Studies on LncRNA Mrhl in Mouse Embryonic Stem Cells and Neuronal Lineage Development

A Thesis Submitted for the Degree of

Doctor of Philosophy

By

Debosree Pal

Molecular Biology and Genetics Unit

Jawaharlal Nehru Centre for Advanced Scientific Research

Bangalore – 560064, India

January 2019

"Science, for me, gives a partial explanation for life. In so far as it goes, it is based on fact, experience and experiment.....In my view all that is necessary for faith is the belief that by doing our best we shall come nearer to success and that success in our aims (the improvement of the lot of mankind, present and future) is worth attaining."

- Rosalind Franklin

Dedicated to my family

"The true laboratory is the mind, where behind illusions we uncover the laws of truth"

- Sir Jagadish Chandra Bose

DECLARATION

I hereby declare that the thesis entitled '**Studies on lncRNA Mrhl in mouse embryonic stem cells and neuronal lineage development**' is an authentic record of research work carried out by me under the supervision of Prof. M.R.S Rao at Chromatin Biology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India. In keeping with the norm of reporting scientific observations, due acknowledgement has been made whenever work described here has been based on the findings of other investigators. Any omission owing to oversight or misjudgement is regretted.

Debosree Pal

Bangalore

18.01.2019

CERTIFICATE

I hereby certify that the thesis entitled '**Studies on lncRNA Mrhl in mouse embryonic stem cells and neuronal lineage development**' has been carried out by Ms. Debosree Pal under my supervision at Chromatin Biology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India. This work has not been submitted for the award of any degree or diploma to any institution or university previously.

Prof. M.R.S Rao

Bangalore

18.01.2019

Acknowledgements

My journey in JNCASR towards pursuing my doctoral studies has been one of the most significant and impactful experiences for me. I would like to take the opportunity herewith to express my gratitude to those who have guided me, helped me and believed in me throughout these years.

I am sincerely grateful to my supervisor Prof. M.R.S Rao for providing me with the opportunity to pursue a challenging and exciting project under his guidance. He has always motivated me to explore and pursue my scientific interests and ideas, been patient and supportive during my difficult times, and constantly encouraged me to develop myself as an independent researcher. He has been a great source of inspiration for me and I am deeply thankful to him for imparting his invaluable knowledge and for being a mentor to me.

I would like to extend my sincere thanks to all esteemed faculty of MBGU and NSU. I would like to thank Prof. Anuranjan Anand, Prof. Udaykumar Ranga, Prof. Maneesha Inamdar, Prof. Tapas K Kundu, Prof. Hemalatha Balaram, Prof. Namita Surolia, Prof. Kaustav Sanyal, Prof. Ravi Manjithaya, Prof. James Chelliah and Prof. Sheeba Vasu for their coursework, their suggestions and constructive criticisms during the annual work presentations and for their guidance and help whenever it has been required.

The work herewith would not have been completed without the excellent central instrumentation facilities of JNCASR. In this regard, I would like to acknowledge the help and support of Ms. Suma B.S of Confocal facility, Dr. R.G. Prakash of Animal facility, Dr. Uttara Chakraborty, Swarupa Yalla and Dr. Narendra Nala of Flow cytometry facility and Ms. Anitha of the Sequencing facility.

I would like to express my sincere appreciation towards all the past and present members of the Chromatin Biology Laboratory for their constant support and help. In this respect, I would like to thank Dr. LN Misra, Dr. Swati, Dr. Nikhil, Dr.Vijay, Dr. Satyakrishna, Dr. Vasudev, Monalisa, Prathima, Anayat, Dr. Roshan, Dr. Shalini, Aditya, Meenakshi, Dhanur, Shubhangini, Bhavana, Raktim, Dr. Arun, Dr. Subendhu, Neha, Utsa, Zenia, Prof. Subbulaksmi, Dr. Sangeeta, Prof. Ramakrishna and Usha. A special thanks to Dhanur, Neha, Utsa and Zenia for their contributions towards this work. I also thank Nishitha, Sharath, Sakthi Veena, and Muniraju for being an integral part of the lab.

My heartfelt thanks to JNCASR administration, academic section, security staff, library, complab, dining hall canteen, utility canteen, hostel staff and Dhanvantri for their support. I would also like to thank past and present members of various student committees for their efforts to make life on campus enjoyable.

I would like to acknowledge a few people in my life who have been there for me through times, good and especially otherwise. I thank Aakash, Ananya, Anindita, Moumita, Meghna, Indira, Bodhisatwa, Rosemary, Semanti and Rohini for being there for me through all these years. I thank my childhood friends Chumki and Amrita for staying by my side. I thank Papri, Sunita, Anindita and Mouli for their words of encouragement and constant support. I am thankful to Janet aunty for her support and Vinay and Akshay for being my friends and for helping me to stay focused during tough times. I am thankful to Arnab, Divyesh, Sutanuka, Sweta, Ronak and Jaspar for going that extra mile to help me and support me whenever needed. I also thank all my colleagues in MBGU and NSU who have helped me in some way or the other in taking a step closer towards the completion of my work.

I would like to extend my sincere thanks to my teachers from school, to all my professors from St. Xavier's college, especially Prof. Uma Sidhhanta, Prof. Dipankar Chakravorty and Prof. Ronita Nag Chaudhuri and to Prof. Nitai P. Bhattacharyya from Saha Institute of Nuclear Physics who have inspired me and have helped me realize my academic interests.

I thank UGC, DBT and JNCASR for funding.

Finally, I express my deepest gratitude to my family. Without them I would not have been able to complete this journey. Their relentless and unwavering support has helped me stay put during times of adversity. They have been my pillars of strength. They have taught me to persevere and always encouraged me to overcome hardships and to keep moving forward. They have believed in me and helped me believe in myself. I am truly grateful to them for everything.

Abbreviations

Airn: Antisense IGF2R RNA

Anril: Antisense non-coding RNA in the Ink4 locus

Bace1-as: β-site amyloid precursor protein cleaving enzyme 1 antisense RNA

BP: Basal progenitor

ChIP: Chromatin IP

ChOP: Chromatin oligoaffinity purification

CRISPR: Clustered regularly interspaced short palindromic repeats

Dali: DNMT1 associated long intergenic non-coding RNA

DAPT: N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester

DEG: Differentially expressed genes

DEPC: Diethyl pyrocarbonate

DMEM: Dulbecco's modified eagle's medium

EB: Embryoid body

EDTA: Ethylenediaminetetracetic acid

EMSA: Electrophoretic mobility shift assay

ENCODE: Encyclopedia of DNA elements

eRNA: Enhancer RNA

ES: Embryonic stem

FANTOM: Functional annotation of the mammalian genome

Fendrr: FOXF1 adjacent non-coding developmental regulatory RNA

FIMO: Find individual motif occurrence

FISH: Fluorescent in situ hybridization

FPKM: Fragments per kilobase of transcript per million mapped reads

Gas5: Growth arrest specific 5

GENCODE: Encyclopedia of genes and gene variants

GO: Gene ontology

Hotair: HOX antisense intergenic RNA

ICM: Inner cell mass

IP: Intermediate progenitor

iPSC: Induced pluripotent stem cells

Jade: Jade1 adjacent regulatory RNA

Kcnq1ot1: KCNQ1 opposite strand/antisense transcript

LIF: Leukemia inhibitory factor

Linc-MD1: Long intergenic non-protein coding RNA, muscle differentiation 1

LincRNA: Long intergenic non-coding RNA

Linc-RoR: Long intergenic non-coding RNA, regulator of reprogramming

LncRNA: Long non-coding RNA

Malat-1: Metastasis associated lung adenocarcinoma transcript 1

Meg3: Maternally expressed 3

mESC: Mouse embryonic stem cell

miRNA: MicroRNA

Mrhl: Meiotic recombination hotspot locus RNA

mRNA: Messenger RNA

Neat-1: Nuclear enriched abundant transcript 1

NEC: Neuroepithelial cell

NICD: Notch intracellular domain

Paupar: PAX6 upstream antisense RNA

PBS: Phosphate buffered saline

qRT-PCR: Real time quantitative PCR

RA: Retinoic acid

RAR: retinoic acid receptor

RepA: Repeat A

RGC: Radial glia cell

RNA Pol II: RNA Polymerase II

RNAi: RNA interference

RNA-Seq: RNA Sequencing

SDS: Sodium dodecyl sulfate

shRNA: Short hairpin RNA

siRNA: Short-interfering RNA

snoRNAs: Small nucleolar RNA

snRNA: Small nuclear RNA

SVZ: Subventricular zone

TE: Trophectoderm

TF: Transcription factor

tRNA: Transfer RNA

TSS: Transcription start site

Tsx: Testis-specific X-linked gene

Tuna: TCL1 upstream neuron associated lincRNA

VZ: Ventricular zone

WB: Western blotting

X_a: Active X-chromosome

X_i: Inactive X-chromosome

Xist: X-inactive specific transcript

List of Figures

List of Tables

Table of Contents

Chapter 1

Introduction

1.1 Non Coding RNAs

The discovery of 'nuclein' by Fredrich Meischer in the 19th century [1] led the ground work for exploring the role of RNAs in living organisms. Initially RNAs were thought to be a link for information flow between the genetic material, DNA and the catalytic components, proteins, as proposed by Francis Crick's central dogma of molecular biology [2]. Since then, the roles of RNAs in the cellular and organismal context have been revised time and again. Simultaneously, the concept of non-coding RNAs had already been introduced, with the discovery of RNA as a component of the ribosome machinery by Georges Palade in 1955 [3] or tRNAs by Mahlon Hoagland and Paul Zamenick in 1957 [4]. Around the same period, the discovery of heterogeneous nuclear RNAs (hnRNAs), small nuclear RNAs (snRNAs which play a role in mRNA splicing) and small nucleolar RNAs (snoRNAs which play a role in ribosomal RNA processing) was also reported. In the early 1970s, David Baltimore along with his peers, discovered retroviruses and reverse transcriptases, thereby proving that indeed RNA acts as the carrier of genetic information [5]. Subsequently, in the 1980s, the discovery of intron splicing by Thomas Cech [6] and the processing of precursor tRNA into a mature tRNA by Sidney Altman, led to the discovery of ribozymes, whereby RNA could act as catalytic molecules [7, 8]. This led to the propagation of the RNA World Hypothesis, originated by Carl Woese but termed by Walter Gilbert, stating that the earliest forms of life on Earth could use RNA both as a carrier of genetic information and as catalytic molecules [9]. About a decade later, RNA interference (RNAi) emerged as a revolutionary concept, underlining the role of RNAs as regulatory molecules. Observations of RNAi in plants by Napoli and Jorgensen and then in *C.elegans* by Ambros *et al* were confirmed by the studies of Fire and Mello in *C.elegans*, leading to the discovery of microRNAs (miRNAs) and short-interfering RNAs (siRNAs) [10].

With the completion of the FANTOM (functional annotation of the mammalian genome) [11] and the ENCODE (encyclopedia of DNA elements) [12] projects, the concept of pervasive transcription was brought into the limelight whereby it was stated that the majority of the bases of the mammalian genome is associated with at least one primary transcript. Although only about 2% of the mammalian genome is transcribed into proteins, to decode the functional significance of the non-coding transcriptome (constituting approximately 90% of the genome) has remained an outright challenge.

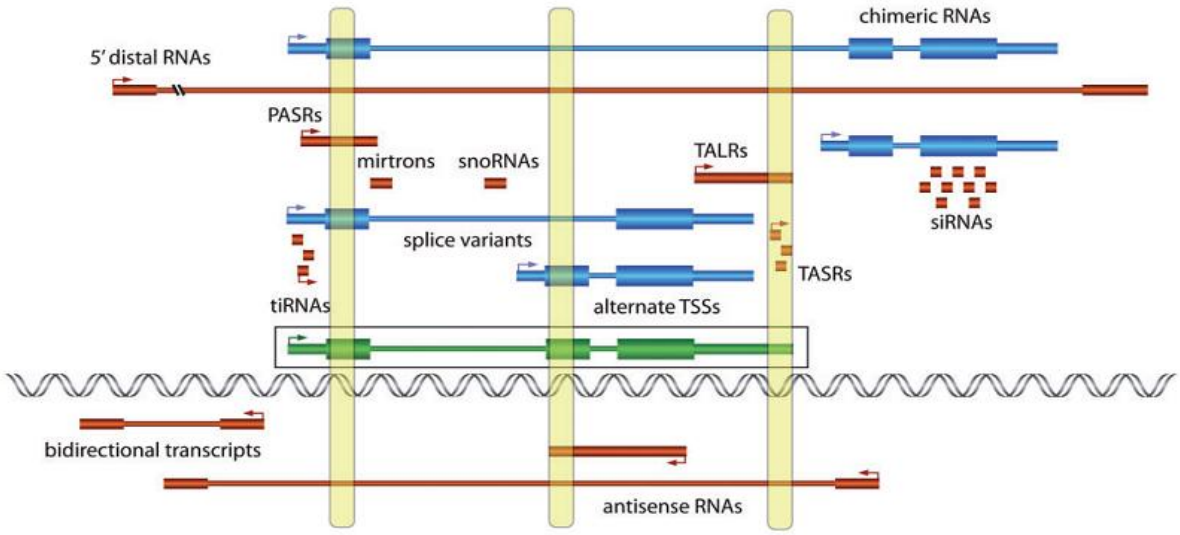


Fig. 1.1 *The complexity of the mammalian genome with respect to the coding and non-coding transcriptome. The boxed region refers to the traditional concept of a gene. The complexity is exemplified through representation of the varied associated non-coding transcripts. Abbreviations: PASRs, promoter-associated small RNAs; TALRs, terminal-associated long RNAs; TASRs, termini-associated short RNAs; tiRNAs, tiny RNAs. Adapted from [13] with permission.*

The mammalian non coding transcriptome can be broadly classified as follows:

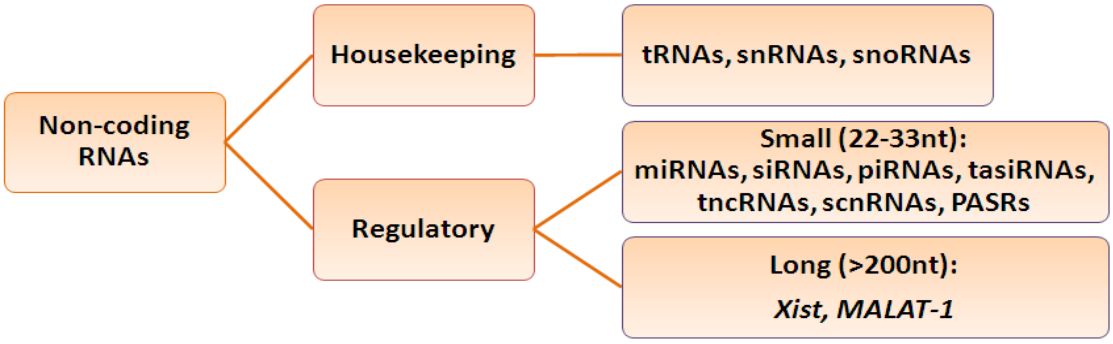


Fig. 1.2 *A broad classification of non-coding RNAs.*

It was only about a decade earlier that the class of germline specific small non-coding RNAs (sncRNAs) was discovered in the form of piwi-interacting RNAs (piRNAs) [14]. The characteristics of small non-coding RNAs are their biogenesis from a double-stranded precursor by RNase III-type enzymes and their association with the Argonaute family of proteins. While miRNAs (endogenous origin) and siRNAs (exogenous or endogenous origin) typically interact with the Ago class of the Argonaute family for their processing, piRNAs

interact with the Piwi class for the same. Also miRNAs and siRNAs are well known for their functions in target gene silencing at the post-transcriptional level, whilst piRNAs have been reported to function mainly in transposon silencing in the eukaryotic germline along with crucial roles in non-gonadal cells like memory storage and epigenetic regulation [15]. The other types of sncRNAs like the trans-acting siRNAs (tasiRNAs) in plants, tiny noncoding RNAs (tncRNAs) in worms, small scan RNAs (scnRNAs) in *Tetrahymena* sp., promoter-associated small RNAs (PASRs) and terminus-associated small RNAs (TASRs) have been shown to possess distinct but overlapping methods of biogenesis. They mediate important cellular functions ranging from genome rearrangement (scnRNAs) to modulation of chromatin modifications and transcription (PASRs) to regulation of copy number of target transcripts (TASRs) [13, 16] (**Figs. 1.1 and 1.2**).

According to the recently updated version of the GENCODE (encyclopedia of genes and gene variants) database, the mouse and the human genome can be divided into coding and non-coding components as follows (**Fig. 1.3**):

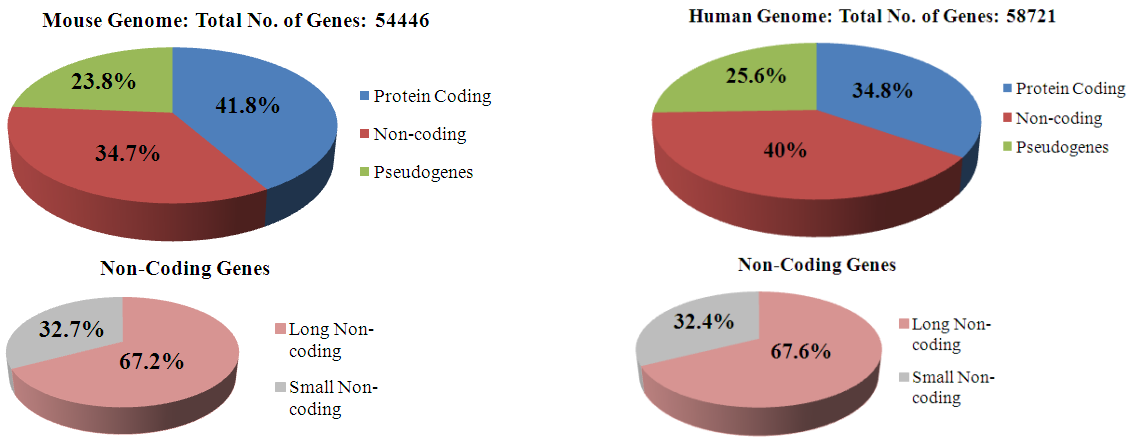


Fig. 1.3 *GENCODE statistics of the percentage of coding and non-coding genes in the mouse and human genome [17]. GENCODE Version M19, Oct 2018 (mouse) and Version 29, Oct 2018 (human).*

The second class of regulatory non-coding RNAs i.e., the long non-coding RNAs (lncRNAs) have gained widespread attention due to their diversity in mechanisms of action and regulation of various physiological processes. The first ever to be discovered lncRNAs were H19 [18] followed by Xist (X-inactive specific transcript) [19] through traditional gene mapping and hybridization array screenings. H19 was implicated in genomic imprinting whereas Xist was shown to function in dosage compensation. Henceforth, in three decades,

an overwhelming number of lncRNAs have been reported with regulatory functions in stem cell biology, disease biology, cancer biology and DNA repair biology, to mention a few, with many more still on the way of being functionally annotated.

1.2 Long Non-Coding RNAs: Characteristics & Classification

Upon propagation of the concept of pervasive transcription, for a long time scientists debated over the functional relevance of long non-coding RNAs and these transcripts were often dubbed as transcriptional noise owing to their abundant numbers, low sequence conservation, high tissue specificity and lower levels of expression as compared to mRNAs. However, it has been shown that the non-coding content of the genome increases with increasing organismal complexity suggesting they are not just an offshoot of the genomic size and they possess distinct features as detailed below [20-22].

LncRNAs have been typically classified based on their size (>200nt). Simultaneously, they possess a set of characteristics such as RNAP II dependant transcription, 5' capping and 3' polyadenylation that make them highly similar to mRNAs. A study by Guttman *et al* [23] revealed that lncRNAs also exhibit chromatin modifications like H3K4me3 at their promoters and H3K36me3 at their gene bodies. Needless to say, such K4-K36 domain containing lncRNA genes are transcriptionally active. However, recent studies have shown that there are subtle and distinct differences between mRNAs and lncRNAs:

1. They are less evolutionarily conserved than mRNAs: About 80% of the lncRNAs are of primate origin. This raised significant concerns over their functional roles. However, it has been observed that a majority of the lncRNAs display conservations in structures, functional domain containing sequences and expression from the syntenic loci. For example, lncRNA Gas5 (growth arrest specific 5) exhibits moderate sequence conservation of ~70% between mouse and human. However, a stem loop structure in the lncRNA that facilitates binding to the glucocorticoid receptor mediating the basic function of Gas5, remains conserved across the two species [24]. Similarly, in Xist, the 5' repetitive element encoding the RepA (repeat A) transcript is the most conserved region of Xist and it has two domains separated by a uracil rich linker. Whilst SUZ12 appears to interact with RepA as a whole, EZH2 and EED have been shown to bind to sub regions of RepA. The linker sequence is, however, divergent across human and mice in sequence but experimental evidence has shown that neither the length nor the sequence of this linker region affects the ability of the RepA transcript to bind to its protein interactors of the PRC2 complex, suggesting conservation at

the structure-function level [25, 26]. Furthermore, Malat-1 (metastasis associated lung adenocarcinoma transcript 1)exhibits high sequence conservation from zebrafish to humans but a knockout model system failed to show any clear phenotypic modifications, indicating that sequence conservation for lncRNAs does not predict their function across species [27].

2. LncRNAs have lower stability and differential chromatin modifications: In a recent review, Ard *et al* [28] have described the properties of lncRNAs with respect to mRNAs. LncRNAs in yeast and humans have been reported to be transcribed slower, display reduced levels of RNA Pol II (RNA Polymerase II) mediated elongation factor recruitment and targeted for degradation faster. They also have reduced levels of transcription active chromatin marks such as H3K4me3 and H3K36me3. Furthermore, certain classes of lncRNAs like enhancer lncRNAs display higher levels of H3K4me1 modification whereas as antisense lncRNAs display higher levels of H3 acetylation. In a report by Sati *et al* [29], lncRNAs in humans showed increased levels of DNA methylation around the TSS (transcription start site) as compared to mRNAs across four different types of tissues (H1 cell line, peripheral blood mononuclear cells, brain cortical tissue and brain germinal matrix tissue). These studies suggest that lncRNAs possess overlapping albeit distinct chromatin signatures as compared to mRNAs (**Fig. 1.4**).

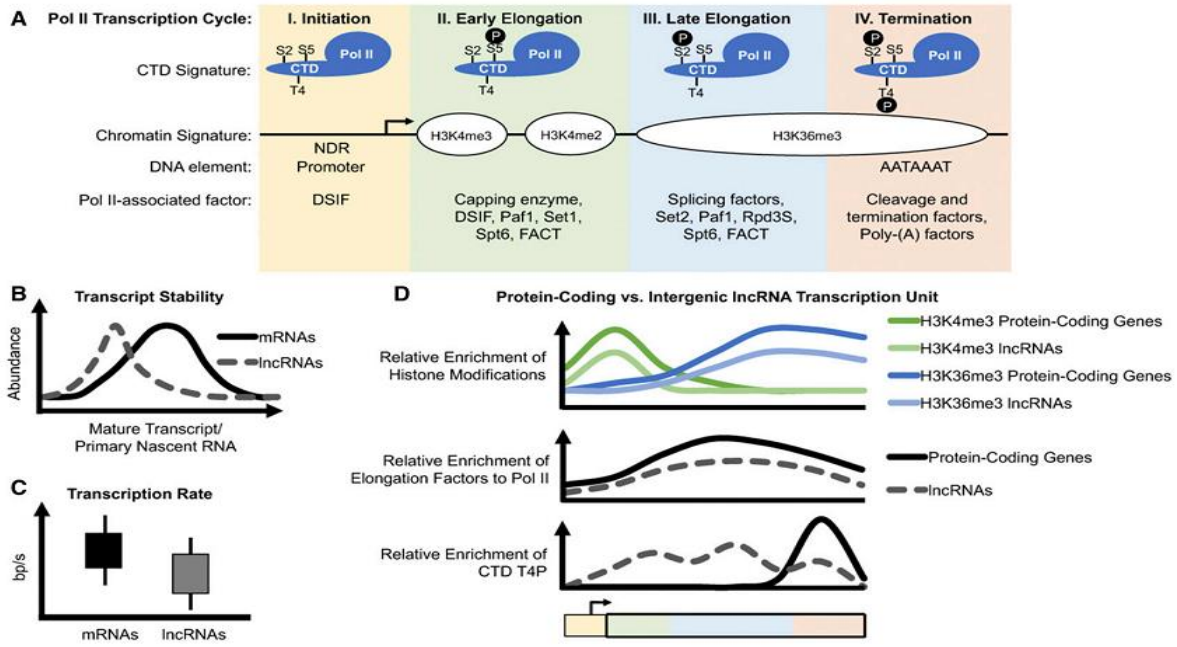


Fig. 1.4 *Comparison of lncRNAs and mRNAs. The RNA Pol II transcription cycle is characterized by four stages and various features as illustrated (A). A comparative analysis*

of the features between lncRNAs and mRNAs is depicted in (B, C, D). Adapted from [28] with permission.

3. LncRNAs exhibit greater tissue specificity: A majority of the lncRNAs have tissue-specific expression, suggesting their tight regulatory controls over biological processes. For example, Fendrr (FOXF1 adjacent non-coding developmental regulatory RNA) [30] and Braveheart [31] are heart-specific lncRNAs known to be crucial for their roles in heart development and differentiation. Similarly, Linc-MD1 (long intergenic non-coding RNA, muscle differentiation 1) [32] is specifically involved in skeletal muscle differentiation and lncRNA Megamind [33] is highly enriched only in the brain with indispensible functions in brain and eye development.

4. LncRNAs can be localized either to the nucleus or to the cytoplasm to perform their regulatory functions: In the nucleus, they can be *cis*-acting or *trans*-acting. LncRNA Kcnq1ot1 is responsible for recruiting G9a and PRC2 histone methyl transferases and subsequent bidirectional silencing of a host of paternally imprinted genes in the *Kcnq1* locus [34]. Xist, similarly acts to induce silencing of the X-chromosome by allowing the PRC2 complex to transfer histone methylation marks at that locus [35, 36]. These are the *cis*-acting lncRNAs. Hotair (HOX antisense intergenic RNA) is exemplary of the *trans*-acting nuclear lncRNAs. It is expressed at the *HoxC* locus but acts to silence the *HoxD* locus [37, 38]. In the cytoplasm, lncRNAs like linc-MD1 [32] and linc-p21 [39] act via the miRNA sponge mechanism or by Staufen-mediated decay respectively, to regulate their respective targets post-transcriptionally.

Hence, lncRNAs are similar yet unique in their characteristics as compared to their protein-coding counterparts.

LncRNAs can be classified based on a number of parameters as has been depicted in the picture below (**Fig. 1.5**)

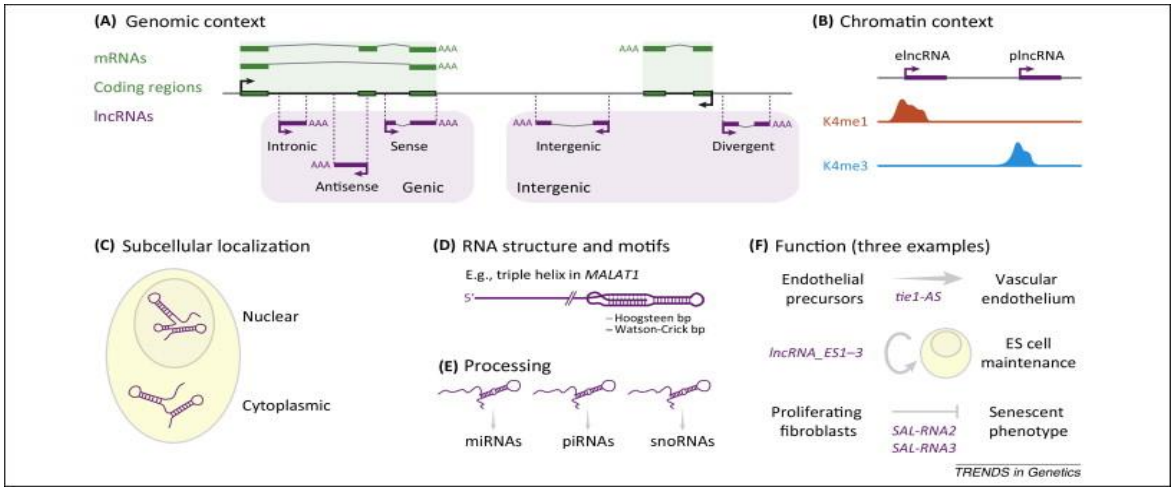


Fig. 1.5 *Classification of lncRNAs. Adapted from [22] with permission.*

Depending upon the genomic context of lncRNAs with respect to protein coding genes, they can be either sense or antisense (direction of transcription), genic or intergenic (overlapping or non-overlapping with protein coding genes), intronic (present within introns of protein coding genes) and divergent (bidirectional promoter transcription). Some of the lncRNAs are transcribed from annotated enhancer loci that possess a higher ratio of the enhancer mark H3K4me1 as compared to H3K4me3. These enhancer lncRNAs (eRNAs) are actively involved in recruiting cofactors, mediating chromosomal looping and hence enhancing transcriptional activity of neighbouring genes. An interesting subclass of these eRNAs are cheRNAs (chromatin associated enhancer lncRNAs) which remain bound to the chromatin and are found in the non-soluble chromatin fractions. Examples include lncRNA Pvt1 (Pvt1 oncogene, non-coding) [40] involved in bridging the chromosome to its target gene MYC and lncRNA Hidalgo [41] required for transcription of the HBG1 gene during the differentiation of K562 cells towards erythroid lineages. Whilst Pvt1 is located 60kb downstream of the *Myc* gene, Hidalgo overlaps in transcription with the *Hbg1* gene. This class of lncRNAs also can be classified as a subclass of the next category based on the subcellular localization namely, nuclear or cytoplasmic. In fact, all these classification categories are non-mutually exclusive. An yet another interesting classification of lncRNAs can be made based on their three-dimensional structural motifs. An example of this would be Malat-1 [42] which has a highly conserved 3' triple helical structure instead of a polyA tail to protect the mature transcript from exonuclease activity. Furthermore, many lncRNAs have been reported to undergo processing to generate small ncRNAs such as miRNAs, piRNAs and snoRNAs and hence lncRNAs can be categorised based on their processing too. An example of this would be

lncRNA Borderline which undergoes processing to generate small RNAs that define epigenetic domains [43]. Other lncRNAs such as Meg3 (maternally expressed 3) [44] and Fendrr [30, 45] form specific structures such as RNA-DNA-DNA triplex via their triplex forming motifs to regulate gene expression. Also, lncRNAs can be further grouped based on their specific mechanisms of action like decoys or miRNA sponges or scaffolds a well as their functions relating to biological phenomena.

