Characterization and Functional Analysis of a Novel Non-Coding RNA (mrhl) from Mouse

A Thesis
Submitted For the Degree of

Doctor of Philosophy

by

G. Gayatri

Molecular Biology and Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
Bangalore – 560 064

SEPTEMBER 2009

Dedicated to my little unborn.....

DECLARATION

I hereby declare that the matter embodied in the thesis entitled "**Characterization and Functional Analysis of a Novel Non-Coding RNA (mrhl) from Mouse**" is the result of investigations carried out by me at the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under the supervision of Prof. M. R. S. Rao and that it has not been submitted elsewhere for the award of any degree or diploma.

In keeping with the general practice in reporting scientific observations, due acknowledgement has been made whenever the work described is based on the findings of other investigators.

G. Gayatri

CERTIFICATE

I hereby certify that the matter embodied in this thesis entitled "**Characterization and Functional Analysis of a Novel Non-Coding RNA (mrhl) from Mouse**" has been carried out by Ms. G. Gayatri at the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under my supervision and that it has not been submitted elsewhere for the award of any degree or diploma.

Prof. M. R. S. Rao
(Research Supervisor)

Acknowledgements

At the outset, I would like to thank my mentor and my thesis supervisor Prof M.R.S Rao, for having given me a very challenging project to work on. I am extremely grateful to him for all the encouragement and support he has extended during my course of investigation. I have always admired his perseverance and passion towards science. His relentless approach to address new questions in biology has exposed me to lot of new things and has given me an opportunity to try different approaches which has been a real learning experience.

I thank all the faculty members of M.B.G.U, the Chairman, Prof. Anuranjan Anand, Prof. Tapas Kundu, Prof. Namita Surolia, Prof. Hemalatha Balaram, Prof. Ranga Uday, Prof. Maneesha Inamdar and Dr. Kaustav Sanyal for the excellent course work that was offered and for their constant encouragement and constructive criticism that has helped me to shape my work in great deal. Specially, I would like to thank Dr. Maneesha for sparing ES cell cultures that were used in the later part of the investigation.

Timely help rendered by Anitha of sequencing facility , Suma of confocal facility, Rupa of mass spec facility and Dr. Prakash of animal house is gratefully acknowledged. Special thanks to Anitha and Suma for extending their help outside the facility as well.

I would like acknowledge Prof Narry V Kim of Seoul National University for her kind gift of the Flag Drosha Construct and the anti Drosha antibodies.

I am extremely thankful to Prof.K.S Narayan and his family for their concern and well wishes.

My due acknowledgement to all my teachers, starting from my school, my under graduation teachers and my post graduation teachers for their motivation and their well wishes. My sincere thanks to my M.Sc guide Dr. S. Karuthapandian for his

words of encouragement and for showing me the path, that research could be so fascinating as well.

I would like to express my sincere thanks to all my lab mates past and present, Dr.Nishant, Dr. Pradeep, Dr.Ramesh, Dr.Jayashree, Dr.Shipra Agrawal, Dr.Arun, Dr.Swathi, Dr.Shritapa, Surbhi, Mishra, Nikhil, Hari, Vijay, Keerthi, Sathish, Manju, Bharath, Sainitin, Uma, Monalisa, Tushna, Priti, Vasudev, Nishitha and Muniraj for all their valuable help and making my stay in the lab and in Bangalore, the most enjoyable one. I will always remember and cherish the wonderful moments shared with them. Special thanks to Jayashree and Sainitin for their help in microarray analysis.

My batch mates Javaid, Vinay, Arif and Srikanth were wonderful buddies. My heartfelt thanks to all of them for the lighter moments that we shared during our course work time. Special thanks to Javaid and Vinay for the brain storming sessions that we have had in science and in general as well.

It was wonderful being a part of MBGU fraternity which was only possible because of all my past and present MBGU friends, Dr.Anand, Dr.Rahul, Dr. Kavitha Siva, Dr. Ashish Dr.V.Swaminathan, Dr. Prashanth, Dr. Krishnapal, Dr. Kiran Batta, Dr. Sonali, Dr. Chandrima, Rinki, Shubro, Vani, Venky, Mamta, Nishtha, Mangai, Bharath, Selvi, Jayasha, Abhishek, Mohan, Mukthi, Mahesh, Karthigeyan, Sujatha, Sanjeev, Parvathy, Varun. Thank you for all the help rendered.

My JNC friends past and present, Manu, Monojit, Sachi, Vini, Saswathi, Saikrishna, Anshuman, Arun, Srini, Shruthi, Snigdhadip, Dinesh, Shampa, Archana, Gopal, Vishwas, Dr Vivek chand Dr. Kabra, Dr. Dhritiman, Dr. Pavan, Dr. Sairam, and many others, were great mates. I would like to express my heartfelt thanks to all of them for making my 5 years of stay very pleasant in the campus.

JNCASR has offered me an extremely good ambience to perform my research peacefully and my stay enjoyable here in the "shrine of science" which would not be possible without the help of the support staff of Administration, Academic, Library, Complab and Mess.

My sincere thanks to my friends Suhasini and Niveditha, with whom, I shared all my high and low of life. My M.Sc friends were great buddies indeed. I would like to express my due acknowledgement to all of them for their well wishes and

prayers.

Arun, my husband and my father in law are the people I am blessed with, in my life. I thank my father in law for all the moral support he has rendered. I am extremely grateful to Arun, for all his unconditional love and support. His encouraging words and his sense of humor has often made my dull days bright. Thank you so much for every good thing that you have brought in my life.

Words fail to express my gratitude that I owe to my parents. They are such wonderful people, and I consider myself extremely lucky to have them in my life. They have given me all the freedom that I wanted. They showered me with love and affection at the same time groomed me as a human being. My sister Swathika, my best friend, I thank her for lending her patient ears and shoulder to lean on at the times of difficulty and for her constant support and well wishes.

Finally, I thank the Almighty, for, without his blessings this endeavor would not have been accomplished.

List of Figures

Abbreviations

APS	-	Ammonium Per Sulphate
ATP	-	Adenosine Tri Phosphate
BLAST	-	Basic Local Alignment Search Tool
CASIRNA	-	cis Acting Silencing RNA
cDNA	-	Complimentary DNA
DEPC	-	DiEthyl Pyro Carbonate
DMEM	-	Dulbecco Modified Eagle Medium
DNA	-	Deoxy ribonucleic Acid
DTT	-	Dithio thritol
ECM	-	Extra Cellular Matrix
EDTA	-	Ethylene Diamine Tetra Acetic acid
EMT	-	Epithelial Mesenchymal Transition
FISH	-	Fluoroscent *In Situ* Hybridization
GO	-	Gene Ontology
LNA	-	Locked Nucleic Acid
miRNA	-	Micro RNA
*mrhl*RNA	-	Meiotic Recombination Hotspot Loci Encoded RNA
mRNA	-	Messenger RNA
NATSRNA	-	Natural Antisense RNA
ncRNA	-	Non-Coding RNA
PAGE	-	Poly Acrylamide Gel Electrophoresis
PBS	-	Phosphate Buffered Saline
PCR	-	Polymerase Chain Reaction
PDGF	-	Platelet Derived Growth Factor
piRNA	-	Piwi interacting RNA

RISC	-	RNA Induced Silencing Complex
RITS	-	RNA Induced Transcriptional Silencing
RNA	-	Ribonucleic Acid
RNAi	-	RNA interference
RNP	-	Ribo Nucleo Protein
rRNA	-	Ribosomal RNA
RTQPCR	-	Reverse Transcription Quantitative Polymerase Chain Reaction
siRNA	-	Small Interfering RNA
snoRNA	-	Small Nucleolar RNA
snRNA	-	Small Nuclear RNA
tRNA	-	Transfer RNA

Contents

Chapter 1

Introduction

Regulation of gene expression is the fundamental aspect of biological phenomena such as responses to environmental conditions, development of multicellular organisms, cellular differentiation, morphology and disease. Gene regulatory patterns are extraordinarily diverse, yet the regulation of each gene is very precise both temporally as well as spatially. Gene regulation is also remarkably flexible to accommodate rapid changes in response to new conditions and evolutionary demands.

Molecular mechanisms underlying regulation of cellular functions were established over 50 years ago, where the central dogma of molecular biology was defined as follows. The genetic information that is stored as DNA is transcribed into a transient messenger RNA and it is decoded on ribosomes with the help of adapter tRNAs to produce a functional protein. Thus, RNA was not believed to play more than an accessory role and proteins were regarded as the ultimate functional entity of the cell. However this model is now challenged to a large extent due to the presence of huge repertoire of RNAs that do not code for proteins and display

remarkable properties more than what is known for mRNA, tRNAs or rRNAs [1].

1.1 Ribonucleic Acid (RNA)

RNA is a ubiquitous biopolymer [2]. It is involved in many aspects of maintenance, processing and transfer of genetic information. It can base pair specifically according to the standard Watson and Crick base pairing. RNAs can assume complex folded conformations that can participate in various cellular processes, independently (RNA) or bound to proteins (RNPs). Higher order structure for many RNAs is largely unknown and the mechanisms of action are poorly understood. Scientists have only recently started appreciating the remarkable structural and functional versatility of RNA. Despite containing only four major nucleotides along with other modified nucleotides, RNA can fold into variety of complex secondary and tertiary structures analogous to proteins [3]. Many fascinating discoveries from the recent past, and the identification of large number of new functional RNAs led to the hypothesis of primordial RNA world where both information and enzymatic functions are carried out by the RNA molecules [4, 5].

1.2 Classification of RNA

Cellular RNAs could be broadly classified into protein coding and protein non-coding RNAs. Messenger RNA qualifies as the only privileged member of protein coding RNAs. The protein non-coding RNAs are further classified into housekeeping RNA and regulatory RNA [6]. Transfer RNA, ribosomal RNA, snRNA, snoRNA, RNase P etc., are the members of the class of housekeeping RNAs that

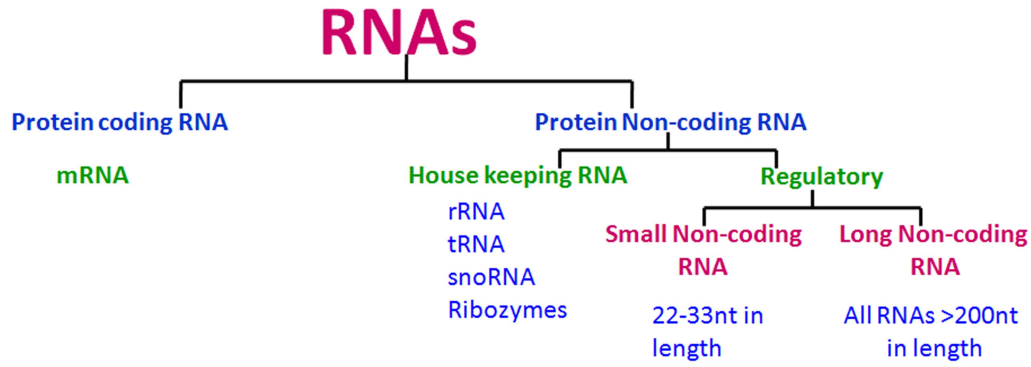

Figure 1.1: *Classification of cellular RNAs.*

function in day to day activities of the cell and are indispensible for cellular survival. The other class of protein non-coding RNA are the regulatory RNAs that function in fine tuning of genetic network and gene regulation. They are further sub classified into long non-coding RNAs and the short non-coding RNA. Long non-coding RNAs are more than 200 nt in length whereas the short RNAs are anywhere between 22-33 nt in length. The examples of each of these RNAs are briefly described in Figure 1.1, which summarizes the classification of the different cellular RNAs.

1.3 Messenger RNA

Messenger RNAs are the transient RNA, which converts the genetic blue print to a chemical blueprint to encode a protein. They are the exclusive members of protein coding RNAs and account for about 2% of total genomic transcription. They are formed in the cell nucleus by post transcriptional processing of primary transcripts. The primary transcripts, also known as heterogenous nuclear RNAs

are transcribed by RNA polymerase II which then undergo extensive post transcriptional modification to generate a mature mRNA that is transported to the cytoplasm and translated to proteins. The post transcriptional modifications include, 5' capping, addition of poly A tails in the 3' end of the RNA and splicing, a process by which the introns are removed from the coding sequence. The process of intron removal from the coding sequence itself involves the participation of ncRNAs such as snRNA which will be discussed shortly. Some mRNAs also undergo a process called RNA editing.

1.4 Non-Coding RNA

The term non-coding RNA (ncRNA) is commonly employed for RNA that does not encode a protein. Originally, the term ncRNAs referred exclusively to the transcripts that are poly adenylated and are transcribed by polymerase II carrying a 7 methyl Guanosine cap structure and lacking an ORF. Nowadays, this designation is extended to all the RNAs that are devoid of protein coding capacity [6]. The huge class of regulatory ncRNAs remained unnoticed untill recent past where scientists have uncovered surprisingly large numbers of regulatory ncRNAs. These results have fuelled speculation that ncRNAs might be important in understanding the increased complexity observed in mammals [7]. Figure 1.2 shows the genomic space for the discovery of novel ncRNA in higher eukaryotes. Estimated sizes of RNA fractions of representative bacterial or eukaryal genomes which are either protein coding or non protein coding are given as percentages of the total size of respective genome. From the Figure 1.2 it is obvious that as the complexity of

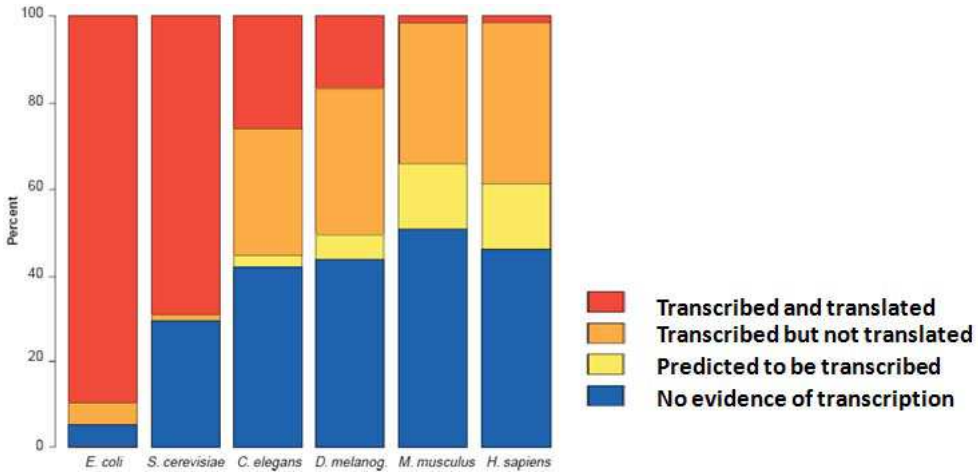

Figure 1.2: *Transciptional profile of lower to higher organisms showing percentage of coding Vs non-coding proportion of the genome. Adapted from Huttenhofer et al. [7].*

the organisms increases the percentage of coding vs. non-coding increases dramatically. In lower prokaryotes while 90% of the genome codes for proteins, the percentage of protein coding genes in higher eukaryotes drops down to as low as 2% while almost 60% of the genome is still transcribed. The recent attempt to identify the functional elements in the human genome the ENCODE project consortium has projected the transcription output of the human genome to be 98%, after analyzing 1% of the total genome [8].

Much of the recent attention has been devoted in understanding two large classes of regulatory RNAs such as long and short non-coding RNAs which will be discussed in detail later. The housekeeping RNAs which also contributes to its share of ncRNAs will be discussed briefly below with few examples of each class.

1.4.1 Housekeeping RNA

These are the huge class of nc RNAs that participate in the infrastructural roles of the cells [2]. They constitute more than 75% of total cellular transcripts with

rRNA and tRNA constituting up to 90 percent.

1.4.1.1 Ribosomal RNA

Ribosomes are the work benches of protein synthesis. It is a huge ribonucleoprotein complex that consists of four ribosomal RNAs in eukaryotes, the 28S, 18S, 5S and 5.8S. The 28S, 18S and 5S rRNA are the product of RNA polymerase I transcription in the nucleolus. The 5.8S rRNA is transcribed by RNA polymerase III. Originally, it was believed that the RNA fraction of the ribosomes are involved in the scaffolding process where it can orient the ribosomal proteins for the catalysis of peptide bond formation [9]. However, with the progression of time, the idea that ribosomal RNAs might participate directly in protein synthesis became entirely credible. The functional importance of rRNA in the peptide bond formation could be best appreciated looking at the active centre of the ribosomes which is purely made up of rRNA rather than the protein. This is well established in the recent past with bulk of information that was obtained with the different crystal structures of ribosomes with its substrate [10].The catalytic activity of rRNA is however astounding, which only further ensures a role of such non-coding RNA in varied cellular function.

1.4.1.2 Transfer RNA

tRNA is the information adapter molecule. It is the direct interface between amino-acid sequence of a protein and the information in mRNA. Therefore, it decodes the information in mRNA. There are $>$ 20 different tRNA molecules. All are between 75-95 nt in length. There are 4 arms and 3 loops, namely the acceptor, D, T pseudouridine C and anticodon arms, and D, T pseudouridine C

and anticodon loops. Sometimes tRNA molecules have an extra or variable loop. Transfer RNA is synthesized in two parts. The body of the tRNA is transcribed from a tRNA gene. The acceptor stem is the same for all tRNA molecules and is added after the body is synthesized. It is replaced often during lifetime of a tRNA molecule.

1.4.1.3 Small nuclear RNA (snRNA)

The removal of the most abundant class of introns requires five major spliceosomal small nuclear ribonucleoprotein particles (snRNPs; U1, U2, U4, U5 and U6) [11]. The spliceosomes recognize 5' and 3' end splice sites which are located at the intron exon boundaries [12]. Each snRNP consists of a uridylic acid-rich small nuclear RNA (U1, U2, U4, U5 and U6 snRNAs) that is post-transcriptionally modified [13]. The 2,2,7-trimethyl guanosine (m3G) capped U1, U2, U4 and U5 snRNAs (Sm snRNAs) contain an Sm site (RAU3-6GR, where R is a purine) flanked by stem-loops, which collectively constitute domain A [14]. The intranuclear trafficking of snRNPs has also been documented, in particular with respect to several organelles, such as Cajal-Bodies, nucleoli and nuclear speckles, which are currently thought to orchestrate several aspects of their maturation, assembly and storage [15].

The central role played by the spliceosomal RNAs in splicing has led to the hypothesis that, like the ribosome, the spliceosome is an RNA-centric enzyme and a relic from the RNA world. Recent structural studies have provided the first glimpses of the structural features of spliceosomal RNAs, and mutational analyses *in vivo* and *in vitro* have uncovered new functional roles for a catalytically essential domain. An emerging model for the active site of group II introns, a closely related

class of natural ribozymes, is likely to provide a wealth of insights on structure and function of the active site of the spliceosome [16].

Atleast two splieceosomal RNAs U2 and U6 have been shown to bind to RNA substrate containing the sequence of the intron branch site and promote a splicing related reaction in the absence of any of the numerous spliceosomal proteins. The reaction product is not the natural branch consisting of nucleotide forming both 2'-5' and 3'-5' phosphodiester bonds but instead a new product consistent with a phosphor triester [17].

1.4.1.4 Ribonuclease P

RNase P, found in all cells, catalyses site specific hydrolysis of precursor RNA substrates including tRNA and 5S RNA and signal recognition particle RNA [18, 19]. RNAse P is an RNA protein complex whose activity resides within the RNA component. Typically they are 300-400 nucleotides in length and comprises of two domains containing the substrate recognition site and the ribosome active site respectively [20]. Recently they have also been shown to play a role in maturation of intron encoded box C/D snoRNA as well [21].

1.4.1.5 Small Nucleolar RNA

The nucleolus is rich in snoRNAs, most of which are $\sim$ 70-250 nt in length. Some snoRNAs have roles in ribosomal RNA processing, but most of them function in rRNA modification [4]. On the basis of weak sequence similarities almost all snoRNA fall into two families: C/D box snoRNA, and the H/ACA snoRNAs [22, 23]. The C/D box snoRNAs use base complimentarity to guide site specific 2' O ribose methylation to rRNA, whereas the H/ACA snoRNAs use base

complementarity to guide site specific pseudouridylation to rRNA. Initially, the targets for snoRNA mediated modification appeared to be restricted to rRNA and their only known subcellular location was nucleolus. However, recently the range of targets has been extended to snRNA and tRNAs where the snoRNAs targeting snRNAs have been found to localize to the Cajal Bodies. In addition, there is evidence for a growing number of orphan snoRNAs which might function distinctly such as guiding modification of other cellular RNAs. Example in this context is the brain specific snoRNA which was proposed to target protein coding mRNA, the serotonin receptor 5-HT2c [24].

1.4.1.6 Telomerase RNA

Maintenance of telomeres by the enzyme telomerase is essential for genomic stability and cell viability in ciliates, vertebrates and yeast. The minimal components of telomerase required for catalytic activity are the telomerase reverse transcriptase (TERT) protein and the template-containing telomerase RNA (TER) [25]. TERs are highly divergent among different species, varying in both size and sequence composition, from 150 nt in ciliates and 450 nt in vertebrates to 930-1300 nt in the budding yeasts. In addition to the template, all TERs contain a 50bp template boundary element (TBE) and a large loop that includes the template, a potential pseudoknot and a loop-closing helical region. In vertebrates, this core domain is required for activity together with TERT, to reconstitute a minimal functional enzyme [26]. Telomerase has been the focus of intense study because of its aberrant up-regulation in the majority (90%) of cancer cell lines and its role in preventing chromosomal instability. Telomerase activity is low or undetectable in most somatic cells, and consequently telomeric DNA repeats are eroded over

successive rounds of DNA templated DNA synthesis, ultimately resulting in shortening of the telomeres below a critical length, which leads to telomere fusion and loss of cell viability. Mutations in TER RNA cause a rare inherited disorder known as Dyskeratosis Congenita (DC) and Aplastic Anemia (AA) [27, 28].

1.4.2 Regulatory Non-Coding RNA

Many of the non-coding RNAs that we have seen till now fulfilled much of the generic functions of the cell such as translation, splicing, RNA modification etc. However, regulatory role for non-coding RNAs had been proposed as early as 1960 where Jacob and Monod [29] who defined the basic principles underlying bacterial gene regulation. They clearly established two types of genes, namely the structural genes that produce the protein coding mRNA and the regulatory genes that might produce regulatory RNA. Though it was later proven that the regulatory gene at least in the case of Lac operon was protein, nevertheless, the concept that RNA could regulate gene expression was quite familiar at that time.

As mentioned previously regulatory non-coding RNA are further classified into two major classes, the small non-coding RNA and the large non-coding RNAs. Each of these families of regulatory RNA is explained fairly in detail in the following sections of this chapter.

1.4.3 Long non-coding RNA

Long non-coding RNAs are those that are greater than 200nt in length and lack an open reading frame greater than 100bp. Long non-coding RNAs were originally identified from large scale full length cDNA library sequencing from mouse. It is

now apparent that such a kind of long non-coding RNAs do exist in all species and is present in large numbers in higher eukaryotes. Given their unexpected abundance, long ncRNAs were initially thought to be spurious transcriptional noise resulting from low RNA polymerase fidelity [30]. However, the expression of many long ncRNAs is restricted to particular developmental contexts and large numbers of mouse ncRNAs are specifically expressed during embryonic stem cell differentiation [31] and in the brain, often exhibiting precise subcellular localization [32].The binding of transcription factors to non-coding loci, together with evidence of purifying selection acting on ncRNA promoters, suggests that this type of expression is explicitly regulated [33, 34].

Nevertheless, with many signatures of functionality, the non-coding RNAs for a very long time were considered as a product of spurious transcription, due to extremely poor conservation across species as compared to protein-coding sequence. One interpretation of this is that, considerably more nucleotide substitutions are deleterious in a protein-coding sequence compared with a non-coding sequence. This would not be too great a surprise given the stringent thermodynamic, structural and functional constraints on protein sequences compared to non-coding RNA sequence. However, this argument now barely holds true for the following reasons. First, it ignores many examples that are conserved, and a recent study ascribes functional roles to a high proportion of such ncRNAs [35]. Second, long ncRNAs are likely to exhibit different patterns of conservation in contrast to protein-coding genes, which are subject to strict functional constraints and must preserve an ORF. By contrast, long ncRNAs can exhibit shorter stretches of sequence that are conserved to maintain functional domains and structures. Indeed, many long ncRNAs

with a known function, such as Xist, only exhibit high conservation over short sections of their length [36]. Third, rather than being indicative of non-functionality, low sequence conservation can also be explained by high rates of primary sequence evolution of long ncRNAs, like promoters and other regulatory elements, that are more plastic in structure-function constraints than proteins [37]. Many conserved regions of the human genome that have been subject to recent and rapid evolutionary change are transcribed into long ncRNAs, including HAR1, a ncRNA expressed in Cajal-Retzius neurons in the developing neocortex [38]. Moreover, the adaptive radiation of non-coding (that is, regulatory) sequences is likely to specify most of the phenotypic differences between, and within, species [39]. Another explanation to the poor conservation of these RNAs is that the process rather than the product of transcription that is actually functional. An example of this case is the non-coding RNA that is transcribed from the fbp1+ promoter in yeast, which is associated with the progressive opening of the promoter and allows the accessibility of transcription factors and RNA polymeraseII [40].

A long non-coding RNA can be placed into one or more of five broad categories:(1) sense, or (2) antisense, when overlapping one or more exons of another transcript on the same, or opposite strand, respectively; (3) bidirectional, when the expression of it and a neighboring coding transcript on the opposite strand is initiated in close genomic proximity, (4) intronic, when it is derived wholly from within an intron of a second transcript (although these, as noted above, sometimes may represent pre-mRNA sequences), or (5) intergenic, when it lies within the genomic interval between two genes [41]. An example representing all these origins of the long non-coding RNA is illustrated from the Pax6 locus. Figure 1.3 shows the origin of long ncRNA from the locus, that spans several kbs and easily reflects

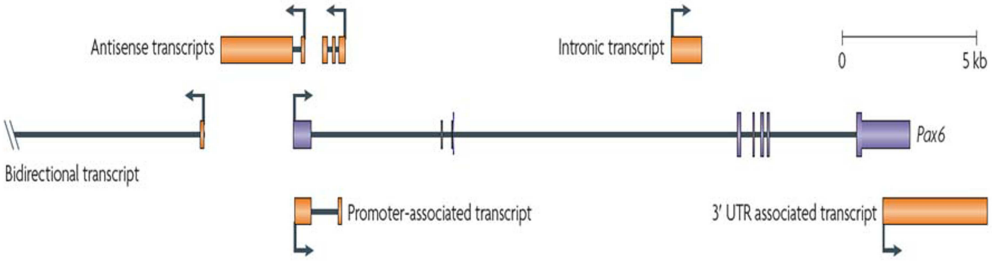

Figure 1.3: *Origins of non-coding RNA from Pax 6 locus. The Blue bars are the protein coding transcripts whereas the orange ones are the protein non-coding transcripts which might be involved in regulation of Pax 6 gene. Adapted from Mercer et al. [37].*

the over lapping expression of protein non-coding RNA with respect to its protein coding transcript of the locus [37].

The long ncRNAs, which are transcribed by RNA polymerase II, spliced and polyadenylated, like Xist, Tsix, Air, H19, Rox, are implicated in variety of regulatory processes, such as imprinting, X-chromosome inactivation, DNA methylation, transcription, RNA interference, chromatin-structure dynamics and antisense regulation [42,43]. In addition, long mRNA-like ncRNAs such as MALAT-1, BC-1 and BC-200 serve as prognostic markers for cancer, whilst the prion-associated RNAs LIT-1, SCA-8 etc. are implicated in a number of neurological disorders [44]. A recent addition to this group of non-coding RNA is the NAT (Natural Antisense Transcripts) that acts in cis and regulate sense /antisense transcripts. They are generally coexpressed with their targets and involved in competitive transcriptional interference [45]. Though the mechanism of many regulatory processes mediated by RNA of different lengths still remains a mystery, it is now becoming clear that RNA is an important cellular regulatory molecule having array of functions within the cell. Some of the process mediated by non-coding RNAs and well studied

examples in the class is explained in the later sections.

1.4.4 Small Non-coding RNA

Since the discovery in 1993 of the first small silencing RNA [46], an array of small RNA classes have been identified, including microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs). These classes differ in their biogenesis, their modes of target regulation and the biological pathways they regulate. There is a growing realization that, despite their differences, these distinct small RNA pathways are interconnected, and that small RNA pathways compete and collaborate as they regulate genes and protect the genome from external and internal threats. The defining features of small silencing RNAs are their short length ($\sim$ 20-30 nucleotides), and their association with members of the Argonaute family of proteins, which they guide to their regulatory targets, typically resulting in reduced expression of target genes. Beyond these defining features, different small RNA classes guide diverse and complex schemes of gene regulation. These different classes of regulatory RNAs also differ in the proteins required for their biogenesis, the constitution of the Argonaute-containing complexes that execute their regulatory functions, their modes of gene regulation and the biological functions in which they participate [47].

Although many classes of small RNAs have emerged, various aspects of their origins, structures, associated effector proteins, and biological roles have led to the general recognition of three main categories: short interfering RNAs (siRNAs), microRNAs (miRNAs), and piwi-interacting RNAs (piRNAs). These RNAs are only known to be present in eukaryotes, although the Argonaute proteins that

Name	Organism	Length (nt)	Proteins	Source of trigger	Function
miRNA	Plants, algae, animals, viruses, protists	20–25	Drosha (animals only) and Dicer	Pol II transcription (pri-miRNAs)	Regulation of mRNA stability, translation
casiRNA	Plants	24	DCL3	Transposons, repeats	Chromatin modification
tasiRNA	Plants	21	DCL4	miRNA-cleaved RNAs from the *TAS* loci	Post-transcriptional regulation
natsiRNA	Plants	22	DCL1	Bidirectional transcripts induced by stress	Regulation of stress-response genes
		24	DCL2		
		21	DCL1 and DCL2		
Exo-siRNA	Animals, fungi, protists	~21	Dicer	Transgenic, viral or other exogenous dsRNA	Post-transcriptional regulation, antiviral defense
	Plants	21 and 24			
Endo-siRNA	Plants, algae, animals, fungi, protists	~21	Dicer (except secondary siRNAs in *C. elegans*, which are products of RdRP transcription, and are therefore not technically siRNAs)	Structured loci, convergent and bidirectional transcription, mRNAs paired to antisense pseudogene transcripts	Post-transcriptional regulation of transcripts and transposons; transcriptional gene silencing
piRNA	Metazoans excluding *Trichoplax adhaerens*	24–30	Dicer-independent	Long, primary transcripts?	Transposon regulation, unknown functions
piRNA-like (soma)	*Drosophila melanogaster*	24–30	Dicer-independent	In *ago2* mutants in *Drosophila*	Unknown
21U-RNA piRNAs	*Caenorhabditis elegans*	21	Dicer-independent	Individual transcription of each piRNA?	Transposon regulation, unknown functions
26G RNA	*Caenorhabditis elegans*	26	RdRP?	Enriched in sperm	Unknown

Figure 1.4: *Different classes of small non-coding RNAs identified till date in different species and their functions. Adapted from Ghildiyal et al. [47].*

function in eukaryotic silencing can also be found in scattered bacterial and archaeal species. The detailed table of small RNAs from different eukaryotic species is enlisted in the Figure 1.4. The following sections will discuss in detail about three well studied class of small RNAs namely the miRNA, siRNA and piRNA and their biogenesis pathways and the protein complexes that modulate their activity.

1.4.4.1 Micro RNA

The first miRNA to be discovered, *lin4*, was identified in a screen for genes that are required for postembryonic development in *C. elegans*. The *lin4* locus produces a 22-nucleotide RNA that is partially complementary to sequences in the 3' UTR of its regulatory target, the *lin14* mRNA [46]. In 2001, tens of miRNAs

were identified in humans, flies and worms by small RNA cloning and sequencing, thereby establishing miRNAs as a new class of small silencing RNAs. miRBase (release 12.0), the registry that coordinates miRNA naming, now lists 1,638 distinct miRNA genes in plants and 6,930 in animals and their viruses [48]. Micro RNAs are the small 20-22 nt RNA that are involved in post transcriptional regulation of messenger RNAs. They are endogenous gene product and are involved in regulation and fine tuning of expression of genes involved in various cellular processes at mRNA level [49].

Biogenesis of miRNAs

miRNAs are derived from the precursor transcripts called primary miRNAs (pri-miRNAs), which are typically transcribed by RNA polymerase II (RNA Pol II), however there are polymerase III transcribed miRNAs as well which has been reported [50–52]. Several miRNA genes are present as clusters in the genome and probably derive from a common pri-miRNA transcript. Generating a 20-24-nucleotide miRNA from its pri-miRNA requires the sequential action of two RNase III endonucleases, assisted by their dsRNA-binding domain (dsRBD) partner proteins. First, the pri-miRNA is processed in the nucleus into a 60-70-nucleotide pre-miRNA by Drosha, acting with its dsRBD partner - DGCR8 in mammals and Pasha in flies. The resulting pre-miRNA has a hairpin structure: a loop flanked by base-paired arms that form a stem [53–55]. Pre-miRNAs have a two-nucleotide overhang at their 3' ends and a 5' phosphate group, which are indicative of their production by an RNase III. The nuclear export protein Exportin 5 carries the pre-miRNA to the cytoplasm bound to Ran, a GTPase that moves RNA and proteins through the nuclear pore [56].

In the cytoplasm, Dicer and its dsRBD partner protein, TRBP in mammals

and loQS in flies, cleaves the pre-miRNA, Dicer, like Argonaute proteins but unlike Drosha - contains a PAZ domain, presumably allowing it to bind the two-nucleotide 3'-overhanging end generated by Drosha [57]. Dicer cleavage generates a duplex containing two strands, termed miRNA and miRNA*, corresponding to the two sides of the base of the stem. These correspond to the guide and passenger strands and thermodynamic criteria influence the choice of miRNA versus miRNA*. miRNAs can arise from either arm of the pre-miRNA stem, and some pre-miRNAs produce mature miRNAs from both arms, whereas others show such pronounced asymmetry that the miRNA* is rarely detected even in high-throughput sequencing experiments [47]. In flies, worms and mammals, a few pre-miRNAs are produced by the nuclear pre-mRNA splicing pathway instead of processing by Drosha. These pre-miRNA-like introns, termed mirtrons, are spliced out of mRNA precursors. The spliced introns first accumulate as lariat products that require 2'-5' debranching by a lariat-debranching enzyme. Debranching yields an authentic pre-miRNA, which can then enter the standard miRNA biogenesis pathway [58, 59]. In plants, DCl1 fills the roles of both Drosha and Dicer, converting pri-miRNAs to miRNA-miRNA* duplexes. DCl1, assisted by its dsRBD partner Hyl1, converts pri-miRNAs to miRNA-miRNA* duplexes in the nucleus, after which the miRNA-miRNA* duplex is thought to be exported to the cytoplasm by HASTy, an Exportin 5 homologue. Unlike animal miRNAs, plant miRNAs are 2'-O-methylated at their 3' ends by HEN1. HEN1 protects plant miRNAs from 3' uridylation, which is thought to be a signal for degradation. HEN1 probably acts before miRNAs are loaded into AGO1, because both miRNA* and miRNA strands are modified in plants [60, 61]. Figure 1.5 summarizes the entire miRNA pathway.

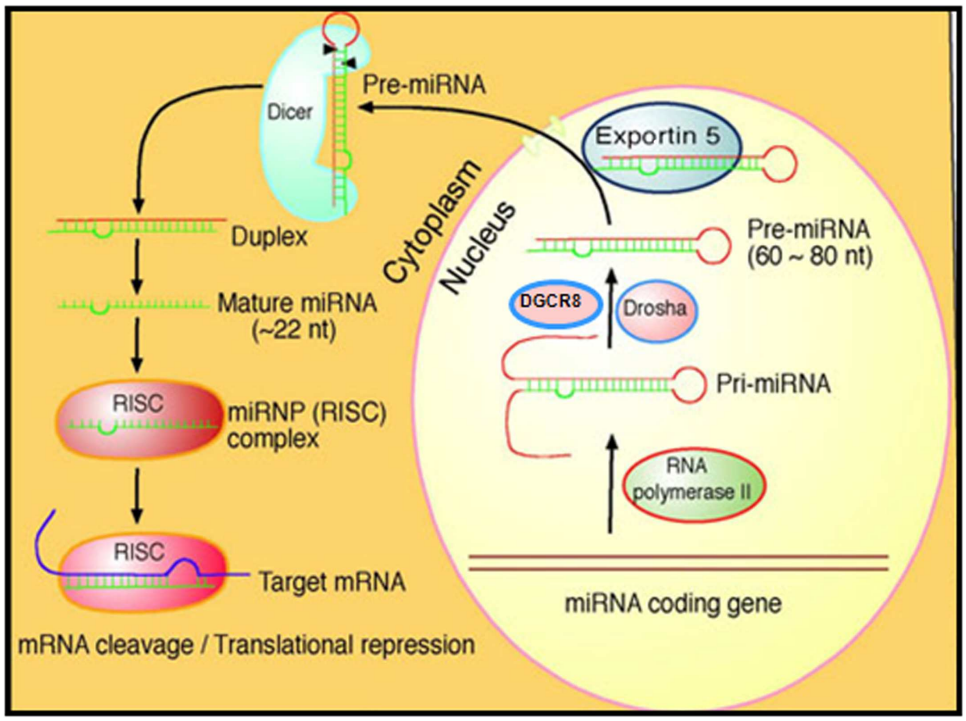

Figure 1.5: *Pathway of miRNA biogenesis.*

Target regulation by miRNAs

The mechanism by which a miRNA regulates its mRNA target reflects both the specific Argonaute protein into which the small RNA is loaded and the extent of complementarity between the miRNA and the mRNA [62, 63]. A few miRNAs in flies and mammals are nearly fully complementary to their mRNA targets which directs endonucleolytic cleavage of the mRNA [64]. Such extensive complementarity is considered the norm in plants, as target cleavage was thought to be the main mode of target regulation in plants. However, in flies and mammals, most miRNAs pair with their targets through only a limited region of sequence at the 5' end of the miRNA called the 'seed region'; these miRNAs repress translation of their mRNA targets. The seed region of all small silencing RNAs contributes most of

the energy for target binding [65]. Thus, the seed is the primary specificity determinant for target selection. The small size of the seed means that a single miRNA can regulate many, even hundreds, of different genes [66]. Intriguingly, recent data suggest that the nuclear transcriptional history of an mRNA influences whether a miRNA represses its translation at the initiation or the elongation step [67].

Major Players of miRNA pathway

Cleavage of dsRNA by enzymes of the RNaseIII family, including Drosha and Dicer, yields products with characteristic termini, with a monophosphate group at the 5' ends, and a two-nucleotide overhang at the 3' ends. These enzymes consist of an RNaseIII domain, which has the catalytic activity, and (generally) a dsRNA-binding domain (dsRBD), and they function as homodimers. The two RNaseIII domains of the dimer associate to form a single processing centre, with each catalytic domain responsible for the hydrolysis of one strand in the duplex. By contrast, both Drosha and Dicer are monomeric and contain two tandemly arranged RNaseIII domains and a single dsRBD [68].

The endonuclease Dicer processes dsRNA substrates (long dsRNAs and pre-miRNAs) into short dsRNA fragments (siRNAs and miRNAs) of defined length, typically 21-25 nucleotides [69]. In addition to two copies of the conserved RNaseIII domain and a dsRBD in the carboxyl terminus, Dicer enzymes usually have an amino-terminal DEXD/H-box domain, followed by a small domain of unknown function (the DUF283 domain) and a PAZ domain. The PAZ domain, which is also found in Argonaute proteins, binds specifically to the 3' end of single-stranded RNA (ssRNA) [70, 71].

The RNaseIII-family member Drosha catalyses the initial processing of pri-miRNAs, yielding pre-miRNAs, which are hairpins with phosphorylated 5' ends

and 3' dinucleotide overhangs [51]. Drosha is a nuclear protein, and its domain structure consists of a proline-rich region and an arginine- and serine-rich region at the N terminus, followed by two RNaseIII domains and a dsRBD. Purified Drosha cleaves dsRNA nonspecifically; specific cleavage of pri-miRNAs requires association with a protein known as DGCR8 (also known as PASHA in invertebrates) in a complex called the microprocessor. DGCR8 binds to the base of the pri-miRNA hairpin, positioning Drosha to cleave the pri-miRNA stem at a distance of 11 base pairs from the junction between the duplex stem and the flanking ssRNA regions. Thus, DGCR8 seems to be a trans-acting specificity determinant, analogous to the PAZ domain of Dicer, which acts in cis [72].

The common feature of RNAi and all related small-RNA-mediated silencing pathways is the association of a small silencing RNA with a protein of the Argonaute family [73]. The resultant protein-RNA complex forms the minimal core of the effector complex known as the RISC. Within the RISC, the small RNA functions as a sequence-specific guide that recruits an Argonaute protein to complementary target transcripts through base-pairing interactions. The target transcripts, typically mRNAs, are then either cleaved or prevented from being translated by ribosomes, leading to their degradation. Argonaute proteins are multidomain proteins that contain an N-terminal domain, and PAZ, middle (MID) and PIWI domains. Crystal structures of prokaryotic Argonaute proteins have revealed a bilobate architecture, with the MID and PIWI domains forming one lobe, and the N-terminal and PAZ domains constituting the other. PIWI domain of an Argonaute protein adopts an RNaseH fold and suggests that Argonaute proteins are responsible for the 'slicer' activity of the RISC [74]. Similarly to the requirements for RNaseH activity, RISC-catalysed RNA cleavage requires divalent metal ions

and yields a 5' product, which has a free 3' hydroxyl group, and a 3' product, which carries a 5' phosphate group [68].

Like transcription factors, miRNAs regulate diverse cellular pathways and are widely believed to regulate most biological processes in plants and animals, ranging from housekeeping functions to responses to environmental stress, development and disease. Some of the well understood functions of few animal miRNA are described in the section on functions of non-coding RNA.

1.4.4.2 Small interfering RNA (siRNA)

The discovery of RNAi in 1998 by Fire and Mellow established dsRNA as a silencing trigger for genes in *C. elegans* [75]. In worms and other animals, siRNA-mediated silencing is termed as RNAi. Remarkably, RNAi is systemic in both plants and nematodes, spreading from cell to cell. In *C. elegans*, RNAi is also heritable: silencing can be transferred to the progeny of the worm that was originally injected with the trigger dsRNA. Viral infection, inverted-repeat transgenes or aberrant transcription products all lead to the production of dsRNA [76]. dsRNA is converted to siRNAs that elicit RNAi. siRNAs were discovered in plants and were later shown in animal extracts to serve as guides that direct endonucleolytic cleavage of their target RNAs. siRNAs can be classified according to the proteins involved in their biogenesis, their mode of regulation or their size, to siRNAs induced by exogenous agents and the endo siRNA [47]. The siRNA biogenesis is schematically represented in Figure 1.6.

siRNAs induced by exogenous agents

Early examples of RNAi were triggered by exogenous dsRNA. In these cases,

long exogenous dsRNA is cleaved into double stranded siRNAs by Dicer, a dsRNA-specific RNase III family ribonuclease [77]. siRNA duplexes produced by Dicer comprise two ~ 21 nucleotide strands, each bearing a 5' phosphate and 3' hydroxyl group, paired in a way that leaves two-nucleotide overhangs at the 3' ends. The strand that directs silencing is called the guide, whereas the other strand, which is ultimately destroyed, is the passenger [78, 79]. Target regulation by siRNAs is mediated by the RNA induced silencing complex (RISC), which is the generic name for an Argonaute-small RNA complex [80]. In addition to an Argonaute protein and a small RNA guide, the RISC might also contain auxiliary proteins that extend or modify its function. For example, some proteins that redirect the target mRNA to a site of general mRNA degradation such as P bodies [81]. Mammals and *C. elegans* each have a single Dicer that makes both microRNAs (miRNAs) and siRNAs, whereas *Drosophila* species have two Dicers: DCR-1 makes miRNAs, whereas DCR-2 is specialized for siRNA production. In *Drosophila*, siRNA pathway is activated upon viral infection, where viral RNAs are targeted and eliminated [51]. R2D2 is a dicer partner that senses the thermodynamic stability between the passenger and the guide strand before loading to RISC and gives preferential signal to Ago2 to cleave the passenger strand [82]. The mature RISC which now contains single stranded guide RNA and is 2'O methylated in its 3' end and this mature RISC targets complementary mRNA for degradation.

Endo-siRNAs

The first endo-siRNAs were detected in plants and *C. elegans* and the recent discovery of endo-siRNAs in flies and mammals suggests that endo-siRNAs are ubiquitous among higher eukaryotes [76]. In plants there are different classes of endo siRNAs, the CASi RNA are the Cis Acting siRNAs that orginate from

the transposons, repeat elements and rDNA repeats [83]. They are involved in heterochromatin formation and DNA silencing by methylation and whereas the TASi RNA that are Trans Acting siRNAs, are generated through miRNA or siRNA pathway where an RNA dependent RNA polymerase enzyme R2D2 copies and amplifies the siRNAs or miRNAs that are generated by the specific pathway [84]. Another class of endo siRNAs in plants are the Natural antisense transcript-derived siRNAs (natsiRNAs) that are produced in response to stress in plants. They are generated from a pair of convergently transcribed RNAs: typically, one transcript is expressed constitutively, whereas the complementary RNA is transcribed only when the plant is subject to an environmental stress [85].

