STUDIES ON THE LONG NON-CODING RNA, *MRHL* IN MOUSE EMBRYONIC STEM CELLS

A thesis submitted for partial fulfillment of degree of

MASTER OF SCIENCE (Biological Sciences)

as part of Integrated PhD program

by

Neha C V

Molecular Biology and Genetics Unit

Jawaharlal Nehru Centre for Advanced Scientific Research

Jakkur, Bangalore-560064

March 2017

DECLARATION

I hereby declare that the work described in this thesis entitled '**Studies on the long non coding RNA, *Mrhl* in mouse embryonic stem cells**' is the result of investigations carried out by myself under the guidance of **Prof. M.R.S Rao** at Chromatin Biology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore-560064, India. This work has not been submitted elsewhere for the award of any other degree.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the work described has been based on findings of other investigators. Any omission, which might have occurred by oversight or misjudgement, is regretted.

Neha C V

Jawaharlal Nehru Center for Advanced Scientific Research
Jakkur Campus, Jakkur Post
Bengaluru 560 064, INDIA

Office Tel: +91 (80) 2208 2754
Office Fax: +91 (80) 2208 2758
+91 (80) 2208 2468
Email:
mrsrao@jncasr.ac.in

Prof. MRS Rao
Honorary Professor and SERB Distinguished Fellow
Molecular Biology and Genetics Unit

CERTIFICATE

This is to certify that the work described in this thesis entitled **'Studies on the long non coding RNA, *Mrhl* in mouse embryonic stem cells'** is the result of investigations carried out by **Ms. C V Neha** at Chromatin Biology laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, under my supervision.

Prof. MRS Rao

ACKNOWLEDGEMENT

I would like to express my gratitude to my mentor, Prof. M R S Rao for giving me an opportunity to work in his lab on a challenging problem. His constant advice and guidance have helped me learn a lot. Discussions with him have lent me a lot of perspective on science in general.

My first foray into the world of scientific investigation was through the POBE programme. I would like to thank Dr. Ravi Manjithaya for his guidance and support throughout the POBE years and my rotation days. He has taught me a lot about experimental design and I am thankful to him for that.

I thank the entire MBGU faculty for their coursework and advice during the course of this project- Prof. Ranga Uday Kumar, Prof. Anuranjan Anand, Prof. Tapas Kundu, Prof. Maneesha Inamdar, Prof. Hemalatha Balaram, Prof. Namitha Surolia, Prof. Kaustuv Sanyal, Dr. Ravi Manjithaya and Dr. James Chelliah. I am also grateful to Prof. Amitabh Joshi and Late Prof. Vijay Kumar Sharma for their enlightening course on Statistics. My special thanks to Prof. Krushnamegh Kunte (NCBS) and Dr. Kshitij Acharya (IBAB) for their enjoyable coursework, which helped me broaden my horizon. I would like to thank Dr. Madhu Bajaj, a professor from Miranda House, for sparking my interest in Genetics and Molecular Biology. It was her teaching, which made a mark on me and steered me towards this field.

The people of the Chromatin Biology Lab have been very helpful and supportive through the course of this project. For this, I thank Debosree, Bhavana, Aditya, Shubhangini, Dhanur and Raktim. I am also thankful to Dr. Arun, Dr. Shalini, Vandana, Dr. Roshan, Monalisa, Divya, Dr. Subhendu, Utsa, Brian and Muniraju for their support.

The Autophagy Lab members have been extremely warm and encouraging and for this, I shall always be grateful. I have learnt a lot from Sunaina, Piyush, Shashank and Suresh and I thank them for this. Additionally, I would like to thank Somya, Dr. Aravinda, Dr. Malini, Gaurav, Lalitha, Veena, Dr. Sreedevi, Vishwa and Vidya for all the good times.

Arun, Mirmira Neelakshi and Srikant have been great friends who have stood by me throughout my journey here and I am thankful to them for all the time spent with

them. My stay at JNCASR has been enjoyable because of a number of people- Sarika, Jigyasa, Akash, Mayank, Sunil, Akshay and Ronak. Thanks for being there.

Rohini and Akshit have been patiently lending me an ear and supporting me through all my difficult times and for this, I shall always hold them dear. Finally, I wish to thank Ayushi, Aditi, Simi, Shweta, Jaya and Tanya for being great friends.

My mother has been my pillar of strength and a friend throughout my life. I can't thank her enough. Nikhil and Namrata have also been very caring and supportive. I am grateful to them. Finally, I would like to thank my grandpa for having inculcated a scientific temper in both my brother and me early in life.

ABBREVIATIONS:

RNA: Ribonucleic Acid

ncRNA : Non-coding RNA

ORF: Open reading frame

tRNA: Transfer RNA

rRNA: Ribosomal RNA

snRNA: Small nuclear RNA

si-RNA: Small interfering RNA

RISC: RNA induced silencing complex

miRNA: Micro RNA

piRNA: Piwi- interacting RNA

lnc RNA: Long non-coding RNA

UTR: Untranslated region

mRNA: Messenger RNA

GAS5: Growth Arrest Specific 5

PRC2: Polycomb Repressive Complex 2

LSD1: Lysine Specific Histone Demethylase 1

REST: Repressor element 1 silencing transcription factor

Co-REST: REST co-repressor 1

MLL1: Myeloid/lymphoid or mixed lineage leukemia 1

CNS: Central Nervous System

TUNA: *Tcl1* Upstream Neuron Associated lincRNA

DEANR1: Definitive Endoderm-Associated lncRNA1

Tug1: Taurine upregulated gene 1

APC: Adenomatous Polyposis Coli

GSK3β: Glycogen Synthase Kinase 3 β

TCF: T-cell factor (family of transcription factors)

LEF: Lymphoid enhancing factor

Mrhl: Meiotic recombination hotspot locus

PHKB: Phosphorylase Kinase Beta

kb: Kilo bases

GAPDH: Glyceraldehyde 3- Phosphate Dehydrogenase

DAPI: 4', 6-diamidino-2-phenylindole

SRA: Steroid Receptor RNA Activator

sh-RNA: short hairpin RNA

Oct4: Octamer binding transcription factor 4

SOX2: SRY (Sex determining region Y) box 2

qRT-PCR: Quantitative real time PCR

LIST OF FIGURES:

CONTENTS:

CHAPTER 1

INTRODUCTION:

Cellular identity, a feature unique to multicellular organisms, gets established over the course of the development of an organism and is achieved through gene expression. Though comprised of the same genome, heterogeneity in terms of different cell types and tissues is established through differential expression profiles. Multiple levels of regulation are exerted on the process of transcription and the events following it in order to achieve these distinct expression patterns [1].

Prior to sequencing of the human genome, proteins were believed to be the main functional players in cellular processes and hence, the majority of the genome would code for them. However, it was the annotation of the genome that surprised many with the observation that a mere two percent of the genome was protein-coding and the rest of the non-coding genome was "junk" [2, 3]. With the advent of deep sequencing, the characterization of the transcriptome became possible and this "junk" DNA was seen to produce a large number of transcripts ubiquitously [4, 5]. Lack of sequence conservation between most of these transcripts across organisms and an absence of apparent function raised questions as to whether these were transcriptional "noise" [2] or some sort of baggage that organisms carried over the course of evolution [6]. This was when the idea of non-coding RNA as regulatory molecules emerged [6].

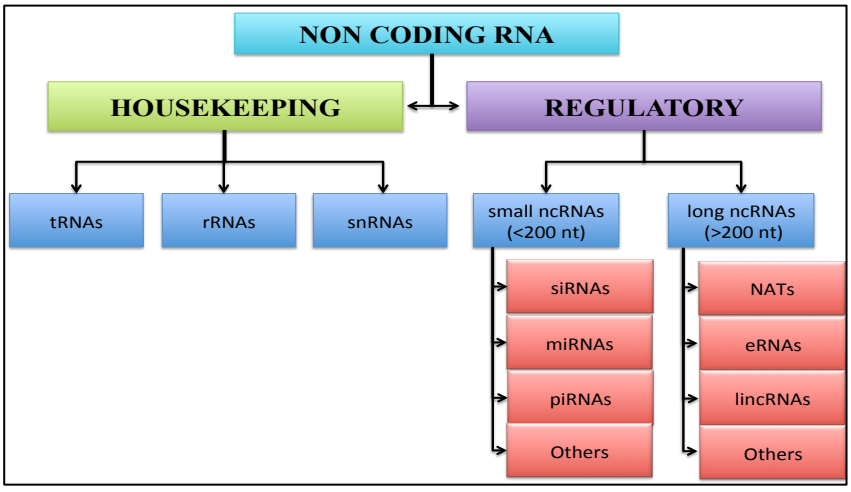


Figure 1.1: Classification of the non-coding transcriptome

NON- CODING RNA:

Conventionally, non-coding RNAs are defined as the transcripts that do not possess an open reading frame (ORF) [7] or have a low coding potential [8]. Non-coding transcripts can be classified as either housekeeping (e.g. tRNA, rRNA, snRNA, etc.) or regulatory in function (e.g., small interfering RNA, piRNA, microRNA, long non-coding RNA and so on). Housekeeping RNAs are critical components of several cellular processes like translation (tRNAs and rRNAs), splicing (snRNAs), etc. [9]. Regulatory RNAs have been described in brief below.

Regulatory non-coding RNAs:

Non-coding RNAs as major players in exerting regulation on several cellular processes represent a class that has gained a lot of attention over the past two decades [6]. Several classes of these regulatory ncRNAs exist, which are broadly classified into two groups (based on their length): small ncRNA (<200 nt) and long ncRNA (>200 nt). Small ncRNAs include small interfering RNAs, microRNAs, piwi-interacting RNAs etc. [6], which are described below:

Small non-coding RNAs

Small Interfering RNAs:

These are small 21-24 nt single-stranded transcripts that are produced from ds-RNA molecules by the action of Dicer complex. Originally discovered in the nematode *C. elegans*, siRNAs play a major role in regulating gene expression. These siRNAs are loaded onto Argonaute of the RISC (RNA-induced silencing complex) and bind to their complementary mRNA sequence, thereby causing its degradation (upon full complementarity) or inhibit translation (if partial complementarity exists) [6].

MicroRNAs:

miRNAs are produced from pre-miRNAs (either from introns of other mRNAs or as independent transcripts) that form short hairpin loops. These loops are acted upon by the DGCR8 of Drosha/ Pasha complex to generate ds-RNA species that get

exported outside the nucleus by a shuttling protein Exportin 5. In the cytoplasm, Dicer acts on the ds-RNAs to generate single stranded transcripts (miRNAs), which get loaded onto RISC and target cognate mRNAs in a manner similar to siRNAs [6].

Piwi -interacting RNAs:

piRNAs (26-30 nt) are germ cell specific regulatory non coding RNAs that interact with a special class of proteins called Piwi proteins. They are responsible for regulation/silencing of retrotransposons and other mobile genetic elements in these cells by epigenetic and post translational modulation and for maintaining the heterochromatin state of certain regions within the genome in the germ cells [6].

Long non-coding RNAs

Long non-coding (lnc) RNAs are arbitrarily defined as the non-protein coding transcripts of length >200 nucleotides [7, 10]. They could be further classified as sense or antisense (based on the orientation of their promoters) or intronic, intergenic, exonic [10], promoter associated or 3'UTR derived (based on the genomic location they are produced from). Based on their mode of action, they could be cis-acting or trans-acting [11].