1.3 Long Non-Coding RNAs: Mechanisms of Gene Regulation

LncRNAs have a diverse array to exert their actions on target genes and target transcripts to mediate regulation of biological processes (**Figs. 1.6 and 1.7**). They can act at the genomic level or at the post-transcriptional level or by modulation of protein activity.

1.3.1 At the genomic level

- **LncRNAs can alter chromatin structure through covalent histone modifications:** The archetype of such a mechanism of action would be Xist and Hotair. In the X chromosome, Xist is present on both the X_a (active X chromosome) and the X_i (inactive X chromosome) but is transcribed from the X_i to aid in the recruitment of the PRC2 complex to to impart dosage compensation in mammals. Interestingly, another lncRNA Tsix is expressed from the X_a itself and acts to prevent the interaction of Xist and PRC2 on the X_a [36]. The 5'end of Xist containing the repeat A motif allows interaction with EZH2 component of the PRC2 complex, directing spreading of the repressive histone modification H3K27me3 onto the X_i locus. Hotair acts in a similar manner but in *trans*, at the *HoxD* locus to mediate silencing of the locus through the recruitment of PRC2 and LSD1 (H3K4me1/2 demthylase) chromatin modifiers. Fendrr interacts with both PRC2 and TrxG/MLL complexes to regulate chromatin structure at their target loci [37, 38] The TrxG/Set1/MLL group of chromatin remodifiers are involved in the activation of gene transcription.
- **LncRNAs regulate chromatin dynamics through the recruitment of chromatin remodelers:** Linc-Cox2, situated downstream of the *Cox2* gene (cyclooxygenase 2), is known to associate with the chromatin remodeler complex SWI/SNF upon LPS stimulation of macrophages [46]. The resulting linc-Cox2/SWI/SNF complex allows chromatin remodelling at target late primary inflammatory response genes, allowing access to the transcription factor NF-κB on its cognate binding sites at the target genes thereby, increasing gene transcription. In a decoy mechanism, lncRNA Mhrt

(Myheart) binds to the ATPase/helicase domain of BRG1, a subunit of the SWI/SNF-like BAF chromatin remodelling complex [47]. In doing so, Mhrt competes out BRG1 from its tethering sites on target gene promoters and prevents it from effecting chromatin remodelling at these loci. The BAF chromatin remodelling complex reprograms cardiac gene expression to induce pathological hypertrophy. Hence, Mhrt acts to protect the heart from stress-induced heart failure and hypertrophy. The first ultraconserved lncRNA, Evf2 [48, 49], known for its function in the development of the mouse forebrain, also interacts with BRG1 in an interesting manner. The DLX1-BRG1-BAF chromatin remodelling complex binds to intergenic enhancers at the *Dlx5/6* locus and activates the transcription of *Dlx5*, *Dlx6* and *Evf2*. The resulting lncRNA Evf2 binds to and directly inhibits the ATPase activity of the bound chromatin remodelling complex, in a negative feedback inhibition mechanism, thereby converting an active enhancer to repressive one.

- Enhancer lncRNAs (eRNAs) can regulate higher order chromatin organization. eRNAs are a distinct class of lncRNAs characterized by very low levels of transcriptional activation modification (H3K4me3) with little bias in direction of transcription and lack of consistent processing phenomena such as splicing and polyadenylation [50]. One of the earliest examples of enhancer-linked non-coding transcription is that from the β-globin locus control region (LCR). The HS2 enhancer regulates transcription of the globin genes 10-50 kb away during erythroid cell differentiation [51]. It undergoes transcription giving rise to several non-coding intergenic RNAs, shorter than 3kb in length. It suggested that such transcription leads to the assembly of a machinery different from the basal promoter machinery and could act in enhancing the transcription at the target genes. In mouse cortical neurons wherein ~12000 neuronal activity regulated enhancer loci were studied, ~3000 were found to be associated with transcription, generating eRNAs <2kb in length [52]. Upon membrane depolarization, the levels of these eRNAs underwent dynamic perturbations and it was reflected in the mRNA levels of the nearby genes. Furthermore, in a specific case of the *Arc* enhancer, in a *Arc* knockout mouse, it was observed that a deletion of the promoter for the *Arc* gene did not lead to a change in the occupancy of RNA Pol II or CBP at the enhancer but there was no detectable eRNA synthesis at the Arc enhancer. This suggested that occupancy of the transcriptional machinery might not be enough to drive eRNA synthesis, rather an interaction of the enhancer with the promoter might be. In a similar study in

lipopolysaccaharide-induced inflammatory response in human monocytes, several non-coding transcripts derived from unidirectional and bidirectional enhancer regions were shown to undego differential expression [53]. Interestingly, knockdown of an eRNA (IL1β-eRNA) and RBT (region of bidirectional transcription, IL1β-Rbt46) selectively reduced LPS-induced expression and release of the proinflammatory mediators, IL1β and CXCL8. An eRNA discovered recently is Leene (lncRNA that enhances eNOS expression) [54]. In endothelial cells, Leene is expressed from a distal enhancer region of eNOS (endothelial nitric oxide synthase). In endothelial cells having a deletion of the Leene enhancer region, overexpression of Leene failed to promote eNOS expression suggesting that interaction between Leene genomic region with that of eNOS promoter region is essential for the transcription of eNOS in endothelial cells. Furthermore, a decreased interaction between Leene transcript and eNOS locus was observed suggesting that both enhancer-promoter physical interaction and eRNA transcript are essential for Leene to enhance eNOS transcription.

- **LncRNAs mediate genomic imprinting:** A host of lncRNAs have been reported to be involved in genome imprinting at various loci. A prominent example is the Kcnq1ot1 (KCNQ1 opposite strand/antisense transcript) lncRNA, expressed from a 1Mb *Kcnq1*/*Cdkn1c* locus on chromosome 7 [34]. It is paternally expressed and acts to imprint a number of protein-coding genes such as *Kcnq1*, *Cdkn1c*, *Phlda2* and *Slc22a18* in all tissues i.e., embryonic and extra-embryonic (ubiquitously imprinted loci) and *Osbpl5*, *Tssc4* and *Ascl2* only in the extra-embryonic tissues (placental-specific imprinted loci). LncRNA Airn (antisense IGF2R RNA) similarly regulates paternal imprinting at *Slc22a2*, *Slc22a3* and *Ifg2r* loci [55, 56]. Airn interacts physically with the *Scl22a3*, which is situated >230 kb away from the Airn locus, and recruits G9a methyltransferase to the *Scl22a3* promoter on the paternal chromosome in *cis* [57]. Xist, as has been discussed earlier, is well-known for X-chromosome inactivation to impart dosage compensation.

1.3.2 At the post-transcriptional level

- **LncRNAs as sponges:** One of the first lncRNAs discovered to be acting as a miRNA sponge was linc-MD1 [32]. Linc-MD1 possesses two binding sites for miR-135 and one binding site for miR-133 and it acts to sequester these miRNAs. miR-135 targets MEF2C and miR-133 targets MAML1. The sequestration of these miRNAs acts to

maintain the expression of their target genes during myogenic differentiation. In the context of gastric cancer, lncRNA BC032469 has been shown to act as a miRNA sponge for miR-1207-5p [58]. This microRNA targets hTERT, the rate-limiting subunit of telomerase, known to be associated with several malignancies. LncRNA BC032469 acts to sequester miR-1207-5p thereby leading to an upregulation of hTERT activity that ultimately leads to cellular proliferation, growth and progression of gastric cancer. Another recently identified lncRNA, Pagbc (prognosis associated gall bladder cancer lncRNA) was shown to be involved in promoting gall bladder tumorigenesis by acting as a sponge or ceRNA (competing endogenous RNA) [59]. LncRNA Pagbc binds to and decoys tumor suppressor miRs-133 and -511. The targets of these two microRNAs are SOX4, which activates transcription at several gene promoters of the PI3K/Akt pathway and PI3KR3 which is involved in the AKT/mTOR pathway. Sequestration of these microRNAs by lncRNA Pagbc leads to the activation of the Akt/mTOR pathway and subsequent malignancy in gall bladder carcinoma.

- **LncRNAs in Staufen-mediated mRNA decay:** LncRNA AF08799, also known as 1/2-sbsRNA1 (half-STAU1-binding site RNA), has been implicated in mediating Staufen-mediated decay (SMD) of its targets SERPINE1 and FLJ21870 [60]. These target mRNAs have a single Alu element in their 3'UTRs that allows lncRNA AF08799 to undergo intermolecular base pairing, thereby activating SMD. Interestingly, not only STAU1 depletion but knockdown of 1/2-sbsRNA1 caused an up regulation in the levels of both SERPINE1 and FLJ21870 mRNAs. Since Alu elements are found only in primates, a study in rodents revealed that in mouse, lncRNAs containing SINE elements (m1/2-sbsRNA) interacts with target mRNAs containing SINE elements to mediate SMD in rodents [61]. In the physiological context, such SMD is elevated during processes like myogenic differentiation of myoblasts to degrade mRNAs of proteins like Pax3 that maintain the myoblasts in an undifferentiated state. The study revealed that one of the lncRNAs, B2 SINE containing m1 / 2-sbsRNA2, promotes SMD of mRNA encoding a protein known as mTRAF6, an E3 ubiquitin ligase which in turn promotes myogenesis. Hence, lncRNA mediated SMD play important roles in cellular processes.
- **LncRNAs as natural antisense transcripts:** Natural antisense transcripts (NATs) are expressed from the opposite strand of a gene and can act either in *cis* or in *trans*. A

primary example of a lncRNA acting as a NAT is Tsix [36]. Tsix is expressed from the future active X-chromosome, it binds to Xist and prevents it from silencing this X-chromosome, thereby maintaining the phenomena of dosage compensation. An interesting *cis*-acting NAT is Bace1-as (β-site amyloid precursor protein cleaving enzyme 1antisense RNA) [62]. It binds to and stabilizes the Bace1 mRNA, contributing towards pathology of Alzheimer's disease (AD) and also acting as a potential candidate for AD prognosis. Nkx2.2-as, as a NAT, also exerts a positive regulatory action on its target NKX2.2 which has been implicated in oligodendrocyte differentiation [63].

1.3.3 At the protein activity level

- LncRNA Mrhl (meiotic recombination hotspot locus) has been shown to be involved in regulating WNT signalling in the context of spermatogenesis through its interaction partner p68 [64]. In spermatogonial cells, Mrhl sequesters p68 in the nucleus, acting as a decoy, preventing the action of p68. However, in cells depleted of Mrhl, p68 translocates to the cytoplasm, stabilizes β-CATENIN and facilitates its nuclear translocation and binding to its target genes along with TCF4 activating gene transcription. LncRNA Concr (cohesion regulated non-coding RNA) regulates sister chromatid adhesion by interacting with DDX11 protein [65]. DDX11 is a DEAD/H box containing DNA-dependant ATPase/helicase involved in the processing of the lagging strand during DNA replication and in the maintenance of the fork structure during the establishment of cohesion. The interaction between DDX11 and lncRNA Concr has been shown to be important for efficient binding of DDX11 to DNA *in vivo*. *In vitro* assays have revealed that Concr enhances the ATPAse activity of DDX11 suggesting a mechanism wherein interaction of lncRNAs with their protein partners can modulate their activity. LncRNA Norad (non-codng RNA activated by DNA damage) [66], similarly acts as a decoy for PUMILIO proteins to maintain genomic stability. In the absence of Norad, PUMILIO proteins bind to and repress mitotic, DNA repair and DNA replication factor genes, driving chromosomal instability.

1.3.4 Regulation of nuclear architecture by lncRNAs

- **LncRNAs as components of nuclear speckles and paraspeckles:** LncRNA Malat-1 has been implicated in the organization of nuclear architecture in eukaryotic cells.

Malat-1 is a component of nuclear speckles, dynamic nuclear structures comprising units of the spliceosome machinery, small nuclear ribonucleoproteins and serine/arginine-rich proteins [67]. Malat-1 is itself recruited to nuclear speckles through interactions with the spliceosome associated proteins and it is also associated with actively transcribing genes. It has been suggested that Malat-1 acts as a scaffold to position nuclear speckles to actively transcribing genes, therefore probably playing a role in the organization of actively transcribing domains in the nucleus. The nuclear enriched abundant transcript 1 (Neat1 or Men ε/β), expressed from a loci adjacent to *Malat-1*, is concentrated in paraspeckles in the nucleus, which are a hub for transcription and RNA processing. Through interactions with several of the paraspeckle proteins like p54nrb/NONO, PSPC1/PSP1, and PSF/SFPQ, it acts to maintain these nuclear sub-acrhitectural compartments and probably is important for regulating gene transcription [68, 69].

- **LncRNAs as organizers of chromatin architectures:** LncRNA Firre is expressed from the X-chromosome and escapes X-chromosome inactivation to remain associated with its site of transcription. LncRNA Firre (functional intergenic RNA repeat element) has been shown to be associated with five other unlinked autosomal loci apart from its own site of transcription. Deletion of Firre or knockdown of its protein interaction partner hnRNP-U abolishes these inter-chromosomal contacts, although it does not affect the expression of genes at these loci [70]. It suggests a mechanism whereby lncRNAs are involved in maintaining higher order chromatin structures. A recent study has shown that deletion of the Firre locus which contains 15 binding sites for CTCF does not affect TAD (topologically associated domain) boundary formation in mouse embryonic stem cells or fibroblasts, However, it does affect the formation a super loop chromatin structure at inactive X-chromosome locus [71]. LncRNA Xist has also been implicated in the organization of the 3D structure of the *Xist* locus. Xist interacts directly with Lamin B receptor (LBR), a transmembrane protein in the inner nuclear membrane, which in turn interacts with Lamin B that is essential to maintain the 3D shape of the nucleus by anchoring chromatin to the nuclear membrane [72]. Evidences have suggested that the Xist coated DNA would scan different nuclear locations and via its interaction with LBR, Xist would mediate the tethering of the X_i to the nuclear lamina. This would not only position the X_i away from the X_a but also bring various regions of the X_i together spatially, facilitating spreading of Xist to the rest of the X_i.

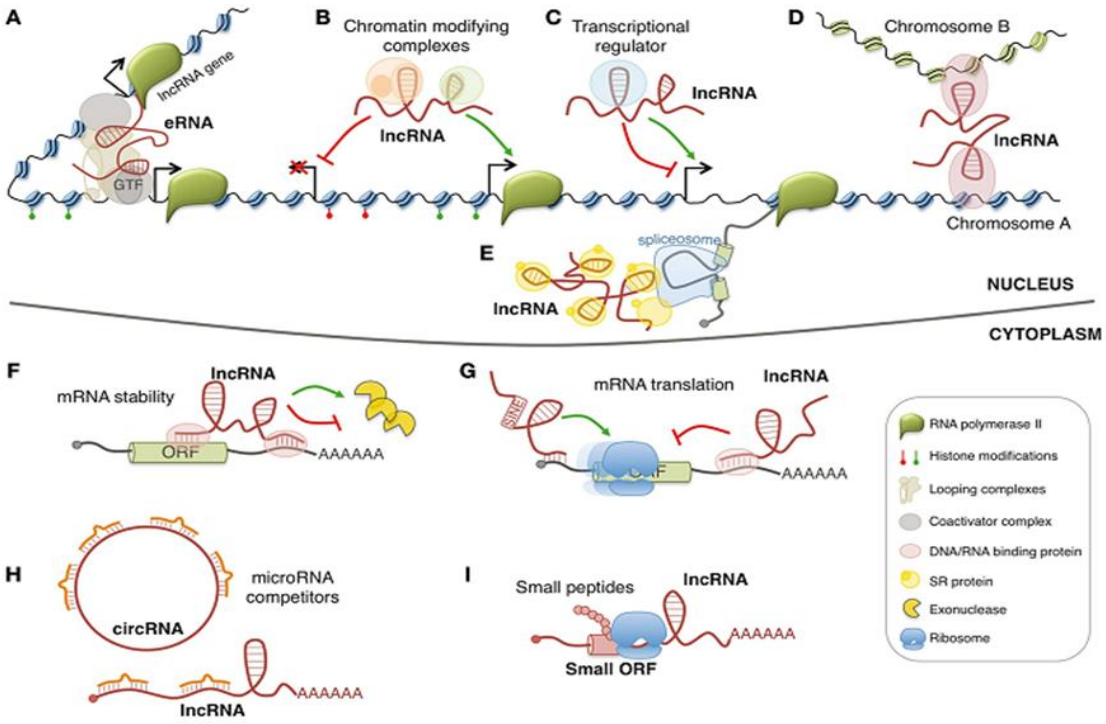


Fig 1.6 *Mechanisms of function of lncRNAs in the nucleus and cytoplasm. Nuclear lncRNAs can act in cis to regulate transcription at target loci by acting as enhancer RNAs (A), recruiting chromatin complexes (B), modulating transcription factor activity (C); in trans by facilitating chromosome conformations (D) or simply by regulating splicing of pre-mRNAs (E). Cytoplasmic lncRNAs influence mRNA stability (F), mRNA translation (G) or act as microRNA competitors (H). Additionally, some lncRNAs can code for functional peptides for regulation (I). Adapted from [73] (open access under CC BY 4.0).*

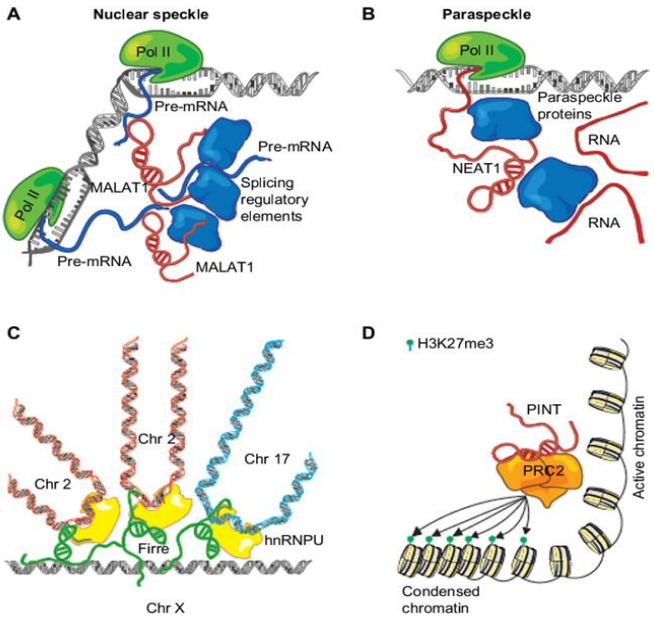


Fig. 1.7 *Mechanisms of action of trans-acting nuclear lncRNAs in maintaining nuclear/nucleolar architecture or chromatin state. Malat-1 and Neat-1 form components of speckles (A) and paraspeckles (B) respectively. Firre interacts with hnRNPU to associate distal chromosomes together (C). Pint confers silencing on target genes via deposition of H3K27me3 repressive mark (D). Adapted from [74] with permission.*

1.4 Long Non-Coding RNAs in Mammalian Physiology and Disease

1.4.1 LncRNAs in dosage compensation

The role of the lncRNAs encoded from the X-chromosome in dosage compensation in females has been discussed in length in the earlier sections. The *Xist* locus is replete with a cohort of lncRNAs, some of which act to activate Xist on the X_i while others act to repress Xist on the X_a (**Fig. 1.8**). The association of lncRNA Xist with PRC2 and in spreading the repressive histone modification throughout the length of the X_i has been well established.

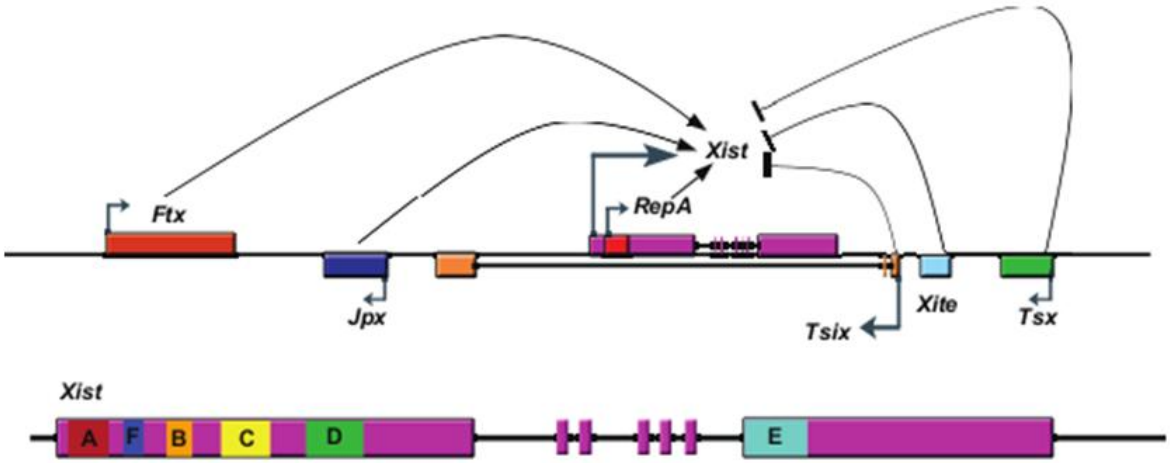


Fig. 1.8 *Schematic for the X-inactivation locus. Xist (X-inactivation specific transcript), Tsix (antisense of Xist), Tsx (testis-specific X-linked gene), Xite (X-inactivation intergenic transcription element), RepA RNA, Jpx [also known as Enox (Expressed Neighbor of Xist)], and Ftx (Five prime to Xist). Ftx, Jpx, and RepA lncRNAs promote Xist transcription, while Tsix, Xite and Tsx inhibit its transcription. Adapted from [75] (open access under CC BY-NC-SA 3.0).*

In a recent study, identification of the protein interaction partners of Xist has revealed SHARP, SAF-A and LBR as three of the major interactors [76]. SHARP interacts with SMRT corepressor which in turn activates the histone deacetylase HDAC3. The recruitment of SHARP acts in a dual mechanism by first deacetylating histones and causing eviction of RNA Pol II from the X_i locus and secondly, by allowing PRC2 to further silence the chromosome through repressive histone methylation marks.

1.4.2 LncRNAs in allelic imprinting

A multitude of lncRNAs have been shown to function in regulating genomic imprinting. Airn is maternally imprinted and mediates imprinting of certain genes in *cis* on the paternal chromosome [55, 56]. Kcnq1ot1 also is expressed paternally and acts to imprint various loci in the paternal genome itself or at certain placental specific loci [34]. LncRNA Gtl2 (also known as Meg3) deletion in mice has revealed similar interesting results. Maternal deletion of the locus caused bidirectional loss of imprinting of genes located in the *Dlk1-Dio3* loci. Such an effect was not observed when the deletion was transmitted paternally, suggesting that lncRNA Gtl2 is involved in maternal imprinting of genes at the mentioned loci [75]. LncRNA H19 is yet another lncRNA implicated in the maternal imprinting of genes surrounding the *Igf2* locus by interacting with a methyl CpG binding domain protein, MBD1 [77].

1.4.3 LncRNAs in body patterning

The *Hox* clusters of genes are involved in regulating positional identity of cells in the developing vertebrate embryo and in determining the anterior-posterior axis of the embryo. LncRNA Hotair is expressed at the *HoxC* locus but acts at the *HoxD* locus *in trans*. Hotair interacts with both PRC2 methylation complex and LSD1 demethylase complex and its conditional knockout leads to derepression of several *HoxD* genes resulting in patterning defects [37, 38]. LncRNA Hottip (HOXA transcript at the distal tip) [78], expressed from the *HoxA* locus, recruits WDR5/MLL complex to activate its neighbouring genes at the distal end by chromosomal looping. Its depletion led to abrogation of expression of genes on the *HoxA* locus with most pronounced effect on *HoxA13* and *HoxA11*, hence acting to regulate distal patterning. LncRNA Frigidair, on the other hand, is expressed from the *HoxC* locus and functions to direct anterior patterning [23].

1.4.4 LncRNAs in DNA repair and centromere functions

Amongst the first few reports of lncRNAs in DDR (DNA damage response) was the detection of a host of non-coding transcripts from the upstream region of the *Ccnd1* promoter following DNA damage signals [79]. The lncRNAs bind to TLS, a RNA-binding protein, which in turn binds to and inhibits the histone acetyl transferase activities of CBP/p300 complex on the *Ccnd1* promoter. This leads to down regulation of the *Ccnd1* gene, cell cycle regulator, upon DNA damage. LncRNA Anril (antisense non-coding RNA in the *Ink4* locus) has been shown

to be upregulated during DNA damage, especially in the late stages in p53 positive cancer cells, in an ATM-dependant manner [80]. Its transcription is mediated by E2F1 and it interacts with PRC1 and PRC2 to repress cyclin dependant kinase inhibitor encoding genes such as *Ink4b*, *Arf* and *Ink4a*. LncRNA Jade (Jade1 adjacent regulatory RNA) has been shown to act in DDR in more than one way [81]. It upregulates the transcription of *Jade1* leading to an increase in JADE1 mediated H4 acetylation and transcriptional activation of genes in response to DNA damage. It also induces G1/S cell cycle arrest, inhibition of apoptosis and positively regulates the DNA damage repair protein MDC1, as a part of response to DNA damage. LncRNA Panda (p21 associated ncRNA DNA damage activated) has been shown to be upregulated in several cancers but to act as a tumor suppressor in diffuse large B-cell lymphoma [82, 83]. It participates in cell cycle arrest and cell survival by repressing pro-apoptotic genes like *Ccnb1*, *Fas*, *Puma*, and *Noxa*. Knockdown of p53 also reduces the induction of Panda upon DNA damage, suggesting that it is a target of p53. Another direct target of p53, lncRNA Pincr (p53 induced non-coding RNA) [84] is induced in expression upon doxorubicin-mediated DNA damage in colorectal cancer lines. Silencing of Pincr leads to increase in cell sensitivity to chemotherapeutics, suggesting Pincr has a pro-survival function in DDR.

LncRNAs have been shown to play essential roles in maintaining centromere integrity as well. In *Drosophila*, SatIII RNA is a lncRNA expressed from the *SatIII* locus, a 359 bp repeat unit containing locus in the X chromosome, known as satellite III [85]. SatIII RNA interacts with the inner kinetochore protein CENP-C for localizing to the centromere. Loss of SatIII RNA causes defects in mitosis and chromosome segregation along with reduced deposition of CENP-A and CENP-C to the centromeres. In mouse, minor satellite repeats encode non-coding transcripts of length around 4 kb that maintain centromeric stability by associating with CENP-A as well as components of the chromosome passenger complex (CPC) such as AuroraB, Survivin and INCENP [86]. In Xenopus, lncRNAs have been demonstrated to bind to CPC *in vitro* and *in vivo* [87] whereas in humans alpha satellite-repeats have been implicated for centromere assembly [88], thereby suggesting that non-coding RNAs are involved in promoting the proper functioning of centromeres.