Since most of the endo siRNA generation atleast in plants and C.elegens is dependent on RDRP, it was believed that such RNAs do not exist in mammals as the enzyme is absent. However, the first mammalian endo-siRNAs to be reported corresponded to the long interspersed nuclear element retrotransposon (L1 repeats) and were detected in cultured human cells [86]. Full-length L1 contains both sense and antisense promoters in its 5' UTR that could, in principle, drive bidirectional transcription, producing overlapping complementary transcripts to be processed into siRNAs by Dicer. However, the precise mechanism by which transposons trigger siRNA production in mammals remains unknown.

Fly endo-siRNAs are derived from transposons, heterochromatic sequences, intergenic regions, long RNA transcripts with extensive structures and they require both dcr2 and ago2 and are involved in transposon and repeat elements silencing in Drosophila [87]. Some of the endo siRNA that are derived from structured loci require additional protein factors such as loquacious. Endo-siRNAs have also been identified in mouse oocytes. As in flies, mouse endo siRNAs are 21 nucleotides,

Dicer-dependent and derived from a variety of genomic sources. The mouse endo siRNAs are bound to AGo2, the only mammalian Argonaute protein thought to mediate target cleavage. However the exact process that triggers the production and their function is unknown at present [88, 89].

1.4.4.3 PIWI interacting RNAs or PiRNA

piRNAs are one of the recently discovered class of small RNAs and, as their name suggests, they bind to the Piwi clade of Argonaute proteins [90,91]. The Piwi clade comprises Piwi, Aubergine (AuB) and AGo3 in flies, MIlI, MIwI and MIwI2in mice and HIlI, HIwI1, HIwI2 and HIwI3 in humans. piRNAs are long 27-35 nt RNA and was initially identified in Drosophila, where it was involved in silencing of repeat elements. It was then named as RASi RNAs or the repeat associated silencing RNAs. Now these are renamed as piRNAs due to their association with PIWI proteins rather than dcr1 or dcr2 [90].

Mammalian piRNAs can be divided into pre-pachytene and pachytene piRNAs, according to the stage of meiosis at which they are expressed in developing spermatocytes [92]. Like piRNAs in flies, pre-pachytene piRNAs predominantly correspond to repetitive sequences and are implicated in silencing transposons, such as L1 and intracisternal A-particle (IAP). In male mice, gametic methylation patterns are established when germ cells arrest their cell cycle 14.5 days post-coitum, resuming cell division 2-3 days after birth [93, 94]. Both MIlI and MIwI2 are expressed during this period, and Miwi2- and Mili-deficient mice lose DNA methylation marks on transposons. The pre-pachytene piRNAs, which bind MIwI2 and MIlI, might serve as guides to direct DNA methylation of transposons. In contrast to pre-pachytene piRNAs, the pachytene piRNAs mainly arise from

unannotated regions of the genome, not transposons, and their function remains unknown.

The current model for piRNA biogenesis was inferred from the sequences of piRNAs that are bound to Piwi, AuB and AGO3. piRNAs bound to Piwi and AuB are typically antisense to transposon mRNAs, whereas AGO3 is loaded with piRNAs corresponding to the transposon mRNAs themselves. Moreover, the first 10 nucleotides of antisense piRNAs are frequently complementary to the sense piRNAs found in AGO3. This unexpected sequence complementarity has been proposed to reflect a feed-forward amplification mechanism - 'piRNA ping-pong' - that is activated only after transcription of transposon mRNA [95]. The mechanism of piRNA biogenesis is given in Figure 1.6.

1.5 Functions of non-coding RNA

Though the mechanism of many regulatory processes mediated by RNA of different lengths still remains a mystery, it is now becoming clear that RNA is an important cellular regulatory molecule having array of functions within the cell. Some of the process mediated by non-coding RNAs and well studied examples in the class is explained briefly below.

1.5.1 Chromatin Modification

The DNA and proteins are closely intertwined to assume a compact structure within the nucleus, known as the chromatin. Nucleosome is the fundamental repeating subunit of chromatin which is made up of histones upon which the DNA is wound. The chromatin by itself is a very dynamic entity and the mere presence of

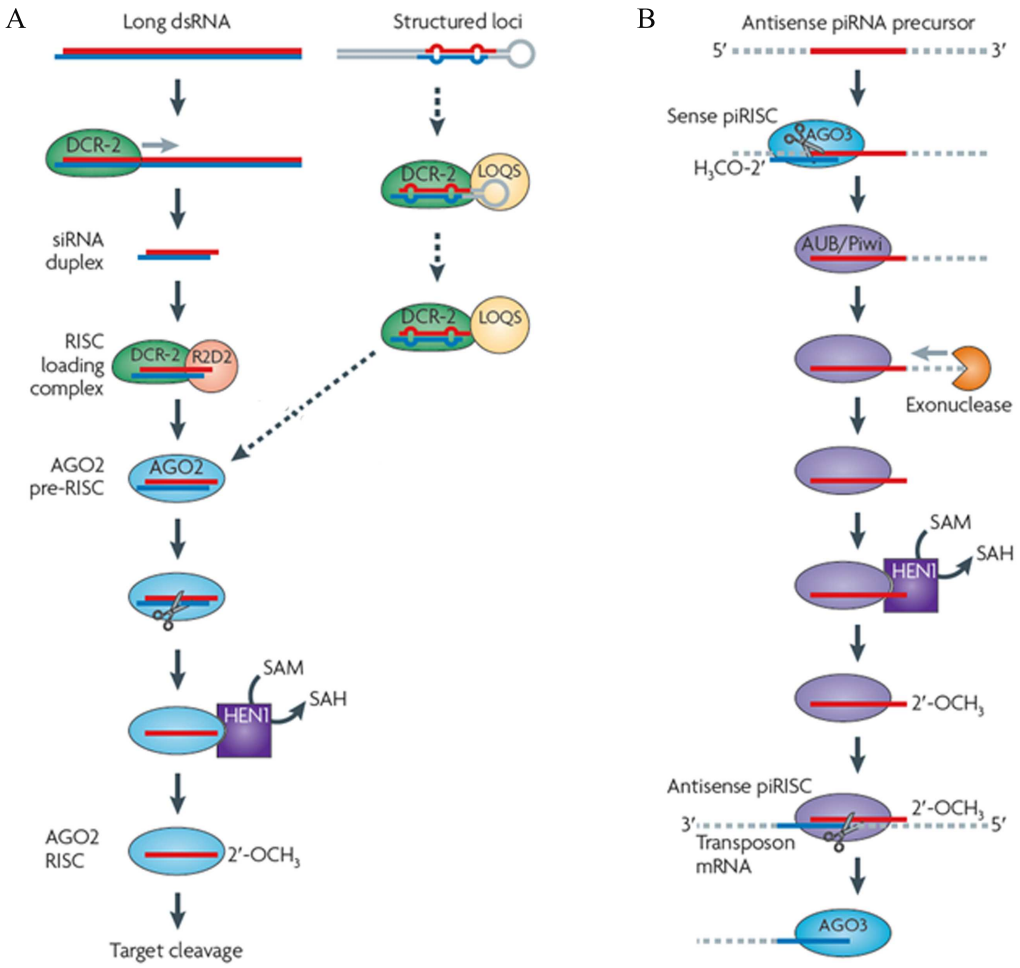

Figure 1.6: *Small RNA biogenesis pathways. (A) siRNA biogenesis pathway. (B) Ping pong mechanism of piRNA biogenesis. Adapted from Ghildiyal et al. [47].*

different states chromatin such as heterochromatin and euchromatin adds further to the complexity of higher order chromatin and thereby to the organism itself [42]. The role of RNA in these chromatin changes had been elusive for a quite some time until recently when scientists have really started appreciating the involvement RNA in various epigenetic states of chromatin. This involves various posttranslational modifications of histones and also deposition of different variant histones on chromatin. Long ncRNAs can also mediate epigenetic changes by recruiting chromatin remodeling complexes to specific genomic loci [37]. The chromatin associated RNA are primarily involved in processes such as genomic imprinting and dosage compensation, although processes where in the ncRNAs act in trans to establish long range chromatin interaction has also been documented.

Dosage compensation

In *Drosophila* dosage compensation is accomplished by hyper transcription in male. The transcription from the single male X chromosome in *Drosophila* is doubled in order to ensure equal levels of X linked genes in males and females [96]. A ribonucleoprotein complex is involved in this process which is known as DCC or the Dosage Compensation Complex, that consists of atleast 5 core proteins generally known as MSL (Male Specific Lethal) proteins: MSL1, MSL2, MSL3, MLE (maleless, an RNA helicase), MOF (chromodomain containing histone acetyl transferase). Two RNAs namely RoX1, and RoX2 (RNA on X chromosome) plays an important role in dosage compensation and essential components of DCC [97]. The expression of these RNAs is male specific and they are functionally redundant. The Rox RNAs associates with the Dosage Compensation Complex (DCC) and is essential for targeting the DCC to the X chromosome and brings about chromatin

modification by hyperacetylation of the H4 Lys 16 by MOF [98].

The epitome of dosage compensation process is seen in mammals, where the dosage compensation between XX females and XY males is achieved by inactivation of one of the two X chromosomes during early female embryogenesis. The role of non protein coding RNA in the process of X chromosome inactivation (XCI) was well established with the discovery of 17.4 kb functional Xist and 40 kb Tsix antisense transcript in mouse [99, 100]. Mammalian X chromosome inactivation is controlled by multilayered silencing pathway involving both short and long non-coding RNA which differentially recruit the epigenetic machinery to establish chromatin asymmetries. In pre XCI ES cells Xist and TSix the non-coding sense and antisense transcript pair are bilaterally transcribed [101]. During ES cell differentiation the X chromosome inactivation is initiated by upregulated expression of Xist on the future inactive X chromosome and Tsix becomes restricted to the future active X chromosome [102, 103]. The high level transcription of Xist on future Xi and coating of the transcript along the length of chromosome leads to the adoption of heterochromatic configuration devoid of any RNA polymerase II [104]. Xist also recruits heterochromatin factors such as PRC2 complex on the Xi [105]. Apart from the long RNA transcripts small RNA are also formed from the XCI locus known as XiRNA for their X inactivation centre origin. These RNAs range from 25 to 35 nt in length and are believed to be formed due to the formation of Xist and Tsix duplex [106]. A recent investigation by Lee and colleagues suggests that the transient heterochromatic configuration of Xist promoter on the future Xi is mediated by the targeted recruitment of the PRC2 complex by 1.6 kb transcript Rep A. This RNA is encoded from an internal promoter of Xist transcript sense strand and covers the A rich repeat region [107]. It is now very obvious that the

X chromosome inactivation itself is brought out by the interplay of many different species of non-coding RNA molecules that originate from the X inactivation centre [99]. All of these RNA dependent pathways however converge on recruitment of chromatin remodeling factors to establish long range repression of genes that are present in the Xi either by marking with repressive histone modification or by replacement of or with histone variants such as macro H2A [108].

Genome imprinting

Another chromatin-associated phenomenon that closely resembles the dosage compensation is the genomic imprinting. It is a process by which modification of one of the two parental alleles of a gene results in preferential silencing of the allele from one parent [109]. In several cases it has been demonstrated that the activity of non-coding RNA genes is essential. H19 is a non-coding RNA that represses maternally derived allele of Igf2 [110, 111]. Similarly, IPW (Imprinted in Prader willi syndrome) is a RNA that regulates transcription in cis in an imprinted region associated with Prader willi syndrome [112]. Another transcript AIR (Antisense to Igf2 receptor) is an 108 kb non-coding RNA that is responsible for imprinting of Igf2R in mouse [113]. Recently genomic imprinting in the Kcnq1ot1 locus has been demonstrated [114–116]. Recent emerging evidence indicates that long ncRNAs such as Kcnq1ot1 and Air which map to the Kcnq1 and Igf2r imprinted gene clusters, respectively, mediate the transcriptional silencing of multiple genes by interacting with chromatin and recruiting the chromatin modifying machinery [99]. There are about 500 genes in mammals that are believed to be imprinted and around 50% of them are silenced by RNA mediated process. Loss of these imprinting associated RNA have also been found in many of the genetic disorders

like Prader willi, Angelman syndromes etc ([117].

Chromatin remodeling

Long ncRNAs can also mediate epigenetic changes by recruiting chromatin remodelling complexes to specific genomic loci. For example, hundreds of long ncRNAs are sequentially expressed along the temporal and spatial developmental axes of the human homeobox (Hox) loci, where they define chromatin domains of differential histone methylation and RNA polymerase accessibility [118]. One of these ncRNAs, Hox transcript antisense RNA (HOTAIR), originates from the HOXC locus and silences transcription across 40 kb of the HOXD locus in trans by inducing a repressive chromatin state, which is proposed to occur by recruitment of the Polycomb chromatin remodelling complex PRC2 by HOTAIR 21 [37, 118].

RNA mediated heterochromatin formation suggests that small RNA could bring about massive down regulation of target genes. The phenomenon of RNAi for the formation of heterochromatin was first studied in the fission yeast centeromeric silencing [119]. RNAi like mechanisms are now known to play a critical role in mediating heterochromatic gene silencing and prevent mobilization of repetitive transposable elements [120,121] in many other organisms as well. A number of protein components of RITS (RNA Induced Transcriptional Silencing Complex) have been purified including their RNA counter parts that includes Argonaute, Dicer and RDRP a novel RNA Dependent RNA polymerase and helicases that collectively bring about the gene silencing at the chromatin level [122].

In some lower organisms such as ciliated protozoans *Tetrahymena* and *Paramecium* RNAi have shown to play a role in a process called DNA elimination. These organisms rearrange their genome during their sexual phase of development [123].

This remarkable downsizing of the genome has been shown to be brought about by RNAi like mechanism [124].

1.5.2 Transcriptional silencing

Non-coding RNAs have been shown to bring about transcriptional gene silencing by several mechanisms. Both long and short non-coding RNAs have been implicated in the process of transcriptional gene silencing.

SiRNAs targeted against promoter have been shown to bring about transcriptional gene silencing. Similarly, the endo siRNA are also involved in many cases of silencing of repeats and transcription of transposable genetic elements as in the case of piRNAs. Transcriptional gene silencing has been very well documented in *Drosophila*, where HP1 and HP2 proteins are recruited by RNAi mediated process, and causes H3K9 methylation pattern which silences the active genes and hetrochromatin marks are acquired [125]. Apart from the generalized roles of these small RNA in transcriptional silencing, some long nc RNA play some role in transcriptional gene silencing by targeting specific gene locus.

Proximal promoters can be transcribed into long ncRNAs that recruit and integrate the functions of RNA binding proteins into the transcriptional programme, as exemplified by the repression of cyclin D1 transcription in human cells [126,127]. Long ncRNAs also act as co-factors to modulate transcription factor activity. For example, in mice, the ncRNA Evf2 is transcribed from an ultraconserved distal enhancer and recruits the binding and action of the transcription factor DlX2 to this same enhancer to induce expression of adjacent protein-coding genes [128]. Some non-coding RNA such as the one that is transcribed from DHFR promoter forms

tritiplex in the promoter region thereby regulating its gene expression [129, 130]. Another example of global gene regulation by non-coding RNA is shown by the pol III transcribed Alu RNA that binds to pol II and regulates pol II dependent transcription [131].

1.5.3 Post transcriptional gene regulation

MicroRNAs and siRNA are classical small RNA that are involved extensively in post transcriptional gene regulation by binding to their cognate mRNAs and target it to repression or degradation. Many microRNAs have been identified in the recent past which have shown to regulate mRNA that are involved in development, differentiation, cell cycle, ES cell differentiation, cancer etc. Apart from miRNA and siRNA long nc RNA such as Zeb2 antisense transcript has been shown to play a role in translational regulation through retainment of intron which is efficiently translated dependent on IRES mediated mechanisms [132, 133]. Figure 1.7 gives an outline of different modes of gene regulation by non-coding RNA.

1.5.4 Regulation by localisation

Many non-coding RNAs are important constituents of subcellular organelles and they are localized to specific cellular compartments thereby regulating the cellular activity. In *Drosophila* hsrω is a non-coding RNA which is upregulated upon stress condition such as heat shock. The hsrω gene has been demonstrated to be crucial for proper development and viability of flies [134]. The long hsrω transcripts are localized to specific subnuclear compartments, ω speckles, with various hnRNP proteins [135, 136]. The ω speckles have been suggested to be the storage site of

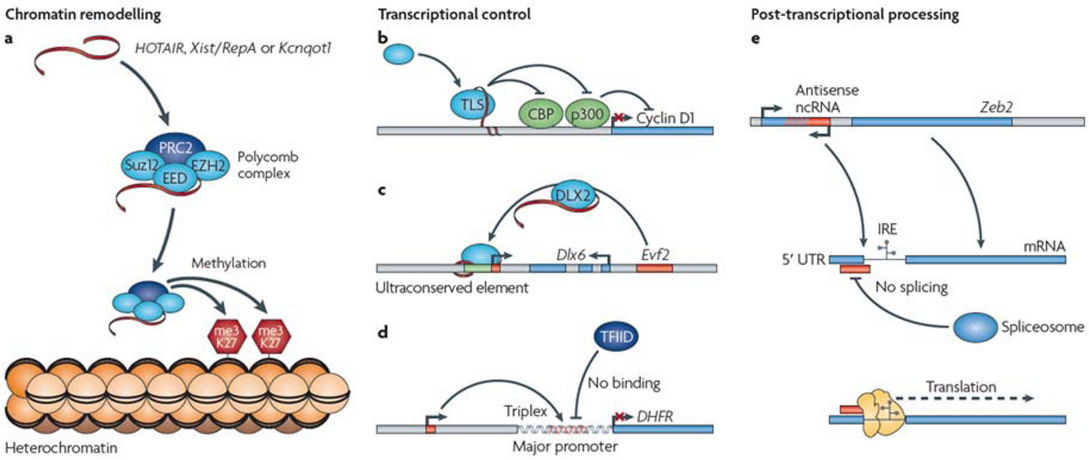

Figure 1.7: *Different modes of target gene regulation by non-coding RNA. Few examples of each type such as chromatin remodeling, transcriptional and post transcriptional silencing are represented. Adapted from Mercer et al. [37].*

hnRNPs, and mutant larval cells lacking functional hsrw transcripts do not form ω speckles, resulting in a diffuse nuclear distribution of the hnRNPs [135]. It has been suggested that the hsrω RNA plays the role of an organizer molecule by regulating the intranuclear trafficking and availability of hnRNPs [134, 135]. Similar to hsrω, non-coding RNAs are also formed from human sat III repeats of chromosome 9 that gets accumulated as large foci as nuclear stress bodies consisting of many RNA binding proteins as well [137, 138]. Neat is a non-coding RNA that is enriched in nuclear speckles. NEAT1 RNA, a highly abundant 4 kb ncRNA, is retained in nuclei in approximately 10 to 20 large foci that are shown to be completely coincident with paraspeckles, the nuclear domains implicated in mRNA nuclear retention. Depletion of NEAT1 RNA via RNAi eradicates paraspeckles, indicating the importance of this RNA in the formation of subnuclear architecture [139, 140]. CTN, another RNA expressed in humans that regulates its protein coding counterpart mCAT-2 mRNA by nuclear retention. CTN is 8 kb RNA and is

enriched in nuclear speckles and believed to regulate its protein coding counterpart through its A to I editing and association with speckle proteins [43, 141].

1.5.5 Disease manifestations

Apart from many functional roles that have been attributed to non-coding RNAs, some of the ncRNAs have also been shown to play a role in disease manifestations. BC1 (150 nt) and BC200 (200 nt) ncRNAs were identified as two cytoplasmic ncRNA transcripts expressed in the mouse and human nervous system, respectively [142, 143]. Interestingly, BC1 RNA is specifically targeted to dendritic domains in neurons. BC1 RNA has been found to interact with the Fragile X mental retardation protein (FMRP) [144]. FMRP is an RNA-binding protein, and mutations associated with the absence of FMRP or altered expression of FMRP lead to fragile X syndrome [145, 146]. Another RNA namely MALAT1 was found to be upregulated during the metastasis [147, 148].

Defects in normal cell processes such as differentiation, proliferation, and apoptosis are all well-known to be involved in cancer pathogenesis. One of the reasons for the connection between miRNA and cancer was initially made because miRNAs were found to be involved in many of these processes [149] This connection initiated research which further reinforced the correlation between miRNAs and cancer development. Researchers discovered that there is aberrant miRNA expression when comparing various types of cancer with normal tissues [150]. The first direct evidence for the key role of miRNAs in cancer came from a study in human chronic lymphocytic leukemia (CLL). Two miRNAs, miR-15 and miR-16 were attributed to disease pathogenesis. Additional investigation of these two miRNAs

revealed that they both showed a significant reduction in expression when compared to their normal tissue counterparts [151] during CLL. Some miRNAs are thought to have oncogenic activity while others have tumor suppressor activity and these are known as oncomiRS [152]. It is important to note that these distinctions may not be so strict and that some miRNAs may express either activity, depending on the situation and tissue type. Recently microRNA expression profiling showed difference in expression in different grades of Glioma allowing them to act as prognostic markers for early detection [153].

1.6 RNA Binding Proteins

RNAs in cells are associated with RNA-binding proteins (RBPs) to form ribonucleoprotein (RNP) complexes. The RBPs influence the structure and interactions of the RNAs and play critical roles in their biogenesis, stability, function, transport cellular localization and regulation [154]. Ribonucleoproteins constitute a huge class of cellular proteins. They possess distinct domains that participate in variety of cellular processes bound to different classes of RNA and proteins. The dynamic association of these proteins with RNA defines the lifetime, cellular localization, processing and the function of the RNA. Although all RBPs bind RNA, they do so with different RNA-sequence specificities and affinities. This activity is mediated by a relatively small number of RNA-binding scaffolds whose properties are further modulated by auxiliary domains. The auxiliary domains can also mediate the interactions of the RBP with other proteins and, in many cases, are subject to regulation by post-translational modification [155, 156]. Nearly 2% and 5-7% of coding genes of yeast and Drosophila respectively encodes for the RNA binding

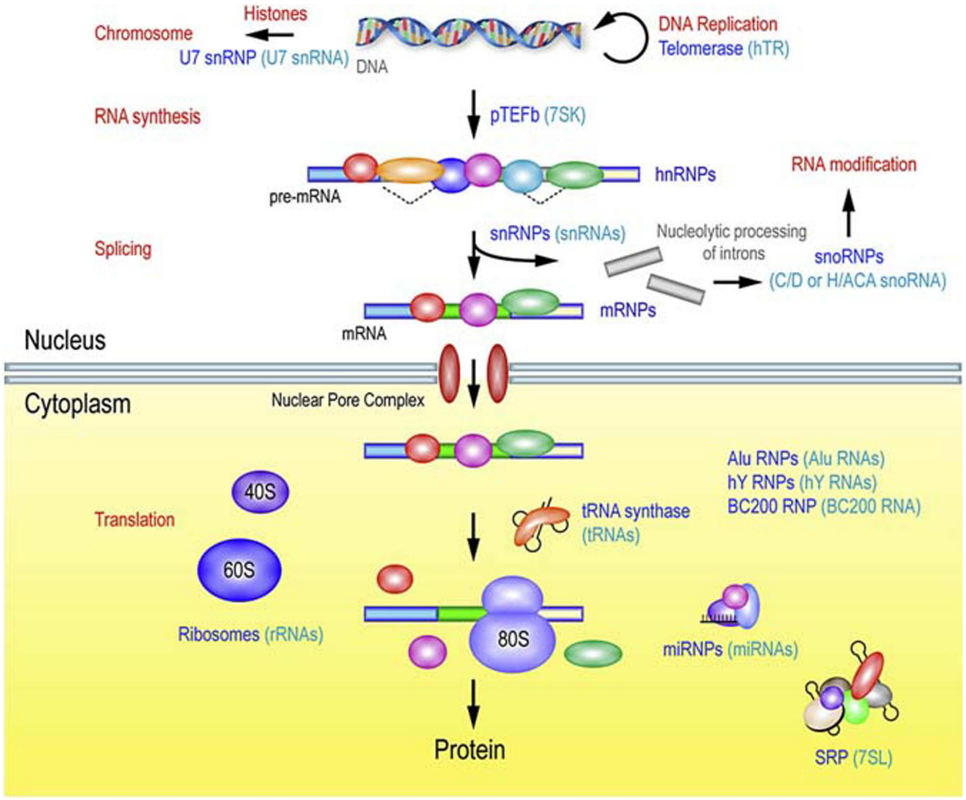

Figure 1.8: *Illustration of functional roles of RNA binding proteins with its active RNA components. These are shown to be functioning in virtually all aspects of cellular physiology both in the nucleus and cytoplasm. Adapted from Glisovic et al. [156].*

proteins [157–159]. Figure 1.8 illustrates the different functional roles participated by the RNA binding proteins along with its active RNA component.

Many of the RNA binding proteins are modular in nature which confers to its unique substrate specificity and increased propensity to different RNA binding [154]. There are different families of RNA binding proteins, some well-characterized RNA-binding domains include the following: RNA-binding domain (RBD, also known as RNP domain and RNA recognition motif, RRM); K-homology (KH) domain (type I and type II); RGG (Arg-Gly-Gly) box; Sm domain; DEAD/DEAH box; zinc finger (ZnFdouble stranded RNA-bindingdomain (dsRBD); cold-shock

domain; Pumilio/FBF (PUF or Pum-HD) domain; and the Piwi/Argonaute/Zwille (PAZ) domain [154,156]. Some of the proteins harbouring dsRBD such as the RNaseIII enzymes and the Piwi/argonaute/PAZ domains in the RNAi processes have been described earlier. The next few sections describe in detail about one such family of RNA binding proteins known as DEAD box helicase.

1.6.1 DEAD Box Helicases

The DExD/H box family of proteins is distinguished by the presence of several conserved motifs, which include the characteristic 'DExD/H' sequence (where x can be any amino acid) and are highly conserved in proteins from viruses and bacteria to humans [160]. The DExD/H box family of proteins includes a large number of proteins that play important roles in RNA metabolism. Members of this family have been shown to act as RNA helicases or unwindases, using the energy from ATP hydrolysis to unwind RNA structures or dissociate RNA-protein complexes in cellular processes that require modulation of RNA structures [161–163]. Apart from the helicase activity they also modulate other activities that allow it to function as "RNA chaperones" [164].

These proteins are known to play important roles in all aspects of RNA synthesis and function, including pre-mRNA processing, ribosome biogenesis, RNA turnover, RNA export and translation, processes that involve multi-step association/dissociation of large RNP complexes as well as the modulation of complex RNA structures. Recently it is also implicated in RNA interference and in micro RNA biogenesis [164]. Apart from its functions that are helicase and unwindase dependent, these proteins have also been implicated in many other functions such

as transcription activation or repression which may or may not require the ATP dependent helicase activity. It is believed that DExD/H helicases may be functioning in transcription, through acting as adaptor or bridging factors to recruit transcription factors or to stabilize transcriptional initiation complexes. The emerging understanding about these classes of proteins is that they are multifunctional and participate in a variety of cellular processes and in regulation of gene expression. DExD/H box helicase consist of two subclasses namely the DExH box that consists of helicases such as Dhx9 and the nuclear DNA helicase. The DEAD box sub family consist of Gemin 3, Ddx5 (p68), Ddx17 (p72) etc. Here, the properties and functions of one of the members of DEAD box family namely the p68 or Ddx5 has been briefly reviewed.

1.6.2 p68/Ddx5 Dead box helicases

p68 (Ddx5) is one of the prototypic members of the DEAD box family of proteins and was one of the first proteins to be shown to exhibit RNA helicase activity *in vitro* [165, 166]. A very closely related protein to p68 is the p72 that exhibits 90% identity in its core region and variation only in their N terminal and C terminal region [167]. They are believed to function together where they are found as heterodimers, p68, although have shown to have independent functions as well [168]. The expression of p68 was shown to be growth and developmentally regulated and to correlate with organ maturation/differentiation [169, 170]. It is highly expressed during embryonic stages and is also a proliferation associated nuclear protein. p68 is a nuclear protein and it is distributed throughout the nucleus and specifically localizes to nucleolus during S phase [171]. p68 knockouts

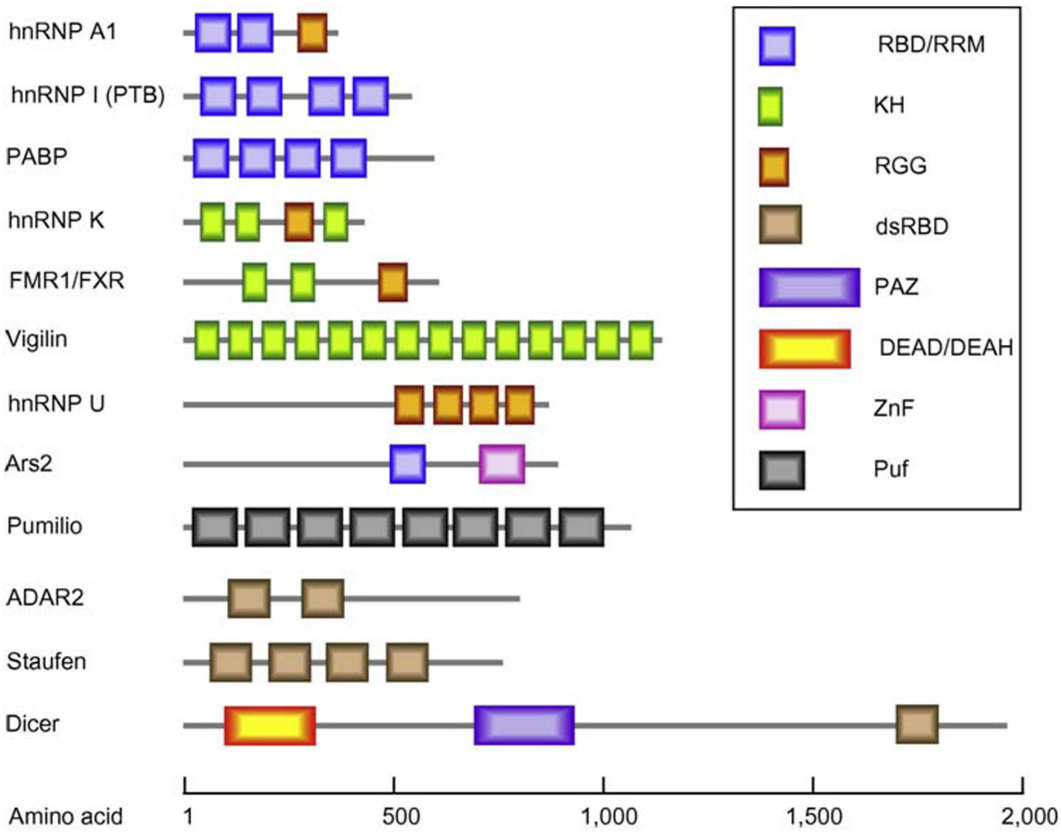

Figure 1.9: *Different domain architecture of RNA binding proteins. Representative proteins harboring those domains are given. Image adapted from Glisovic et al. [156].*

are embryonic lethal suggesting a vital role for this DEAD box helicase. p68 functions in plethora of cellular processes. Initially p68 was identified as a splicing protein and it was purified from spliceosomal complex where it was shown to interact with U1 SnRNA 5'site duplex [172]. Several reports have implicated both p68 and p72 in pre-mRNA processing and alternative splicing, and the yeast p68/p72 (Dbp2) was found to be important for ribosomal RNA processing [173]. More recently, both proteins were found to be components of the mouse Drosha complex and to be required for some if not all primary miRNA (microRNA) and ribosomal RNA processing in mammals [174, 175].

1.6.3 p68 role in transcriptional regulation

A study of the *Drosophila* p68 orthologue (Rm62) demonstrated a function for p68 in clearance of mRNA transcripts from transcription sites and suggested that p68 was required to facilitate release of RNA to allow the chromatin to be reset to an inactive state, resulting in gene deactivation [176]. This was the first report suggesting a role for p68 in transcription *per se*. Recent literature evidences also gives insight into this protein playing both transcriptional activation as well as repression that is promoter dependent. The first indication that p68 may function in transcription in mammals came from a report showing that p68 interacts with and specifically co-activates ERα. This function was totally independent of its helicase activity [177]. Later on, it was shown that p72/p68 together coactivate ER alpha which is dependent on steroid receptor co activator and SRA (steroid receptor RNA activator) a non-coding RNA [89, 178, 179]. A

chromatin immunoprecipitation-based study demonstrated the cyclical, combinatorial, recruitment of cofactors to the promoter of the ERα-target gene pS2 in response to oestrogen, p68 was found to be recruited to this promoter in the first transcriptionally productive cycle together with ERα and the transcription factors TBP (TATAbox-binding protein) and TFIIA (transcription factor IIA), consistent with a key role in transcription initiation [180]. p68 has also been shown to co activate p53 tumor suppressor gene. p68 can bind to p53 *in vitro* as well as *in vivo* and activates genes that are p53 dependent [181]. A recent study has shown that p68 and p72 synergize with SRA to activate MyoD transcriptional activity [182]. In another study, p68 was shown to interact with and to co-activate Runx2, and was found to be recruited to a region of the osteocalcin promoter that contains a Runx-binding element [183]. Another function implicated for this group of proteins is that the role as transcriptional repressors where it can bind to HDAC and repress transcription in a promoter specific manner [184].

Apart from transcriptional coactivator function, p68 has been shown to play a role in cellular signaling events as well. During epithelial mesenchymal transition that is induced by PDGF stimulation, p68 which gets phosphorylated in tyrosine 305 was shown to translocate to the cytoplasm where it stabilizes β-catenin and subsequently p68-β-catenin translocates back to the nucleus bringing about downstream transcriptional regulation [185].

These findings, coupled with previous reports of p68 interactions with components of the transcriptional machinery have suggested models in which p68 and p72 may be important for the recruitment of specific components of the transcription machinery, including chromatin remodeling factors, and may facilitate formation and stabilization of the initiation complex. The recent identification of its role in

cellular signaling process such as wnt has given more insights into varied function of this protein that needs to be explored further.

1.7 Developmental signaling pathways

Thus it is evident from the above discussion, that non-coding RNAs and the RNA binding proteins are indeed a major determinant factor in disease conditions by targeting many cellular signaling processes. Cell signaling is part of a complex system of communication that governs basic cellular activities and coordinates cellular functions. The ability of cells to perceive and correctly respond to their microenvironment is the basis of development, tissue repair and immunity as well as normal tissue homeostasis. Errors in cellular information processing are responsible for diseases such as cancer, autoimmunity, and diabetes. There are different types of cell signaling events in response to various physiological and environmental conditions.

Many cellular signaling events are mediated through their receptors that are coupled to enzymatic activity, steroid hormones, or heterotrimeric G protein. The examples of which include the TGFβ signaling, MAPP kinase signaling, the signaling that involves receptor tyrosine kinases, or the JakStat signaling in response to immunoregulatory stimulus. Apart from the ones mentioned above there are signaling pathways that are intrinsically coupled to cellular development and differentiation. They are the Wnt signaling, Notch signaling, sonic hedgehog signaling. The wnt signaling is described fairly in detail in the following sections.

1.7.1 Wnt signaling

Wnt is an extremely conserved developmental signal pathway. Wnt proteins are secreted signal molecules that act as local mediators to control many aspects of development in all animals. They were independently identified in flies and mice. In *Drosophila*, the wingless gene came to light because of its role in wing development while in mouse Int1 gene,was found because it promoted the development of breast tumor when integration of a viral DNA next to it. The Wnt-β-catenin pathway regulates cell adhesion, morphology, proliferation, migration and structural remodeling [186, 187].

Wnt signaling activates several distinct intracellular pathways, which are important for cell proliferation, differentiation, and polarity. In the canonical wnt signaling wnt proteins, interacts with the transmembrane receptor Frizzled. Following the interaction with Frizzled, the downstream effect of the most widely studied Wnt pathway is stabilization and nuclear translocation of the cytosolic protein, β-catenin. There also exist two other types of wnt signaling independent of β-catenin: 1) Wnt/planar cell polarity (PCP), a Wnt pathway that signals through the small GTPases, Rho and Rac, to promote changes in the actin cytoskeleton, and 2) Wnt/Ca2+, a Wnt pathway that promotes intracellular calcium levels and negatively regulates the Wnt/β-catenin pathway [186, 188].

1.7.2 Canonical Wnt Signaling

The best understood Wnt signal transduction cascade is the Wnt-β-catenin pathway. When this path way operates within a homeostatic range, it regulates cell fate, proliferation and self-renewal of stem and progenitor cells throughout the

lifespans of metazoa. The importance of staying in such a short range is emphasized by the evidence that links hyperactive or hypo active signaling to various diseases [189]. The canonical Wnt signal transduction pathway is schematically illustrated in Figure 1.10.

In the absence of Wnt ligand, intracellular levels of β-catenin are regulated by a multiprotein complex encompassing kinases such as GSK3b (glycogen synthase kinase-3b) and CK1 (casein kinase 1), and the scaffolding proteins APC (adenomatous polyposis coli), Axin1 and Axin2 (conductin). This 'destruction complex' binds and phosphorylates β-catenin at serine and threonine residues, thus targeting it for ubiquitination and proteolytic degradation. In the presence of Wnt ligands, co-activation of the Frizzled and LRP (low-density lipoprotein receptor-related proteins) receptors leads to inhibition of the destruction complex and the consequent stabilization of β-catenin. Intracellular β-catenin accumulation eventually results in its nuclear translocation. In the nucleus, β-catenin binds to members of the TCF/LEF family of transcription factors, thus modulating expression of a broad range of target genes. The components of the signaling pathway are briefly outlined below [186, 188].

1.7.2.1 Extracellular inhibitors

At least three classes of Wnt antagonists are reported in *Xenopus*, and in humans. However, none of them have been identified in *Drosophila* or *C. elegans*.

The first class, *secreted frizzled-related proteins (sFRPs)*, are also called as secreted apoptosis-related proteins (SARPs). The sFRPs thus compete with the Fz proteins for binding to secreted Wnt ligands and antagonize the Wnt function. However, a contradictory effect of the sFRPs has also been described, in which

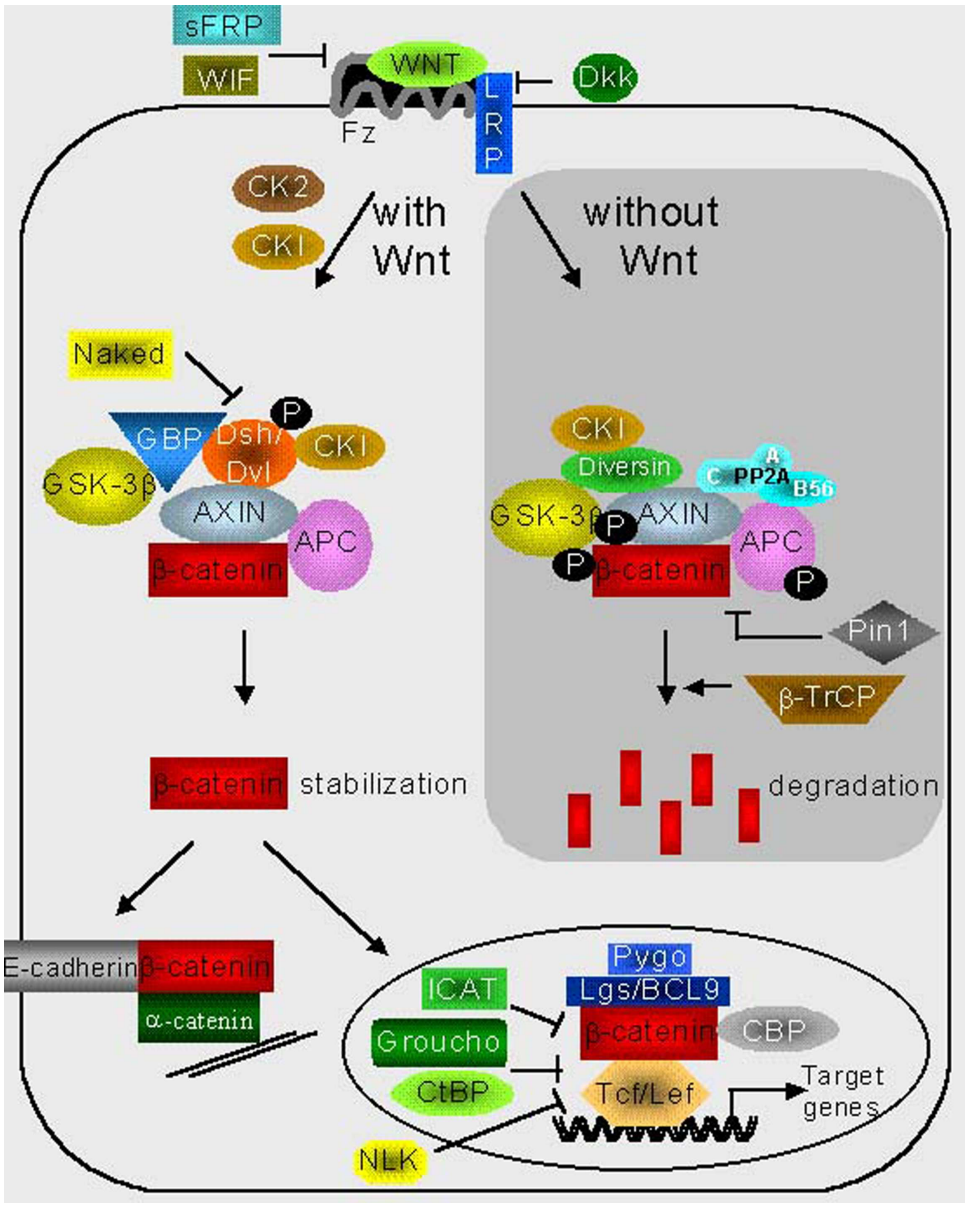

Figure 1.10: *Canonical wnt signaling with and without the ligand. Adapted from wnt home page /stanford.edu.*

the sFRPs enhance the Wnt activity by facilitating the presentation of the ligand to the Fz receptors [190]. *Wnt-inhibitory factor-1 (WIF-1)* represents the second class of secreted Wnt antagonists, and in *Xenopus* WIF-1 binds to Wnt proteins and inhibits their activities by preventing access to cell surface receptors [191]. The third type of secreted antagonists, Dickkopf (Dkk), includes four known human proteins namely DKK1, 2, 3 and 4 [192]. In *Xenopus*, Dkk does not inhibit Wnt ligands directly, but interacts with the Wnt co-receptor, LRP, and prevents formation of an active Wnt-Fz-LRP receptor complex [193]. Recently, it was found that Kremen1 and Kremen2 worked as Dkk receptors [194] and a ternary complex between Kremen2, Dkk1 and LRP6 lead to endocytosis and thus removal of LRP6 from the plasma membrane [194].

1.7.2.2 Ligands and receptors

Wnt ligands belong to a family of proto-oncogenes expressed in several species ranging from the fruit fly to man. This large family of secreted glycoproteins is considered to be one of the major families of signaling molecules. The first Wnt gene, mouse Int-1, was identified by its ability to form mammary tumors in mice when activated by integration of the mouse mammary tumor virus (MMTV) [195]. Int-1 was later renamed Wnt-1 due to the relationship between this gene and the Wg gene in *Drosophila* [196]. At present, 19 human WNT genes are characterized (http://www.stanford.edu/~rnusse/wntwindow.html). Although the individual members of this family are structurally related, they are not functionally equivalent and each may have distinct biological properties [197].

In *Drosophila*, the Fz genes play an essential role in development of tissue polarity. The Fz genes code for seven-trans membrane proteins and evidences show that

Fz proteins function as receptors for Wg in *Drosophila* [198]. Several mammalian homologues have been identified (http://www.stanford.edu/~rnusse/wntwindow.html). Both the extracellular cysteine rich domain and the transmembrane segment are strongly conserved, but nevertheless, the Fz proteins differ in both function and ligand specificity [199]. Although it is known that the Wnts interact with the Fz receptor, the mechanism of Fz signaling is not fully understood [190].

Arrow (*Drosophila*)/LRP (in vertebrates) is required during Wnt signaling, possibly by acting as a co-receptor for Wnt [200–202]. The LRP gene encodes a long single-pass transmembrane protein, and the extracellular domain binds Wnt directly making a ternary complex with the Fz receptor [201]. Recently, it was observed that the intracellular part of LRP binds Axin [193]. Two mammalian homologues have been described, LRP5 and LRP6 [203].

1.7.2.3 Downstream of the receptor complex

Dsh in *Drosophila* and Dvl in vertebrates encode a cytoplasmic phosphoprotein and is a positive mediator of Wnt signaling [204]. In the human genome, three homologues have been described, DVL1-3 [205, 206]. Dsh/Dvl works downstream of Fz receptor, but upstream of β-catenin [190]. However, its exact mechanism of action remains unknown, but several binding partners have been detected. The DIX domain of Dvl binds Axin. This binding inhibits Axin promoted GSK-3b dependent phosphorylation of β-catenin [207]. Upon Wnt stimulation, Dsh/Dvl binds CK1. This binding probably inhibits phosphorylation of the serine (Ser) 45 site in β-catenin causing stabilization of β-catenin and activation of the Wnt pathway [208, 209]. However, the exact mechanism of this event is yet unknown.