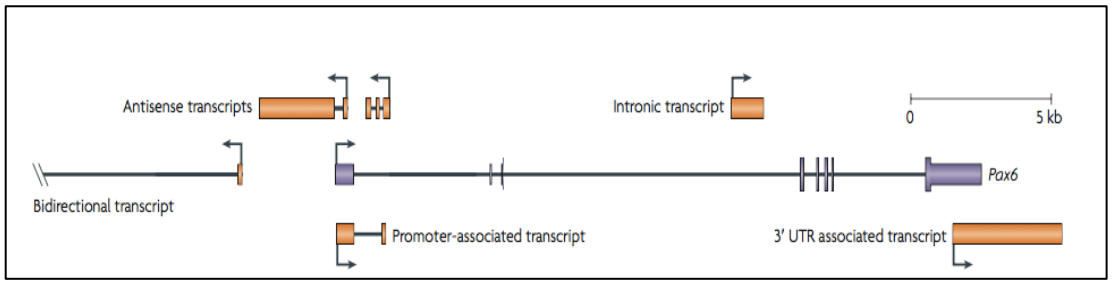


Figure 1.2: Classification of long non-coding RNAs based on their genomic architecture *(Adapted from Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. Nature Reviews Genetics, 10(3), 155-159).*

LncRNAs: Similarities and differences as compared to mRNA

Although lncRNAs appear to differ in an obvious manner from mRNAs in terms of their protein coding potential [8], they share a number of similarities with mRNAs. Most of them are RNA polymerase II transcribed, possess a 5' methyl guanosine cap and a 3' poly-A tail [12]. Additionally, promoters of lncRNA genes possess histone marks quite similar to those of protein coding genes- H3K4me3 (active gene mark at transcription start site) and H3K27ac (activation mark in the enhancer region) [13, 14]. However, not all is similar between these two classes of transcripts. While mRNAs show high sequence conservation across species, lncRNAs lack an overall sequence conservation. But there may exist highly conserved stretches within the sequence. Conserved secondary structure, function and/or syntenic genomic location of lncRNA genes may also be found across different species suggesting that these components may have been retained over the course of evolution due their functional significance [8]. Messenger RNAs are localised predominantly to the cytoplasm, where they associate with the ribosomes to generate proteins, lncRNAs show nuclear and/or cytoplasmic localisation in a context dependent manner [14]. LncRNAs exhibit exosome mediated degradation in the nucleus in addition to the usual degradation mechanism that protein coding transcripts undergo in the cytoplasm by decapping and subsequent action of exonucleases [14]. Thus, lncRNAs are highly regulated transcripts both spatially and temporally due to the fact that they themselves modulate a number of processes within the cell [14].

Diverse Mechanisms of Action of lncRNAs

Since their discovery, many long non-coding RNA have been shown to play important roles in enforcing regulation in different cellular processes like genetic imprinting, differentiation and development, maintenance of pluripotency, reprogramming of cells, etc. All of this is achieved by versatile mechanisms including recruitment of chromatin remodelers, silencing of mRNA transcripts by RNA interference, serving as micro RNA sponges and RNA decoys, assisting in alternative splicing, masking transcription factor binding sites, inhibition of translation, altering the localization of proteins and many others [2, 3, 12, 15].These

functions, though not always pertaining to only one category, can mainly be classified as:

Signal: A number of lncRNAs show a restricted tissue specific and time specific expression pattern, and this control over their transcription is modulated by various internal and external inputs of the cell. This makes them ideal to serve as signals for the transcription of certain genes, in response to these cues [16]. A well known example of this is the *Xist* RNA, which facilitates dosage compensation in females by inactivation of the extra copy of X chromosome. This time and developmental stage specific expression of the *Xist* RNA causes the recruitment of the repressive complexes to the chromosome to be silenced and resulting in a chromosome-wide repression of gene expression [2, 16].

Decoy: The lncRNAs of this category serve as decoys that prevent their targets (proteins or miRNAs) from binding to their cognate sequences and hence, regulate gene expression [16]. *GAS5* (Growth Arrest Specific 5) is a lncRNA that modulates activity of the glucocorticoid receptor. It has an RNA stem loop which mimics the DNA motif present in the glucocorticoid responsive gene promoters, enabling it to titrate away the glucocorticoid receptor and prevent it from binding to its recognition sequences in the promoters [13, 15, 16].

Guide: RNA, being a versatile biomolecule, has the ability to interact with two other major biomolecules- DNA and protein. Utilising the complementary base pairing ability, RNA can pair up with DNA sequences. It can also fold upon itself to form secondary structures and interact with proteins. These properties make it an ideal candidate to guide protein complexes to certain locations on the chromatin, where RNA provides additional specificity to the ribonucleoprotein complex [9, 15, 16]. An example of this is *Xis*t RNA as explained earlier.

Scaffold: Many lncRNAs serve as platforms, where different protein complexes are brought together. Sometimes, the complexes involved might even be antagonistic to each other. An example in this category is *Hotair* that binds to PRC2, a repressive complex which silences gene expression by depositing H3K27me3 marks. At its 3' end, it also binds to LSD1 of the CoREST/REST complex that demethylates H3K4 marks to antagonise gene activation [2, 16].

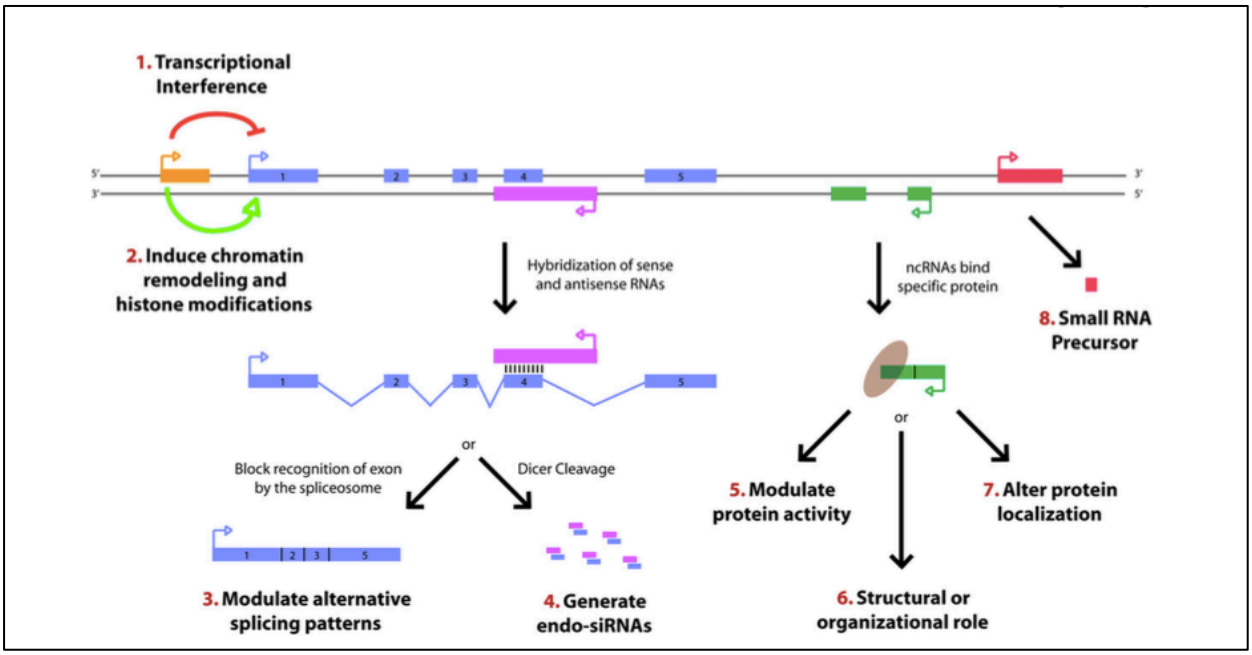


Figure 1.3: Versatile mechanisms of regulation exerted by long non-coding RNAs in various cellular processes *(Adapted from Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes & development, 23(13), 1494-1504).*

LncRNAs in Development and Differentiation:

Mammalian development is a highly complex process regulated at several levels. It has been noted that as developmental complexity increases, the number of lncRNAs also increases, indicating that these long non-coding transcripts exert tight control over the different stages of development [17]. Increasing reports over the past few years have implicated lncRNAs in the maintenance of pluripotency, regulation of differentiation, lineage specification, cell fate decisions, genomic imprinting etc. [1, 2, 18].

Many long non-coding RNA have been implicated in orchestrating different epigenetic modifications in diverse systems, including embryonic stem cells [19, 20]. Embryonic stem cells represent a unique model system because their genome undergoes extensive remodeling based on the cues that they receive [21]. Post each

round of division, a half of the daughter cells undergo differentiation while the other half retains their pluripotent nature, and hence, their genome bears a unique epigenetic signature consisting of bivalent chromatin marks, i.e., the presence of both activating and repressive marks at the same genomic location. This helps them maintain the plasticity of the genome for both pluripotency and differentiation. Also, due to the bivalent marks, cells are poised to respond to external cues without any lag [22, 23, 24].

Non-coding RNA profiling in mouse embryonic stem cells using microarray has revealed two novel long non-coding RNA, *Evx1as* and *Hoxb5/6as,* that are associated with *MLL1* (a histone methyltransferase) and trimethylated H3K4 histones, indicating their potential role in epigenetic control of lineage specification and pluripotency at the homeotic loci [1]. Genome wide screening has revealed that some long non-coding RNAs are central to the regulation of Oct4-Nanog transcriptional network in mESCs and are, in turn, transcriptionally regulated by them, establishing a regulatory feedback loop [25].

Other well-known example of a lncRNA that is crucial for development is *Hotair*. It is transcribed from the HoxC cluster and exerts its function in trans on the HoxD cluster. Hox genes are homeobox containing transcription factors that are involved in the formation of the general body plan and patterning. *Hotair* binds to two chromatin modifying complexes- PRC2 at 5'end and LSD1 at 3'end and epigenetically regulates the HoxD genes [26, 27].

A recent report [28] identified that *TUNA/ Megamind*, an evolutionarily conserved lncRNA, is essential for the maintenance of pluripotency of mouse embryonic stem cells and is highly expressed in the CNS of both mouse and zebrafish, playing a key role in the commitment of stem cells to the neural lineage. This example illustrates how the same lncRNA can have a context dependent role along the different stages of development.

Additionally, a plethora of long non-coding transcripts have been implicated in regulating differentiation of different lineages- *Braveheart* (cardiovascular development) [29], *Fendr* (heart and body wall formation) [30], *DEANR1* (endoderm

differentiation) [31], *Tug1* for retina development [32] etc. indicating the importance of lncRNAs as a class in orchestrating different stages of development.

Table 1 | **lncRNA manipulation and resulting phenotypes in model animal systems**

lncRNA	Process	Site of action	Loss-of-function methods	Phenotype	Refs
Mouse					
Xist	Dosage compensation	Nucleus	Gene disruption in embryo	Embryonic lethality	57
			Conditional disruption in haematopoietic stem cells	Aberrant haematopoiesis and blood cell cancer	63
Tsix	Dosage compensation	Nucleus	Embryonic gene inactivation by either promoter deletions or premature termination	Embryonic lethality	60,61
Kcnq1ot1	Genomic imprinting	Nucleus	Embryonic gene inactivation by either promoter deletions or premature termination	Growth defects	64,66
Airn	Genomic imprinting	Nucleus	Embryonic gene inactivation by either promoter deletions or premature termination	Growth defects	29,65
Fendrr	Gene expression regulation in mesoderm	Nucleus	Gene disruption in embryo	Embryonic lethality	95
			60% reduction in embryo by RNA interference	Normal development	95
Hotair	Hox gene regulation	Nucleus	Gene disruption in embryo	Defects in skeletal system development	72
Dlx1os	Homeodomain transcription factor regulation in developing forebrain	Nucleus	Embryonic gene inactivation by premature termination	Morphologically normal with mild skull and neurological defects	92
Dlx6os1	Homeodomain transcription factor regulation in developing forebrain	Nucleus	Embryonic gene inactivation by premature termination	Morphologically normal with altered GABAergic interneuron development	90
Malat1	Tumorigenesis	Nucleus	Gene disruption in embryo	Normal development	82
Miat	Retina development	Nucleus	Knockdown and overexpression in neonatal retina	Defects in specification of retina cell types	126
Six3os1	Retina development	Nucleus	Knockdown and overexpression in neonatal retina	Defects in specification of retina cell types	127
Tug1	Retina development	Nucleus	Knockdown in neonatal retina	Defects in differentiation of photoreceptor progenitor cells	128
Vax2os	Retina development	Nucleus	Overexpression in neonatal retina	Defects in differentiation of photoreceptor progenitor cells	129
Zebrafish					
Cyrano	Embryogenesis	Not analysed	Knockdown and functional inactivation in embryo by morpholino oligonucleotides	Developmental defects	110
Megamind	Embryogenesis	Not analysed	Knockdown and functional inactivation in embryo by morpholino oligonucleotides	Defects in brain morphogenesis and in eye development	110
Chicken					
HOTTIP	HOXA regulation	Nucleus	Knockdown in chick embryos by RNA interference	Altered limb morphology	30

Table 1: A list of some of the lncRNAs that are involved in the regulation of various developmental processes in mammals *(Adapted from Fatica, A., & Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nature Reviews Genetics, 15(1), 7-21).*

Wnt SIGNALLING: A CONSERVED SIGNALLING PATHWAY THAT IS CRUCIAL IN DEVELOPMENT

Wnt signaling is a conserved pathway that features throughout the course of mammalian development. It plays an important role in body axis patterning, lineage specification and differentiation [33]. Additionally, it is reported to maintain the fine balance between pluripotency and differentiation, mediated through the transcription network consisting of Oct4, Nanog and Sox2 [33].