1.4.5 LncRNAs in immunological development

A plethora of lncRNAs have been implicated to function during innate and adaptive immune responses. Linc-Cox2 has been shown to be highly up regulated in expression upon LPS

stimulation of mouse $CD11C1^+$ bone marrow derived dendritic cells in a TLR4-NFκB mediated signaling response [46]. It acts to negatively regulate inflammation associated interferon genes such as *Ccl5* but activates *Tlr1*, *Il6* and *Il23a*. It mediates its action of transcriptional interference through its interaction with hnRNP-A/B and hnRNP-A2/B1 protein partners. Pacer (p50-associated Cox-2 extragenic RNA) is also associated with the *Cox2* locus [89]. LncRNA Pacer acts to replace the repressor complex p50-p50 with the activator complex p50-p65 to mediate p300 binding and activation of *Cox2* gene expression in *cis*, in primary human mammary epithelial cells and in PMA-stimulated human monocyte–macrophage cell lines. In a genome-wide study targeted to analyze perturbations in expression of lincRNAs upon LPS stimulation of THP1 macrophages, linc1992 or Thril (TNFα and hnRNPL related immunoregulatory lincRNA) was identified as an interesting candidate [90]. Thril was shown to function through its interaction partner hnRNP-L to regulate TNFα and a host of other innate immune response genes. In host-pathogen responses, lncRNA Neat1 was one of the first lncRNAs to be implicated in HIV-1 replication [91]. Neat1 levels were seen to be up regulated during HIV-1 infection and it acts to reduce virus production by decreasing nucleus-to-cytoplasm export of Rev dependent instability element (INS)-containing HIV-1 transcripts during posttranscriptional regulation. LncRNA Nron (non-coding repressor of Nuclear Factor of Activated T-cells) was shown to be down regulated upon HIV-1 infection as it modulates HIV-1 replication [92]. Knockdown of Nron in Jurkat cells infected with HIV-1 led to increase in the activity of NFAT and HIV1-LTR activity that led to an increase in HIV-1 replication as compared to control cells infected with HIV-1. Thus, lncRNAs regulate a diverse array of immunological responses.

1.4.6 LncRNAs in cancer

The role of lncRNAs in cancer is overwhelming wherein they exhibit cancer type specificity (**Fig. 1.9**) and act as both tumor suppressors and oncongenes. The first of the lncRNAs to be associated with cancer prognosis were Pca3 (prostate cancer associated 3 or DD3) and Pcgem1 (prostate specific transcript 1) in a screen of lncRNAs in prostate cancer versus normal tissues, with Pca3 being the first FDA approved lncRNA to be used as a non-invasive marker of prostate cancer. LncRNA Malat-1 is also implicated to be involved in cell proliferation in a diverse range of cancers such as liver, breast and colon, with lung carcinoma being the first of the cancer types in which it was used for prognosis. LncRNA Hotair is over expressed in metatstatic breasts, lung and pancreatic tumors and has been shown to promote metastasis (reviewed in [93]). LncRNA Schlap1 (second chromosome

locus associated with prostate-1) is another of the lncRNAs that is up regulated in prostate cancer and promotes tumor invasion and metastasis by preventing the action of the SWI/SNF chromatin remodeler complex at a genome-wide level [94]. LncRNA Anril or Cdkn2b-As1 has been shown to be associated with disease-related SNPs and it acts to silence its neighbouring gene (a cyclin dependant kinase inhibitor) by interaction with CBX7 and PRC2, thereby promoting cellular proliferation [95, 96]. Amongst the tumor suppressors, lncRNA Gas5 has been shown to be down regulated in breast cancer. It competes out the binding of the glucocorticoid receptor to its cognate DNA sites and induces cell arrest and cell sensitivity to apoptosis. Linc-Pint (long intergenic non-protein coding RNA, p53 induced transcript) interacts with PRC2 complex and silences target genes; it is down regulated in colorectal cancer. LncRNA Ptenp1 (phosphatase and tensin homologue pseudogene 1) acts as miRNA sponge that target PTEN, a known tumor suppressor protein that can impede cellular proliferation. Ptenp1 has been shown to inhibit cellular proliferation, tumor growth and invasion and its locus is selectively lost in sporadic colon cancer, prostate cancer and melanoma. LncRNA Nbat1 (neuroblastoma associated transcript 1) can be used as a clinical marker for neuroblastomas. It acts to prevent tumor growth and invasion and also aids in proper neuronal differentiation of precursors by targeting the PRC2 complex to silence REST, a protein known to repress neuronal genes. In neuroblastomas, Nbat1 undergoes CpG methylation and is also known to be associated with a high-risk single-nucleotide polymorphism on chromosome 6p22 that reduces its expression levels, thereby leading to aggressive neuroblastoma and impaired neuronal differentiation (reviewed in [97]. The extensive interactions of lncRNAs with proteins to regulate cancer physiology serve as a promising tool for cancer therapeutics. With the help of antisense oligonucleotides (ASOs) or nanoparticle mediated deliver of siRNAs or small molecule inhibitors of lncRNAs, lncRNAs can be targeted in specific cancers as an alternative to chemotherapeutic drugs which often confer resistance. A few approaches in this direction have been undertaken such as the supplementation of Gas5 mimicking locked nucleic acids to hormone receptors has been shown to induce cell apoptosis [98]. ASOs against Malat-1 have also been shown to be effective in a preclinical mouse MMTV-PyMT breast cancer model by causing cystic differentiation, increased cell adhesion and reduction in cell migration [99]. Thus, lncRNAs hold unforeseen prospective in being used as prognostic markers as well as therapeutic targets in cancer pathobiology.

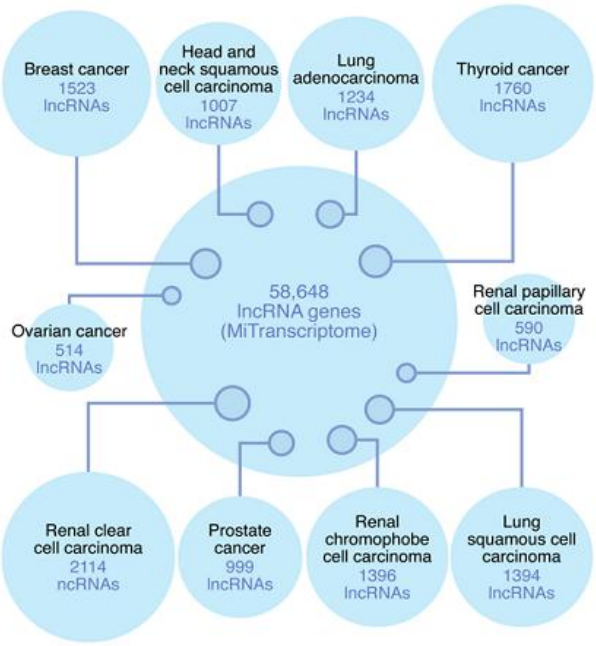


Fig. 1.9 *LncRNA genes in cancer. The MiTranscriptome bioinformatics effort identified ~60,000 lncRNAs involved in cancer with 8,000 of them being cancer/lineage specific. Adapted from [93].*

1.5 Long Non-Coding RNAs: Techniques

1.5.1 Identification of lncRNAs: transcript sequence assembly

As a first step towards annotating lncRNAs, their identification becomes crucial. Several techniques exist that help in the discovery of lncRNAs in various cellular and physiological contexts. Over the years, the traditional cDNA library and hybridization arrays have been replaced by more high-throughput next generation sequencing techniques that provide various advantages such higher signal-to-noise ratio, lower amounts of starting material and transcript analysis at the single cell level. The techniques with respect to the discovery of lncRNAs are outlined as follows:

Technique	**Methodology**	**Application**	**Advantage/Disadvantage**
cDNA Library and Tiling Arrays	cDNA is generated and Sanger sequencing is performed to identify the transcript repertoire in a cell or tissue specific manner. In tiling	The FANTOM project dedicated to identify and annotate functional elements in the mouse genome was the first large-scale project where 60,770 cDNAs were subjected to Sanger sequencing	High noise to signal ratio and cross-hybridization.

	arrays, cDNAs are allowed to hybridize onto microarray slides probed with tiled oligonucleotides designed to represent non-repetitive sequences across several chromosomes or the genome as a whole.	revealing ~70% of the genome to be non-coding in nature [11]. Identification of transcripts from the four human *Hox* gene clusters and discovery of lncRNA Hotair was performed using tiling arrays [37].	
Serial Analysis of Gene Expression (SAGE)	Short stretches of cDNA sequences containing restriction enzyme sites at their 3'ends are concatenated and subjected to cloning and Sanger sequencing.	About 272 SAGE experiments performed across normal and cancerous human samples revealed that lncRNAs are distributed across all the chromosomes with only 1% being ubiquitously expressed in all the investigated tissues, suggesting the tissue-specific expression manner of lncRNAs [100].	Allowed the identification of new transcripts as well as quantification of the existing ones.
RNA-Sequencing (RNA-Seq)	Libraries are generated with the help of oligodT primers to enrich for polyadenylated transcripts ; random priming was introduced as an alternative; single-	Around 1000 lincRNAs were identified through RNA-Seq across three different cell types (embryonic stem cells, neuronal precursors and lung fibroblasts) much of which was not previously known [101]. In a parallel	Derivatives of this technique such as RNA Capture Sequencing (a combination of tiling arrays and RNASeq), ChIP-RNA Seq [for transcription and

	end and paired-end libraries are also generated nowadays wherein in paired-end sequencing, the same transcript is sequenced from either directions, allowing more information recovery such as splice variants and identification of boundaries of transcription units in cases of multiple start and polyadenylation signals.	study, 8000 lincRNAs were catalogued across 24 human tissues and cell types from 4 billion RNA-seq reads [102].	epigenetic regulators], ChIA-RNA Seq [Chromatin Interaction Analysis] and DNA-RNA Seq have been developed and extensively used to suit the requirements of specific studies.
Cap Analysis of Gene Expression (CAGE)	Allows the identification of 5'cap transcripts narrowing down the repertoire to Pol-II transcribed genes.	A CAGE study performed on human and mouse induced pluripotent stem cells, embryonic stem cells and differentiated cells led to the identification of close to 400 novel transcripts in humans and about 8000 in mice that were named as non-annotated stem transcripts (NASTs) [103]. It was also discovered that a large	An improved version of this technique wherein sequencing could be performed with very small amounts of starting material was nanoCAGE and CAGEScan.

		number of these transcripts were generated from long repeat regions that constitute promoter and enhancer regions.	
Single-Cell RNA Sequencing (scRNA-Seq)	Isolation of single cells using FACS or commercial microfluidics systems followed by SMART-Seq, DP-Seq or Quartz-Seq.	In mouse blastocysts, it was discovered that heterogeneity in the transcriptomes of individual cells was a pre-requisite for lineage specific differentiation [104]. sc-RNA Seq of individual cells of the developing brain neocortex revealed that many lncRNAs are abundantly expressed in individual cells and their expression was very cell-type specific [105].	Due to scarcity of starting material, scRNA-Seq faces high variability in terms of technical noise (RNA capture efficiency, random dropouts during libraray generation) and biological noise (cell sizes, cell cycle states and stochastic gene expression).
Global Run-On Sequencing (Gro-Seq)	Extracted nuclei is subjected to run-on assays by subjecting them to Br-UTP labelled nucleotides and the detergent Sarkosyl (causes RNA Pol II release) thereby generating nascent transcripts which are ~100t in length and labelled.	RNA-Seq and Gro-Seq studies have revealed that divergent transcription occurs at the promoter regions of active protein-coding genes [106].	Sample preparation and extraction of Br-UTP labelled RNA remains technically challenging.

	The labelled transcripts are captured by anti-Br-UTP antibody and subjected to deep sequencing.		

1.5.2 Localization of LncRNAs

For decoding the function of lncRNAs, it becomes a prerequisite to understand their subcellular localization. As explained previously, lncRNAs can act through diverse yet distinct mechanisms depending on whether they are localized to the nucleus or to the cytoplasm. Inside the nucleus, localization studies may also shed light upon whether the lncRNA under investigation occupies defined territories or specific domains. The major technique available in this regard is mentioned below:

Technique	**Methodology**	**Application**	**Advantage/Disadvantage**
Fluorescence In Situ Hybridization (FISH)	Probes bearing fluorescent tags and complementary to a target of interest are allowed to hybridize and then detected by imaging.	LncRNA Malat-1 was found to be localized in speckles in nuclei of interphase cells and interchromatin granule clusters in mitotic cells [107]. LncRNA Neat1 is found to be localized in paraspeckles [108]. LncRNA Kcnq1ot1 was found to form domains within which the genes can undergo epigenetic inactivation [109].	Cis or trans action of lncRNAs is not decipherable. smRNA (single molecule) RNA FISH helps detect low abundant lncRNAs by using a pool of probes scanning the length of the lncRNA.

1.5.3 Function of LncRNAs

1.5.3.1 Loss-of-function studies

In forward genetics, genes responsible for an aberrant phenotype are identified and characterized. In reverse genetics, a gene of interest is either knocked out or its transcript levels are depleted to understand resulting phenotype(s). In case of lncRNAs, reverse genetics is very useful in decoding their target genes/pathways for regulation or their functional relevance with respect to the resultant phenotypes. The techniques in this regard are as follows:

Technique	**Methodology**	**Application**	**Advantage/Disadvantage**
RNA Interference (RNAi)	Short interfering RNAs (siRNAs), Short hairpin RNAs (shRNAs) or Antisense oligonucleotides (ASOs) induce degradation of the target RNA by the Drosha/Dicer machinery or by the RNAseH dependant degradation pathway respectively.	Depletion of endogeneous lncRNAs by RNAi, in combination with other techniques, provides in depth insights about its mechanism of action and the target genes that it regulates.	Off-target effects of siRNAs or shRNAs. Cellular localization and for the lncRNAs has also to be taken into consideration.
Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)	CRISPR involves Cas9 nuclease mediated targeted editing or deletion of specific sequences in the genome. CRISPRi	A CRISPR/cas9 mediated deletion of candidate lncRNAs in murine macrophages led to the identification of two lincRNAs, lincRNA-Cox2 and lincRNA-AK170409	The neighbouring genes or loci should be taken into consideration when knocking out lncRNA loci since many

	or CRISPRa involves a modified Cas9 targeted to inhibit or activate genes at the transcription level.	[110]. They were shown to regulate NF-κB signalling through distinct pathways. In a genome-wide CRISPRi screen across 7 cell types including transformed and iPS cells, 499 lncRNA loci were identified out of 16401 which were involved, often in a cell-type specific manner, in regulating cellular growth and transcriptional networks [111]. A CRISPRa screen in melanoma cells similarly led to the identification of 11 lncRNAs, one of them being Emiceri, that confer resistance phenotype [112].	lncRNAs act as enhancers. Verification of off-target effects is also important.

1.5.3.2 LncRNA-Interactome Studies

LncRNAs act to regulate their targets through interactions with protein partners such as directing chromatin remodelers to their sites of action or through acting as decoys for proteins such as transcription factors. They also interact with DNA at their sites of action through protein partners or through the formation of RNA-DNA-DNA triplex structures. Furthermore, they have been shown to interact with other RNA molecules for the regulation of various processes. Outlined below are some of the techniques in this regard:

RNA-Protein Interactions:

Technique	Methodology	Application	Advantage/Disadvantage
RNA Immunoprecipitation (RIP)	RIP is used to identify RNAs that interact with a protein of interest using a specific antibody.	PRC2 was identified to be the interacting partner of lncRNA Xist which led to the deduction of the epigenetic mechanism of X-chromosome inactivation [36].	Cross-linking of cells is not required thereby reducing artefacts. Use of nucleases can help understand if protein interaction is direct or if the RNA interacts as a single strand or in specific conformations.
High-throughput sequencing cross-linking immunoprecipitation (HITS-CLIP)	Also known as CLIP-Seq, this technique involves UV crosslinking followed by RNAse and proteinase digestion and purification. The purified RNA segments are then subjected to high-throughput sequencing.	A cohort of lncRNAs was identified in human colorectal cancer cells hat interact with PRC2 [113]. CLIP-Seq on RNA Pol II in HeLa and HEK293 cells has identified 100 novel exon-intron cricular RNAs [114].	UV crosslinking may lead to base mutations. Low-abundant lncRNAs also may evade detection.
Photoactivatable ribonulceotide-enhanced cross-linking immunoprecipita-	Photoreactive ribonucleoside analogs, 4-Thiouridine (4SU) and 6-thioguanosine	The interaction of JARID2, an accessory component of the PRC2 complex, with many lncRNA	The analog induced mutations allow the detection of the specific binding sites for the

tion (PAR-CLIP)	(6SG), are applied to living cells to facilitate stronger cross-linking. These analogs upon UV exposure lead to a T to C or G to A mutation.	including Meg3 was identified in embryonic stem cells through PAR-CLIP [115].	proteins on the RNAs. iCLIP Individual nucleotide resolution CLIP) allows the detection of binding sites at single nucleotide resolution.

RNA-DNA Interactions:

Technique	**Methodology**	**Application**	**Advantage/Disadvantage**
Chromatin Isolation by RNA Purification (ChIRP)	Cells are cross-linked with formaldehyde and the chromatin is isolated and sonicated following which oligos against the RNA of interest are used to pull down the chromatin. The DNA is then purified from the chromatin and subjected to sequencing. Alternatively, proteins can be purified and subjected to mass-	LncRNA Hotair was shown to preferentially associate at GA-rich regions in the genome using this technique [116]. ChIRP can be combined with mass spectrometry as well LncRNA Mrhl was shown to bind to 1370 loci in the genome in mouse spermatogonial cells and it was shown to interact with proteins like hnRNAP A/B, hnRNPA2/B1 and PC4 [117]. Xist lncRNA has been shown to interact with	Use of tiled oligonucleotides covering the entire length of the RNA is required since RNA conformations in vivo can be dynamic.

	spectometry.	81 different proteins [118].	
RNA-Antisense Purification (RAP)	This technique is similar to ChIRP except that longer oligos complementary to the RNA of interest are used.	The precise binding locations of lncRNA Xist on the inactive X-chromosome were identified [119].	Longer oligos ensure higher specificity and greater signal-to-noise ratio.
Capture Hybridization Analysis of RNA Targets (CHART)	Similar to ChIRP and RAP, CHART utilizes afiinity tagged oligos. The oligos are specifically designed against open regions of the RNA, determined initially by oligo binding and RNAseH mapping.	CHART was used to functionally characterize lncRNA Rox2, involved in dosage compensation in *Drosophila* [120]. With this technique and a time-course analysis, it was also shown that lncRNA Xist first binds to gene-rich clusters and then spreads to gene-poor regions [121].	More focused design of probes eliminates the need to synthesize a larger number to cover the entire length of the RNA.

RNA-RNA Interactions:

Technique	**Methodology**	**Application**	**Advantage/Disadvantage**
RAP-RNA	Similar to RAP but the cross-linking methods are different in this technique. RAP-	U1 snRNA was shown to hybridize to the 5'splice sites of nascent RNAs. LncRNA Malat-1 was	RAP-RNA[FA-DSG]. involves combined treatment with FA and disuccinimidyl glutarate (FA-DSG)

	RNA[AMT] can be used to detect direct RNA–RNA interaction by utilizing the cross-linker 4'aminomethyltrioxalen (AMT), which generates specific uridine bases cross-links. RAP-RNA[FA], and RAP-RNA[FA-DSG]. are the other variations.	shown to interact with pre-mRNAs indirectly through protein complexes [122].	that creates strong cross-links and can capture RNAs bound by multiple protein partners.
Cross-linking, ligation and sequencing of hybrids (CLASH)	UV cross-linking followed by pulldown of RNA-RNA complexes using affinity-tagged protein (that binds to either of the RNAs).	snoRNA-rRNA interactions in yeast and miRNA [123] interactome of AGO1 protein were identified through this technique [124].	By varying the time of UV crosslinking, kinetics of RNA-RNA interactions can be studied.

1.5.4 Secondary Structure of lncRNAs

LncRNAs often display conservation at the structure-function level than at the sequence-function level. They can adopt specific motifs or form distinct secondary structures that can mediate their functions in a specific context. Following are the techniques for understanding the secondary structure of lncRNAs:

Technique	**Methodology**	**Application**	**Advantage/Disadvantage**
Selective 2′ -hydroxyl acylation by primer extension (SHAPE)	In this technique, N-methylisotoic anhydride (NMIA) and 1-methyl-7-nitroisatoic anhydride (1 M7) reagents are used to selectively modify RNA to form a 2′-O-adduct. Modified positions are then detected using primer extension.	The secondary structure of the entire RNA genome of HIV has been decoded [125].	Whether the base-pairing interactions are at close or long range cannot be determined. Shotgun secondary structure (3S) technique is used to fragment the RNA and then compare the profile against the whole RNA SHAPE profile to obtain data on local structures as well as the overall structure.
Fragmentation Sequencing (Frag-Seq)	RNA is digested with nuclease P1 and then subjected to NGS mediated structure probing. P1 cleaves single stranded nucleic acids making it possible to probe into secondary structures of RNA at a precise resolution.	The nuclear transcriptome of ES and neuronal precursor cells was analyzed to identify single-stranded stretches for previousy known lncRNAs. The structure of U15b C/D box snoRNA was also determined [126].	Its a high throughput genome-wide technique. But it does not provide resolution at a single-nucleotide level.

Parallel Analysis of RNA Structure (PARS)	RNA is digested with single or double strand specific nucleases, converted to cDNA and subjected to deep sequencing.	Structure of lncRNAs in the budding yeast transcriptome was mapped [127]. PARS on a family (father, mother and child) revealed that 15% of transcripts bearing single-nucleotide changes alter the RNA secondary structure and many of such transcripts (known as riboSNitches) have been linked with diseases [128].	Low abundant transcripts are barely detected.

1.5.5 Ribosome Profiling

Ribosome profiling techniques were developed to better understand the non-coding nature of lncRNAs:

Technique	**Methodology**	**Application**	**Advantage/Disadvantage**
Ribosome Profiling	RNA-ribosome complexes are stabilized by cycloheximide treatment followed by RNA digestion. RNA fragments protected by ribosome interactions are	This method revealed that the frequency of lncRNAs engaged by ribosome is comparable to that of 5'UTR of protein coding genes [129]. UCHL1 antisense lncRNA enhances translation of the	Relatively short fragments are recovered which might pose a problem during mapping for repetitive sequences. Possibility of recovering RNAs

	purified and subjected to high-throughput sequencing.	protein-coding counterpart by directing the sense mRNA to the polysomal fraction [130].	protected by proteins other than ribosomes.
Translating Ribosome Affinity Purification (TRAP)	Genetically modified mouse encoding eGFP tagged ribosomal proteins allow the recovery of mRNAs associated with at least one 80S ribosome. The recovered RNAs are then subjected to microarray or sequencing.	TRAP analysis on embryonic brain and kidney tissues revealed that 85 ncRNAs in brain and 60 in kidney are actually associated with translating ribosomes [131].	

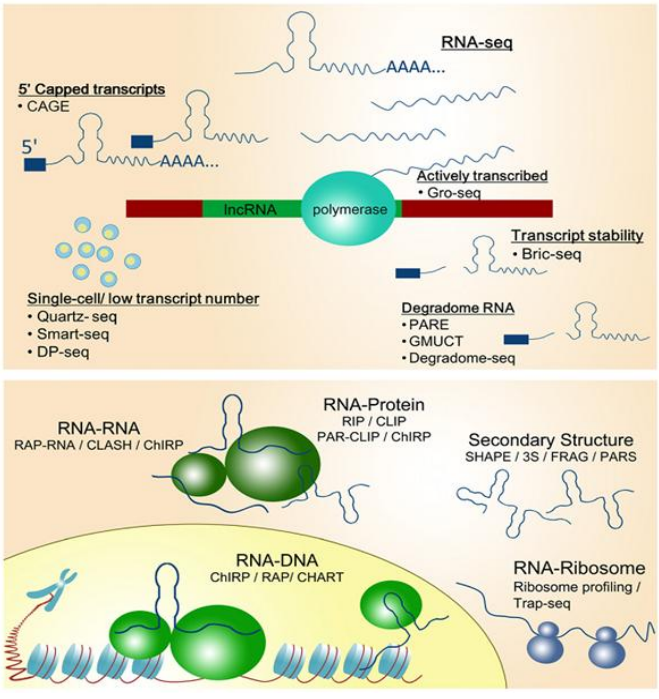


Fig 1.10 *A comprehensive understanding of the non-coding transcriptome would require multiple approaches as listed above and as illustrated in the picture. Adapted from [132] with permission.*

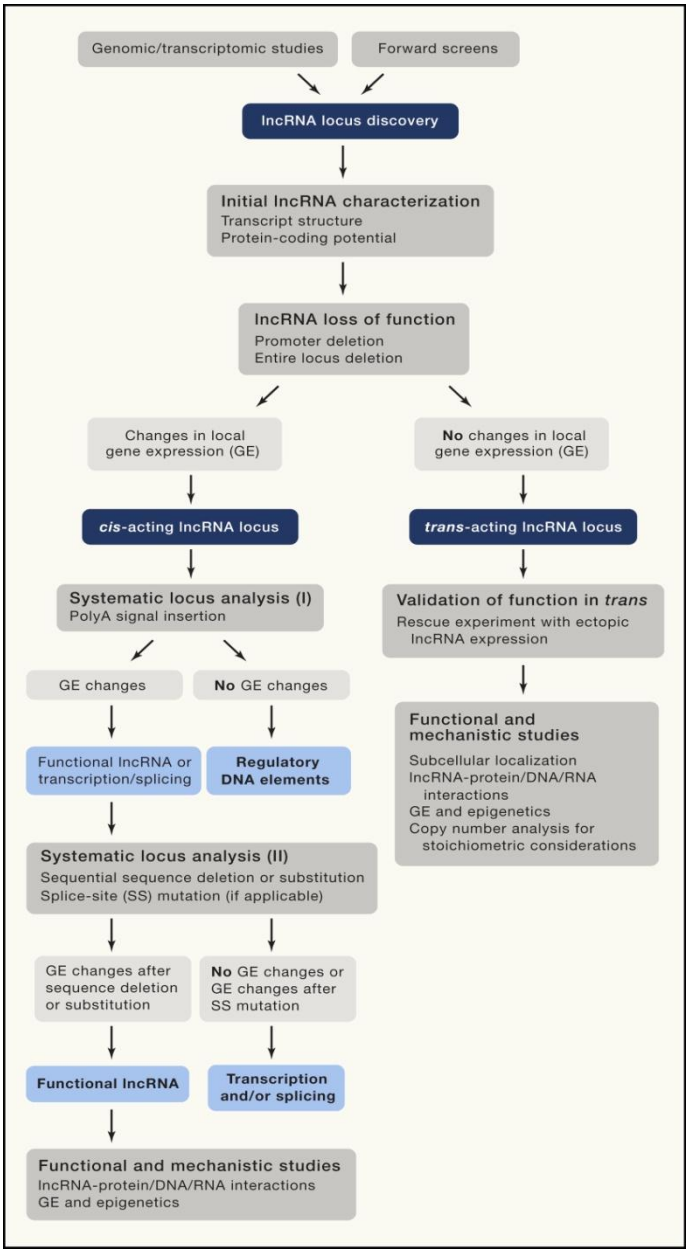


Fig. 1.11 *A probable pipeline for understanding functional relevance of lncRNAs. Genomic or transcriptomic screens lead to the identification of lncRNAs. Initial characterization of lncRNAs is performed with respect to their structure, coding potential and subcellular localization. Loss-of-function studies reveal their regulation of local or distal genes classifying them as cis- or trans-acting lncRNAs. Systematic analyses of cis-acting lncRNAs would reveal if the transcript or their act of transcription is necessary for regulation of their target loci. Furthermore, studies towards understanding their interactions with proteins, DNA or RNA would provide insights into their mechanisms. Trans-acting lncRNAs can be further validated for their function through rescue experiments and subsequently assayed for mechanisms through similar approaches. Adapted from [133] with permission.*

1.6 Aims and Scope of the Current Study

As detailed in the earlier sections, multi-dimensional approaches have unravelled a plethora of lncRNAs with diverse roles in aspects of development and cancer. Our lab discovered one such lncRNA during characterization of the meiotic recombination hotspot locus of mouse chromosome 8 and hence termed it as Mrhl [134]. Mrhl is a sense, intronic, single-exonic 2.4 kb long lncRNA, located within the 15th intron of *Phkb* and is conserved syntenically in humans. It possesses a poly (A) tail and exhibits tissue specific expression in adult mouse.