In the absence of a Wnt signal, CK1 associates and cooperates with Axin, probably through diversin. This drives the phosphorylation and degradation cascade of β-catenin, and subsequently inhibits the Wnt signaling pathway [209]. One theory suggests that in the presence of a Wnt signal, Dvl recruits GBP to the multiprotein complex. GBP then titrates GSK-3b from Axin leading to accumulation of β-catenin in the cytoplasm [210]. The human homologue of GBP is frequently rearranged in advanced T-cell lymphomas (FRAT). Newly, FRAT/GBP was shown to contain a nuclear export sequence, leading to nuclear export of itself as well as the bound GSK-3b. Thus, FRAT/GBP is involved in regulating the accessibility of cytoplasmic GSK-3b [211].

1.7.2.4 The multiprotein complex

The stability of β-catenin (encoded by the gene CTNNB1) is regulated by a multiprotein complex consisting of β-catenin, Axin/Conductin, APC, and GSK-3b [212]. In this scaffolding complex, GSK-3b phosphorylates primed β-catenin, thus marking β-catenin for ubiquitylation and subsequent proteasome mediated degradation. During the last few years, several novel players that interact with the components of the multiprotein complex have emerged. Still, the exact mechanisms of action of the multiprotein complex need further clarification.

β-catenin

β-catenin was first described in humans as a protein which interacts with the cytoplasmic domain of E-cadherin and with α-catenin, anchoring the cadherin complex to the actin cytoskeleton [213]. Subsequently, the homology between β-catenin and the Armadillo (Arm) of *Drosophila* and β-catenin in *Xenopus* lead to the discovery of an additional role for mammalian β-catenin, namely as the key

mediator of Wnt signaling [214]. The primary structure of β-catenin comprises an amino-terminal domain of approximately 130 amino acids, a central region of 12 imperfect repeats of 42 amino acids known as arm repeats (since they show homology with repeats found in the Arm protein of *Drosophila*), and a carboxy-terminal domain of 110 amino acids. The amino-terminus of β-catenin is important for regulating its stability, whereas the carboxyl terminus works as a transcriptional activator domain [215]. Interestingly, plakoglobin, also called g-catenin, shares overall 70% amino acid identity with β-catenin and as much as 80% within the arm repeat domain [216]. Plakoglobin binds E-cadherin, a-catenin, APC, Axin and Tcf/Lef transcription factors, and is involved in cell adhesion as well as Wnt signaling. However, differences between β-catenin and plakoglobin in these processes exist. β-catenin activity is controlled by a large number of binding partners that affect the stability and localization of β-catenin. Binding of a-catenin to the N-terminal region of β-catenin and E-cadherin to the arm repeat [216] connects β-catenin to cell adhesion [150].

The arm repeat domain of β-catenin mediates binding of cadherins, APC, Axin and Tcf/Lef family of transcription factors [150]. E-cadherin, APC and Tcf/Lef interact with this domain of β-catenin in an overlapping and mutually exclusive manner [215]. Two ubiquitin-mediated degradation systems are involved in the catabolism of β-catenin. Both F-box proteins, b-TrCP and Ebi, recognize and bind to the same sites on the N-terminal domain of β-catenin [217]. However, unlike b-TrCP, Ebi probably does not require phosphorylation of β-catenin for recognition. Ebi works in complex with SIAH-1, a TP53 induced protein, linking activation of TP53 to the degradation of β-catenin [218]. Both the degradation systems require an intact APC protein [217]. GSK-3b sequentially phosphorylates threonine (Thr)

41, Ser 35, and Ser 33 of β-catenin after β-catenin has been primed (phosphorylated at Ser 45) by CK1 [209, 212]. In *Drosophila*, Legless (Lgs) and Pygopus (Pygo) have recently been shown to be required for Arm to function as a transcriptional co-activator in the Wg signaling pathway. Lgs encodes the homologue of human BCL-9, whereas Pygo codes for a PHD (plant homology domain) finger protein and two human homologues have been identified, hPYGO1 and hPYGO2 [219]. In the mouse system, inhibitor of β-catenin and TCF-4 (ICAT) binds the C-terminal domain of β-catenin and inhibits its interaction with TCF-4. β-catenin-TCF-4 mediated transactivation of Wnt target genes is then repressed [220].

APC

The first clue of the mode of action of APC came from studies showing that APC binds β-catenin. Later it has been demonstrated that APC plays a central role in regulating the β-catenin level in the Wnt signaling pathway in addition to be involved in cell migration, cytoskeleton regulation, chromosome segregation and in tumors. APC encodes a large protein consisting of several distinct conserved domains interacting with a number of different proteins [221].

Axin

Axin works as a scaffold protein involved in forming the multiprotein complex leading to phosphorylation and degradation of β-catenin and thereby acts as a negative regulator of Wnt signaling. Later, a homologue of Axin in mouse, Conductin, and in rat, Axil, were identified [222]. Two human homologues also exist, namely, AXIN1(the Axin homologue) and AXIN2(the Conductin/Axil homologue). Axin and Conductin share 45% amino acid identity. Interestingly, Axin is ubiquitously expressed, whereas Conductin is more selectively expressed in specific tissues. Recently, it was shown that Conductin is a downstream target

gene of the Wnt pathway and might work in a negative feedback loop controlling Wnt signaling activity [223]. GSK-3b is recruited to the multiprotein complex by Axin [222] and then phosphorylates Axin and APC and thereby increasing their interaction with β-catenin. Subsequently, primed β-catenin bound to Axin and APC, is phosphorylated by GSK-3b, marking β-catenin for proteasome mediated degradation.

GSK-3b

GSK-3b, zw3 or shaggy in *Drosophila*, is a member of the Ser/Thr family of protein kinases. This protein is a key enzyme in the Wnt signaling pathway. As outlined above, GSK-3b phosphorylates primed β-catenin prior to proteasome degradation, and it phosphorylates Axin and APC and enhances their interaction with β-catenin. Unlike most protein kinases, GSK-3b is constitutively active and phosphorylation of GSK-3b leads to inhibition of its activity [224]. Two highly related human homologues, GSK-3a and GSK-3b have been identified, and these two isoforms are more than 95% identical in the protein kinase catalytic domain. Consequently, GSK-3a can substitute for many, but not all of the functions of GSK-3b in the Wnt signaling pathway [225].

1.7.2.5 Nuclear components of β-catenin signal

The stabilized cytosolic β-catenin gets translocated into the nucleus, but how β-catenin enters the nucleus is not yet fully understood. Nuclear β-catenin associates with the family of Tcf/Lef transcription factors, and is therefore a key factor for expression of Wnt downstream genes.

Tcf/Lef

The Tcf/Lef proteins are a class of related high mobility group (HMG)-box

of transcription factors. Four human homologues of Tcf/Lef have been identified, namely LEF1, TCF1, TCF3 and TCF4. They all recognize the same DNA sequences, however they display tissue specific expression patterns. Upon Wnt signal, Tcf/Lef acts in a complex with β-catenin, BCL-9, Pygo and CBP and target genes like c-MYC, cyclin D1, WNT inducible signaling pathway protein (WISP)-3, and matrix metalloproteinase (MMP)-7, are expressed. Without the presence of Wnt stimulation, the Tcf/Lef proteins repress transcription of the Wnt target genes by binding to co-repressors like Groucho and C-terminal binding protein (CtBP). Recently it was also shown that Brg1(Smarca4) is also another nuclear component in β-catenin dependent transcription whereby, Brg1 is required for the chromatin remodeling activity in association with CBP and nuclear β-catenin [226, 227].

1.7.3 Non canonical wnt signaling

The non canonical wnt signaling is essential for determining the cell polarity in the lower as well as higher organisms and hence it is also known as planar cell polarity pathway. The molecular and genetic dissection of the so-called noncanonical Wnt/PCP pathway has been increasingly emerging, as this genetic pathway is involved in several diverged processes that require the coordination of cell polarity and cell cohesion within a cluster of cells in vertebrates. These processes include convergent extension (CE), neural tube closure, ear hair cell orientation, ciliogenesis, and hindbrain neuron migration. Most recently, it has been suggested that the PCP pathway is linked with human birth defects and diseases [228]. Non canonical wnt signaling is activated upon binding of ligands such as wnt 5b, wnt11 and activation of disheveled, the key player in non canonical wnt signaling events.

This would eventually activate the Rho/Rac GTPases that brings about actin reorganization and establish epithelial cell polarity in higher vertebrates. Available evidence suggests that Frizzled is activated by the interaction between the non classical cadherin-like cell adhesion proteins, Fat and Dachsous in *Drosophila* and some mammalian protocadherins seems to be homologues of these Drosophila proteins [150].

1.8 Aim and Scope of the Present Investigation

Eukaryotic genome encodes a huge repertoire of non-coding RNA molecules. A number of new non coding RNAs are being discovered in the last few years and new functions are being assigned to them. The discovery of these molecules has made a tremendous impact in our understanding of the higher order genome regulation and the complexity of an organism itself. The ENCODE project consortium which aims in identification of functional elements in the genome has concluded that almost 90% of any genomic output are non protein coding RNAs, in contrast to protein coding part of the genome which is only 2% in higher mammalian species. Though large proportion of this non-coding RNAs has been dismissed as spurious transcription, it is becoming increasingly clear that the vast majority of these could be functional transcripts.

Non-coding RNAs of varying sizes such as small and long non-coding RNA, that were discussed above, have shown to participate in plethora of cellular processes such as transcriptional and post transcriptional gene regulation, chromatin remodeling, and regulation of protein function and so on and so forth. However, a systematic study in terms of characterizing these species and identification

of functional roles for these RNAs is far from complete. Very few mammalian non-coding RNAs have been extensively characterized with their functions partly understood. Bulk of many other non-coding RNA species have been identified by genome sequencing projects or discovered accidently, which needs much more detailed characterization in terms of its properties and functional roles. This study aims in understanding the role of one such non coding RNA from the mouse.

Earlier work from our laboratory that involved the study of a meiotic recombination hotspots in mouse identified a novel hot spot locus [229]. This locus was mapped to mouse chromosome 8 CD near to a meiotic DNA repair site. When analyzed for the transcription property of this locus, it was serendipitously found that the hotspot locus encodes a non protein coding RNA. The RNA was found to be 2.4 kb in length. Initial analysis showed that the RNA lacked any open reading frame and it was expressed in multiple tissues. Interestingly, the RNA also assumed a very stable secondary structure when predicted using secondary structure prediction programs such as mFold. It was also found to be a putative RNA polymerase II transcription product.

In this present investigation we have undertaken a detailed study of this transcript with the following objectives in mind.

1. Characterization of the non-coding RNA.

2. Function role of the non-coding RNA in cellular context and identification of the non-coding RNA interacting partners.

The following sections describe in detail on the experimental approaches undertaken and the results obtained in addressing the objectives.

Chapter 2

Materials and Methods

2.1 Materials

All the fine chemicals that were used were purchased from Sigma Chemical Company, St.Louis, MO, U.S.A. or from GIBCO BRL, Life Technologies Inc. MD, U.S.A. Restriction enzymes used were obtained from New England Biolabs Inc. MA, U.S.A. Thermostable Thermoscript reverse transcriptase was obtained from Invitrogen Corp CA, U.S.A. Nylon membranes were obtained from Amersham International, England. Taq DNA polymerase was obtained either from New England Biolabs, Inc. MA, U.S.A. or from Bangalore Genei, Bangalore, India. Fast Link DNA ligase was procured from Epicentre Biotechnologies, WI, and USA. The primers used were synthesized from Microsynth, Switzerland or from Sigma, U.S.A. All the bacterial growth media were obtained from Hi Media, Bangalore, India. The Enhanced ChemiLuminscence (ECL) kit was purchased from Pierce, U.S.A. The entire analytical grade chemicals were purchased from Ranbaxy Laboratories,

India and Merck (India) Ltd. X-ray films for ECL were obtained from Kodak. ^{35}S-cysteine methionine and α^{32}P-UTP, α^{32}P-dATP and α^{32}P-ATP were purchased from BRIT-India. siRNAs designed for *mrhl* were purchased from Dharmacon, U.S.A. SiRNA against Drosha were purchased from Qiagen, Germany. The modified nucleotides and enzymes used for *in vitro* transcription were purchased from Roche Life sciences. Cy3 and cy5 labeled locked nucleic acid probes (LNA) were procured from Exiqon, Denmark. 5' dideoxy oligonucleotides were procured from Integrated DNA technologies IDT, U.S.A. All the reagents used for cell culture studies were obtained from Sigma or Invitrogen. Fetal bovine serum was purchased from Sigma or US Biologicals. Transfection reagents were obtained from Invitrogen. Recombinant Dicer enzyme was obtained from Ambion, USA. Antibodies were obtained from Santa Cruz Biotechnology Corporation (Nucleolin, Lamin B), Upstate (β-Catenin), Novus Biologicals (p68), Sigma Aldrich, U.S.A (Anti-Flag) or Cell Signalling technologies (phosphor-tyr 100). Drosha antibody was a gift from Narry Kim, Seoul National University. All the reagents, enzymes and buffers used for microarray analysis were purchased from Affymetrix technologies. The mouse 43.0K expression arrays were obtained from Affymetrix technologies.

2.1.1 Animals

The animals used in this study were male Balb/C mice or CD1mice. The animals were procured from Animal Facility, JNCASR Bangalore, India.

2.1.2 *E.coli* Strains

DH5 α: F'endA1 hsd R17 (r_k^- m_k^-) glnV44 thi1 recA1 gyrA (Nal^r) relA1 Δ(lacIZYA - argF) U169 deoR (ϕ80dlacΔ(lacZ M15)

XL1 Blue: endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 $proAB^+$ $lacI^q$ Δ(lacZ)M15] hsdR17(r_K^- m_K^+)

2.1.3 Cell lines

Cell lines used in this study were Gc1-Spg mouse spermatogonia (type B), RAG1 (Mouse Renal Adeno Carcinoma), HEK293T (Human embryonic kidney), HeLa (Human Cervical Carcinoma), C6(Rat Glioma) all of which were obtained from American Type Culture Collection (ATCC), U.S.A. All the cell lines were maintained in DMEM high glucose supplemented with 10% fetal bovine serum containing recommended concentration of antibiotics (Penicillin, Streptomycin and Amphotericin B) and L.Glutamine was added to a final concentration of 1 mM.

2.1.4 Plasmids

1. pGem3Zf+ - Amp resistance - Full length *mrhl* RNA
2. pBlue Script - Amp resistance - 3.4 kb contig from mouse chromosome 8
3. pCK Drosha Flag - Kan resistance - Full length human Drosha with C terminal flag
4. pCDNA3.1 - Amp resistance - Full length *mrhl* RNA

Full length *mrhl* RNA with 3' s1Aptamer fusion

2.2 Methods

2.2.1 Extraction of total RNA

DEPC treated water was used for all the RNA work. 0.01% DEPC was added to autoclaved water. After 12 hours, it was autoclaved again and stored at -20 C. Trizol (Invitrogen Corp.) was used for RNA isolation from tissues and cell lines. Initial homogenization of tissues was done in a TISSUE TEAROR (Biospec Instruments Inc.). The protocol for RNA isolation was as follows. The tissue after homogenization in TRIzol was (1 ml TRIzol for 100 mg tissue) was incubated for 10 minutes at room temperature. Two hundred microlitres of chloroform was added to the homogenate and mixed thoroughly. The mixture was spun at 12000 g for 15 minutes. The aqueous phase was separated and 500 μl of isopropanol was added to the aqueous phase to precipitate the RNA. The RNA was collected after spinning down at 12000 g for 15 minutes at 4 C. The clear RNA pellet was washed once with 75% ethanol and then finally resuspended in nuclease free water.

2.2.2 Formaldehyde agarose gel

The formaldehyde agarose gel was prepared by melting agarose (GIBCO-BRL) for a 1% gel in DEPC treated water and after cooling to 60 C, 5X formaldehyde gel running buffer (0.1 M MOPS pH 7.0, 40 mM Sodium acetate, 5 mM EDTA pH 8.0) and formaldehyde were added to a final concentration of 1X and 2.2M respectively. The gel was cast and allowed to solidify for at least 30 minutes and the gel was pre-run for 5 minutes at 50 volts in 1X formaldehyde gel running buffer.

Total RNA sample was prepared as follows for loading into the gel: RNA 4.5

μl, 5X formaldehyde gel running buffer 2.0 μl, formaldehyde 3.5 μl and formamide 10.0 μl. The RNA was denatured by heating to 65 C for 15 minutes and then chilled on ice for 5 minutes. Prior to loading, the 6X loading dye was added to the RNA sample (50% glycerol, 1 mM EDTA pH 8.0, 0.25% bromophenol blue and 0.25% xylene cyanol FF). The RNA sample was then loaded into the pre-run gel and electrophoresed until the bromophenol blue dye reached the bottom of the gel.

2.2.3 Synthesis of first strand of cDNA

About 3 μg of total RNA was reverse transcribed using the Thermoscript RT-PCR system. In a 0.5 ml microfuge tube the total RNA and 1 μl of 10 μM 3'-gene specific primer were mixed. The total volume was adjusted to 10 μl using DEPC-treated water. The RNA and the primer were denatured by heating the mixture at 65 C for 5 minutes and then snap chilled. To the denatured template and primer, 1X cDNA synthesis buffer, 0.005 M DTT, 40 U of RNase inhibitor, I0 mM dNTP mix and 15U of Thermoscript Reverse Transcriptase were added. The reaction mixture was incubated at 65 C for 1 hour after adjusting the total volume to 20 μl in a thermal cycler preheated to 65 C. The reaction was terminated by incubating at 85 C for 5 minutes. The cDNA synthesis reaction mixture was stored at -20 C or used for PCR immediately. When required, 1/10th of the reaction contents were used for subsequent Polymerase chain reaction.

2.2.4 Polymerase chain reaction (PCR)

In a typical PCR about 10 ng of DNA (1/10th of RT product, cDNA in the case of RT-PCR), dNTPs to a final concentration of 200 μM, 10 μM stock of the sense and antisense primers, 5 μl of 10X Taq DNA polymerase buffer, water to a total volume of 50 μl and 1 unit of Taq polymerase were added. The PCR was performed for 35 cycles and the cycling conditions were 94 C/1 minute to denature the template, annealing temperature of 5 C below the Tm of the primer pairs for 1 minute and extension at 72 C for 1 minute/kb with a final extension at 72 C for 10 minutes. For cloning purposes the Phusion (New England Biolabs) or Platinum Taq DNA polymerase-HiFi (Invitrogen corp.) was used and the manufactures standard protocol was followed.

2.2.5 Synthesis of second strand of cDNA by PCR

In a typical second strand synthesis reaction only 10% of the first strand synthesis reaction (2 μl) was used. To the first strand synthesis reaction, dNTPs to a final concentration of 200 μM, sense and antisense primers to a final concentration 200 μM, 1X polymerase buffer, water to a total volume of 50 μl and the respective polymerase were added. The PCR was performed for 30 cycles by following the recommended thermal cycling conditions. The final PCR product was checked on a 1% agarose gel with appropriate molecular weight markers.

2.2.6 Agarose gel electrophoresis

The flat bed agarose gel electrophoresis was conducted to visualize DNA samples. The agarose gel of required percentage was cast in 1X TAE buffer (40 mM Tris

acetate and 1mM EDTA pH 8.0) containing 10 μg/100 ml of ethidium bromide. DNA was mixed with 6X loading buffer (50% glycerol containing a pinch of either bromophenol blue or Xylene cyanol) and loaded into the wells. Electrophoresis was carried out at 80-100 volts for 1 hour and the DNA was visualized by placing the gel on a UV trans-illuminator and photographs were taken using BioRad-Gel documentation system 1000.

2.2.7 Restriction digestion of DNA

The plasmid DNA obtained by alkaline lysis method [230] or the PCR products carrying the suitable restriction enzyme sites were digested with the respective restriction enzymes for at least 2-3 hours at 37 C in their respective buffers. The digested DNA was again separated on agarose gels. The digested PCR products or vectors were separated on a 1% agarose gel and subsequently eluted for further use.

2.2.8 Elution of DNA from agarose gels

Gel elution was performed using Qiagen Gel Elution Kit following manufacturer's protocol. The DNA pellet was air dried and resuspended in the required amount of sterile water.

2.2.9 Ligation of digested DNA into plasmid vector

A typical ligation reaction was setup with gel-purified, digested vector and an insert of interest at a molar ratio of 1:3, 1X ligation buffer containing 10 mM ATP and 0.5 μl of Fast Link DNA ligase (15 units/μl - Epicentre Biotechnologies) in a

total reaction volume of 15 μl. The reaction mix was incubated at 25 C over night.

2.2.10 Preparation of media

The medium used for the bacterial work was LB broth (10 g bactotryptone, 5 g yeast extract, 10 g NaCl for one liter; autoclaved at 15 lbs for 20 minutes). To prepare LB-agar, bacteriological agar was added to the LB broth to a final concentration of 1.5%. Before adding the appropriate antibiotic, the medium was cooled to around 50 C. Ampicillin was used at a final concentration of 0.1 mg/ml.

2.2.11 Preparation of competent cells by and transformation of bacterial cells with plasmid DNA

Competent cells were prepared by one step protocol developed by Chung *et al.* [231] with minor modifications. The *E.coli* strain DH5α or XL1 Blue was grown overnight in 3 ml LB broth without ampicillin. A 200 μl aliquot of the overnight culture was transferred into 50 ml of LB broth without ampicillin and incubated at 37 C with shaking. The cells were grown till an O.D600 of 0.3-0.4. The cells were chilled, transferred to an Oakridge tube and pelleted at 2,500 g for 10 minutes at 4 C. After removing the supernatant, the pellet was gently resuspended in 2 ml of filter sterilized ice cold 1X TSS (Transformation and Storage Solution) (1 ml TSS was prepared by mixing 500 μl of 2X LB broth, 400 μl of 25% PEG 3350, 50 μl of 1 M $MgCl_2$ and 50 μl of DMSO) and 200 μl aliquots were directly used for bacterial transformation or frozen in liquid N_2 for future use.

For transformation of ligation mix or positive transformation an aliquot of the competent cells was thawed on ice and approximately 50 ng of DNA in minimal

volume was added. The cells were kept on ice for 30 minutes and subjected to heat shock for 90 seconds at 42 C. Immediately, the cells were placed on ice for 5 minutes and 0.8 ml of sterile LB broth without ampicillin was added and allowed for recovery at 37 C for 45 minutes. The cells were pelleted at 2,500 g for 3 minutes. at room temperature. The pellet was resuspended in 200 μl of the supernatant and spread on LB-agar plate with appropriate antibiotic and incubated at 37 C overnight.

2.2.12 Screening of recombinant clones by mini preparation of plasmid DNA by alkaline lysis method

The bacterial colonies obtained in the above protocol were randomly picked and inoculated into 3 ml LB-broth with ampicillin and grown to saturation. The culture was centrifuged at 12,000 rpm for one minute at 4 C. The cell pellet was suspended in 100 μl of solution I (50 mM glucose, 25 mM Tris pH 8.0, 10 mM EDTA pH 8.0) by vortexing. A 200 μl aliquout of freshly prepared solution II (0.2 N NaOH, 1% SDS) was added and mixed the contents of the tube gently and the tubes were placed on ice for 5 minutes. 150 μl of solution III (3M potassium acetate, pH 4.8) was added and after gentle mixing the tubes were placed on ice for a further period of 10 minutes. The contents were spun at 11,000 g for 10 minutes at 4 C and the supernatant containing the plasmid was transferred to a fresh microfuge tube. The nucleic acids were precipitated by adding equal volume of isopropanol at room temperature for 20 minutes. The pellet was resuspended in 50 μl of T10E1. RNA was removed by digestion with RNase A (1 mg/ml) at 37 C for 20 minutes. The DNA was extracted twice with phenol:chloroform ratio

and finally with chloroform. The extracted DNA was precipitated with one-tenth the volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ice-cold ethanol and stored at -20 C for at least an hour. The DNA was pelleted by centrifuging at 11,000 g for 10 minutes. at 4 C. The pellet was washed with 70% alcohol, dried and resuspended in a minimum volume of autoclaved water. An aliquot of the DNA was then digested with suitable restriction enzymes and checked for release of the insert on a 1-1.5% agarose gel.

2.2.13 Purification of plasmid DNA by PEG precipitation for sequencing

The plasmid DNA prepared by the alkaline lysis method after confirming for the release of proper size insert by appropriate restriction enzyme digestion was purified by PEG precipitation. The DNA pellet after the mini-preparation was resuspended in 32 μl of autoclaved water. Subsequently 8 μl 4 M NaCl and equal volume of 13% PEG 8000 were added and the tube was incubated on ice for at least 1-2 hours. The DNA was centrifuged at 11,000 g for 15 minutes at 4 C and the pellet was washed with 70% ethanol after aspirating the supernatant and again centrifuged under similar conditions. The pellet was then dried, dissolved in autoclaved water and an aliquot was checked on agarose gel to check the quality of the DNA.

2.2.14 SDS poly acrylamide gel electrophoresis

Proteins were analyzed on SDS poly-acrylamide gels (SDS-PAGE). A 10% SDS-PAGE were used for analyzing the proteins. The composition of the running gel

and stacking gels is given for 10 ml.

2.2.15 Running gel composition

Water	-	4.0 ml
29:1 Acrylamide-bisacrylamide mix	-	3.3 ml
1.5 M Tris.Cl (pH 8.8)	-	2.5 ml
10% SDS	-	0.1 ml
10% Ammonium persulphate	-	0.1 ml
TEMED	-	0.004 ml

2.2.16 Stacking gel composition

Water	-	6.8 ml
29:1 Acrylamide-bisacrylamide mix	-	1.7 ml
1.5 M Tris.Cl (pH 6.8)	-	1.25 ml
10% SDS	-	0.1 ml
10% Ammonium persulphate	-	0.1 ml
TEMED	-	0.01 ml

After polymerization, the gel was mounted in the electrophoresis apparatus and electrophoresed in 1X Tris glycine SDS buffer (25 mM Tris/250 mM glycine pH 8.3/0.1% SDS) at 100 V till the bromophenol blue dye reached the bottom of the gel. The samples were prepared in 50 mM Tris.Cl pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue/10% glycerol. After electrophoresis gels were stained with 0.4% Coomassie Brilliant Blue in 40% methanol/10% acetic acid for 4 hours. Destaining was carried out in the same solution without Coomassie Brilliant Blue.

2.2.17 Western blot analysis and detection

The protein bands resolved by 10% SDS-PAGE were transferred on to nitrocellulose membrane using Hoeffer semidry transfer apparatus. Before transfer, the gel and nitrocellulose membrane were soaked thoroughly in the transfer buffer (0.025 M Tris/0.192 M glycine/20% methanol). The proteins were transferred using the above-mentioned apparatus electrophoretically at 2 mA/sq cm of current for 90 minutes. After transfer, the membrane was blocked in the rinse buffer (10 mM sodium phosphate, pH 7.2/0.9% NaCl/0.05% Tween 20) containing 5% milk powder at room temperature for overnight. Subsequently the membrane was rinsed with 3-4 changes of rinse buffer for 30 minutes. The membrane was then incubated in the rinse buffer containing 1% milk powder and 1μg of appropriate primary antibody for 2 hours at room temperature. After washing with rinse buffer as before, it was incubated with 1:5000 diluted HRP-conjugated secondary antibodies (Bangalore-Genei) at room temperature with gentle shaking for 90 minutes. The membrane was again washed with several changes of rinse buffer. The detection of antibody reaction was done by Enhanced ChemiLuminescence (ECL) of Amersham Biosciences following manufacturer's protocol or using ECL kit of Thermo scientific corporation.

2.2.18 Cloning of *mrhl* gene in pGEM3zf+

The total RNA from mouse testis was reverse transcribed using oligo dT reverse primer. The cDNA generated was used as a template for PCR using *mrhl* gene specific primers containing adapter sequences with restriction site.

mrhl F: 5'TACT**GGATCC**GTGACTTGCTCTTCATTAGAT 3'

mrhl R: 5'GTAC**CTCGAG**TGGGGTAGTCTCTGGATAGT 3'

The pGem 3Zf+ was digested with BamHI and SalI and the amplified insert DNA was digested with BamHI and XhoI. The respective digested products were gel eluted and ligated in 1:3 molar ratio of vector to insert. The recombinant clones were screened on LB Ampicillin plates after transformation in DH5α competent cells. The clones containing the insert were verified using PCR, restriction digestion and sequencing.

2.2.19 RNA isolation from nuclear and cytoplasmic fraction

RNA was isolated from the nucleus and cytoplasm from liver and testis by the following protocol. The tissue was homogenised in buffer A (10 mM Tris-Cl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 40 mM $NaHSO_3$, 0.1 mM PMSF) containing 0.34 M sucrose and 0.5 U/μL RNAsin and 2 mM Vanadyl Ribonuclease inhibitor complex (VRC). The homogenate was spun at 1000 g to separate nuclear and cytoplasmic fractions. The nuclear fraction was suspended in buffer A containing 2.3 M sucrose and spun at 100,000 g for 1 hour. The nuclear pellet was washed once in buffer A containing 0.34 M sucrose and the RNA was isolated using TRIzol reagent. For cytoplasmic RNA isolation the supernatant after separating the nuclear pellet was taken and the RNA was extracted by using phenol: chloroform followed by addition of 2.5 volume of ethanol and was allowed to precipitate at -20 C overnight. After incubation the RNA was precipitated by spinning down at 14,000 g for 30 minutes at 4 C. The RNA pellet was washed once in 75% ethanol and then finally resuspended in nuclease free water. RNA obtained from fractionated tissues was

used for RT PCR or northern analysis.

2.2.20 Subnuclear fractionation

Nuclei isolated from the tissues as described previously were resuspended in buffer A and sonicated for 5 sec at 25 amplitude in a sonicator (Vibra sonics) and layered on 30% sucrose cushion containing 2 mM VRC and centrifuged at 1000 g for 20 minutes to fractionate it into chromatin and nucleoplasm. RNA was isolated from both the fractions after digestion with RNase free DNAse (Sigma), by the protocol described above. An aliquot of the input was also used for profiling the protein composition to check the purity of the preparation.

2.2.21 *In vitro* coupled transcription/translation

The *in vitro* coupled transcription/translation was done using T7 rabbit reticulolysate system (Promega). Briefly, the gene cloned downstream of the T7 promoter in the pGem3Zf+ vector, was added to rabbit reticulolysate along with 35S-methionine and incubated for 90 minutes at 30 C. The *in vitro* translated products were separated in 15% SDS-PAGE. The gel was dried and exposed to phosphorimager screen. RNA was also isolated from the same mixture after DNase treatment using TRIzol and RT PCR was performed for the specific genes of interest.

2.2.22 *In vitro* transcription

The 2.4 kb *mrhl* cloned in pGem vector under T7 promoter was linearised using XhoI, purified and used as a template for *in vitro* transcription. The template was incubated with 1 mM each of ATP, CTP, GTP and UTP or α^{32}P UTP and T7

RNA polymerase in 1X transcription buffer. The reaction mixture was incubated at 37 C for 2 hours and the *in vitro* transcribed RNA was purified using RNA easy purification kit (Qiagen, Germany). For labeling the transcripts with biotin the UTP concentration was reduced to 0.7 mM and biotin UTP was added to a final concentration of 0.3 mM.

2.2.23 Urea PAGE

The small RNAs were separated in 15% PAGE containing 7M Urea as follows. The ultrapure urea was weighed and added to 15% acrylamide /bis acrylamide mix from a 40% stock solution. The urea was allowed to dissolve completely and 0.5X TBE(stock 5X) was added to the mixture. Then APS and TEMED were added and the gel was allowed to set in the cast for atleast 2-3 hours. The gel was prerun at 50 V for 30 minutes before loading the samples.

2.2.24 Random primer labeling

About 50 ng of PCR purified template DNA was subjected to random primer labeling using NEB random primer labeling kit according to manufactures instructions. Briefly, the template was denatured at 95 C and snap chilled. To this 1mM each of dGTP, dCTP, dTTP were added followed by addition of 10 μci of α^{32}PATP. About 10 units of Klenow polymerase were added to the mixture along with the buffer containing random hexamer. The reaction mix was incubated at 37 C for 1 hour and the reaction was inactivated by the addition of 0.2 mM EDTA. The labeled DNA was separated from unlabeled free nucleotides using sephadex G 50 column.

2.2.25 Small RNA Northern Hybridization

The small RNA PAGE was carried out as described by [232]. Approximately 30 μg of RNA in loading buffer was denatured at 65 C for 15 minutes and loaded on the 15% PAGE containing 7 M Urea gel. The EtBr stained gel was used to check the integrity and the quantity of RNA. The small RNAs were transferred to Nylon membrane and immoblised under UV. The membrane was prehybridized at 42 C for 2 hours in prehybridization buffer (7% SDS, 5X SSC, 2X Denharts reagent, 0.2 M sodium phosphate and 100 g/ml of denatured single stranded DNA). The amplified PCR product representing full length or part of the *mrhl* gene was gel purified and was used for probe generation by Klenow labelling using random primers. Alternatively RNA probes derived from *in vitro* transcription in the presence α^{32}PUTP were also used for strand specific northern analysis. Hybridisation was carried out at 42 C overnight. The membrane was then washed twice in 2X SSC/0.5% SDS at room temperature and once each in 1X SSC/0.1% SDS at 42 C and 0.1X SSC/0.1% SDS at 42 C and exposed to phosphorimager screen.

2.2.26 Splinted Ligation

Splinted ligation was carried out as described by Maroney *et al.* [233]. The sequence of the oligos used in the experiment is given below. The Linker oligo was purchased from IDT, while the Bridge oligos were from Sigma.

Linker Oligo: 5'CGCTTATGACATTddC3'

Bridge *mrhl* 1: 5'GAATGTCATAAGCGAAGCACTTACTGTCACTTGAT-GCA3'

Bridge *mrhl* 2: 5'GAATGTCATAAGCGCAAAATTAAAACAGCTTCTAC-CTCT3'

Bridge Let 7: 5'GAATGTCATAAGCGAACTATACAACCTACTACCTCA3'

The ligation oligonucleotide was 5'-end radiolabeled by incubating 20 pmol DNA, 2.5 μM [γ-32P]ATP (150 mCi/mL, Perkin-Elmer), 50 mM Tris-HCI (pH 7.5), 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 5% glycerol, and 20 units T4 poly nucleotide kinase (Promega) in a reaction volume of 20 μL for 30 minutes at 37 C. The volume of labeled products was adjusted to 100 μL with RNase-free water and was purified in a Sephadex G25 coloumn.

For ligation, all reactions consisted of 100 fmol bridge oligonucleotide, 100 fmol radiolabeled ligation oligonucleotide, total RNA sample(2-5 μg), 75 mM KCl, 20 mM Tris (pH 8.0), 10 units T4 DNA ligase (Promega), and 1X ligation buffer (Promega) in a reaction volume of 15 μL. Before adding T4 DNA ligase and Ligate-IT buffer, the reaction mixture was denatured at 95 C for 1 minute. After annealing at 65 C for 2 minutes and 37 C for 10 minutes, the ligase and the buffer were added to the reaction mixture and incubated at 30 C for 1 hour. Reactions were terminated by heat inactivation at 75 C for 15 minutes and then treated with 1 unit of calf intestinal alkaline phosphatase (Promega) at 37 C for 15 minutes. Reaction products were separated using denaturing 15% urea-polyacrylamide gels and quantified using a PhosphorImager.

2.2.27 miRNA processing Assay

The miRNA-processing assay was carried out using mouse testicular total cell lysates. Briefly, the testis was decapsulated and treated with collagenase IV

(Sigma). The cell pellet was resuspended in buffer B (20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 5% glycerol) and sonicated to lyse the cells. The processing reaction was carried out as described by Han *et al.* [234]. Briefly, 30 μL of processing reaction contained 15 μL of whole-cell extract or the beads, 6.4 mM $MgCl_2$, 1 U/μL of Ribonuclease Inhibitor (Sigma), and the labeled transcripts of 1×10^4 to 1×10^5 cpm. The reaction mixture was incubated at 37 C for 90 minutes. RNA was extracted from the reaction mixture by phenol extraction and analyzed on a 10%-12.5% denaturing polyacrylamide gel.

2.2.28 *In vitro* Drosha processing Assay

HEK293T cells were transfected with pCK Drosha Flag or pCK Flag. The cells were harvested 48 hours post transfection and the pellet was suspended in buffer B. The cells were sonicated on ice and centrifuged at 13,200 rpm for 15 minutes at 4 C. The supernatant was incubated with 10 μL of anti-Flag antibody conjugated to agarose beads (anti-Flag M2 affinity gel, Sigma) with constant rotation for 90 minutes at 4 C. The beads were washed six times in buffer B, drained, and used for *in vitro* processing.

2.2.29 *In vitro* Dicer Assay

The *in vitro* transcribed sense (T7) or antisense (SP6) transcript of the purified and labeled $\sim$ 80 nt region was used for Dicer assay as per the manufacturers protocol (Ambion). An aliquot of the reaction was taken at different time points and resolved on a 15% Urea-PAGE. The gel was dried and autoradiographed.

2.2.30 Culturing of mammalian cells

The mammalian cells were maintained aseptically in DMEM supplemented with 10% FBS containing 2 mM L-Glutamine and concentration of 1 mg/ml Penicillin and Streptomycin(Sigma 100 mg/ml stock) and 1 mg/ml Amphotericin B (Fungizone, sigma 100 mg/ml stock)). The Cells were grown at 37 C at 5% CO_2 in water jacketed CO_2 incubator (Binder) until it is 90% confluent. Passaging of cells were carried out after removing the cells from the culture flask by trypsinisation (0.25% Trypsin and 0.02% EDTA in DMEM). The trypsinisation was arrested by the addition of complete DMEM containing and 10% FBS and the cells were removed and pelleted at 200 g for 3 minutes. The cell pellet was resuspended in appropriate amount of complete medium and replated in culture flasks.

2.2.31 Freezing of mammalian cells

The mammalian cells grown to confluency were washed once in PBS and trypsinised and pelleted. To the cell pellet the ice cold freezing mix (90% serum + 10% DMSO) was added drop by drop and immediately transferred to cryovials at a cell density of about 10^6 cells/Vial. The vials are stored at -80 C in a freeze boy for 2 days and then transferred to liquid nitrogen.

2.2.32 Thawing of mammalian cells

The frozen stocks were thawed after incubating them for 2-5 minutes in a circulating water bath set at 37 C. To the thawed cells 5ml of complete DMEM was added and spun down at 200 g for 3 minutes. The cell pellet was resuspended gently in complete DMEM and plated on to the tissue culture flasks.

2.2.33 Over expression studies in cell lines

The day before transfection, cells were plated upto 80% confluency. The medium was changed just 4 hours before transfection. *Mrhl* gene cloned in pCDNA 3.1 vector under CMV promoter was transfected into 80% confluent Gc1-Spg cells using lipofectamine 2000 as per manufacturer's instructions. The concentration of lipofectamine to DNA ratio is given below for different culture volume.

Surface area	Transfection volume	Lipofectamine concentration	DNA concentration
24well plate	100 μl	2 μl	0.8 μg
6well plate	500 μl	6 μl	3 μg
90mm Dish	3 ml	35 μl	15 μg

The DNA and lipofectamine were added in respective volume of DMEM without FBS and left for 5 minutes. After the incubation, the DNA and the lipofectamine mixtures were mixed together and incubated for 20 minutes for DNA lipofectamine complex formation. The complex was directly plated on the cells grown on culture dishes. The cells were harvested at various time points post transfection and RNA was extracted using TRIzol reagent.

2.2.34 Silencing experiments

The transfection of siRNAs was carried out as described above. Instead of plasmid constructs siRNAs were used for transfection with lipofectamine 2000. The concentrations and time point of each siRNAs were titrated to obtain optimum silencing efficiency. siRNA concentration of 100-150 nM were used for different experiments. The cells were harvested 48hours post transfection and RNA was

extracted using TRIzol reagent for checking the down regulation efficiency by RT-QPCR. For western analysis the cells were harvested after 60 hours post transfection and lysed in 1X RIPA buffer and loaded on SDS-PAGE gel for detection using specific antibodies.

2.2.35 *In situ* RNA hybridisation with LNA probes

The LNA probes specific for the *mrhl* RNA and the processed transcript were labeled with Cy3 and Cy5 (Exiqon) and hybridised *in situ* on mouse cell lines as described by Prasanth *et al.* [141] with some modifications. Briefly, the cells grown on the cover slips were washed once with PBS and fixed with 4% formaldehyde and 10% acetic acid for 10 minute at room temperature. The cells were dehydrated in 70% ethanol, subsequently rehydrated in PBS and permeablised in PBS with 0.01% Triton X-100. After washing twice in PBS, the cells were prehybridised for about 1 hour in solution containing 10% dextran sulphate, 2 mM vanadyl ribonucleotide complex, 0.2% BSA, 40 μg E.coli tRNA, 2X SSC, 50% formamide and 200 μg/ml of sheared single stranded salmon sperm DNA. The prehybridisation was followed by hybridisation with 10nM concentration of the probe in the same buffer at 50 C in a moist chamber for 8hours. The cells were then washed twice in buffer containing 50% formamide and 2X SSC followed by once in 50% formamide in 0.5X SSC and 0.1% SDS. The cover slips were mounted subsequently on 60% glycerol in PBS containing 10 μg/μl DAPI.

For immunofluorescence, the cells after hybridisation washes were once again washed with PBS and appropriate dilutions of the primary antibody (mouse monoclonal anti Nucleolin, Santa Cruz and rabbit polyclonal anti Drosha, a gift from

Prof. Narry kim) was added and incubated for 1 hour at room temperature in a moist chamber followed by three washes in PBS containing 0.1% Tween 20. This was followed by incubation with appropriate secondary antibody conjugated with Alexa fluor for 1 hour at room temperature in a moist chamber. The cells were washed thrice in PBS containing 0.1% Tween 20 and dried. The cover slips were mounted on 60% glycerol in PBS containing DAPI. The images were acquired in a LSM 10 Meta confocal microscope (Carl Zeiss) and images were analysed by image analysis software provided by Carl Zeiss.

2.2.36 Preparation of whole cell extracts from cultured cells

For harvesting, cells were placed on ice and the media was removed by aspiration. The cell monolayers were washed twice with 5 ml of ice cold PBS buffer. Cells were removed by scraping using a cell scrapper in 5 ml of ice-cold PBS. Five ml cell mixture was transferred to a sterile 15 ml centrifuge tube and centrifuged at 200 g for 5 minutes at 4 C. The supernatant was removed by aspiration and 1 ml of ice cold PBS was added and the cells were transferred to a 1.5 ml sterile centrifuge tube and centrifuged briefly at 200 g at 4 C. The supernatant was removed by aspiration and cells were lysed in 150 μl of Lysis solution (0.15 M Tris-HCl, pH 6.7/5% sodium dodecyl sulfate (SDS), and 30% glycerol) sonicated briefly and then diluted 1:10 in PBS-0.5% Nonidet P-40 (NP-40) supplemented with complete protease inhibitor (Sigma) and centrifuged at 16,000 g for 10 minutes at 4 C to remove cellular debris.

2.2.37 Northwestern blotting

The total, cytoplasmic or nuclear cell lysates that were separated on a 10% SDS polyacrylamide gel was transferred to nitocellulose membrane.The proteins were allowed to renature on the membrane by overnight incubation in the renaturation buffer containing 10 mM HEPES pH 7.9, 40 mM KCl, 5% glycerol, 0.2% NP40, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 5 mg/ml BSA. After renaturation the membrane was put in a RNA binding buffer (buffer C, 10 mM HEPES pH 7.9, 150 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 0.2 mM DTT, and 0.05% NP40 containing 50 μg/ml tRNA, 500 μg/ml Heparin) containing 50 000 cpm of probe prepared by *in vitro* transcription described previously. The membrane was incubated for 3 hours at 20 C and washed twice in buffer C. The membrane was then processed for phosphorimager.

2.2.38 siRNA mediated silencing of *mrhl* RNA and microarray analysis

The Gc1-Spg cells were grown a day prior to transfection to reach upto 70-80% confluency. Cells were then transfected with 100 nM *mrhl* siRNA pool (Dharmacon) using lipofectamine 2000 (Invitrogen). Cells were harvested 48 hours post transfection, and the RNA was isolated using the TriZol method. The down regulation of *mrhl* RNA expression was checked using RT QPCR method using specific primers against *mrhl*. For control experiments scrambled siRNA pool were used.

Total RNA after silencing of *mrhl* expression along with the control RNA from scrambled siRNA treated cells, of three independent experiments were subjected to microarray expression profiling using Affymetrix mouse 43.0 K expression arrays.

The entire protocol followed was as per the Affymetrix recommendation. The arrays after hybridization and washes were scanned and the preliminary analyses were done by using Array Assist software using PLIER algorithm. The genes were grouped based on P value and fold change. Majority of the analysis were performed using the genes that showed a P value of $\leq$0.05 and fold change $\geq$2. For some analysis the genes that showed P value of $\leq$0.05 and fold change of $\geq$1.5 were also taken. The biological functional analysis was done using DAVID EASE, GOTM. Hierarchical clustering was done using JAVA Tree View software.

Gotm: http://bioinfo.vanderbilt.edu/gotm/

David ease: http://david.abcc.ncifcrf.gov/

Java tree view: http://jtreeview.sourceforge.net/

2.2.39 RNA affinity pull down

The *in vitro* transcribed 2.4 kb RNA that was biotin labeled was denatured at 65 C and slowly cooled to room temperature to facilitate gain of optimum secondary structures. Freshly harvested Gc1-Spg cells were washed once in ice cold PBS and lysed in lysis buffer containing 10 mM Tris pH 8.0, 150 mM NaCl, 5 mm $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 0.5% Triton X 100 and 5% glycerol containing 1 mM PMSF and 1X Protease inhibitor cocktail (Sigma 1000X stock). The cell lysate was precleared for about 1 hour at 4 C in streptavidin agarose beads (Invitrogen) equilibrated in lysis buffer. The precleared cell lysate was incubated with *in vitro* transcribed and biotynalated RNA along with 0.5 units/μl RNAse inhibitor (Promega), 80 μg/ml tRNA for overnight followed by incubation with equilibrated streptavidin agarose beads for 2 hours. The beads containing the

RNA protein complex were washed thrice in the lysis buffer and the beads were directly boiled in SDS-gel loading dye and loaded on a 10% SDS PAGE gel.