Canonical Wnt signaling involves a ligand- Wnt (*Wingless* in Drosophila and *Int* in mouse), the binding of which to its cognate receptor Frizzled (along with Disheveled as a co-receptor), leads to the activation of the signaling cascade. This causes the dismantling of the destruction complex (comprising of Axin, adenomatous polyposis coli (APC), glycogen synthase kinase (GSK) β and casein kinase), ultimately leading to the stabilization of the effector molecule β-catenin. β-catenin translocates to the nucleus, where it associates with TCF/LEF family of transcription factors to facilitate the activation of Wnt- target genes. However, in the absence of the Wnt ligand, the destruction complex remains active and phosphorylates β-catenin, targeting it for degradation by the proteasomal machinery [34].

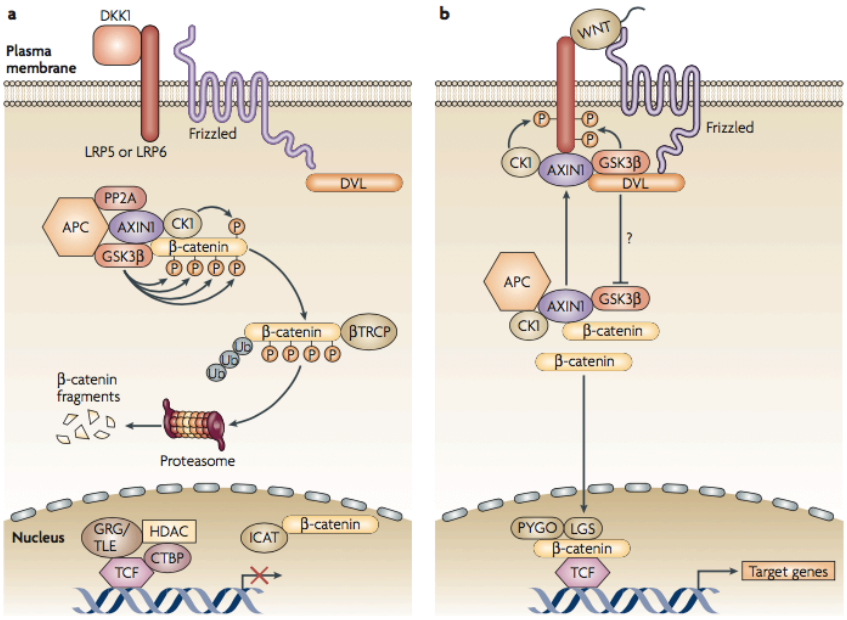


Figure 1.4: Canonical Wnt signal transduction pathway *(Adapted from* Staal, F. J., Luis, T. C., & Tiemessen, M. M. (2008). WNT signalling in the immune system: WNT is spreading its wings. *Nature Reviews Immunology*, *8*(8), 581-593.

Beta catenin, the key effector molecule in this pathway, promotes self –renewal capability of cells by associating with Oct4 on one hand and also drives differentiation of cells with TCF/LEF as its partner in a context-dependent manner [33]. In naïve pluripotent cells (e.g., mouse embryonic stem cells), Oct4 titrates away beta-catenin and exhibit a membranous localization (complexed with E-cadherin), making beta catenin unavailable for its transcriptionally active function with TCF/LEF family of transcription factors. Simultaneously, a different pool of beta-catenin represses TCF3, enhancing the effects of Oct4-Nanog driven network. A portion of the Oct4-beta-catenin complexes also occupies the promoters of certain pluripotency genes and drives their expression [35]. Alternatively, upon activation of Wnt-signalling, beta catenin serves as a co-activator for TCF1/TCF4/LEF and initiates the differentiation regime [36]. The drop in the available beta-catenin level frees TCF3, which represses the pluripotency circuitry [35]. Thus, it is observed that the effect of beta-catenin may depend on the state of pluripotency the cells are in [37].

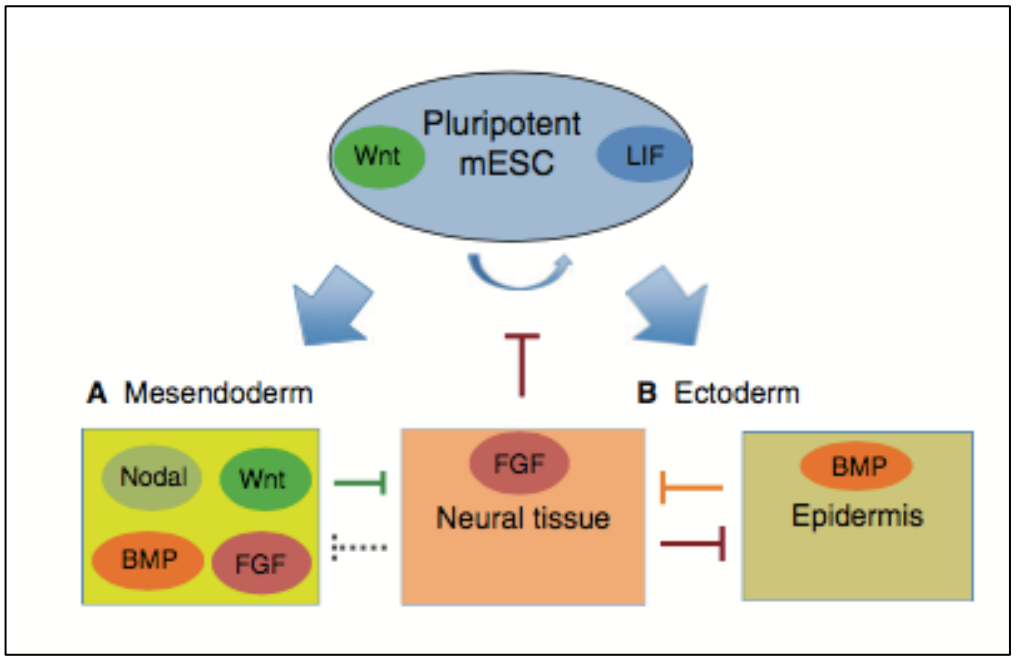


Figure 1.5: Crosstalk between different signaling pathways that determines the control between pluripotency and lineage commitment in mouse embryonic stem cells *(Adapted from Sokol, S. Y. (2011). Maintaining embryonic stem cell pluripotency with Wnt signaling. Development, 138(20), 4341-4350.)*

AIM AND SCOPE OF THE PRESENT INVESTIGATION

Earlier work from our laboratory had identified a 2.4 kb long non-coding RNA named *Mrhl* (Meiotic recombination hotspot locus), which is produced from the 15th intron of *PHKB* gene housed within the meiotic recombination hotspot locus located on chromosome 8 in mouse. It was shown to be RNA polymerase II transcribed, unspliced and polyadenylated [38]. It was expressed across multiple tissues in an adult mouse like testis, spleen, liver and kidneys but not in other tissues like heart, brain, lung and skeletal muscles [38]. In mouse spermatogonial cells, the 2.4 kb transcript showed a nuclear restricted localization [39]. Within the nucleus, it was found to be present within the nucleolus [39]. Further, the primary transcript was processed to an 80-nucleotide intermediate by Drosha machinery [39].

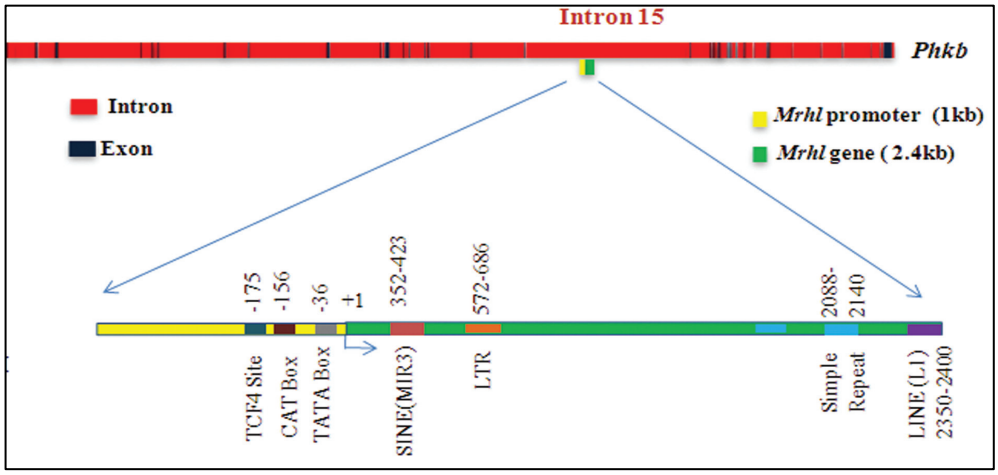


Figure 1.6: The genomic architecture of the mrhl RNA gene within the 15th intron of PHKB gene. *(Adapted from Akhade, V. S., Dighe, S. N., Kataruka, S., & Rao, M. R. S. (2016). Mechanism of Wnt signaling induced down regulation of mrhl long non-coding RNA in mouse spermatogonial cells. Nucleic acids research, 44(1), 387-401).*

For the functional characterization of *Mrhl* in mouse spermatogonial cells, microarray analysis upon *Mrhl* down regulation was performed which revealed the role of the RNA in various cellular processes including signaling [40]. A number of Wnt-signaling related genes were observed to be perturbed upon *Mrhl* down regulation. Further analysis concluded that *Mrhl* negatively regulates Wnt-signaling via its interacting partner p68, a DEAD box helicase. Upon *Mrhl* down

regulation, p68 localizes to the cytoplasm and aids in the shuttling of beta-catenin to the nucleus, which is a hallmark of Wnt-signaling activation. This nuclear beta catenin, in association with TCF 4, occupies the promoters of Wnt- target genes and activates their transcription. This Wnt activation, in turn, negatively regulates the expression of *Mrhl* [40], forming a feedback loop. This interplay between *Mrhl* and Wnt-signaling has been extensively studied and is a key event that drives mouse spermatogonial cells towards meiotic commitment [41].

Further characterization of *Mrhl* showed that it was chromatin associated and it occupied around 1370 genomic loci. Upon superimposing this dataset with the list of perturbed genes obtained from microarray analysis, 37 loci were obtained where gene regulation was achieved by the physical association of *Mrhl*, termed GRPAM loci [42]. One of these GRPAM loci was Sox8, a key transcription factor during spermatogenesis, which is regulated by *Mrhl* through promoter interaction (*unpublished data; Kataruka, S.*).