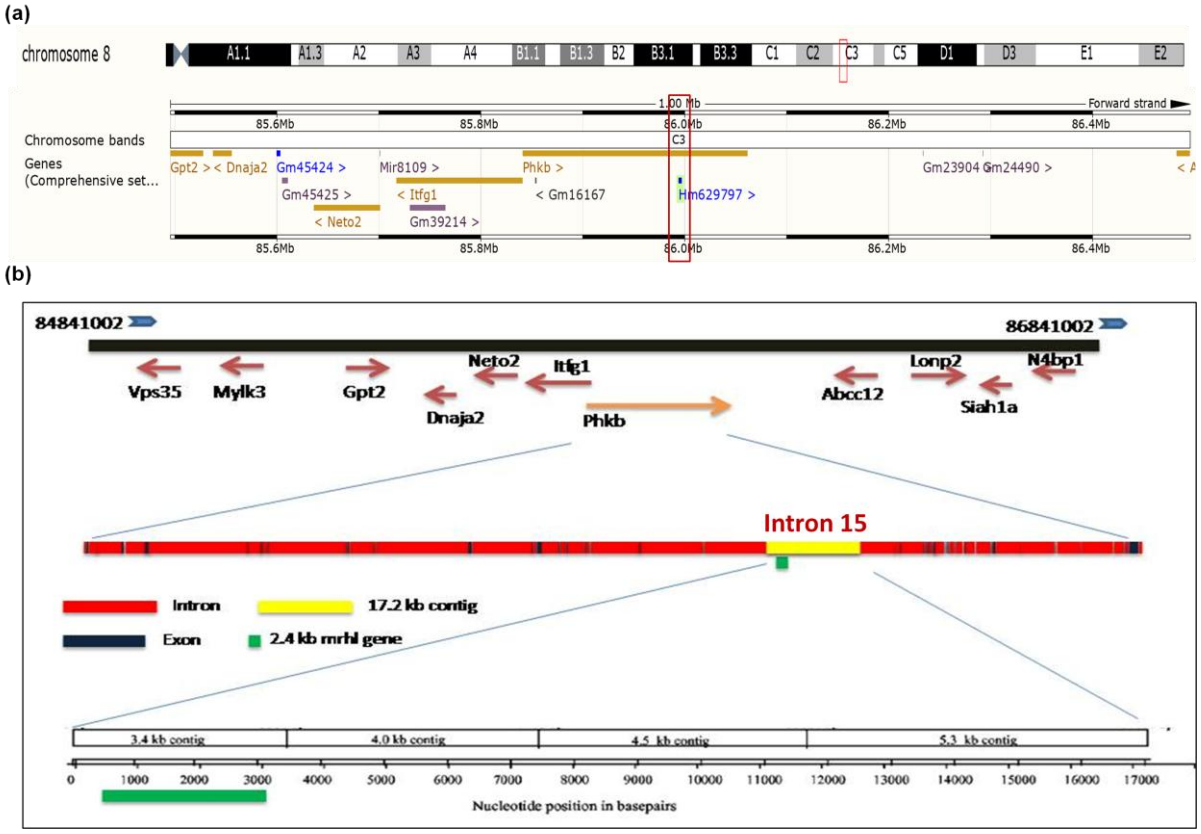


Fig. 1.12 *Genomic location of Mrhl. (a) Annotation of Mrhl in mouse (Hm629797) as per the Ensembl database [135]. (b) Schematic representing location of Mrhl with respect to Phkb and chromosome 8 (schematic by Dr. Vijay Akhade).*

The 2.4 kb transcript is nuclear-localized in mouse spermatogonial progenitor cells (GC1-Spg cell line), exhibits a stable predicted secondary structure and processed into a ~80 nt precursor transcript which is acted upon by the Dicer machinery to generate a smaller 22nt mature miRNA *in vitro*, although the miRNA has not been detected *in vivo* in spermatogonial progenitor cells probably due to context-dependant regulation [136]. Studies were undertaken towards decoding its function in spermatogenesis. It was discovered to be a negative

regulator of WNT signaling pathway in association with p68 as its interaction partner in spermatogonial progenitor cells [64]. Subsequent ChOP-Seq studies revealed genes relevant to the WNT pathway and spermatogenesis and WNT signaling to be its chromatin targets, SOX8 being a prevalent one [117, 137]. It was also shown to be negatively regulated by TCF4/β-CATENIN itself during meiotic commitment of the progenitors [138]. Current studies are being undertaken towards understanding its regulation of SOX8 in the context of spermatogenesis. In purview of the role of lncRNA Mrhl in meiotic commitment of spermatogonial progenitor cells, feedback regulation of a highly conserved context-dependant developmental pathway such as the WNT pathway and the functions of lncRNAs in diverse developmental phenomena (as discussed earlier), we aimed towards exploring Mrhl and its functional relevance in mouse embryonic stem cells and neuronal commitment as our two model systems of study. The rationale behind choosing these two model systems have been discussed in the relevant chapters. Hence, we define the objectives of our current study as follows:

- **Understanding the role of lncRNA Mrhl in mouse embryonic stem cells (mESCs):** Characterization of the RNA and performing an in-depth study of the perturbed genes, pathways and processes upon knockdown of Mrhl in mESCs.
- **Deciphering lncRNA Mrhl as a molecular player in neuronal lineage development:** Characterization of the RNA during neuronal development and decoding its regulators in neuronal progenitors.
- **Gaining insights into the cellular/biological phenotypes upon knockdown/knockout of lncRNA Mrhl:** Generation of stable knockdown lines for Mrhl in mESCs, generation of Mrhl knockout mice and their preliminary characterization.

Chapter 2

Materials and Methods

2.1 Antibodies, Plasmids and Chemicals

2.1.1 Antibodies

Anti-GAPDH (Abeomics, ABM22C5), anti-H3 (Abcam, ab46765), anti β-CATENIN (Abcam), anti p68 (Novus Biologicals), anti-OCT4 (Abcam, ab27985), anti-NESTIN (Abcam, ab11306), anti-TUJ1 (Abcam, ab14545), anti-H3K4me3 (Abcam, ab12209), anti-NICD (Abcam, ab8915), anti-PAX6 (Abcam, ab5790), anti-E-CADHERIN (BD Biosciences, 610182).

2.1.2 Plasmids

Scrambled and Mrhl shRNA plasmids 1, 2, 3 and 4 were custom made from Sigma in the pLKO.1-Puro-CMV-tGFP vector backbone. The sequences of the shRNAs are as follows:

Mrhl shRNA#1: 5'GCACATACATACATACACATATATT 3'

Mrhl shRNA#2: 5'GTGAAATGACTGTGCTTTATT 3'

Mrhl shRNA#3: 5'CAAGTTGACTGCTGATTTATT 3'

Mrhl shRNA#4: 5'GGAGAAACCCTCAAAAGTATT 3'

p3x-FLAG-CMV-10 (Sigma) was used for the generation of Pax6 and Pax6 (5a) expression vectors.

2.1.3 Chemicals

All fine chemicals were obtained from Sigma, Himedia and Emparta, Merck unless otherwise mentioned. For all RNA work, buffers and reagents were either treated with DEPC (Sigma, D5758) or prepared in DEPC-treated water.

2.2 Cell Lines

E14tg2a feeder independent mESC line was a kind gift from Prof. Tapas K. Kundu's lab (JNCASR, India). P19EC line was a kind gift from Prof. Kumar Somasundaram's lab (IISc, India). Additionally, HEK293T cell line (ATCC, U.S.A) was also used. E14tg2a cells were maintained on 0.2% gelatine coated dishes (Himedia, TC041) in mESC medium containing DMEM (Dulbecco's modified Eagles' medium), high glucose (Sigma, D1152), 15% FBS

(Gibco, 16000-044), 1X non-essential amino acids (Sigma, M7145), 0.1 mM β-mercaptoethanol and 1X penicillin-streptomycin (Sigma, P4333). P19EC cells were maintained in DMEM, 10% FBS, 0.1 mM β-mercaptoethanol and 1X penicillin-streptomycin. HEK293T cells were maintained in DMEM, 10%FBS and 1X penicillin-streptomycin.

All transfections for E14tg2a cells were performed using Trans IT-X2 reagent (Mirus, MIR 6003) as per the manufacturer's protocol. Briefly, the DNA and the Trans IT-X2 reagent were diluted in serum-free DMEM, mixed gently yet thoroughly, incubated for 30 minutes at room temperature and added to the cells directly. Cells were harvested at the desired time point for further analysis. Transfections for P19EC and HEK293T cells were performed using Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) as per the manufacturer's protocol. Briefly, appropriate amounts of DNA and lipofectamine were diluted in serum-free DMEM and incubated for 5 minutes at room temperature. The DNA was then added to the lipofectamine and mixed gently yet thoroughly and incubated for an additional 20 minutes at room temperature. The DNA-lipofectamine mix was then added directly to the cells and incubated for the desired amount of time after which cells were harvested and taken forward for suitable analysis.

2.3 Differentiation of mESCs

2.3.1 Embryoid body differentiation

For embryoid body (EB) differentiation, E14tg2a cells were trypsinized using 0.05% trypsin (Sigma, T4799) and $2.5X10^5$ cells were plated onto 35 mm bacteriological grade dishes (Tarsons, 460035) in EB differentiation medium containing DMEM, 10% FBS, 0.1 mM β-mercaptoethanol and 1X penicillin-streptomycin. EBs were harvested at different time points and processed accordingly for further analysis.

2.3.2 Retinoic acid mediated neuronal differentiation

For retinoic acid mediated neuronal differentiation, the protocol by Bibel *et al* [139] was followed. Briefly, $4X10^6$ cells were plated onto 100 mm bacteriological grade dishes (Greiner Bio-One, 633102) in EB differentiation medium and allowed to grow for 4 days. Retinoic acid [(RA), Sigma, R2625] or dimethyl sulfoxide [(DMSO), Sigma, D2650] was added to the EBs on the 4^{th} day at a final concentration of 5 μM. The EBs were allowed to grow for an

additional 4 days. After 8 days of differentiation (4-/4+ RA), the EBs were harvested by gravity sedimentation and subjected to trypsinization with 0.05% freshly made trypsin for 3 minutes at 37°C. Following this, the EBs were dissociated by pipetting and filtered through 40 μm filters (Falcon, 352340) to remove undissociated EBs. Cells were then collected by centrifugation at 1150 rpm for 5 minutes. At this point, cells were either taken for relevant analysis or were further plated for neuron generation. For the latter, appropriate tissue culture grade petri plates or coverslips were coated with 100 μg/ml poly-D-lysine (Merck Millipore, A-003-E) at 37°C overnight. Next day, plates were washed thoroughly three times with deionized sterile water and dried. The plates were then coated with laminin solution (Roche, 11243217001) for 2 hours at 37°C and kept ready for use. The cell pellet obtained above was resuspended in N2 medium containing DMEM/F12 (Sigma, D2906), 1X N2 (Thermo Fisher Scientific, 17502048) and 1X penicillin-streptomycin and plated at a density of 90,000 cells/cm^2. N2 medium was changed 2 hours after plating and then again after 24 hours. For further differentiation, medium was changed after an additional 24 hours to N2B27 medium containing DMEM/F12, 1X N2, 1X B27 (Thermo Fisher Scientific, 17504044) and 1X penicillin-streptomycin. Cells were harvested at different time points for appropriate analyses.

2.4 Cell Culture Assays

For measurement of half-life of Mrhl, E14tg2a cells, at a confluency of ~80-90%, were treated with 10 μM actinomycin D (Sigma, A9415). Cells were harvested at different time points for further analysis.

For NOTCH inhibition studies, EBs were treated with NOTCH inhibitor DAPT **(**N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester, Sigma D5942) at a concentration of 10 μM on the 4^{th} day of differentiation along with RA.

Alkaline phosphatase assay was performed as per the manufacturer's protocol (Sigma, 86R). Briefly, cells were fixed in a citrate-acetone-formaldehyde fixative for 30 seconds at room temperature and rinsed in water for 45 seconds. Alkaline dye mix (napthol AS-BI alakine solution + diazonium salt) was added to the cells and incubated for 15 minutes at room temperature. The cells were then washed in deionized water for 2 minutes and mounted in an appropriate aqueous mounting medium.

2.5 Generation of Stable Lines

Stable knockdown lines for Mrhl were generated in E14tg2a cells as per the protocol of Pijnappel *et al* [140] with some modifications Briefly, viral particles were generated in HEK293T cells by transfection of 5 μg scrambled or Mrhl shRNA plasmids, 2.5 μg pSPAX2, 1.75 μg pVSVG and 0.75μg pRev along with lipofectamine as per the manufacturer's protocol. The media containing viral particles was harvested 48 hours after transfection and centrifuged at 4,000 rpm for 5 minutes to remove cellular debris. This was the first round of viral particles. Fresh mESC medium was added and harvested after an additional 24 hours to collect second round of viral particles. The viral supernatants were stored in -80°C if necessary.

E14tg2a cells were plated such that they reached a confluency of ~60-70% on the day of transduction. The viral supernatant was mixed with 8 μg/ml DEAE-dextran and 1000 units/ml ESGRO (Merck Millipore, ESG1107) and added directly to the E14tg2a cells. If the viral supernatant containing medium was too yellow, then 80% of the supernatant was mixed with 20% of fresh mESC medium. Transduction was performed for 24 hours with the first round of viral particles and an additional 24 hours with the second round of viral particles. The transduced cells were then subjected to puromycin selection (1.5 μg/ml puromycin, Sigma, P8833) for a week.

2.6 Animals

All animal work was performed at JNCASR, Bengaluru in accordance with the procedures and guidelines of the Institutional Animal Ethics Committee (IAEC).

Mrhl knockout mice were generated by Cyagen and handled by Vivo Biotech for delivery. Exon 1 of Mrhl (Hm629797gene; Ensembl: ENSMUSG00000098439; Transcript: Hm629797-001 ENSMUST00000183645) was selected as the targeting region for knockout. To engineer the targeting vector, homology arms were generated by PCR using BAC cloneRP23-60C11 and RP23-429O11 from the C57BL/6J library as template. In the targeting vector, the Neo cassette was flanked by LoxP sites and DTA was used for negative selection The linearized vector was subsequently delivered to ES cells (C57BL/6) via electroporation, followed by drug selection, PCR screening, and Southern Blot confirmation. After gaining 278 drug-resistant clones, 7 potentially targeted clones were confirmed, 5 of which were

expanded for Southern Blotting. The KO mice were then obtained by mating with mice expressing Cre recombinase.

2.7 RNA Fluorescent In Situ Hybridization (FISH) and Immunofluorescence (IF)

RNA FISH followed by IF was performed as per the protocol of de Planell-Saguer *et al* [141] with modifications. The probes used for RNA FISH studies were Cy5 labelled locked nucleic acid probes procured from Exiqon (reported in [64]).

2.7.1 FISH

Fixation and permeabilization: Cells were either grown on coverslips or cell smears were prepared. EBs or brain tissues were fixed in 4% paraformaldehyde for 7-8 hours at 4°C followed by equilibration in 20% sucrose solution overnight and embedded in tissue freezing medium (Leica, 14020108926). The embedded tissues were then cryo-sectioned and collected on Superfrost slides (Fisher Scientific, 12-550-15). For cells, a brief wash was given with 1X PBS (phosphate buffered saline, pH 7.4) followed by fixation with 2% formaldehyde for 10 minutes at room temperature. The cells were then washed with 1X PBS three times for 1 minute each and permeabilization buffer (1X PBS, 0.5% Triton X-100) was added for 5 minutes and incubated at 4°C. The permeabilization buffer was removed and cells were washed briefly with 1X PBS for three times at room temperature. For tissue sections, antigen retrieval was performed by boiling the sections in 0.01M citrate buffer (ph 6) for 10 minutes. The sections were allowed to cool, washed in distilled water three times for 5 minutes each and then in 1X PBS for 5 minutes, each time with gentle shaking.

Hybridization: The samples were then blocked in prehybridization buffer [3% BSA, 4X SSC (saline sodium citrate, pH 7)] for 40 minutes at 50°C. Hybridization (Mrhl probes tagged with Cy5, final concentration 95nM) was performed with prewarmed hybridization buffer (10% dextran sulphate in 4X SSC) for 1 hour at 50°C. After hybridization, slides were washed four times for 6 minutes each with wash buffer I (4X SSC, 0.1% Tween-20) at 50°C followed by two washes with wash buffer II (2X SSC) for 6 minutes each at 50°C. The samples were then washed with wash buffer III (1X SSC) once for 5 minutes at 50°C followed by one wash with 1X PBS at room temperature. For tissue sections, all washes were performed as mentioned above with a time of 4 minutes for buffers I-III. The samples were then processed for IF.

2.7.2 IF

Samples were blocked with IF Blocking buffer (4% BSA, 1X PBS) for 1 hour at 4°C. Primary antibody solution (2% BSA, 1X PBS) was prepared containing the appropriate dilution of desired primary antibody and the samples were incubated in it for 12 hours at 4°C. Next day, the samples were washed with IF wash buffer (0.2% BSA, 1X PBS) three times for 5 minutes each with gentle shaking. The samples were incubated in secondary antibody for 45 minutes at room temperature and washed with 1X PBS three times for 10 minutes each with gentle shaking. The samples were finally mounted in mounting medium containing glycerol and DAPI.

2.7.3 Preparation of cell smears

EBs were trypsinized and filtered as described earlier and cross-linked with 1% formaldehyde for 8 minutes at room temperature. Formaldehyde was quenched using $1/7^{th}$ volume of 1M glycine for 5 minutes at room temperature and cells were collected by centrifugation at 1500 rpm for 5 minutes at 4°C. The cells were washed 3 times with ice cold 1X PBS and collected by centrifugation each time. A 10 µl aliquot of the cell suspension in 1X PBS was taken on a slide, covered with a coverslip and dipped in liquid N_2 to fix the cells on the slide (aliquot volume can be determined empirically to obtain desired cell density per aliquot). The cover slip was removed immediately thereafter with the help of scalpel blade. The slides were hydrated in 1X PBS for 3-5 minutes and dehydrated in graded ethanol from 50%, 75%, 95% to 100%. The slides can be stored in 100% ethanol at this point. Before proceeding for RNA FISH or IF, slides were rehydrated in the reverse order of ethanol series with a final rehydration step in 1X PBS.

2.8 Biochemical Fractionation

2.8.1 Cell fractionation

Approximately 5-10 million cells were lysed using lysis buffer (0.8 M sucrose, 150 mM KCl, 5 mM $MgCl_2$, 6 mM β-mercaptoethanol and 0.5% NP-40) supplemented with 75 units/ml RNAse inhibitor (Thermo Fisher Scientific, EO0381) and 1X mammalian protease inhibitor cocktail (Roche, 04693159001) and centrifuged at 10,000 g for 5 minutes at 4 °C. The supernatant containing cytoplasmic fraction was mixed with 3 volumes of TRIzol for

RNA extraction or with Laemmli buffer for western blotting as described later. The resultant pellet was washed twice with lysis buffer and was subjected to RNA or protein extraction.

2.8.2 Sub-nuclear fractionation

Approximately 10 million cells were lysed with hypotonic lysis buffer (10 mM Tris-HCl ph 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.3% v/v NP-40 and 10% v/v/ glycerol) supplemented with RNAse inhibitor and mammalian protease inhibitor cocktail and centrifuged at 1,000 g for 5 minutes at 4 °C. The supernatant comprising the cytoplasmic fraction was kept aside and the nuclear pellet was washed twice with hypotonic lysis buffer. The nuclear pellet was then resuspended in modified Wuarin-Schibler buffer (10 mM Tris-HCl pH 7.0, 4 mM EDTA, 300 mM NaCl, 1 M urea and 1% NP-40) supplemented with RNAse inhibitor and mammalian protease inhibitor cocktail and vortexed for 10 minutes. Nucleoplasmic and chromatin fractions were separated by centrifugation at 1,000 g for 5 minutes at 4 °C. The chromatin pellet was resuspended in sonication buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 3 mM $MgCl_2$, 0.5 mM PMSF, 75 units/ml RNAse inhibitor) and sonicated for 10 minutes. The chromatin was then obtained as the supernatant following centrifugation at 18,000 g for 10 minutes at 4 °C to remove all debris. The resultant nucleoplasmic and chromatin fractions were then subjected to RNA or protein extraction as described later.

2.9 Immunoprecipitation (IP)

2.9.1 p68 IP

Cells were lysed in hypotonic lysis buffer (10 mM Tris-HCl ph 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.3% NP-40, 10% glycerol) supplemented with RNase inhibitor, mammalian protease inhibitor cocktail and 1mM PMSF. Nuclei were pelleted down at 1200 g for 10 minutes at 4°C and subsequently lysed in nuclear lysis buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5% NP-40) supplemented with RNase inhibitor, mammalian protease inhibitor cocktail and PMSF. The debris was removed by centrifugation at 15,000 g for 10 minutes at 4°C and the supernatant nuclear fraction was collected. To 1 mg of the nuclear fraction containing proteins, 7 µg of either pre-immune serum or p68 antibody was added and incubated overnight at 4°C. Next day, the fraction was incubated with protein A dynabeads for 3 hours at 4°C. The beads were washed with wash buffer (20 mM Tris-HCl pH 7.4, 2 mM $MgCl_2$, 10 mM KCL, 150 mM NaCl, 10% glycerol, 0.2% NP-40) supplemented with RNase inhibitor, mammalian protease inhibitor cocktail and PMSF. Subsequently, the beads were

washed twice with wash buffer (as above with 0.5% NP-40) and collected. The beads were then resuspended either in RNAiso Plus and subjected to RNA isolation for qRT-PCR analysis or in Laemmli buffer and resolved on a 10% SDS-PAGE gel for western blotting analysis as described later.

2.9.2 Chromatin IP (ChIP)

ChIP protocol was performed according to Cotney and Noonan's protocol [142]. Protein A or protein G dynabeads (Thermo Fisher Scientific, 10001D, 10003D) were used for rabbit or mouse antibodies respectively.

Preparation of antibody beads: 25 µl dynabeads were used for each ChIP reaction. The beads were first washed with 1 ml of bead binding buffer (1X PBS, 0.2% Tween-20) and resuspended in 200 µl of the buffer per reaction. H3K4me3 (2 µg) or Pax6 (4 µg) antibodies or their isotype controls were added to the beads in separate tubes and incubated for ~16 hours at 4°C on an end-to-end rotor at 10 rpm. The next day, beads were washed with 1ml of bead binding buffer followed by 1ml of dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl). The beads were then resuspended in 25 µl dilution buffer per ChIP reaction and stored at 4°C for further use.

Chromatin extraction and quantification: Brain tissues from E14.5 dpc embryos or EBs were harvested in serum-free DMEM. Brain samples were minced with a scalpel blade and pipetted a few times whereas EBs were trypsinized and filtered as described earlier. The samples were then subjected to cross-linking using 1% formaldehyde (Sigma, F8775) for 10 minutes at room temperature following which samples were quenched with glycine at a final concentration of 0.125M for 10 minutes at room temperature. The samples were then centrifuged at 2,000 g for 10 minutes at 4°C and washed twice with 1X ice-cold PBS. Finally, the pellets were either flash frozen in liquid N_2 and stored in -80°C or processed for chromatin extraction.

The cross-linked pellet was resuspended in six volumes of ice cold cell lysis buffer (50 mM Tris-HCl pH 8.0, 140 mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) supplemented with 1X mammalian protease inhibitor cocktail and incubated on ice for 20 minutes. For brain samples, the pellets were homogenized once after addition of lysis buffer with 30-40 strokes (BioSpec tissue tearor, 985370) and once after the incubation period was over. The nuclei were then harvested by centrifugation at 2,000 g for 5 minutes at

4°C. The supernatant was removed, the nuclei were resuspended in five volumes of ice-cold nuclear lysis buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 0.5% SDS) supplemented with 1X mammalian protease inhibitor cocktail and incubated on ice for 20 minutes. The samples were then sonicated in Bioruptor (Diagenode, UCD-200) for 35 cycles (at pulses of 30sec on and 30sec off) and centrifuged at 16,000 g for 10 minutes at 4°C to remove insoluble material. The centrifuged samples were then transferred to fresh tubes, aliquoted as per requirement, flash frozen in liquid N_2 and stored in -80°C.

For monitoring sonication efficiency and DNA quantification, 10 µl aliquots was kept aside, diluted in 10 µl of TE buffer and treated with 10 µg of RNaseA (Sigma, R6513) for 30 minutes at 37°C followed by 20 µg of proteinase K (Thermo Fisher Scientific, EO0491) for 1 hour at 55°C . The aliquots were then subjected to reverse cross-linking for 5 minutes at 95°C, allowed to cool to room temperature slowly and analyzed on 1% agarose gel for sonication efficiency or subjected to DNA isolation by phenol-chloroform method for quantification.

Immunoprecipitation: Approximately 10-25 µg of chromatin was diluted with dilution buffer supplemented with 1X mammalian protease inhibitor cocktail to reduce the SDS concentration to <0.1% and achieve a final volume of 450 µl. 5% of the dilution was stored as input at 4°C. 25 µl of antibody or isotype control beads were added to the chromatin dilution and incubated for 12-16 hours at 4°C. Next day, the beads were washed with 1 ml wash buffer (100 mM Tris-HCl pH 8.0, 500 mM LiCl, 1% NP-40, 1% deoxycholic acid) supplemented with 1X mammalian protease inhibitor cocktail for 5 times at room temperature. The beads were given a final wash with 1 ml of TE buffer and resuspended in 85 µl of elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS). Elution was performed twice for 10 minutes each at 65°C under constant agitation in a thermo mixer. All ChIP and input samples (volume made upto 170 µl with dilution buffer for input samples) were subjected to reverse cross-linking for 12 hours at 65°C.

Chromatin purification and analysis: Next day, all samples were treated with 10 µg of RNAse A for 1 hour at 37°C followed by 200 µg proteinase K for 2 hours at 55°C and DNA was extracted by the phenol-chloroform method. The DNA was precipitated with $1/10^{th}$ volume 3M sodium acetate (ph 5.2), 3 volumes of 100% ethanol and glycogen at a final

concentration of 0.5 µg/µl overnight at -20°C. Next day, DNA was pelleted, washed with 75% ethanol, dried, dissolved in sterile deionized water and subjected to qRT-PCR analysis.

2.10 Generation of Plasmid Clones

2.10.1 pGL4 .10-Mrhl promoter constructs

For the Mrhl promoter constructs in luciferase vector backbone, 1.25 kb, 1.5 kb or 3.25 kb of the Mrhl promoter (250 bp were taken downstream of the transcription start site in each case) were amplified from genomic DNA. The 1.25 kb and 1.5 kb amplicons were subjected to restriction digestion with NheI (NEB, R0131S) and XhoI (NEB, R0146S). The 3.25 kb amplicon was subjected to digestion with KpnI (NEB, R0142S) and XhoI. The pGL4.10 (luc2) vector (Promega, E6651) was also subjected to digestion with the corresponding restriction enzymes. All digested DNA were subsequently checked on agarose gel and column purified (Macherey-Nagel, 740609.50). Ligation was performed with T4 DNA ligase (NEB, M0202S) for 2 hours at room temperature. The ligated clones were then transformed into SURE competent cells (Stratagene, 200238) and plated onto LB agar plates with 100 µg/ml ampicillin at 37°C overnight. Next day, the clones were screened by colony PCR. The positive clones were sequence confirmed and purified using columns (Macherey-Nagel, 740412.50) to obtain the pGL4.10-Mrhl 1.25, 1.5 or 3.25 constructs.

2.10.2 p3X-FLAG-CMV-10-Pax6 and -Pax6(5a) constructs

Pax6 and Pax6(5a) were amplified from cDNA of E14.5 mouse embryonic brains. The amplicon and the plasmid were digested with HindIII (NEB, R0104S) and BamHI (R0136S) and subjected to ligation with T4 DNA ligase. The ligated clones were transformed into XL1 Blue competent cells and plated onto LB agar plates containing 100 µg/ml ampicillin. The clones were screened by colony PCR and Pax6 or Pax6 (5a) clones were confirmed by sequencing.

2.11 Electrophoretic Mobility Shift Assay (EMSA)

2.11.1 Purification of PAX6 and PAX6(5A) proteins: p3X-FLAG-CMV-10-Pax6 and -Pax6(5a) constructs were transfected into HEK293T cells. 24 hours after transfection, cells were harvested and lysed in five volumes of FLAG lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) supplemented with 1X mammalian protease inhibitor cocktail on an end-to-end rotor at 15 rpm for 1 hour at 4°C. The samples were then

centrifuged at 12,800 rpm for 30 minutes at 4°C to remove all debris and the supernatant was collected. To 500 µl of supernatant containing proteins, 30 µl M2 agarose beads (Sigma, A220) were added and incubated on an end-to-end rotor for 3 hours at 12 rpm at 4°C. The beads were then washed in 1 ml of FLAG wash buffer (50 mM Tris-HCl ph 7.4, 150 mM NaCl) twice with pipetting followed by centrifugation at 2000 rpm for 3 minutes at 4°C. The FLAG tagged proteins were then eluted in the FLAG wash buffer containing a final concentration of 500 ng/µl of FLAG peptide (Sigma, F4799) on an end-to-end rotor at 12 rpm for 1 hour at 4°C. The beads were separated by centrifugation and the supernatant was collected. Protein concentration was estimated using Bradford reagent (Biorad), an aliquot was taken and tested by western blotting for protein expression.

