGF family. Although currently the function of FGF11 is unknown, FGF family members are known to be involved in a range of biological processes especially in cell proliferation [247]. Onc3 is also located downstream of *Tmem102*, like *Fgf11*, *Tmem102* has not been well characterised.

ISH analysis of *AK015136* (OncK) showed this long ncRNA was expressed in the somatic cells of 13.5 and 14.5 dpc XX gonads. Using the UCSC genome browser, *Tnc10* was found to overlap with several long ncRNAs including coiled-coil domain containing 41, opposite strand (*Ccdc41os1*).

## 2. Testis-enriched long ncRNAs (Tncs)

In contrast to the ovary, a number of morphological changes and cell differentiation occur in the testis from 11.5 dpc in the mouse [18]. Upon *Sry* expression, cell migration, testis cord formation, steroidogenesis and mitotic arrest of germ cells take place in the XY gonads between 11.5 to 13.5 dpc [18]. It is possible that testis-enriched long ncRNAs expressed at these developmental stages play roles in testis differentiation through regulation of gene expression.

According to qRT-PCR results obtained in the current study, *AK013819* (Tnc6) was expressed at high levels in the Sertoli and/or germ cells in the XY gonads from 11.5 to 13.5 dpc then expression decreased at 14.5 dpc. From the expression pattern of Tnc6, it can be predicted that Tnc6 plays a role in one or several of the testis-specific events that take place during 11.5 and 13.5 dpc. Tnc6 expression overlaps with the intron of neurotrophin 3 (*Ntf3*), NTF3 is a neurotrophin involved in control of progenitor cell fate in cortical postmitotic neurons [248]. A recent study using rat as a model system showed that SRY directly regulates *Ntf3* during sex determination, and NTF3 is essential for testis cord formation as it acts as a chemo attractant during cell migration [249]. In rats, *Sry* expression peaks at embryonic day 13 (E13) then at E14 *Ntf3* expression follows [249]. Therefore, given the possibility that both Tnc6 and *Ntf3* were expressed within Sertoli cells and both genes were expressed directly after *Sry* expression, Tnc6 may be a possible target of SRY in mice developing gonads during sex determination.

*AK034891* (Tnc3) was highly expressed in the testis at 11.5 dpc then expression declined afterwards. 11.5 dpc is also the peak expression stage for *Sry* [250]. The unique expression pattern of Tnc3 suggests it may be required in the down-regulation of *Sry* expression. As such, mutation of Tnc3 would lead to prolonged *Sry* expression in the testis. Likewise, it is also possible that *Sry* regulates Tnc3 expression in the developing testis. Located on chromosome 16, Tnc3 partially overlaps with roundabout homolog 2 (*Robo2*). *Robo2* was found to be important for brain, eye and kidney development in mice [251-253]. A recent study showed WNT4 signalling was inhibited by SLIT/ROBO2 signalling to promote mammary stem cell senescence [254]. Inhibition of WNT4/RSPO1 signalling pathway is essential for testis development. *Wnt4* expression becomes ovary-specific at 11.5 dpc [75], the same time point as when Tnc3 was highly expressed. It can be speculated that Tnc3 exerts inhibitory effect on the WNT signalling pathway by regulating expression of neighbouring gene *Robo2*.

For *AK013488* (Tnc7) and *AK005877* (Tnc9), sexually dimorphic expression was not observed at the developmental stages investigated in the current study. In the testis, Tnc7 showed higher expression at 12.5 dpc and Tnc9 at 11.5 dpc compared to ovary. According to the developmental stages they were expressed, it is possible that Tnc7 and Tnc9 are involved in the process of testis cord formation and cell migration respectively. Additionally, Tnc9 is located on chromosome X, downstream of DMRT-like family C1a (*Dmrtc1a*), which is also named as *Dmrt8.1*. Expression analysis on *Dmrt8.1* has found *Dmrt8.1* is expressed exclusively in the Sertoli cells of the adult but not embryonic testis. More experiments would be necessary to determine if Tnc9 is expressed in the adult testis and whether it can regulate *Dmrt8.1* expression in adulthood.

ISH analysis of *AK043086* (Tnc13) showed this long ncRNA was predominantly expressed in the coelomic epithelium. Tnc13 is conserved at the 3' end and is downstream of *Kif26b*, a known regulator of embryonic kidney mesenchyme in the mouse [255]. It is possible that Tnc13 regulates *Kif26b* expression during embryonic development. However, expression or function of *Kif26b* within the developing testis has not been investigated to date.

Through ISH, *AK045786* (Tnc10) expression was detected from 12.5 to 14.5 dpc XY gonads. Tnc10 is upstream of transmembrane channel-like gene family 5 (*Tmc5*), although the exact function of Tmc5 is currently unknown, it is possible that it can act as ion channels similar to TMC1 [256].

Most of the long ncRNAs analysed in this study are conserved between mouse, rat and human, therefore expression analysis in mouse models can be highly clinically relevant. In order to reach an overall conclusion on the expression pattern of long ncRNAs during gonad differentiation, more repeats of quantitative analysis needs to be done to identify the reason for discrepancy between the biological replicates as well as between the past and current study.

ISH and qRT-PCR results analysed in the current study validated microarray data of six Oncs conducted on embryonic mice gonads that was previously presented, and provided temporal expression data for one additional Onc and six Tncs, which supports the hypothesis that long ncRNAs are expressed during gonad differentiation.

## II. FUNCTIONAL ANALYSIS

Through the cloning of expression constructs, the effect of overexpression of long ncRNA *AK020106* (*Onc3*) was further examined.

Cell culture experiments are often the first step to functionally analyse long ncRNAs. For example, *Neat1* is a very stable, highly expressed long ncRNA in mammary gland and its expression is turned off at lactation [257]. Cell culture experiments with mammary epithelial cells showed knockdown of this long ncRNA using siRNAs increased the proliferation of cells, which correlates with the reduced cell proliferation rate at lactation [257]. This observation suggested that expression of *Neat1* needs to be repressed at lactation in order for the cells to stop proliferating.

Another example of functional analysis of long ncRNA in cell culture is the knockdown of long ncRNA *BANCR* in melanoma cells. The knockdown of *BANCR* induced by shRNA transfection reduced the invasiveness as shown by the scratch assay. Moreover, overexpression of this ncRNA increases the migration speed of the cells [258]. The results indicate *BANCR* is likely to control, at least partly, the invasiveness of melanoma cells.

Both examples are associated with metastatic behaviour of genes even though the RNAs are not protein-coding. These ncRNAs control the behaviour of cells, presumably by control patterns of gene expression through proteins. In the current study, the long ncRNA *Onc3* expression construct was transfected into HEK293, TM3 and GC2 cells to observe changes in cellular events such as proliferation, migration and apoptosis.

Migration/proliferation analysis of HEK293 cells showed that cells that express *Onc3* (3PC) had a significant reduction in cell migration and/or proliferation compared to the control cells. It is difficult to determine if it is due to reduced proliferation and/or migration because the closing of the gap in this assay could be due to either of the two options or a combination of both. Interestingly, HEK293 cells stably transfected with the *Onc3* expression construct but without cumate treatment (3P) showed increased migration/proliferation compared to 3PC cells, but reduced migration/proliferation compared to PC cells. It is possible that the cumate treatment had direct effect on HEK293 cells, or that the PiggyBac vector system is leaky, leading to expression of *Onc3* without cumate treatment.

Apoptosis analysis showed that *Onc3* did not induce apoptosis in HEK293 cells, as there was no difference in the number of apoptotic cells between test and control cells. Further replications of TUNEL experiments are required to confirm that there is no change in cell apoptosis.

Analysis of cell culture results for TM3 cells showed a different effect of Onc3 expression. In cell migration/proliferation analysis, both Onc3 expression (3PC) and empty PiggyBac vector with cumate treatment (PC) showed a reduced migration/proliferation rate compared to no cumate treatment (3P), suggesting that the cumate exerts a direct effects on TM3 cells. Interestingly, although the direct effect of cumate was also observed in HEK293 cells and GC2 cells, the effect was not as significant compared to TM3 cells, therefore it is possible TM3 cells are more susceptible to the effect of cumate.

Interestingly, MTT assay showed that TM3 cells that expressed Onc3 proliferated faster than the control cells, including cells stably transfected with the empty PiggyBac vector and treated with cumate. If the cumate treatment has a direct effect on closing of the gap in the “scratch” assay, then this could suggest that the migration rate of TM3 cells expressing Onc3 cells was significantly reduced.

The direct effect of cumate on TM3 cells was further analysed through TUNEL assay, where cells with cumate showed an increase in apoptosis compared to cells without cumate, therefore it is likely that the increase in proliferation as measured by the MTT assay, which detects live cells, is even greater than determined. Increase in cell proliferation is a key phenotype of cancer cells, therefore it is possible that Onc3 can function as an oncogene through promoting cell proliferation, in a similar way as *PCAT1* in prostate cancer [189].