These studies indicate that as mouse spermatogonial cells (which are stem-cell like) undergo commitment to meiosis, *Mrhl* plays an indispensable role in regulating the transcriptional programme. Other work in our laboratory indicates that *Mrhl* is involved in the retinoic acid induced differentiation of mouse embryonic stem cells into neurons (*unpublished data*, *Pal, D*). Thus, the RNA seems to play a crucial role in different aspects of development like commitment and differentiation. Further, its interplay with Wnt-signalling is an interesting feature as this signaling pathway is a common theme during mammalian development and it exhibits context specific roles.

Since it was known from previous work in the laboratory that *Mrhl* is expressed in mouse embryonic stem cells, we were interested in understanding what its probable role could be in these cells. With this broad question in mind, the objectives of the present investigation were:

1. **Characterization of *Mrhl* lncRNA in mouse embryonic stem cells.**
2. **Deciphering the role of *Mrhl* in mouse embryonic stem cell circuitry and identification of interacting partners of the RNA in this context.**

CHAPTER 2

MATERIALS AND METHODS:

2.1 Cell lines, antibodies and reagents:

E14.tg2A cells were a kind gift from Prof. Tapas Kundu (JNCASR, India). siRNA scrambled and *Mrhl* pools were purchased from Dharmacon.

All fine chemicals were purchased from Sigma. DNase I enzyme was obtained from New England Biolabs (M0303S) and Human recombinant LIF from Merck-Millipore (ESG1106). The following antibodies have been used in the current study: Anti-H3 antibody from Abcam (ab46765); Anti-GAPDH antibody from Abeomics (ABM22C5).

The list of primers used in this study:

Gene	Forward Primer	Reverse Primer
Mrhl	TGAGGACCATGGCTGGACTCT	AGATGCAGTTTCCAATGTCCAAAT
Beta actin	AGGTCATCACTATTGGCAACG	TACTCCTGCTTGCTGATCCAC
U1 snRNA	CTTACCTGGCAGGGGAGAT	CAGTCCCCCACTACCACAA
SRA	TCCACCTCCTTCAAGTAAGC	GACCTCAGTCACATGGTCAACC
Oct3/4	ACCACCATCTGTCGCTTC	CCACATCCTTCTCTAGCC
Nanog	AGGGTCTGCTACTGAGATGCTCTG	CAACCACTGGTTTTTCTGCCACCG
Rex 1	ACCCTAAAGCAAGACGAGGC	GACTCGAGAAGGGAACTCGC
Sox17	GTAAAGGTGAAAGGCGAGGTG	GTCAACGCCTTCCAAGACTTG
Fgf5	GTAGCGCGACGTTTTCTTCG	AATTTGGCTTAACACACTGGC
Goosecoid	GAAGCCCTGGAGAACCTCTT	CAGTCCTGGGCCTGTACATT
Brachyury	CAGCCCACCTACTGGCTCTA	CCCCTTCATACATCGGAGAA

2.2 Cell culture and transfection:

E14.tg2A cells were cultured in 0.2 % gelatin coated dishes containing Dulbecco's Modified Eagle's Medium (DMEM, Sigma) with 15 % FBS (Life Technologies, #10082147), 0.1 mM non-essential amino acids (Life

Technologies), 0.1 mM 2-mercaptoethanol, 100 units/ml penicillin-streptomycin (Sigma) and 1000 units/ml Leukemia Inhibitory Factor (LIF) in 5 percent humidified chamber at 37 °C.

For transfection, cells were plated such that they attain 60-70 % confluence at the time of transfection. Transfection was carried out with either scrambled or pooled *Mrhl* si-RNA (100 nM) in 5% serum containing medium using Lipofectamine 2000 as per the instructions provided. Cells were harvested after 48 hours.

2.3 RNA isolation, c-DNA synthesis and qRT-PCR:

Total RNA was isolated from cells using TRIzol (Ambion) reagent using the instructions provided. c-DNA was synthesized using approximately 2.5 μg of RNA as template with oligo dT primers. For quantitative real-time PCR (qRT PCR), c-DNA was diluted 1:1 and 1/16 of it was used as template along with gene specific primers and SensiFAST SYBR No Rox RT-mix (Bioline). qRT PCR was performed using Rotor gene 6000 machine.

2.4 Cell fractionation:

Approximately, 5-10 million cells were lysed using the lysis buffer (0.8 M sucrose, 150 mM KCl, 5 mM $MgCl_2$, 6 mM 2-mercaptoethanol and 0.5% NP-40) supplemented with RNase inhibitor (75 units/ml; Thermo Fischer Scientific) and mammalian protease inhibitor cocktail (1X; Sigma) and centrifuged at 10,000 g (4°C) for 5 minutes. The supernatant containing cytoplasmic fraction was mixed with 3 volumes of TRIzol (Ambion) (for RNA extraction) or with Laemelli buffer (for immunoblotting analysis). The pellet was washed twice with the lysis buffer and utilized for RNA or protein extraction as described for the cytoplasmic fraction.

2.5 Subnuclear fractionation:

Cells (~ 10 million) were lysed with the hypotonic lysis buffer containing 10 mM Tris (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.3% (vol/vol) NP-40 and 10% (vol/vol) glycerol supplemented with RNase inhibitor (75 units/ml;

Thermo Fischer Scientific) and mammalian protease inhibitor cocktail (1X; Sigma) and centrifuged at 1000 g for 5 minutes (4°C). The supernatant comprises of the cytoplasmic fraction. The nuclear pellet was washed twice with the hypotonic lysis buffer and then resuspended in Modified Wuarin-Schibler buffer (10mM Tris-HCl (pH 7.0), 4 mM EDTA, 0.3 M NaCl, 1 M urea, and 1% (vol/vol) NP-40) along with RNase inhibitor (75 units/ml; Thermo Fischer Scientific) and mammalian protease inhibitor cocktail (1X; Sigma) and vortexed for 10 minutes. Nucleoplasmic and chromatin fractions were separated by centrifuging at 1000 g for 5 minutes (4°C). The chromatin pellet was resuspended in the sonication buffer (20mM Tris HCl, pH 7.5, 150 mM NaCl, 3mM $MgCl_2$, 0.5 mM PMSF and 75 units/ml RNase inhibitor (Thermo Fischer Scientific)), sonicated for 10 minutes and supernatant chromatin was obtained by centrifugation at 18,000 g for 10 minutes. The fractions were analysed by both immunoblotting and real time quantitative PCR analysis.

2.6 Chromatin Immunoprecipitation:

Approximately 6-7 million cells were taken and crosslinked using 1% formaldehyde for 10 minutes with gentle shaking. Crosslinking was quenched using 125 mM glycine (final concentration) for 5 minutes. Cells were washed twice with ice-cold PBS and lysed using hypotonic lysis buffer containing 10 mM Tris (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.3% (vol/vol) NP-40 and 10% (vol/vol) glycerol supplemented with RNase inhibitors (75 units/ml), 1 mM phenylmethylsulfonyl fluoride (PMSF) and mammalian protease inhibitor cocktail (1X). Nuclei were pelleted by centrifuging at 1200 g for 10 minutes (4 °C). The pellet was resuspended in the nuclear lysis buffer containing 0.1% SDS, 0.5% Triton X-100, 20 mM Tris-HCl (pH 7.5), 150 mM NaCl supplemented with RNase inhibitor (75 units/ml; Thermo Fischer Scientific), 1 mM phenylmethylsulfonyl fluoride (PMSF) and mammalian protease inhibitor cocktail (1X; Sigma). Sonication was performed (Biorupter, 25 cycles) to obtain fragments ranging from ~200 bp to 1 kb. Approximately, 15 µg of chromatin was incubated with either 4 µg of pre-immune serum or anti-H3 antibody overnight at 4 °C,

followed by incubation with protein A Dynabeads (Thermo Fischer Scientific) for 3 hours at 4 °C. The beads were washed sequentially with each of the following wash buffers for 5 minutes: Wash buffer 1: nuclear lysis buffer; Wash buffer 2: nuclear lysis buffer with 500mM NaCl; Wash buffer 3: 10 mM Tris-Cl (pH 8.0), 0.5 % NP-40, 0.5 % sodium deoxycholate; 1 mM EDTA supplemented with RNase inhibitor (75 units/ml; Thermo Fischer Scientific). Beads were used either for protein analysis or RNA isolation. For RNA isolation, beads were mixed with elution buffer: 100 mM NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.5 % SDS and 100 μg/ml Proteinase K (Thermo Fischer Scientific) and incubated at 55°C for an hour. Crosslinking was reversed by heating at 95°C for 10 minutes. Chromatin bound RNA was isolated using Trizol (Ambion), followed by DNase I treatment to remove genomic DNA and analyzed further by quantitative real time PCR.

2.7 p68 Immunoprecipitation:

Cells were lysed in the hypotonic lysis buffer containing 10 mM Tris (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.3% (vol/vol) NP-40 and 10% (vol/vol) glycerol supplemented with RNase inhibitors (75 units/ml), 1 mM phenylmethylsulfonyl fluoride (PMSF) and mammalian protease inhibitor cocktail (1X). Nuclei were pelleted down at 1200 *g* for 10 minutes (4°C) and subsequently, lysed in the nuclear lysis buffer (150 mM KCl, 25 mM Tris (pH 7.4), 5mM EDTA, 0.5 % NP-40, 1X mammalian protease inhibitor cocktail, 1 mM phenylmethylsulfonyl fluoride (PMSF), 75 units/ml RNase inhibitor). The debris were removed by centrifuging at 15000 *g* and the nuclear lysate was utilized for immunoprecipitation. To 1 mg of nuclear lysate, 7 μg of either pre-immune serum or p68 polyclonal antibody (rabbit) was added and incubated overnight at 4 °C, followed by incubation with protein A Dynabeads (Thermo Fisher Scientific) for 3 hours at 4 °C. The beads were washed with wash buffer (20 mM Tris, pH 7.4, 2 mM magnesium chloride, 10 mM potassium chloride, 150 mM sodium chloride, 10% glycerol, 1X mammalian protease inhibitor cocktail, 1 mM phenylmethylsulfonyl fluoride (PMSF), 75 units/ml RNase inhibitor

(Thermo Fischer Scientific)) containing 0.2 % Nonidet P-40 (octylphenoxypolyethoxyethanol) and subsequently, twice with wash buffer containing 0.5 % Nonidet P- 40. The proteins were eluted from the beads in Laemelli buffer and separated on a 10 % SDS-PAGE gel. The proteins were transferred onto a nitrocellulose membrane. The blot, after blocking in 5% skim milk, was incubated overnight with the primary p68 antibody (1:1000 dilution in 1% skim milk in 0.05% Tween 20 in PBS) at 4 °C. The blot was washed once with 0.05% Tween 20 in PBS and incubated with HRP-conjugated goat anti-rabbit antibody (1:4000 dilution in 1% skim milk in 0.05% Tween 20 in PBS) at room temperature for 1 hour. Subsequently, the blot was washed thrice with 0.05% Tween 20 in PBS). Using luminol as substrate, the blot was visualised using chemiluminescence-based detection.

2.8 LIF withdrawal mediated differentiation:

E14.tg2A cells were cultured for two passages as described above. The cells were then transferred to the culture medium without LIF and were harvested subsequently at day 0, 2, 3, 4 and 5. Gene expression analysis was performed by total RNA isolation and quantitative real time PCR.