2.2.40 S1 Aptamer pull down

For *in vivo* RNA pull down experiments, the S1 aptamer approach as described by Srisawat and Engelke [235] was followed. The S1 aptamer sequence was ligated to the 3' end of 2.4 kb *mrhl* by sticky end ligation after digestion with XhoI enzyme. The pCDNA 3.1 vector containing *mrhl* was treated with calf intestinal alkaline phosphatase (CIAP) and was purified by phenol:chloroform extraction followed by ethanol precipitation. S1 aptamer digested with XhoI was purified and ligated to CIAP treated pCDNA *mrhl* downstream of CMV promoter. The orientation of the clones was checked by sequencing. The construct was transfected into Gc1 Spg cells and 36 hours post transfection, the cells were harvested and lysed in lysis buffer. The cell lysate along with RNAsin and tRNA were incubated with equilibrated streptavidin beads for overnight and the washes were performed as described above. The RNA bound proteins were subjected to 10% SDS PAGE analysis.

2.2.41 Mass spectrometry

The SDS PAGE gel after separation of proteins was stained with EZ blue (Sigma) or silver stained. The bands that were present were excised along with control where no proteins were present. In gel tryptic digestion was performed according to protocol provided by the Bruker. The tryptic digest was analyzed in Bruker mass spectrometer. MALDI TOF was done initially to identify the

peptide mass fingerprint. Specific peaks were lifted further to perform MS/MS analysis to identify the peptide sequence. Proteins were identified using MASCOT software(http://www.matrixscience.com/). Carbamidation was set as fixed modification and the methionine oxidation was set as variable modification. The threshold cut off was set as 0.6 kDa. Proteins for which at least 2 peptides matched were considered to be significant.

2.2.42 Cellular fractionations

The proteins from nuclear and cytoplasmic fractions were prepared by using NE-PER reagent of Pierce as per the manufacturer's protocol. The proteins isolated from each of the fractions were subjected to western or immunoprecipitation using appropriate antibodies.

2.2.43 Immunoprecipitation

Immunoprecipitation experiments were performed as described below. Briefly, cells lysed in IP buffer (20 mM Hepes-KOH pH 7.4, 100 mM NaCl, 10% glycerol, 2 mM DTT, 5 mM $MgCl_2$, 0.1 mM PMSF 0.5% Triton X) was precleared and incubated with primary antibody (rabbit poly clonal p68) or rabbit pre immune sera for 4 hours, followed by incubation for 2 hours with protein A agarose. Immunoprecipitated proteins were separated in 10% SDS PAGE gel and transferred to nitrocellulose membrane. The membrane after blocking in 5% skim milk in PBS was incubated with primary antibody for 4 hours to overnight at 4 C. The membrane was then washed thrice in PBST followed by incubation in HRP conjugated

secondary antibody. The blot after washes was subjected to chemiluminiscence using luminol as substrate. For detection of phospho proteins TBS was used instead of PBS.

2.2.44 Immunofluroscence

For immunofluorescence experiments, the cells that were directly grown on cover slips were washed with PBS. The cells were fixed in 1% formaldehyde and 10% acetic acid mix for 10 minutes, followed by washes in PBS. The cells were blocked in PBS containing 1% BSA for 45 minutes in a moist chamber. Appropriate dilutions of the primary antibody (rabbit polyclonal β-catenin from Upstate or rabbit polyclonal p68 from Novus Biologicals were added and incubated for 1 hour at room temperature in a moist chamber followed by three washes in PBS containing 0.1% Tween 20. This was followed by incubation with appropriate secondary antibody conjugated with Alexa fluor for 1 hour at room temperature in a moist chamber. The cells were washed thrice in PBS containing 0.1% Tween 20 and dried. The cover slips were mounted on 60% glycerol in PBS containing 1 μg/ml DAPI. The images were acquired in a LSM 10 Meta confocal microscope (Carl Zeiss) and analysed by image analysis software provided by Carl Zeiss.

Chapter 3

Characterization of the non-coding *mrhl* RNA

3.1 Introduction

An earlier study from our laboratory involved a detailed characterization of a mouse meiotic recombination hotspot locus mapping to mouse chromosome 8CD (Nishant *et al.* 2004) [229]. Transcriptional analysis of this region as probed by a northern analysis for the total RNA from testis and liver revealed a positive signal to one of the contigs at 2.4 kb region. The *in silico* analysis of this transcript did not reveal any open reading frame that was greater than 100 base pairs. The largest ORF also did not seem to have any Kozak sequences at its immediate 5′ proximal region. Thus it was concluded that this RNA could be a non-coding transcript (Figure 3.1). The RNA was then subjected to secondary structure prediction analysis by Mfold program and found to have considerable propensity to form stable secondary structure at 37 C at physiological ionic strength (Figure

3.2). Expression of this RNA was also studied by northern analysis on multiple tissues, which showed that the *mrhl* RNA is expressed only in testis, liver, kidney and spleen and it is absent in brain, lungs, heart, and skeletal muscles (Nishant *et al.* 2004) (Figure 3.3) [229].

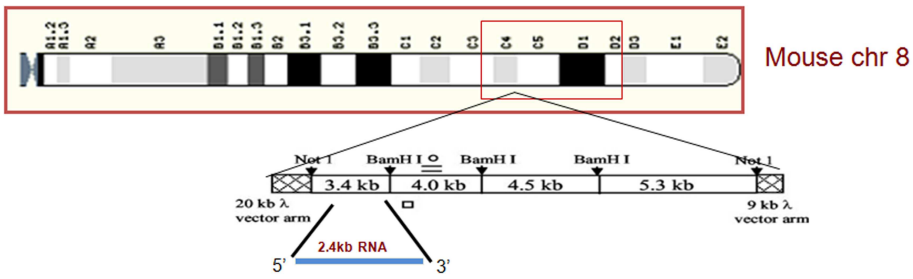

Figure 3.1: *Identification of 2.4 kb mrhl RNA. Chromosomal position of 17.2 kb mouse meiotic hot spot locus from mouse chromosome 8 and the contig map (Nishant et al. [229]).*

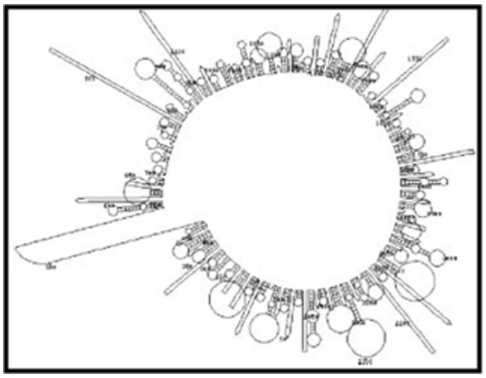

Figure 3.2: *Secondary structure prediction of mrhl RNA given by MFold program. The folding was performed at 150 mM salt at 37 C. Optimum structure with maximum δG is given.*

The bioinformatic analysis of the *mrhl* proximal promoter region showed the presence of TATA box at -35 position and a CAAT box at -155 position. A polyadenylation signal sequence was also found with a single nucleotide variation

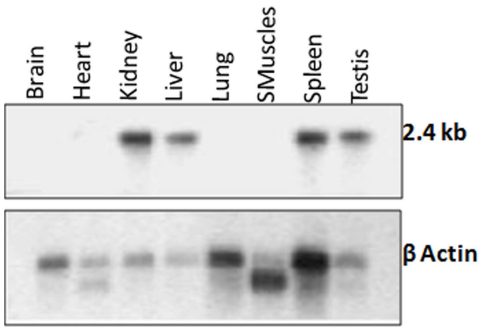

Figure 3.3: *Multiple tissue northern blot. Northern hybridization of polyadenylated RNA obtained from mouse multiple tissues using full length mrhl probe. Positive hybridization signal is seen in Testis, spleen, kidney and liver where as the RNA is absent in other tissues. β actin in the bottom pane was the loading control.*

at its 3′ end. Further downstream is the meiotic DNA repair site that was identified previously in our laboratory [236]. The characteristics promoter features and polyadenylation of RNA shows that the *mrhl* is a putative polymerase II transcribed product. *Mrhl* is also an unspliced RNA and lacks any introns. The schematic of the proximal promoter elements is described in Figure 3.4.

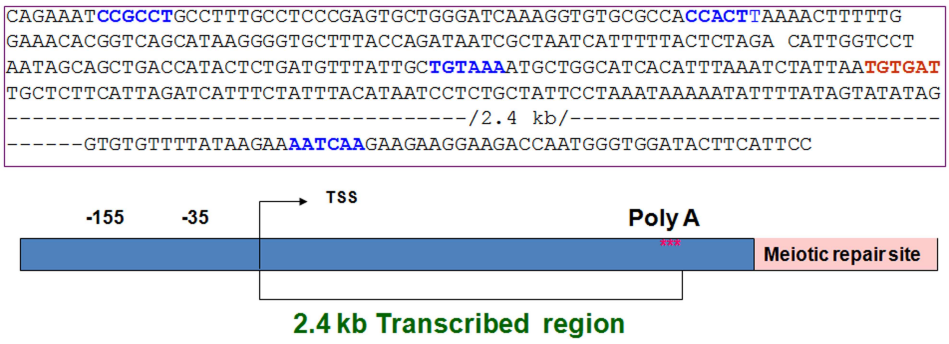

Figure 3.4: *Proximal regulatory elements of mrhl RNA. The proximal promoter elements of mrhl gene such as the TATA box, CAAT box and the polyA signal sequences are represented in nucleotide sequence (top) as well as the schematic (bottom). The meiotc repair site downstream of the mrhl gene is also highlighted.*

3.2 Cloning of *mrhl* in pGEM 3Zf+

The full length *mrhl* RNA was cloned from mouse testis cDNA. The *mrhl* was amplified using specific adapters containing BamHI and XhoI site which was cloned in the pGEM3Zf+ vector that was linearised with the enzymes BamHI and XhoI. The recombinant clones that were grown in ampicillin containing LB plate were confirmed by restriction digestion and PCR amplification. Figure 3.5 is the restriction analysis confirmation and PCR amplification of full length *mrhl* from pGmrhl plasmid. This plasmid was used for many of the downstream experiments such as *in vitro* transcription and transcription coupled translation etc.

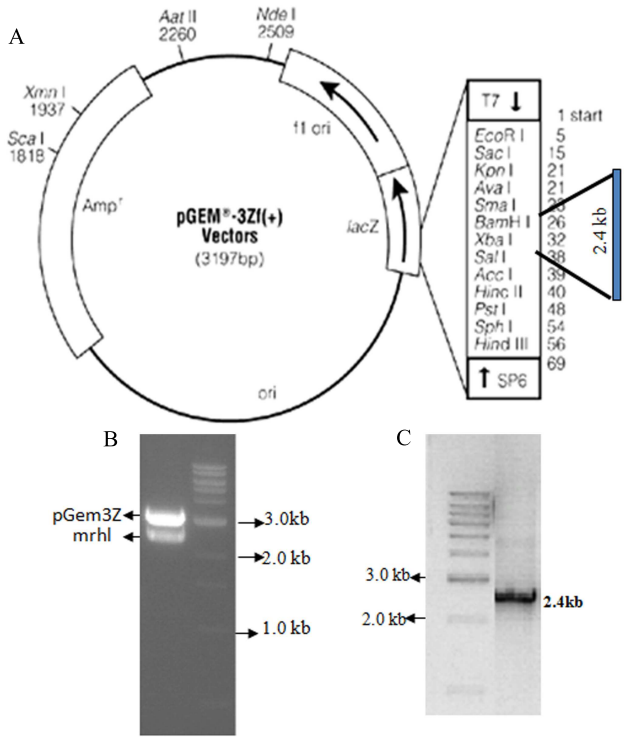

Figure 3.5: *Cloning of mrhl gene in pGEM vector. (A) The vector map of pGEM3Zf+ showing the restriction sites in the MCS which was used to clone 2.4 kb mrhl gene. (B) Restriction digestion confirmation of the clone. (C) PCR amplification of mrhl from the plasmid pGmrhl giving a specific amplicon of 2.4 kb.*

3.3 *Mrhl* RNA does not code for protein

In order to ensure that the RNA does not code for any protein, an *in vitro* coupled transcription and translation assay was carried out. The 2.4 kb non-coding RNA gene that was cloned under a T7 promoter was used for the reaction in the presence of ^{35}S methionine using rabbit reticulolysate system. As shown in Figure 3.6A, there was no protein product detected in the lane of pGmrhl, where as the control plasmids that harbour luciferase and the mouse poly-pyrimidine tract binding protein gave respective protein products. The negative control that contained only empty vector also did not give any protein product. In order to verify that the gene is indeed transcribed full length in the *in vitro* transcription reaction, a RT-PCR was performed to amplify the *mrhl* transcript by isolating RNA from the same reaction. Figure 3.6B shows the PCR amplification of the gene product, which shows that the 2.4 kb RNA gene is indeed transcribed but not translated in the *in vitro* coupled transcription and translation experiment. A no RT reaction did not give any amplification, eliminating the possibility of the plasmid getting amplified in the reaction.

3.4 *Mrhl* is poorly conserved across species

The conservation of *mrhl* RNA across many different species was performed by BLAST against the different genome databases and significant homology of *mrhl* was found only in the rat genome. In rat, the gene is located on chromosome 19 and the gene shares about 80% homology to the mouse gene, and most interestingly was absent in the human genome (Figure 3.7). An extensive bioinformatics search was

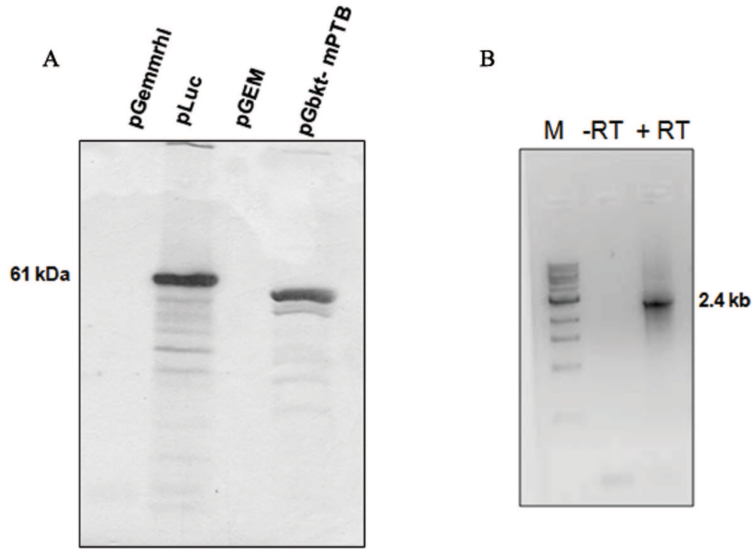

Figure 3.6: *In vitro transcription coupled translation. (A) Lane 1, in vitro transcription coupled translation assay of pGmrhl (2.4 kb RNA gene cloned into pGEM3Z vector downstream of T7 promoter); lane 2, positive control luciferace encoding 61 kDa protein; lane 3, empty vector negative control and lane 4, pGbktT7-mPTB encoding 58 kDa mouse poly pyrimidine tract binding protein. (B) RT-PCR of RNA isolated from the in vitro transcription coupled translation of pGmrhl in the presence and absence of RT. M is the 1kb DNA ladder.*

carried out to identify the homologs of *mrhl* RNA in other species by looking at the syntenic region across different organisms. In mouse, the *mrhl* RNA is adjacent to the phosphorylase kinase beta (phkb) gene on chromosome 8. Analysed up to 100 kb region flanking the phkb gene was carried out in various species using low stringency BLAST and also by using BLAST two sequence options. The extensive search also did not identify any significant homologs in any other species except in the case of rat genome.

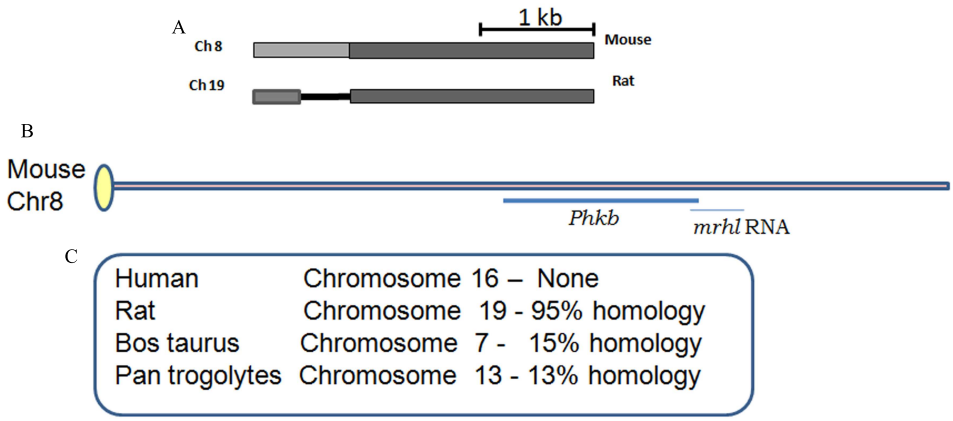

Figure 3.7: *Mrhl is poorly conserved across species. (A) Schematic map of homology between mouse and rat mrhl gene. (B) Position of mouse mrhl gene with respect to its syntenic region harbouring phkb gene from mouse chromosome 8. (C) Percentage of conservation of the mrhl and the syntenic region across different species.*

3.5 The 2.4 kb *mrhl* RNA is nuclear localised

The sub cellular localization pattern of the *mrhl* RNA was then investigated to study the preferential localization of *mrhl* in cellular compartments. The testis and liver tissues from the mouse were fractionated to nucleus and cytoplasm by differential centrifugation and the RNA was isolated. RT-PCR was done subsequently using the gene specific primers for the non-coding RNA. Figure 3.8 shows the result where the amplification of *mrhl* RNAis seen in total RNA as well as in the nuclear fraction but not in the cytoplasmic fraction. The β actin mRNA was amplified in all the fractions as probed by single exonic primers. The U1 snRNA was used to assess the purity of the nuclear and cytoplasmic fractions. It was amplified only in the nuclear fraction but not in the cytoplasm. A no RT control was also performed in order to eliminate any false amplification that might arise due to DNA contamination. Thus the *mrhl* RNA appears to be nuclear restricted and does not get transported into the cytoplasm. The nuclear localization of *mrhl*

RNA was not surprising, as many non-coding RNAs have been shown earlier to exhibit definite nuclear localization pattern mediating potent regulatory events such as regulation of coding genes (CTN) [141], imprinting (H19, Air) [110, 113], and dosage compensation (Xist, roX) [97, 108, 109]

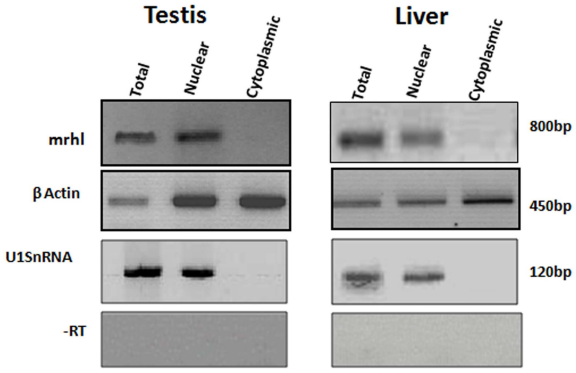

Figure 3.8: *Mrhl is a nuclear restricted RNA. RT PCR of the mrhl RNA from testis and liver total, nuclear and cytoplasmic RNA. The mrhl RNA was found to be restricted to nuclear fraction and absent in cytoplasm. β actin was used as positive control. U1 snRNA was also used as a control to assess the purity of fractions. A no RT reaction was used as a negative control to eliminate any false amplification from DNA.*

3.6 The 2.4 kb *mrhl* RNA is processed further to an intermediate RNA species

The 2.4 kb *mrhl* RNA showed an extensive secondary structure comprising of stem loop and hairpin loops (Nishant *et al.* 2004) [229]. It is now well established that many of the microRNAs are transcribed as a primary transcript by polymerase II. This primary transcript posseses stem loop structure that will be recognized and processed by nuclear RNase III enzyme Drosha in the nucleus and subsequently to a 22 nt miRNA by Dicer in the cytoplasm [50, 53, 55]. Hence, it was examined

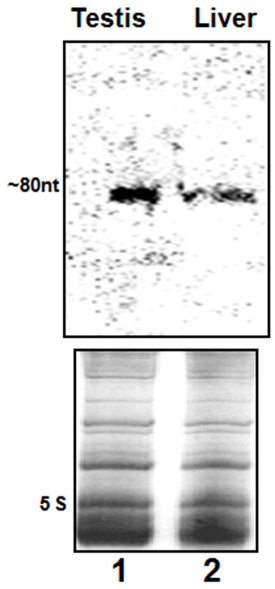

Figure 3.9: *Mrhl is processed to 80 nt intermediate RNA. Small RNA northern blot of total RNA separated in 15% polyacrylamide-urea gel probed with 2.4 kb full length gene (probe D) by Klenow labeling using random primers on mouse testis, (lane 1) and liver total RNA, (lane 2). Bottom panel shows ethidium bromide stained pattern for loading control.*

whether the 2.4 kb non-coding RNA also gets processed further. For this purpose total RNA was separated in 15% polyacrylamide urea gel and northern blot was carried out on both testis and liver RNA. The blot was hybridised using body labelled probe against purified 2.4 kb PCR product. Interestingly, a signal from both the testis as well as the liver RNA at approximately 80 nt position was observed (Figure 3.9).

In order to narrow down the region that actually generates the 80 nt product we amplified different regions of the non-coding RNA gene and northern analysis was carried out using each of them as a probe against total RNA extracted from liver and testis. Figure 3.10A shows the different primer positions used in generating the amplicons of different regions of the transcript. Figure 3.10B shows the

autoradiogram of such a northern analysis. The blots showing the hybridisation signal are derived from the amplicons that shared 380bp common region that has been highlighted in Figure 3.10A. Hybridisation with probe E failed to show any signal, as it does not overlap with the small region identified and represented the far end of the gene. The signal that arises due to the presence of 380bp common region was further confirmed by using that region (probe A) as a probe (Figure 3.11). Thus the region that gave signal was mapped to 380bp from the original 2.4 kb.

We further narrowed down the region corresponding to 80 nucleotides by carrying out northern blot analysis by using amplicons of 100bp fragment each. The shorter regions from the 380bp fragment used for northern is shown in Figure 3.12A and the sequences in Figure 3.12B. The specific region of 96 bp was identified as the one giving rise to the signal in the northern analysis (Figure 3.12C). However, one cannot rule out the possibility of few molecules of the 80 nt processed RNA originating from other regions of the primary transcript that fall below the level of detection by hybridisation. Additional northern hybridization experiment was carried out using the identified small region by using radiolabelled RNA probe transcribed either in sense or antisense direction. This would ensure that the signal in the northern was truly arising from the same strand as the primary transcript. Figure 3.13 clearly shows that the hybridization occurs with the antisense transcript as a probe but not with the sense probe confirming that *in vivo* the 80 nt RNA is of the same strand and is the processed product of the primary transcript.

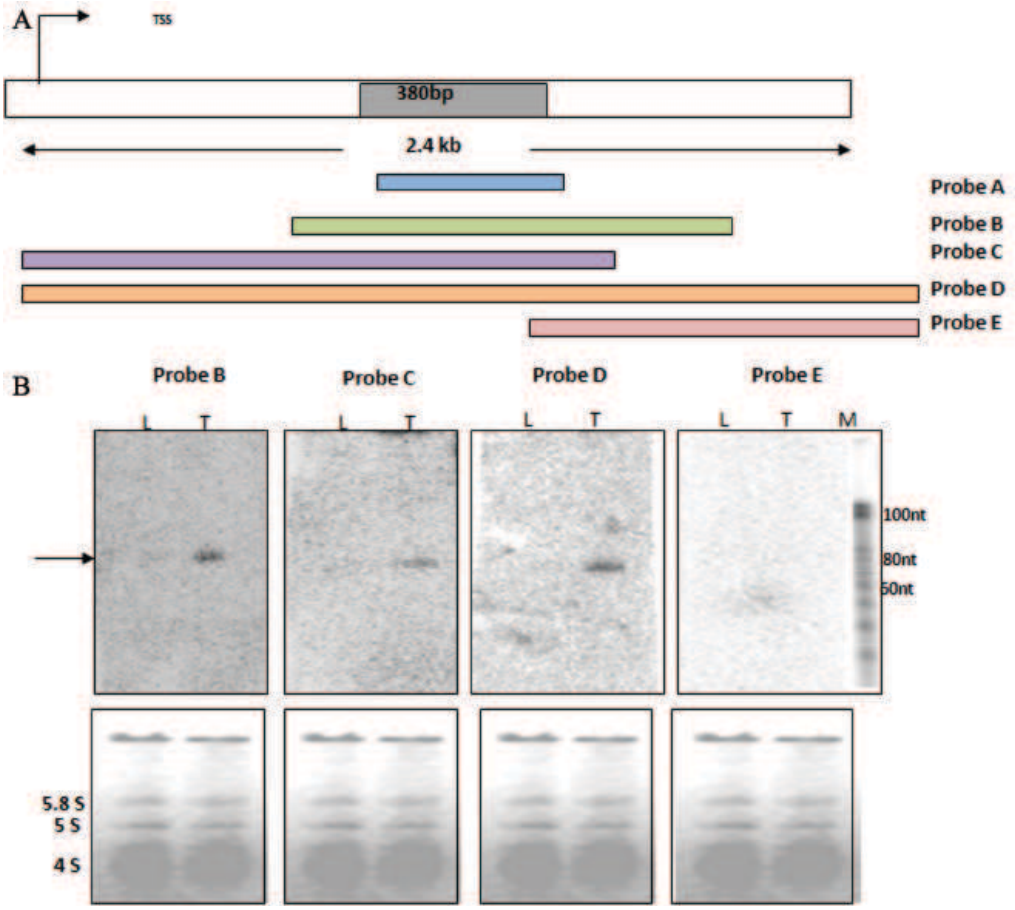

Figure 3.10: *(A) Scheme of different regions of mrhl used for probe generation and its position. The highlighted region represents the sequence shared between probes B, C, D giving positive signal. (B) Northern blot on liver, (lane 1) and testis RNA (Lane 2) separated on 15% polyacrylamide Urea gel using each of the probes B, C, D, E. Bottom panel shows the ethidium bromide stained pattern for loading control.*

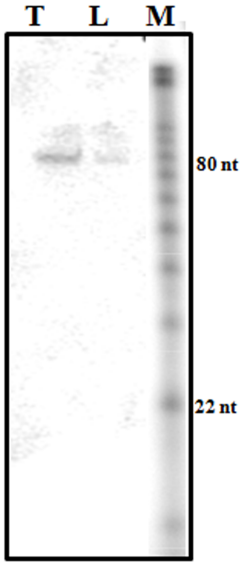

Figure 3.11: *Small RNA northern blot using probe A spanning the 380bp region that corresponds to the hybridization signal. Lane 1 and 2 corresponds to testis and liver RNA respectively. Bottom panel is the ethidium bromide stained pattern. M is the RNA ladder (Ambion).*

3.7 Drosha mediates processing of 2.4 kb *mrhl* RNA

The microRNAs that are 22 nt RNA molecules regulating mRNA expression post transcriptionaly are derived from a pre miRNA which are about 60 to 80 nt in length by the action of Dicer [50, 237]. The pre miRNA itself is a product of cleavage of larger transcript by the enzyme Drosha [52] The basic requirements in a particular RNA species to be processed by Drosha have been identified. The RNAs to be processed by Drosha often fold back to give a stem loop structure with a large stem, an internal bulge and a large loop region. Apart from this, Drosha also requires flanking non-structured RNA for its cleavage [238]. Interestingly the sequence of 2.4 kb when folded using a Mfold program also generated a stem

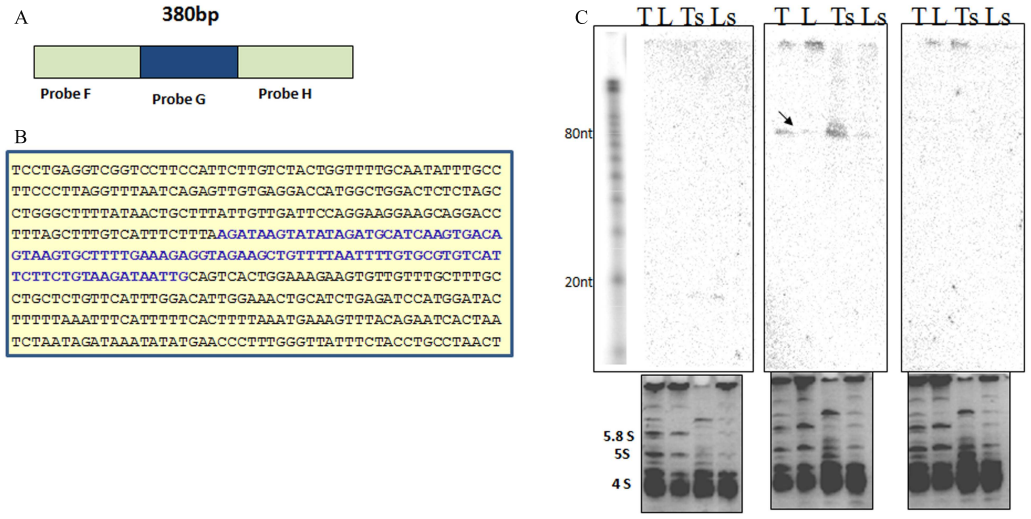

Figure 3.12: *Northern of narrowed down transcript from mrhl. (A) Probes used for identification of ∼ 80 nt region within the narrowed down 380 bp region (Figure 3.11A). (B) Small RNA northern blot of total RNA from testis (lanes 1, 5 and 9), liver (lanes 2, 6 and 10) and small RNAs enriched (Ambion miRNA isolation kit) from testis (lanes 3, 7 and 11) and liver (lanes 4, 8 and 12) using the probes F, G and H. Bottom panel represents ethidium bromide stained pattern. (C) Sequence of the 380bp region that was narrowed down from the 2.4 kb region is shown and highlighted is the 90nt region that gave the signal in northern blots.*

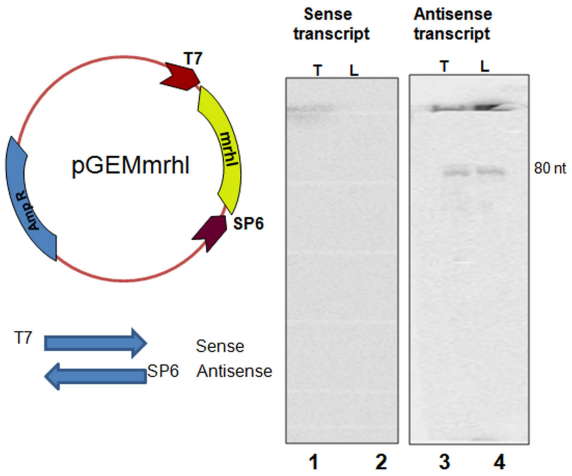

Figure 3.13: *The 80 nt intermediate is derived from sense strand. Northern blot on mouse testis and liver RNA by using sense probe (lane 1, 2) and antisense probe (lane 3, 4) derived from T7 and SP6 promoters respectively. The positive signal was obtained only in antisense probe showing that in vivo the RNA is present in sense orientation.*

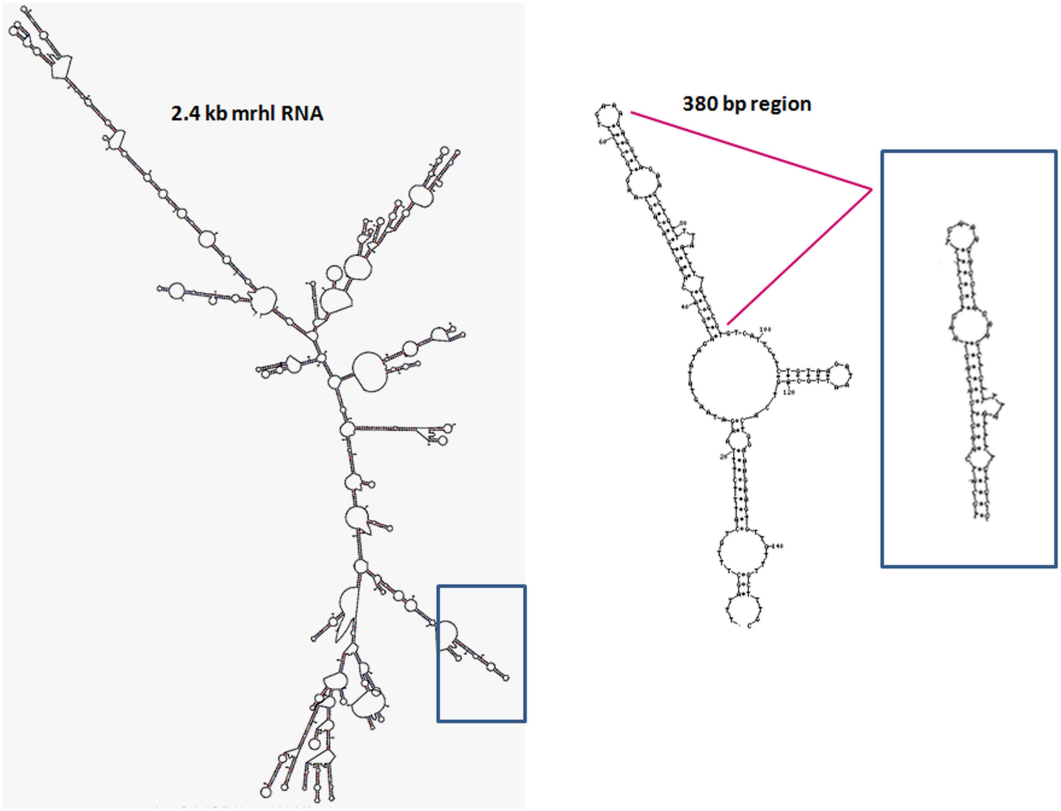

Figure 3.14: *The 80 nt intermediate RNA assumes stem loop structure. (A) Secondary structure predicted for 2.4 kb full length mrhl RNA. (B) Secondary structure of the 380 nt region showing stem loop structures predicted by Mfold that maps to the region corresponding to hybridization signal. The inset is the ∼ 80 nt region that possess the stem loop structure which is also highlighted in the 2.4 kb full length transcipt.*

loop structure with a large stem and an internal bulge and a loop region along with flanking unstructured RNA (Figure 3.14). This region of the sequence also corresponded to the ∼ 80 nt fragment that we detected *in vivo* as shown in Figure 3.14B suggesting that the 2.4 kb transcript may serve as a substrate for the Drosha machinery.

Subsequently, it was examined, whether the 2.4 kb transcript can generate the 80 nt fragment in an *in vitro* assay using a cell free system. For this purpose the *in vitro* processing assay was performed according to the method described

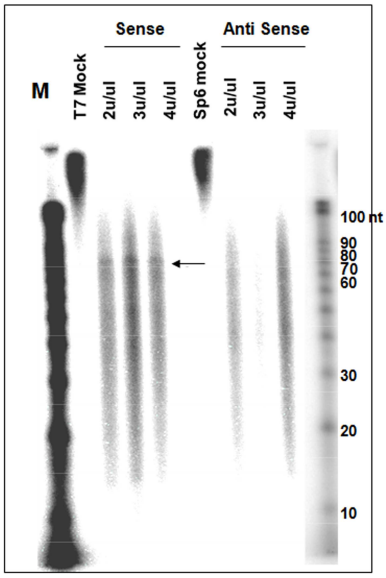

Figure 3.15: *MicroRNA processing assay. RNA processing assay of the 2.4 kb transcript derived from T7(sense) and SP6 (antisense) promoters using total testicular cell lysate. Lanes 2 and 6 represent the primary 2.4 kb transcript obtained from T7 or SP6 polymerase mediated transcription respectively. Processing assay was carried out with 2 (lanes 3 and 7) 3 (lanes 4 and 8) and 4 units (lanes 5 and 9) of RNasin/µl. The processed ∼ 80 nt RNA was seen only in the lanes 3, 4 and 5 containing the sense transcripts. M represents the molecular size marker.*

for miRNA processing by Han *et al.* [234]. The gene cloned between the T7 and SP6 promoters in pGmrhl was transcribed in both sense and antisense orientation and were incubated with whole cell extract from testis. The processed RNA was separated on a 15% polyacrylamide gel containing 7M urea. The results presented in Figure 3.15 show an ∼ 80 nt processed RNA band was seen only from the sense transcript but not from the antisense transcript.

After establishing that the 2.4 kb transcript can be processed to generate an ∼ 80 nt fragment in a cell free extract, it was further examined whether this can be demonstrated with Drosha (RNasen in mammals). For this purpose, an *in vitro* Drosha assay was carried out in which the plasmid pCK harboring Drosha Flag was

transfected into HEK 293T and the Drosha flag complex was pulled down using anti Flag agarose beads. It was essential that Drosha complexed with another RNA binding protein DGCR8 was used, since the primary miRNA cleavage is brought about by Drosha-DGCR8 complex also known as microprocessor complex. The beads containing the complex were directly used for Drosha assay using *in vitro* transcribed sense or antisense *mrhl* 2.4 kb RNA as the substrate. A western blot analysis using anti Drosha antibody confirmed the presence of Drosha in the input cell lysates of Flag Drosha transfected cell line (Figure 3.16A). The mock transfected cell line also showed the presence of endogenous Drosha. Western blot analysis on the immunoprecipitated beads using anti Flag antibody detected the Drosha Flag fusion protein at 140 kDa, only in the immuno pull down of pCK Drosha Flag transfected cell line but not in the mock pCK flag transfected cell line. Figure 3.16B shows the autoradiogram of processing assay and it can be seen that, only sense transcript is cleaved by the Drosha complex to generate the $\sim$ 80 nt fragment but not the antisense transcript. As a control, mock-transfected cells were also used, which did not show any processed $\sim$ 80 nt RNA product.

In order to unequivocally demonstrate that Drosha machinery is involved in the processing of 2.4 kb *mrhl* RNA primary transcript *in vivo*, siRNA approach was employed to down regulate Drosha and look for the $\sim$ 80 nt processed intermediate. This experiment was carried out on Gc1 Spg cell line of mouse spermatogonial cells. Figure 3.17A shows the RT PCR analysis of the 2.4 kb RNA from the Gc1-Spg cell line in the presence and absence of RT where the signal was seen only in the RT lane demonstrating that this cell line does express the 2.4 kb non-coding RNA. A northern blot of RNA from the same cell line was also done in order to confirm the presence of processed $\sim$ 80 nt intermediate RNA. As can be seen

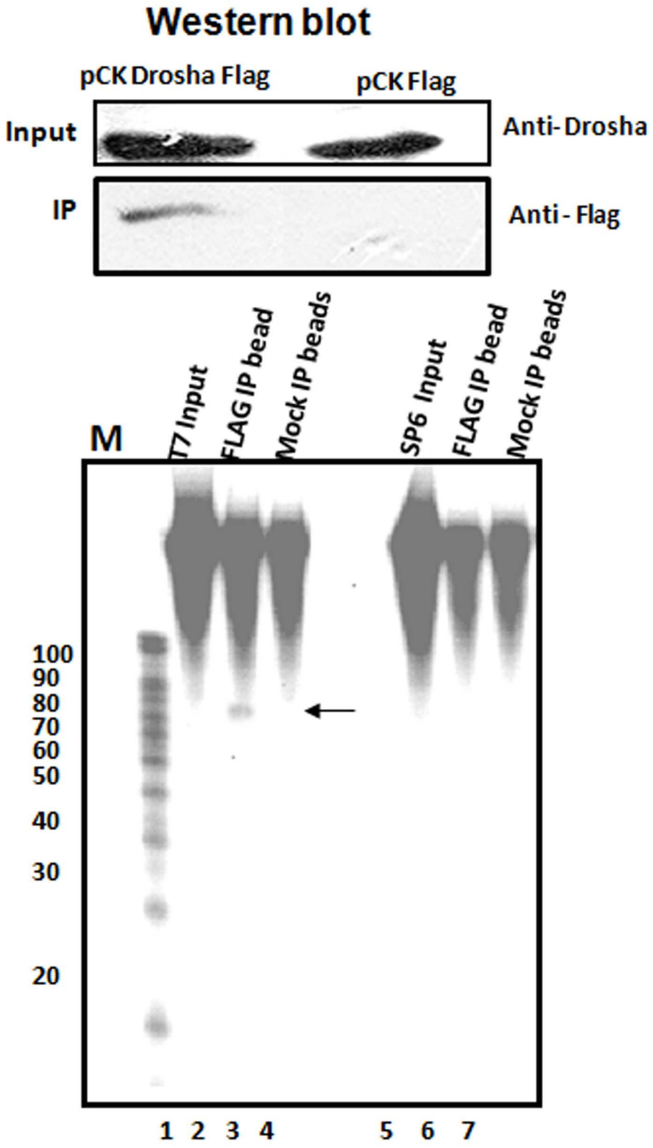

Figure 3.16: *In vitro Drosha assay. Flag IP was performed from the HEK 293T cells transfected with pCK-Drosha - Flag construct. Mock transfection contains the empty vector pCK Flag without the insert. (A) The upper panel represents the western blot of the input cell lysates using Drosha antibody while the bottom panel represent western blot of Flag IP beads using flag antibody. Flag Drosha fusion protein at 140 kDa is seen only in the IP beads from pCK drosha Flag transfected cell lysate but not from pCK Flag transfected cell lysate. (B) Processing assay using the Flag immunoprecipitated beads containing Drosha complex. Lanes 2 and 5 represent the input T7 and SP6 polymerase derived RNA from 2.4 kb primary transcript. Mock IP lanes 4 and 7 represent the processing reaction carried out using mock-transfected beads. Lanes 3 and 6 represent the reaction carried out with Flag IP beads. The products were separated on 15% Urea PAGE and autoradiographed. M is the 10 nt RNA ladder.*

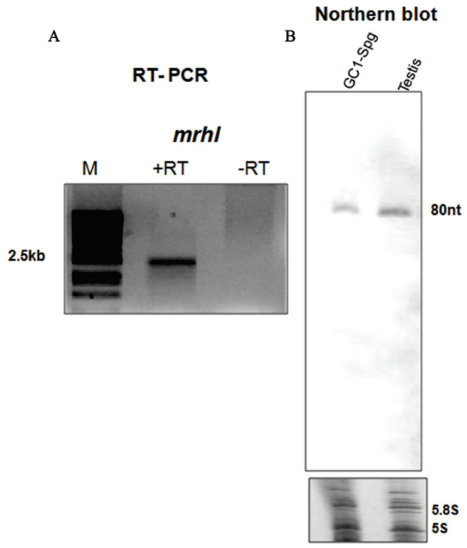

Figure 3.17: *Expression of mrhl in mouse Gc1-Spg cell line. (A) RT-PCR of mrhl RNA from mouse Gc1-Spg cell line in the presence and absence of RT. (B) Small RNA northern blot on mouse Gc1-spg cell line (lane 1) and RNA isolated from 10 day old mouse testis (lane 2) probed with probe G. Bottom panel is the ethidium bromide stained pattern.*

in the Figure 3.17B the 80 nt RNA is detected in mouse Gc1 Spg cells (lane 1). The $\sim$ 80 nucleotide processed RNA intermediate was also observed in the testicular cells from 10 day old mouse containing predominantly spermatogonial cells (lane 2). The results of the siRNA experiment using Gc1-Spg cell line is shown in Figure 3.18. The densitometric analysis showed that, Drosha was down regulated to almost 70% after 60 hours post transfection. At the same time the $\sim$ 80 nt processed intermediate was also reduced to approximately 66% to that of the control. Under these conditions of Drosha silencing, it was also observed that there is no change in the 2.4 kb transcript level. Thus, this experiment demonstrates Drosha's involvement in the processing of 2.4 kb *mrhl* RNA to $\sim$ 80 nt intermediate RNA.