Last but not least, migration/proliferation analysis showed that expression of Onc3 increased GC2 cell motility as both GC2 cells that expressed Onc3 (3PC) and GC2 cells transfected with the Onc3 expression construct but without cumate treatment (3P) migrated/proliferated faster compared to GC2 cells with empty PiggyBac vector with cumate treatment (PC) and empty PiggyBac vector without cumate treatment (P) cells. The direct effect of cumate may explain the significant difference in migration/proliferation between cells that expressed Onc3 (3PC) and without cumate treatment (3P). Comparison between GC2 cells transfected with Onc3 without cumate treatment (3P) and empty PiggyBac vector without cumate treatment (P) cells were made to eliminate the direct effect of cumate on GC2 cells, and Onc3 without cumate treatment (3P) cells were found to migrate/proliferate significantly than empty PiggyBac vector without cumate treatment (P) cells, further confirming the assumption that Onc3 enhances migration/proliferation of GC2 cells.

It was later found that the changes in cell movement, observed in the *in vitro* scratch assay, were mainly the effect changes in cell migration and not proliferation, no differences in cell proliferation was measured in the different cell types. Previous studies have identified several long ncRNAs that potentiates cell migration and invasion in cancer [259, 260]. *HOTAIR* is one of the long ncRNAs that promotes hepatocellular carcinoma cell migration

through inhibition of the RNA binding motif protein 38 (RBM38) [260]. The increase in migration of GC2 cells after the expression of Onc3 correlates with the previous finding in TM3 cells, which together suggested ectopic expression of this ovary-enriched long ncRNA promotes cell proliferation/migration, possibly through the regulation of a downstream gene such as *Fgf11*. As described earlier, FGFs are the key players in cell proliferation and migration [247]. During development, angiogenesis and cancer, FGFs induce both mitogenesis and cell motility through the same FGF receptor on the same cell, however the processes cannot occur simultaneously [247]. The length of FGF exposure is essential as it determines whether the cell will be proliferative or migratory: prolonged FGF exposure activates the MAP kinase pathway via ERKs, leading to a proliferative phenotype, while transient stimulation of FGF triggers the SRC pathway and the p38 MAP kinase pathway, which results in increased cell migration [261]. In the current study, it can be speculated that positive regulation of *Fgf11* expression by Onc3 activates the MAP kinase pathway via ERKs in TM3 cells, while in GC2 cells Onc3 regulation of *Fgf11* resulted in increased cell migration through the SRC pathway and the p38 MAP kinase pathway.

Similar to HEK293 cells, apoptotic activity did not vary for GC2 cells between the test and control cells. The TUNEL assay results was variable between individual experiments, therefore repetitions of TUNEL assay is necessary for statistical analysis of the results.

# FUTURE DIRECTIONS

Expression and functional analysis performed in the current study were only the early stages of the investigation of the role of long ncRNAs in gonad development. More experiments needs to be completed in order to address questions not yet answered in the current study.

## I. EXPRESSION ANALYSIS

First of all, in order to identify if *Tnc* expression is dependent on germ cells, germ cell depleted *W<sup>e</sup>* mouse strain gonads can be used to assess the expression of *Tncs*. Furthermore, ISH/immunohistochemistry (IHC) dual staining experiments will identify co-localisation of long ncRNAs with markers of different cell types such as *SRY*, *SOX9*, *E-cadherin* (germ cell marker) or *FOXL2* (somatic cell marker) expression, which will allow the identification of the cell lineage that express the long ncRNA..

It is also important to examine expression changes of long ncRNAs in mice with a known gonadal phenotype. Using both qRT-PCR and section ISH on mice with mutations in genes known to be important for gonad differentiation will provide further information on the role of long ncRNAs within the gene regulatory network underlying gonad development. For example, both *Tnc3* and *Sry* were highly expressed at 11.5 dpc in the XY gonad hence *Tnc3* may be directly regulated by *SRY* or *SOX9*. While in the XX gonads, both *OncF* and *Foxl2* were expressed from 12.5 dpc onwards. Based on these expression patterns, embryonic gonads from *Sox9* (for *Tncs*) and *Foxl2*, *Wnt4*, *Rspo1* (for *Oncs*) knockout mice can be used to determine whether the expression of long ncRNAs is affected. For instance, if *Tnc3* expression is reduced in the *Sox9* knockout mice, then it is possible that *Sox9* regulates the expression of *Tnc3* or at least it shows that the expression of *Tnc3* is downstream of *Sox9*.

In addition, the expression of genes located upstream, downstream or overlap with long ncRNA candidates such as *Robo2* and *Ntf3* needs to be analysed both qualitatively and quantitatively through ISH and qRT-PCR. It will be useful to determine if these genes are expressed in the embryonic gonad from 11.5 to 14.5 dpc, if they are expressed in a sexually dimorphic pattern, and if, the expression pattern correlates with their nearby long ncRNAs.

## II. FUNCTIONAL ANALYSIS

In order to analyse the functions of long ncRNA candidates other than *Onc3* (*AK020106*), it is important to clone more long ncRNA expression and knockdown constructs. To overexpress a long ncRNA, stable transfection of the expression construct into cell lines can be performed especially if the long ncRNA is not expressed in that cell line. On the other hand, if the long ncRNA is expressed in the cell line, a knockdown construct can be transfected instead. After transfection, the expressions of long ncRNAs can be examined and cellular events such as proliferation, migration and apoptosis will be monitored through MTT assay, *in vitro* scratch assay and TUNEL assay respectively.

In addition to testicular Leydig and germ cell lines, the ovarian cell line 12XG can be used in cell culture to analyse the functions of both ovary and testis-enriched long ncRNAs. While I was unable to analyse cellular events in 12XG cells due to time constraints and transfection difficulties, I believe ectopic expression of *Tncs* and overexpression of *Oncs* in 12XG cells are likely to result in changes in cell proliferation and migration similar to the phenotypes observed in the current study. However, it is also possible that the changes in proliferation and migration in *Onc3* expressed cells are specific to *Onc3* and each long ncRNAs have their specific effects on cellular events.

The expression and knockdown constructs can also be used in *ex vivo* gain- and loss-of-function experiments. At 11.5dpc, mouse genital ridges will be explanted and transfected with the expression or knockdown constructs and cultured on agar blocks [262]. This culture system can be used to mimic *in vivo* gonad development and allow the transfection of DNA expression constructs into the developing gonad to target gene expression.

Moreover, knockout studies can put be into context with genes known to be important for gonad differentiation and thereby provide further information on the role of long ncRNAs within the gene regulatory network. The effect of overexpression of long ncRNAs in mouse embryonic development can be analysed through generation of long ncRNA knockout transgenic mice with TALEN or CRISPR/CAS9 based technologies [263] and characterise the mutant phenotypes. In cases where knockout of the gene is embryonically lethal, generation of conditional tissue-specific knockout mouse models through the use of site-specific Cre-lox recombination system can be used [264].

Furthermore, transgenic mice with delayed or early expression of genes known to be important for gonad differentiation can be used to determine whether a regulatory relationship exist between the gene and long ncRNA. For example, transgenic mice with normal *Sox9* expression but early *Sry* expression [265] can be used to determine if *Sry*

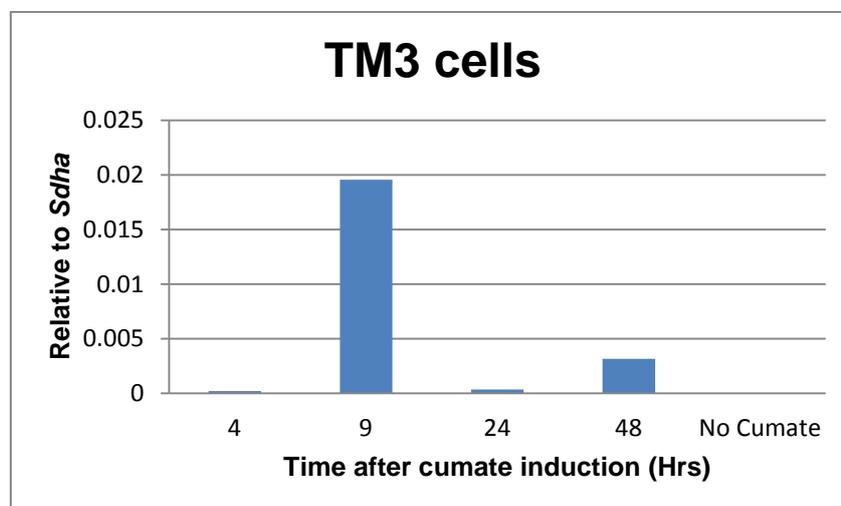
regulates Tnc3. If Tnc3 is regulated by Sry, the expression of Tnc3 should also be earlier than normal.

The gain- and loss-of-function experiments described above will hopefully provide evidence as to how long ncRNAs function in gonad development in the near future.