2.9 S1 aptamer based RNA pulldown:

E14.tg2A cells were seeded such that they reach 60-70 percent confluence at the time of transfection. Transfection was performed with either S1 aptamer tagged mrhl cloned into pcDNA3.1 or the empty vector for mock pulldown (1.5 μg/ml) using Lipofectamine 2000 as per the instructions provided. Cells were harvested after 48 hours and crosslinked with 1% formaldehyde for 10 minutes with gentle agitation. Crosslinking was quenched with 125mM glycine (final concentration) for 5 minutes. Cells were washed twice with ice cold PBS. Lysis was performed using 10mM Tris-Cl (pH 8.0), 150 mM NaCl, 5mM $MgCl_2$, 0.1mM EDTA, 0.8 % Triton X-100, 5 % glycerol, 1X mammalian protease inhibitor cocktail (Sigma), 1 mM phenylmethylsulfonyl fluoride (PMSF), 75 units/ml RNase inhibitor (Thermo Fischer Scientific) and centrifuged at 1000 g 10 minutes (4 °C). The lysate was incubated with streptavidin-agarose beads (Thermo Fischer

Scientific) in the presence of yeast tRNA (100 μg/ml) and RNase inhibitor (75 units/ml; Thermo Fischer Scientific) overnight at 4 °C. Subsequently, the beads were washed thrice with the lysis buffer for 5 minutes each. For protein analysis, beads were mixed with Laemelli buffer, boiled and separated on a 10% SDS-polyacrylamide gel. Silver staining was performed using Silver Quest staining kit (Invitrogen) to identify all the interacting proteins. For RNA isolation, beads were directly mixed with TRIzol (Ambion) and processed further as per instructions provided.

CHAPTER 3

RESULTS AND DISCUSSION:

3.1 BASIC CHARACTERISATION OF *Mrhl* RNA IN MOUSE EMBRYONIC STEM CELLS

3.1.1 *Mrhl* RNA is expressed in mouse embryonic stem cells at a comparable level to that of mouse spermatogonial cells:

Mrhl RNA showed a high level of expression in adult testis and hence, the initial characterization of the RNA was performed in the mouse spermatogonial stem cell line GC1-spg. Since we knew that Mrhl was expressed in mouse embryonic stem cell line E14.tg2A, we wished to know whether the expression level was similar to that of GC1-spg. By quantitative real time PCR, it was seen that there was no significant difference between the expression levels of *Mrhl* RNA in both the cell lines indicating that it is expressed to a similar level in mouse embryonic stem cells as compared to mouse spermatogonial cells (Figure 3.1.1).

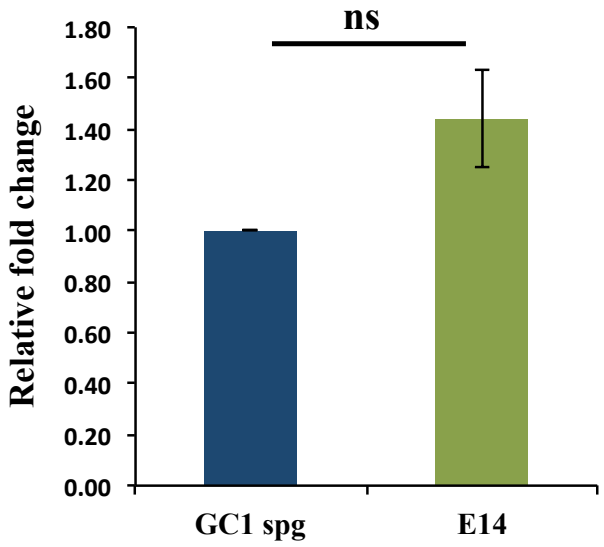


Figure 3.1.1: Expression of *Mrhl* RNA in mouse embryonic stem cell line E14.tg2A as compared to mouse spermatogonial stem cell line GC1-spg. *(Results representative of three independent biological replicates and error bars represent standard deviation).*

3.1.2 *Mrhl* is nuclear restricted in mouse embryonic stem cells:

Long non-coding RNAs show context specific functions, which are primarily dependent on the compartment they are localized to. We were interested in knowing where *Mrhl* RNA is localized within the cell and hence, we resorted to subcellular fractionation to address this. It was observed that *Mrhl* is nuclear restricted as its expression pattern matched that of *U1 snRNA*, which is a bona-fide nuclear limited RNA. The purity of the fractions was assessed by immunoblotting for which cytoplasmic marker was Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and the nuclear marker was histone H3 (Figure 3.1.2).

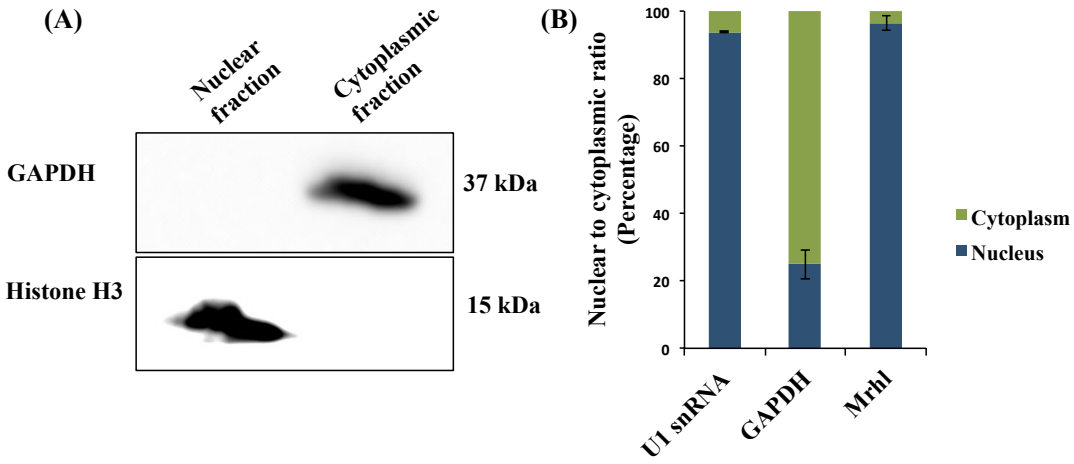


Figure 3.1.2 (I): *Mrhl* is nuclear restricted in mouse embryonic stem cells. (A) Immunoblotting of the subcellular fractions using *GAPDH* as a cytoplasmic marker and *histone H3* as a nuclear marker. (B) Quantitative real time PCR analysis of the percentage localization of *Mrhl* RNA. *GAPDH* and *U1 snRNA* have been used as references. *(Results representative of three independent biological replicates and error bars represent standard deviation)*.

Previously, fluorescence in situ hybridization experiments had been carried out in E14.tg2A cells using an antisense probe tagged with Cy5 fluorophore, which showed that *Mrhl* localized clearly to the nucleus (Figure 3.1.2 (II),

unpublished data; Iyer, D.), thus confirming the findings of the above experiment.

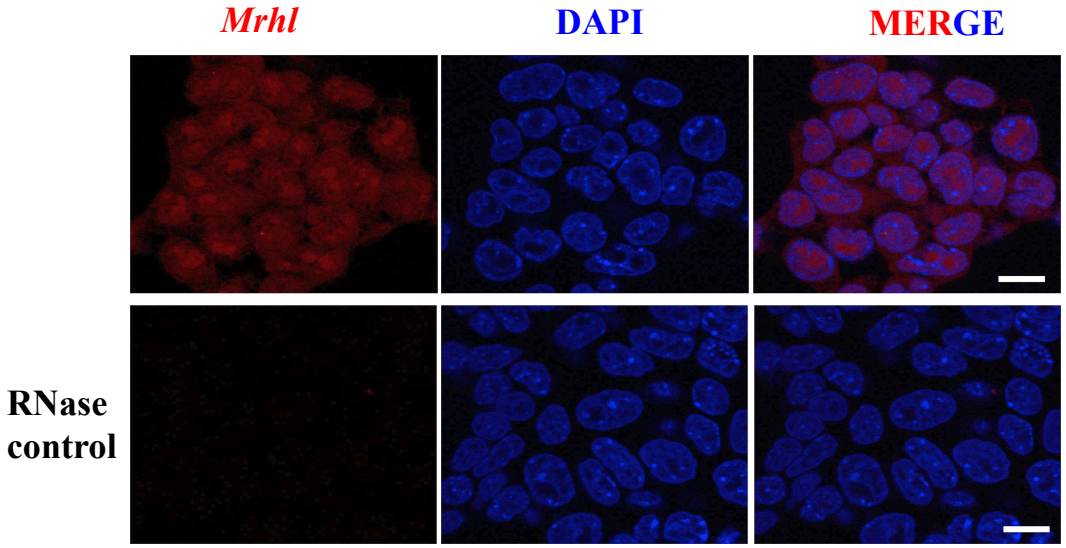


Figure 3.1.2 (II): *Mrhl* is nuclear restricted in mouse embryonic stem cells. Fluorescence in-situ hybridization images showing nuclear localization of *Mrhl* RNA. RNase A treated sample showing the specificity of the fluorescence probe against *Mrhl*. Scale bar = 10 µm. *(Unpublished data; Iyer, D.)*

3.1.3 *Mrhl* RNA is chromatin associated in mouse embryonic stem cells

Since *Mrhl* showed nuclear localization, we wished to probe whether it was chromatin bound as in the case of mouse spermatogonial cells or not. Upon subnuclear fractionation, we found that *Mrhl* was indeed associated with the chromatin (a small fraction was also present in nucleoplasm). The purity of the fractions was analysed by using specific markers namely Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) for cytoplasm and histone H3 for the chromatin fraction. *U1 snRNA* was used as a positive marker, which shows a clear chromatin association but a fraction of it is also present in the nucleoplasm (Figure 3.1.3 (I)).

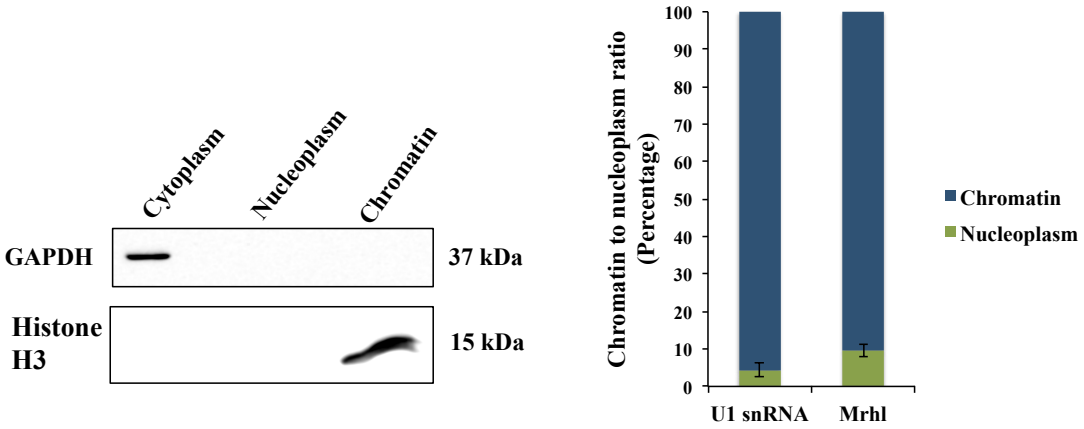


Figure 3.1.3 (I): *Mrhl* is a chromatin associated RNA. (A) Immunoblotting analysis to show the purity of the fractions upon subnuclear fractionation: *GAPDH* and *histone H3* have been used as markers for cytoplasmic and chromatin fractions respectively. (B) Quantitative real time PCR analysis of the percentage localization of *Mrhl* in different fractions. *U1 snRNA* has been used as a positive control. *(Results representative of two independent biological replicates, each in duplicates and error bars represent standard deviation).*

To further confirm that *Mrhl* is chromatin bound, we performed a histone H3 chromatin immunoprecipitation and scored for the presence of the RNA. *Mrhl* showed a clear association with the chromatin similar to that of *U1 snRNA* (Figure 3.1.3 (II)).