2.11.2 EMSA

End labelling of probe and purification: Oligos were designed such that they harboured the PAX6/PAX6(5A) motifs on Mrhl promoter. All forward oligos were subjected to end labelling with ɤ-32P-ATP in polynucleotide kinase reaction buffer [PNK buffer in a final concentration of 1X, 40pmol oligo, 20 units PNK enzyme (NEB, M0201S), 30 µCi ɤ-32P-ATP, volume made upto 50 µl with nuclease free water] for 30 minutes at 37°C. The enzyme was heat inactivated for 20 minutes at 65°C. The oligos were purified by the phenol-chloroform method and precipitated in $1/10^{th}$ volume of 3M sodium acetate, ph 5.2, 3 volumes of 100% ethanol and 10 µg yeast tRNA for 1 hour at -80°C. The oligos were pelleted by centrifugation at 12,800 rpm for 10 minutes, washed in 75% ethanol and dried. They were then dissolved in 1X annealing buffer (10 mM Tris-HCl ph 8.0, 20 mM NaCl). To each of the forward oligos, three times excess of the reverse oligos were added, heated for 10 minutes at 95°C and allowed to cool slowly over several hours to overnight at room temperature. The annealed probes were purified using sephadex C-50 columns. Glass wool was packed near the tip of a 1 ml syringe till 0.1 ml and the syringe was packed with C-50 beads slurry till 1 ml with centrifugation at 1800 rpm for 3 minutes at 4°C. A final centrifugation was performed for the packed beads to remove all water before loading the annealed probes onto the column. The annealed probes were collected again by centrifugation and their activity was noted as counts per minute (cpm) in a scintillation counter. Unlabelled forward oligos were also annealed to their corresponding reverse oligos and purified in a similar manner for competition assay.

Binding reaction and electrophoresis: The purified proteins were allowed to bind to the oligos in a reaction mix containing 5X EMSA Buffer (60 mM Hepes-KOH ph 7.9, 300 mM KCl, 15 mM $MgCl_2$, 2.5 mM DTT, 20% w/v Ficoll 400, 1 mg/ml BSA, 1X mammalian protease inhibitor cocktail added just before use) to a final concentration of 1X, PAX6 or PAX6 (5A) purified proteins (100ng to 400ng in increasing concentrations), 400 ng/µl salmon sperm DNA (Sigma, D9156) and ɤ-32P-ATP labelled double stranded oligo (20,000 cpm). The volume was made upto 12.5 µl with nuclease free water. The reaction was incubated for 40 minutes at 37°C. For competition assay, 125X molar excess of unlabelled double stranded probe was used. After incubation, loading dye was added to the samples (containing bromophenol blue, xylene cyanol and without SDS) and the samples were run on a 5% native gel (40% acrylamide stock in 30:08 ratio of acrylamide to bis-acrylamide) prepared in 0.5X TBE (10X TBE stock: 1.8M tris base, 90mM boric acid, 2.5mM EDTA) at 150V at 4°C. The gel was then dried and exposed to X-ray film for 24 hours.

2.12 Luciferase and β-galactosidase Assay

2.12.1 Luciferase assay

P19EC line was used for luciferase assays. Cells were plated such that they reached a density of ~70% on the day of transfection. 2 µg of pGL4-Mrhl promoter constructs were transfected with 200 ng of PAX6 or PAX6(5A) expression plasmids and 50 ng of β-gal plasmid as an internal control. Cells were harvested 24 hours post transfection and luciferase assay was performed as per the manufacturer's protocol (Promega, E2610). Briefly, cells were lysed in five times volume of 1X reporter lysis buffer and vortexed for 10 seconds. The debris was pelleted by centrifugation at 12,000 rpm for 15 seconds at room temeprature and the supernatant was collected. The supernatant and Bright-Glo reagent were mixed in 1:1 ratio and subjected to measurement for luminescence in luminometer (Berthold, Sirius L).

2.12.2 β-galactosidase assay

The β-gal assay as performed as per Uchil *et al* [143]. Briefly, 30 µl of the cell extract was mixed with 3 µl 100X Mg^{+2} solution (0.1 M $MgCl_2$, 4.5 M beta-mercaptoethanol), 66 µl 1X ONPG (4 mg/ml ONPG in 0.1 M phosphate buffer pH 7.5) and 201 µl 0.1 M sodium phosphate, pH 7.5. The reactions were incubated for 30 minutes to a few hours at 37°C until a

faint yellow colour developed. The reactions were stopped by adding 500 µl 1 M Na_2CO_3 and the absorbance was recorded at 420 nm in a spectrophotometer.

2.13 RNA Isolation and PCR

Total RNA was isolated from cells or tissues using TRIzol (Thermo Fisher Scientific, 15596026) for RNA-sequencing and IP or using RNAiso Plus (Takara Bio, 9108) for analysis by qRT-PCR as per the manufacturer's protocol. Briefly, cells or tissues were homogenized in TRIzol or RNAiso Plus using a tissue tearor with 30-40 strokes. The samples were incubated for 5 minutes at room temperature following which 200 µl chloroform was added per 1 ml of homogenized sample, mixed thoroughly and incubated for 2 minutes at room temperature. The samples were then centrifuged at 12,000 g for 15 minutes at 4°C and the aqueous layer was collected. Equal volume of 100% isopropanol was added to the aqueous samples and incubated for 1 hour to overnight at -20°C. The samples were then centrifuged at 12,000 g for 10 minutes at 4°C and the supernatant was discarded without dislodging the RNA pellet. The pellet was then washed in 75% ethanol at 10,000 g for 5 minutes at 4°C, the supernatant was discarded carefully and the pellet was dried for 5 minutes at 37°C. The pellet was then dissolved in an appropriate amount of DEPC-treated water. The RNA samples were then subjected to DNase treatment (NEB, M0303S), re-precipitated in $1/10^{th}$ volume of 3M sodium acetate, pH 5.2 and 3 volumes of 100% ethanol, washed, dried and dissolved appropriately. About 1-3.5 µg of the RNA was taken for cDNA synthesis using oligodT primers (Thermo Fisher Scientific, SO132), RevertAid reverse transcriptase (Thermo Fisher Scientific, EP0442) and RNase inhibitor (Takara Bio, 2313A). The cDNA was diluted 1:1 with nuclease free water and subjected to qRT-PCR using SyBr green mix (Takara, RR820A) in real-time PCR machine (BioRad CFX96). All semi-quantitative PCR was performed in thermal cycler machine (BioRad, Tetrad2) using Taq polymerase (Takara, R001A).

2.14 Western Blotting

Cells or tissues were lysed in five times volume of RIPA buffer **(**150 mM sodium chloride, 1.0% NP-40 or Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA) supplemented with 1 mM PMSF and 1X mammalian protease inhibitor cocktail for 15 minutes on ice with occasional vortexing. Samples were then centrifuged at 12,000 rpm for 10 minutes at 4°C to remove all debris. The supernatant was collected and protein concentration was estimated using Bradford reagent. Suitable

amounts of protein samples were then resolved on 10-12% SDS-PAGE gel and transferred onto nitro cellulose membrane in transfer buffer (25 mM Tris-HCl pH 7.6, 192 mM glycine, 0.03% SDS and 20% methanol). The membrane was blocked in 5% skim milk/1X PBS solution for 1 hour at room temperature or for 4 hours at 4°C. The membrane was then incubated in 1% skim milk/1X PBST (1X PBS, 0.05% Tween-20) solution containing an appropriate dilution of the desired primary antibody overnight at 4°C. Next day, the membrane was washed twice with 1X PBST for 8 minutes each and then incubated in 1% skim milk/1X PBST containing an appropriate dilution of the desired HRP-conjugated secondary antibody for 1 hour at room temperature or 4 hours at 4°C. The membrane wsas washed twice in 1X PBST for 8 minutes each and analyzed using chemiluminescence (Millipore, Luminata Forte, WBLUF0100) and exposure to X-ray flims.

2.15 Systems Analysis

2.15.1 RNA-Seq

E14tg2a cells treated with scrambled or Mrhl shRNA (shRNA 4) were subjected to RNA isolation and quality check. RNA samples were then subjected to library preparation in duplicates and sequenced on Illumina Hi-Seq 2500 platform. mm10 Genome was downloaded from GENCODE and indexed using Bowtie2-build with default parameters. Adapter ligation was done using Trim Galore (v 0.4.4) and each of the raw Fastq files were passed through a quality check using the FastQC. PCR duplicates were removed using the Samtools 1.3.1 with the help of ‘rmdup’ option. Each of the raw files were then aligned to mm10 genome assembly using TopHat with default parameters for paired-end sequencing as described in [144]. After aligning, quantification of transcripts was performed using Cufflinks and then Cuffmerge was used to create merged transcriptome annotation. Finally differentially expressed genes were identified using Cuffdiff. The threshold for DE genes was log2 (fold change) >1.5 for up regulated genes and log2 (fold change) <1.5 for down regulated genes. The DE genes were analyzed further using R CummeRbund package.

2.15.2 GO enrichment analysis

Gene Ontology (GO) analysis was performed in PANTHER [145]. Significant enrichment test was performed with the set of differentially expressed genes in PANTHER and Bonferroni correction method was applied to get the best result of significantly enriched biological processes.

2.15.3 Fisher's exact test

Fisher's exact test was performed in PANTHER Gene Ontology (GO) where p-value significance was calculated based on the ratio of obtained number of genes to the expected number of genes (O/E) considering the total number of genes for the respective pathway in *Mus musculus*.

2.15.4 Cluster analysis

Hierarchical clustering method was performed using Cluster 3.0 [146]. Gene expression data (FPKM of all samples i.e, scrambled and shRNA treated) was taken and log2 transformed. Low expressed (FPKM<0.05) and invariant genes were removed. Then genes were centered and clustering was performed based on differential expression pattern of genes and fold change. Genes were grouped in 9 clusters and visualized as a network in Cytoscape [147]. Functional enrichment of each cluster was performed using the Gene Mania Tool [148].

2.15.5 TF network analysis

Motifs were downloaded for all transcription factors from JASPAR [149] and sequence of interest for each TF (1.5 kb upstream & 500bp downstream of TSS) was extracted using BedtoFasta of the Bedtools suite [150] . Then each motif was scanned across the sequence of all TFs to create the table matrix that reflects the number of binding sites for each TF across the other TFs using MEME suite [151] with a e-value of 1E-04. Finally the heatmap was generated from the table matrix using R 3.3.2. TFs were fed into STRING [152] to visualize the known interactions from the experimental data and hierarchy was setup manually as per the interaction among given TFs (proteins).

2.15.6 FIMO analysis

3 kb promoter sequence of Mrhl (upstream from its transcription start site or TSS) was taken along with the motifs for PAX6 binding as described in a report by Xie and Cvekl [153]. FIMO (Find Individual Motif Occurrences) [154] scanning was then performed with a default cutoff value of 1E-04.

2.16 Primer and Oligo Sequences

2.16.1 qRT-PCR primers

Gene Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
Oct4	ACCACCATCTGTCGCTTC	CCACATCCTTCTCTAGCC
Sox2	GAGTGGAAACTTTTGTCCGAGA	GAAGCGTGTACTTATCCTTCTTCAT
Nanog	AGGGTCTGCTACTGAGATGCTCTG	CAACCACTGGTTTTTCTGCCACCG
Fgf5	GTAGCGCGACGTTTTCTTCG	AATTTGGCTTAACACACTGGC
Gata3	GGCTACGGTGCAGAGGTATC	GATGGACGTCTTGGAGAAGG
T	CAGCCCACCTACTGGCTCTA	CCCCTTCATACATCGGAGAA
β-globin	CTCAAGGAGACCTTTGCTCA	AGTCCCCATGGAGTCAAAGA
α-fetoprotein	CCTGTGAACTCTGGTATCAG	GCTCACACCAAAGCGTCAAC
Nestin	GTCGCTTAGAGGTGCAGCAG	TTCCAGGATCTGAGCGATCT
Pax6	TCCAGGTGCTGGACAATGAA	GCTTTTCGCTAGCCAGGTTG
Ascl1	TTGCGGCTGCTTTCCTTTTC	CTGCGAAGCACGATCAAAGG
Tuj1	TGAGGCCTCCTCTCACAAGT	CCAGCACCACTCTGACCAAA
GFAP	CACGAACGAGTCCCTAGAGC	GTGGCCTTCTGACACGGATT
Hes5	GAGATGCTCAGTCCCAAGGAG	GCGAAGGCTTTGCTGTGTTT
Mrhl	TGAGGACCATGGCTGGACTCT	AGATGCAGTTTCCAATGTCCAAAT
FoxG1	GGCAAGGGCAACTACTGGAT	CGTGGTCCCGTTGTAACTCA
En1	TCCTACTCATGGGTTCGGCT	CTGGAACTCCGCCTTGAGTC
Gbx2	CAACTTCGACAAAGCCGAGG	TGACTCGTCTTTCCCTTGCC
β-actin	AGGTCATCACTATTGGCAACG	TACTCCTGCTTGCTGATCCAC
GAPDH	GGGAAATGAGAGAGGCCCAG	TACGGCCAAATCCGTTCACA
U1snRNA	CTTACCTGGCAGGGGAGAT	CAGTCCCCCACTACCACAA
SRA	TCCACCTCCTTCAAGTAAGC	GACCTCAGTCACATGGTCAACC

2.16.2 ChIP primers

Gene Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
Mrhl Pax6 ChIP Site 1	TCCACCTTACTTGCCTTGGA	CAGTCATTTGGGAAGGTTGCT

Mrhl Pax6 ChIP Site 2	GGTCTTCTTTCTCTTGAGTTTAATGTG	GGGTTGCTATCTCTGCCACT
Mrhl Pax6 ChIP Site 3+4	TTCATGTGTCTGTGTATACCTCTTCA	TTTCCTTCCTTAAACTGTTTCTCTG
Mrhl Proximal Promoter (Histone ChIP)	GAACTCCAGCCTATCTGATGGT	GCGATTATCTGGTAAAGCACACC
Actin ChIP	CACCCATCGCCAAAACTCTTCAT	CGCACAGTGCAGCATTTTTTTAC

2.16.3 Cloning primers

Gene Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
Mrhl 1.25kb_NheI	CTAGCTAGCTAGAAACTG AATGCCCCAACTCTC	
Mrhl 1.5kb_NheI	CTAGCTAGCTAGCTCTT GAGTTTAATGTGTCATAC	
Mrhl 3.25kb_KpnI	TGGGTACCCCCTCAACCT GTTCCTTTAGCCTG	
Mrhl Cloning_XhoI		CCGCTCGAGCGGGCTTCTACA GGTGTTTCAAAAGA
Mouse Pax6 Cloning	CCCAAGCTTATGCAGAACAGTCA CAGCGGAG	GGGGTACCTTACTGTAATCGAGGC CAGTACTG

2.16.4 Genotyping primers

Hm629797_F1: GGCTCACAGATTAAGGGTACTGTT

Hm629797_R1: TTTCTCCCAGAACAGCTTCAGGAA

Hm629797_R2: GAGTATGGTCAGCTGCTATT

2.16.5 EMSA oligos (5' to 3'):

Site 1:

Probe 1_F: CTGTAATCTTAAATTATGCATCCTTCCGTGCACCTGACAC

Probe 1_R: GTGTCAGGTGCACGGAAGGATGCATAATTTAAGATTACAG

Site 2:

Probe 2_F: AGAAAAAATCATCAGTCTAGGAGAAAATGATAAATTTACT

Probe2_R: AGTAAATTTATCATTTTCTCCTAGACTGATGATTTTTTCT

Site 3:

Probe 3_F: CATTATCTTTCTAAAATTCTTCATGTGTCTGTGTATACCTC

Probe 3_R: GAGGTATACACAGACACATGAAGAATTTTAGAAAGATAATG

Site 4:

Probe 4_F: TCTCCTTAGCAAGAGCAGTGGTTACTTTTTGTGTTTTTAT

Probe 4_R: ATAAAAACACAAAAAGTAACCACTGCTCTTGCTAAGGAGA

Positive control for Pax6:

P6CON_F: GGATGCAATTTCACGCATGAGTGCCTCGAGGGATCCACGTCGA

P6CON_R: TCGACGTGGATCCCTCGAGGCACTCATGCGTGAAATTGCATCC

Chapter 3

Understanding the Role of LncRNA Mrhl in Mouse Embryonic Stem Cells

3.1 Introduction

3.1.1 Mouse embryonic stem cells: discovery and characteristics

Post fertilization in mouse, the embryo undergoes three rounds of cell division to reach the eight-cell stage post which the cells undergo compaction and develop apical-basal polarity giving rise to the morula by day E2.0-E2.5. Two rounds of asymmetric cell division then gives rise to the early blastocysts by E3.5 resulting in the sorting of the embryonic cells into 'outside' and 'inside' cells. The outside cells give rise to the extraembryonic trophectoderm (TE) whilst the inside cells form the pluripotent inner cell mass (ICM) (reviewed in [155]). Embryonic stem (ES) cells are derived from the ICM of pre-implantation blastocyst stage embryos (**Fig. 3.1**) and were first reported to be isolated and maintained in culture by Evans and Kaufman in one report [156] and by GR Martin in an independent report [157]. ES cells are characterized by their capacity to proliferate and self-renew indefinitely as well as differentiate to give rise to any cell type pertaining to the three germ layers i.e., ectoderm, endoderm and mesoderm.

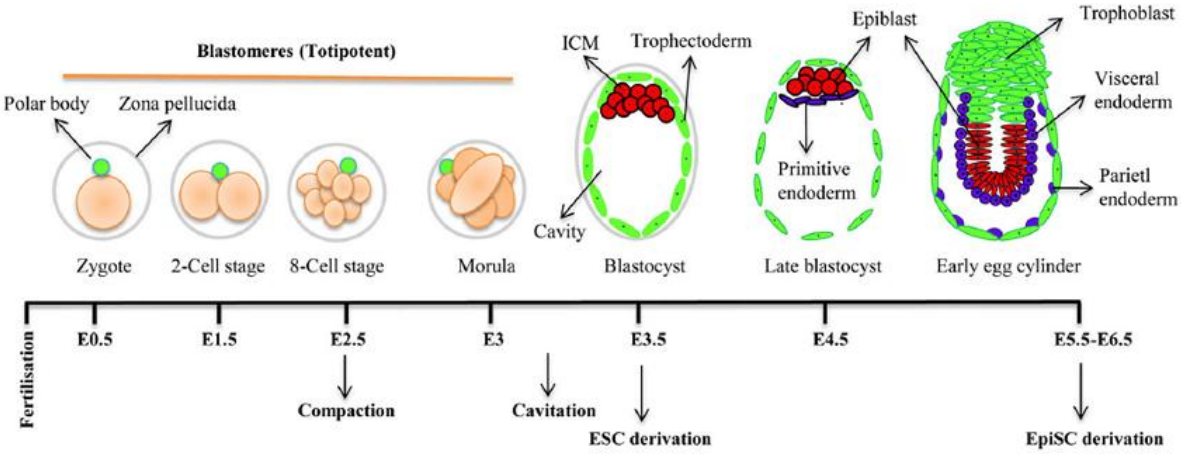


Fig 3.1 *Derivation of ESCs. The zygote undergoes divisions to give rise to the morula. The morula undergoes cavitation, compaction and divisions to form the blastocysts (E3.5). The inner cell mass (ICM) of the blastocysts stage embryo is the origin for ESCs. At E5.5-6.5, the embryo forms the egg cylinder where germ layer specification takes place. Epiblast stem cells (EpiSCs) are derived from this stage. Adapted from [158] with permission.*

3.1.2 Pathways and mechanisms regulating pluripotency/differentiation of ES cells

I. Transcription factors

The discovery of induced pluripotent stem cells (iPSCs) independently by James Thomson [159] and Shinya Yamanaka [160] led to the identification of the core pluripotency transcription factors i.e., OCT3/4 (POU5F1), SOX2, c-MYC and KLF4, now referred to as the Yamanaka factors. Across mouse, rat and human ES cells, a common set of TFs (OCT3/4, SOX2 and NANOG) act to regulate the pluripotency of the cells. OCT4 deficient embryos survive until the morula stage but fail to form the ICM *in vivo* and ES cell pluripotent colonies *in vitro*. It is actually expressed throughout the pre-implantation stages of the developing embryo and it reappears in the germ cell progenitors. The levels of OCT4 is however very crucial to regulate the 'stemness' properties of ES cells. Overexpression of OCT4 has been shown to differentiate ES cells into the primitive endoderm and mesoderm lineages. Also, OCT4 is not exclusively expressed in the pluripotent epiblast cells of the developing embryo and hence is not the sole determining factor of pluripotency, although it is one of the major ones (reviewed in [158]).

SOX2 belongs to the Sox family of transcription factors which contain a highly conserved HMG (high mobility group) DNA-binding domain. Deletion of SOX2 *in vivo* causes death of embryos immediately post-implantation. *In vitro*, depletion of SOX2 results in the differentiation of ES cells into the trophoblast lineage. This phenotype is similar to that of OCT4 since they often act in concert as a heterodimer to regulate other genes such as *Fgf4*, *Nanog*, *Lefty1* along with *Oct4* and *Sox2* themselves. However again, the expression of SOX2 is not restricted only to the pluripotent mass of cells, with it being expressed in neural as well as extra-embryonic tissues (reviewed in [158]).

NANOG is a homeodomain containing transcription factor that acts in concert with OCT4 and SOX2 to maintain the ES cell properties. *In vivo*, NANOG null embryos give rise to pluripotent cells but differentiate into extra-embryonic endoderm lineage immediately after. The same observation is made *in vitro* probably because NANOG acts to negatively regulate the primitive endoderm transcription factor GATA6. The levels of NANOG expression varies considerably in ES cells that contributes to population heterogeneity. Also, ES cells from NANOG null embryos have been successfully derived as well as iPSCs from NANOG null somatic cells, suggesting that NANOG is not essential to maintain stem cell properties of ES cells. However, NANOG over expressing ES cells have been shown to stabilize the

undifferentiated state of ES cells allowing them to propagate independent of feeder layers or small molecule growth factors (reviewed in [158, 161].

c-MYC belongs to the MYC family of basic helix-loop-helix leucine zipper transcription factors and has been shown to be a direct target of LIF/STAT3 signaling pathway which is well-known for maintaining ES cell pluripotency. MYC is highly stable in ES cells and is required for sustaining canonical PI3K signaling pathway and suppressing GSK-3β activity, preventing early differentiation of ES cells. Coordinated loss of c-MYC and n-MYC leads to loss of self-renewal and pluripotency as well as differentiation of ES cells towards primitive endoderm and mesoderm lineages, GATA6 being one of the negatively regulated target genes of MYC. MYC has also been shown to coordinate with BMP-4 signaling to repress MAPK pathways since activation of the MAPK cascade by fibroblast growth factors acts as a differentiation cue for ES cells (reviewed in [162], [163]). Furthermore, it has been shown that deletion of c-MYC and n-MYC or their pharmacological inhibition arrests ES cells or blastocysts in a reversible dormant phase, suggesting that MYC controls the biosynthetic machinery of ES cells [164].

KLF4 belongs to the Kruppel-like factor family of conserved zinc finger transcription factors. Along with OCT4, SOX2 and c-MYC, KLF4 reprograms adult fibroblasts into induced pluripotent stem cells. Studies have shown that knockdown of Klf4 induces differentiation of ES cells whereas over expression maintains the pluripotent self-renewing state of the cells and abolishes the requirement of LIF (leukemia inhibitory factor) in *in vitro* cultures. Interestingly, over expression of NANOG rescues KLF4 knockdown mediated differentiation phenotype of ES cells suggesting that KLF4 acts upstream of NANOG. KLF4 has, in fact, been shown to bind to the proximal and distal elements of NANOG promoter to regulate its expression and has also been shown to form a protein-protein complex with OCT4 and SOX2. KLF4 has been predicted to be a direct downstream target of the LIF-STAT3 signaling pathway and in the absence of KLF4, OCT4 and SOX2 bind to the NANOG promoter but are unable to maintain the pluripotent status of ES cells, suggesting KFL4 as a crucial regulator of the pluripotency pathway (reviewed in [158], [165]).

II. Signaling pathways

LIF/JAK/STAT3 pathway: LIF belongs to the IL-6 family of cytokines that regulate a wide variety of physiological processes pertaining to immune responses and development. LIF binds to the LIF receptor which forms a heteromeric complex with an IL-6 family co-receptor

subunit gp130. Upon binding, LIF activates the Janus kinase family of tyrosine kinases (JAKs). JAKs phosphorylate the tyrosine residues on the cytoplasmic tail of the receptor which form the docking site for STAT3 in the cytoplasm. STAT3 then undergoes phosphorylation itself by JAKs and forms dimers via its SH2 domain. Following dimerization, it translocates to the nucleus to activate target gene transcription. STAT3 has been shown to be a critical regulator of self-renewal capacity of ES cells. A dominant negative form of STAT3 (where the tyrosine residue is replaced with a phenylalanine residue) impairs self-renewal properties of ES cells even in the presence of LIF [166]. An inducible active form of STAT3 when expressed in ES cells showed that ES cells could maintain their properties without LIF [167]. LIF triggers three major signaling pathways, the JAK/STAT3 pathway, the PI3K/AKT pathway and the SHP2/MAPK pathway but the JAK/STAT3 pathway is the one that is necessary for the maintenance of ESCs. Amongst the target genes of the JAK/STAT3 pathway, c-MYC is one of the most important ones as it is actively involved in maintaining the pluripotent and undifferentiated state of ES cells. MYC and STAT3 have also been shown to occupy the promoters of ES cell enriched factors such as KLF4 and TFCP211. It has been further shown that depletion of TFCP211 is able to impair the self-renewal and reprogramming properties of ES cells conferred by STAT3 activation. Interestingly, TFCP211 integrates into the core pluripotency network by activating NANOG and TBX3 and its own activity is further enhanced by OCT4, SOX2 and ESSRB, providing a crucial link between the LIF signaling pathway and the core pluripotency pathway (reviewed in [158]).

WNT/β-CATENIN signaling pathway: In attempts to better establish ground state growth conditions for ES cells, the 3i/2i culture conditions were discovered [168]. This culture condition consists of N2B27 medium supplemented with inhibitors against GSK3β and FGF/ERK signaling pathway. The 3i/2i media abolishes the necessity of LIF for the maintenance of ES cells. Inhibition of GSK3β supports ES cells by stabilizing cytoplasmic β-CATENIN, one of the main components of the canonical WNT pathway. In the absence of WNT ligand, GSK3β forms a destruction complex with APC, CK1 and AXIN and phosphorylates β-CATENIN for ubiquitination and proteasome-mediated degradation. Upon binding of WNT to its receptor Frizzled and coreceptor LRP5/6, the destruction complex formation is inhibited and β-CATENIN translocates to the nucleus where it interacts with the TCF/LEF family of transcription factors to activate its target set of genes. Additionally, β-CATENIN also localizes to the cell membrane along with E-CADHERIN and α-CATENIN

and plays a key role in the formation of adherent junctions to regulate cell adhesion properties. Addition of the GSK3β inhibitor CHIR99021 in the 3i/2i media has been shown to maintain self-renewal properties of ES cells suggesting that β-CATENIN is essential for the maintenance of ES cells. In fact, over expression of β-CATENIN has been shown to mimic the effects of CHIR99021 in mouse and rat ES cells. Mechanistically, β-CATENIN has been reported to mediate the degradation of TCF3 leading to abrogation of TCF3 mediated transcriptional repression of important pluripotent associated genes such as *Oct4*, *Nanog*, *Tfcp211* and *Essrb* (reviewed in [158]). However, excess activity of β-CATENIN caused by over inhibition of GSK3 β leads to the differentiation of ES cells since β-CATENIN/LEF1 complex is known to up regulate lineage specific genes like *Cdx2* and *T* [169].

FGF/MEK/ERK signaling pathway: FGF4 signaling is active throughout the various stages of early mouse embryonic development from the one-celled stage to the primitive streak stage. In blastocysts, FGF4 is produced by the inner cell mass where it is required for growth and proliferation but interestingly, FGF signaling has been implicated in the proper formation and differentiation of the trophectoderm and the primitive endoderm. In ES cells, FGF signaling is active although FGF null ES cells do not show any phenotypic defects. However, such cells are resistant towards differentiation into neural and mesodermal lineages (reviewed in [158]). In a genome-wide siRNA screen, it was found that many of the differentiation associated genes such as *Gmnn*, *Psmb3* and *Ifna14* contribute to ERK activation by down regulating MAP kinase phosphatases such as DUSP1 and/or DUSP6 which are in turn inhibitors of the FGF/MEK/ERK pathway [170]. Furthermore, FGF/ERK has been shown to phosphorylate STAT3 at Ser^{727} to prime mouse ESCs for commitment towards the neural lineage [171] and inhibit the activity of Klf4 by binding to its activation domain and phosphorylating it at Ser^{123}. Thus FGF/MEK/ERK pathway is necessary for acting as a cue for exit of ES cells from the pluripotent state.