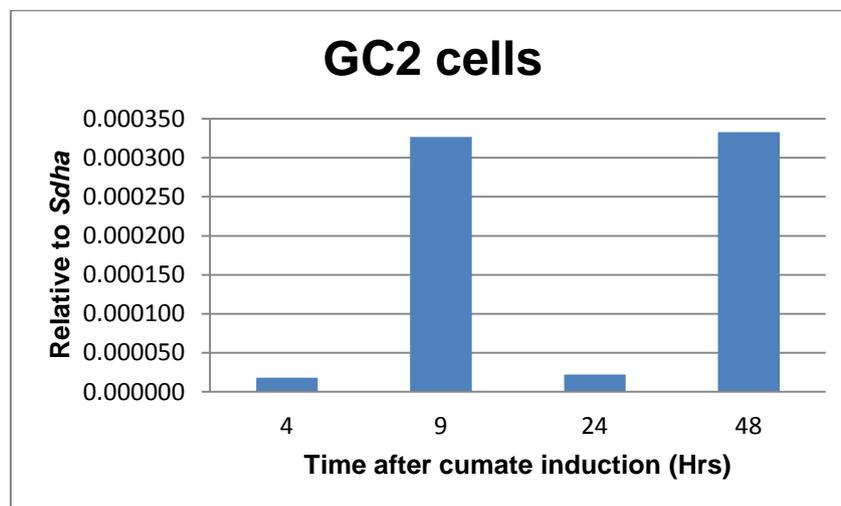
# CONCLUSION

Sexual reproduction is a highly regulated process used by most animals. It starts with the formation of gametes through meiosis, followed by fertilisation and sex determination. The correct development of the gonad is determined early in fetal life. Using microarrays techniques, previous studies identified a number of long ncRNAs, which displayed sexual dimorphic expression during mouse gonad development, but no function has been assigned. Although the analyses in this thesis do not directly provide information on the functionality and role of these long ncRNAs in sex development, long ncRNA candidates in this study was shown to be expressed sexually dimorphically in either the testis or the ovary at particular stages during early gonad development, and this expression pattern hints to an involvement of long ncRNAs in gonad differentiation and development. Ovary-enriched long ncRNA *AK020106* was found to up-regulate proliferation of a testis-derived cell line while it increases cell migration in a germ cell line. More experiments needs to be conducted in order to elucidate the function of long ncRNAs in embryonic gonad and whether they play a role in sex determination and gonad differentiation.

# SUPPLYMENTARY DATA



**Figure 35. qRT-PCR quantification of Onc3 expression in TM3 cells.** qRT-PCR analysis of expression of Onc3 4,9,24 and 48 hours after cumate induction in TM3 cells using gene specific primers for Onc3 relative to *Sdha*. Negative control of cells without cumate treatment was included. Statistically analysis was not performed as the experiment was only completed once.



**Figure 36. qRT-PCR quantification of Onc3 expression in GC2 cells.** qRT-PCR analysis of expression of Onc3 4,9,24 and 48 hours after cumate induction in GC2 cells using gene specific primers for Onc3 relative to *Sdha*. Statistically analysis was not performed as the experiment was only completed once.

# REFERENCES

1. A, J., *Recherches sur la differentiation sexuelle de l'embryon de lapin. Role des gonades foetales dans la differentiation sexuelle somatique.* Arch Anat Microsc Morphol Exp, 1947. **36**: p. 271-315.
2. Sinclair, A.H., et al., *A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif.* Nature, 1990. **346**(6281): p. 240-4.
3. Gubbay, J., et al., *A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes.* Nature, 1990. **346**(6281): p. 245-50.
4. Hacker, A., et al., *Expression of Sry, the mouse sex determining gene.* Development, 1995. **121**(6): p. 1603-14.
5. Hanley, N.A., et al., *SRY, SOX9, and DAX1 expression patterns during human sex determination and gonadal development.* Mech Dev, 2000. **91**(1-2): p. 403-7.
6. McCarrey, J.R. and U.K. Abbott, *Mechanisms of genetic sex determination, gonadal sex differentiation, and germ-cell development in animals.* 1979. p. 217-290.
7. Svingen, T. and P. Koopman, *Building the mammalian testis: origins, differentiation, and assembly of the component cell populations.* Genes Dev, 2013. **27**(22): p. 2409-26.
8. Saxen, L. and H. Sariola, *Early organogenesis of the kidney.* Pediatr Nephrol, 1987. **1**(3): p. 385-92.
9. Ginsburg, M., M.H. Snow, and A. McLaren, *Primordial germ cells in the mouse embryo during gastrulation.* Development, 1990. **110**(2): p. 521-8.
10. Fujimoto, T., Y. Miyayama, and M. Fuyuta, *The origin, migration and fine morphology of human primordial germ cells.* Anat Rec, 1977. **188**(3): p. 315-30.
11. Byskov, A.G., *Differentiation of mammalian embryonic gonad.* Physiol Rev, 1986. **66**(1): p. 71-117.
12. Cinalli, R.M., P. Rangan, and R. Lehmann, *Germ cells are forever.* Cell, 2008. **132**(4): p. 559-62.
13. Wilhelm, D., S. Palmer, and P. Koopman, *Sex determination and gonadal development in mammals.* Physiol Rev, 2007. **87**(1): p. 1-28.
14. Hiraoka, N., et al., *Matrix metalloproteinases regulate neovascularization by acting as pericellular fibrinolysins.* Cell, 1998. **95**(3): p. 365-77.
15. Bitgood, M.J., L. Shen, and A.P. McMahon, *Sertoli cell signaling by Desert hedgehog regulates the male germline.* Curr Biol, 1996. **6**(3): p. 298-304.
16. Barsoum, I. and H.H. Yao, *The road to maleness: from testis to Wolffian duct.* Trends Endocrinol Metab, 2006. **17**(6): p. 223-8.
17. Nef, S. and L.F. Parada, *Cryptorchidism in mice mutant for *Ins13*.* Nat Genet, 1999. **22**(3): p. 295-9.
18. Cool, J., T. DeFalco, and B. Capel, *Testis formation in the fetal mouse: dynamic and complex de novo tubulogenesis.* Wiley Interdiscip Rev Dev Biol, 2012. **1**(6): p. 847-59.
19. Bott, R.C., et al., *Vascular endothelial growth factor and kinase domain region receptor are involved in both seminiferous cord formation and vascular development during testis morphogenesis in the rat.* Biol Reprod, 2006. **75**(1): p. 56-67.
20. Buehr, M., S. Gu, and A. McLaren, *Mesonephric contribution to testis differentiation in the fetal mouse.* Development, 1993. **117**(1): p. 273-81.
21. Stacker, S.A. and M.G. Achen, *The vascular endothelial growth factor family: signalling for vascular development.* Growth Factors, 1999. **17**(1): p. 1-11.
22. Combes, A.N., et al., *Endothelial cell migration directs testis cord formation.* Dev Biol, 2009. **326**(1): p. 112-20.
23. Capel, B., et al., *Migration of mesonephric cells into the mammalian gonad depends on Sry.* Mech Dev, 1999. **84**(1-2): p. 127-31.

24. Uzumcu, M., et al., *Embryonic testis cord formation and mesonephric cell migration requires the phosphatidylinositol 3-kinase signaling pathway*. Biol Reprod, 2002. **67**(6): p. 1927-35.
25. Bowles, J., et al., *Retinoid signaling determines germ cell fate in mice*. Science, 2006. **312**(5773): p. 596-600.
26. Menke, D.B., J. Koubova, and D.C. Page, *Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave*. Dev Biol, 2003. **262**(2): p. 303-12.
27. Racki, W.J. and J.D. Richter, *CPEB controls oocyte growth and follicle development in the mouse*. Development, 2006. **133**(22): p. 4527-37.
28. Young, J.M. and A.S. McNeilly, *Theca: the forgotten cell of the ovarian follicle*. Reproduction, 2010. **140**(4): p. 489-504.
29. Ono, M. and V.R. Harley, *Disorders of sex development: new genes, new concepts*. Nat Rev Endocrinol, 2013. **9**(2): p. 79-91.
30. Martineau, J., et al., *Male-specific cell migration into the developing gonad*. Curr Biol, 1997. **7**(12): p. 958-68.
31. Brennan, J. and B. Capel, *One tissue, two fates: molecular genetic events that underlie testis versus ovary development*. Nat Rev Genet, 2004. **5**(7): p. 509-21.
32. Borum, K., *Oogenesis in the mouse. A study of the meiotic prophase*. Exp Cell Res, 1961. **24**: p. 495-507.
33. Richings, N.M., et al., *Growth and histology of ovarian follicles after cold storage in the tamar wallaby*. Reprod Fertil Dev, 2006. **18**(6): p. 677-88.
34. Jager, R.J., et al., *A human XY female with a frame shift mutation in the candidate testis-determining gene SRY*. Nature, 1990. **348**(6300): p. 452-4.
35. Veitia, R., et al., *Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46,XY female phenotype*. Hum Genet, 1997. **99**(5): p. 648-52.
36. Koopman, P., et al., *Male development of chromosomally female mice transgenic for Sry*. Nature, 1991. **351**(6322): p. 117-21.
37. Harley, V.R., et al., *DNA binding activity of recombinant SRY from normal males and XY females*. Science, 1992. **255**(5043): p. 453-6.
38. Ely, D., et al., *Review of the Y chromosome, Sry and hypertension*. Steroids, 2010. **75**(11): p. 747-53.
39. Tiersch, T.R., M.J. Mitchell, and S.S. Wachtel, *Studies on the phylogenetic conservation of the SRY gene*. Hum Genet, 1991. **87**(5): p. 571-3.
40. Hiramatsu, R., et al., *A critical time window of Sry action in gonadal sex determination in mice*. Development, 2009. **136**(1): p. 129-38.
41. Nef, S., et al., *Testis determination requires insulin receptor family function in mice*. Nature, 2003. **426**(6964): p. 291-5.
42. Tevosian, S.G., et al., *Gonadal differentiation, sex determination and normal Sry expression in mice require direct interaction between transcription partners GATA4 and FOG2*. Development, 2002. **129**(19): p. 4627-34.
43. Gierl, M.S., et al., *GADD45G functions in male sex determination by promoting p38 signaling and Sry expression*. Dev Cell, 2012. **23**(5): p. 1032-42.
44. Karl, J. and B. Capel, *Sertoli cells of the mouse testis originate from the coelomic epithelium*. Dev Biol, 1998. **203**(2): p. 323-33.
45. Knower, K.C., et al., *Failure of SOX9 regulation in 46XY disorders of sex development with SRY, SOX9 and SF1 mutations*. PLoS One, 2011. **6**(3): p. e17751.
46. Sekido, R. and R. Lovell-Badge, *Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer*. Nature, 2008. **453**(7197): p. 930-4.
47. Kim, Y., et al., *Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination*. PLoS Biol, 2006. **4**(6): p. e187.