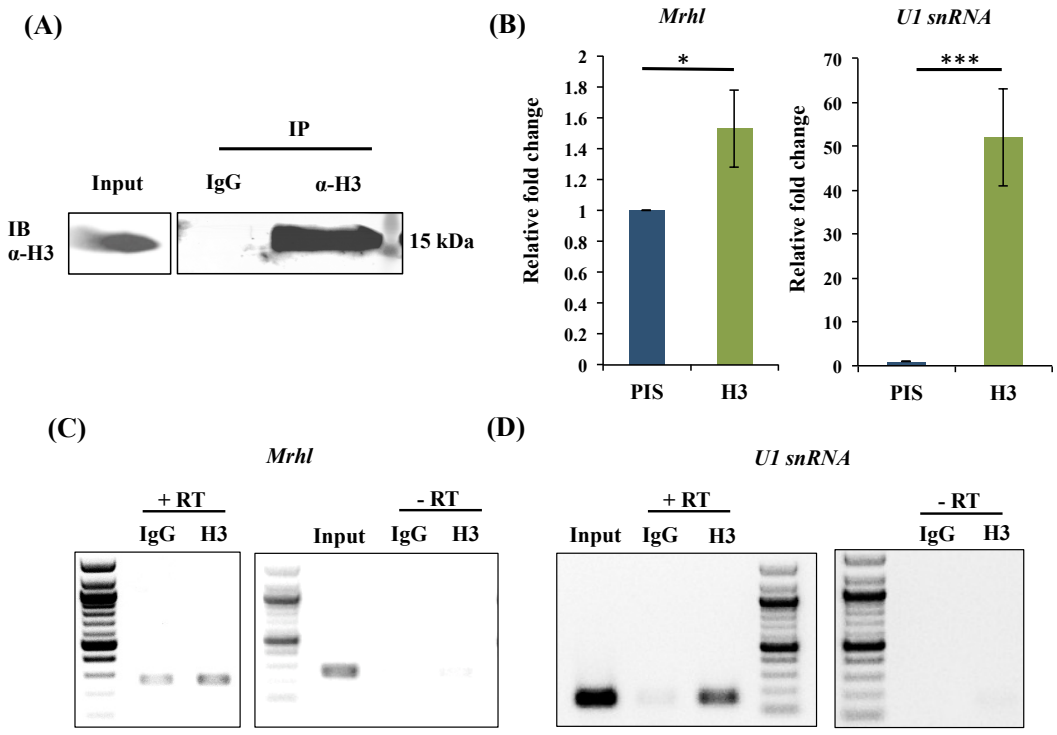


Figure 3.1.3 (II): *Mrhl* is associated with oligonucleosomal chromatin. (A) Immunoblot (IB) showing the specificity of H3 immunoprecipitation (IP) with antibody against H3 or pre-immune IgG (negative control). Input serves as a positive control. (B) Quantitative real time PCR analysis showing enrichment of *Mrhl* and *U1 snRNA* (positive control) upon H3 chromatin immunoprecipitation using H3 antibody as compared to pre-immune serum. (C) and (D) Semi-quantitative PCR analysis showing the enrichment of *Mrhl* and *U1 snRNA* respectively upon H3 chromatin immunoprecipitation. (–RT (minus reverse transcriptase) lanes serve as controls to check for genomic DNA contamination). *(Results representative of three independent biological replicates and error bars represent standard deviation. Statistical significance calculated using two tailed Students' t-test; * represents $p<0.05$; *** represents $p<0.001$).*

3.1.4 p68, a DEAD box helicase, does not interact with *Mrhl* RNA in mES cells

In mouse spermatogonial cells, p68, a DEAD box helicase was identified as a novel interacting partner of *Mrhl* RNA. So, we probed whether this interaction exists in the context of embryonic stem cells. Upon

immunoprecipitation of p68, we found that *Mrhl* does not interact significantly with p68 unlike *SRA*, which is a known interacting partner of the DEAD box helicase (Figure 3.1.4).

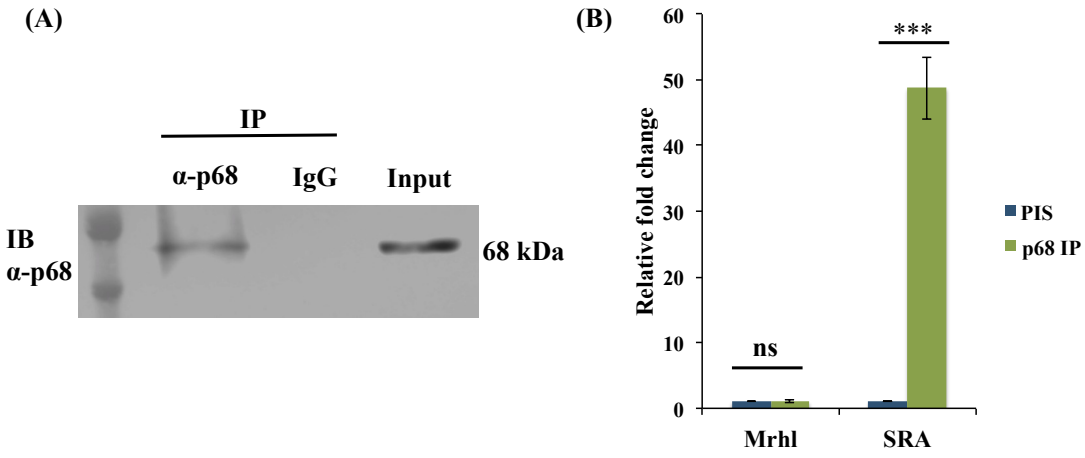


Figure 3.1.4: p68, a DEAD box helicase, does not interact with *Mrhl* in mES cells. (A) Nuclear lysate of E14.tg2A cells immunoprecipitated (IP) using p68 antibody or pre-immune IgG (negative control) and immunoblotted with p68 antibody to check for the specificity of the pulldown. Input has been used as a positive control. (B) Quantitative real-time PCR analysis showing the enrichment of *Mrhl* or *SRA* (positive control) RNA upon p68 immunoprecipitation as compared to pre-immune serum. *(Results representative of three independent biological replicates and error bars represent standard deviation. Statistical significance calculated using two tailed Students' t-test; ns represents p>0.05; *** represents p<0.001).*

3.1.5 Wnt signaling status upon *Mrhl* down regulation

As previously mentioned, *Mrhl* negatively regulates Wnt-signaling through its interaction with p68 in mouse spermatogonial cells. Since, the interaction of *Mrhl* and p68 is absent in the context of embryonic stem cells, we were interested in knowing the status of Wnt- signaling in these cells. Beta- catenin, the effector molecule of Wnt- signalling, translocates to the nucleus upon Wnt activation. Upon shRNA mediated

down regulation of *Mrhl*, it was seen that beta- catenin exhibits a membranous localization (without any signal from the nucleus) similar to control cells (Figure 3.1.5*, unpublished data; Iyer, D.*). This indicates that *Mrhl* RNA might not be involved in regulating Wnt- signaling in mouse embryonic stem cells and might have a context specific role in different cell types.

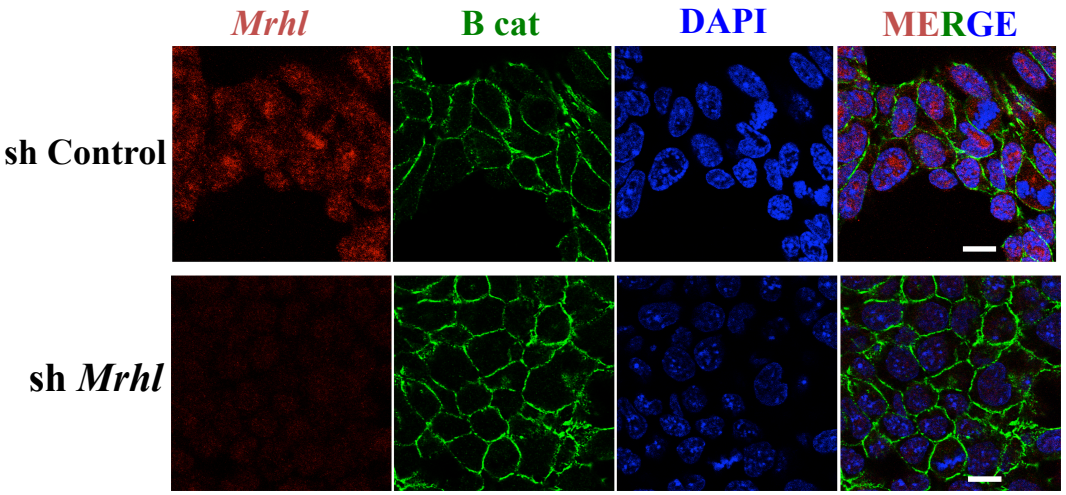


Figure 3.1.5: Wnt-signaling status upon *Mrhl* down regulation. Fluorescence in-situ hybridization and immunofluorescence images showing a membranous localization of beta-catenin in *Mrhl* sh-RNA transfected cells similar to control cells (scrambled shRNA). Scale bar = 10 µm. *(Unpublished data; Iyer, D.)*

3.2 DECIPHERING THE ROLE OF Mrhl RNA IN MOUSE EMBRYONIC STEM CELL CIRCUITRY AND IDENTIFICATION OF THE INTERACTING PARTNERS

To determine the functions exerted by a lncRNA in the cellular context, there are largely two approaches employed. First approach involves depletion of the RNA of interest to ascertain the possible

processes/pathways that the RNA might be regulating or is involved in. Alternately, identification of the interacting protein partners of the RNA by itself can give us an insight into which processes the lncRNA might be playing a part in. We employed both these methods to decipher the role of *Mrhl* in mouse embryonic stem cells.

3.2.1 Down regulation of *Mrhl* RNA and global gene expression analysis

To deplete the RNA, we resorted to transient knockdown using siRNA. Post 48 hours of transfection, we observed up to 60 percent down regulation (Figure 3.2.1 (I) A). To check whether this down regulation of *Mrhl* affects the pluripotency of mouse embryonic stem cells, we scored for the markers *Oct4*, *Nanog* and *Rex1*.