TGF-β/SMAD signaling pathway: This pathway is divided into the TGF-β/ACTIVIN/NODAL pathway and the BMP/GDF/MIS pathway. TGF- β signaling pathway is activated by ligand binding to the Type I and Type II transmembrane receptors. Following ligand binding, the serine/threonine kinase activity at the intracellular domains of these receptors gets activated and they phosphorylate the SMAD proteins. SMADs 1,2,3,5 and 8 form homotrimers and interact with SMAD4 wherein they activate their target genes in the nucleus. BMP signaling has been shown to maintain ES cell characteristics by suppressing neural differentiation. It acts in conjunction with LIF signaling because LIF pathway prevents

differentiation of ES cells into mesoderm and endoderm lineages. Under serum free culture conditions, LIF promoted neural differentiation of ES cells but the addition of BMP4 suppressed it and maintained the undifferentiated state of the cells in such conditions. BMP4 also induces the expression of inhibitor of differentiation (Id) factor which is a suppressor of proneural basic helix-loop-helix transcription factors. Furthermore, mouse embryonic fibroblast feeder layers secrete BMP4 wherein it inhibits ERK phosphorylation through DUSP9 activation. DUSP9 over expression was shown to substitute BMP4 in the maintenance of ES cells. However, BMP4 alone induces mesodermal differentiation of mouse ES cells suggesting that a fine balance between LIF and BMP signaling pathways are crucial towards maintaining the undifferentiated state of mouse ES cells [reviewed in [158, 161].

Phosphatidyl inositol 3 kinase (PI3) pathway: PI3 kinases catalyze the phosphorylation of their phospholipid substrate inositol at the third carbon position. They are categorized into three major classes depending on their substrate specificity, amino acid sequence and homology of their lipid kinase domains. The class 1A PI3 kinases consist of heterodimers of regulatory and catalytic subunits. They are activated by various receptor tyrosine kinases that bind to a spectrum of growth factors such as FGF, EGF and PDGF. Upon activation, PI3 kinases phosporylate PIP2 to generate the second messenger PIP3 (phosphatidyl inositol 3,4,5-tris-phosphate). AKT1 binds to PIP3, is translocated to the inner cell membrane and is activated by the serine threonine kinase PDK1. Activation of AKT1 in turn leads to the activation of various cellular pathways that regulate phenomena such as cellular survival and proliferation. It was shown that the treatment of ES cells with LY294002, a PI3K inhibitor, prohibited the progression of cells from G1 to S phase thereby decreasing cellular proliferation [172]. Targeted disruption of PTEN, a negative regulator of PI3K pathway, led to an increase in ES cell proliferation and tumorigenicity [173] suggesting that the PI3K pathway is crucial for maintaining the proliferative state of ES cells.

RAS/RAF/ERK pathway: RAS proteins are a part of a superfamily of low molecular weight GTP binding proteins that switch between an active GTP bound state and an inactive GDP bound state with the help of GTP exchange factors. They have been implicated in cellular proliferation and differentiation. Binding of growth factors to receptor tyrosine kinases leads to their activation via autophosphorylation which then form the binding sites for SHP2 and GRB2 which in turn activate RAS through the GTP exchange factor SOS. Activated RAS binds to RAF serine/threonine kinases which activate ERK pathway that regulates genes

involved in cell proliferation such as *c-Jun*, *c-Fos*, *Ets* and *Elk*. RAS/RAF/ERK pathways have, however, been implicated in promoting differentiation of ES cells. ERK inhibitor PD98059 has been shown to promote the efficient derivation of ES cells from blastocysts. On the other hand, *Grb2*-null mice resist differentiation into endoderm lineage whilst the ectopic expression of HRas in ES cells causes them to differentiate massively. Hence, this pathway is involved in supporting differentiation of ES cells (reviewed in [161]).

Thus, a multitude of pathways and transcription factors act in a coordinated network to regulate ES cell characteristics and differentiation.

3.1.3 Lineage differentiation of ES cells

ES cells have long been an appealing model system of study for development owing to their ability to differentiate into diverse variety of cell types belonging to any of the three germ layers. Such procedures of differentiation enable the study of the molecular pathways and players involved in the formation and specification of differentiated cell types. The three major ways to differentiate ES cells include growing them in suspension as three dimensional spheroids known as embryoid bodies (EBs), co-culture with stromal cells and withdrawal of LIF (**Fig. 3.2 a**).

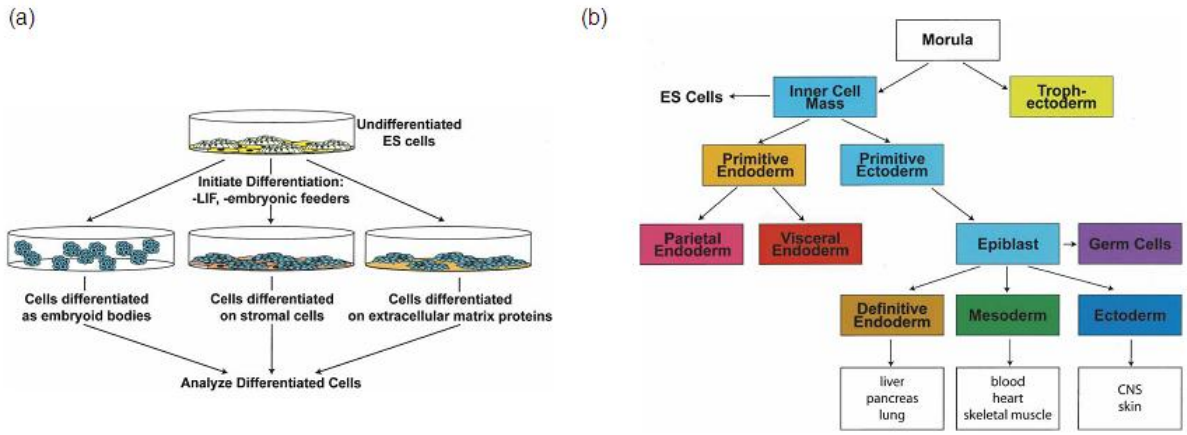


Fig. 3.2 *Lineage differentiation of mouse ES cells. (a) Different methods to induce differentiation of ES cells; (b) Inner cell mass gives rise to cell types of all lineages. ES cells are derived from the ICM. Adapted from [174] (access under CC BY-NC 4.0).*

Ectoderm lineage: Cells of the nervous system and the skin are derived from the ectoderm lineage which in turn arises from the epiblast at the anterior portion of the of the developing mouse embryo (**Fig. 3.2 b**). Mouse ES cells have been routinely cultured under various

neural induction cues such as retinoic acid treatment of EBs or serum free culture as monolayers to generate neural progenitors that can generate neurons or astrocytes under appropriate differentiation signals [139, 175, 176]. Alternatively, specialized protocols guiding the differentiation of mouse ES cells into specific types of neurons such as glutamatergic neurons, dopaminergic neurons, precursors of cortical interneurons or spinal cord motor neurons have also been developed [177-180]. Whilst BMP, WNT and ACTIVIN/NODAL signaling pathways have been implicated to suppress neural differentiation, NOTCH signaling has been established to play crucial roles in the formation of cells of the neural lineage (reviewed in [181]). Interestingly, addition of BMP4 to neuroectoderm cultures promotes the formation of the epidermal lineage comprising keratinocytes. Mouse ES cells grown on an extracellular matrix substrate comprising collagen in the presence of BMP4 differentiated into functional keratinocytes and were shown to form an equivalent of a stratified epithelium [182]. Hence, ES cells can be differentiated into both neural and epidermal cells types of the ectoderm lineage.

Endoderm lineage: Studies have suggested that during the formation of the primitive streak at the gastrulation stage of embryos, a population of cells positive for T and FOXA2 arise representing the mesendodermal progenitors. Whereas ACTIVIN/NODAL signaling facilitates the formation of these common progenitors, sustained ACTIVIN signaling leads to the development of the definitive endoderm. Furthermore, FGF and BMP4 pathways induce the formation of hepatic specification in ACTIVIN induced endoderm whereas retinoic acid and Sonic Hedgehog lead to the formation of pancreatic cell types [181, 183]. Mouse ES cells grown as monolayers in the presence of retinoic acid, bFGF and dibutyryl cAMP were shown to give rise to definitive endoderm lineage that had the potential to form pancreatic precursors [184]. Embryoid bodies cultured for 2 weeks and then treated with hepatotrophic factors have also been shown to generate hepatocytes [185]. Further, BMP4 along with ACTIVIN and NODAL has been shown to generate NKX2.1 positive lung progenitor cells in embryoid bodies [186]. Thus through a combination of factors, ES cells can be differentiated into cells of the endoderm and more specifically, the hepatic, pancreatic or lung lineages (**Fig.3.2 b**).

Mesoderm lineage: The mesoderm lineage in the early mouse embryo gives rise to the cardiovascular and skeletal systems during development (**Fig. 3.2 b**). Flk-1^{+} mesodermal populations generate cells of the haematopoietic lineage upon induction with BMP signaling factors (reviewed in [181]). However, a combination of WNT, ACTIVIN and BMP signaling pathways has been shown to facilitate the formation of the haematopoietic mesoderm

whereas inhibition of WNT signaling has been implicated to promote the development of the cardiac lineage at later stages. Co-culture of mouse ES cells with OP9 cells lacking the capability to generate M-CSF (macrophage colony stimulating factor) [187] or growing the ES cells in conditioned medium from HepG2 cells have successfully led to the formation of precursors of the haematopoietic lineage [188]. Reports have also focussed on the generation of cardiac progenitors, cardiomyocytes and engineered cardiac tissues from mouse ES cells either from embryoid bodies spontaneously or with the help of defined growth factors and culture conditions [189-191].

Germ cell lineage: In the developing mouse embryo, primordial germ cells (PGCs) are first distinguished at the gastrulation stages in the extraembryonic allantois mesoderm. BMP4 and -8b factors have been demonstrated to be necessary for the specification of PGCs from the embryonic epiblast and theses observations have been recapitulated in *in vitro* cultures from ES cells [192]. In a study by Kimura *et al*, ES cells were induced to form mesoderm by co-cultue with OP9 cells and then led to generate PGCs by ERK signaling inhibition [193]. However, other reports have established the derivation of PGC-like cells (PGCLC) by treatment of ES cell derived embryoid bodies with BMP4 and WNT3 or by over expression of germ cell specific genes such as *Dazl* and *Mvh* (reviewed in [194]). In two reports by Hayashi *et al*, a combination of ACTIVIN, BMP, FGF and LIF signaling pathways were used to obtain PGCLCs through the formation of epiblast like stem cells (**Fig. 3.3**). These PGCLCs were shown to be capable of integrating into the germ line and give rise to functional differentiated sperms and oocytes after *in vivo* transplantation [195, 196].

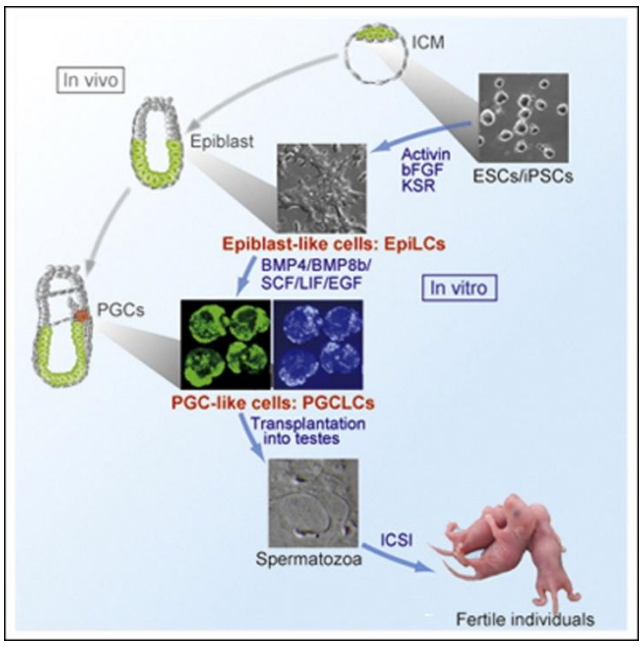


Fig. 3.3 *Differentiation of ES cells into germ line lineage cells. ES cells are first differentiated into EpiLCs through a combination of signaling factors followed by the formation of PGCLCs which can give rise to mature sperm (or oocytes) in vivo. Adapted from [196] with permission.*

Along with serving as a model system to study the molecular players of development and differentiation, ES cell derived cell types of various lineages possess the potential to be developed as targets for clinical and therapeutic purposes. In this context, it becomes highly important to decode and understand the regulatory pathways and mechanisms operating in ES cells for their functioning.

3.1.4 Long non-coding RNAs in ES cell pluripotency and lineage specification

I. LncRNAs in pluripotency network in ES cells

Over the past few decades, lncRNAs have been extensively implicated in regulating diverse physiological phenomena with their roles in stem cell circuitry, cell commitment and fate specification being of paramount importance due to the applications of stem cells in regenerative medicine. One of the genome-wide studies in mouse ES cells led to the identification of a class of lncRNAs i.e., lincRNAs to have crucial roles in maintaining pluripotency of stem cells [197]. Out of 226 lincRNAs, knockdown of 26 of them in stem cells were shown to affect ES cell morphology and reduction in the expression of the core pluripotency factors like OCT4 and NANOG, suggesting an exit from pluripotency state of ES cells in the absence of these lincRNAs. Furthermore, it was observed that ~12% of all ES cell lincRNAs and ~50% of all lincRNAs regulating pluripotency were bound by one or the other core pluripotency transcription factors. This established that lincRNAs are intricately involved in the entire ES cell transcriptional network. Additionally, it was found that knockdown of 30 of the lincRNAs triggered the ES cells to express specific markers corresponding to the ectoderm, endoderm or mesodermal lineages, suggesting that these lincRNAs indirectly maintain pluripotency by repressing lineage-specific genes.

An earlier study addressed the importance of lncRNAs during ES cell differentiation into embryoid bodies which possess cell types representing all the three lineages [198]. About 945 ncRNAs were identified in this study, with 174 of them showing differential expression and correlation with either pluripotency state or specific lineage states. Specifically, two of the lncRNAs, Evx1as and Hoxb5/6as were up regulated along with their genomically associated protein-coding genes during the primitive streak differentiation phase of embryoid bodies. Both of these lncRNAs showed colocalized expression with EVX1 and HOXB6 respectively in mouse embryos. It was also observed that these lncRNAs co-precipitated with the chromatin modifier MLL1, suggesting their probable function in directing MLL1 activity at

its target loci which includes several *Hox* genes and other developmentally important genes in human and mouse cells.

Studies directed towards the identification of individual lncRNAs in stem cell biology led to the discovery of lncRNA AK028326 which is a part of the annotated lncRNA Gomafu/Miat (myocardial infarction associated transcript) [199]. Depletion of lncRNA AK028326/Gomafu resulted in loss of ES cell morphology and stem cell characteristics. This lncRNA has been implicated in regulating the pluripotency of mouse embryonic stem cells through an atuofeedback loop in association with OCT4. Its knockdown in ES cells not only causes reduction in expression of the core pluripotency factors but also an up regulation of trophoblast specific markers such as CDX2, HAND1, EOMES and GATA3, suggesting its dual role in maintaining pluripotency directly through the core transcription factor network and also as a repressor of the trophoblast lineage. Interestingly, Gomafu has also been shown to be involved in repressing osteogenic lineage program in stem cells [111] but contrarily shown to be associated with oligodendrocyte specification [200]. However, over expression of AK028326 in ES cells led to an increase in the expression levels of ectodermal lineage markers and reduction in mesodermal and primitive streak ones [199], suggesting that lncRNA AK028326 by itself can alter the differentiation status of ES cells, although it is a part of the Gomafu transcriptional unit. LncRNA Tuna (TCL1 upstream neuron associated lincRNA) has been shown to play a role in the formation of neural precursors from stem cells and its depletion inhibits differentiation of stem cells into the neural lineage [201]. However, Tuna has also been shown to maintain pluripotency of stem cells as its loss of function alters ES cell morphology and proliferation. Tuna forms a multi-protein complex with RNA binding proteins PTBP1, hnRNP-K and NCL which occupy promoters of *Nanog*, *Sox2* and *Fgf4* to maintain the pluripotent nature of stem cells.

In human embryonic stem cells, lncRNA Gas5 has been shown to promote self-renewal in a positive feedback loop with the core pluripotency factors OCT4 and SOX2 [105]. It also attenuates miRNAs that target the TGFβ receptor family ligand NODAL which also leads to maintenance of the self-renewing state of human ES cells. In a screen for lincRNAs in human induced pluripotent cells (iPSCs), lincST8SIA3 was identified as crucial lncRNA involved in the formation and maintenance of iPSCs and was termed as linc-RoR (regulator of reprogramming) [101]. In an independent study, linc-RoR was shown to be acting as a miRNA sponge to decoy miRNAs binding to core pluripotency factors, OCT4, SOX2 and

NANOG, as their targets, thereby helping in the maintenance of the pluripotent state of human ES cells [202].

II. LncRNAs in lineage commitment

Neural lineage: LncRNAs have been largely implicated to perform crucial regulatory roles in multipotent or unipotent stem cells or in the differentiation of embryonic stem cells into specific lineages. LncRNA Paupar (PAX6 upstream antisense RNA) has been shown to play a role in maintaining the self-renewal properties of neuroblastoma cells, which are neural progenitors [203]. Its depletion from the cell led to an increase in neurite growth and increase in neuronal differentiation in the cells. Paupar was shown to regulate around 942 genes in neural progenitors relating to synaptic control and cell cycle regulation and further studies revealed that it co-regulated a specific subset of genes along with PAX6, underlying its importance in neural progenitors. LncRNA Pnky has been shown to maintain a specific class of $Sox2^+$ neural stem cells in the mouse embryonic forebrain [204]. It was shown that Pnky interacts with PTBP1, a splicing regulator and a repressor of neuronal differentiation to regulate a common set of target genes and splice variants. On the other hand, lncRNA Dali (DNMT1 associated long intergenic RNA), encoded downstream of POU3F3, a protein known for its role in the development of the nervous system, was shown to be important for the differentiation of neural progenitors through knockdown studies in neuroblastoma cells [205]. Genomic studies showed that Dali regulates genes such as *E2f2*, *Fam5b*, *Sparc* and *Dkk1* that encode pro-differentiation factors and negatively regulates genes that prevent the formation of neurites. Dali was shown to regulate neuronal specification of progenitors by regulating *Pou3f3* in *cis*, several neuronal differentiation, neuronal projection and cell cycle genes in *trans* and methylation status at *Dlgap5*, *Hmgb2* and *Nos1* genes by interacting with DNMT1, a methyl transferase.

Hematopoeitic lineage: In the hematopoietic lineage, lincRNA-Eps (lincRNA erythroid prosurvival) has been shown to be involved in the maintenance of erythroid progenitors by regulating their proliferation [206]. It represses *Pycard*, a pro-apoptotic gene that activates caspase during apoptosis hence, preventing undue apoptosis in the progenitors. LncRNA Ego (eosinophil granule ontogeny) was found to be over expressed upon stimulation of umbilical cord blood cells or bone marrow cells ($CD34^+$) with IL-5 and only slightly induced in the presence of other cytokines like epoeitin-α, SCF, GM-CSF etc [207]. Depletion of lncRNA Ego from $CD34^+$ umbilical cord blood cells resulted in their death within 5 days of growth in

IL-5 medium and reduced expression of the eosinophil proteins MBP and EDN suggesting the importance of lncRNA Ego in the proper development of eosinophil lineage of cells. Similarly, lncRNA Hotairm1 (HOXA transcript antisense RNA myeloid-specific 1) has been implicated to be regulated by the transcription factor PU.1 in the process of development of myeloid cells [208, 209].

Muscle lineage: A host of lncRNAs have been shown to play important regulatory roles in the maintenance or differentiation of muscle stem cells. LncRNA Munc (MYOD upstream non-coding, also known as DRR^{eRNA}) was identified in a screen for lncRNAs during the differentiation of myoblasts to myotubes [210]. Depletion of lncRNA Munc in myoblasts led to reduction of MYOD association to the DRR enhancer and to the promoter of *Myogenin*, leading to impairment of proper differentiation of myoblasts. These results were replicated *in vivo* as well, with knockdown of MUNC displaying defects in muscle regeneration. Linc-MD1 has also been shown to be associated with myogenesis [32]. By acting as a ceRNA for miR-133, it acts to prevent degradation of the microRNA targets MEF2C and MAML-1, both of which are important as transcription factors during myogenesis. A recently identified lncRNA with established roles in myogenic differentiation is Lnc-Mg (myogenesis-associated lncRNA) [211]. It acts as a ceRNA for miR-125b and in turn regulates the abundance of insulin like growth factor 2. Its over expression in *in vitro* cultures improves cellular differentiation whereas conditional knockout mice display muscle atrophy and loss of endurance during exercise, highlighting the importance of lncRNAs in cell fate specification.

Epidermal lineage: Ancr (antidifferentiation non-coding RNA) was reported to be present in epidermal progenitor cells and underwent significant reduction in expression levels upon differentiation [212]. In a regenerated organotypic culture system recapitulating the human epidermis, the levels of Ancr were found to be reduced, with even the basal layer expressing differentiation genes. On the contrary, lncRNA Tincr (terminal differentiation induced non-coding RNA) was found to be expressed in organotypic cultures and its depletion led to perturbations in a number of differentiation genes [213]. Tincr was shown to act through STAU1 mediated stabilization of several key mRNAs encoding structural and regulatory proteins necessary for keratinocyte differentiation. Both Ancr and Tincr play crucial roles in the maintenance or differentiation of adult stem cell population of the epidermis.

Male germ cell lineage: LncRNAs have also been widely implicated to have important regulatory roles in the progenitor cells of the neonatal testes or during later stages of

differentiation. LncRNA Mrhl has been shown to regulate WNT signalling negatively in mouse spermatogonial cells. Depletion of Mrhl in spermatogonial cells, led to the activation of WNT signalling as a primary effect through p68 mediated nuclear translocation of the WNT signalling effector β-CATENIN [64]. Genome-wide chromatin occupancy studies revealed that Mrhl occupies and regulates key genes involved in spermatogenesis, Sox8 being a predominant one [117]. Interestingly, WNT and Mrhl act in a negative feedback loop through the recruitment of CTBP1 as the corepressor at the Mrhl locus, to regulate the meiotic commitment phenomenon of male germ cells [138]. LncRNA 033862 has also been shown to be involved in the self-renewal and maintenance of spermatogonial stem cells (SSC) in neonatal testes and in *in vitro* cultures [72]. Knockdown of this lncRNA from SSC cultures led to a reduction in colony size and morphology and decrease in expression levels of key stem cell and proliferation genes such as *Bcl6b*, *Ccnd2* and *Pou5f1*. Chromatin isolation by RNA purification experiments revealed that lncRNA 033862 physically occupies and regulates the expression of *Gfra1* (GDNF family receptor alpha 1), a gene known to be important for GDNF (glial cell line derived neurotrophic factor) signalling that is required for the self-renewal capacity of spermatogonial stem cells. On the contrary, lncRNA Tsx (testis specific X-linked) has been shown to have roles in later stages of spermatogenesis mainly, in the pachytene spermatocytes [214].

The list of lncRNAs involved in coordinating and regulating cell fate decisions or stem cell programming is exhaustive, with the above being a few representative examples. They act in concert with proteins, ranging from chromatin modifiers to enzymes to transcription factors, to fine-tune diverse phenomena with respect to development and differentiation.

3.2 Rationale of Current Study

Earlier studies from our lab have characterized lncRNA Mrhl extensively in the context of male germ cell commitment and spermatogenesis. In one of the earliest studies it was revealed that Mrhl negatively regulates WNT signaling in B type spermatogonial progenitor cells (Gc1-Spg cell line) in conjunction with its interacting partner p68 [64]. It acts to sequester p68 in the nucleus and keeps the WNT signaling in check in the progenitor cells. Upon down regulation of Mrhl, p68 translocates to the cytoplasm and stabilizes β-CATENIN therein allowing its transport to the nucleus whereby it binds to TCF4 at the target genes leading to their activation. Genome-wide chromatin occupancy studies have shown that Mrhl physically occupies 1370 loci on the chromatin and directly regulates 37 of them that have

been referred to as the GRPAM loci (Genes Regulated by the Physical Association with Mrhl) [117]. Many of the GRPAM loci belong to WNT signaling and spermatogenesis pathways. Furthermore, Mrhl was shown to be required for p68 occupancy at majority of these GRPAM loci. Interestingly, WNT has been shown to in turn negatively regulate Mrhl during later stages of meiotic commitment of spermatogonial progenitors [138]. These studies unravelled a negative feedback loop between lncRNA Mrhl and the WNT pathway involved in a chromatin-mediated fine regulation of progenitor sustenance and meiotic differentiation during spermatogenesis.

The role of lncRNAs in embryonic stem cell network and lineage differentiation/commitment has been already discussed in detail. The role of WNT signaling in mouse ES cells has been outlined whereas its role in development in invertebrates and vertebrates has been widely established. In vertebrates, it has been shown to be involved in the establishment of the dorsoventral and anteroposterior axes during very early stages of embryonic development. During gastrulation, WNT signaling is however inhibited in the anterior portion and is activated in the posterior region of the embryo. A loss in this tight control of WNT expression leads to axial duplications. During later stages of embryonic development, WNT has been shown to be expressed in various organ systems such as the central nervous system, skeletal system and somite formation emphasizing its context-dependant regulatory functions during vertebrate development. In the adult tissues, WNT signaling has been extensively implicated in proliferative and homeostatic roles in the skin, intestine and bone (reviewed in [215-217]).

In the context of the roles of lncRNAs and WNT signaling in development and differentiation, we sought to address lncRNA Mrhl as a player in mouse embryonic development and cell fate specification using mouse ES cells (mESCs) as our model system of study.

(All materials and methods have been detailed in Chapter 2).

3.3 Results

3.3.1 Mrhl is a nuclear-localized lncRNA, is associated with the chromatin and exhibits moderate stability in mESCs: We analyzed poly (A) RNA-Seq datasets from the ENCODE database and observed that Mrhl is expressed predominantly and at higher levels in the embryonic stages of tissues of various lineages. In the brain, heart and lung, Mrhl is

expressed all throughout different stages of embryos whereas its expression is almost null postnatally (**Fig. 3.4 a-e**). However, in the liver and the kidney, Mrhl expression is down regulated but not completely abrogated in the postnatal stages (**Fig. 3.4 f, g**). From E8.5 onwards, the mouse embryo undergoes a surge of differentiation, cell specification and organogenesis phenomena. Our data analysis suggested that Mrhl might have a selective role to play in these processes in the context of mouse embryonic development. To further address this role, we selected mESCs as our model system of study. In the blastocyst stage around E3.5-4.5, the embryo harbours two distinct cell populations i.e., the ICM and the TE. Whilst the TE gives rise to the extra-embryonic tissues, the ICM is the origin for all cell types of various tissues. mESCs derived from the ICM hence serve as an excellent system to understand molecular events related to development.