48. Kim, Y., et al., *Fibroblast growth factor receptor 2 regulates proliferation and Sertoli differentiation during male sex determination*. Proc Natl Acad Sci U S A, 2007. **104**(42): p. 16558-63.
49. Wilhelm, D., et al., *SOX9 regulates prostaglandin D synthase gene transcription in vivo to ensure testis development*. J Biol Chem, 2007. **282**(14): p. 10553-60.
50. De Santa Barbara, P., et al., *Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Mullerian hormone gene*. Mol Cell Biol, 1998. **18**(11): p. 6653-65.
51. Schmahl, J., et al., *Fgf9 induces proliferation and nuclear localization of FGFR2 in Sertoli precursors during male sex determination*. Development, 2004. **131**(15): p. 3627-36.
52. Bagheri-Fam, S., et al., *Loss of Fgfr2 leads to partial XY sex reversal*. Dev Biol, 2008. **314**(1): p. 71-83.
53. Bagheri-Fam, S., et al., *Conserved regulatory modules in the Sox9 testis-specific enhancer predict roles for SOX, TCF/LEF, Forkhead, DMRT, and GATA proteins in vertebrate sex determination*. Int J Biochem Cell Biol, 2010. **42**(3): p. 472-7.
54. Houston, C.S., et al., *The campomelic syndrome: review, report of 17 cases, and follow-up on the currently 17-year-old boy first reported by Maroteaux et al in 1971*. Am J Med Genet, 1983. **15**(1): p. 3-28.
55. Foster, J.W., et al., *Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene*. Nature, 1994. **372**(6506): p. 525-30.
56. Barrionuevo, F., et al., *Homozygous inactivation of Sox9 causes complete XY sex reversal in mice*. Biol Reprod, 2006. **74**(1): p. 195-201.
57. Bi, W., et al., *Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization*. Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6698-703.
58. Vidal, V.P., et al., *Sox9 induces testis development in XX transgenic mice*. Nat Genet, 2001. **28**(3): p. 216-7.
59. Polanco, J.C., et al., *Sox10 gain-of-function causes XX sex reversal in mice: implications for human 22q-linked disorders of sex development*. Hum Mol Genet, 2010. **19**(3): p. 506-16.
60. Sutton, E., et al., *Identification of SOX3 as an XX male sex reversal gene in mice and humans*. J Clin Invest, 2011. **121**(1): p. 328-41.
61. Stevanovic, M., et al., *SOX3 is an X-linked gene related to SRY*. Hum Mol Genet, 1993. **2**(12): p. 2013-8.
62. Weiss, J., et al., *Sox3 is required for gonadal function, but not sex determination, in males and females*. Mol Cell Biol, 2003. **23**(22): p. 8084-91.
63. Stankiewicz, P., et al., *Duplication of Xq26.2-q27.1, including SOX3, in a mother and daughter with short stature and dyslalia*. Am J Med Genet A, 2005. **138**(1): p. 11-7.
64. Woods, K.S., et al., *Over- and underdosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism*. Am J Hum Genet, 2005. **76**(5): p. 833-49.
65. Sato, Y., et al., *The male-determining gene SRY is a hybrid of DGCR8 and SOX3, and is regulated by the transcription factor CP2*. Mol Cell Biochem, 2010. **337**(1-2): p. 267-75.
66. Lee, K., et al., *Congenital hydrocephalus and abnormal subcommissural organ development in Sox3 transgenic mice*. PLoS One, 2012. **7**(1): p. e29041.
67. Stolt, C.C. and M. Wegner, *SoxE function in vertebrate nervous system development*. Int J Biochem Cell Biol, 2010. **42**(3): p. 437-40.
68. Ottolenghi, C., et al., *Loss of Wnt4 and Foxl2 leads to female-to-male sex reversal extending to germ cells*. Hum Mol Genet, 2007. **16**(23): p. 2795-804.
69. Chassot, A.A., et al., *Activation of beta-catenin signaling by Rspo1 controls differentiation of the mammalian ovary*. Hum Mol Genet, 2008. **17**(9): p. 1264-77.
70. Uda, M., et al., *Foxl2 disruption causes mouse ovarian failure by pervasive blockage of follicle development*. Hum Mol Genet, 2004. **13**(11): p. 1171-81.

71. Tomizuka, K., et al., *R-spondin1 plays an essential role in ovarian development through positively regulating Wnt-4 signaling*. Hum Mol Genet, 2008. **17**(9): p. 1278-91.
72. Parma, P., et al., *R-spondin1 is essential in sex determination, skin differentiation and malignancy*. Nat Genet, 2006. **38**(11): p. 1304-9.
73. Bernard, P., et al., *Wnt signaling in ovarian development inhibits Sf1 activation of Sox9 via the Tesco enhancer*. Endocrinology, 2012. **153**(2): p. 901-12.
74. Slusarski, D.C., V.G. Corces, and R.T. Moon, *Interaction of Wnt and a Frizzled homologue triggers G-protein-linked phosphatidylinositol signalling*. Nature, 1997. **390**(6658): p. 410-3.
75. Kashimada, K., et al., *FOXL2 and BMP2 act cooperatively to regulate follistatin gene expression during ovarian development*. Endocrinology, 2011. **152**(1): p. 272-80.
76. Mandel, H., et al., *SERKAL syndrome: an autosomal-recessive disorder caused by a loss-of-function mutation in WNT4*. Am J Hum Genet, 2008. **82**(1): p. 39-47.
77. Yao, H.H., J. Aardema, and K. Holthusen, *Sexually dimorphic regulation of inhibin beta B in establishing gonadal vasculature in mice*. Biol Reprod, 2006. **74**(5): p. 978-83.
78. Ueno, N., et al., *Isolation and partial characterization of follistatin: a single-chain Mr 35,000 monomeric protein that inhibits the release of follicle-stimulating hormone*. Proc Natl Acad Sci U S A, 1987. **84**(23): p. 8282-6.
79. Yao, H.H., et al., *Follistatin operates downstream of Wnt4 in mammalian ovary organogenesis*. Dev Dyn, 2004. **230**(2): p. 210-5.
80. De Baere, E., et al., *Spectrum of FOXL2 gene mutations in blepharophimosis-ptosis-epicanthus inversus (BPES) families demonstrates a genotype--phenotype correlation*. Hum Mol Genet, 2001. **10**(15): p. 1591-600.
81. Cocquet, J., et al., *Evolution and expression of FOXL2*. J Med Genet, 2002. **39**(12): p. 916-21.
82. Crisponi, L., et al., *The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome*. Nat Genet, 2001. **27**(2): p. 159-66.
83. Boulanger, L., et al., *FOXL2 is a female sex-determining gene in the goat*. Curr Biol, 2014. **24**(4): p. 404-8.
84. Mork, L., et al., *Temporal differences in granulosa cell specification in the ovary reflect distinct follicle fates in mice*. Biol Reprod, 2012. **86**(2): p. 37.
85. Zheng, W., et al., *Two classes of ovarian primordial follicles exhibit distinct developmental dynamics and physiological functions*. Hum Mol Genet, 2014. **23**(4): p. 920-8.
86. Rastetter, R.H., et al., *Marker genes identify three somatic cell types in the fetal mouse ovary*. Dev Biol, 2014. **394**(2): p. 242-52.
87. Haegerbarth, A. and H. Clevers, *Wnt signaling, lgr5, and stem cells in the intestine and skin*. Am J Pathol, 2009. **174**(3): p. 715-21.
88. Moniot, B., et al., *Male specific expression suggests role of DMRT1 in human sex determination*. Mech Dev, 2000. **91**(1-2): p. 323-5.
89. Minkina, A., et al., *DMRT1 protects male gonadal cells from retinoid-dependent sexual transdifferentiation*. Dev Cell, 2014. **29**(5): p. 511-20.
90. Matson, C.K., et al., *DMRT1 prevents female reprogramming in the postnatal mammalian testis*. Nature, 2011. **476**(7358): p. 101-4.
91. Tannour-Louet, M., et al., *Identification of de novo copy number variants associated with human disorders of sexual development*. PLoS One, 2010. **5**(10): p. e15392.
92. Crawford, P.A., et al., *Nuclear receptor DAX-1 recruits nuclear receptor corepressor N-CoR to steroidogenic factor 1*. Mol Cell Biol, 1998. **18**(5): p. 2949-56.
93. Muscatelli, F., et al., *Mutations in the DAX-1 gene give rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism*. Nature, 1994. **372**(6507): p. 672-6.
94. Meeks, J.J., et al., *Dax1 regulates testis cord organization during gonadal differentiation*. Development, 2003. **130**(5): p. 1029-36.