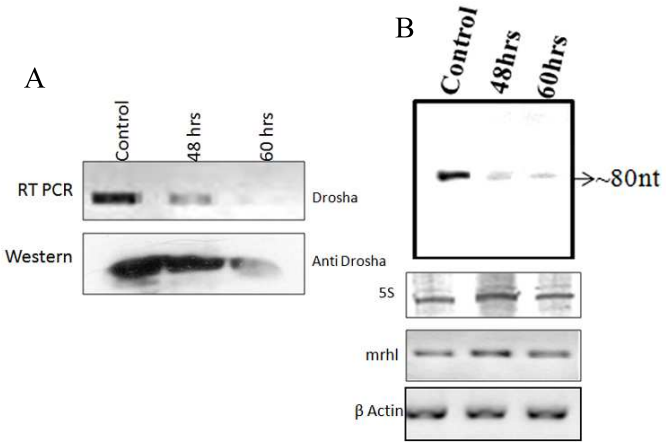

Figure 3.18: *Impaired processing of mrhl primary transcript in Drosha silenced condition. (A) Drosha silencing was confirmed both by RT PCR using specific drosha primers (top panel) and western blot using anti Drosha antibodies showing reduced Drosha RNA and protein respectively by 48 hours and 60 hours of post transfection with Drosha siRNA as against lipofectamine treated control. (B) Small RNA northern blot on Gc1 Spg total RNA probed for 80 nt RNA after Drosha down regulation. Bottom panel (5S RNA) shows the ethidium bromide stained gel pattern that was used as loading control. The bottom 2 panels show the RT PCR analysis of* mrhl *and βactin using specific primers.*

3.8 The 22 nt mature miRNA is not generated *in vivo*

The above experiments have shown that the 2.4 kb *mrhl* RNA is processed into a $\sim$ 80 nt nucleotide intermediate by the Drosha enzyme machinery. As mentioned earlier in the miRNA processing pathway the 60-80 nt intermediate RNA is further processed into small 20-22 nt miRNA species by the Dicer machinery. It is interesting to note that, the mature small 22 nt RNA was not observed in the small RNA northern blot described in previous experiments indicating that *in vivo* this RNA may not be processed towards miRNA species. As a positive control presence of Let-7 miRNA and Mmu miR1a were probed in the RNA samples and the result shown in Figure 3.19 also confirms that there is no technical problem

in detecting the mature microRNA in our northern blot experiments. The same blot was stripped and reprobed using *mrhl* probe, where again only the $\sim$ 80 nt intermediate RNA was detected but not the 22 nt mature miRNA. As mentioned earlier the entire process of miRNA maturation happens partially in the nucleus where Drosha is present and precursor to the mature miRNA takes place with the aid of cytoplasmic Dicer. The localization of the $\sim$ 80 nt RNA in sub cellular fractions was examined. The nuclear and the cytoplasmic fractions were separated from testis and liver tissues as described previously and a small RNA northern analysis was performed by the same probe that was used to detect the $\sim$ 80 nt RNA. Again only the 80 nt RNA was detected in the nuclear compartment but not in the cytoplasm (Figure 3.20). Some additional bands were also observed in 150 nt size region in the nuclear fraction which is presumed to be the partially processed primary transcript in the nucleus. Thus the 80 nt intermediate RNA is also nuclear restricted and not being transported to cytoplasm.

In order to confirm that the 22 nt is not been generated *in vivo* we resorted to the recently described sensitive technique of splinted ligation method [233]. The results of this experiment described in Figure 3.21 indicate that the positive control Let 7 is detected by the splinted ligation method with 2μg of total RNA, where as the *mrhl* RNA was not detected even at 4 μg of RNA. Two bridge oligos spanning each of the two complementary strands of the $\sim$ 80 nt *mrhl* small RNA were used as we were not sure of the mature miRNA strand that could be generated *in vivo*. The negative control experiment which did not have either ligase, or RNA or the Bridge oligo did not give rise to any signal.

Furthermore, over expression of the 2.4 kb *mrhl* RNA in the Gc1-Spg cell line under CMV promoter was also performed and tried to detect the 22 nt RNA,

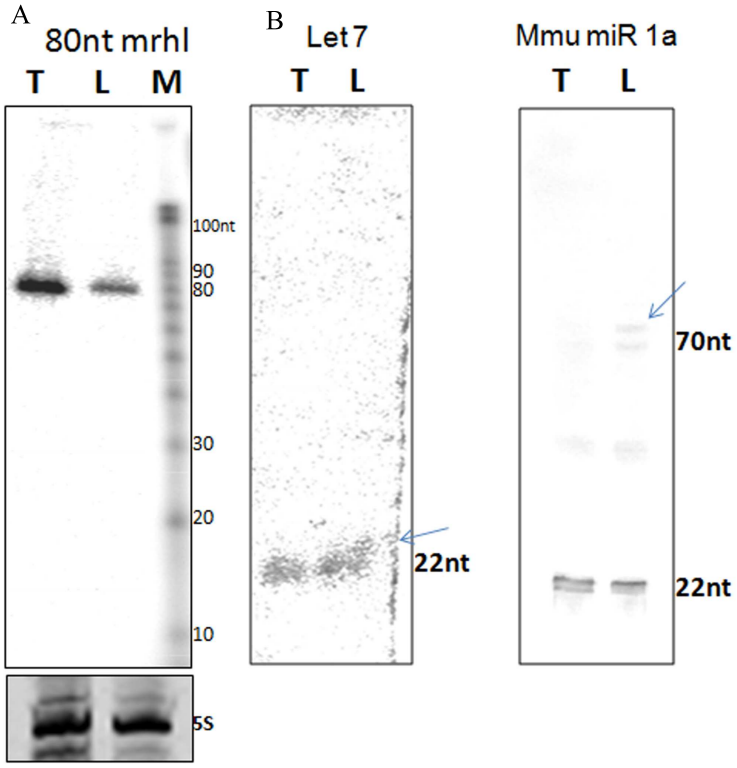

Figure 3.19: *Expression of mature miRNA. (A) Small RNA northern blot using probe that corresponds to 80 nt intermediate region. Lane 1 and 2 corresponds to testis and liver RNA respectively. M is the RNA ladder (Ambion). The 22 nt mature miRNA was detected using mrhl probes. (B) Positive control northern hybridization with Let-7 and Mmu miR1a gave respective 22 nt signal using specific probes.*

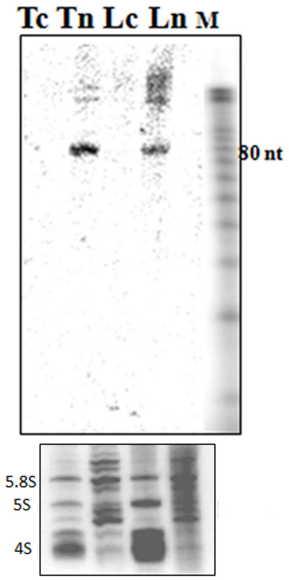

Figure 3.20: *The 80 nt intermediate RNA is also nuclear restricted. Northern blot on the RNA isolated from nuclei and cytoplasm of testis (Tn,Tc)and liver tissues (Ln , Lc) with probe G showing signal only in the nuclear lanes but not in the cytoplasm. The bottom panel is the ethidium bromide stained gel pattern.*

so that the primary transcript is available in abundance to generate detectable quantity of the mature miRNA. The cells were harvested at 12 hr, 24 hr and 36 hrs post transfection with pCDNA *mrhl* construct and carried out northern blot using $\sim$ 80 nt region specific probe (probe G). As seen from Figure 3.22, though there is a marginal increase in 80 nt precursor form by 24 hours and 36 hourrs post transfection, the 22 nt mature RNA could not be observed as monitored by northern analysis which also confirms our conclusion that the 22 nt may not be formed in the *in vivo* scenario. The expression pattern of the 2.4 kb *mrhl* primary transcript under the over expressed conditions was also monitored by RT PCR.

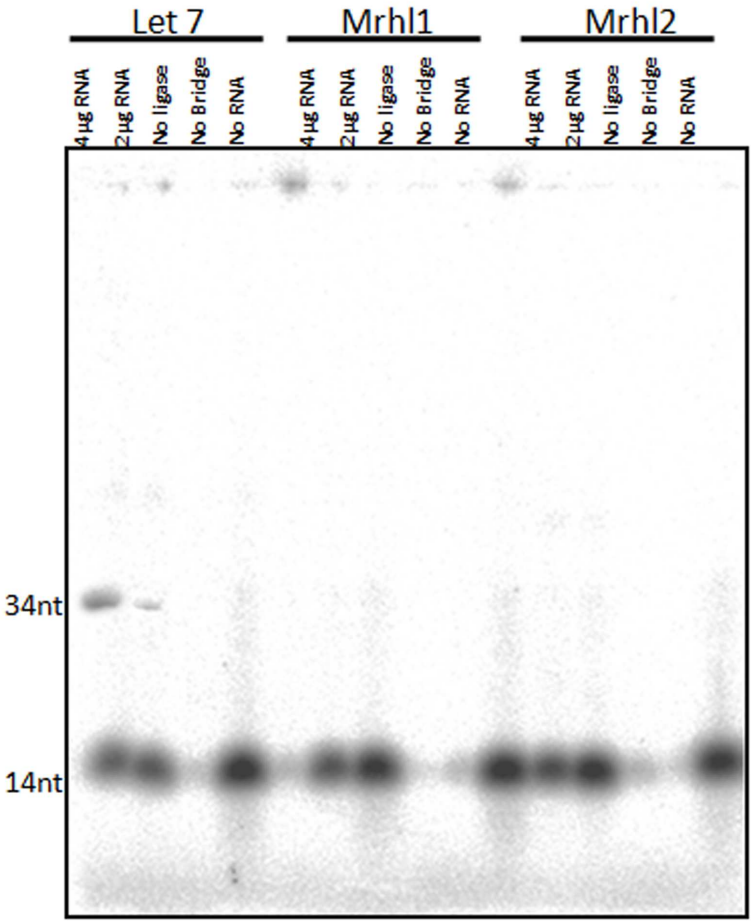

Figure 3.21: *Splinted Ligation. Splinted ligation experiment on total testis RNA for detection of microRNAs (Maroney et al. 2007). Let7, positive control is detected at 4 and 2 µg of RNA concentration, where as the mrhl bridge oligos 1 and 2 representing each strand of the 80 nt intermediate RNA failed to give positive signal. The negative controls with no ligase or no bridge oligo or no RNA did not give any signal.*

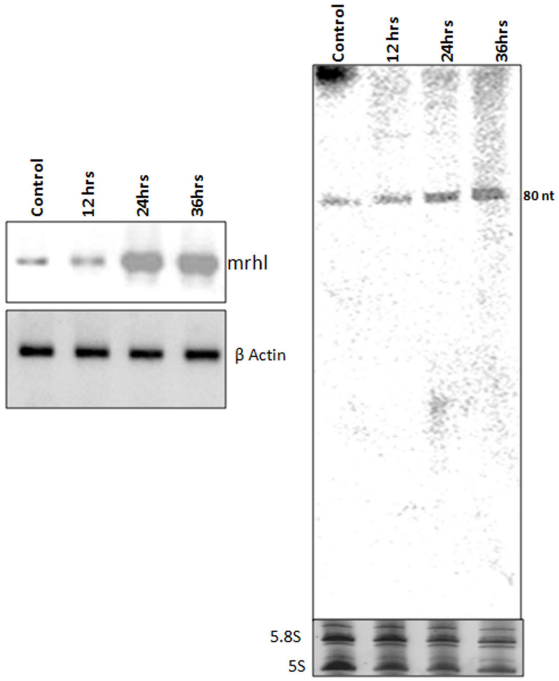

Figure 3.22: *Mrhl overexpression in cell lines. Over expression of the 2.4 kb mrhl RNA in the mouse Gc1 Spg cell line using the pCDNA mrhl construct . Small RNA northern blot using 80 nt region as probe on total RNA isolated from cells 12hrs, 24hrs and 36 hrs post transfection.The RT PCR of 2.4 kb RNA overexpression at various time points is shown with β actin control. Bottom panel represents the ethidium bromide stained gel showing 5S RNA.*

3.9 Dicer can process the 80 nt intermediate RNA to 22 nt RNA *in vitro*

However, we were still curious to know whether this ∼ 80 nt RNA species can be cleaved *in vitro* by Dicer to generate the 22 nt miRNA. Towards this direction, we amplified the region that corresponds to 80 nt region using T7 and SP6 adapter primers for generating sense and antisense transcripts respectively. These *in vitro* transcribed RNAs were used as substrates for an *in vitro* Dicer assay using recombinant human Dicer. As can be seen from Figure 3.23, a 22 nt RNA fragment was generated from only the sense transcript but not from the antisense transcript. Thus the ∼ 80 nt RNA that has been identified does have the propensity to act as a substrate for Dicer in an *in vitro* scenario.

3.10 The *mrhl* RNA localises to specific nuclear compartment

Since the experiments described above showed that both the 2.4 kb primary transcript and the processed ∼ 80 nuclear extracts were fractionated into nucleoplasmic and chromatin fractions and the isolated RNA were separated on a 15% and 5% polyacrylamide containing 7M urea and hybridised with probe G that encompassed the ∼ 80 nt region with flanking nucleotide stretch which will hybridise both to the ∼ 80 as well as the 2.4 kb primary transcript. As can be seen in the Figure 3.24A and 3.24B, the 2.4 kb primary transcript was enriched in the nucleoplasmic fraction while the ∼ 80 nt fragment was enriched in the chromatin

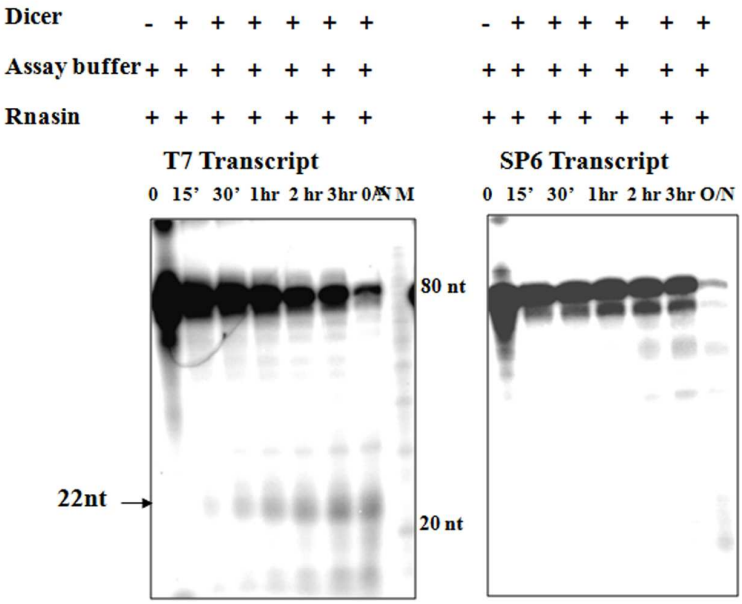

Figure 3.23: *In vitro Dicer assay. Dicer assay for in vitro transcribed 80 nt RNA from T7 (sense) and SP6 (antisense) promoter using recombinant human Dicer. The reaction was carried out for different time periods as indicated. The sequence corresponding to the 80 nt processed RNA was amplified using T7 and SP6 adapter primers and the RNA was in vitro transcribed using T7 (sense) or SP6 (antisense) polymerase and used as substrates for Dicer. M is the 10nt RNA ladder. 22 nt mature miRNA was accumulated over a period of time from sense transcript only.*

fraction. The integrity of the RNA was also checked in ethidium bromide stained gel as shown in Figure 3.24C. The purity of the subnuclear fractions were confirmed by analysing their protein composition on 10% SDS PAGE. The protein pattern showed in Figure 3.24D clearly reveals the presence of histones only in the chromatin fraction and absent in nucleoplasm.

The *in vivo* localization of this non-coding RNA within the nucleus was next examined by fluorescence RNA *in situ* hybridisation experiment. Two probes spanning different regions of the primary transcript, one falling within the $\sim$ 80 nt region and another away from that region to score for full-length 2.4 kb transcript. Each of the probes was synthesised using LNA technology for better hybridisation stability and specificity [239]. They were labelled with Cy3 for the probe 1 falling within the $\sim$ 80 nt region and with Cy5 for the probe 2 falling outside the region but within the 2.4 kb transcript. This experiment was carried out on Gc1 Spg cell line of mouse spermatogonial cells which expresses both the 2.4 kb primary transcript as well as the processed $\sim$ 80 nt RNA. The *in situ* hybridisations were performed in the presence of 50% formamide concentration and high stringency washes. The *in situ* fluorescence images are shown in Figure 3.25 panels A-F. Most interestingly clear large 3-4 punctate signals were observed that were hybridising to both the probes. When the cells were hybridised simultaneously, both the probes showed co localisation pattern (Figure 3.25 panels G-J). The signal observed in the large punctuate foci using the $\sim$ 80 nt probe 1 does not distinguish between hybridisation of the probe to either 2.4 kb primary transcript or the processed $\sim$ 80 nt intermediate *per se*. However, there were some additional signals from probe 1 that corresponds to the processed $\sim$ 80 nt RNA that did not co localise. In a control experiment we treated the cells with RNase A prior to hybridisation in

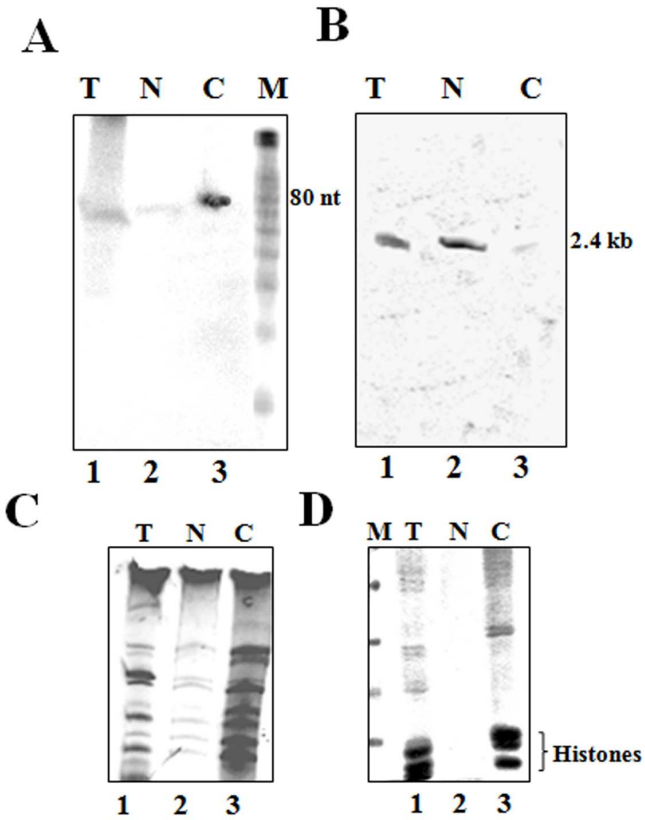

Figure 3.24: *Distribution of precursor and the processed RNA between nucleoplasm and chromatin. Testicular nuclei were processed to isolate the nucleoplasmic and chromatin fraction and the total RNA was hybridised with probes against 2.4 kb primary transcript and 80 nucleotide processed product. (A) Northern blot using probe G of RNA from total (T, lane 1), nucleoplasm(N, lane 2) and chromatin(C, lane 3) of testis separated on 15% PAGE containing 7M urea showing bands at 80 nt position. M is the 10nt RNA ladder. (B) Northern blot using probe G of RNA from total (T, lane 1), nucleoplasm (N, lane 2) and chromatin (C, lane 3) of testis separated on 5% PAGE containing 7M urea showing bands at 2.4 kb position. (C) RNA profile of total (T), nucleoplasm (N) and chromatin (C) of mouse testis on a 15% PAGE containing 7M urea loaded in lanes 1 through 3. (D) Protein profile by 15% SDS PAGE of the fractionated total nucleus (lane 1), nucleoplasm (lane 2) and chromatin (lane 3) showing enrichment of histones in chromatin fraction.*

which case no fluorescence signal was seen (Figure 3.25 panels K-P).

3.11 Nucleolus is the site of processing of the primary transcript

This observation was extended further to look at the localisation of this 2.4 kb *mrhl* transcript with any of the known nuclear protein(s). Initially, the experiment was performed with Drosha to see the colocalisation of Drosha with the RNA and it was found indeed that the Drosha and the 2.4 kb primary transcript colocalise to each other (Figure 3.26A-D) strengthening the earlier findings of Drosha's involvement in the processing of the 2.4 kb *mrhl* RNA. Some reports have shown the presence of Drosha in the nucleolus [175]. Hence, colocalisation of Drosha and Nucleolin, a nucleolus specific protein were carried out. Figure 3.26E-H shows the colocalisation of Drosha and Nucleolin. It was observed that Drosha and Nucleolin colocalised with each other. We further went ahead to study the colocalisation of the *mrhl* RNA to Nucleolus. Thus a colocalisation analysis of the RNA with the Nucleolin was performed. Interestingly, the RNA and the Nucleolin also did colocalise together showing the presence of the RNA in nucleolus (Figure 3.26I-L). A similar experiment was performed in RAG1 cell line (mouse renal adeno carcinoma cell line) where the presence of *mrhl* RNA in a similar pattern was observed with with predominant localisation in the nucleolus, showing the results presented so far is a general theme and not an artefactual signal owing to some property of the cell line itself. As mentioned earlier, the 2.4 kb RNA coding gene is absent in humans and our earlier experiment showed that this gene is not expressed in

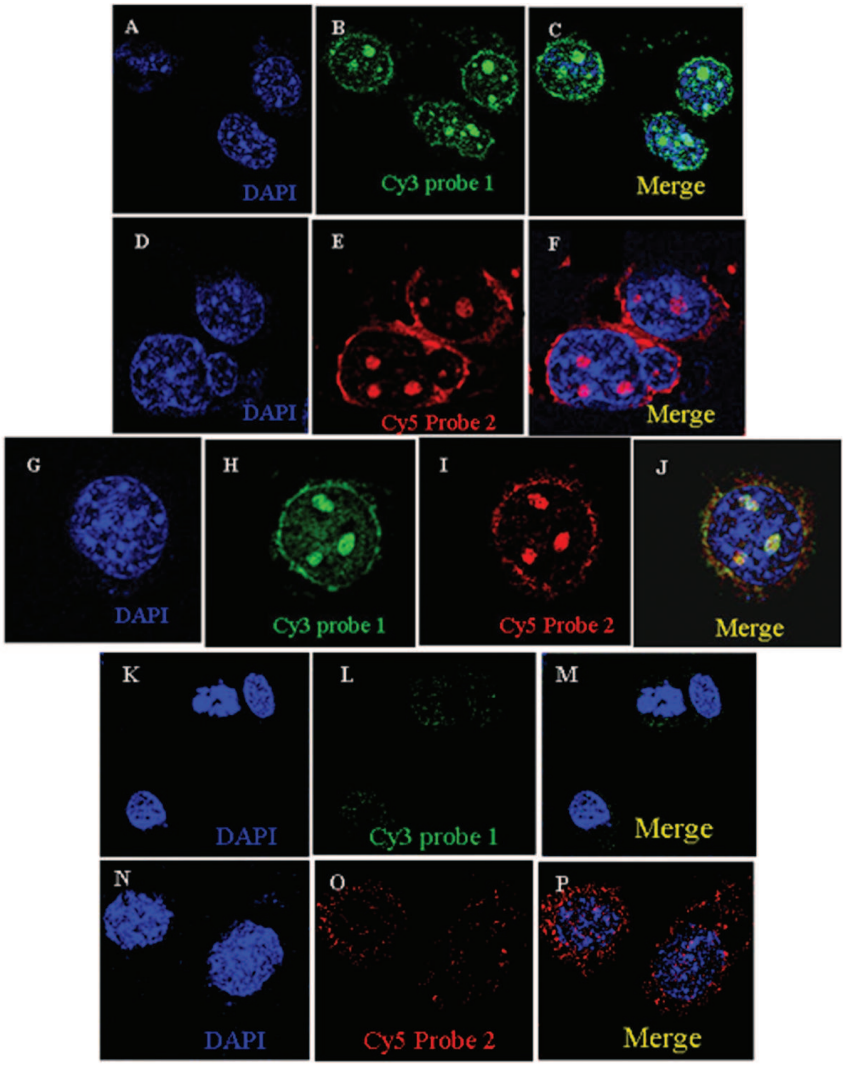

Figure 3.25: *The mrhl RNA localises to nucleus. RNA in situ on mouse Gc1 Spg cells using LNA probes. Panels A-C represent RNA in situ performed using probe 1 (encompassing the sequence within 80 nt region) labelled with Cy3. Panel B shows punctate nuclear localization of the RNA. Panel C is the merge image. Panels D-F represent RNA in situ hybridization using probe 2 labelled with Cy5 encompassing the region outside the 80 nt region but within 2.4 kb mrhl RNA. Panels G-J represent the colocalisation of probe 1 (panel H) and probe 2 (panel I) showing clear overlapping pattern (panel J). Cy3 shows additional diffused signal apart from the colocalised spots. Panels K-M and N-P are the RNA in situ of cells using probe 1 and 2 respectively where cells are treated with RNase A prior to hybridisation showing no signal. Nucleus is stained with DAPI.*

brain tissue in mouse. As an additional negative control similar RNA *in situ* in HeLa (human) as well as C6 rat glioma (brain) cell lines were carried out. As can be seen in Figure 3.26M-R and 3.26S-X no signals could be detected in both these cell lines.

3.12 Discussion

3.12.1 Characterization of *mrhl* non-coding RNA

Mrhl is a 2.4 kb non-coding RNA identified in our laboratory earlier, that is unspliced and polyadenylated [229]. Here experimental evidence has been provided, that, it is a non-coding RNA species by carrying out *in vitro* transcription coupled translation assay. Extensive bioinformatic search was performed to identify homologue of *mrhl* RNA across different species. The gene encoding this non-coding RNA transcript is also found in rat genome on chromosome 19 with 80% identity. It is interesting to note that this gene was not found in human genome database by bioinformatics search. This is not completely surprising since recent evidences show that non-coding RNA genes may not be conserved across species [38]. Many long ncRNAs fail to show any conservation of long stretches across species. This is true even with Xist which is essential for mammalian dosage compensation which is conserved only for a very short stretch [99]. Even the miRNAs where evolutionary conservation was considered to be a major factor for classifying as a micro RNA, recent reports suggesting that there are number of authentic miRNAs that fail to satisfy this criterion and these RNAs are in fact believed to play some role

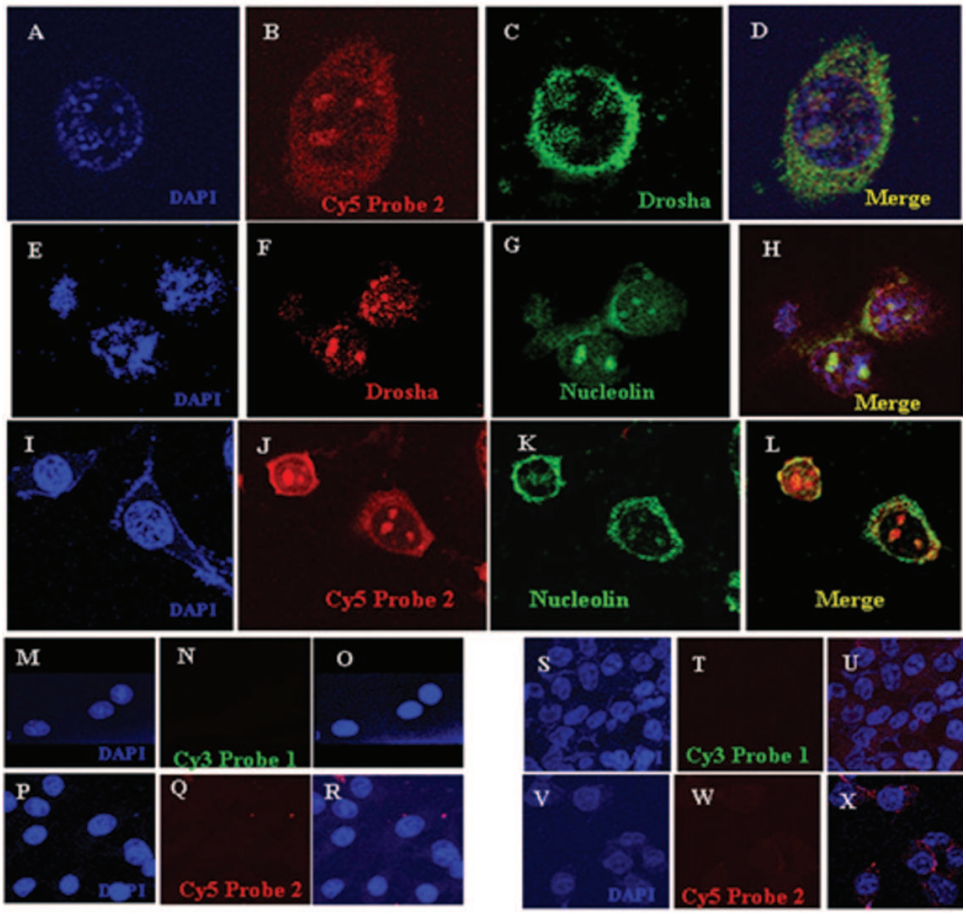

Figure 3.26: *Nucleolus is the site of processing of mrhl RNA. Immunolocalisation combined with RNA in situ on the mouse cells. Panels A-D RNA in situ on the mouse Gc1 Spg cells using LNA probe 2 labelled with Cy5 (panel B) and immunofluroscence using anti Drosha followed by anti rabbit Alexa 488 (panel C). Panel D is the colocalisation of both the signals showing overlapping regions in the nucleus. Immunolocalisation of Drosha (panel F) and nucleolin (panel G) in the nucleus using anti rabbit Alexa 568 and anti mouse Alexa 488 secondary antibodies respectively. Panel H is the merge of the two images showing colocalisation. Colocalisation of* mrhl *RNA hybridised to cy5 labeled LNA probe 2 (panel J) with nucleolin (panel K) showing overlapping signals (panel L). Panels M-R and S-X are negative controls using HeLa and rat C6 glioma cells respectively that do not express the RNA, where no hybridisation of the LNA probes could be visualized. The nucleus is stained with DAPI.*

in evolution of that species as well [49]. As described in the introduction chapter, non-coding RNAs are subjected to less evolutionary constraints compared to the protein coding counterpart, which are subjected to more evolutionary stress of retaining the ORFs. Thus, it would be best to consider the possibility that secondary structure of the RNA that could be conserved without any primary sequence conservation across the species. It would be very challenging to identify a structural homologue of this RNA in humans.

3.12.2 *Mrhl* is processed by the miRNA machinery

The extensive secondary structure of the 2.4 kb *mrhl* RNA prompted, to look for its putative RNA substrate in the miRNA pathway. Processing of this large RNA into a small intermediate RNA of 80 nt in length was indeed observed *in vivo*. The 80 nt small intermediate mouse RNA is also found to have 95% identity to the rat RNA sequences. The $\sim$ 80 nt RNA region from the long transcript was further delineated and no short ORF within this region was found thus eliminating the possibility of any small peptides been generated. This exercise was necessary because recently some non-coding RNA were shown to encode small peptides. Similarly the Steroid receptor RNA activator (SRA) is an RNA coactivator of steroid hormone dependent transcription which also encodes a small peptide though the SRA is also indispensible for the transcription coactivity [240].

Analysis of secondary structure of the *mrhl* RNA was performed and a fold back structure like stem loop was predicted which would serve as a potential pre miRNA. Pre miRNA are 70-80 nt small RNA that are derived from a longer primary transcript by the action of nuclear RNaseIII Drosha [234]. The *in vivo*

identification of an $\sim$ 80 nt processed species which has stem loop structure was also supported by *in vitro* experiments in which 2.4 kb transcript was cleaved to an $\sim$ 80 nt small RNA by total cell free lysate as well as by immune purified Drosha complex. The Drosha silencing also resulted in the reduced accumulation of the processed transcript which ensures that the *mrhl* processing is not a random degradation event but a specific regulated process. Interestingly, we also found that the 80 nt small RNA species can be cleaved by Dicer *in vitro* to a 22 nt micro RNA species. The northern blot on fractionated cellular compartments also showed the presence of the $\sim$ 80 nt RNA in the nucleus. This was very interesting because the pre miRNA generated in the nucleus will have to be transported to cytoplasm for its processing by Dicer, but our experiments did not find any cytoplasmic pre miRNA. The lack of detection of the 22 nt mature RNA processed product *in vivo* can be explained in two scenarios, where, in one the RNA being export defective and cannot bind to exportin 5 for its nuclear cytoplasmic transport as majority of the pre miRNA requires exportin 5 for its transport to cytoplasm [56]. The other reason may be, it is transiently expressed at some developmental or differentiation time point, which is not been scored for at present. It is becoming very clear that there exists multiple level of regulation operating within a miRNA pathway. For example, the *C. elegans* miR38 is expressed only in embryo whereas the precursor is ubiquitously detected in all stages of development, indicating that the maturation might be temporally regulated [241]. Since, some of the miRNA are known to determine major cell fate decisions, it has become vital for the system to tightly regulate its expression in highly temporal and spatial manner [232, 242]. It is still plausible that under certain physiological or pathological condition this 22 nt species may be generated *in vivo* and found in cytoplasm targeting specific

mRNAs.

3.12.3 *Mrhl* is a putative nuclear regulatory RNA

The next major question that was addressed in the present investigation was a possible role of the $\sim$ 80 nt RNA within the nucleus itself. There are numerous RNAs reported in the literature that elicit its function at nuclear level. The well-studied regulatory RNA such as Xist in mammals brings about long-range chromatin interactions and is indispensable for the establishment of dosage compensation by forming inactive X chromosome [111, 243]. Prasanth *et al.* have also reported a non-coding RNA CTN, which is activated upon stress, and regulates its mRNA counterpart, which is cation-activated channel protein mRNA by alternate promoter and poly A site usage. The 8 kb CTN RNAs are localised specifically in sub nuclear compartments like nuclear speckles and is post transcriptionaly cleaved to produce the protein coding mRNA [141]. Similarly, a screen for nuclear transcripts have identified two non-coding RNAs that are linked and are associated with S35 splicing domains in nucleus [139]. One of these RNAs named NEAT1 is shown to function in the formation of nuclear speckles [140]. These evidences suggest that RNA that are involved in multitude of functions are actually targeted to specific compartments in the cells and do not show a generalised uniform localisation pattern [244].

The fluorescent RNA *in situ* experiments using LNA probes also show a specific localisation pattern of both the 2.4 kb transcript and the processed $\sim$ 80 nt RNA species to nucleolus. There were also additional extra nucleolar signal observed with probe 1 labelled with Cy3 apart from the one that was colocalising with

Nucleolin and Drosha. The *mrhl* belonging to the family of snoRNA due to its specific localization in the nucleolus is also ruled out because of the absence of any C/D or H/ACA motifs present in it [23]. Furthermore, biochemical fractionation of nucleus has also shown that the processed $\sim$ 80 nt RNA species is associated with chromatin fraction. Analysis of these extra nucleolar chromatin domain(s) harbouring the processed $\sim$ 80 nt RNA species will give valuable insights into possible function of the 2.4 kb *mrhl* RNA.

A number of nuclear non-coding RNAs have been identified in the recent past enforcing the idea that the nuclear non-coding RNAs are one of the major determinants of nuclear functions that include transcription, post transcriptional gene regulation, regulation of protein functions, chromatin remodeling etc [37]. *Mrhl* is now an authentic nuclear non-coding RNA in the mouse. *Mrhl* also has a processed counterpart which is 80 nt in length. *Mrhl* thus, qualifies as a bonafide miRNA primary transcript as it could act as a substrate for RNaseIII enzymes. An interesting question that arises at this juncture is to identify which of the RNA species namely the primary or the processed 80 nt transcript is involved in the functional role. In the case of MALAT1, a nuclear retained non-coding RNA the primary transcript is misregulated in many cancers. It was recently shown that 3' end of this RNA is processed to yield a tRNA like RNA that is transported to cytoplasm and regulates protein synthesis. Both the transcripts are found to be stable and functional [245]. Though in *mrhl*, both the RNAs are nuclear localized, the 80 nt RNA shows enrichment in the chromatin that favors a nuclear regulatory function for the processed 80 nt species. It remains to be seen whether the longer primary transcript also posses any regulatory role within the nucleus as it is also one of the stable species apart from the 80 nt intermediate RNA.

With an increasing proportion of functional non-coding transcripts, it is very tempting to speculate that the *mrhl* could also belong to the member of nuclear non-coding RNA. The stability of the RNA, distinct sub nuclear localization, the ability to get processed by RNaseIII enzyme to a highly structured intermediate species and tissue specific expression of *mrhl* RNA clearly favors the argument presented above. The next chapter explains the efforts undertaken in establishing the functional role of this RNA.

Chapter 4

Functional Analysis of mrhl RNA

4.1 Introduction

In recent years, there has been an explosion in the discovery of several classes of non-coding RNAs which constitute a huge family of gene regulatory molecules in higher eukaryotes. Non coding RNAs function in diverse cellular processes. Many such functions are mediated through their protein binding partner wherein the non-coding RNAs are indispensible component of such complexes. For example, small RNAs (siRNA and miRNA) are often found in association with the RISC (RNA Induced Silencing Complex) that consists of argonaute and other proteins [246–248]. In case of piRNAs, which are involved in transposon regulation, the RNA component is associated with Piwi proteins such as Miwi or Mili in mouse and Hiwi in humans [91, 95]. Long non-coding RNAs also have functional protein partners. For example, in the case of XIST, the RNA is associated with the macro H2A during the process of X chromosome inactivation [249, 250].

p68 also referred to as Ddx5 is a founding member of large family of DEAD box

helicases [165]. The proteins of this family consist of DEAD box motif which is the signature sequence of these members. They also possess other conserved sequences, including an ATPase domain and an RNA helicase domain [164]. These proteins have been shown to play important roles in diverse biological processes such as development [169], regulation of transcription [170], RNA processing [172], ribosome biogenesis and also in the miRNA pathway [175, 251]. Apart from the functions described above, p68 has also been shown to play a role in signaling events. During epithelial mesenchymal transition that is induced by PDGF stimulation, p68 which gets phosphorylated in tyrosine 305 was shown to translocate to the cytoplasm where it stabilizes β-catenin and subsequently p68-β-catenin translocates back to the nucleus bringing about downstream transcriptional regulation [185].

Wnt signaling is a highly conserved developmental signaling pathway involving the major effector protein β-catenin [215]. There are two major types of wnt signaling, the canonical and the non canonical signaling. The canonical wnt signaling gets activated upon binding of wnt ligands to its receptors Frizzled or LRP which results in activation of the pathway, where β-catenin gets stabilized in the cytoplasm. The stabilized β-catenin translocates to the nucleus whereby it binds to TCF/LEF family of transcription factors and activates the wnt target genes. In wnt uninduced condition, the β-catenin is highly unstable in the cytoplasm and it is phosphorylated by GSK3β/axin/APC complex that targets β-catenin to ubiquitin mediated proteolysis. The wnt signaling is very essential in development, cell proliferation and ES cell differentiation and it has been discussed in detail in the Chapter 1.

As demonstrated in Chapter 3, *mrhl* RNA is a 2.4 kb non-coding RNA which is processed in the nucleus by the Drosha machinery to generate a 80 nt processed

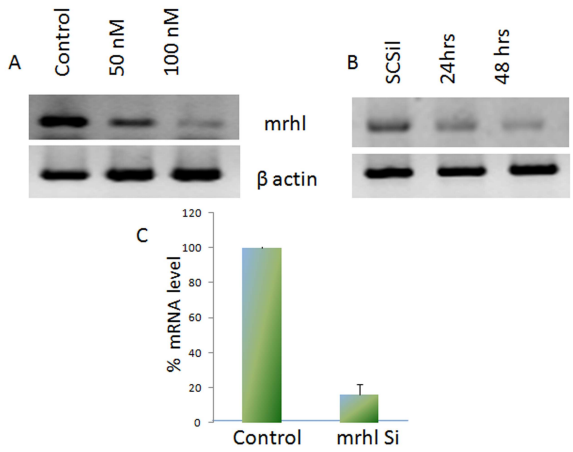

Figure 4.1: *Mrhl down regulation on Gc1-Spg cells. (A) Semi quantitative RT PCR showing down regulation of mrhl transcript in siRNA treated cells as against the control. (B) Quantitative Real time PCR analysis showing the percentage of mRNA decrease in the siRNA experiment as against the control scrambled siRNA treated cells.*

RNA intermediate that is restricted to nucleus and does not get exported to the cytoplasm. Based on this observation, it was speculated that this non-coding RNA might have a nuclear function. To address the function of this RNA in the nucleus, two approaches were undertaken. Initially, the *mrhl* RNA was down regulated and the global gene expression change due to silencing of this RNA was studied. In another approach, the *mrhl* RNA binding proteins were identified. Later the cross talk between the *mrhl*-protein complex and the gene expression change upon *mrhl* down regulation including cell signaling pathways were established.

4.2 Effect of down regulation of *mrhl* RNA on global gene expression

As a first step towards understanding the biological function(s) of *mrhl* non-coding RNA, silencing of this RNA was performed to analyze its effect on global gene expression using microarray technology. For this purpose, Gc1-Spg cell line which was derived from mouse spermatogonia was used, where, the 2.4 kb *mrhl* non-coding RNA expression was established as described in Chapter 3. It is generally believed that siRNA mediated down regulation of its target mRNA is a cytoplasmic event. Although the non-coding RNA and its processed 80 nt intermediate are nuclear restricted, some recent reports suggest that siRNA mediated down regulation also takes place in the nucleus, though the mechanism of this down regulation of nuclear restricted RNA is not clear at present [252, 253]. Four siRNAs targeting to 4 different regions of *mrhl* primary transcripts were designed using Dharmacon web based tool. The designed siRNAs were searched against mouse genome database using BLAST to verify the specificity of the designed sequences. Initial siRNA experiments were carried out using individual siRNA and then as a pool with different concentrations and time points to arrive at maximum efficiency of down regulation. The effective concentration was then determined to be 100 nM using siRNA pools at 48 hours post transfection where we obtained nearly 70-80% efficiency of down regulation (Figure 4.1A and B).

Total RNA samples were obtained from 3 independently performed siRNA experiments where the efficiency of down regulation was determined by RT QPCR and was found to be nearly the same. In a control experiment, we used scrambled siRNA pool (Dharmacon)(Figure 4.1C). The RNA from the control and the *mrhl*

siRNA treated cells were processed further for expression profiling using Affymetrix 43.0 K mouse genome expression array. The arrays were scanned and preliminary analysis was done to check for the proper working of internal controls in the microarray experiment. The values of all the negative and positive controls of 3 sets of the experiments were within the limits of Affymetrix standards and quality.

A thorough analysis of the differentially regulated genes was carried out, where, the genes were initially sorted based on fold change and significance values. PLIER algorithm of array assist software was used to perform this analysis and the number of genes that showed differential pattern of regulation across all three experiments is given in Figure 4.2A which are based on fold changes and statistical significance. For all our analysis the genes that are differentially perturbed more than 1.5 fold with a P value of < 0.05 has been used. Three sets of genes which satisfied these criteria were compared, and the Venn diagram, shown in Figure 4.2B represents the overlap of number of genes that were falling under each of the three sets of genes. A heat map of co-regulated and clustered genes was generated using java tree view software. Figure 4.3 shows representation of the heat map generated. It is clear from the figure that the genes from siRNA treated samples grouped together separately, whereas the control scrambled siRNA treated samples fell into a separate group without any overlap between the experiments and controls showing the reliability of the data obtained on differential gene expression. It is also seen that many genes cluster within themselves indicative of co-regulation or their involvement in a similar pathway. The gene sequence of *mrhl* RNA, being a non-coding RNA was not spotted on the affymetrix array and hence could not be further validated in this array experiment other than the real time PCR analysis shown in Figure 4.1B. A complete list of genes that are perturbed by *mrhl* RNA

P value/ Fold change	≤0.05	≤0.02	≤0.01	≤0.005	≤0.0010
≥1.5	652	261	155	79	13
≥2	315	139	54	41	6
≥3	176	80	51	25	5

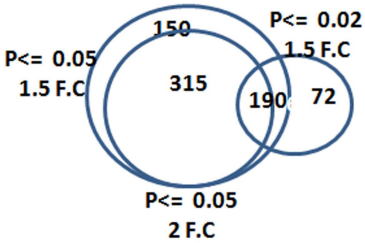

Figure 4.2: *Preliminary grouping of perturbed genes in microarray. Genes short listed based on fold change and significance value as analysed by array assist are given (Top). Comparative Venn diagram of overlap of the genes from the three list that were chosen for further analysis (Bottom).*

silencing is given in Annexure 1.