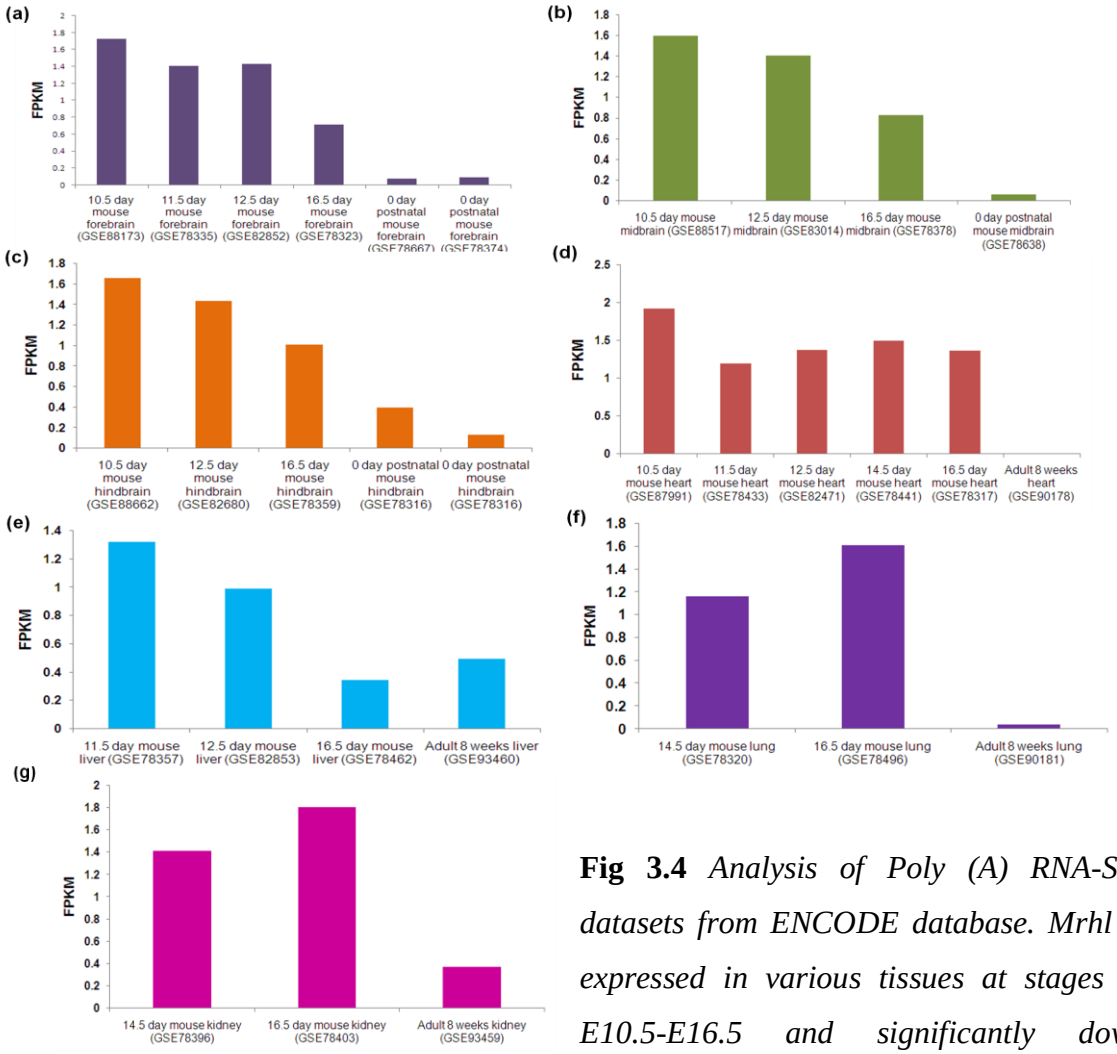


Fig 3.4 *Analysis of Poly (A) RNA-Seq datasets from ENCODE database. Mrhl is expressed in various tissues at stages of E10.5-E16.5 and significantly down regulated in postnatal stages or adult stages. (a-c) Fore-, mid-, hindbrains; (d) heart; (e) liver; (f) lung; (g) kidney.*

To understand the role of Mrhl in mESCs, we first studied its localization, chromatin association and stability because it would shed light onto its functional relevance in mESCs. RNA FISH studies revealed that Mrhl is largely localized in the nucleus and this was confirmed with the help of biochemical fractionation studies (**Fig. 3.5 a, b**). In purview of the role of Mrhl as a chromatin regulatory lncRNA in spermatogonial progenitors, we further fractionated the nuclear extract into chromatin and nucleoplasm fractions and observed that Mrhl predominantly localizes to the chromatin in mESCs (**Fig. 3.5 c**). To understand whether Mrhl is indeed bound to the chromatin, H3 ChIP was performed. We observed that Mrhl was significantly pulled down along with H3 suggesting that it is physically associated with the chromatin (**Fig. 3.5 d**). Furthermore, we also performed an analysis for the stability of Mrhl is mESCs and observed that Mrhl exhibits moderate stability with a half-life of 2.73 hours (**Fig. 3.5 e**). The above studies revealed that Mrhl is stably expressed in mESCs with a predominant association to the chromatin whereas the *in vivo* data analysis suggested its involvement in cellular differentiation. Hence, we were intrigued to further address the role of Mrhl in mESCs.

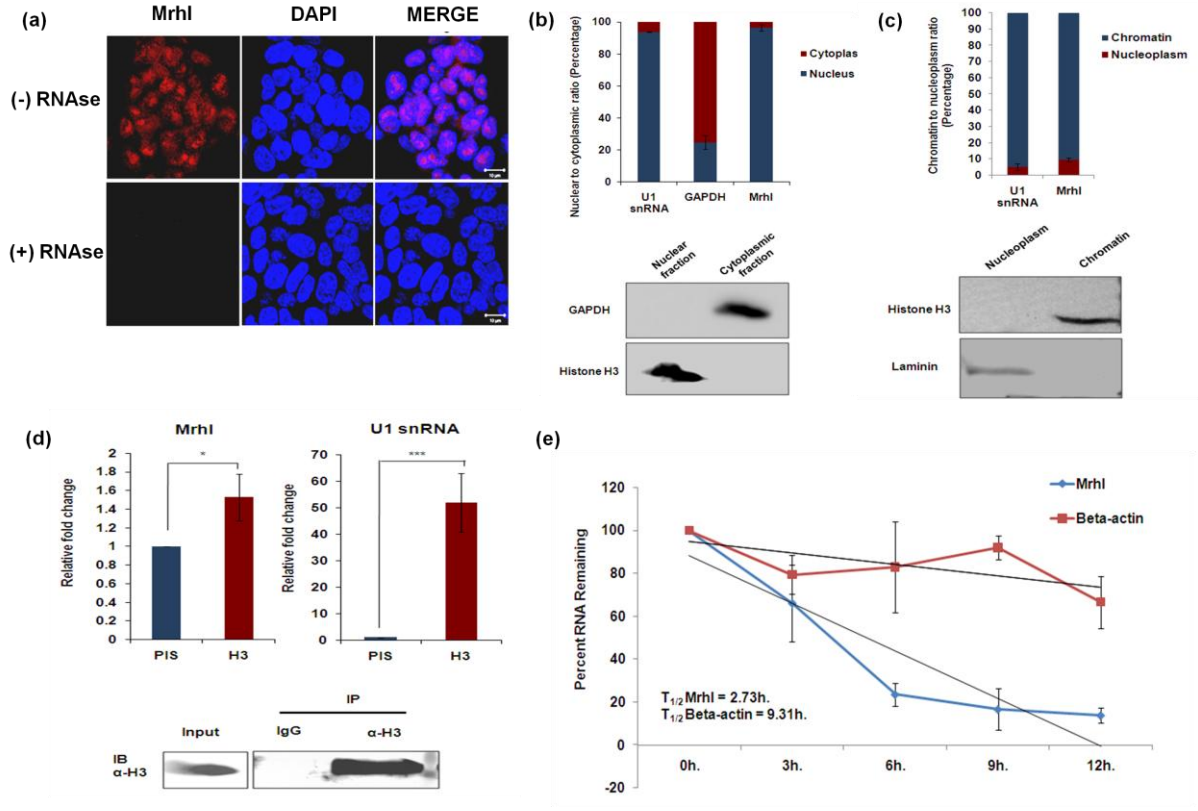


Fig. 3.5 *Mrhl is a nuclear localized chromatin bound stable lncRNA in mESCs. (a) RNA FISH studies showed Mrhl to be predominantly localized to the nucleus in mESCs. (b) Nuclear cytoplasmic fractionation to confirm nuclear localization of Mrhl; western blot*

shows purity of fractions. (c) Nucleoplasmic chromatin fractionation shows predominant localization of Mrhl in the chromatin fraction; western blot shows purity of fractions. (d) H3 ChIP confirmed chromatin association of Mrhl; western blots shows validation for H3 pulldown. (e) Actinomycin D studies revealed Mrhl to have a half-life of 2.73 hours. (*Figs. 1b, 1c and 1d have been adapted from M.S. thesis, Neha C.V.).*

*Error bars indicate standard deviation from three independent experiments. *p<0.05, **p<0.01, ***p<0.001, student's t-test. Scale bar = 10μm.*

3.3.2 Mrhl levels are perturbed upon differentiation of mESCs: Our observations with respect to the prevalent expression of Mrhl in various stages of mouse embryonic organogenesis prompted us to understand its status under conditions of differentiation of mESCs. We undertook the embryoid body method of differentiation since EBs have been widely characterized with respect to their differentiation into all three germ layers in a time dependant manner. During the first 2-4 days of differentiation, EBs recapitulate early events corresponding to gastrulation of mouse embryos. From day 6 and onwards, EBs give rise to specified cells and structures recapitulating organogenesis events and thereafter [218-220]. We differentiated mESCs over a time course of 8 days and observed that the expression levels of Mrhl are perturbed with a preferential increased expression during days 4 and 6 of EB differentiation (**Fig. 3.6 a**). An analysis of the expression of various lineage related markers suggested that during days 4-6 of EB formation, neuroectoderm and mesodermal lineages are formed predominantly (**Fig. 3.6 a**) with earlier reports indicating the formation of early endoderm lineage as well. Whilst it still remains a challenge to strictly correlate *in vitro* differentiation with *in vivo* embryonic development, our results here reveal that Mrhl might be involved in the early stages of specification of various lineages as well as in the later stages of organogenesis in mouse embryonic development. The results here intrigued us to understand the role of Mrhl in mESCs per se.

3.3.3 Knockdown of Mrhl in mESCs does not perturb canonical WNT signaling pathway: The role of canonical WNT signaling in mESCs and mouse early embryonic development is highly context-dependant and often a fine balance is necessary to regulate a pluripotent/proliferative state versus a differentiated state. Earlier studies have revealed a negative feedback loop operating between Mrhl and the canonical WNT pathway in the context of male germ cell differentiation in conjunction with an interaction partner p68. In purview of these reports, we sought to find out whether Mrhl functions through the WNT

pathway in mESCs. We performed shRNA mediated knockdown of Mrhl in mESCs and scored for the status of the canonical WNT effector β-CATENIN which is known to translocate to the nucleus upon WNT pathway activation in spermatogonial progenitors. RNA FISH confirmed our down regulation of Mrhl in shRNA treated cells. However, we did not observe any change in the localization of β-CATENIN (**Fig. 3.6 b**). Furthermore, p68 immunopulldown showed that Mrhl does not interact with p68 in the context of mESCs (**Fig. 3.6 c**). These results suggested that Mrhl probably functions through different pathways and targets depending on the context although its basic mechanisms such as regulation through association with protein partners or chromatin mediated regulation of the target genes might still be the same. The above observations intrigued us to further address the regulatory pathways of Mrhl in mESCs.

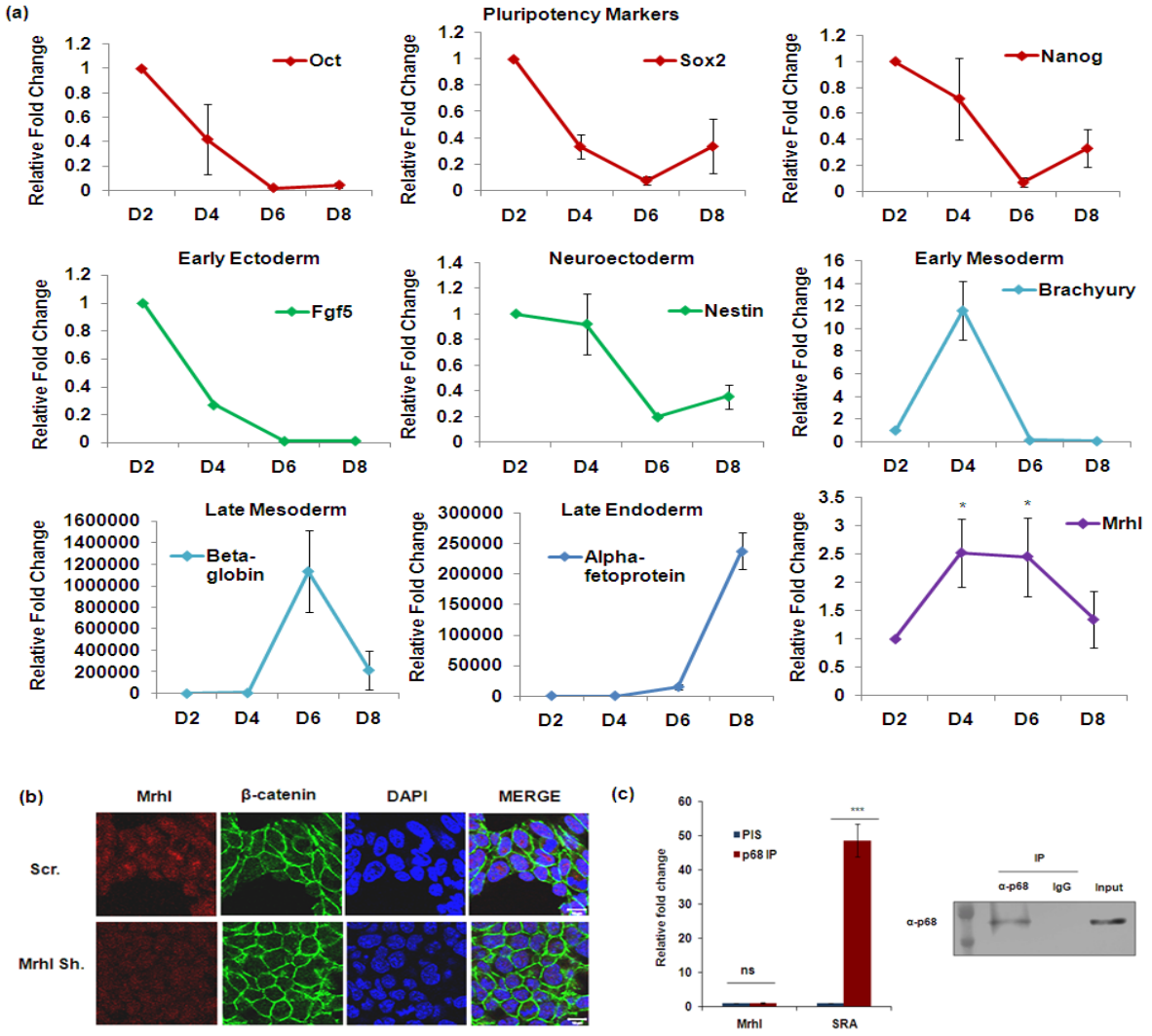


Fig. 3.6 *Mrhl in context of differentiation of mESCs and WNT signaling. (a) Mrhl levels are up regulated during days 4 and 6 of embryoid body differentiation; various lineage markers*

demonstrate EB differentiation. (b) Nuclear localization of β-CATENIN in scrambled versus Mrhl shRNA cells revealed that Mrhl does not function through the WNT pathway in mESCs. (c) p68 immunopulldown showed Mrhl does not interact with p68 in the context of mESCs.

*Error bars indicate standard deviation from three independent experiments. *p<0.05, **p<0.01, ***p<0.001, student's t-test. Scale bar = 10μm.*

3.3.4 Upon depletion of Mrhl in mESCs, genes belonging to developmental processes are significantly perturbed: To understand the role of Mrhl in mESCs, we carried out shRNA mediated transient knockdown of Mrhl in these cells using four independent shRNA constructs i.e., sh.1, sh.2 sh.3 and sh.4 to rule out off-target effects. We obtained a statistically significant down regulation of ~50% using shRNAs 1 and 4 (**Fig. 3.7 a**). Interestingly, a qRT-PCR analysis of the pluripotent markers Oct4, Sox2 and Nanog revealed no significant differences in the expression levels between scr. (scrambled) and sh. treated cells (**Fig. 3.7 b**). We then subjected scr. and sh.4 treated mESCs to RNA-sequencing on Illumina Hi-Seq 2500 platform in duplicates. The results were analyzed as has been detailed in the methodology section. To further validate our experiment, we analyzed poly (A) RNA-Seq datasets for mESCs from the ENCODE database and observed that Mrhl displayed similar FPKM values in earlier reports as well. Furthermore, the FPKM values between sh.4 and scr. treated cells confirmed Mrhl down regulation levels (**Fig. 3.7 c**). An analysis of the differentially expressed genes (DEG) revealed 1143 genes to be perturbed in expression upon Mrhl knockdown in mESCs comprising of 729 down regulated and 414 up regulated (**Annexure 1**). A classification of the DEG into coding or non-coding categories revealed that Mrhl knockdown majorly affected protein coding genes along with a substantial number of lincRNAs. Few of the DEG were also those coding for transcription factors, antisense RNAs and miRNAs (**Fig. 3.7 d**). Gene ontology (GO) showed us a diverse variety of molecular functions and biological processes to be affected in the Mrhl silenced mESCs (**Fig. 3.7 e**). Majority of the DEG belonged to molecular functions such as binding (23.6%), catalytic activity (18.3%), receptor activity (8.8%) and signal transducer activity (7.4%) whereas in the biological processes they majorly belonged to categories such as cellular processes (40.9%), biological regulation (15.8%), metabolic process (25.7%), developmental process (11.7%) and multicellular organismal process (10.7%). To assess amongst the DEG if a subset showed a statistical over representation over the others, a GO enrichment analysis was performed with a p-value<0.05. The enrichment analysis revealed positive regulation of developmental

processes and positive regulation of multicellular organismal processes as the top two hits with respect to the number of perturbed genes and the p-value (**Fig. 3.7 f**).

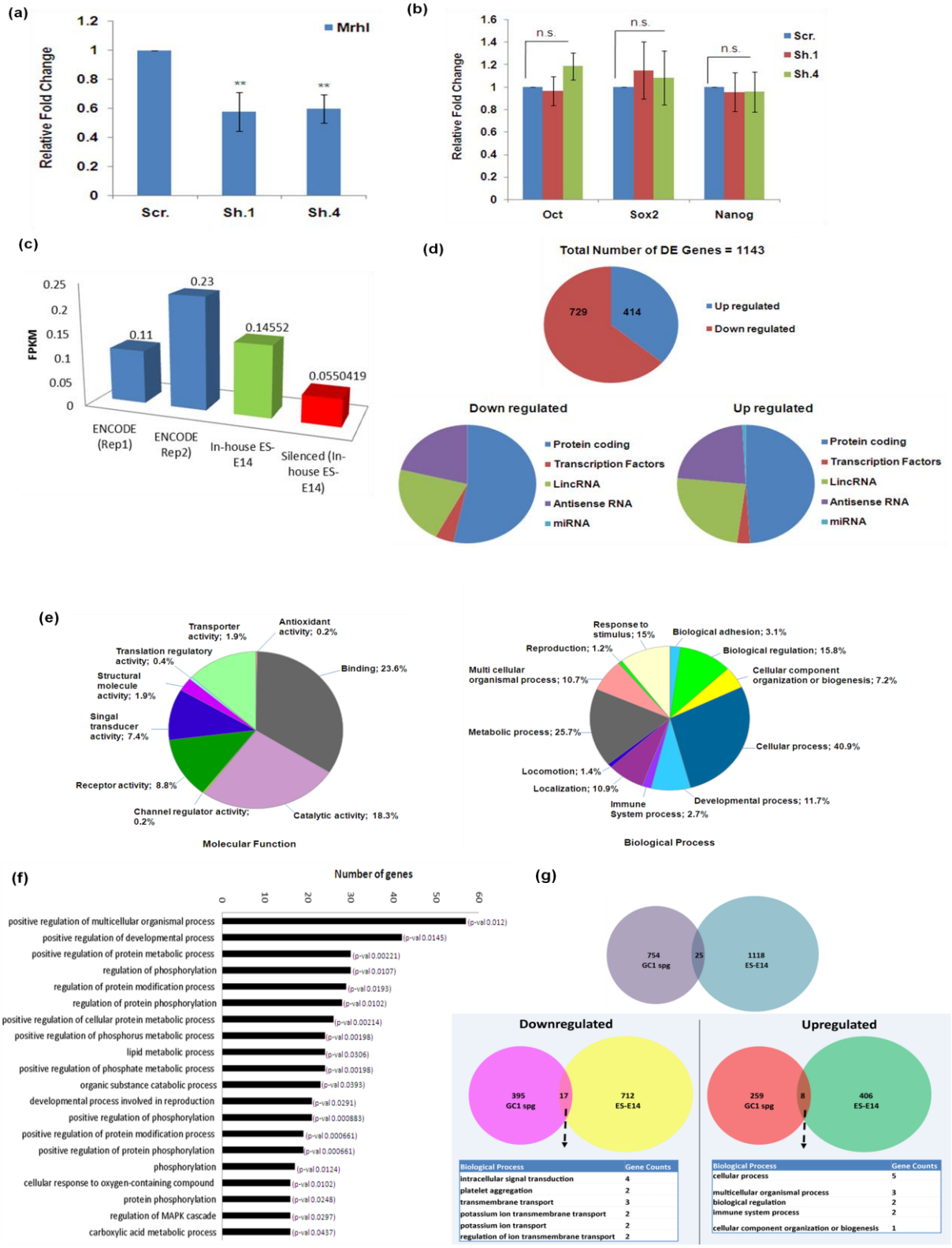


Fig. 3.7 *RNA-Seq and gene ontology analysis in Mrhl knockdown versus control mESCs.*

(a) Knockdown efficiency as scored by qRT-PCR in Mrhl shRNA 1 and 4 (sh. 1 and sh.4) treated cells as compared to scrambled shRNA (scr.) treated cells. (b) Assay for pluripotency markers in Mrhl knockdown cells. (c) Comparison of FPKM values from publicly available ENCODE datasets and in-house RNA-Seq analysis. (d) Representation of differentially expressed genes (DEG). (e) Gene ontology (GO) analysis for perturbed molecular functions and biological processes in Panther. (f) GO enrichment analysis (p<0.0.5) for DEG in Panther. (g) Comparison of DEG between mESCs and Gc1-Spg spermatogonial progenitor line.

A number of the genes also belonged to categories such as protein metabolic processes, lipid metabolic processes, phosphate metabolic processes, protein modification processes and MAPK cascade. We further performed a comparative analysis between the DEG in mESCs and GC1-Spg spermatogonial progenitors. Interestingly, a vast proportion of the DEG were very different between the two cellular systems **(Fig. 3.7 g**). Only 25 genes were however common to both the DEG sets and GO analysis revealed that the down regulated common ones (17 genes) belonged to categories such as signal transduction and transport activity whereas the up regulated common ones (8 genes) belonged majorly to cellular processes and organismal processes with a smaller number being grouped into immune processes and biological regulation. These observations imply that whilst Mrhl is a highly context dependant molecular player, it might be regulating a common indispensible set of targets to confer its role in development and cellular differentiation.

In order to further narrow down into more significant categories amidst the aforementioned DEG ones, Fisher's exact test was performed with a p-value<0.001. This allowed us to understand the perturbed pathways and processes that are enriched or over represented with a higher statistical significance. Interestingly, we observed that the DEG were grouped into more specified categories such as cell-cell signaling, ion transport, synaptic transmission, response to endogenous stimuli, ectoderm development, anion transport, cell-cell adhesion and neuromuscular synaptic transmission (**Table 1)**. The category of developmental process (GO: 0032502) had the maximum number of genes as 60 (**Appendix Table 1**) with a significant fold change of 1.72 and p-value of 6.44E-05. A closer inspection of the DEG belonging to this category revealed genes coding for lineage-specific transcription factors pertaining to neuronal lineage, hematopoietic and vascular lineage, cardiac lineage, skeletal lineage, mesodermal lineage and pancreatic lineage along with cell adhesion genes and receptor activity related genes with functions in migration, axon guidance, signaling, growth

and differentiation, structural roles and cellular proliferation amongst others (**Tables 2a** and **2b**).

GO-Slim Process	Obtained Number of Genes	Expected Number of Genes	P-Value	FDR	Fold Change	Obtained Over/under Expected
Neuromuscular synaptic transmission (GO:0007274)	6	1.08	1.19E-03	3.22E-02	5.54	+
Ectoderm development (GO:0007398)	16	4.79	5.25E-05	4.27E-03	3.34	+
Response to endogenous stimulus (GO:0009719)	17	5.28	4.68E-05	5.72E-03	3.22	+
Ion transport (GO:0006811)	28	9.01	3.57E-07	8.71E-05	3.11	+
cell-cell adhesion (GO:0016337)	11	3.62	1.54E-03	3.41E-02	3.04	+
Anion transport (GO:0006820)	16	5.67	3.21E-04	1.30E-02	2.82	+
Synaptic transmission (GO:0007268)	20	9.03	1.25E-03	3.04E-02	2.21	+
Cell-cell signaling (GO:0007267)	29	13.55	2.54E-04	1.24E-02	2.14	+
Developmental process (GO:0032502)	**60**	**34.89**	**6.44E-05**	**3.93E-03**	1.72	+
G-protein coupled receptor signaling pathway (GO:0007186)	6	20.16	4.36E-04	1.52E-02	0.3	-
immune response (GO:0006955)	3	14.52	5.72E-04	1.75E-02	0.21	-

Table 1

Fisher's exact test and the corresponding GO processes.

Functions	Transcription Factors
Neuronal lineage	Atoh1, Tbr1, Dlx3, Lhx1, Vsx2
Cardiac lineage	Myocd
Hematopoeitic and vascular lineage	Erg
Skeletal morphogenesis and osteoblast lineage	Runx2
Myeloid and B-cell lymphoid lineage	Spi1
Axis specification and positional identities	Hoxc4
Mesoderm differentiation and limb patterning	Tbx2
Trophoblast giant cells differentiation and cardiac morphogenesis	Hand1
Epithelialization of somitic mesoderm and development of cardiac mesoderm	Mesp1
Notch signaling	HeyL
Regulation of developmental processes	Tbx19
Pancreatic development and maintenance	Pdx1
Roles in epidermis and probable roles spermatogenesis, oocyte differentiation and embryo development	Bnc1

Table 2a *Transcription factors from the DEG belonging to developmental processes (GO: 0032502) involved in lineage specific functions.*

Functions	Cell Adhesion and Receptor Activity related genes
Migration of epidermal cells and cerebellar development	Fat2
Cell adhesion regulation during development	Col24a1, Col22a1, Lamc3
Neuromuscular circuit development, axon guidance	Epha3
Ca (+2) dependant cell adhesion and neurite growth in hippocampal neurons	Lrfn5
Axon guidance and neuronal cell survival	Uncd5
T-cell biology and neuron regeneration	Nav3
NMDA receptor signaling and regulation of synapses	Dlg2
Angiogenesis and spinal cord neurogenesis	Dll4
Growth and differentiation of neuronal, glial, epithelial and other cell types	Nrg2
Hematopoeisis	Flt3
Acrosome expansion during spermatogenesis	Spaca1
Muscle contraction and structural role in muscles	Mybpc3, Mybpc1
Component of basement membranes; role in hearing and vision	Ush2a
Cellular proliferation and differentiation during retina and bone formation	Gdf6
Osteogenesis and adipogenesis	Gdf10

Table 2b *Cell adhesion and receptor activity related genes from the DEG belonging to developmental processes (GO: 0032502) also involved in lineage specific functions.*

Thus, from these analyses, we concluded that depletion of Mrhl in mESCs causes dysregulation of major development and lineage-specific genes and processes.

3.3.5 Gene co-expression analysis in Mrhl knockdown conditions revealed several co-expression modules with distinct functional enrichment of various processes: One of the popular ways to derive relationships between genes and their functions and between genes and their regulation of physiological states from a genome-wide data is to perform a gene co-expression analysis [221]. In such an analysis, statistically correlated genes or statistically co-expressed genes are grouped together into modules. Subsequently, the modules are interrogated to identify gene-gene interactions leading to the construction of a network within a module or to understand if the correlated genes within a specific module define a particular function or phenotype. Gene co-expression analysis or protein co-expression analysis organizes the transcriptome or proteome in biologically meaningful modules of co-expressed genes or proteins associated with specific cellular states or disease phenotypes. In a study by Seyfried *et al*, clustering analysis followed by co-expression analysis for proteins across

healthy patients, patients with Alzheimer's disease (AD) and patients with asymptomatic AD provided insights into differentially expressed proteins across these patients and helped to identify a molecular signature defining each of the AD or asymptomatic AD conditions [222]. We performed hierarchical clustering of the DEG and obtained nine such co-expression modules. In each of the modules, the functional interactions between the genes were visualized with the help of Cytoscape. To further understand the biological significance of these clusters, GeneMania tool was used to generate the functional enrichment of the correlated genes in each cluster (**Fig. 3.8 a**). Whilst clusters 1 and 9 did not show any specific functional enrichment, cluster 2 displayed a diverse enrichment of functions including syncytium formation by plasma membrane fusion, protein A kinase binding, positive regulation of neurological processes and metabolic/biosynthetic processes. In cluster 3, a predominant category of functional enrichment was metabolic processes pertaining to organic molecules along with structural constituent of eye lens forming another functional category. Clusters 4 and 5 could be appropriately grouped into ion channel activity and transporter activity respectively. Cluster 6 represented a functional enrichment of various nervous system related processes such as glial lineage differentiation, peripheral nervous system development and axon ensheathment. Cluster 7 showed a functional enrichment of immune system processes such as host-cell interactions and responses to interferon-beta. Cluster 8 represented responses to xenobiotic stimuli as the functional classification.