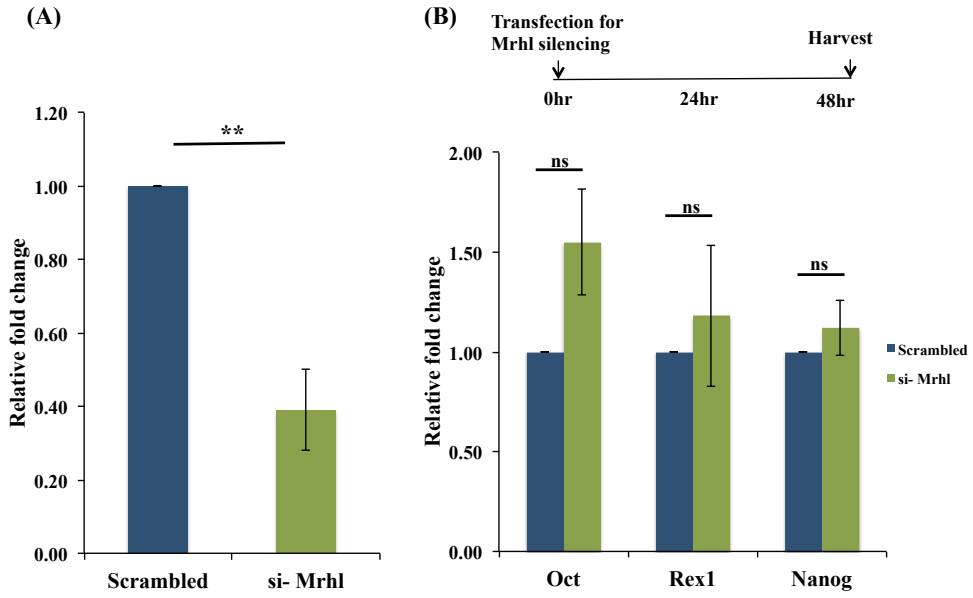
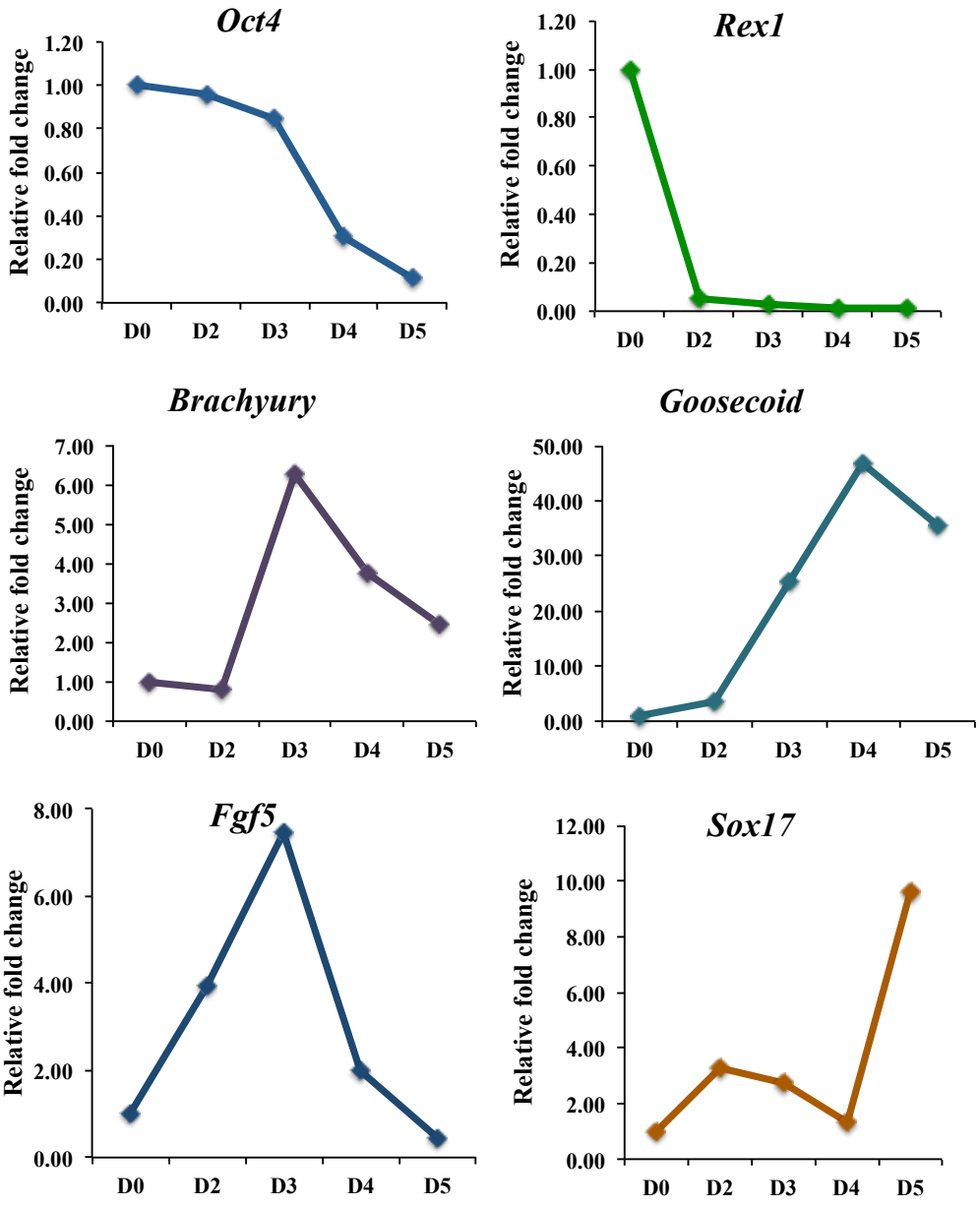


Figure 3.2.1 (I): Down regulation of *Mrhl* RNA and expression analysis of pluripotency markers. (A) Quantitative real time PCR analysis showing down regulation of *Mrhl* using siRNA. (B) Expression analysis of *Oct4*, *Rex1* and *Nanog* (pluripotency markers) upon down regulation of *Mrhl (*as compared to scrambled siRNA). *(Results representative of three independent biological replicates and error bars represent standard deviation. Statistical significance calculated using two tailed Students' t-test; ** represents p<0.01; ns represents p>0.05).*

We found no significant changes in the pluripotency marker levels (Figure 3.2.1 (I) B) indicating that this siRNA-mediated down regulation of *Mrhl* RNA does not directly affect the pluripotent status of the mES cells. Additionally, it also suggests that *Mrhl* is not directly involved in the maintenance of pluripotency. We are in the process of identifying the processes/ pathways perturbed upon *Mrhl* down regulation by global gene expression profiling.

Meanwhile, to decipher whether *Mrhl* was involved in lineage specification, we resorted to LIF withdrawal mediated differentiation of mouse embryonic stem cells. Removal of LIF serves as a cue for differentiation to set in to form derivatives of the three germ layers- namely ectoderm, endoderm and mesoderm. Upon inducing differentiation, we observed that the expression profiles of some of the lineage specific transcription factors like *Oct4*, *Rex1*, *Brachyury*, *Goosecoid*, *Sox17* and *Fgf5* (Figure 3.2.1 (II) A) matched those reported previously [43, 44]. Additionally, in this system, we scored for the expression of *Mrhl.*

(A)



(Figure legend on the next page)

(B)

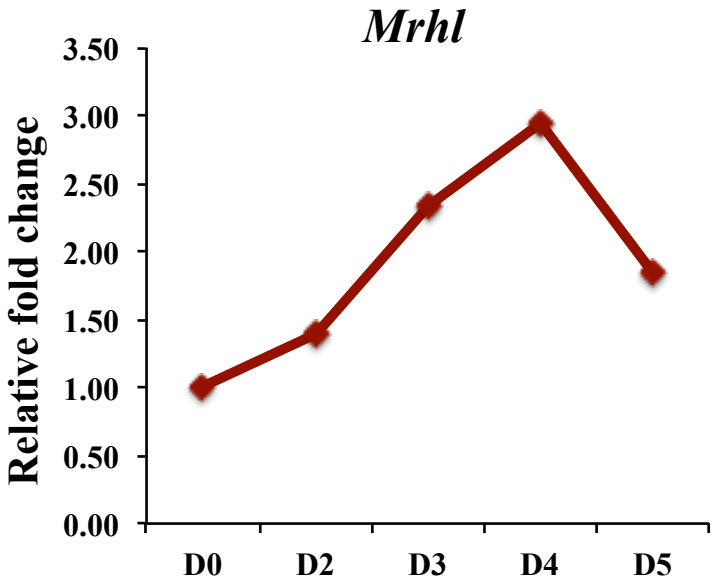


Figure 3.2.1 (II): *Mrhl* is upregulated as mouse embryonic stem cells undergo differentiation. (A) Expression profiles of markers of different lineages- *Oct4* (pluripotency), *Rex1* (pluripotency), *Brachyury* (primitive streak, mesendoderm), *Goosecoid* (mesoderm), *Fgf5* (ectoderm) and *Sox17* (endoderm) over the course of day 0 to day 5 post withdrawal of leukemia inhibitory factor (LIF). (B) Expression profile of *Mrhl* over the course of differentiation. (*Results represented an average of two independent experiments).*

Upon inducing differentiation by LIF withdrawal, we found that *Mrhl* is up regulated and follows the trend of other transcription factors like *Brachyury* and *Goosecoid,* indicating that the RNA might be involved in lineage specification or differentiation. Global gene expression profiling might lend some insights into this aspect as well.

3.2.2 Identification of interacting protein partners of *Mrhl* in mESCs

LncRNAs along with proteins form ribonucleoprotein complexes through which they exert a number of functions. Hence, identifying the

protein partners of the lncRNA of interest might shed light on the sort of cellular processes it might be involved in. Towards this goal, we tagged mrhl with S1 aptamer tag towards its 3' end in pcDNA 3.1 vector. S1 aptamers are high affinity RNA tags that fold to form a secondary structure, which specifically binds streptavidin.

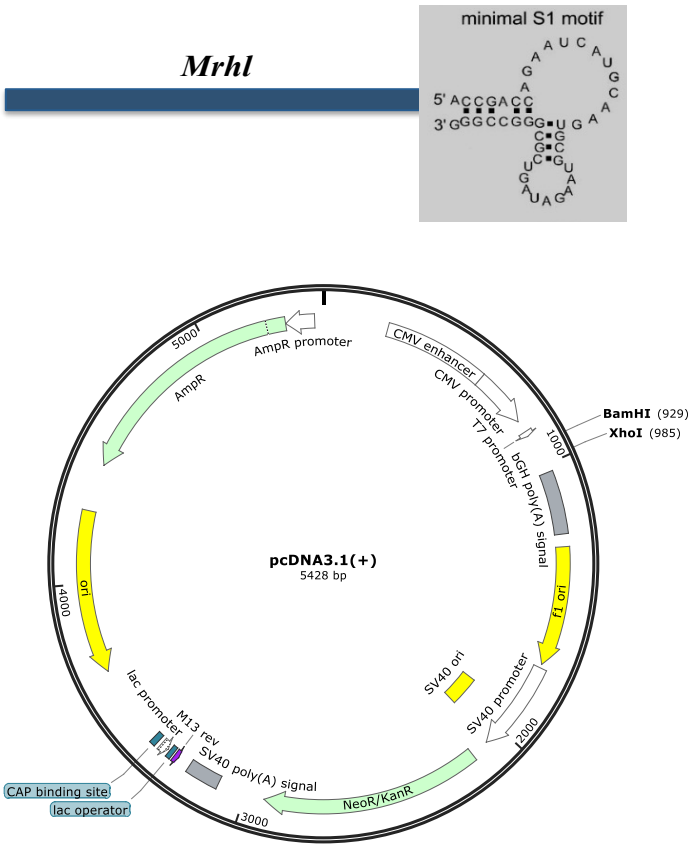


Figure 3.2.2(I): Vector map of pCDNA3.1 in which mrhl gene tagged with S1-aptamer (towards 3' end) was cloned under CMV promoter.

Upon overexpressing this RNA and its subsequent pull-down, we observed high enrichment of the *Mrhl* as compared to the mock pull-down (vector only) indicating that the pull-down with streptavidin beads is specific. We observe a number of protein bands (approximately 45, 50, 72, 90, 100, 105 kDa bands as shown by arrows in Figure 3.2.2 (II)) that are enriched in the Mrhl-S1 aptamer pull-down compared to the mock pull-down as seen in the silver stained gel. We plan to identify these proteins by mass spectrometric analysis.