95. Uhlenhaut, N.H., et al., *Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation*. Cell, 2009. **139**(6): p. 1130-42.
96. German, J.L., 3rd, A.G. Bearn, and G.J. Mc, *Chromosomal studies of three hermaphrodites*. Am J Med, 1962. **33**: p. 83-7.
97. Hughes, I.A., et al., *Consensus statement on management of intersex disorders*. J Pediatr Urol, 2006. **2**(3): p. 148-62.
98. MacLaughlin, D.T. and P.K. Donahoe, *Sex determination and differentiation*. N Engl J Med, 2004. **350**(4): p. 367-78.
99. Visootsak, J. and J.M. Graham, Jr., *Klinefelter syndrome and other sex chromosomal aneuploidies*. Orphanet J Rare Dis, 2006. **1**: p. 42.
100. Klee, P., et al., *A novel SRY mutation leads to asymmetric SOX9 activation and is responsible for mixed 46,XY gonadal dysgenesis*. Horm Res Paediatr, 2012. **78**(3): p. 188-92.
101. Mallet, D., et al., *Gonadal dysgenesis without adrenal insufficiency in a 46, XY patient heterozygous for the nonsense C16X mutation: a case of SF1 haploinsufficiency*. J Clin Endocrinol Metab, 2004. **89**(10): p. 4829-32.
102. Rizvi, A.A., *46, XX man with SRY gene translocation: cytogenetic characteristics, clinical features and management*. Am J Med Sci, 2008. **335**(4): p. 307-9.
103. Cox, J.J., et al., *A SOX9 duplication and familial 46,XX developmental testicular disorder*. N Engl J Med, 2011. **364**(1): p. 91-3.
104. de Lau, W.B., B. Snel, and H.C. Clevers, *The R-spondin protein family*. Genome Biol, 2012. **13**(3): p. 242.
105. Harrison, S.M., et al., *Screening and familial characterization of copy-number variations in NR5A1 in 46,XY disorders of sex development and premature ovarian failure*. Am J Med Genet A, 2013. **161A**(10): p. 2487-94.
106. Dahm, R., *Friedrich Miescher and the discovery of DNA*. Developmental Biology, 2005. **278**(2): p. 274-288.
107. Allen, F.W., *The Biochemistry of the Nucleic Acids, Purines, and Pyrimidines*. Annual Review of Biochemistry, 1941. **10**(1): p. 221-244.
108. Crick, F., *Central dogma of molecular biology*. Nature, 1970. **227**(5258): p. 561-3.
109. Kozak, M., *Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles*. Microbiological Reviews, 1983. **47**(1): p. 1-45.
110. Zaug, A.J. and T.R. Cech, *In vitro splicing of the ribosomal RNA precursor in nuclei of Tetrahymena*. Cell, 1980. **19**(2): p. 331-8.
111. Guerrier-Takada, C., et al., *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme*. Cell, 1983. **35**(3 Pt 2): p. 849-57.
112. Doudna, J.A. and T.R. Cech, *The chemical repertoire of natural ribozymes*. Nature, 2002. **418**(6894): p. 222-8.
113. Pederson, T., *Regulatory RNAs derived from transfer RNA? RNA*, 2010. **16**(10): p. 1865-9.
114. Katayama, S., et al., *Antisense transcription in the mammalian transcriptome*. Science, 2005. **309**(5740): p. 1564-6.
115. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
116. Venter, J.C., et al., *The sequence of the human genome*. Science, 2001. **291**(5507): p. 1304-51.
117. Carninci, P., et al., *The transcriptional landscape of the mammalian genome*. Science, 2005. **309**(5740): p. 1559-63.
118. Consortium, E.P., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project*. Nature, 2007. **447**(7146): p. 799-816.
119. Kapranov, P., A.T. Willingham, and T.R. Gingeras, *Genome-wide transcription and the implications for genomic organization*. Nat Rev Genet, 2007. **8**(6): p. 413-23.

120. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function*. Cell, 2004. **116**(2): p. 281-97.
121. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions*. Nat Rev Genet, 2009. **10**(3): p. 155-9.
122. Taft, R.J., M. Pheasant, and J.S. Mattick, *The relationship between non-protein-coding DNA and eukaryotic complexity*. Bioessays, 2007. **29**(3): p. 288-99.
123. Mattick, J.S., *Non-coding RNAs: the architects of eukaryotic complexity*. EMBO Rep, 2001. **2**(11): p. 986-91.
124. Shabalina, S.A. and N.A. Spiridonov, *The mammalian transcriptome and the function of non-coding DNA sequences*. Genome Biol, 2004. **5**(4): p. 105.
125. Carninci, P., *Molecular biology: The long and short of RNAs*. Nature, 2009. **457**(7232): p. 974-5.
126. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome*. Nature, 2012. **489**(7414): p. 57-74.
127. Rana, T.M., *Illuminating the silence: understanding the structure and function of small RNAs*. Nat Rev Mol Cell Biol, 2007. **8**(1): p. 23-36.
128. Bartel, D.P., *MicroRNAs: Target Recognition and Regulatory Functions*. Cell, 2009. **136**(2): p. 215-233.
129. Mendell, J.T., *MicroRNAs: critical regulators of development, cellular physiology and malignancy*. Cell Cycle, 2005. **4**(9): p. 1179-84.
130. Girard, A., et al., *A germline-specific class of small RNAs binds mammalian Piwi proteins*. Nature, 2006. **442**(7099): p. 199-202.
131. Aravin, A.A., et al., *Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline*. Curr Biol, 2001. **11**(13): p. 1017-27.
132. Lau, N.C., et al., *Characterization of the piRNA complex from rat testes*. Science, 2006. **313**(5785): p. 363-7.
133. Ghildiyal, M. and P.D. Zamore, *Small silencing RNAs: an expanding universe*. Nat Rev Genet, 2009. **10**(2): p. 94-108.
134. Okamura, K. and E.C. Lai, *Endogenous small interfering RNAs in animals*. Nat Rev Mol Cell Biol, 2008. **9**(9): p. 673-8.
135. Ni, J., A.L. Tien, and M.J. Fournier, *Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA*. Cell, 1997. **89**(4): p. 565-73.
136. King, T.H., et al., *Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center*. Mol Cell, 2003. **11**(2): p. 425-35.
137. Kiss-Laszlo, Z., et al., *Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs*. Cell, 1996. **85**(7): p. 1077-88.
138. Bernstein, E. and C.D. Allis, *RNA meets chromatin*. Genes Dev, 2005. **19**(14): p. 1635-55.
139. Umlauf, D., P. Fraser, and T. Nagano, *The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme*. Biol Chem, 2008. **389**(4): p. 323-31.
140. Jia, W., W. Chen, and J. Kang, *The functions of microRNAs and long non-coding RNAs in embryonic and induced pluripotent stem cells*. Genomics Proteomics Bioinformatics, 2013. **11**(5): p. 275-83.
141. Pagni, M., et al., *Role of microRNAs and long-non-coding RNAs in CD4(+) T-cell differentiation*. Immunol Rev, 2013. **253**(1): p. 82-96.
142. Cesana, M., et al., *A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA*. Cell, 2011. **147**(2): p. 358-69.
143. Prensner, J.R. and A.M. Chinnaiyan, *The emergence of lncRNAs in cancer biology*. Cancer Discov, 2011. **1**(5): p. 391-407.
144. Huttenhofer, A. and J. Vogel, *Experimental approaches to identify non-coding RNAs*. Nucleic Acids Res, 2006. **34**(2): p. 635-46.