The classification of the over represented class of perturbed genes based on its GO annotation and biological function was carried out. David Ease as well as GOTM softwares were used to group the genes based on functional annotation category. The pie chart generated of the genes grouped under various functional categories is given in Figure 4.4. It can be seen that nearly 45% of the genes that are perturbed belong to cell adhesion (14%), signalling (17%) and development and differentiation (14%). Interestingly 13% of the genes belong to transcription process comprising of transcription factors (8%) and transcription regulation (5%). Since, some non-coding RNA also function in cis, the chromosome wide distribution of the genes that are perturbed following silencing of *mrhl* RNA were examined. We could not find any enrichment of genes in any particular chromosome,

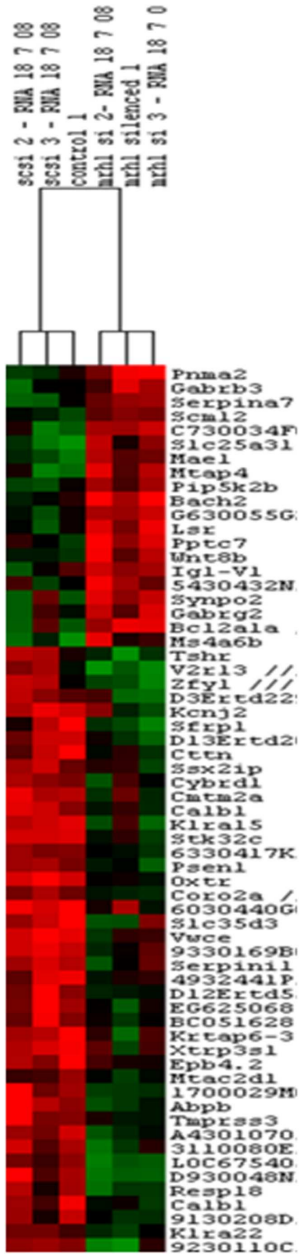

Figure 4.3: *Cluster analysis of mrhl perturbed genes. Representative heat map of the genes that are up or down regulated in the microarray experiment showing of clustering of experiment and controls in the same group.*

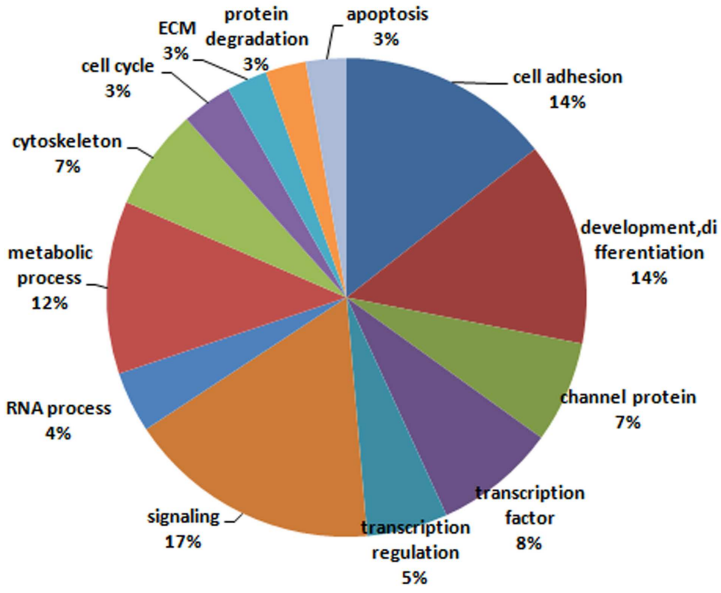

Figure 4.4: *Classification of genes based on GO terms showing over represented class of genes in the microarray experiment.*

and these genes were distributed throughout the genome (Figure 4.5). However, some genes that were present as gene clusters in the genome that was in the array list including Protocadherins and Integrins were also found to be perturbed upon *mrhl* silencing. Taking together the observations, it is speculated that *mrhl* RNA preferentially regulates the cell adhesion mediated signaling that control development and differentiation with the involvement of specific transcription factors. In addition to these classes of molecules, it was also observed that genes belonging to metabolic process, cell cycle are perturbed upon *mrhl* RNA silencing.

The next interest was to look into closely at the transcription factors that were perturbed after *mrhl* silencing. The rationale behind looking at transcription factors was that, if we can identify those transcription factors controlling a set of downstream genes which can also be identified them from the microarray gene list shown in Table 4.3. About 11 transcription factors that were differentially

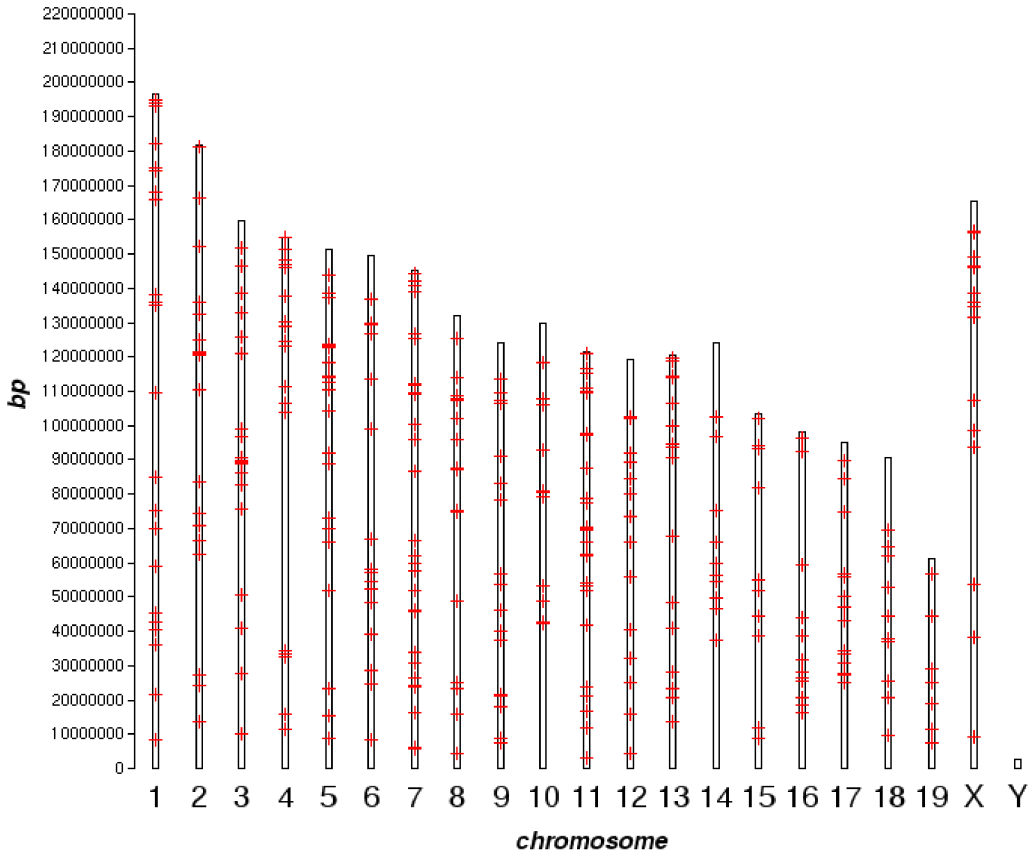

Figure 4.5: *Chromosome wide distribution of perturbed genes upon mrhl silencing.*

regulated in our global gene expression profiling (Table 4.1) were identified. Interestingly, transcription factors that are involved in developmental program such as Sox8 and Sox12 and the members of Fork head box group of transcription factors such as FOXD4 and FOXP1 were found to be differentially regulated. Further validation of the perturbation of these transcription factors were carried out by real time PCR analysis following *mrhl* silencing and the data are presented in Figure 4.6. The transcription factors Sox12, FOXP1, Sox8, Bach2, NFIC, FOXD4, Scml2 were up regulated several fold upon silencing of *mrhl* RNA. On the other hand, transcription factors AP2A and Odz2, were down regulated six to nine fold upon silencing of *mrhl* RNA. When re-examined for the list of genes that were perturbed upon *mrhl* silencing, it was observed that many of the target genes of AP2A, FOXP1 and Sox12 were present indicating that the effect of *mrhl* RNA gene silencing on these genes are probably mediated by the primary effect of *mrhl* RNA on these transcription factors (Figure 4.7).

Gene	Fold Change
Tsc22d1	-5.82216
Tcfap2a	-4.346088
Odz4	-2.29
Nfic	2.232845
Scml4	2.2474325
Sox8	2.362174
Sox12	2.7187328
Foxp1	2.9280138
Foxd4	3.314626
Scml2	3.4461977
Bach2	7.095153

Table 4.1: *List of transcription factors.*

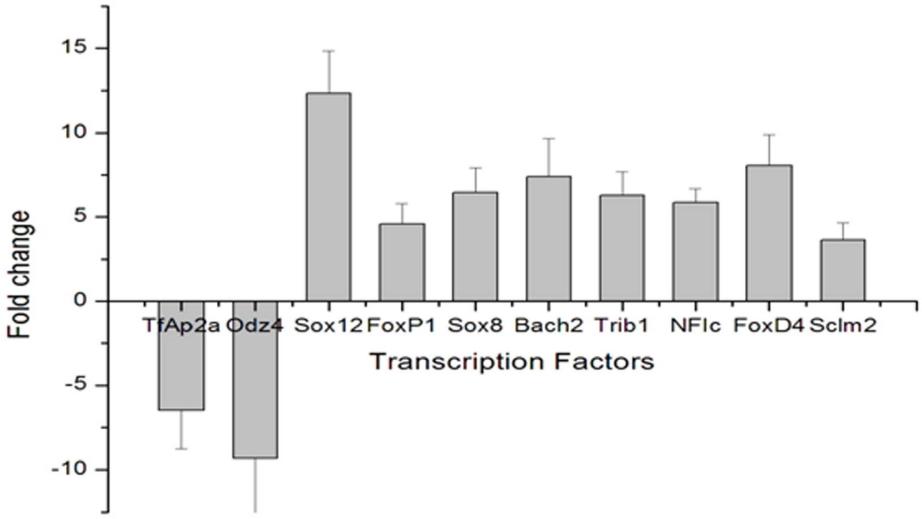

Figure 4.6: *Differential expression of transcription factors upon mrhl silencing. Real time PCR validation of transcription factors that are differentially expressed in mrhl down regulated cells compared against scrambled siRNA treated cells.*

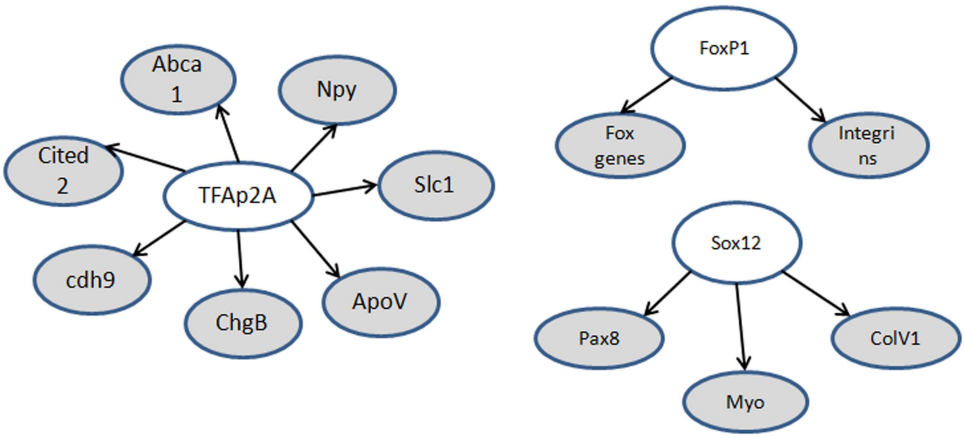

Figure 4.7: *Network of targets of transcription factors. Network of target genes of the transcription factors both of which are perturbed in mrhl down regulated condition.*

4.3 Activation of wnt signalling in *mrhl* silenced cells

The next goal was to see the enrichment of specific pathway genes that are present in the expression profiling data. It was observed that the genes involved in wnt signaling were over represented in the array data compared to other genes of any specific pathway (Table 4.2). RT QPCR validation of those genes involved in wnt signaling was carried out and the results are presented in Figure 4.8A. Wnt8B was significantly up regulated both in the microarray as well as RT PCR whereas the wnt negative regulator SFRP1 was down regulated. Other genes that belong to the wnt pathway such as DKK3, TCF4 were also differentially regulated. Two other genes that are part of non canonical wnt signaling namely Prickle1 and Rac1 were also found to be differentially regulated in the *mrhl* silenced condition. Smarca4 or Brg1 which is directly not involved in wnt signaling but is essential for the chromatin remodeling activity of wnt targets in complex with TCF family members was found to be up regulated in *mrhl* down regulated conditions.

Gene	Fold Change
SFRP1	-7.4
RAC1	-2.3
DKK3	-1.9
Frizzled6	-1.5
Cyclin D3	1.5
TCF4	1.6
Smarca4	2.4
Prickel1	3.6
Wnt 8B	5.4

Table 4.2: *Wnt Pathway genes.*

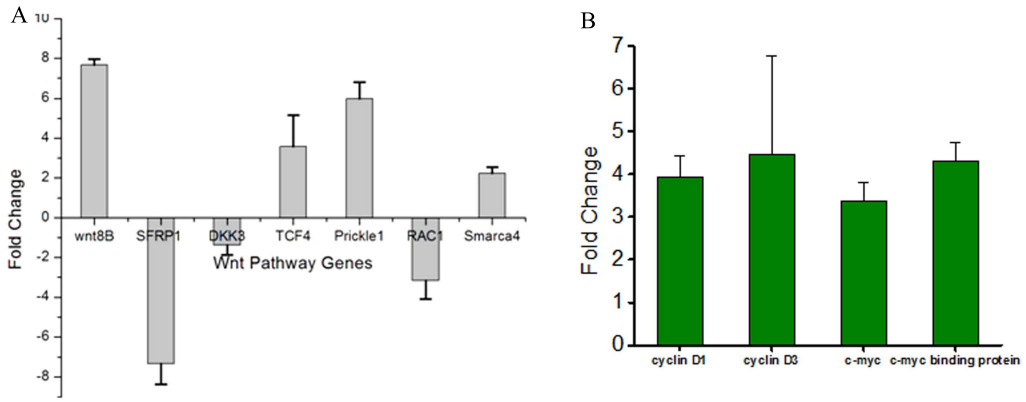

Figure 4.8: *Wnt signaling is perturbed upon mrhl down regulation. (A) Realtime PCR validation of wnt pathway genes present in the microarray gene list. Negative regulators of wnt are down regulated whereas the positive regulators are up regulated. (B) Real time QPCR of some β-catenin regulated genes in the siRNA treated cells compared with scrambled siRNA treated controls.*

Gene symbol	**Fold Change**	**Entrez gene** I.D
Cell Adhesion		
Igl-V1	4	16142
Vav1	2.86	22324
Cdh9	2.81	12565
Selp	2.42	20344
Tro	2.01	56191
Nrxn1	-2.09	18189
Klre1	-2.2	243655
Thbs4	-2.36	21828
Dscaml1	-2.43	114873
Igsf4b	-2.55	94332
Pcdha11	-2.58	12942
Igsf11	-2.81	207683
Klra22	-2.9	93969
Itgb2l	-3.02	16415
Pcdhb22	-3.24	93893
Dsg2	-3.44	13511
Cldn16	-4.02	114141
Panx3	-4.93	208098

Klra15	-7.21	27423
Signaling		
Wnt8b	5.44	22423
Gabrg2	4.37	14406
Asb15	4.28	78910
Pip5k2b	4.24	108083
Prickle1	3.65	106042
Gpr22	3.49	73010
Gpr45	-2.07	93690
Ppp2r5a	-2.08	226849
Pskh1	-2.22	244631
Arl4a	-2.28	11861
Txk	-2.37	22165
Npr3	-2.37	18162
Als2	-2.39	74018
Npy2r	-2.46	18167
Diras1	-3.11	208666
Ltbp1	-3.58	268977
Tshr	-5.3	22095
Stk32c	-6.29	57740
Sfrp1	-7.45	20377
Ghrl	-8.82	58991
Development and Differentiation		
Lsr	6.03	54135
Fabp9	4.36	21884
Aard	4.02	239435
Epb4.2	3.83	13828
Cspg4	3.26	121021
Bex1	3.22	19716
Odam	3.17	69592
Tnnt3	2.83	21957
Elk3	2.76	13713
Eda2r	2.1	245527
Vgf	2.06	381677
Chrdl1	-2.2	83453
Nrg3	-2.23	18183
Odz4	-2.29	23966
Ndn	-3.55	17984
Adamts20	-3.68	223838
Serpini1	-4.33	20713

Psen1	-4.76	19164
Vwce	-6.86	71768
Transcription		
Bach2	7.09	12014
Zfp286	4.52	192651
Zswim2	3.91	71861
Scml2	3.44	107815
Foxd4	3.31	14237
Zfp703	3.09	353310
Foxp1	2.92	108655
Sox12	2.71	20667
Sox8	2.36	20681
Smarca4	2.29	20586
Scml4	2.24	268297
Nfic	2.23	18029
Mkrn1	-2.27	54484
Ppargc1a	-2.36	19017
Klf12	-2.6	16597
Zfp708	-3.6	432769
Tcfap2a	-4.34	21418
Cmtm2a	-4.75	73381
Tsc22d1	-5.82	21807
Zfy1	-6.08	22767
Metabolism		
Ddc	3.71	13195
Ugt8a	3.7	22239
Hsd3b3	2.4	15494
Agmat	2.27	75986
ORF34	2.05	207375
Apoa5	-2.19	66113
Aldh1a3	-2.33	56847
Arg2	-2.42	11847
St8sia6	-2.46	241230
Folh1	-2.51	53320
Adh5	-2.57	11532
Acacb	-3.15	100705
Otc	-3.37	18416
Sncaip	-4.32	67847
Cytoskeleton Organization		
Mtap4	6.77	17758
Synpo2	5.64	118449

Spna1	4.15	20739
Ktn1	2.57	16709
Ttll10	2.25	330010
Rac1	-2.35	19353
Dnhd1	-3.15	77505
Plek	-3.43	56193
Coro2a	-3.82	107684
Ablim3	-4.78	319713
Cttn	-4.85	13043
Krtap6-3	-5.46	16702
Channel protein		
Kctd12b	3.77	207474
Scn7a	2.66	20272
Kcna6	-2.66	16494
Kcnh1	-2.88	16510
Hcn3	-3.25	15168
Scn3b	-3.31	235281
Kcne2	-4.48	246133
Kcnj2	-5.01	16518
LOC675405	-5.41	675405
RNA process		
Mael	5.22	98558
Skiv2l2	3	72198
Rex2	-2.09	19715
Gm817	-2.28	329207
Snrpn	-4.6	20646
Rps6ka6	-6.56	67071
Apoptosis		
Bcl2a1a	6.98	12044
Gzma	2.013	14938
Tnfrsf19	-2.29	29820
Bcl11a	-2.36	14025
ECM		
Col3a1	6.65	12825
Dmp1	4.63	13406
Col3a1	4.11	12825
Mmp1b	3.4	83996

Table 4.3: *List of genes perturbed upon mrhl silencing ($FC \geq 2$, P value ≤ 0.05) classified based on its GO term.*

The downstream targets of wnt signaling in the list of differentially expressed genes upon *mrhl* silencing were examined. Interestingly, many wnt target genes that were also differentially regulated in the *mrhl* silenced cells as against control cells. The Real time PCR validation of some of these genes is given in Figure 4.8B. This included cyclin D1 and c myc which are the known targets of wnt β-catenin pathway. The other genes that are targets of wnt such as Claudin 1, Ephrin B were also found to get up regulated several fold in the microarray experiment (Table 4.3). From the results presented so far it is inferred that the wnt signaling, one of the major developmental signaling pathway, is affected upon *mrhl* silencing with the effects manifested both in wnt upstream as well as downstream genes. However, the ultimate wnt effector molecule β-catenin mRNA was not found to be differentially regulated in our experiment. It is well known that β-catenin regulation is mediated predominantly at the level of protein than at the transcription level [254]. So the stability of β-catenin in *mrhl* down regulated cells was examined as against controls that were either lipofectamine treated or scrambled siRNA treated. This was done by immuno-cytochemistry whereby the cells were stained with β-catenin antibody. siRNA transfections were carried out, and the cells harvested 48 hours post transfection were fixed and the immunofluroscence experiments were performed. The images were captured using confocal microscope and the representative field images of a 40× and 100× magnification are given in Figure 4.9. It can be seen from the figure that β-catenin was found to be stabilized in the *mrhl* si treated cells with several foci of cytoplasmic β-catenin accumulation. Nuclear β-catenin was also observed in many of the cells showing an active wnt signaling in these cells. The controls in contrast had uniform staining of β-catenin only in the cell membrane near adherens junction [255] where it is present

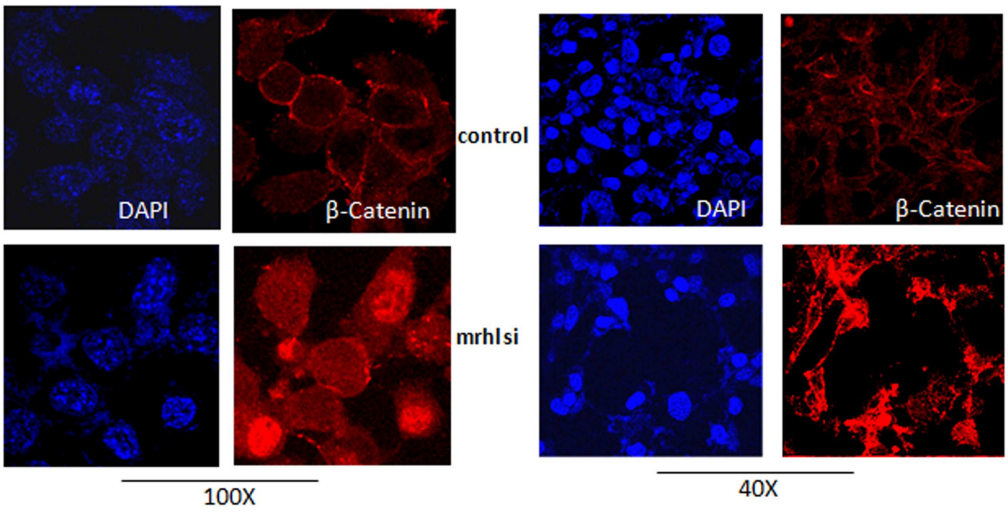

Figure 4.9: *Mis-localization of β-catenin in the mrhl siRNA treated cells. The β-catenin stains more in the cytoplasm and to some extent in the nucleus. The control cells are either lipofectamine treated or scrambled siRNA treated where β-catenin exclusively localized to adherens junctions in the plasma membrane. Nucleus is stained with DAPI.*

in association with cadherins. Our results thus demonstrate an active wnt signaling that involves cytoplasmic stabilization and eventual nuclear translocation of *β*-catenin in *mrhl* silenced cells which in turn bring about differential regulation of downstream targets of wnt signaling such as cyclin D1 and c-myc.

4.4 Identification of *mrhl* RNA interacting proteins

As mentioned in the introduction many of the RNA function are mediated through its interaction with proteins. Therefore, it was investigated, whether *mrhl* could bind to any protein within the cell. This was necessitated because, complimentary sequence of *mrhl* RNA present in the wnt pathway genes that were perturbed in this study was not found. Furthermore, it has been seen, from the

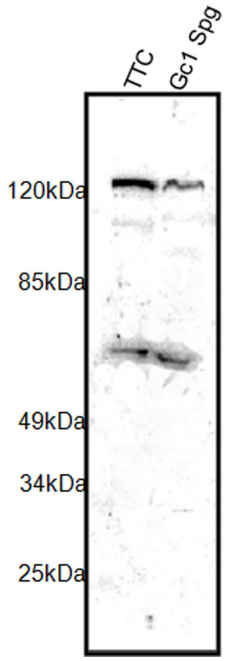

Figure 4.10: *Northwestern blot probed with full length in vitro transcribed mrhl transcript on mouse total testicular cells and the Gc1-Spg cell line.*

previous results presented in chapter3 that *mrhl* primary transcript is processed to an 80 nt intermediate RNA within the nucleus and a mature 22 nt RNA could not be detected *in vivo*. The absence of mature 22 nt RNA *in vivo* suggested a different post transcriptional regulation for the mechanism of down regulation of gene expression by this *mrhl* RNA. Thus, experiments were performed to identify *mrhl* RNA interacting proteins. For this purpose, a north-western analysis was carried out, where the full length 2.4 kb *mrhl* RNA was used as a probe. As shown in Figure 4.10, at least 3 interacting proteins around $\sim$ 120 kDa, $\sim$ 100 kDa and at $\sim$ 65 kDa region were identified, among which 120 kDa and 65 kDa were the more prominent ones. The experiment was done both in total testicular lysate as well as the in the Gc1-Spg cell lysate, where similar profiles were observed in both the cases (Figure 4.10).

However, from an experiment like northwestern blot only molecular size information could be obtained and the identity of the protein cannot be established. In order to identify these proteins, we resorted to a more sensitive experiment like RNA affinity pull down. The 2.4 kb *mrhl* transcript was *in vitro* transcribed in the sense orientation from the transcript cloned in pGem vector under T7 promoter in the presence of biotinylated UTP so that the resulting transcript would be labeled with biotin. The *in vitro* transcribed RNA was completely denatured and then slowly renatured to room temperature to preserve all the potential secondary structures. The pull down was performed using total cell lysate of Gc1-Spg cells. The mock pull down was done only with the streptavidin beads. The proteins that were pulled down were resolved on an SDS-PAGE and proteins were visualized after silver staining (Figure 4.11). The proteins were then identified by MALDI_TOF mass spectrometry. At least four proteins could be identified by MS/MS analysis and the proteins with its respective molecular weight and the peptides that were identified are given in the Table 4.4. Among the four proteins identified, further analysis was carried out for one of the RNA binding proteins, namely, Ddx5 also known as DEAD box helicase p68.

4.5 p68 or Ddx5, a DEAD box helicase, is a putative *mrhl* interacting protein

p68 or Ddx5, belongs to the family of DEAD box helicases which can unwind RNA in an ATP dependent manner [164] and participate in diverse biological processes. Recent evidence also suggests its role in the regulation of steroid receptors

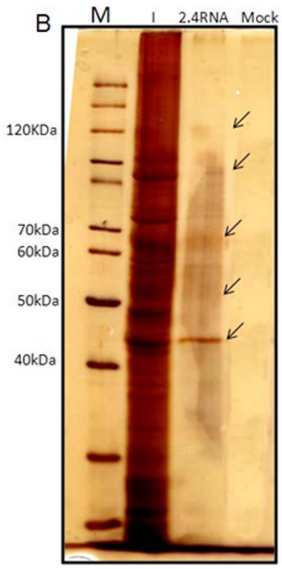

Figure 4.11: *Identification of mrhl binding proteins by affinity pull down. Lane 1: marker Lane 2 is input, Lane 3 is mrhl RNA pull down and lane 4, mock pull down. The arrow mark refers to interacting proteins that were subjected to mass spectrometry.*

such as androgen and estrogen receptor mediated transcription [256,257]. In order to confirm the result obtained above from the pull down and mass spectrometric analysis, western blot analysis using anti p68 antibody on the affinity pull down proteins was performed. In such an analysis, it was observed that the p68 was picked up in the input cell lysate and the *mrhl* pull down sample where as it was not observed in a mock pull down where antisense *mrhl* was used or only the beads (Figure 4.12A). This observation was extended further to check for an *in vivo* interaction between *mrhl* RNA and p68 by doing an S1 aptamer pull down. S1 aptamers are high affinity streptavidin aptamers that can be tagged to RNA/DNA sequences and pull down can be performed using streptavidin beads [235,258]. The *mrhl* gene was tagged to S1 aptamer and was cloned in pCDNA 3.1 mammalian expression vector. The *mrhl*S1 fusion construct was transfected in to Gc1-Spg cells and the cells were lysed after 36 hours post transfection. The cell lysate containing

Molecular Weight	Protein	Identified Peptides
120 kDa	Lim domain containing protein	K.FGE FCHGCSLLMT GPAMVAGEFK K.RYTVVGNP YWMAPEMLNG K K.RPPVEKATT K
68 kDa	DDX5	DR.GFGAPRFGGSKAGPLSG.K R.GVEICIATPGK
50 kDa	Tubulin beta	K.FWEVISDEHGIDPTGTYHGDSDLQLDR.I R.ISVYYNEATGGK.Y K.YVPR.A R.AILVDLEPGTMDSVR.S R.SGPFGQIFRPDNFVFGQSGAGNNWAK.G K.GHYTEGAELVDSVLDVVR.K K.IREEYPDR.I R.IMNTFSVVPSPK.V R.FPGQLNADLR.K
42 kDa	Dusp11	R.VFSGYSSAK.K K.SFEKHLAPEECFSPLDLFN.K R.YLIDVEGMRPDDAIELFS.R R.GFEDSTHMMEPVFTAT.K

Table 4.4: *Mrhl interacting proteins identified through mass spectrometry.*

RNasin and protease inhibitors were directly bound to streptavidin agarose beads. Controls were performed using cell line transfected with empty vector and only the beads. The proteins were eluted and it was subjected to western analysis with anti p68 antibodies. As can be seen from the Figure 4.12B, the p68 is present in the input cell lystae of *mrhl*S1 and mock transfected cells and in the affinity pull down proteins of mrhls1 transfected cells only and it is absent in the mock transfected pull down proteins. Beads which were bound to the lysate acted as negative control and did not yield any signal. This confirms that *mrhl* does interact with p68 protein *in vivo*.

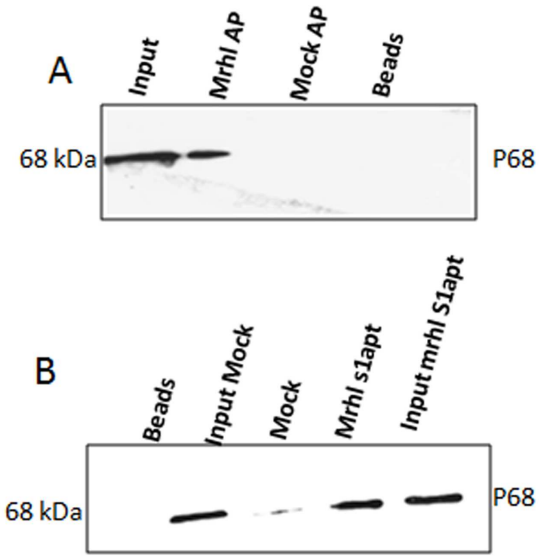

Figure 4.12: *p68 is a putative mrhl interacting partner. (A) Western blot analysis of in vitro pull down experiment with Gc1-Spg cells using biotinylated mrhl sense RNA using p68 antibody. (B) Western blot of in vivo pull down in Gc1-Spg cells transfected with plasmid containing mrhl fused with S1 aptamer using p68 antibody. Mock is the pull down experiment carried out in Gc1-Spg cells transfected with plasmid containing antisense mrhl fused with S1 aptamer. Lane 1 and 4 are input of each of the experiment. Lane 5 is the pull down with only beads using wild type Gc1-Spg cell lysate.*

It is also known from the previous reports that the p68 can regulate transcription by binding with a non-coding RNA SRA (Steroid Receptor RNA activator). Caretti and group have analysed the secondary structure of the SRA and have suggested some potential secondary structure that may be responsible for p68 binding and its co activator function (Figure 4.13A) [182]. It was interesting to note that one such motif was also present in the *mrhl* RNA and particularly in the 80 nt intermediate RNA (Figure 4.13B). In order to see whether the 80 nt intermediate RNA could also bind to p68, similar RNA affinity pull down experiment was carried out using *in vitro* transcribed 80 nt intermediate *mrhl*. The interacting proteins were probed by western analysis using anti p68 antibodies. As can be seen from Figure 4.13C, the 80 nt RNA species can also interact to p68, but with less efficiency as compared to full length 2.4 which was used as a control.

4.6 *Mrhl* RNA is involved in nuclear retention of p68

Having confirmed that p68 is a putative interacting partner for *mrhl* both *in vitro* and *in vivo*, the next interest was to know whether p68 is a part of activation of the wnt signaling pathway that has been seen in the global expression profiling following down regulation of *mrhl* RNA. A previous report from Yang *et al.* [185] has shown that there is indeed a direct correlation between p68 and β-catenin activation at least in the colon cancer cells particularly during the process of epithelial mesenchymal transition. It has been shown by this group that the p68 gets phosphorylated in the tyr 305 residue and the phosphorylated p68 can

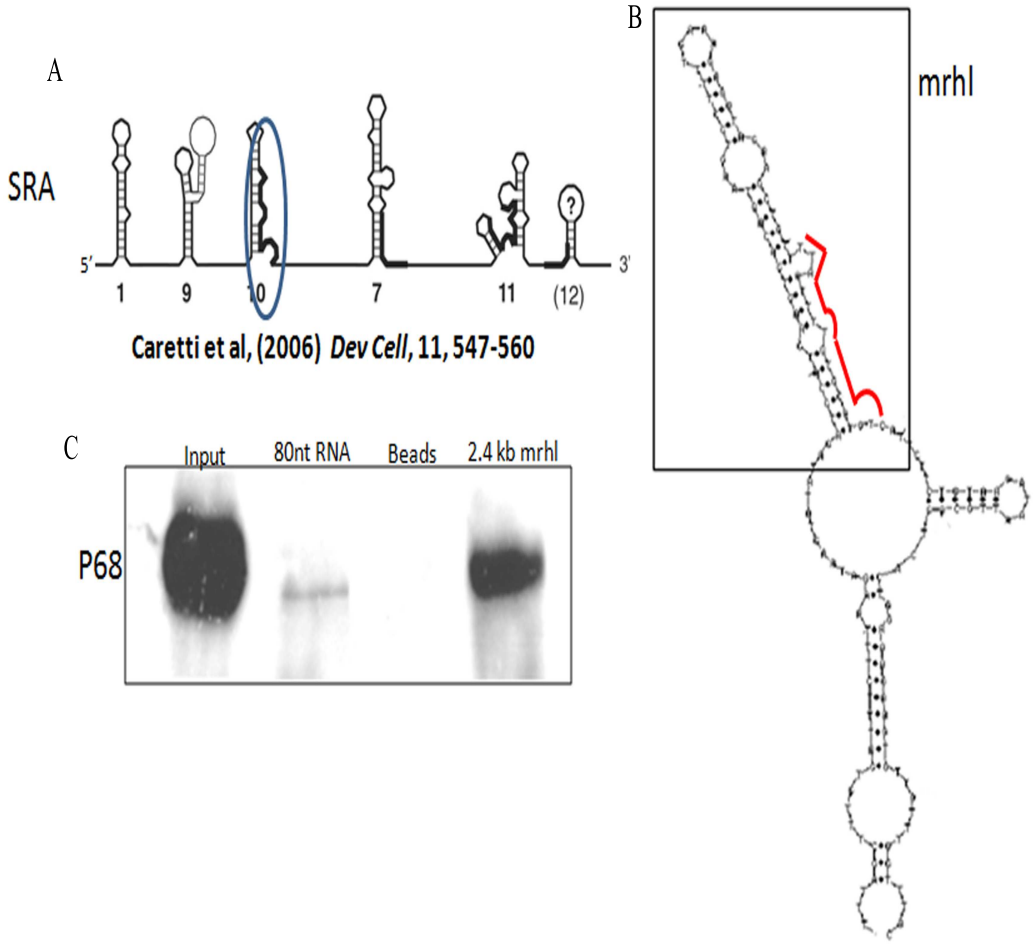

Figure 4.13: *p68 can interact with 80 nt intermediate mrhl RNA. (A) RNA secondary structure motifs present in SRA RNA(Caretti 2006). (B) Secondary structure of a part of mrhl RNA consisiting of the 80 nt intermediate RNA. Highlighted in red is the similar motif that is observed in mrhl. (C) RNA affinity pull down using in vitro transcribed 80 nt intermediate RNA followed by western using p68 antibodies that shows interaction with p68. Input is 1% of the total protein. Mock is the pull down carried out only with beads. Full length 2.4 kb mrhl was also used as an internal control.*

translocate to cytoplasm and displace axin GSK complex associated with β-catenin degradation, thereby stabilizing cytoplasmic β-catenin. The stabilized p68 and β-catenin then localizes back to the nucleus and reprogram the genes that are involved in epithelial mesenchymal transition.

Based on the results so far on the relation between *mrhl* RNA, wnt signaling and β-catenin, the subcellular distribution of p68 in normal cells and in the *mrhl* down regulated cells was examined. Having confirmed the *mrhl* down regulation by RT PCR, 48 hours after siRNA transfection (Figure 4.14A), the cells were fractionated to yield nuclear and cytoplasmic fractions. The mock transfection was performed only on lipofectamine transfected cells and was treated the same way as the experiment. The nuclear and cytoplasmic lysate of the experiment and the control were subjected to western analysis with anti p68 antibodies. As seen from Figure 4.14B, a significant amount of p68 is present in the cytoplasmic fraction of *mrhl* down regulated cells as compared to mock transfected cells. The lamin control was used to assess the purity of nuclear fraction while GAPDH served as a cytoplasmic loading control. An immuno pull down of the cytoplasmic fraction with p68 antibody was also carried followed by a western analysis with phospho tyrosine antibody of the immune precipitated cytoplasmic fraction to check whether the cytoplasmic p68 was phosphorylated. A signal in 68 kDa position with the phosphor tyrosine antibody in *mrhl* down regulated cells but not in the control was observed (Figure 4.14C). The immune precipitated proteins were also subjected to western analysis using p68 antibody which was picked up only in the *mrhl* down regulated cytoplasmic fraction as expected but not in the control (Figure 4.14C).

Further, an immuno fluoroscence experiments in Gc1-Spg cells was carried out which was transfected with *mrhl* siRNA and stained with p68 antibody. The

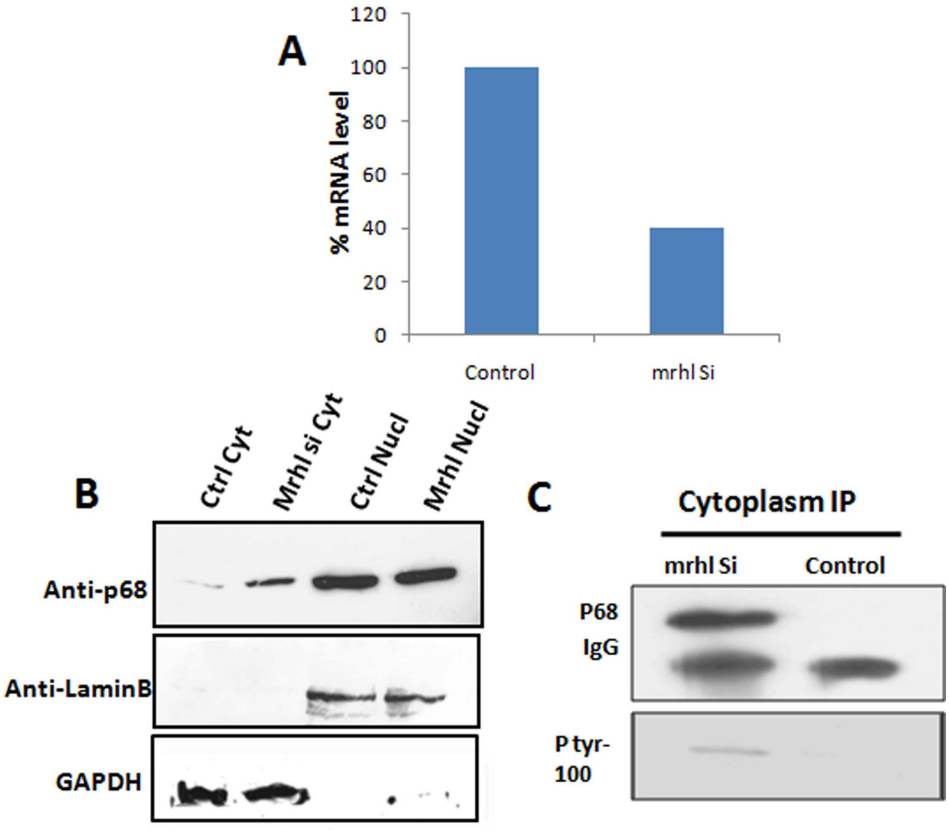

Figure 4.14: *mrhl is required for nuclear retention of p68. (A) mrhl downregulation probed by RT quantitaive PCR. (B) Western blot analysis of p68 in fractionated nucleus and cytoplasm of Gc1-Spg cells transfected with mrhl siRNA or control scrambled siRNA. Lamin was used to assess the purity of the fractions and GAPDH was used as the loading control. (C) Cytoplasmic p68 under mrhlRNA down regulation is phosphorylated at tyrosine which was probed with phospho Tyr antibody. Top panel western of p68 IP in the mrhl siRNA transfected cytoplasm and control cytoplasm. Bottom panel, p68 immunoprecipitated protein were subjected to western using phopho Tyrosine antibodies in mrhl si treated cells and the control cells.*

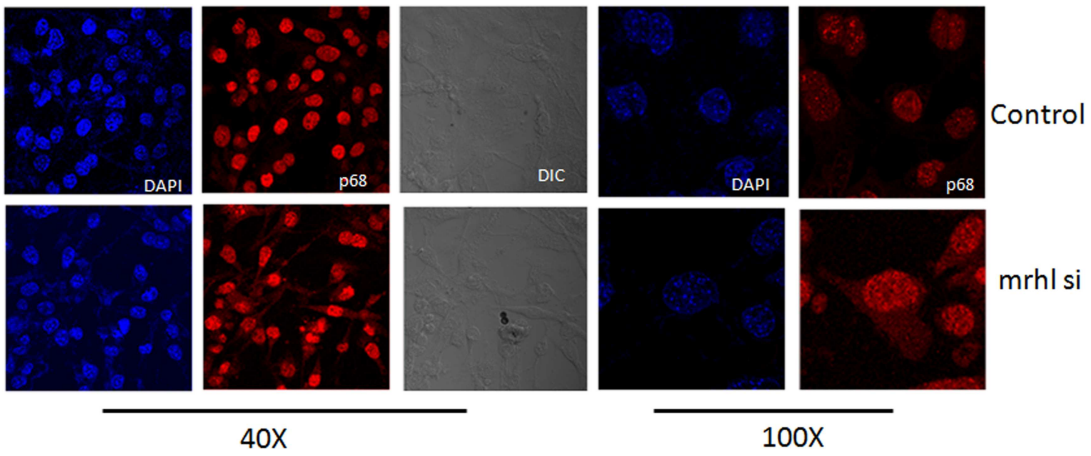

Figure 4.15: *Cytoplasmic localisation of p68 in mrhl siRNA treated cells as probed by immuno fluoroscence using anti p68 antibodies followed by Alexa 568 labeled secondary antibody. Nuclei are stained with DAPI. The representative field of magnification 40X and 100X are given.*

control was lipofectamine treated cells where immuno fluoroscence was performed in a similar way. As can be seen from the Figure 4.15 the cytoplasmic staining of p68 was observed, in addition to the nuclear staining in *mrhl* silenced cells but not in the control. Thus, this experiment conclusively shows that p68 can translocate to cytoplasm and cytoplasmic p68 in its phophorylated form probably binds to β-catenin and translocate β-catenin in to the nucleus eventually as suggested by Yang *et al.* [185].

4.7 Discussion

As has been discussed in the introduction, non-coding RNAs have been shown to play multiple roles in many cellular processes. *Mrhl* is one such RNA that has been identified and characterized as novel non-coding RNA gene in the mouse genome. From the results discussed earlier, in Chapter 3, it was presumed that

the *mrhl* could belong to the member of nuclear regulatory RNAs. Subsequently the biological function of this non-coding RNA transcript is being addressed. Two approaches were taken to address the function of this non-coding RNA. The first approach, was the silencing of *mrhl* and studying the global gene expression change using a mouse expression array. In another approach, the question of interacting partners of *mrhl* has been addressed.

4.7.1 *Mrhl* down regulation perturbs the expression of signaling and differentiation genes

As an initial step in this direction, perturbation of global gene expression following the down regulation of *mrhl* RNA by using siRNA approach was carried out. Nuclear silencing of non-coding RNAs is now fairly established though the mechanisms of nuclear RNA silencing remains to be understood completely [253]. *Mrhl* down regulation was carried out using siRNAs that target different regions of *mrhl* on the mouse spermatogonial cells. *Mrhl* silencing has resulted in marked perturbation of genes that are involved cell adhesion, cell signaling, development and differentiation. It is well known that these processes are intimately involved in cellular differentiation, embryonic development. Many non-coding RNAs identified in recent past have shown to play role in development. The first miRNAs lin4 and Let7 that were identified in *C.elegans* were involved in regulation timing of developmental signals [46]. Other non-coding RNAs such as HOTAIR that was found to regulate in trans, the HOX locus, the homeotic cluster that is involved in body patterning [118]. Many small and long non-coding RNAs have also been associated with brain development [259]. Considering the fact that the *mrhl* down regulation

also results is differential expression of genes involved in such developmental program, it would be interesting to demonstrate any developmental molecules being regulated by *mrhl*.

4.7.2 Transcription factors involved in development are upregulated

Another group of genes that are affected by *mrhl* down regulation are the transcription factors and transcription associated genes which accounted for almost 13 percent of the total perturbed genes. It was also deciphered from the list of perturbed genes (Table 4.3) that the effect of down regulation of *mrhl* RNA of some of the genes might be indirect targets of these transcription factors which might be the primary targets. Importantly, the transcription factors belonging to Bach 2, Sox and Fox group were found to be up regulated. These transcription factors again are the ones that are involved in regulating the genes involved in pluripotency and differentiation [260–262]. Many of the transcription factors were found to be upregulated upon *mrhl* silencing showing that *mrhl* should act as a repressor of these genes in the normal condition. It is also important to bring to the attention here, that the Activator protein TfAP2a was downregulated 2.5 fold in *mrhl* down regulated condition. TFAP2a is also known to be a negative regulator of wnt signaling [263], whereas the Sox transcription factors positively regulate wnt signaling and also enters into a positive feedback mechanism in the wnt pathway [261]. The significance of these findings are discussed below.

4.7.3 Wnt signaling is activated in *mrhl* down regulated cells

One of the important clues that has been obtained in this study is the possible involvement of *mrhl* non-coding RNA in the wnt signaling pathway. Many of the genes belonging to this pathway are perturbed under the condition of *mrhl* down regulation. Wnt 8B, the ligand itself, was majorly up regulated and wnt negative regulator SRFP1 was down regulated. When screened for targets of wnt signaling, upregulation of wnt target genes such as cyclin D1, and c myc along with other targets such as claudin, ephrin, c-myc binding protein etc, were observed. Other members of wnt signaling also includes the players of non canonical wnt pathway namely the Rac1 and Prickle which were also differentially regulated upon *mrhl* silencing. Intrestingly, some of the protocadherins were also found to be down regulated in the microarray experiment. Recently, it has been found that mammalian protocadherins are homologues of *Drosophila* Fat which is involved in wnt/PCP pathway [264]. The planar cell polarity pathway governs the cell movement and migration as discussed previously by reorganizing the cytoskeleton. The microarray results also show differential regulation of many of the cytoskeletal proteins(7%). Thus, from the microarray study it was clear that many genes both upstream and downstream of wnt signaling both canonical as well as non canonical pathways were perturbed upon *mrhl* silencing.