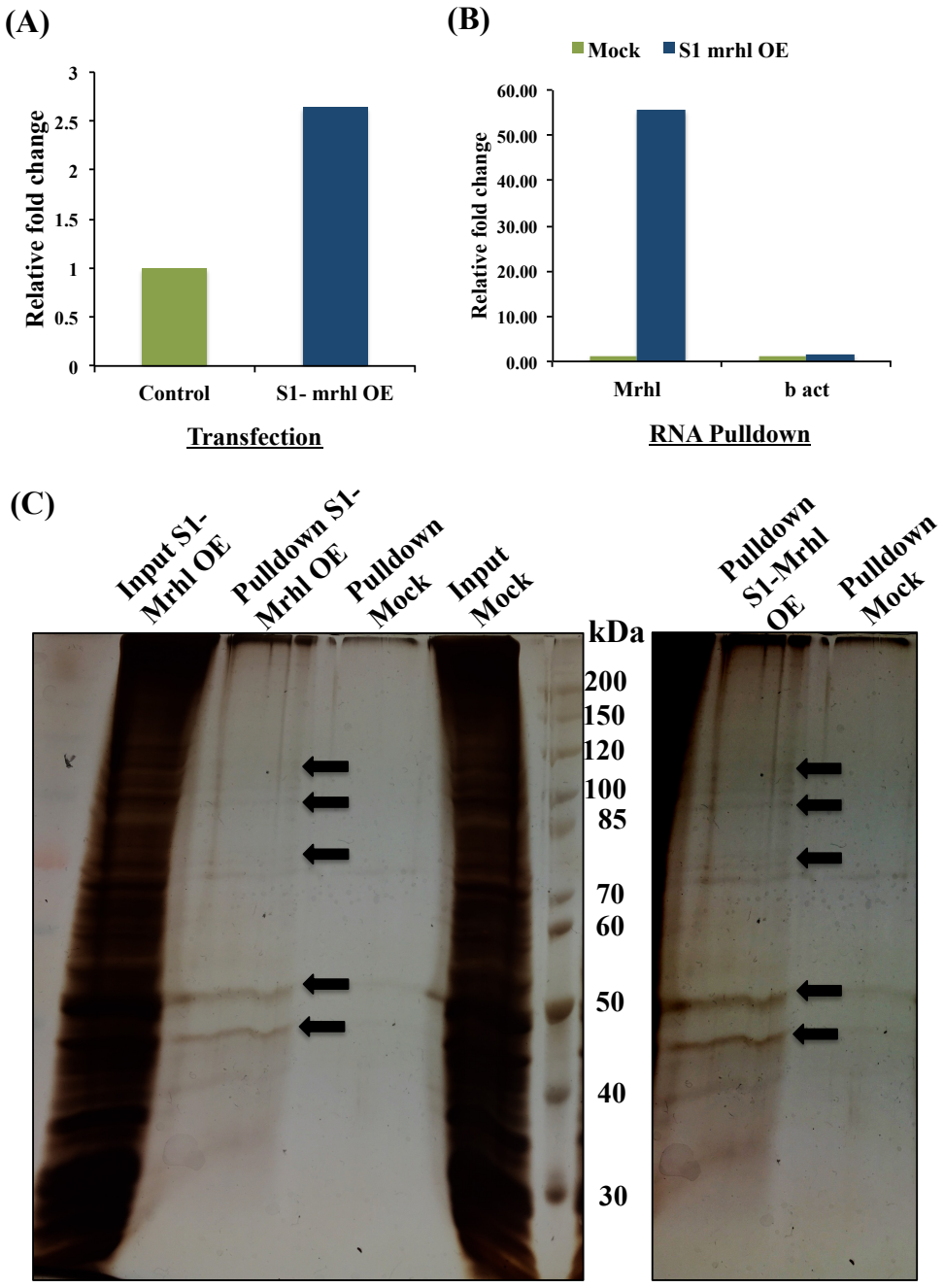


Figure 3.2.2(II): Identification of interacting protein partners of *Mrhl* in mouse embryonic stem cells. (A) Quantitative real time PCR analysis showing overexpression of S1-aptamer tagged *mrhl* vector as compared to mock vector (control) transfection. (B) Enrichment of *Mrhl* upon streptavidin-agarose pull-down in S1-mrhl transfected cells compared to mock pull-down indicating the specificity of the RNA pull-down. Beta actin serves as a negative control. (C) Silver stained gel showing enriched protein bands (black arrows) in S1-aptamer tagged *Mrhl* pull-down compared to mock pull-down. (Pull-down lanes shown on the right side upon increased exposure).

CHAPTER 4

SUMMARY:

LncRNAs have emerged as sentinels in the regulation of various cellular processes and hence, it is not surprising that the number of long non-coding RNA genes correlate with developmental complexity over the course of evolution. The versatility of RNA as a biomolecule has been instrumental in its participation in diverse cellular contexts.

Meiotic recombination hotspot locus *(Mrhl)* is a long non-coding RNA that was discovered in our laboratory. It showed a tissue specific expression pattern and was highly expressed in testis, liver, spleen and kidneys [(38)]. Owing to its high expression in adult testis, the initial characterization of the RNA was carried out in a B-type mouse spermatogonial cell line GC1-spg. In these cells, Mrhl was seen to be a nuclear restricted RNA that was processed to a 80 nucleotide transcript by Drosha machinery [(39)]. Further, it was shown to be a chromatin associated RNA with a role in regulating gene expression of several loci (some of which were exerted by its physical association) [(42)]. Functional analysis of the RNA revealed its role in the regulation of one of the important signaling pathway, Wnt, which is crucial all along the course of development. Wnt signaling in turn negatively regulates *Mrhl*, forming a feedback loop [(40)]. This interplay between *Mrhl* and Wnt has been extensively characterized in our laboratory and has been found to be a key step in meiotic commitment of mouse spermatogonial cells [(41)].

Since long non-coding RNAs exhibit tissue specific and context specific function, the present study involves the basic characterization of *Mrhl* in mouse embryonic stem cells. We find that similar to mouse spermatogonial cells, *Mrhl* is a nuclear restricted and chromatin associated RNA. However, unlike the case in Gc1-spg cells, *Mrhl* does not interact with p68, a DEAD box helicase. This interaction of *Mrhl* with p68 is instrumental in its regulation of Wnt- signaling. Hence, it was not surprising to find that in mouse embryonic stem cells, down regulation of *Mrhl* does not result in the nuclear translocation of beta- catenin, which is a hallmark of canonical Wnt signaling activation (*unpublished work; Iyer, D.*).

To understand the function exerted by *Mrhl* in mouse embryonic stem cells, we resorted to two approaches. Firstly, upon down regulation of the RNA, we find no change in the pluripotency markers indicating that *Mrhl* is not directly involved in the maintenance of pluripotency. Further insights into the functions of the RNA will be obtained by global gene expression profiling upon down regulation of *Mrhl*. Additionally, preliminary results suggest that *Mrhl* is up regulated as mouse embryonic stem cells undergo differentiation. The second approach employed was to determine the interacting protein partners of the *Mrhl* by RNA pull-down followed by mass spectrometric analysis.

It will be interesting to know the repertoire of interacting partners the RNA can have and the list of processes it might be regulating in this context. Since *Mrhl* is chromatin bound in mouse embryonic stem cells too, it will be enlightening to map the genome wide occupancy of this RNA in this system, which is epigenetically quite unique. Also, these studies will also provide us with a comparative analysis of the common and exclusive roles *Mrhl* can carry out in these two different systems.

REFERENCES:

1. Dinger, M., Amaral, P. & Mercer, T. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 1433–1445 (2008).
2. Fatica, A., & Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics*, *15*(1), 7-21.
3. Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes & development*, *23*(13), 1494-1504.
4. Clark, B. S., & Blackshaw, S. (2015). Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. *Emerging Roles of Long Noncoding RNAs in Neurological Diseases and Metabolic Disorders*, 32.
5. Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution—trashing old rules to forge new ones. *Cell*, *157*(1), 77-94.
6. Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, *15*(6), 423.
7. Kornienko, A. E., Guenzl, P. M., Barlow, D. P., & Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC biology*,*11*(1), 1.
8. Diederichs, S. (2014). The four dimensions of noncoding RNA conservation.*Trends in Genetics*, *30*(4), 121-123.
9. Eddy, S. R. (2001). Non–coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, *2*(12), 919-929.
10. Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA biology*, *10*(6), 924-933.
11. Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*, *10*(3), 155-159.
12. Fatima, R., Akhade, V. S., Pal, D., & Rao, S. M. (2015). Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. *Molecular and cellular therapies*, *3*(1), 1.

13. Guttman, M., & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature*, *482*(7385), 339-346.
14. Quinn, J. J., & Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics*, *17*(1), 47-62.
15. Geisler, S., & Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature reviews Molecular cell biology*, *14*(11), 699-712.
16. Wang, K. C., & Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Molecular cell*, *43*(6), 904-914.
17. Mattick, J. S. (2007). A new paradigm for developmental biology. *Journal of Experimental Biology*, *210*(9), 1526-1547.
18. Flynn, R. A., & Chang, H. Y. (2014). Long noncoding RNAs in cell-fate programming and reprogramming. *Cell stem cell*, *14*(6), 752-761.
19. Rosa, A., & Ballarino, M. (2015). Long Noncoding RNA Regulation of Pluripotency. *Stem cells international*, *2016*.
20. Lee, J. T. (2012). Epigenetic regulation by long noncoding RNAs. *Science*,*338*(6113), 1435-1439.
21. Rasmussen, T. P. (2003). Embryonic stem cell differentiation: a chromatin perspective. *Reprod Biol Endocrinol*, *1*(1), 100.
22. Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., ... & Fisher, A. G. (2006). Chromatin signatures of pluripotent cell lines.*Nature cell biology*, *8*(5), 532-538.
23. Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., ... & Jaenisch, R. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*(2), 315-326.
24. Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., ... & Lee, W. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(7153), 553-560.
25. Mohamed, J. S., Gaughwin, P. M., Lim, B., Robson, P., & Lipovich, L. (2010). Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA*, *16*(2), 324-337.

26. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., ... & Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, *129*(7), 1311-1323.
27. Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., ... & Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, *329*(5992), 689-693.
28. Lin, N., Chang, K. Y., Li, Z., Gates, K., Rana, Z. A., Dang, J., ... & Head, S. R. (2014). An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Molecular cell*, *53*(6), 1005-1019.
29. Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., Steinhauser, M. L., ... & Abo, R. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*, *152*(3), 570-583.
30. Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., ... & Herrmann, B. G. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental cell*, *24*(2), 206-214.
31. Jiang, W., Liu, Y., Liu, R., Zhang, K., & Zhang, Y. (2015). The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. *Cell reports*, *11*(1), 137-148.
32. Young, T. L., Matsuda, T., & Cepko, C. L. (2005). The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Current Biology*, *15*(6), 501-512.
33. Sokol, S. Y. (2011). Maintaining embryonic stem cell pluripotency with Wnt signaling. *Development*, *138*(20), 4341-4350.
34. Staal, F. J., & Clevers, H. C. (2005). WNT signalling and haematopoiesis: a WNT–WNT situation. *Nature Reviews Immunology*, *5*(1), 21-30.
35. Faunes, F., Hayward, P., Descalzo, S. M., Chatterjee, S. S., Balayo, T., Trott, J., ... & Arias, A. M. (2013). A membrane-associated β-catenin/Oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells. *Development*, *140*(6), 1171-1183.

36. Gordon, M. D., & Nusse, R. (2006). Wnt signaling: multiple pathways, multiple receptors, and multiple transcription factors. *Journal of Biological Chemistry*, *281*(32), 22429-22433.
37. Lien, W. H., & Fuchs, E. (2014). Wnt some lose some: transcriptional governance of stem cells by Wnt/β-catenin signaling. *Genes & development*, *28*(14), 1517-1532.
38. Nishant, K. T., Ravishankar, H., & Rao, M. R. S. (2004). Characterization of a mouse recombination hot spot locus encoding a novel non-protein-coding RNA. *Molecular and cellular biology*, *24*(12), 5620-5634.
39. Ganesan, G., & Rao, S. M. (2008). A novel noncoding RNA processed by Drosha is restricted to nucleus in mouse. *Rna*, *14*(7), 1399-1410.
40. Arun, G., Akhade, V. S., Donakonda, S., & Rao, M. R. S. (2012). mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells. *Molecular and cellular biology*, *32*(15), 3140-3152.
41. Akhade, V. S., Dighe, S. N., Kataruka, S., & Rao, M. R. S. (2016). Mechanism of Wnt signaling induced down regulation of mrhl long non-coding RNA in mouse spermatogonial cells. *Nucleic acids research*, *44*(1), 387-401.
42. Akhade, V. S., Arun, G., Donakonda, S., & Satyanarayana Rao, M. R. (2014). Genome wide chromatin occupancy of mrhl RNA and its role in gene regulation in mouse spermatogonial cells. *RNA biology*, *11*(10), 1262-1279.
43. Shigeta, M., Ohtsuka, S., Nishikawa-Torikai, S., Yamane, M., Fujii, S., Murakami, K., & Niwa, H. (2013). Maintenance of pluripotency in mouse ES cells without Trp53. *Scientific reports*, *3*.
44. Lindsley, R. C., Gill, J. G., Kyba, M., Murphy, T. L., & Murphy, K. M. (2006). Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. *Development*, *133*(19), 3787-3796.