145. Lee, J.T., *Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome*. *Genes Dev*, 2009. **23**(16): p. 1831-42.
146. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs*. *Cell*, 2007. **129**(7): p. 1311-23.
147. Goodrich, J.A. and J.F. Kugel, *Non-coding-RNA regulators of RNA polymerase II transcription*. *Nat Rev Mol Cell Biol*, 2006. **7**(8): p. 612-6.
148. Beltran, M., et al., *A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition*. *Genes Dev*, 2008. **22**(6): p. 756-69.
149. Yoon, J.H., K. Abdelmohsen, and M. Gorospe, *Posttranscriptional gene regulation by long noncoding RNA*. *J Mol Biol*, 2013. **425**(19): p. 3723-30.
150. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals*. *Nature*, 2009. **458**(7235): p. 223-7.
151. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis*. *Nature*, 2010. **464**(7291): p. 1071-6.
152. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response*. *Cell*, 2010. **142**(3): p. 409-19.
153. Sauvageau, M., et al., *Multiple knockout mouse models reveal lincRNAs are required for life and brain development*. *Elife*, 2013. **2**(0): p. e01749.
154. Scaruffi, P., *The transcribed-ultraconserved regions: a novel class of long noncoding RNAs involved in cancer susceptibility*. *ScientificWorldJournal*, 2011. **11**: p. 340-52.
155. Calin, G.A., et al., *Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas*. *Cancer Cell*, 2007. **12**(3): p. 215-29.
156. Feuerhahn, S., et al., *TERRA biogenesis, turnover and implications for function*. *FEBS Lett*, 2010. **584**(17): p. 3812-8.
157. Amaral, P.P. and J.S. Mattick, *Noncoding RNA in development*. *Mamm Genome*, 2008. **19**(7-8): p. 454-92.
158. Redon, S., P. Reichenbach, and J. Lingner, *The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase*. *Nucleic Acids Res*, 2010. **38**(17): p. 5797-806.
159. Cawley, S., et al., *Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs*. *Cell*, 2004. **116**(4): p. 499-509.
160. Clark, M.B., et al., *Genome-wide analysis of long noncoding RNA stability*. *Genome Res*, 2012. **22**(5): p. 885-98.
161. Ravasi, T., et al., *Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome*. *Genome Res*, 2006. **16**(1): p. 11-9.
162. Tupy, J.L., et al., *Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster*. *Proc Natl Acad Sci U S A*, 2005. **102**(15): p. 5495-500.
163. Martens, C., et al., *Protein-protein interactions and transcriptional antagonism between the subfamily of NGFI-B/Nur77 orphan nuclear receptors and glucocorticoid receptor*. *Mol Endocrinol*, 2005. **19**(4): p. 885-97.
164. Djebali, S., et al., *Evidence for transcript networks composed of chimeric RNAs in human cells*. *PLoS One*, 2012. **7**(1): p. e28213.
165. Wilusz, J.E., et al., *A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails*. *Genes Dev*, 2012. **26**(21): p. 2392-407.
166. Brown, J.D., S.E. Mitchell, and R.J. O'Neill, *Making a long story short: noncoding RNAs and chromosome change*. *Heredity (Edinb)*, 2012. **108**(1): p. 42-9.
167. Mitton-Fry, R.M., et al., *Poly(A) tail recognition by a viral RNA element through assembly of a triple helix*. *Science*, 2010. **330**(6008): p. 1244-7.
168. Tycowski, K.T., et al., *Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses*. *Cell Rep*, 2012. **2**(1): p. 26-32.

169. Louro, R., A.S. Smirnova, and S. Verjovski-Almeida, *Long intronic noncoding RNA transcription: expression noise or expression choice?* Genomics, 2009. **93**(4): p. 291-8.
170. Nakaya, H.I., et al., *Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription.* Genome Biol, 2007. **8**(3): p. R43.
171. Yin, Q.F., et al., *Long noncoding RNAs with snoRNA ends.* Mol Cell, 2012. **48**(2): p. 219-30.
172. Salzman, J., et al., *Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types.* PLoS One, 2012. **7**(2): p. e30733.
173. Jeck, W.R., et al., *Circular RNAs are abundant, conserved, and associated with ALU repeats.* RNA, 2013. **19**(2): p. 141-57.
174. Hansen, T.B., J. Kjems, and C.K. Damgaard, *Circular RNA and miR-7 in cancer.* Cancer Res, 2013. **73**(18): p. 5609-12.
175. Memczak, S., et al., *Circular RNAs are a large class of animal RNAs with regulatory potency.* Nature, 2013. **495**(7441): p. 333-8.
176. Zhang, Y., et al., *Circular intronic long noncoding RNAs.* Mol Cell, 2013. **51**(6): p. 792-806.
177. Gibb, E.A., C.J. Brown, and W.L. Lam, *The functional role of long non-coding RNA in human carcinomas.* Mol Cancer, 2011. **10**: p. 38.
178. Huarte, M. and J.L. Rinn, *Large non-coding RNAs: missing links in cancer?* Hum Mol Genet, 2010. **19**(R2): p. R152-61.
179. Faulkner, G.J., et al., *The regulated retrotransposon transcriptome of mammalian cells.* Nat Genet, 2009. **41**(5): p. 563-71.
180. Kim, T.K., et al., *Widespread transcription at neuronal activity-regulated enhancers.* Nature, 2010. **465**(7295): p. 182-7.
181. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.
182. Reis, E.M., et al., *Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer.* Oncogene, 2004. **23**(39): p. 6684-92.
183. Perez, D.S., et al., *Long, abundantly expressed non-coding transcripts are altered in cancer.* Hum Mol Genet, 2008. **17**(5): p. 642-55.
184. Ji, P., et al., *MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer.* Oncogene, 2003. **22**(39): p. 8031-41.
185. Tano, K., et al., *MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes.* FEBS Lett, 2010. **584**(22): p. 4575-80.
186. Schmidt, L.H., et al., *The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth.* J Thorac Oncol, 2011. **6**(12): p. 1984-92.
187. Gutschner, T., et al., *The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells.* Cancer Res, 2013. **73**(3): p. 1180-9.
188. Wu, L., et al., *Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins.* Biochemistry, 2013. **52**(52): p. 9519-27.
189. Prensner, J.R., et al., *Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression.* Nat Biotechnol, 2011. **29**(8): p. 742-9.
190. Prensner, J.R., et al., *PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer.* Cancer Res, 2014. **74**(6): p. 1651-60.
191. Kino, T., et al., *Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor.* Sci Signal, 2010. **3**(107): p. ra8.
192. Mourtada-Maarabouni, M., et al., *GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer.* Oncogene, 2009. **28**(2): p. 195-208.

193. Matouk, I.J., et al., *The H19 non-coding RNA is essential for human tumor growth*. PLoS One, 2007. **2**(9): p. e845.
194. Lottin, S., et al., *Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells*. Carcinogenesis, 2002. **23**(11): p. 1885-95.
195. Gabory, A., H. Jammes, and L. Dandolo, *The H19 locus: role of an imprinted non-coding RNA in growth and development*. Bioessays, 2010. **32**(6): p. 473-80.
196. Braconi, C., et al., *Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma*. Proc Natl Acad Sci U S A, 2011. **108**(2): p. 786-91.
197. Faber, J., et al., *HOXA9 is required for survival in human MLL-rearranged acute leukemias*. Blood, 2009. **113**(11): p. 2375-85.
198. Poliseno, L., et al., *A coding-independent function of gene and pseudogene mRNAs regulates tumour biology*. Nature, 2010. **465**(7301): p. 1033-8.
199. Massone, S., et al., *17A, a novel non-coding RNA, regulates GABA B alternative splicing and signaling in response to inflammatory stimuli and in Alzheimer disease*. Neurobiol Dis, 2011. **41**(2): p. 308-17.
200. Faghihi, M.A., et al., *Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase*. Nat Med, 2008. **14**(7): p. 723-30.
201. Modarresi, F., et al., *Knockdown of BACE1-AS Nonprotein-Coding Transcript Modulates Beta-Amyloid-Related Hippocampal Neurogenesis*. Int J Alzheimers Dis, 2011. **2011**: p. 929042.
202. Lanz, R.B., et al., *A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex*. Cell, 1999. **97**(1): p. 17-27.
203. Leygue, E., *Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer*. Nucl Recept Signal, 2007. **5**: p. e006.
204. Yao, H., et al., *Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA*. Genes Dev, 2010. **24**(22): p. 2543-55.
205. Uhler, J.P., C. Hertel, and J.Q. Svejstrup, *A role for noncoding transcription in activation of the yeast PHO5 gene*. Proc Natl Acad Sci U S A, 2007. **104**(19): p. 8011-6.
206. Keller, C., et al., *Noncoding RNAs prevent spreading of a repressive histone mark*. Nat Struct Mol Biol, 2013. **20**(8): p. 994-1000.
207. Hawkins, P.G. and K.V. Morris, *Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5*. Transcription, 2010. **1**(3): p. 165-175.
208. Yang, F., et al., *Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans*. Hepatology, 2011. **54**(5): p. 1679-89.
209. Kogo, R., et al., *Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers*. Cancer Res, 2011. **71**(20): p. 6320-6.
210. Li, L., et al., *Targeted disruption of Hotair leads to homeotic transformation and gene derepression*. Cell Rep, 2013. **5**(1): p. 3-12.
211. Tripathi, V., et al., *The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation*. Mol Cell, 2010. **39**(6): p. 925-38.
212. Liu, X., et al., *Long non-coding RNA gadd7 interacts with TDP-43 and regulates Cdk6 mRNA decay*. EMBO J, 2012. **31**(23): p. 4415-27.
213. Sone, M., et al., *The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons*. J Cell Sci, 2007. **120**(Pt 15): p. 2498-506.
214. Clemson, C.M., et al., *An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles*. Mol Cell, 2009. **33**(6): p. 717-26.
215. Guttman, M., et al., *lincRNAs act in the circuitry controlling pluripotency and differentiation*. Nature, 2011. **477**(7364): p. 295-300.