4.7.4 DEAD box helicase p68 is a *mrhl* interacting protein

A critical step in wnt signaling is the nuclear translocation of β-catenin. An earlier report form Yang *et al.* [185], has shown the involvement of a p68/Ddx5

which is an RNA helicase in this pathway during PDGF mediated epithelial mesenchymal transition in colon cancer cells. p68 belongs to family of DEAD box helicase which plays crucial roles in transcriptional as well as posttranscriptional regulation [170]. More importantly p68 itself is developmentally regulated and is a proliferation associated nuclear protein [169]. The experimental evidences presented here, by using pull down assays clearly demonstrate that *mrhl* RNA interacts with p68 helicase both *in vitro* and *in vivo*. Another major observation made in the present investigation is that down regulation of *mrhl* RNA results in the cytoplasmic localization of p68. This series of experimental observations are in agreement with the sequence of events that occurs during PDGF mediated epithelial mesenchymal transition. The report by Yang *et al.* also showed that the cytoplasmic p68 is phosphorylated at its tyrosine 305 residue. It was also observed in the present study that the cytoplasmic p68 under the condition of *mrhl* down regulation is also phosphorylated at the tyrosine residue. It was also demonstrated that β-catenin could translocate to nucleus in *mrhl* downregulated which is also consistent with the observation of Yang *et al.* However, it should be brought to the attention here that there has been a report by Stucke *et al.* [265] contradicting the model of Yang *et al.* But the observations made in the present investigation are in complete agreement with that of Yang *et al.* and hence it is suggested that *mrhl* RNA is an important component of the process of development and differentiation involving wnt signaling pathway and β-catenin as the effector molecule. Wnt signaling is also an important signaling pathway in the differentiation of embryonic stem cells and in this context it is interesting to point out that *mrhl* RNA is also expressed in mouse embryonic stem cells.

Although the domain of the non-coding RNA which interacts with p68 in the

present work has not been addressed, a preliminary computational exercise shows that a domain with a particular stem loop secondary structure of SRA is shown to have a co activator role with p68 [182, 240] is also observed in the *mrhl* non-coding RNA. SRA is a non-coding RNA partner of p68 and together they are involved in regulation of transcription of myoD and steroid receptor dependent transcription. From the literature, it is known that the p68 is regulates diverse cellular processes independently and also together with protein or RNA binding partners [170]. In this context, identification of *mrhl* as a putative partner could give valuable insight to the role of p68 and *mrhl* when present as a complex in the nucleus. Together they can have different chromatin substrates and transcription targets which needs to be uncovered. It is interesting to note here that in human cells p68 down regulation has resulted in upregulation of SFRP2 and down regulation of troponin and other genes involved in muscle differentiation [182], whereas *mrhl* silencing results in downregulation of SFRP and up regulation of troponin, and myo(synaptopodin). In addition to the involvement of *mrhl* p68 complex in wnt signaling and β-catenin mediated regulation of gene expression, it is very likely that the complex of p68 and *mrhl* RNA may have additional direct roles in transcriptional regulation of specific chromatin domains. It was shown in chapter 3 that the *mrhl* RNA is also associated chromatin fraction and is localized at several foci within the nucleus. Characterization of these chromatin domains bound by the *mrhl* RNA will give valuable insights into the biological function of this mouse specific non-coding RNA.

It would also be very interesting to identify and characterize in detail the complex of *mrhl* RNA with p68, with particular reference to its components and the dynamics that leads to phophorylation/dissociation and cytoplasmic translocation

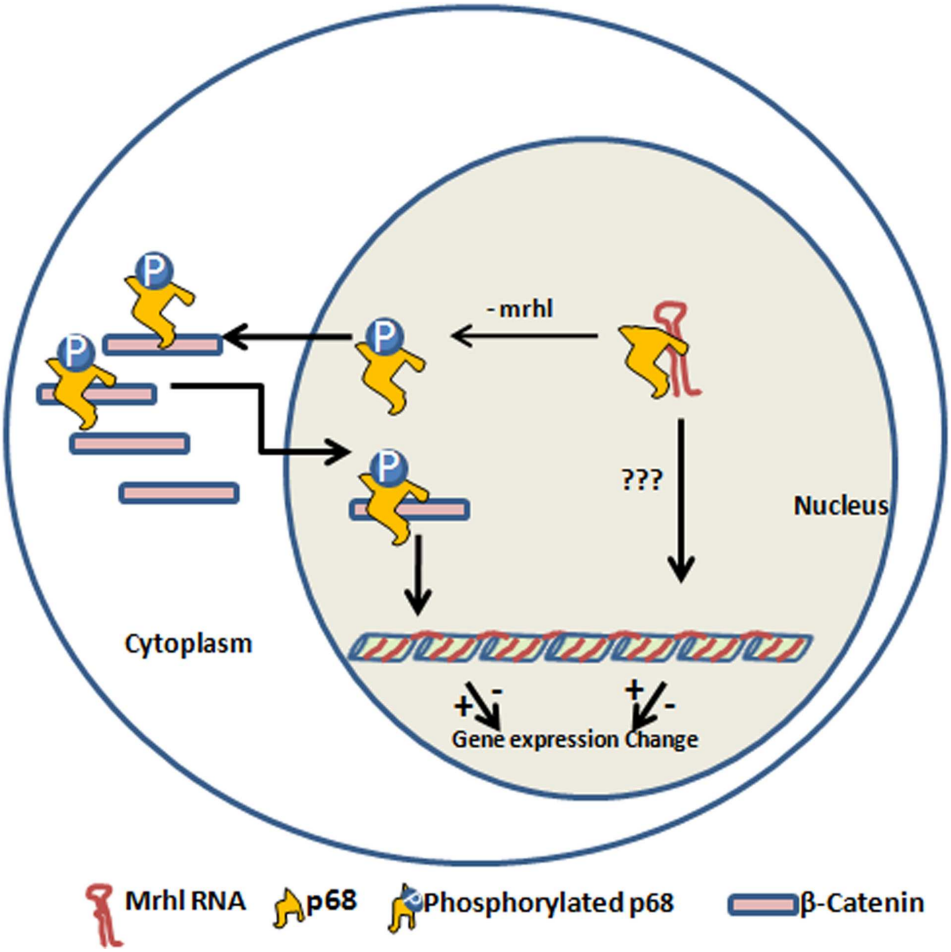

Figure 4.16: *Hypothetical model depicting the role of mrhl in cells. In mrhl down regulated condition, there is a p68 phosphorylation and translocation of p68 to cytoplasm from nucleus. The translocated p68 binds to β-catenin and results in cytoplasmic stablisation and nuclear translocation of β-catenin which eventually activates or repress genes that are β-catenin dependent. On the other hand, it is possible, when mrhl RNA is present it is bound to p68 and the mrhl-p68 complex could act on chromatin and bring about gene expression changes.*

of p68 to trigger the nuclear translocation of β-catenin. It is also interesting to note that Dusp11, a dual specificity phosphatase, which interacts with RNA/RNP, was picked up as one of the interacting protein of *mrhl* RNA [266]. It is known that Abl kinase phosphorylates p68 in the nucleus [185]. Since phosphorylation of p68 promotes cytoplasmic translocation of the protein, it would be of interest to identify the mechanism that is operated via *mrhl* that would retain the p68 in the nucleus. In this context, identification of Dusp11 a dual specificity phosphatase as one of the binding partners is very interesting. Dusp11 is a RNA/RNP binding phosphatase, one can envisage that *mrhl* binding to p68 could allow the Dusp11 also to bind to the complex as it is a RNP binding phosphatase. The Dusp11 could maintain p68 in dephosphorylated condition in the presence of *mrhl* and prevent the translocation of p68 to cytoplasm which is promoted in the *mrhl* silenced condition. Thus identification of entire *mrhl* binding complex will allow us to understand in more detail the exact mechanism and the components underlying p68 regulation and wnt signaling.

Based on all these observations a hypothetical model showing the involvement of *mrhl* RNA regulation of gene expression of subset of genes is proposed which is shown in Figure 4.16 where in two possible mechanisms are implicated, one through the involvement of β-catenin and the other, as direct effect of p68 and *mrhl* RNA complex on a subset of chromatin domains. A more thorough investigation is required to validate this model and show a novel nuclear regulatory function that is coupled to transcription and cell signaling for this *mrhl* non-coding RNA.

Chapter 5

Summary

Analysis of the draft sequence of the human and mouse genome has revealed that coding sequences account for less than 2% of its total transcriptome. Several non-coding RNAs (ncRNAs) have been identified whose function are yet to be determined. By definition, non-coding RNA refers to the RNA that are transcribed but not translated to proteins. Many possible functions for these non-coding RNA have been postulated. The eukaryotic genome encodes two distinct categories of other ncRNAs, other than the classical rRNA, tRNA, snRNA, referred to as small ncRNAs and long mRNA-like ncRNAs. Much of the recent attention has been focused on the small RNA that are 22-33 nucleotide in length which includes miRNAs, siRNAs, piRNA etc. Majority of the small RNAs play their functional role in gene silencing by interaction with chromatin or by base pairing with complementary mRNA or DNA. The long ncRNAs, which are transcribed by RNA polymerase II, spliced and polyadenylated, like Xist, Tsix, Air, H19 and Rox etc., are implicated in variety of regulatory processes such as imprinting, X-chromosome inactivation, DNA methylation, transcription, RNA interference, chromatin-structure dynamics

and antisense regulation.

Previously, a non-coding RNA that was encoded in mouse chromosome 8 was identified in a locus harbouring meiotic recombination hot spot activity in our laboratory. This 2.4 kb transcript (*mrhl*) was shown to be an unspliced, polyadenylated transcript expressed in multiple mouse tissues. The transcript does not possess a significant open reading frame but had considerable propensity to form a stable secondary structure. We have since then begun to probe the function of this non-coding RNA. The present thesis is a step towards elucidating the function of the non-coding RNA that is transcribed from mouse chromosome 8 and the objectives laid were to a) characterize the 2.4 kb non-coding *mrhl* RNA, b) identify the interacting partners and c) understand the function of this RNA in cellular context.

We have found in the present study that this 2.4 kb *mrhl* RNA is a nuclear restricted RNA and is not translated *in vitro* in a coupled transcription and translation assay. Thus it was concluded that *mrhl* is an authentic nuclear non-coding RNA. It was also observed that the 2.4 kb RNA is processed further to yield an 80 nt intermediate RNA by northern analysis. This processed region has been delineated from the 2.4 kb primary transcript by multiple northern blot analyses using probes spanning different regions of the 2.4 kb RNA. RNA secondary structure predicts a stem loop in this 80 nt intermediate RNA sequence. miRNA processing assay was performed to demonstrate that this 80 nt intermediate RNA is indeed generated by Drosha, a nuclear RNaseIII enzyme. Drosha silencing also resulted in reduced accumulation of this RNA *in vivo*. It could also be demonstrated that the 80 nt RNA could be processed further to 22 nt RNA by recombinant Dicer in an *in vitro* reaction. However, the mature 22 nt miRNA could not be detected

in vivo even using sensitive techniques such as splinted ligation or upon over expressing the *mrhl* primary transcript in mouse spermatogonial GC1-Spg cell lines suggesting a possible regulation at the nucleus. Fractionation study was performed to look at the enrichment of the 80 nt RNA in the sub cellular fractions and we observe that the 80 nt intermediate RNA was also nuclear restricted and was not found in cytoplasm explaining the absence of 22 nt RNA, which is a cytoplasmic reaction. Localization of both primary and the intermediate *mrhl* in the GC1-Spg mouse spermatogonial cells by RNA FISH using LNA probes shows a distinct punctuate nuclear localization signal for both 2.4 kb primary transcript as well as processed 80 nt species where we could see colocalisation. However, the 80 nt intermediate RNA indeed had more foci that did not colocalise with the 2.4 kb primary transcript. Combined RNA FISH with immunofluroscence with Drosha and Nucleolin demonstrated that nucleolus is the possible site of processing of the primary *mrhl* transcript to the 80 nt RNA. The subnuclear fractions of chromatin and nucleoplasm as probed by northern on 5% and 15% Urea PAGE using respective probes, showed enrichment of 80 nt RNA in the chromatin and 2.4 kb RNA in the nucleoplasm. The results thus obtained on the characterization of *mrhl* RNA favors a potential regulatory role for the RNA within the nucleus.

Our next aim was to study the functional role of the RNA in cellular context. In order to address this we took two approaches. First silencing experiment was performed by which we down regulated *mrhl* and performed a global gene expression perturbation using Affymetrix 43K mouse expression array. A list of about 260 genes was shortlisted which had $P \leq 0.05$ and showed fold change of more than 2. The genes were classified based on its GO term. Interestingly, it was observed that several of the cell adhesion molecules were differentially regulated

upon silencing of *mrhl*. It was also observed that genes involved in cell differentiation, signaling and development and particularly genes involved in wnt signaling to show differential expression upon *mrhl* down regulation as against the scrambled siRNA control. This gave us a possible clue into the function of *mrhl* as a developmental regulator that could act via the wnt pathway. The wnt pathway genes were validated by RT QPCR and wnt targets that were represented in the array were also identified from the microarray data. Up regulation of Cyclin D1, c-myc and other wnt targets were observed as well. Further investigation was carried out in this direction, and studied the effect of the wnt activation on the fate of β-catenin, the effector molecule of wnt signaling in *mrhl* down regulated condition. Wnt is a very conserved signaling pathway and activation of wnt results in cytoplasmic stabilization of β-catenin which then translocates to nucleus and activates genes responsive to wnt binding to TCF/LEF family transcription factors. In the absence of wnt ligand, it rapidly undergoes degradation via proteosome mediated pathway. In the experiments performed it was observed that the, cytoplasmic β-catenin was stablized and there was also nuclear accumulation of the β-catenin upon *mrhl* down regulation as revealed by immuno fluorescence.

Many of the RNAs discovered recently have been shown to interact with huge protein complexes or a part of the complex itself for their functional manifestation. There are non-coding RNAs that interact with the polycomb proteins known as PcG RNA that are involved in embryonic stem cell differentiation. Thus it is becoming apparent that RNA interaction with proteins could mediate potential regulatory roles. In this direction, to study the *mrhl* RNA interacting proteins, initially a northwestern blot was performed using full length *mrhl* RNA identified at least 4 interacting proteins in the molecular weight range of 120 kDa,

100 kDa, 50kDa, 65 kDa. In order to identify the proteins a RNA affinity pull down experiment was performed. Here, the *in vitro* transcribed biotinylated RNA was incubated with cell lysate and RNA protein complex were pulled down using streptavidin beads and the interacting proteins were identified through mass spectrometric analysis. p68 or DDX5, a DEAD box RNA helicase was identified being a putative RNA binding protein partner. This interaction was validated further through western blot analysis using antibody against p68 *in vitro*. *In vivo* validation was performed by streptavidin aptamer pull down in *mrhl* S1aptamer transfected cell line. The *mrhl* S1 pulled down proteins were subjected to western blot analysis which also confirmed the p68 *mrhl* interaction *in vivo*.

The interaction studies of *mrhl* resulted in identification of p68 as putative protein partner whereas, down regulation of *mrhl* resulted in the activation of wnt signaling. In order to study the cross talk between these observations a working hypothesis was developed based on a previous report by Yang *et al.* [185] where they have shown that p68 is tyrosine phophorylated and the phosphorylated p68 could translocate to cytoplasm. The cytoplasmic p68 interacts with β-catenin and dissociate Axin GSK complex thereby stablising the cytoplasmic β-catenin. The β-catenin and p68 then translocates back to the nucleus and activates the genes that program epithelial mesenchymal transition which includes various cell adhesion molecules and the genes that are β-catenin dependent. Thus to substantiate the results, based on this previous finding, *mrhl* was down regulated in GC1 spg cells. In the *mrhl* down regulated background we did find cytoplasmic localization of p68. The tyrosine phosphorylation status of the p68 probed using phospho tyrosinr antibody on immunoprecipitated cytoplasmic p68 in *mrhl* down regulated cells, showed that the cytoplasmic p68 was indeed phosphorylated in its tyrosine residue.

In conclusion, *mrhl* is a nuclear non-coding RNA that is processed to a 80 nt RNA product via miRNA processing machinery involving Drosha. The 80 nt RNA has a stem loop structure, and thus it qualifies to be a bonafide primary miRNA processed intermediate transcript. However the mature miRNA of 22 nt in length is not found *in vivo*. The distinct punctate nuclear localization of the primary and processed RNA favors a vital regulatory function in the nucleus. It has also been shown in this study, the involvement of a nuclear non-coding RNA in wnt signaling pathway mediated through its interacting partner p68. We surmise, from the results obtained that the *mrhl* is involved in nuclear retention of p68 a DEAD box helicase protein which is also involved in various transcriptional and post transcriptional regulatory events.

It is also interesting to note here, that *mrhl* down regulation also has influence on the expression of many transcription factors which are involved in developmental program, such as FOX and Sox group of factors. From the chromatin fractionation study, we know that the *mrhl* RNA is enriched in the chromatin fraction. Based on these results, it would be a fruitful exercise to address the chromatin domains that are bound with this RNA. An experiment like large scale sequencing of the chromatin that is enriched with *mrhl* RNA can give valuable insights to the *mrhl* RNA bound regions of the chromatin.

Another interesting observation from the present work is that the activation of wnt signaling in the *mrhl* down regulation cells. It is well established that the wnt signaling is a very fundamental signaling involved in the development of multicellular organisms. Perturbation in this signal results in severe developmental defects. Keeping in mind the importance of this pathway in differentiation process, it would be interesting to study the role of *mrhl* in the embryonic stem cells. Interestingly,

it has been observed that the *mrhl* RNA both the primary transcript as well as the processed 80 nt intermediate is expressed in the mouse ES cells. Establishing the function of this RNA in the cellular differentiation process would have a profound impact in understanding the involvement of such non-coding RNAs in development. It has to be brought to the attention that this RNA is a poorly conserved RNA having homologue only in the rat species. However, it is now well established that such non conserved functional RNAs do exist as discussed in the Chapter 1. The non conservation of *mrhl* could be explained in two ways. One, in which the proteins that are interacting with the RNA are much conserved across species such as p68. One should consider here the secondary structure of the RNA, which could be conserved without primary nucleotide sequence conservation, and such motifs that are conserved should be identified from the other species which would eventually result in identification of novel non-coding RNAs involved in similar cellular aspects. In this context, it would be interesting to study the region of *mrhl* harbouring secondary structure that might be involved in p68 binding by carrying out extensive domain wise characterization of the p68 protein as well the *mrhl* RNA.

Another, hypothesis could be that the non conservation could reflect on the evolution of the species. If one considers the protein coding vs non-coding regions across species, it is the non-coding transcripts that show a gradual increase in number in contrast to protein coding genes which is almost constant in all higher eukaryotes. Thus, such a paradox in the complexity and evolution of species could be explained by the identification of such non-coding RNA particularly when they function in developmental program such as *mrhl*.

The significance of non-protein-coding RNAs as central components of various

cellular processes is being increasingly appreciated over the recent years. One particular intriguing aspect of the non-coding transcriptome is its potential to fill the regulatory gap created by the surprisingly low number of protein-coding genes in higher organisms. Given the documented versatility and catalytic power of RNA as well as the existence of numerous RNAs with little or no known function, it seems certain that additional roles for these transcripts will throw light on the complex gene regulatory networks at cellular as well as organismal level. We, in this study have only addressed a tip of an iceberg, where, further research awaits in understanding the global picture of these non-coding RNAs in cellular homeostasis.

Annexure 1

Complete microarray list of differentially expressed genes upon mrhl down regulation. P Value ≤ 0.05, Fold Change ≥ 1.5.

Gene Title	Gene Symbol	Fold Change	regulation
ghrelin	Ghrl	8.824283	down
androgen binding protein beta	Abpb	8.539614	down
secreted frizzled-related sequence protein 1	Sfrp1	7.453029	down
killer cell lectin-like receptor, subfamily A, member 15	Klra15	7.2197695	down
oxytocin receptor	Oxtr	7.1553125	down
von Willebrand factor C and EGF domains	Vwce	6.866967	down
ribosomal protein S6 kinase polypeptide 6	Rps6ka6	6.567401	down
serine/threonine kinase 32C	Stk32c	6.2927437	down
Mammary gland RCB-0526 Jyg-MC(A) cDNA		6.2230244	down
solute carrier family 35, member D3	Slc35d3	6.217865	down
zinc finger protein 1, Y linked	Zfy1	6.059308	down
DNA segment, Chr 13, ERATO Doi 205, expressed	D13Ertd205e	5.863972	down
vomeronasal 2, receptor, 13	V2r13	5.8313546	down
TSC22 domain family, member 1	Tsc22d1	5.8221645	down
keratin associated protein 6-3	Krtap6-3	5.4691353	down
similar to voltage gated channel like 1	LOC675405	5.4171133	down
thyroid stimulating hormone receptor	Tshr	5.3076963	down
Predicted gene, EG625068	EG625068	5.2122216	down
potassium inwardly-rectifying channel, subfamily J, member 2	Kcnj2	5.011891	down
regulated endocrine-specific protein 18	Resp18	4.9556894	down
pannexin 3	Panx3	4.9348826	down
Forkhead box L2 opposite strand transcript	Foxl2os	4.907441	down
cortactin	Cttn	4.8588705	down
calbindin-28K	Calb1	4.784588	down
Actin binding LIM protein family, member 3	Ablim3	4.7800193	down
presenilin 1	Psen1	4.7645454	down
CKLF-like MARVEL transmembrane domain containing 2A	Cmtm2a	4.758477	down
IQ motif containing F3	Iqcf3	4.7094636	down
Small nuclear ribonucleoprotein N	Snrpn	4.605408	down
potassium voltage-gated channel, Isk-related subfamily, gene 2	Kcne2	4.4854107	down
X transporter protein 3 similar 1 gene	Xtrp3s1	4.4779925	down
interleukin 17D	Il17d	4.4444876	down
transcription factor AP-2, alpha	Tcfap2a	4.346088	down
Serine (or cysteine) peptidase inhibitor, clade I, member 1	Serpini1	4.335833	down
synuclein, alpha interacting protein (synphilin)	Sncaip	4.3234158	down
transmembrane protease, serine 3	Tmprss3	4.241809	down
calbindin-28K	Calb1	4.1369605	down
cDNA sequence BC051628	BC051628	4.0682096	down
claudin 16	Cldn16	4.0280867	down
alanine and arginine rich domain containing protein	Aard	4.024166	down
erythrocyte protein band 4.2	Epb4.2	3.8372006	down

Coronin, actin binding protein 2A	Coro2a	3.8262446	down
a disintegrin-like and metallopeptidase with thrombospondin type 1 motif, 20	Adamts20	3.6854324	down
vomeronasal 1 receptor, C8	V1rc8	3.6835787	down
zinc finger protein 708	Zfp708	3.6378329	down
latent transforming growth factor beta binding protein 1	Ltbp1	3.584756	down
necdin	Ndn	3.5554175	down
synovial sarcoma, X breakpoint 2 interacting protein	Ssx2ip	3.4693851	down
desmoglein 2	Dsg2	3.4438908	down
pleckstrin	Plek	3.4369526	down
cytochrome b reductase 1	Cybrd1	3.4317017	down
ornithine transcarbamylase	Otc	3.37796	down
sodium channel, voltage-gated, type III, beta	Scn3b	3.3172169	down
hyperpolarization-activated, cyclic nucleotide-gated K+ 3	Hcn3	3.2536252	down
protocadherin beta 22	Pcdhb22	3.247978	down
odontogenic, ameloblast asssociated	Odam	3.1740217	down
Acetyl-Coenzyme A carboxylase beta	Acacb	3.1598532	down
dynein heavy chain domain 1	Dnhd1	3.156872	down
DIRAS family, GTP-binding RAS-like 1	Diras1	3.1172512	down
integrin beta 2-like	Itgb2l	3.0264037	down
osteoclast associated receptor	Oscar	3.0000458	down
minichromosome maintenance deficient domain containing 1	Mcmdc1	2.935403	down
killer cell lectin-like receptor subfamily A, member 22	Klra22	2.9088423	down
potassium voltage-gated channel, subfamily H (eag-related), member 1	Kcnh1	2.8852658	down
Immunoglobulin superfamily, member 11	Igsf11	2.8157163	down
centrin 1	Cetn1	2.7973218	down
potassium voltage-gated channel, shaker-related, subfamily, member 6	Kcna6	2.6637957	down
membrane targeting (tandem) C2 domain containing 1	Mtac2d1	2.604103	down
Kruppel-like factor 12	Klf12	2.6019516	down
protocadherin alpha 11	Pcdha11	2.5869217	down
alcohol dehydrogenase 5 (class III), chi polypeptide	Adh5	2.5710175	down
immunoglobulin superfamily, member 4B	Igsf4b	2.5517106	down
RIKEN cDNA 9430020K01 gene	9430020K01Rik	2.5287862	down
folate hydrolase	Folh1	2.5101964	down
neuropeptide Y receptor Y2	Npy2r	2.4657125	down
ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 6	St8sia6	2.464091	down
Down syndrome cell adhesion molecule-like 1	Dscaml1	2.4388938	down
DNA segment, Chr 7, Wayne State University 130, expressed	D7Wsu130e	2.4378269	down
arginase type II	Arg2	2.4290082	down
transmembrane protein 56	Tmem56	2.418868	down
amyotrophic lateral sclerosis 2 (juvenile) homolog (human)	Als2	2.3935883	down
TXK tyrosine kinase	Txk	2.3792949	down
natriuretic peptide receptor 3	Npr3	2.3734481	down
B-cell CLL/lymphoma 11A (zinc finger protein)	Bcl11a	2.367807	down
thrombospondin 4	Thbs4	2.3644867	down
peroxisome proliferative activated receptor, gamma, coactivator 1 alpha	Ppargc1a	2.360956	down
hypothetical protein LOC574403	MGC99845	2.3542194	down
RAS-related C3 botulinum substrate 1	Rac1	2.3511782	down
aldehyde dehydrogenase family 1, subfamily A3	Aldh1a3	2.3363826	down

histocompatibility 28	H28	2.2999513	down
Odd Oz/ten-m homolog 4 (Drosophila)	Odz4	2.2945127	down
tumor necrosis factor receptor superfamily, member 19	Tnfrsf19	2.292577	down
ADP-ribosylation factor-like 4A	Arl4a	2.285053	down
makorin, ring finger protein, 1	Mkrn1	2.2702432	down
neuregulin 3	Nrg3	2.237481	down
Protein serine kinase H1	Pskh1	2.2227247	down
killer cell lectin-like receptor family E member 1	Klre1	2.2084706	down
chordin-like 1	Chrdl1	2.2056174	down
expressed sequence BB182387	BB182387	2.2019928	down
apolipoprotein A-V	Apoa5	2.19067	down
T-cell receptor gamma chain	Tcrg	2.182569	down
transmembrane protein 167	Tmem167	2.1728487	down
Ribosomal protein L27a	Rpl27a	2.1153889	down
melanoma antigen, family B, 3	Mageb3	2.1092572	down
Fanconi anemia, complementation group M	Fancm	2.1047366	down
neurexin I	Nrxn1	2.0947363	down
reduced expression 2	Rex2	2.0922077	down
Protein phosphatase 2, regulatory subunit B (B56), alpha isoform	Ppp2r5a	2.0873513	down
G protein-coupled receptor 45	Gpr45	2.07894	down
Exocyst complex component 6B	Exoc6b	2.050361	down
solute carrier family 5 (sodium/glucose cotransporter), member 12	Slc5a12	2.0459132	down
solute carrier family 27 (fatty acid transporter), member 3	Slc27a3	2.0409288	down
chemokine (C-X-C motif) ligand 15	Cxcl15	2.0307386	down
huntingtin interacting protein 2	Hip2	2.024571	down
potassium voltage gated channel, Shaw-related subfamily, member 1	Kcnc1	1.992777	down
ATPase, Na+/K+ transporting, alpha 2 polypeptide	Atp1a2	1.9898292	down
homeo box D4	Hoxd4	1.9792013	down
coiled-coil domain containing 19	Ccdc19	1.9711522	down
dickkopf homolog 3 (Xenopus laevis)	Dkk3	1.9689758	down
testis expressed gene 101	Tex101	1.9656849	down
RAD51-like 1 (S. cerevisiae)	Rad51l1	1.9592172	down
bruno-like 4, RNA binding protein (Drosophila)	Brunol4	1.9586202	down
solute carrier family 7 (cationic amino acid transporter, y+ system), member 11	Slc7a11	1.9582862	down
cytochrome P450, family 2, subfamily g, polypeptide 1	Cyp2g1	1.9424107	down
Src-like-adaptor 2	Sla2	1.9356142	down
zinc and ring finger 2	Znrf2	1.9305444	down
paired box gene 8	Pax8	1.9058213	down
Progesterone receptor	Pgr	1.9008353	down
UDP-glucose ceramide glucosyltransferase-like 1	Ugcgl1	1.8914635	down
glucosamine-6-phosphate deaminase 2	Gnpda2	1.8871695	down
phospholipase C-like 2	Plcl2	1.879651	down
ATP-binding cassette, sub-family B (MDR/TAP), member 1B	Abcb1b	1.8746468	down
0 day neonate skin cDNA		1.860754	down
cadherin 8	Cdh8	1.8598398	down
cannabinoid receptor 1 (brain)	Cnr1	1.8528757	down
sperm adhesion molecule 1	Spam1	1.8502244	down
nuclear receptor interacting protein 3	Nrip3	1.8415687	down
Ring finger protein 43	Rnf43	1.8332926	down

Rab40b, member RAS oncogene family	Rab40b	1.8261486	down
cystin 1	Cys1	1.8185573	down
mutated in colorectal cancers	Mcc	1.8151777	down
dachshund 1 (Drosophila)	Dach1	1.8130869	down
Vacuolar protein sorting 13B (yeast)	Vps13b	1.8123155	down
gene model 1960, (NCBI)	Gm1960	1.8095125	down
sorting nexin 6	Snx6	1.8056029	down
glutamate receptor, metabotropic 4	Grm4	1.7973747	down
crystallin, beta A4	Cryba4	1.7940987	down
serine (or cysteine) peptdiase inhibitor, clade B, member 8	Serpinb8	1.7929515	down
RAB30, member RAS oncogene family	Rab30	1.7916994	down
Interleukin 1 receptor accessory protein-like 1	Il1rapl1	1.7888286	down
Adult male corpora quadrigemina cDNA		1.7807056	down
Pseudouridylate synthase 7 homolog (S. cerevisiae)	Pus7	1.7736651	down
matrilin 1, cartilage matrix protein 1	Matn1	1.7715887	down
Potassium voltage-gated channel, subfamily Q, member 5	Kcnq5	1.7656866	down
ELMO domain containing 1	Elmod1	1.7643262	down
gene model 550, (NCBI)	Gm550	1.7629039	down
potassium voltage-gated channel, subfamily Q, member 5	Kcnq5	1.7553399	down
hypothetical gene supported by AK052160	LOC432868	1.7545301	down
defensin beta 35	Defb35	1.7542205	down
DEAD (Asp-Glu-Ala-Asp) box polypeptide 19b	Ddx19b	1.7510388	down
transformation related protein 53 inducible protein 5	Trp53i5	1.7459875	down
sodium channel, voltage-gated, type II, alpha 1	Scn2a1	1.7420155	down
leucine rich repeat containing 4	Lrrc4	1.7299373	down
carboxylesterase 3	Ces3	1.7273406	down
guanine nucleotide binding protein (G protein), gamma 4 subunit	Gng4	1.7244552	down
adenosine A2b receptor	Adora2b	1.7168142	down
3-hydroxyisobutyrate dehydrogenase	Hibadh	1.714085	down
protein phosphatase 1A, magnesium dependent, alpha isoform	Ppm1a	1.7060726	down
nephronophthisis 4 (juvenile) homolog (human)	Nphp4	1.7060224	down
microtubule associated serine/threonine kinase family member 4	Mast4	1.7015119	down
RNA imprinted and accumulated in nucleus	Rian	1.6961976	down
tyrosine kinase, non-receptor, 1	Tnk1	1.6924574	down
SLAIN motif family, member 1	Slain1	1.6858076	down
RIKEN cDNA 1110021J02 gene	1110021J02Rik	1.6843935	down
proteasome (prosome, macropain) 26S subunit, non-ATPase, 9	Psmd9	1.6816317	down
cytochrome P450, family 2.	Cyp2c37	1.6793411	down
mitochondrial ribosomal protein L32	Mrpl32	1.679179	down
osteopetrosis associated transmembrane protein 1	Ostm1	1.6787639	down
shroom family member 1	Shroom1	1.6719311	down
peroxisome proliferative activated receptor, gamma, coactivator 1 alpha	Ppargc1a	1.6676657	down
Cysteine and histidine-rich domain (CHORD)-containing, zinc-binding protein 1	Chordc1	1.6660378	down
5,10-methylenetetrahydrofolate reductase	Mthfr	1.6645987	down
CLIP associating protein 2	Clasp2	1.6590799	down
Aconitase 2, mitochondrial	Aco2	1.6588173	down
Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6	Arhgef6	1.6557909	down
5-hydroxytryptamine (serotonin) receptor 1A	Htr1a	1.6454349	down
hyperpolarization-activated, cyclic nucleotide-gated K+ 1	Hcn1	1.6440294	down

DNA segment, Chr 2, ERATO Doi 239, expressed	D2Ertd239e	1.6424857	down
calcium/calmodulin-dependent protein kinase II inhibitor 1	Camk2n1	1.6361611	down
solute carrier family 5 (sodium/glucose cotransporter), member 9	Slc5a9	1.6335927	down
growth hormone secretagogue receptor	Ghsr	1.63355	down
Ras-related GTP binding D	Rragd	1.6273764	down
chloride channel Ka	Clcnka	1.6248251	down
midline 2	Mid2	1.6220553	down
SH3 domain binding glutamic acid-rich protein like 2	Sh3bgrl2	1.6209261	down
WNK lysine deficient protein kinase 3	Wnk3	1.6161704	down
fibromodulin	Fmod	1.611188	down
abhydrolase domain containing 13	Abhd13	1.6105531	down
histocompatibility 28	H28	1.6090543	down
zinc finger and BTB domain containing 8 opposite strand	Zbtb8os	1.6069704	down
LIM motif-containing protein kinase 2	Limk2	1.6068318	down
Zinc finger protein 93	Zfp93	1.605346	down
myc induced nuclear antigen	Mina	1.6021843	down
GTP binding protein (gene overexpressed in skeletal muscle)	Gem	1.6020193	down
hypothetical protein B930075F07	B930075F07	1.5992023	down
cleavage and polyadenylation specific factor 2	Cpsf2	1.5990644	down
Frizzled homolog 6 (Drosophila)	Fzd6	1.5974157	down
CDC14 cell division cycle 14 homolog A (S. cerevisiae)	Cdc14a	1.5925483	down
ATP-binding cassette, sub-family A (ABC1), member 5	Abca5	1.5921804	down
synaptotagmin XI	Syt11	1.5877944	down
ankyrin repeat and sterile alpha motif domain containing 1B	Anks1b	1.5847445	down
G protein-coupled receptor 155	Gpr155	1.5817263	down
transformation related protein 53 regulating kinase	Trp53rk	1.5798626	down
BAT2 domain containing 1	Bat2d	1.5755575	down
protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 2	Ppfia2	1.5739301	down
c-myc binding protein	Mycbp	1.5732329	down
interleukin 1 receptor-like 2	Il1rl2	1.5711889	down
ATPase, class I, type 8B, member 1	Atp8b1	1.5710179	down
transmembrane protein 174	Tmem174	1.5706334	down
Transformation related protein 53 binding protein 1	Trp53bp1	1.5697446	down
guanine nucleotide binding protein (G protein), gamma 12	Gng12	1.566291	down
CD320 antigen	Cd320	1.5648824	down
Zinc finger protein 787	Zfp787	1.5622659	down
potassium voltage-gated channel, subfamily H (eag-related), member 7	Kcnh7	1.5619024	down
matrix Gla protein	Mgp	1.560794	down
cannabinoid receptor 1 (brain)	Cnr1	1.560368	down
retinoschisis (X-linked, juvenile) 1 (human)	Rs1	1.5601835	down
glycoprotein 1b, alpha polypeptide	Gp1ba	1.5575476	down
D site albumin promoter binding protein	Dbp	1.5534114	down
phosphoglucomutase 2-like 1	Pgm2l1	1.552995	down
tumor necrosis factor receptor superfamily, member 18	Tnfrsf18	1.5514305	down
Glucocorticoid induced transcript 1	Glcci1	1.5495039	down
SCO-spondin	Sspo	1.5489168	down
calcium channel, voltage-dependent, alpha2/delta subunit 1	Cacna2d1	1.5489028	down
homer homolog 1 (Drosophila)	Homer1	1.5419091	down
Cyclin D3	Ccnd3	1.5398498	down

gene model 50, (NCBI)	Gm50	1.5367199	down
RAR-related orphan receptor beta	Rorb	1.5333772	down
gametogenetin binding protein 1	Ggnbp1	1.5329114	down
phosphatidylinositol-specific phospholipase C, X domain containing 1	Plcxd1	1.5276464	down
solute carrier family 12, member 1	Slc12a1	1.5268003	down
CD160 antigen	Cd160	1.5250115	down
solute carrier family 25, member 27	Slc25a27	1.5249126	down
c-myc binding protein	Mycbp	1.5101599	down
Visual cortex cDNA, RIKEN full-length enriched library,		1.5005847	down
BTB and CNC homology 2	Bach2	7.095153	up
B-cell leukemia/lymphoma 2 related protein A1a	Bcl2a1a	6.9877853	up
Microtubule-associated protein 4	Mtap4	6.771951	up
procollagen, type III, alpha 1	Col3a1	6.6534495	up
paraneoplastic antigen MA2	Pnma2	6.644585	up
solute carrier family 25, member 31	Slc25a31	6.2134767	up
lipolysis stimulated lipoprotein receptor	Lsr	6.032483	up
PTC7 protein phosphatase homolog (S. cerevisiae)	Pptc7	6.025744	up
tribbles homolog 2 (Drosophila)	Trib2	5.869255	up
synaptopodin 2	Synpo2	5.645597	up
wingless related MMTV integration site 8b	Wnt8b	5.4451303	up
maelstrom homolog (Drosophila)	Mael	5.2248325	up
membrane-spanning 4-domains, subfamily A, member 6B	Ms4a6b	5.21284	up
Serine (or cysteine) peptidase inhibitor, clade A member 7	Serpina7	5.0807915	up
RAB8A, member RAS oncogene family	Rab8a	4.9656444	up
gamma-aminobutyric acid (GABA-A) receptor, subunit beta 3	Gabrb3	4.656861	up
dentin matrix protein 1	Dmp1	4.632175	up
dynamin 1-like	Dnm1l	4.537594	up
zinc finger protein 286	Zfp286	4.5276074	up
gamma-aminobutyric acid (GABA-A) receptor, subunit gamma 2	Gabrg2	4.373157	up
ATP-binding cassette, sub-family G (WHITE), member 5	Abcg5	4.3656216	up
fatty acid binding protein 9, testis	Fabp9	4.363583	up
ankyrin repeat and SOCS box-containing protein 15	Asb15	4.285139	up
Phosphatidylinositol-4-phosphate 5-kinase, type II, beta	Pip5k2b	4.2436166	up
spectrin alpha 1 /// similar to spectrin alpha 1	Spna1	4.159283	up
Gene model 879, (NCBI)	Gm879	4.1485305	up
procollagen, type III, alpha 1	Col3a1	4.114069	up
immunoglobulin lambda chain, variable 1	Igl-V1	4.0002275	up
prolactin-like protein A	Prlpa	3.93325	up
zinc finger, SWIM domain containing 2	Zswim2	3.9162233	up
potassium channel tetramerisation domain containing 12b	Kctd12b	3.7762604	up
Pellino 1	Peli1	3.767381	up
Dopa decarboxylase	Ddc	3.71953	up
UDP galactosyltransferase 8A	Ugt8a	3.7045178	up
Prickle like 1 (Drosophila)	Prickle1	3.6549535	up
G protein-coupled receptor 22	Gpr22	3.4932404	up
sex comb on midleg-like 2 (Drosophila)	Scml2	3.4461977	up
matrix metallopeptidase 1b (interstitial collagenase)	Mmp1b	3.4065428	up
Calmodulin 3	Calm3	3.4056456	up
forkhead box D4	Foxd4	3.314626	up
F-box protein 36	Fbxo36	3.3036625	up

chondroitin sulfate proteoglycan 4	Cspg4	3.265387	up
brain expressed gene 1	Bex1	3.227367	up
zinc finger protein 703	Zfp703	3.0947416	up
p21 (CDKN1A)-activated kinase 7	Pak7	3.0701518	up
Acyloxyacyl hydrolase	Aoah	3.0436742	up
Superkiller viralicidic activity 2-like 2 (S. cerevisiae)	Skiv2l2	3.0077672	up
Forkhead box P1	Foxp1	2.9280138	up
vav 1 oncogene	Vav1	2.862455	up
troponin T3, skeletal, fast	Tnnt3	2.8335156	up
cadherin 9	Cdh9	2.8135777	up
ELK3, member of ETS oncogene family	Elk3	2.766346	up
Synaptotagmin I	Syt1	2.755157	up
SRY-box containing gene 12	Sox12	2.7187328	up
solute carrier family 1 (glial high affinity glutamate transporter), member 3	Slc1a3	2.6908576	up
arrestin domain containing 5	Arrdc5	2.6715667	up
sodium channel, voltage-gated, type VII, alpha	Scn7a	2.6620193	up
Olfactory receptor 1344	Olfr1344	2.6557763	up
Kinectin 1	Ktn1	2.5734076	up
Ubiquitin protein ligase E3 component n-recognin 1	Ubr1	2.5113926	up
retinol binding protein 7, cellular	Rbp7	2.506909	up
selectin, platelet	Selp	2.4285843	up
cDNA sequence BC015286	BC015286	2.4190724	up
hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 3	Hsd3b3	2.4092853	up
aminopeptidase puromycin sensitive	Npepps	2.3771298	up
chromogranin B	Chgb	2.3765488	up
SRY-box containing gene 8	Sox8	2.362174	up
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	Smarca4	2.2920327	up
agmatine ureohydrolase (agmatinase)	Agmat	2.2769406	up
tubulin tyrosine ligase-like family, member 10	Ttll10	2.2522202	up
Sex comb on midleg-like 4 (Drosophila)	Scml4	2.2474325	up
nuclear factor I/C	Nfic	2.232845	up
glypican 2 (cerebroglycan)	Gpc2	2.2157204	up
SID1 transmembrane family, member 1	Sidt1	2.2031376	up
ectodysplasin A2 isoform receptor	Eda2r	2.1057427	up
VGF nerve growth factor inducible	Vgf	2.0619547	up
Open reading frame 34	ORF34	2.0588498	up
trophinin	Tro	2.0180752	up
granzyme A	Gzma	2.0132134	up
nuclear factor I/C	Nfic	1.965658	up
myosin XVIIIb	Myo18b	1.9573606	up
glutathione S-transferase, alpha 1 (Ya)	Gsta1	1.944705	up
nephronectin	Npnt	1.8896224	up
NHL repeat containing 2	Nhlrc2	1.884267	up
glutathione S-transferase, alpha 2 (Yc2)	Gsta2	1.875826	up
transglutaminase 1, K polypeptide	Tgm1	1.8677815	up
dynein, axonemal, heavy chain 7c	Dnahc7c	1.8662553	up
WD repeat domain, phosphoinositide interacting 1	Wipi1	1.8652867	up
otopetrin 2	Otop2	1.8632149	up
zinc finger protein of the cerebellum 4	Zic4	1.8487298	up

FCH domain only 1	Fcho1	1.8462025	up
olfactory receptor 749	Olfr749	1.8401613	up
Ring finger protein 34	Rnf34	1.829579	up
solute carrier family 1 , member 1	Slc1a1	1.8284004	up
solute carrier family 2 (facilitated glucose transporter), member 4	Slc2a4	1.8278917	up
LPS-responsive beige-like anchor	Lrba	1.8207219	up
kinesin family member 21B	Kif21b	1.8101343	up
microtubule associated monoxygenase, LIM domain containing 2	Mical2	1.7998582	up
mitogen activated protein kinase kinase 7	Map2k7	1.7947936	up
laminin, beta 3	Lamb3	1.7906799	up
UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 4	B3gnt4	1.7729676	up
Zinc and ring finger 3	Znrf3	1.7695706	up
bromodomain containing 3	Brd3	1.76783	up
tescalcin	Tesc	1.7442336	up
enhancer of yellow 2 homolog (Drosophila)	Eny2	1.7368585	up
BTB and CNC homology 2	Bach2	1.7320616	up
Transcription factor 4	Tcf4	1.6999027	up
neurexin III	Nrxn3	1.6975553	up
growth arrest specific 2	Gas2	1.6974478	up
ubiquitin-conjugating enzyme E2B, RAD6 homology (S. cerevisiae)	Ube2b	1.6914178	up
dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2	Dyrk2	1.6899979	up
eukaryotic translation initiation factor 4, gamma 1	Eif4g1	1.6885562	up
syntrophin, gamma 1	Sntg1	1.6837147	up
storkhead box 2	Stox2	1.6727942	up
vomeronasal 1 receptor, C5	V1rc5	1.6645459	up
transformation related protein 63	Trp63	1.6626483	up
Ankyrin repeat and SAM domain containing 1	Anks1	1.6624368	up
Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 1	Cited1	1.6591775	up
FAST kinase domains 2	Fastkd2	1.654636	up
T-box 6	Tbx6	1.6506313	up
guanine nucleotide binding protein, alpha inhibiting 2	Gnai2	1.650558	up
kinase suppressor of ras 1	Ksr1	1.6434165	up
chemokine (C-C motif) ligand 28	Ccl28	1.6388469	up
retinoic acid receptor, gamma	Rarg	1.6295984	up
golgi transport 1 homolog A (S. cerevisiae)	Golt1a	1.6292521	up
fibroblast growth factor 10	Fgf10	1.6268995	up
CUB and Sushi multiple domains 1	Csmd1	1.6219556	up
Coiled-coil domain containing 130	Ccdc130	1.620044	up
zinc finger protein 169	Zfp169	1.6188608	up
Glioma tumor suppressor candidate region gene 1	Gltscr1	1.6186901	up
G protein-coupled receptor 125	Gpr125	1.6113076	up
fibroblast growth factor 12	Fgf12	1.6085553	up
ephrin A3	Efna3	1.6026549	up
tenascin XB	Tnxb	1.6008724	up
sperm acrosome associated 1	Spaca1	1.5992742	up
cell growth regulator with EF hand domain 1	Cgref1	1.5985755	up
claudin 1	Cldn1	1.5982128	up
presenilin 2	Psen2	1.5975857	up
disabled homolog 1 (Drosophila)	Dab1	1.5947363	up

procollagen, type XIV, alpha 1	Col14a1	1.5920473	up
Ring finger protein 166	Rnf166	1.5871948	Up
protease, serine, 29	Prss29	1.5867528	Up
general transcription factor III C 1	Gtf3c1	1.579731	Up
kelch-like ECH-associated protein 1	Keap1	1.5756584	Up
LIM and SH3 protein 1	Lasp1	1.5743945	Up
baculoviral IAP repeat-containing 4	Birc4	1.5727795	Up
predicted gene, EG627821	EG627821	1.5679657	Up
angiogenin, ribonuclease A family, member 3	Ang3	1.566327	Up
squamous cell carcinoma antigen recognized by T-cells 3	Sart3	1.5640463	Up
T-box 1	Tbx1	1.5620521	Up
Tnf receptor-associated factor 5	Traf5	1.5616097	Up
Glutamate receptor, ionotropic, kainate 2 (beta 2)	Grik2	1.5587282	Up
interleukin 3	Il3	1.5546032	Up
cytochrome c oxidase, subunit VIIIa	Cox8a	1.5543281	Up
testis expressed gene 22	Tex22	1.55334	Up
Single-stranded DNA binding protein 3	Ssbp3	1.550755	Up
Nuclear receptor coactivator 1	Ncoa1	1.5448709	Up
butyrophilin, subfamily 1, member A1	Btn1a1	1.541489	Up
Nuclear fragile X mental retardation protein interacting protein 2	Nufip2	1.5355133	Up
phosphatidylinositol glycan anchor biosynthesis, class Z	Pigz	1.5341334	Up
Kell blood group precursor (McLeod phenotype) homolog	Xk	1.5333047	Up
BTB (POZ) domain containing 14B	Btbd14b	1.5274377	Up
Paralemmin	Palm	1.525702	Up
predicted gene, EG666504	EG666504	1.5256082	Up
interleukin 6 receptor, alpha	Il6ra	1.5254347	Up
ATP-binding cassette, sub-family B (MDR/TAP), member 1B	Abcb1b	1.5233356	Up
sperm associated antigen 16	Spag16	1.5198305	Up
cytoglobin	Cygb	1.5189643	Up
diacylglycerol kinase, iota	Dgki	1.5146734	Up
zinc finger protein 455	Zfp455	1.5033585	Up

Annexure 2

List of primer pair used for real time PCR analysis

Gene	Primer Sequence 5'-> 3'
FrizzledF	TTTTCTCACTTGCCCTTTTC
FrizzledR	TGGATCTCAAATAAGCAGCA
CMYCF	CCACTCACCAGCACAACTAC
CMYCR	ACTGTTCTCGTCGTTTCCTC
DKK1F	CATTCTGGCCAACTCTTTCT
DKK1R	TGGAGTACAGGTGAAGCAAA
NFiCF	CAGACAGCCTCCACCTACTT
NFiCR	TTTGTCCTTTTCTCCTCCTG
TCF4F	CAACTTGAGGGACGACTTCT
TCF4R	CATCCATCTGTCCCTCCTAC
CYCLIND1F	GGCTGGTAGTATGAGGTGCT
CYCLIND1R	CCCAATGAAAGACCAATCTC
CYCLIND3F	GTAAAATCCACACACCAGCA
CYCLIND3R	GACCCCACAATCACAAGAA
CMYCBPF	TGCTGGACACGCTGACGA
CMYCBPR	AAGTGTGATGAGTGGATTTA
Wnt 8BF	GTTTCTTATGAAGCCCGTGT
Wnt 8BR	ATTGTTGTGCAGATTCATGG
Sox12F	AAGTCTGCTCCACTGGTCTC
Sox12R	TCACTCACATGACGTTTTCC

List of primer pair used for real time PCR analysis

Sox8F	ACCAAGCTGTGAAAGGAGAG
Sox8R	AAAGTTAGCATTGCCTGGTC
Foxp1F	CCCCTGAATCTGGTATCAAG
Foxp1R	TTATGACGCACTGCATTCTT
Foxd4F	CAGAGCACTTTGGCTGATTA
Foxd4R	TCCTCGTCATCCTCTTCTTC
Tfap2aF	CACTCCTACTGCTGCTGCT
Tfap2aR	AAGTTCAAGTGGGTGGTTG
bach2F	ATGTGAGCAGTCTTACGGA
bach2R	GTTTCCGTATTTCACATTCCA
prickle1F	GTAGTATTTCATAGACCTAAGTGT
prickle1R	CACAGAAAGAGCTCATCGC
rac1F	TTAACGAGCAATATGTCTGAC
rac1R	ATTCTTGCCAGTGAGTTAGG
sfrp1F	TCTCCTGAGTGTGTAGATCT
sfrp1R	GAAATCCTCACAGGTCCGC
trib2F	CAGAAAGAGTAAACGGGCAA
trib2R	AGGTACTGGGAACACAAATA

Bibliography

[1] Szymanski M; Barciszewska MZ; Zywicki M; Barciszewski J; *J. Appl. Genet.*, **2003**, *44*, 1–19.