216. Wang, X., et al., *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription*. Nature, 2008. **454**(7200): p. 126-30.
217. Martianov, I., et al., *Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript*. Nature, 2007. **445**(7128): p. 666-70.
218. Feng, J., et al., *The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator*. Genes Dev, 2006. **20**(11): p. 1470-84.
219. Bond, A.M., et al., *Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry*. Nat Neurosci, 2009. **12**(8): p. 1020-7.
220. Camblong, J., et al., *Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae*. Genes Dev, 2009. **23**(13): p. 1534-45.
221. Koerner, M.V., et al., *The function of non-coding RNAs in genomic imprinting*. Development, 2009. **136**(11): p. 1771-83.
222. Sleutels, F., R. Zwart, and D.P. Barlow, *The non-coding Air RNA is required for silencing autosomal imprinted genes*. Nature, 2002. **415**(6873): p. 810-3.
223. Birger, Y., et al., *The imprinting box of the mouse Igf2r gene*. Nature, 1999. **397**(6714): p. 84-8.
224. Kanduri, C., *Kcnq1ot1: a chromatin regulatory RNA*. Semin Cell Dev Biol, 2011. **22**(4): p. 343-50.
225. Navarro, P., et al., *Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program*. Genes Dev, 2006. **20**(20): p. 2787-92.
226. Penny, G.D., et al., *Requirement for Xist in X chromosome inactivation*. Nature, 1996. **379**(6561): p. 131-7.
227. Tian, D., S. Sun, and J.T. Lee, *The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation*. Cell, 2010. **143**(3): p. 390-403.
228. Plath, K., et al., *Role of histone H3 lysine 27 methylation in X inactivation*. Science, 2003. **300**(5616): p. 131-5.
229. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome*. Science, 2008. **322**(5902): p. 750-6.
230. Shen, Y., et al., *X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations*. Proc Natl Acad Sci U S A, 2008. **105**(12): p. 4709-14.
231. Ogawa, Y., B.K. Sun, and J.T. Lee, *Intersection of the RNA interference and X-inactivation pathways*. Science, 2008. **320**(5881): p. 1336-41.
232. Castillo, A.F., et al., *Hormone-dependent expression of a steroidogenic acute regulatory protein natural antisense transcript in MA-10 mouse tumor Leydig cells*. PLoS One, 2011. **6**(8): p. e22822.
233. Chen, H., et al., *Identification of novel markers of mouse fetal ovary development*. PLoS One, 2012. **7**(7): p. e41683.
234. McFarlane, L., et al., *Novel PCR assay for determining the genetic sex of mice*. Sex Dev, 2013. **7**(4): p. 207-11.
235. Rolland, A.D., et al., *Uncovering gene regulatory networks during mouse fetal germ cell development*. Biol Reprod, 2011. **84**(4): p. 790-800.
236. Dagleish, R., *The human type I collagen mutation database*. Nucleic Acids Res, 1997. **25**(1): p. 181-7.
237. Gubern, C., et al., *Validation of housekeeping genes for quantitative real-time PCR in in-vivo and in-vitro models of cerebral ischaemia*. BMC Mol Biol, 2009. **10**: p. 57.
238. Dinger, M.E., et al., *NRED: a database of long noncoding RNA expression*. Nucleic Acids Res, 2009. **37**(Database issue): p. D122-6.
239. Kornienko, A.E., et al., *Gene regulation by the act of long non-coding RNA transcription*. BMC Biol, 2013. **11**: p. 59.

240. Chen, L.L. and G.G. Carmichael, *Decoding the function of nuclear long non-coding RNAs*. *Curr Opin Cell Biol*, 2010. **22**(3): p. 357-64.
241. Costa, F.F., *Non-coding RNAs, epigenetics and complexity*. *Gene*, 2008. **410**(1): p. 9-17.
242. Kraus, P., et al., *Mouse strain specific gene expression differences for illumina microarray expression profiling in embryos*. *BMC Res Notes*, 2012. **5**: p. 232.
243. Ramaswamy, G., et al., *PPARalpha controls the intracellular coenzyme A concentration via regulation of PANK1alpha gene expression*. *J Lipid Res*, 2004. **45**(1): p. 17-31.
244. Leonardi, R., et al., *Pantothenate kinase 1 is required to support the metabolic transition from the fed to the fasted state*. *PLoS One*, 2010. **5**(6): p. e11107.
245. Bullejos, M. and P. Koopman, *Germ cells enter meiosis in a rostro-caudal wave during development of the mouse ovary*. *Mol Reprod Dev*, 2004. **68**(4): p. 422-8.
246. Kashef, J., et al., *Expression of the tetraspanin family members Tspan3, Tspan4, Tspan5 and Tspan7 during Xenopus laevis embryonic development*. *Gene Expr Patterns*, 2013. **13**(1-2): p. 1-11.
247. Boilly, B., et al., *FGF signals for cell proliferation and migration through different pathways*. *Cytokine Growth Factor Rev*, 2000. **11**(4): p. 295-302.
248. Parthasarathy, S., et al., *Ntf3 acts downstream of Sip1 in cortical postmitotic neurons to control progenitor cell fate through feedback signaling*. *Development*, 2014. **141**(17): p. 3324-30.
249. Clement, T.M., et al., *SRY directly regulates the neurotrophin 3 promoter during male sex determination and testis development in rats*. *Biol Reprod*, 2011. **85**(2): p. 277-84.
250. Lovell-Badge, R. and A. Hacker, *The molecular genetics of Sry and its role in mammalian sex determination*. *Philos Trans R Soc Lond B Biol Sci*, 1995. **350**(1333): p. 205-14.
251. Grieshammer, U., et al., *SLIT2-mediated ROBO2 signaling restricts kidney induction to a single site*. *Dev Cell*, 2004. **6**(5): p. 709-17.
252. Yue, Y., et al., *Isolation and differential expression of two isoforms of the ROBO2/Robo2 axon guidance receptor gene in humans and mice*. *Genomics*, 2006. **88**(6): p. 772-8.
253. Thompson, H., et al., *Robo2 is required for Slit-mediated intraretinal axon guidance*. *Dev Biol*, 2009. **335**(2): p. 418-26.
254. Harburg, G., et al., *SLIT/ROBO2 signaling promotes mammary stem cell senescence by inhibiting Wnt signaling*. *Stem Cell Reports*, 2014. **3**(3): p. 385-93.
255. Uchiyama, Y., et al., *Kif26b, a kinesin family gene, regulates adhesion of the embryonic kidney mesenchyme*. *Proc Natl Acad Sci U S A*, 2010. **107**(20): p. 9240-5.
256. Chatzigeorgiou, M., et al., *tmc-1 encodes a sodium-sensitive channel required for salt chemosensation in C. elegans*. *Nature*, 2013. **494**(7435): p. 95-9.
257. Standaert, L., et al., *The long noncoding RNA Neat1 is required for mammary gland development and lactation*. *RNA*, 2014. **20**(12): p. 1844-9.
258. Li, R., et al., *Long non-coding RNA BANCR promotes proliferation in malignant melanoma by regulating MAPK pathway activation*. *PLoS One*, 2014. **9**(6): p. e100893.
259. Yang, S., et al., *Inhibition of SCAMP1 suppresses cell migration and invasion in human pancreatic and gallbladder cancer cells*. *Tumour Biol*, 2013. **34**(5): p. 2731-9.
260. Ding, C., et al., *Long non-coding RNA HOTAIR promotes cell migration and invasion via down-regulation of RNA binding motif protein 38 in hepatocellular carcinoma cells*. *Int J Mol Sci*, 2014. **15**(3): p. 4060-76.
261. LaVallee, T.M., et al., *Activation of the MAP kinase pathway by FGF-1 correlates with cell proliferation induction while activation of the Src pathway correlates with migration*. *J Cell Biol*, 1998. **141**(7): p. 1647-58.
262. Svingen, T., et al., *Ex vivo magnetofection: a novel strategy for the study of gene function in mouse organogenesis*. *Dev Dyn*, 2009. **238**(4): p. 956-64.
263. Gaj, T., C.A. Gersbach, and C.F. Barbas, 3rd, *ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering*. *Trends Biotechnol*, 2013. **31**(7): p. 397-405.