[2] Mattick JS; *Embo Reports*, **2001**, *2*, 986–991.

[3] Doudna AJ; CechTR; *Nature*, **2002**, *418*, 222–228.

[4] Eddy SR; *Nat. Rev. Genet.*, **2001**, *2*, 919–929.

[5] Joyce GF; *Nature*, **2002**, *418*, 214–221.

[6] Morey C; Avner P; *FEBS Lett.*, **2004**, *567*, 27–34.

[7] Huttenhofer A; Schattner P; Polacek N; *Trends Genet.*, **2005**, *21*, 289–297.

[8] The ENCODE Project Consortium; *Nature*, **2007**, *447*, 799–816.

[9] Tissieres A; Cold Spring Harbour Laboratory Press, NY, 1974.

[10] Schmeing TM; Seila AC; Hansen JL; Freeborn B; Soukup JK; Scaringe SA; Strobel SA; Moore PB; Steitz TA; *Nat. Struct. Biol.*, **2002**, *9*, 225–230.

[11] Nagai K; Muto Y; Krummel DAP; Kambach C; Ignjatovic T; Walke S; Kuglstatter A; *Biochem. Soc. Trans.*, **2001**, *29*, 15–26.

[12] Jurica MS; Moore MJ; *Mol. Cell*, **2003**, *12*, 5–14.

[13] Branlant C; Krol A; Ebel JP; Lazar E; Haendler B; Jacob M; *EMBO J.*, **1982**, *1*, 1259–1265.

[14] Singh R; Reddy R; *Proc. Natl. Acad. Sci. USA*, **1989**, *86*, 8280–8283.

[15] Patel SB; Bellini M; *Nucl. Acid Res.*, **2008**, *36*, 6482–6493.

[16] Collins CA; Guthrie C; *Nat. Struct. Biol.*, **2000**, *7*, 850–854.

[17] Valadhkhan S; Manley JL; *Nature*, **2001**, *413*, 701–707.

[18] Frank DN; Pace NR; *Annu. Rev. Biochem.*, **1998**, *67*, 153–180.

[19] Morl M; Marchfielder A; *EMBO Rep.*, **2001**, *2*, 17–20.

[20] Marvini MC; Engelke DR; *RNA biology*, **2009**, *6*, 40–42.

[21] Coughlin DJ; Pleiss JA; Walker SC; Whitworth GB; Engelke DR; *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 12218–12223.

[22] Bachellerie JP; Cavaille J; Httenhofer A; *Biochimie.*, **2002**, *84*, 775–790.

[23] Kiss T; *J. Cell Sci.*, **2004**, *117*, 5949–5951.

[24] Cavaille J; Buiting K; Kiefmann M; Lalande M; Brannan CI; Horsthemke B; Bachellerie JP; Brosius J; Httenhofer A; *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 14311–14316.

[25] Greider CW; Blackburn EH; *Nature*, **1989**, *337*, 331–337.

[26] Lin J; Ly H; Hussain A; Abraham M; Pearl S; Tzfati Y; Parslow TG; Blackburn EH; *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 14713–14718.

[27] Wong JMY; Collins K; *Lancet*, **2003**, *362*, 983–988.

[28] Theimer CA; Feigon J; *Curr. Opin. Struct. Biol.*, **2006**, *16*, 307–318.

[29] Jacob F; Perrin D; Sanchez C; Monod J; *C R Hebd Seances Acad. Sci.*, **1960**, *250*, 1727–1729.

[30] Struhl K; *Nat. Struct. Mol. Biol.*, **2007**, *14*, 103–105.

[31] Dinger ME; Amaral PP; Mercer TR; Pang KC; Bruce SJ; Gardiner BB; Askarian-Amiri ME; Ru K; Solda G; Simons C *et al.; Genome Res.*, **2008**, *18*, 1433–1445.

[32] Mercer TR; Dinger ME; Sunkin SM; Mehler MF; Mattick JS; *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 716–721.

[33] Cawley S; Bekiranov S; Ng HH; Kapranov P; Sekinger EA; Kampa D; Piccolboni A; Sementchenko V; Cheng J; Williams AJ *et al.; Cell*, **2004**, *116*, 499–509.

[34] Ponjavic J; Ponting CP; Lunter G; *Genome Res.*, **2007**, *17*, 556–565.

[35] Guttman M; Amit I; Garber M; French C; Lin MF; Feldser D; Huarte M; Zuk O; Carey BW; Cassady JP *et al.; Nature*, **2009**, *458*, 223–227.

[36] Pang KC; Frith MC; Mattick JS; *Trends Genet.*, **2006**, *22*, 1–5.

[37] Mercer TR; Dinger ME; Mattick JS; *Nat. Rev. Genet.*, **2009**, *10*, 155–159.

[38] Pollard KS; Salama SR; Lambert N; Lambot M; Coppens S; Pedersen JS; Katzman S; King B; Onodera C; Siepel A *et al.; Nature*, **2006**, *443*, 167–172.

[39] Carroll SB; *Cell*, **2008**, *134*, 25–36.

[40] Hirota K; Miyoshi T; Kugou K; Hoffman CS; Shibata T; Ohta K; *Nature*, **2008**, *456*, 130–134.

[41] Ponting CP; Oliver PL; Reik W; *Cell*, **2009**, *136*, 629–641.

[42] Bernstein E; Allis CD; *Genes Dev.*, **2005**, *19*, 1635–1655.

[43] Prasanth KV; Spector DL; *Genes Dev.*, **2007**, *21*, 11–42.

[44] Costa FF; *Gene*, **2005**, *357*, 83–94.

[45] Lapidot M; Pilpel Y; *EMBO Rep.*, **2006**, *7*, 1216–1222.

[46] Lee RC; Feinbaum RL; Ambros V; *Cell*, **1993**, *75*, 843–854.

[47] Ghildiyal M; Phillip D; *Nat. Rev. Genet.*, **2009**, *10*, 94–106.

[48] Griffiths-Jones S; Saini HK; van Dongen S; Enright AJ; *Nucl. Acids Res.*, **2008**, *36*, 154–158.

[49] Ambros V; Lee RC; Lavanway A; Williams PT; Jewell D; *Curr. Biol.*, **2003**, *13*, 807–818.

[50] Lee Y; Ahn C; Han J; Choi H; Kim J; Yim J; Lee J; Provost P; Radmark O; Kim S; Kim VN; *Nature*, **2003**, *425*, 415–419.

[51] Lee Y; Kim M; Han J; Yeom K; Lee S; Baek SH; Kim VN; *EMBO J.*, **2004**, *23*, 4051–4060.

[52] Cai X; Hagedorn CH; Cullen BR; *RNA*, **2004**, *10*, 1957–1966.

[53] Lee Y; Ahn C; Han J; Choi H; Kim J; Yim J; Lee J; Provost P; Radmark O; Kim S; Kim VN; *Nature*, **2003**, *425*, 415–419.

[54] Denli AM; Tops BB; Plasterk RH; Ketting RF; Hannon GJ; *Nature*, **2004**, *432*, 231–235.

[55] Gregory RI; Yan K; Amuthan G; Chendrimada TP; Doratotaj B; Cooch N; Shiekhattar R; *Nature*, **2004**, *432*, 235–240.

[56] Bohnsack MT; Czaplinski K; Grlich D; *RNA*, **2004**, *10*, 185–191.

[57] Chendrimada TP; Gregory RI; Kumaraswamy E; Norman J; Cooch N; Nishikura K; Shiekhattar R; *Nature*, **2005**, *436*, 740–744.

[58] Okamura K; Hagen JW; Duan H; Tyler DM; Lai EC; *Cell*, **2007**, *130*, 89–100.

[59] Ruby JG; Jan CH; Bartel DP; **2007**, *448*, 83–86.

[60] Park W; Li J; Song R; Messing J; Chen X; *Curr. Biol.*, **2002**, *12*, 1484–1495.

[61] Reinhart BJ; Weinstein EG; Rhoades MW; Bartel B; Bartel DP; *Genes Dev.*, **2002**, *16*, 1616–1626.

[62] Meister G; Landthaler M; Patkaniowska A; Dorsett Y; Teng G; Tuschl T; *Mol. Cell*, **2004**, *15*, 185–197.

[63] Liu J; Carmell MA; Rivas FV; Marsden CG; Thomson JM; Song J; Hammond SM; Joshua-Tor L; Hannon GJ; *Science*, **2004**, *305*, 1437–1441.

[64] Yekta S; Shih IH; Bartel DP; *Science*, **2004**, *304*, 594–596.

[65] Brennecke J; Stark A; Russell RB; Cohen SM; *PLoS Biol.*, **2005**, *3*, E85.

[66] Lewis BP; Burge CB; Bartel DP; *Cell*, **2005**, *120*, 15–20.

[67] Kong YW; Cannell IG; de Moor CH; Hill K; Garside PG; Hamilton TL; Meijer HA; Dobbyn HC; Stoneley M; Spriggs KA *et al.; Proc. Natl Acad. Sci. USA*, **2008**, *105*, 8866–8871.

[68] Jinek M; Doudna JZ; *Nature*, **2009**, *457*, 405–412.

[69] Bernstein E; Caudy AA; Hammond SM; Hannon GJ; *Nature*, **2001**, *409*, 363–366.

[70] Zhang H; Kolb FA; Jaskiewicz L; Westhof E; Filipowicz W; *Cell*, **2004**, *118*, 57–68.

[71] Gan J; Tropea JE; Austin BP; Court DL; Waugh DS; Ji X; *Cell*, **2006**, *124*, 355–366.

[72] Yeom KH; Lee Y; Han J; Suh MR; Kim VN; *Nucl. Acids Res.*, **2006**, *34*, 4622–4629.

[73] Hutvagner G; Simard MJ; *Nat. Rev. Mol. Cell Biol.*, **2008**, *9*, 22–32.

[74] Song JJ; Smith SK; Hannon GJ; Joshua-Tor L; *Science*, **2004**, *305*, 1434–1437.

[75] Fire A; Xu S; Montgomery MK; Kostas SA; Driver SE; Mellow CC; *Nature*, **1998**, *391*, 806–811.

[76] Hamilton AJ; Baulcombe DC; *Science*, **1999**, *286*, 950–952.

[77] Bernstein E; Caudy AA; Hammond SM; Hannon GJ; *Nature*, **2001**, *409*, 363–366.

[78] Elbashir SM; Lendeckel W; Tuschl T; *Genes Dev.*, **2001**, *15*, 188–200.

[79] Zamore PD; Tuschl T; Sharp PA; Bartel DP; *Cell*, **2000**, *101*, 25–33.

[80] Hammond SM; Bernstein E; Beach D; Hannon GJ; *Nature*, **2000**, *404*, 293–296.

[81] Liu J; Rivas FV; Wohlschlegel J; Yates III JR; Parker R; Hannon GJ; *Nat. Cell Biol.*, **2005**, *7*, 1261–1266.

[82] Liu Q; Rand TA; Kalidas S; Du F; Kim H; Smith DP; Wang X; *Science*, **2003**, *301*, 1921–1925.

[83] Xie Z; Johansen LK; Gustafson AM; Kasschau KD; Lellis AD; Zilberman D; Jacobsen SE; Carrington JC; *PLoS Biol.*, **2004**, *2*, E104.

[84] Yoshikawa M; Peragine A; Park MY; Poethig RS; *Genes Dev.*, **2005**, *19*, 2164–2175.

[85] Borsani O; Zhu J; Verslues PE; Sunkar R; Zhu JK; *Cell*, **2005**, *123*, 1279–1291.

[86] Yang N; Kazazian HHL; *Nat. Struct. Mol. Biol.*, **2006**, *13*, 763–771.

[87] Okamura K; Chung W; Ruby JG; Guo H; Bartel DP; La EC; *Nature*, **2008**, *453*, 803–806.

[88] Tam OH; Aravin AA; Stein P; Girard A; Murchison EP; Cheloufi S; Hodges E; Anger M; Sachidanandam R; Schultz RM; Hannon GJ; *Nature*, **2008**, *453*, 534–538.

[89] Watanabe T; Totoki Y; Toyoda A; Kaneda M; Kuramochi-Miyagawa S; Obata Y; Chiba H; Kohara Y; Kono T; Nakano T *et al.;* *Nature*, **2008**, *453*, 539–543.

[90] Aravin AA; Naumova NM; Tulin AV; Vagin VV; Rozovsky YM; Gvozdev VA; *Curr. Biol.*, **2001**, *11*, 1017–1027.

[91] Girard A; Sachidanandam R; Hannon GJ; Carmell MA; *Nature*, **2006**, *442*, 199–202.

[92] Lau NC; Seto AG; Kim J; Kuramochi-Miyagawa S; Nakano T; Bartel DP; Kingston RE; *Science*, **2006**, *313*, 363–367.

[93] Aravin AA; Sachidanandam R; Bourchis D; Schaefer C; Pezic D; Toth KF; Bestor T; Hannon GJ; *Mol. Cell*, **2008**, *31*, 785–799.

[94] Grivna ST; Pyhtila B; Lin H; *Proc. Natl Acad. Sci. USA*, **2006**, *103*, 13415–13420.

[95] Brennecke J; Aravin AA; Stark A; Dus M; Kellis M; Sachidanandam R; Hannon GJ; *Cell*, **2007**, *128*, 1089–1103.

[96] Ilik I; Akhtar A; *RNA Biol.*, **2009**, *6*, 113–121.

[97] Smith ER; Pannuti A; Gu W; Steurnagel A; Cook RG; Allis CD; Lucchesi JC; *Mol. Cell Biol.*, **2000**, *20*, 312–318.

[98] Gilfillan GD; Dahlsveen IK; Becker PD; *FEBS Lett.*, **2004**, *567*, 8–14.

[99] Kanduri C; Whitehead J; Mohammad F; *FEBS Lett.*, **2009**, *583*, 857–864.

[100] Brockdorff N; Ashworth A; Kay GF; Cooper P; Smith S; McCabe VM; Norris DP; Penny GD; Patel D; Rastan S; *Nature*, **1991**, *351*, 329–331.

[101] Navarro P; Chambers I; Karwacki-Neisius V; Chureau C; Morey C; Rougeulle C; Avne P; *Science*, **2008**, *321*, 1693–1695.

[102] Chaumeil J; Le Baccon P; Wutz A; Heard E; *Genes Dev.*, **2006**, *20*, 2223–2237.

[103] Sado T; Hoki Y; Sasaki H; *Dev. Cell.*, **2005**, *9*, 159–165.

[104] Kohlmaier A; Savarese F; Lachner M; Martens J; Jenuwein T; Wutz A; *PLoS Biol.*, **2004**, *2*, E171.

[105] Wang J; Mager J; Chen Y; Schneider E; Cross JC; Nagy A; Magnuson T; *Nat. Genet.*, **2001**, *28*, 371–375.

[106] Zhao J; Sun BK; Erwin JA; Song JJ; Lee JT; *Science*, **2008**, *322*, 750–756.

[107] Sun BK; Deaton AM; Lee JT; *Mol. Cell*, **2006**, *21*, 617–628.

[108] Gilbert SL; Pehrson JR; Sharp PA; *J. Biol. Chem.*, **2000**, *275*, 36491–36494.

[109] Reik W; Walter J; *Nat. Rev. Genet.*, **2001**, *2*, 21–32.

[110] Gabory A; Ripoche MA; Yoshimizu T; Dandolo L; *Cytogenet. Genome Res.*, **2006**, *128*, 188–193.

[111] Constancia M; Dean W; Lopes S; Moore T; Kelsey G; Reik W; *Nat. Genet.*, **2000**, *26*, 203–6.

[112] Wevrick R; Francke U; *Hum. Mol. Genet.*, **1997**, *6*, 325–332.

[113] Sleutels F; Zwartand R; Barlow DP; *Nature*, **2002**, *415*, 68–73.

[114] Pandey RR; Mondal T; Mohammad F; Enroth S; Redrup L; Komorowski J; Nagano T; Mancini-Dinardo D; Kanduri C; *Mol. Cell*, **2008**, *32*, 232–246.

[115] Mohammad F; Mondal T; Kanduri C; *Epigenetics*, **2009**, *4*, 277–286.

[116] Umlauf D; Goto Y; Cao R; Cerqueira F; Wagschal A; Zhang Y; Feil R; *Nat. Genet.*, **2004**, *36*, 1296–1300.

[117] Wevrick R; Kerns JA; Francke U; *Hum. Mol. Genet.*, **1994**, *3*, 1877–1882.

[118] Rinn JL; Kertesz M; Wang J; Brunghman S; Goodnough H; Wong B; Cleary M; Helms JA; Segal E; Chang HY; *Cell*, **2007**, *129*, 1311–1323.

[119] Elgin SC; Grewal IS; *CurR. Biol.*, **2003**, *13*, R895– R898.

[120] Matzke MA; Mette M; Matzke AJM; *Plant Mol. Biol.*, **2000**, *43*, 402–415.

[121] Sijen T; Plasterk RH; *Nature*, **2003**, *426*, 310–314.

[122] Motamedi MR; Verdel A; Colmenares SU; Gerber SA; Gygi SP; Moazed D; *Cell*, **2004**, *119*, 789–802.

[123] Mochizuki K; Fine NA; Fujisawa T; Gorovsky MA; *Cell*, **2002**, *110*, 689–699.

[124] Mochizuki K; Gorovsky MA; *Curr. Opin. Genet. Dev.*, **2004**, *14*, 1–7.

[125] Pal-Bhadra M; Leibovitch BA; Gandhi SG; Rao M; Bhadra U; Birchler JA; Elgin SC; *Science*, **2004**, *303*, 669–672.

[126] Wang X; Arai S; Song X; Reocjart D; Du K; Pascual G; Tempst P; Rosenfeld MG; Glass CK; Kurokawa R; *Nature*, **2008**, *454*, 126–130.

[127] Guenther MG; Levine SS; Boyer LA; Jaenisch R; Young RA; *Cell*, **2007**, *130*, 77–88.

[128] Feng J; Bi C; Clark BS; Mady R; Shah P; Kohtz JD; *Genes Dev.*, **2006**, *20*, 1470–1484.

[129] Martianov I; Ramadass A; Barros AS; Chow N; Akoulitchev A; *Nature*, **2007**, *445*, 666–670.

[130] Ohno M; Fukagawa T; Lee JS; Ikemura T; *Chromosoma*, **2002**, *111*, 201–213.

[131] Mariner PD; Walters RD; Espinoza CA; Drullinger LF; Wagner SD; Kugel JF; Goodrich JA; *Mol. Cell*, **2008**, *29*, 499–509.

[132] Beltran M; Piug I; Pena C; Garcia JM; Alvares AV; Pena R; Bonilla F; de Herreros AG; *Genes Dev.*, **2008**, *22*, 756–769.

[133] Willingham AT; Orth PA; Batalov S; Peters EC; Wen BG; Aza-Blanc P; Hogenesch JB; Schultz PG; *Science*, **2005**, *309*, 1570–1573.

[134] Lakhotia SC; Kluwer Acdemic/Plenum Publishers, Georgetown, 1974.

[135] Prasanth KV; Rajendra TK; Lal AK; Lakhotia SC; *J. Cell Sci.*, **2000**, *113*, 3485–3497.

[136] Bendena WG; Ayme-Southgate A; Garbe JC; Pardue ML; *Dev. Biol.*, **1991**, *144*, 65–77.

[137] Jolly C; Metz A; Govin J; Vigneron M; Turner BM; Khochbin S; Vourch C; *J. Cell Biol.*, **4200**, *164*, 25–33.

[138] Jolly C; Lakhotia SC; *Nucleic Acids Res.*, **2006**, *34*, 5508–5514.

[139] Hutchinson JN; Ensminger AW; Clemson CM; Lynch CR; Lawrence JB; Chess AA; *BMC Genomics*, **2007**, *8*, 39–49.

[140] Clemson CMZ; Hutchinson JN; Sara SA; Ensminger AW; Fox AH; Chess A; Lawrence JB; *Mol. Cell*, **2009**, *33*, 717–726.

[141] Prasanth KV; Prasanth SG; Xuan Z; Hearn S; Freier SM; Bennett CF; Zhang MQ; Spector DL; *Cell*, **2005**, *123*, 249–263.

[142] Iacoangeli A; Rozhdestvensky TS; Dolzhanskaya N; Tournier B; Schtt J; Brosius J; Denman RB; Khandjian EW; Kindler S; Tiedge H; *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 734–739.

[143] Skryabin BV; Sukonina V; Jordan U; Lewejohann L; Sachser N; Muslimov I; Tiedge H; Brosius J; *Mol. Cell Biol.*, **2003**, *23*, 6435–6441.

[144] Bagni C; *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, E19.

[145] Muslimov IA; Iacoangeli A; Brosius J; Tiedge H; *J. Cell Biol.*, **2006**, *175*, 427–439.

[146] Cao X; Yeo G; Muotri AR; Kuwabara T; Gage FH; *Annu. Rev. Neurosci.*, **2006**, *29*, 77–103.

[147] Luo JH; Ren B; Keryanov S; Tseng GC; Rao UN; Monga SP; Strom S; Demetris AJ; Nalesnik J; Yu YP; Ranganathan S; Michalopoulos GK; *Hepatology*, **2006**, *44*, 1012–1024.

[148] Ji P; Diederichs S; Wang W; Boing S; Metzger R; Schneider PM; Tidow N; Brandt B; Buerger H; Bulk E *et al.; Oncogene*, **2003**, *22*, 6087–6097.

[149] Negrini M; Nicoloso MS; Calin GA; *Curr. Opin. Cell Biol.*, **2009**, *3*, 470–479.

[150] Nelson KM; Weiss GJ; *Mol. Cancer Ther.*, **2008**, *7*, 3655–3660.

[151] Nicoloso MS; Kipps TJ; Croce CM; Calin GA; *Br. J. Haematol.*, **2007**, *139*, 709–716.

[152] Esquela-Kerscher A; Slack FJ; *Nat. Rev. Cancer*, **2006**, *4*, 259–269.

[153] Novakova J; Slaby O; Vyzula R; Michalek J; *Biochem. Biophys. Res. Commun.*, **2009**, *386*, 1–5.

[154] Lunde BM; Moore C; Varani G; *Nat. Rev. Mol.Cell Biol.*, **2007**, *8*, 479–490.

[155] Glisovic T; Bachorik JL; Yong J; Dreyfuss G; *FEBS Lett.*, **2008**, *14*, 1977–1986.

[156] Chen Y; Varani G; *FEBS J.*, **2005**, *272*, 2088–2097.

[157] Keene JD; *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 7018–7024.

[158] Lasko P; *J. Cell Biol.*, **2000**, *150*, F51–F56.

[159] Lee MH; Schedl T; *WormBook*, **2006**, pages 1–13.

[160] Linder P; Lasko PF; Leroy P; Nielsen PJ; Nishi K; Schnier J; Slominski PP; *Nature*, **1989**, *337*, 121–122.

[161] Jankowsky E; Gross CH; Shuman S; Pyle AM; *Science*, **2001**, *291*, 121–125.

[162] Schwer B; *Nat. Struct. Biol.*, **2001**, *8*, 113–116.

[163] Fairman ME; Maroney PA; Wang W; Bowers HA; Gollnick P; Nilsen TW; Jankowsky E; *Science*, **2004**, *304*, 730–734.

[164] Fuller-Pace FV; *Nucl. Acids Res.*, **2006**, *34*, 4206–4215.

[165] Lane DP; Hoeffler WK; *Nature*, **1980**, *288*, 167–170.

[166] Hirling H; Scheffner M; Restle T; Stahl H; *Nature*, **1989**, *288*, 562–564.

[167] Lamm GM; Nicol SM; Fuller-Pace FV; Lamond AI; *Nucl. Acids Res.*, **1996**, *24*, 3739–3747.

[168] Ogilvie VC; Wilson BJ; Nicol SM; Morrice NA; Saunders LR; Barber GN; Fuller-Pace FV; *Nucl. Acids Res.*, **2003**, *31*, 1470–1480.

[169] Stevenson RJ; Hamilton SJ; MacCallum DE; Hall PA; Fuller-Pace FV; *J. Pathol.*, **1998**, *184*, 351–359.

[170] Fuller-Pace FV; Ali S; *Biochem. Soc. Trans*, **2008**, *36*, 609–612.

[171] Kahlina K; Goren I; Pfeilschifter J; Frank S; *J. Biol. Chem.*, **2004**, *279*, 44872–44882.

[172] Liu ZR; *Mol. Cell Biol.*, **2002**, *22*, 5443–5450.

[173] Lee CG; *J. Biol. Chem.*, **2002**, *277*, 39679–39683.

[174] Jalal C; Uhlmann-Schiffler H; Stahl H; *Nucl. Acids Res.*, **2007**, *35*, 3590–3601.

[175] Fukuda T; Yamagata K; Fujiyama S; Matsumoto T; Koshida I; Yoshimura K; Mihara M; Naitou M; Endoh H; Nakamura T *et al.;* *Nat. Cell Biol.*, **2007**, *9*, 604–611.

[176] Buszczak M; Spradling AC; *Genes Dev.*, **2006**, *20*, 977–989.

[177] Endoh H; Maruyama K; Masuhiro Y; Kobayashi Y; Goto M; Tai H; Yanagisawa J; Metzger D; Hashimoto S; Kato S; *Mol. Cell. Biol.*, **1999**, *19*, 5363–5372.

[178] Caretti G; Lei EP; Sartorelli V; *Cell Cycle*, **2007**, *6*, 1172–1176.

[179] Honig A; Berget SM; O'Malley BW Auboeuf D;; *Science*, **2002**, *298*, 416–419.

[180] Metivier R; Penot G; Hubner MR; Reid G; Brand H; Kos M; Gannon F; *Cell*, **2003**, *115*, 751–763.

[181] Bates GJ; Nicol SM; Wilson BJ; Jacobs AM; Bourdon JC; Wardrop J; Gregory DJ; Lane DP; Perkins ND; Fuller-Pace FV; *EMBO J.*, **2005**, *24*, 543–553.

[182] Caretti G; Schiltz RL; Dilworth FJ; Di Padova M; Zhao P; Ogryzko V; Fuller-Pace FV; Hoffman EP; Tapscott SJ; Sartorelli V; *Dev. Cell*, **2006**, *11*, 547–560.

[183] ED; Niu L; Caretti G; Nicol SM; Teplyuk N; Stein GS; Sartorelli V; van Wijnen AJ; Fuller-Pace FV; Westendorf JJ Jensen; *J. Cell. Biochem.*, **2008**, *103*, 1438–1451.

[184] BJ; Bates GJ; Nicol SM; Gregory DJ; Perkins ND; Fuller-Pace FV Wilson.

[185] Yang L; Lin C; Liu ZR; *Cell*, **2006**, *127*, 139–155.

[186] Cadigan KM; Nusse R; *Genes Dev.*, **1997**, *11*, 3286–3305.

[187] Reya T; Clevers H; *Nature*, **2005**, *434*, 843–850.

[188] Nusse R; *Cell Res.*, **2008**, *18*, 523–527.

[189] Logan CY; Nusse R; *Annu. Rev. Cell Dev. Biol.*, **2004**, *20*, 781–810.

[190] Uthoff SM; Eichenberger MR; McAuliffe TL; Hamilton CJ; Galandiuk S; *Mol. Carcin.*, **2001**, *31*, 56–62.

[191] Hsieh JC; *Nature*, **1999**, *398*, 431–436.

[192] Krupnik VE; Sharp JD; Jiang C; Robison K; Chickering TW; Amaravadi L; Brown DE; Guyot D; Mays G; Leiby K; emphet al.; *Gene*, **1999**, *238*, 301–313.

[193] Mao B; Wu W; Li Y; Hoppe D; Stannek P; Glinka A; Niehrs C; *Nature*, **2001**, *411*, 321–325.

[194] Mao B; Wu W; Davidson G; Marhold J; Li M; Mechler BM; Delius H; Hoppe D; Stannek P; Walter C *et al.; Nature*, **2002**, *417*, 664–667.

[195] Nusse R; Varmus HE; *Cell*, **1982**, *31*, 99–109.

[196] Nusse R; Brown A; Papkoff J; Scambler P; Shackleford G; McMahon A; Moon R; Varmus H; *Cell*, **1991**, *64*, 231–237.

[197] Dimitriadis A; Vincan E; Mohammed IM; Roczo N; Phillips WA; Baindur-Hudson S; *Cancer Lett.*, **2001**, *166*, 185–191.

[198] Bhanot P; Brink M; Samos CH; Hsieh JC; Wang Y; Macke JP; Andrew D; Nathans J; Nusse R; *Nature*, **1996**, *382*, 225–230.

[199] Wang Y; Macke JP; Abella BS; Andreasson K; Worley P; Gilbert DJ; Copeland NG; Jenkins NA; Nathans J; *J. Biol. Chem.*, **1996**, *271*, 4468–4476.

[200] Pinson KI; Brennan J; Monkley S; Avery BJ; Skarnes WC; *Nature*, **2000**, *407*, 535–538.

[201] Tamai K; Semenov M; Kato Y; Spokony R; Liu C; Katsuyama Y; Hess F; Saint-Jeannet JP; He X; *Nature*, **2000**, *407*, 530–535.

[202] Wehrli M; Dougan ST; Caldwell K; O'Keefe L; Schwartz S; Vaizel-Ohayon D; Schejter E; Tomlinson A; DiNardo S; *Nature*, **2000**, *407*, 535–538.

[203] Brown SD; Twells RC; Hey PJ; Cox RD; Levy ER; Soderman AR; Metzker ML; Caskey CT; Todd JA; Hess JF; *Biochem. Biophys. Res. Comm.*, **1998**, *248*, 879–888.

[204] Klingensmith J; Nusse R; Perrimon N; *Genes Dev.*, **1994**, *8*, 118–130.

[205] Pizzuti A; Amati F; Calabrese G; Mari A; Colosimo A; Silani V; Giardino L; Ratti A; Penso D; Calzy L *et al.; Mol. Genet.*, **1996**, *5*, 953–958.

[206] Bui TD; Beier DR; Jonssen M; Smith K; Dorrington SM; Kaklamanis L; Kearney L; Regan R; Sussman DJ; Harris AL; *Biochem. Biophys. Res. Comm.*, **1997**, *239*, 510–516.

[207] Kishida S; Yamamoto H; Hino S; Ikeda S; Kishida M; Kikuchi A; *Mol. Cell Biol.*, **1999**, *19*, 4414–4422.

[208] Sakanaka C; Leong P; Xu L; Harrison SD; Williams LT; *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 1066–1076.

[209] Amit S; Hatzubai A; Birman Y; Andersen JS; Ben-Shushan E; Mann M; Ben-Neriah Y; Alkalay I; *Genes Dev.*, **2002**, *16*, 1066–1076.

[210] Fraser E; Young N; Dajani R; Franca-Koh J; Ryves J; Williams RS; Yeo M; Webster MT; Richardson C; Smalley MJ *et al.; J. Biol. Chem.*, **2002**, *277*, 2176–2185.

[211] Franca-Koh J; Yeo M; Fraser E; Young N; Dale TC; *J. Biol. Chem.*, **2002**, *277*, 2176–2185.

[212] Schwarz-Romond T; Asbrand C; Bakkers J; Kohl M; Schaeffer HJ; Huelsken J; Behrens J; Hammerschmidt M; Birchmeier W; *Genes Dev.*, **2002**, *16*, 2073–2081.

[213] Kemler R; Ozawa M; *BioEssays*, **1989**, *11*, 88–91.

[214] McCrea PD; Turck CW; Gumbiner B; *Science*, **1991**, *254*, 1359–1361.

[215] Willert K; Nusse R; *Curr. Opin. Genet. Dev.*, **1998**, *8*, 95–102.

[216] Huber AH; Weis WI; *Cell*, **2004**, *303*, 1483–1489.

[217] Polakis P; *Cell*, **2001**, *105*, 563–566.

[218] Liu J; Stevens J; Rote CA; Yost HJ; Hu Y; Neufeld KL; White RL; Matsunami N; *Mol. Cell*, **2001**, *7*, 927–936.

[219] Kramps T; Peter O; Brunner E; Nellen D; Froesch B; Chatterjee S; Murone M; Zollig S; Basler K; *Cell*, **2002**, *109*, 47–60.

[220] Tago K; Nakamura T; Nishita M; Hyodo J; Nagai S; Murata Y; Adachi S; Ohwada S; Morishita Y; Shibuya H *et al.; Genes Dev.*, **2000**, *14*, 1741–1749.

[221] Aoki K; Taketo MM; *J. Cell Sci.*, **2007**, *120*, 3327–3335.

[222] Behrens J; Jerchow BA; Wortele M; Grimm J; Asbrand C; Wirtz R; Kohl M; Wedlich D; Birchmeier W; *Science*, **1998**, *280*, 596–599.

[223] Lustig B; Jerchow B; Sachs M; Weiler S; Pietsch T; Karsten U; van de Wetering M; Clevers H; Schlag PM; Birchmeier W *et al.; J. Mol. Cell. Biol.*, **2002**, *22*, 1184–1193.

[224] Manoukian AS; Woodgett JR; *Adv. Cancer Res.*, **2002**, *84*, 203–229.

[225] Frame S; Cohen P; *Biochem. J.*, **2001**, *359*, 1–16.

[226] He X; *Cell*, **2006**, *127*, 40–42.

[227] Mosimann C; Hausmann G; Basler K; *Nat. Rev. Mol. Cell Biol.*, **2009**, *10*, 276–286.

[228] Katoh M; *Oncol. Rep.*, **2005**, *14*, 1583–1588.

[229] Nishant KT; Ravishankar H; Rao MRS; *Mol. Cell Biol.*, **2004**, *24*, 5620–34.

[230] Birnboim HC; Dolly J; *Nucl. Acids Res.*, **1979**, *7*, 1513–1523.

[231] Chung CT; Niemela SL; Miller RH; *Proc. Natl. Acad. Sci USA*, **1979**, *86*, 2172–2175.

[232] Lagos-Quintana M; Rauhut R; Yalcin A; Meyer J; Lendeckel W; Tuschl T; *Curr. Biol.*, **2002**, *12*, 735–739.

[233] Maroney PA; Chamnongpol S; Souret F; Nilsen TW; *RNA*, **2007**, *13*, 930–936.

[234] Han J; Lee Y; Yeom KH; Nam JW; Kim YK; Jin H; Kim VN; *Genes Dev.*, **2004**, *18*, 3016–3027.

[235] Srisawat C; Engelke DR; *RNA*, **2001**, *7*, 632–641.

[236] Ramachandra L; Rao MRS; *Chromosoma*, **1994**, *7*, 486–501.

[237] Kim VN; *Mol. Cell*, **2005**, *19*, 1–15.

[238] Sohn SY; Cho Y; Zhang BT; Kim VN Han J; Lee Y; Yeom KH; Nam JW; Heo I; Rhee JK;; *Cell*, **2006**, *125*, 887–901.

[239] Elmqn J; Thonberg H; Ljungberg K; Frieden M; Westergaard M; Xu Y; Wahren B; Liang Z; Prum H; Koch T; Wahlestedt C; *Nucl. Acids Res.*, **2005**, *33*, 439–447.

[240] Colley SM; Iyer KR; Leedman PJ; *IUBMB Life*, **2008**, *60*, 159–164.

[241] Ambros V; Lee RC; Lavanway A; Williams PT; Jewell D; *Curr. Biol.*, **2003**, *13*, 807–818.

[242] He L; Thomson JM; Hemann MT; Hernando-Monge E; Mu D; Goodson S; Powers S; Cordon-Cardo C; Lowe SW; Hannon GJ; Hammond SM; *Nature*, **2005**, *435*, 828–833.

[243] Heard E; *Curr. Opin. Cell Biol.*, **2004**, *16*, 247–255.

[244] Pagano A; Castelnuovo M; Tortelli F; Ferrari R; Dieci G; *PLoS Genetics*, **2007**, *3*, 174–184.

[245] Wilusz JE; Freier SM; Spector DL; *Cell*, **2008**, *135*, 919–932.

[246] Iwasaki S; Kawamata T; Tomari Y; *Mol. Cell.*, **2009**, *34*, 58–67.

[247] Farazi TA; Juranek SA; Tuschl T; *Development*, **2008**, *135*, 1201–1214.

[248] Su H; Trombly MI; Chen J; Wang X; *Genes Dev.*, **2009**, *23*, 303–317.

[249] Nesterova TB; Mermoud JE; Hilton K; Pehrson J; Surani MA; McLaren A; Brockdorff N; *Differentiation*, **2002**, *69*, 216–225.

[250] Rasmussen TP; Mastrangelo MA; Eden A; Pehrson JR; Jaenisch R; *J. Cell Biol.*, **2000**, *150*, 1189–1198.

[251] Salzman DW; Shubert-Coleman J; Furneaux H; *J. Biol. Chem.*, **2007**, *282*, 32771–32779.

[252] Berezhna SY; Supekova L; Supek F; Schultz PG; Deniz AA; *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 7682–7687.

[253] Robb GB; Brown KM; Khurana J; Rana TM; *Nat. Struct. Mol. Biol.*, **2005**, *12*, 133–137.

[254] Rubinfeld B; Robbins P; El-Gamil M; Albert I; Porfiri E; Polakis P; *Science*, **1997**, *275*, 1790–1792.

[255] Aberle H; Schwartz H; Kemler R; *J. Cell Biochem.*, **1996**, *61*, 514–523.

[256] Clark EL; Coulson A; Dalgliesh C; Rajan P; Nicol SM; Fleming S; Heer R; Gaughan L; Leung HY; Elliott DJ *et al.; Cancer Res.*, **2008**, *68*, 7938–7946.

[257] Watanabe M; Yanagisawa J; Kitagawa H; Takeyama K; Ogawa S; Arao Y; Suzawa M; Kobayashi Y; Yano T; Yoshikawa H *et al.; EMBO*, **2001**, *20*, 1341–1352.

[258] Walker SC; Scott FH; Srisawat C; Engelke DR; *Methods Mol. Biol.*, **2008**, *488*, 23–40.

[259] Qureshi IA; Mehler MF; *Neurosci. Lett.*, **2009**, *In press*.

[260] Hannenhalli Sl Kaestner KH; *Nat. Rev. Genet.*, **2009**, *10*, 233–240.

[261] Kormish JD; Sinner D; Zorn AM; *Dev. Dyn.*, **2009**.

[262] Igarashi K; Ochiai K; Muto A; *J. Biochem.*, **2007**, *141*, 783–789.

[263] Li Q; Dashwood RH; *J. Biol. Chem.*, **2004**, *279*, 45669–45675.

[264] Tissir F; Bar I; Jossin Y; De Backer O; Goffinet AM; *Nat. Neurosci.*, **2005**, *8*, 451–457.

[265] Stucke VM; Gorses D; Hofmann F; *Cell Cycle*, **2008**, *7*, 830–832.

[266] Yuan Y; Li DM; Sun H; *J. Biol. Chem.*, **1998**, *273*, 20347–20353.

List of Publications

Gayatri G and Rao MRS, "A Novel non-coding RNA processed by Drosha is restriced to nucleus in mouse." *RNA*, **14**, 1399-1410 (2008).

Gayatri G and Rao MRS, "*Mrhl* non-coding RNA regulates wnt signalling through its interaction with p68, a DEAD box helicase." *Manuscript in communication.*

Keerthi TC, **Gayatri G** and Rao MRS, "Identification of a novel nucleolin related protein (NRP) gene expressed during rat spermatogenesis" *BMC Mol. Biol.*, **10**, 64 (2009).