264. Sauer, B., *Inducible gene targeting in mice using the Cre/lox system*. *Methods*, 1998. **14**(4): p. 381-92.
265. Kumar, T.R., et al., *Transgenic mouse technology: principles and methods*. *Methods Mol Biol*, 2009. **590**: p. 335-62.

# APPENDICES

## APPENDIX 1: PRIMERS

### 1. qRT-PCR primers

<i>Shda</i> _F	5'-TG TTCAGTTCCACCCACA-3'
<i>Shda</i> _R	5'-TCTCCACGACACCCTTCTGT-3'
<i>Foxl2</i> _F	5'-GCTACCCAGCCCGAAGAC-3'
<i>Foxl2</i> _R	5'-GTGTTGTCCCGCCTCCCTTG-3'
<i>Amh</i> _F	5'-CGAGCTCTTGCTGAAGTTCCA-3'
<i>Amh</i> _R	5'-GTGTTGTCCCGCCTCCCTTG-3'
<i>AK019493 (OncP)</i> _F	5'-GCTTCCATTGCCACGTCTCAAT-3'
<i>AK019493 (OncP)</i> _R	5'-CCTGTGAAATCAAGCTGGAGGA-3'
<i>AK015184 (OncQ)</i> _F	5'-CAGGGGCAGCTATACTCTTAA-3'
<i>AK015184 (OncQ)</i> _R	5'-CTAGAGGTAGACTAGGGTCAA-3'
<i>AK182836 (OncF)</i> _F	5'-CCTGACTCTGTGCCTAACAGAA-3'
<i>AK182836 (OncF)</i> _R	5'-CTTTTGTGGGAGGGACAGAGAA-3'
<i>AK044909 (OncL)</i> _F	5'-CGTAGACTGGGCTGGTATCAA-3'
<i>AK044909 (OncL)</i> _R	5'-GTGTCTTCCACGTTTCTGTCCT-3'
<i>AK034891 (Tnc3)</i> _F	5'-CTGTGACTATTTTGTGGCCCAAT-3'
<i>AK034891 (Tnc3)</i> _R	5'-CTTGCAGCCTTTATTGCTGTCTT-3'
<i>AK013819 (Tnc6)</i> _F	5'-CGTATTGTAGCTCCAGGCTCAA-3'
<i>AK013819 (Tnc6)</i> _R	5'-CTGGCTGCTATGAACACCAGAA-3'
<i>AK013488 (Tnc7)</i> _F	5'-GCAGAGAGCTCTTGGACAGAAA-3'
<i>AK013488 (Tnc7)</i> _R	5'-CCACACATTTTCCCCTGCACAA-3'
<i>AK005877 (Tnc9)</i> _F	5'-GCTCACACCATGGAGAACTCAT-3'
<i>AK005877 (Tnc9)</i> _R	5'-GGCATCCTCTCCTGAAGAACTT-3'
<i>AK036014 (OncB)</i> _F	5'-CTTTATCCCTGCCTGGGCTGA-3'
<i>AK036014 (OncB)</i> _R	5'-GCATGTCCACAAAGGGCACTCT-3'
<i>AK020106 (Onc3)</i> _F	5'-TATAGCAGTCCCAGGGAGGA-3'
<i>AK020106 (Onc3)</i> _R	5'-CACCCACGTTGTCAACACTC-3'

### 2. ISH probe primers

<i>AK034891 (Tnc3)</i> _F	5'-CTTCTCCACAGGCAAGCATAGAA-3'
<i>AK034891 (Tnc3)</i> _R	5'-CAGCAGTATTGCCTACGAGTGAAA-3'
<i>AK013819 (Tnc6)</i> _F	5'-GTCACCTGTTCTGGGAATTCACAA-3'
<i>AK013819 (Tnc6)</i> _R	5'-GGCTGAGTCTTAGCACCTAA-3'
<i>AK013488 (Tnc7)</i> _F	5'-GATGGGAGAGCTTCTGGCAGAA-3'
<i>AK013488 (Tnc7)</i> _R	5'-GGAAGGCAAACAGGGTGGTTTGAA-3'
<i>AK005877 (Tnc9)</i> _F	5'-CAGCTGTCTGTCCCTGGAGCAA-3'

<b>AK005877 (Tnc9)_R</b>	5'-CATCACTGGACAAGATTCCCTTGA-3'
<b>AK045786 (Tnc10)_F</b>	5'-GCCAGCCAGCATGGCACAGAA-3'
<b>AK045786 (Tnc10)_R</b>	5'-GGAACATGTATGGAGATCCAAGAA-3'
<b>AK043086 (Tnc13)_F</b>	5'-GGTCTCCACTCAGCCCCTGTAGAA-3'
<b>AK043086 (Tnc13)_R</b>	5'-GCTCTCGTCACCAAAGTTGGAAGGA-3'
<b>AK020106 (Onc3)_F</b>	5'-GTGCTCTGAATTTGCACGAAC-3'
<b>AK020106 (Onc3)_R</b>	5'-TGGGTCTCCCACTTTCTTCA-3'
<b>AK036014 (OncB)_F</b>	5'-CAGTACTGTGGACCACGTCAAA-3'
<b>AK036014 (OncB)_R</b>	5'-GGAAAGGGATGAAGGCCTCAAA-3'
<b>AK182836 (OncF)_F</b>	5'-CACTCCACCCTGTTTCCAGT-3'
<b>AK182836 (OncF)_R</b>	5'-GTCGTGGGCTCTGAGAGAAG-3'
<b>AK015136 (OncK)_F</b>	5'-TGTGGTTCACGGTCACAGTT-3'
<b>AK015136 (OncK)_R</b>	5'-AGCTGCGGTTTGCCTTAATA-3'
<b>AK044909 (OncL)_F</b>	5'-GGTATGTTCTGGAGTGGAAACATAA-3'
<b>AK044909 (OncL)_R</b>	5'-CAGACTCAGTCCATCACAGCTAA-3'

### 3. PCR primers

<b>Sex_F</b>	5'-GATGATTTGAGTGGAAATGTGAGGTA-3'
<b>Sex_R</b>	5'-CTTATGTTTATAGGCATGCACCATGTA-3'
<b>M13_F</b>	5'-GTAAAACGACGGCCAG-3'
<b>M13_R</b>	5'-CAGGAAACAGCTATGAC-3'

### 4. Expression constructs primers

<b>AK020106 (Onc3)_F</b>	5'-TTGCTAGCGGCCTCTGTGCGCCGCCGCGCCT-3'
<b>AK020106 (Onc3)_R</b>	5'-TTGCGGCCGCATACTCTAAATGGAGTGGGTCTCCCACTT-3'

## APPENDIX 2: BUFFERS AND SOLUTIONS

### 1. Sex genotyping

<b>Tail lysis buffer</b>	50mM Tris (pH 8.0), 100mM EDTA, 100mM NaCl and 1% SDS
<b>Neutralisation solution</b>	1:10 1M Tris-HCl pH8 / Lysis buffer
<b>10X DNA loading buffer</b>	2.5g Ficoll 400 + 25mg bromophenol blue + 25mg xylene cyanol + water to 50 ml
<b>50X TAE</b>	2M Tris [Emresco] + 1M acetic acid [Sigma] + 50mM EDTA-HCl pH8 [Sigma]

### 2. ISH

<b>PBS</b>	1 phosphate buffer saline tablet [Astral scientific] in 500ml water (equivalent to 0.14M NaCl, 0.0027M KCl, 0.01M phosphate buffer pH7.4)
<b>Acetylation mix</b>	196ml water + 2.6ml triethanolamine [Sigma] + 350µl HCl [Sigma] + 500µl acetic anhydride [Sigma]
<b>50X Denhardts</b>	0.5g Ficoll (Type 400), 0.5g polyvinylpyrrolidone + 0.5g bovine serum albumin (Fraction V) + water to 50ml
<b>Prehybridisation mix</b>	50% formamide [Sigma] + 5X SSC, 5X Denhardts + 250µg/ml yeast RNA + 500µg/ml herring sperm DNA
<b>20X SSC</b>	175.3g NaCl [Merck Millipore] + 88.2 g sodium citrate [Merck Millipore] /1L of water
<b>NT Buffer</b>	50mM Tris-HCl pH7.5 + 150mM NaCl
<b>Blocking solution</b>	10% heat-inactivated horse serum (HS) in NT buffer
<b>Anti-DIG antibody</b>	1:2000 dilution of anti-DIG antibody in 1% HS/NT
<b>NTM</b>	100mM Tris pH9.5 + 100mM NaCl + 50mM MgCl <sub>2</sub> [Amresco]
<b>Colour Reaction</b>	3.5µl NBT (4-Nitro blue tetrazolium chloride) [Roche] + 3.5µl BCIP (5-bromo-4-chloro-3-indoyl-phosphate toluidine salt) [Roche] / 1ml of NTM

### 3. Plasmid preparation

<b>LB Medium</b>	10g bactotrypton + 10g NaCl + 5g yeast extract + water to 1L
<b>DNA loading buffer</b>	2.5g Ficoll 400 + 25mg bromphenol blue + 25mg xylene cyanol + water to 50ml