An interesting observation from the Fisher's exact test analysis and the gene co-expression analysis was an over representation of nervous system related genes and processes (**Table 2 a,b; Fig. 3.8**). We sought to analyze this observation in further detail and performed a study on the gene list belonging to neuronal lineage from the statistical test and Cluster 6 from the gene co-expression analysis. The neuronal lineage related genes formed ~20% of the gene set belonging to developmental processes (GO: 0032502). Amongst them, ~70% of the genes were down regulated in expression (*Atoh1*, *Tbr1*, *Unc5d*, *Rarb*, *Lhx1*, *Epha3*, *Dlx3*, *Fat2*) and were observed to have functions in neuronal cell survival and/or migration, cell type specification, brain development and axon guidance or synapse functioning (**Appendix table 2a**). On the other hand, the cluster 6 genes with functional enrichment of glial lineage differentiation and axon ensheathment were all up regulated in expression (except *Wscd2*) under conditions of Mrhl knockdown in mESCs (**Appendix table 2b**). Of notable importance owing to their highest perturbation in expression were two transcription factors ATOH1 (GO: 0032502), down regulated in expression by ~2.2 fold and known to play important roles in

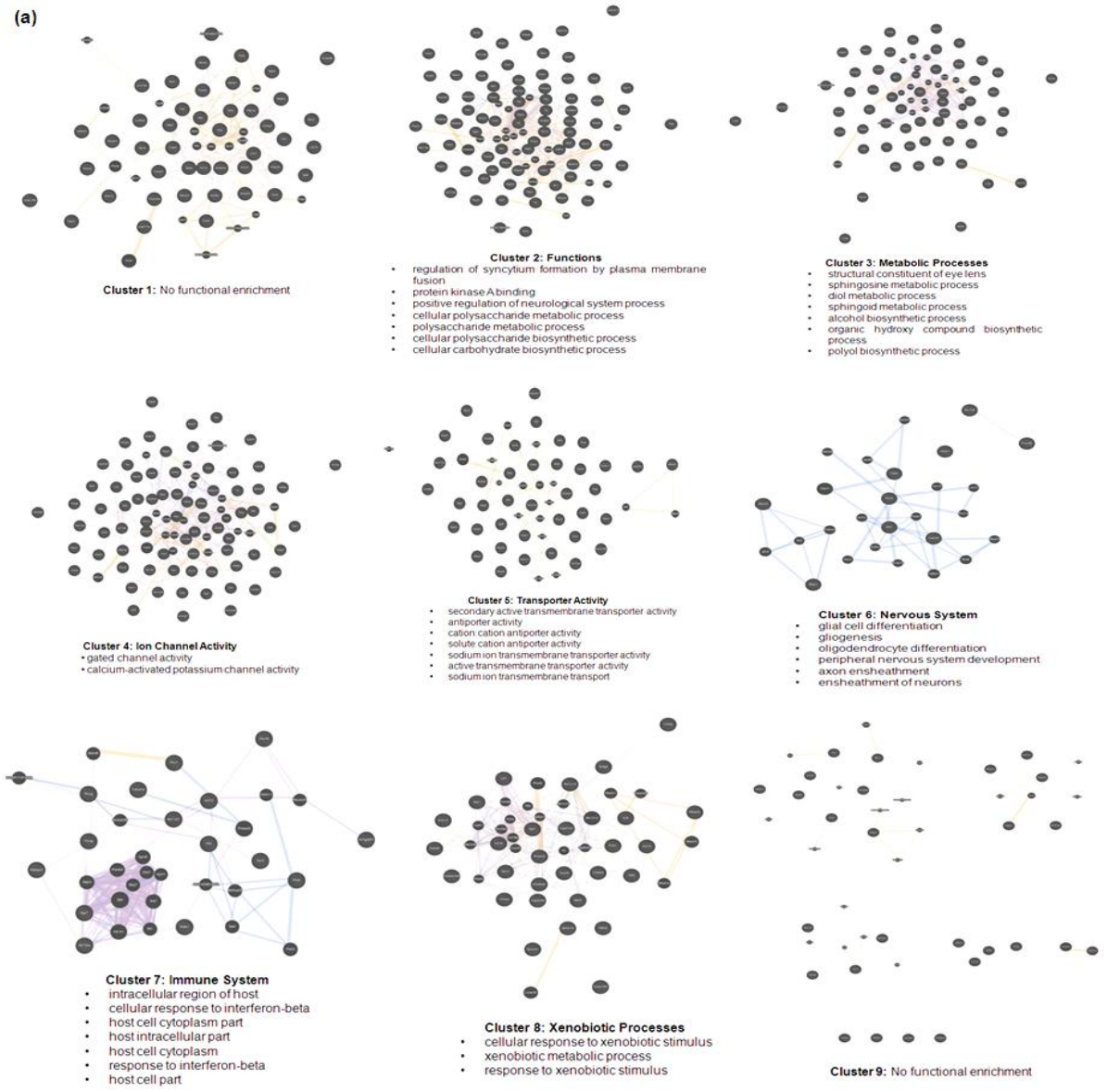
neuronal differentiation and cerebellum development [223, 224] and POU3F2 (Cluster 6), up regulated in expression by ~2.5 fold and established to reprogram fibroblasts into neurons in conjunction with ASCL1 and MYT1L [225]. An interrogation of their protein interaction partners from the STRING database revealed networks involving factors crucial for the regulation of various aspects of neuronal development. ATOH1 was observed to have known interactions with SOX2, established to play a role in the maintenance of neural stem cells/progenitors [226] along with ZIC1 [227-229], TCF3 [230] and TCF12 [231, 232] which are involved extensively in neural progenitor regulation and neuronal differentiation (**Fig. 3.8 b**). POU3F2 protein network revealed its interaction with PAX6, well known for its role in the development of the eye and brain [233, 234], TBR1 also known for its role in brain development [235, 236], ASCL1, a pioneer transcription factor involved in neuronal differentiation [225] and PAX3 [237, 238] and FOXA2 [239-241] which have roles in neuronal development and differentiation (Fig. 3.8 c). Although POU3F2 has been extensively established to play crucial roles in the neuronal lineage, a study by Schreiber *et al* revealed that POU3F2 is expressed in primary oligodendrocytes as well as in oligodendrocyte-like CG4 cells. POU3F2 is thought to compensate for the function of OCT6/POU3F1 in myelin forming oligodendrcoytes along with POU3F3 in the central nervous system. Although the exact role of POU3F2 in the glial lineage remains to be studied [242] it might explain its interaction with glial lineage related genes in Cluster 6. The study herewith showed that when Mrhl is depleted in mESCs, genes related to neuronal lineage and brain development are over represented amongst the list of perturbed genes and nervous system processes are functionally enriched in statistically co-expressed genes suggesting that Mrhl might have a role to play in the development of the brain and/or in neuronal lineage specification.

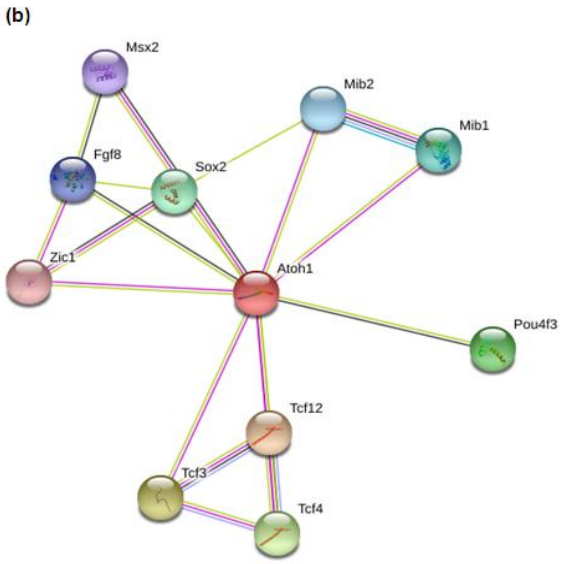
In addition to the analyses performed herewith with respect to the perturbed pathways and processes in Mrhl depleted mESCs, we also carried out a transcription factor (TF) network construction to address the potential interactions between the differentially expressed TFs and to identify a master TF through which Mrhl might be acting to regulate the TF network. A transcription factor network is a mode of concerted gene expression and interdependent regulation that often defines a particular cellular fate or disease state [243]. 'Endogenous' networks function in cells under basal conditions to define cellular architectures whereas 'exogenous' networks are those that function under an external stimuli or altered environments to determine the cellular output or response to the external factor [244]. In a

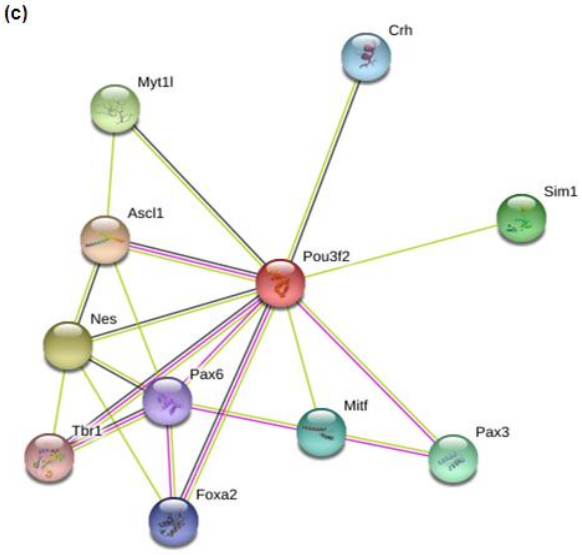
study by Dunn *et al*, a data-constrained computational approach was adopted to define the minimal TF network operating in mESCs [245]. Gene expression for 17 factors known to play roles in ES cell maintenance were interrogated across different culture conditions and a TF network was constructed using high Pearson correlation coefficients to narrow down the most essential factors involved in defining ES cell characteristics. It was observed that the cellular architecture of ES cells was defined by 12 TFs with 16 interactions across three different signal environments as inputs thus suggesting that the makeup of ES cells can be attributed to a simple core module rather than a vast genome-wide network. In a similar study performed by Goode *et al*, a host of transcriptome data was analyzed for 6 different TFs to understand the differentiation program in action from ES cells to blood precursor cells to terminally differentiated macrophages [246]. The study addressed the complexity of development/differentiation stage specific TF assembly regulating the relevant process, provided clues about TFs such as TAL1 and LMO2 that act as master TFs to alter the chromatin and transcriptional landscape of hematopoietic genes and identified novel factors such as TEAD and the cofactor YAP required for the specification of hematopoietic precursor cells. In our study, we narrowed down the list of TFs from our set of DEG in Mrhl knockdown condition of mESCs and obtained 38 TFs showing perturbation in expression. All TFs were observed to be lineage-specific or developmental-stage specific in functions (**Appendix table 3**). This observation was confirmed by performing a gene ontology analysis for these TFs which revealed developmental and metabolic processes to be the highest functional enrichments (**Fig. 3.8 d**). To address the potential cross-regulation between the perturbed TFs, motif for each TF was scanned across promoter sequence of each TF and a matrix was generated and visualized as a heat map (**Fig. 3.8 e, f**). The matrix heat map revealed RUNX2 as a potential master TF with binding motifs on the maximum number of TFs followed by ERG. Whilst RUNX2 is known for its role in development of the osteoblast and chondrocyte lineage [247] and ERG for its roles in the vascular lineage [248], the matrix identified a novel regulatory TF network that might be functioning in mESCs in the context of Mrhl depletion. To further understand this potential TF matrix in terms of their known interactions at the protein level, the TFs were provided as input into the STRING database and a simplified TF hierarchy was generated (**Fig. 3.9**). This exercise validated RUNX2 as the probable master TF at the top of the hierarchy with RB1, ERG, FEV, SMAD9 and HEYL being in the first tier and undergoing regulation via RUNX2, although their direct regulation by Mrhl cannot be ruled out. An overlap of the predicted matrix and STRING hierarchy revealed HEYL and HOXB7 to be the common regulated TFs of the potential TF network.

HeyL is a target of the NOTCH signaling pathway and interestingly, has been shown to promote the differentiation of neural progenitor cells [249] whilst *Hoxb7* is a homeobox containing gene that has been implicated for its roles in embryonic and multipotent stem cells [250, 251]. Our analyses with respect to TFs in mESCs under Mrhl knockdown conditions reveal two important perspectives; firstly, the differentially expressed TFs are all involved in regulating various aspects of development, patterning or defining specific cellular fates and secondly, the probable cross-talk or network operating amongst these TFs suggests that Mrhl might be involved in defining a state for the mESCs wherein they are primed to develop or differentiate into one or the other lineage or adopt a specific cellular identity from a more naive state and this regulation by Mrhl may either be in a positive or negative manner depending on the cue and the corresponding target genes which are involved.

Thus, the transcriptome gene ontology enrichment analyses, Fisher's statistical significance test, gene co-expression modules and TF hierarchy network strongly suggest that Mrhl is involved in determining the identity of mESCs with respect to differentiation and cell fate specification since lack of Mrhl expression leads to positive or negative regulation of genes and processes that define specific cellular states or lineages. One of the predominant states amongst these appears to be the neuronal lineage or neuronal identity specification.


(a)
Cluster 1: No functional enrichment
Cluster 2: Functions
regulation of syncytium formation by plasma membrane fusion
protein kinase A binding
positive regulation of neurological system process
cellular polysaccharide metabolic process
polysaccharide metabolic process
cellular polysaccharide biosynthetic process
cellular carbohydrate biosynthetic process
Cluster 3: Metabolic Processes
structural constituent of eye lens
sphingosine metabolic process
diol metabolic process
sphingoid metabolic process
alcohol biosynthetic process
organic hydroxy compound biosynthetic process
polyol biosynthetic process
Cluster 4: Ion Channel Activity
gated channel activity
calcium-activated potassium channel activity
Cluster 5: Transporter Activity
secondary active transmembrane transporter activity
antiporter activity
cation cation antiporter activity
solute cation antiporter activity
sodium ion transmembrane transporter activity
active transmembrane transporter activity
sodium ion transmembrane transport
Cluster 6: Nervous System
glial cell differentiation
gliogenesis
oligodendrocyte differentiation
peripheral nervous system development
axon ensheathment
ensheathment of neurons
Cluster 7: Immune System
intracellular region of host
cellular response to interferon-beta
host cell cytoplasm part
host intracellular part
host cell cytoplasm
response to interferon-beta
host cell part
Cluster 8: Xenobiotic Processes
cellular response to xenobiotic stimulus
xenobiotic metabolic process
response to xenobiotic stimulus
Cluster 9: No functional enrichment


(b)
Msx2
Mib2
Mib1
Fgf8
Sox2
Zic1
Atoh1
Pou4f3
Tcf12
Tcf3
Tcf4


(c)
Crh
Myt1l
Sim1
Ascl1
Pou3f2
Nes
Pax6
Mitf
Tbr1
Pax3
Foxa2

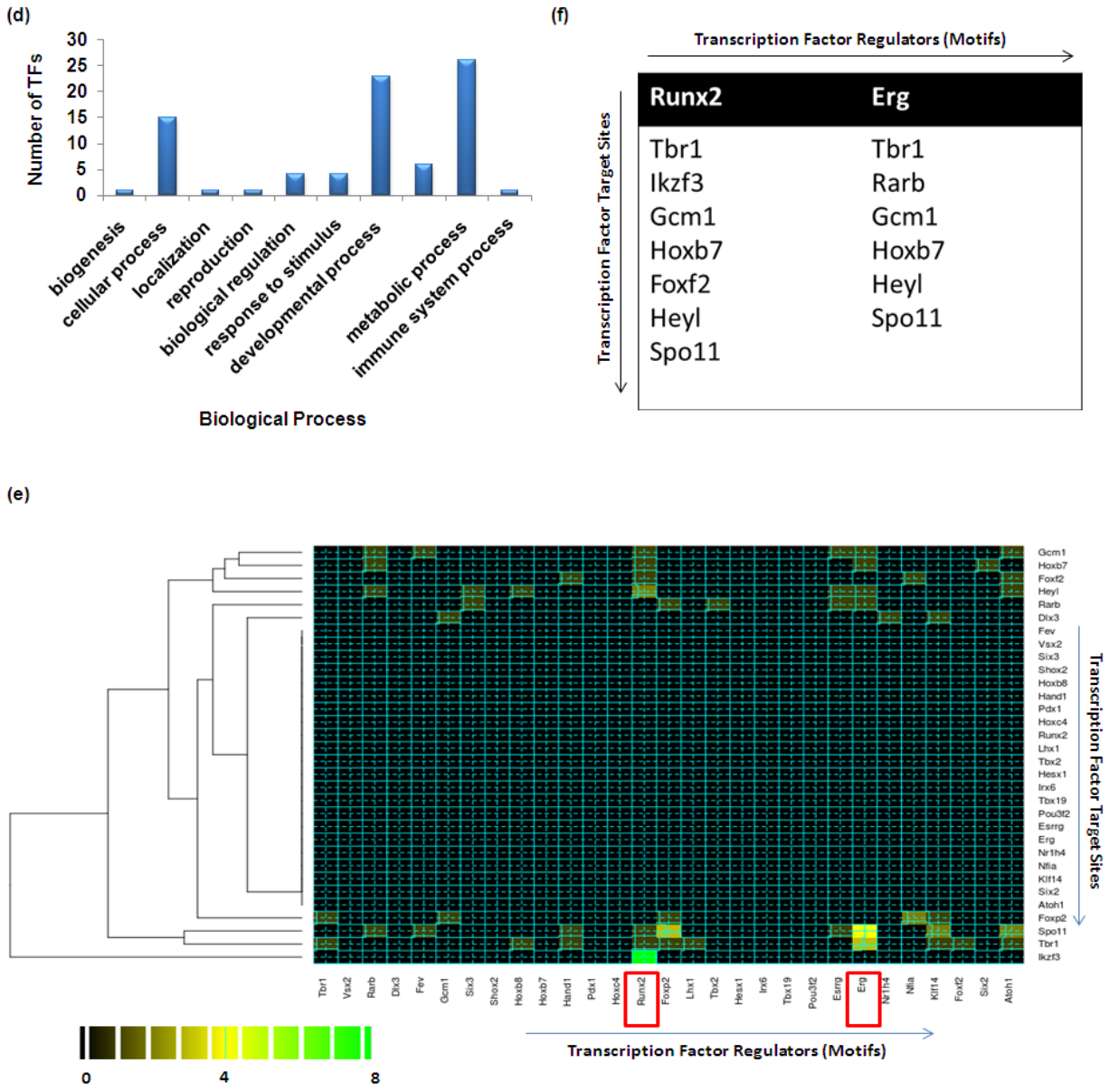


Fig. 3.8 *Gene co-expression and TF network analysis. (a) Hierarchical clustering using Cluster 3.0 revealed 9 gene co-expression modules with Clusters 2-8 having specific functional enrichments. (b,c) Interrogation of protein interaction partners of ATOH1 and POU3F2 from the STRING database; (d) Gene ontology for perturbed TFs; (e) TF matrix as visualized in R. (f) Consolidated view of TF matrix; RUNX2 and ERG displayed the highest number of binding motifs for other TFs.*

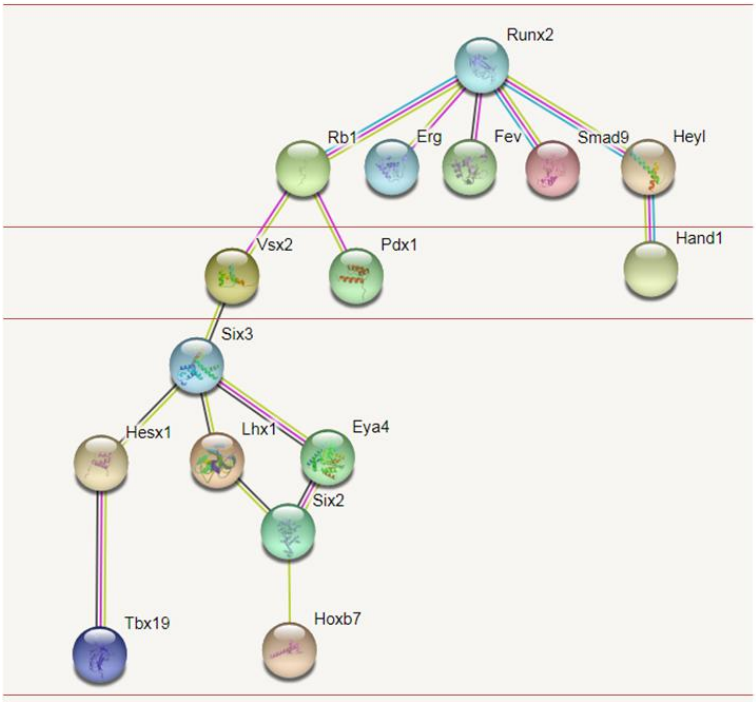


Fig. 3.9 *TF hierarchy generated upon analyzing interactions between the TFs in STRING.*

3.4 Summary: The study herewith was directed towards understanding lncRNA Mrhl as a molecular player in the pluripotency/differentiation circuit of mESCs. We observed that Mrhl is nuclear localized, associated with the chromatin and a moderately stable lncRNA in mESCs. We further observed that upon differentiation of mESCs into the three germ layers, Mrhl levels are up regulated during specific time points of differentiation in embryoid bodies. In conjunction with the feedback loop between Mrhl and WNT signaling in male germ cell meiotic commitment and the role of a conserved pathway such as WNT in development and differentiation, we found that Mrhl does not regulate WNT in the context of mESCs. Knockdown of Mrhl did not alter localization status of the WNT effector β-CATENIN. During differentiation into the various lineages, it still remains to be understood if Mrhl and WNT have a probable functional regulatory loop. However, it indicated that Mrhl acts through different pathways and targets depending on the cellular context. Towards understanding the role of Mrhl in mESCs per se, our shRNA mediated knockdown experiments followed by RNA-sequencing and gene ontology analysis revealed that 1143 genes to be perturbed in expression with broad molecular functions and biological processes being affected. GO enrichment analysis provided insights into developmental and multicellular organismal processes being the top amongst the perturbed processes. A further stringent Fisher's test showed developmental processes to be one of the most significant of the affected processes with the maximum number of perturbed genes. Interestingly, majority of these genes were studied to be lineage-specific transcription factors, cell adhesion and

receptor activity related genes important in various developmental processes. A comparison of the DEG between mESCs and Gc1-Spg spermatogonial progenitor line showed distinct and common targets of Mrhl emphasizing its context dependant role. Hierarchical clustering and gene co-expression analysis revealed statistically correlated genes acting in concert to regulate diverse functions such as ion channel/transporter activity, nervous system processes, immune system processes and response to xenobiotic stimuli. Fisher's test revealed ~20% of the perturbed gene set to be involved in neuronal lineage and brain development related function with ~70% of them being down regulated in expression whereas the gene co-expression studies revealed a functional enrichment of glial related processes with majority of the genes being up regulated in expression under conditions of Mrhl knockdown in mESCs. *Atoh1* and *Pou3f2* were noteworthy due to their highest perturbation in expression. Furthermore, an analysis of their protein interaction partners through STRING database revealed important transcription factors involved in neural stem cell/progenitor maintenance and/or proliferation, neurogenesis and brain development. An interrogation of the potential TF network operating in mESCs under conditions of Mrhl depletion showed that the TFs are majorly enriched for functions pertaining to development and metabolism with RUNX2 being a probable master TF in the network.

The major findings and implications of this study are as follows:

- Mrhl is a nuclear localized, chromatin bound, stable lncRNA expressed in mESCs.
- Mrhl exhibits differential expression upon embryoid body formation and does not function through the WNT/p68 cascade in mESCs.
- Depletion of lncRNA Mrhl in mESCs leads to perturbation of genes over representing developmental processes. Our systems approach involving gene ontology, gene co-expression and TF network analysis suggests that in mESCs, Mrhl might be involved in maintaining a more primed state of the cells which can readily respond to environmental cues for differentiation and development. This inference is supported by the fact that Mrhl is up regulated in expression itself during EB differentiation of mESCs. Interestingly, pluripotency related pathways and genes were not seen to be the targets of Mrhl in mESCs suggesting Mrhl as a player in cellular differentiation and fate specification.
- A significant proportion of the differentially expressed genes, processes and networks belonged to neuronal lineage and brain development. The study directed us towards

understanding the role of Mrhl in the context of neuronal lineage specification and brain development in further detail.

3.5 Future Perspectives

- Since Mrhl is predominantly nuclear localized and chromatin bound in mESCs, it would be interesting to address the chromatin occupancy at target loci for Mrhl in mESCs to understand the mechanisms of regulation of target genes by Mrhl. ChOP-Seq (chromatin oligo affinity purification followed by high throughput sequencing) has been performed in this regard.
- It would also be interesting to understand if target gene regulation at the chromatin level by Mrhl in mESCs is mediated directly through RNA-DNA triplex formation or through protein interaction partners. Our preliminary analysis has revealed that regulation of key development related genes by Mrhl might be through the former mechanism.

Appendix Table 1 *DEG belonging to developmental process (GO: 0032502)*

Gene Symbol	Gene Title	Fold Change
Ppef1	Serine/threonine-protein phosphatase with EF-hand 1	-2.42731
Tmod4	Tropomodulin-4	1.89531
Pink1	Serine/threonine-protein kinase , mitochondrial	-3.6133
Myocd	Myocardin	-2.38788
Nav3	Neuron navigator 3	3.12058
Lrfn5	Leucine-rich repeat and fibronectin type-III domain-containing protein 5	-3.25557
Fat2	Protocadherin	-2.15291
Atoh1	Protein atonal homolog 1	-2.19323
Hoxc4	Homeobox protein	-1.87857
Acer1	Alkaline ceramidase 1	-1.52245
Tbr1	T-box brain protein 1	-2.09846
Irx6	Iroquois-class homeodomain protein	2.17312
Erg	ETS transcriptional factor ERG	-3.22962
Heyl	Hairy/enhancer-of-split related with YRPW motif-like protein	-1.8545
Unc5d	Netrin receptor UNC5D	-1.9187
Col24a1	Collagen alpha-1(XXIV) chain	-2.7645
Eya4	Eyes absent homolog 4	-1.82276
Rarb	Retinoic acid receptor beta	-1.85496
Col22a1	Collagen, type XXII, alpha 1	-1.83229
Lamc3	Laminin subunit gamma-3	-2.77708
Tnf	Tumor necrosis factor	-2.12012
Lst1	Leukocyte-specific transcript 1 protein	2.20903
Apoa1	Apolipoprotein A-I	2.80639
Lhx1	LIM/homeobox protein	-1.77181
Runx2	Runt-related transcription factor 2	-2.50902
Tbx2	T-box transcription factor	1.50887
Spi1	Transcription factor PU.1	1.88953
Epha3	Ephrin type-A receptor 3	-2.93209
Olfml3	Olfactomedin-like protein 3	-1.86368
Dlg2	Disks large homolog 2	3.23586
Fev	Protein FEV	-1.69091
Nap1l5	Nucleosome assembly protein 1-like 5	1.86934
Gcm1	Chorion-specific transcription factor	-2.00664
Lrrc32	Leucine-rich repeat-containing 32	-1.68589
Hand1	Heart- and neural crest derivatives-expressed protein 1	-1.69772
Dlx3	Homeobox protein	-1.85345
Rb1	Retinoblastoma-associated protein	2.53923
Tbx19	T-box transcription factor	-1.92391
Vsx2	Visual system homeobox 2	1.55701
Gdf10	Growth/differentiation factor 10	-1.95133
Pdx1	Pancreas/duodenum homeobox protein 1	2.14924
Nr1h4	Bile acid receptor	-2.79377
Dll4	Delta-like protein 4	2.46391

Nrg2	Pro-neuregulin-2, membrane-bound isoform	1.59667
Insrr	Insulin receptor-related protein	-2.87654
Flt3	Receptor-type tyrosine-protein kinase	1.50741
Birc7	Baculoviral IAP repeat-containing protein 7	-1.95604
Bnc1	Zinc finger protein basonuclin-1	-2.85009
Cxcr6	C-X-C chemokine receptor type 6	-1.79159
Ifi27l2a	Interferon alpha-inducible protein 27-like protein 2A	1.89083
Spaca1	Sperm acrosome membrane-associated protein 1	2.16592
Mybpc3	Myosin-binding protein C, cardiac-type	2.13581
Ush2a	Usherin	4.59849
Gdf6	Growth/differentiation factor 6	1.60351
Vipr1	Vasoactive intestinal polypeptide receptor 1	-2.0256
Mst1	Serine/threonine-protein kinase 4	-1.95824
Mesp1	Mesoderm posterior protein 1	2.12023
Mybpc1	Myosin-binding protein C, slow-type	1.52215
Fat4	Protocadherin Fat 4	-1.62977

Appendix Table 2a *List of neuronal lineage related genes and functions from GO: 0032502*

Gene Name	Log2(Fold Change)	Function
Nav3	3.12058	Neuronal migration; neurite outgrowth [252]
Atoh1	-2.19323	Differentiation of subsets of neural cells [223, 224, 253]
Tbr1	-2.09846	Required for brain development [235, 236, 254, 255]
Unc5d	-1.9187	Involved in neuronal cell survival and axon guidance [256, 257]
Rarb	-1.85496	Neuronal differentiation [258, 259]
Lhx1	-1.77181	Neurogenesis and migration [260-263]
Epha3	-2.93209	Neuromuscular circuit development, axon guidance [264-267]
Dlg2	3.23586	NMDA receptor signaling and regulation of synapses [268]
Rb1	2.53923	Neuronal differentiation, growth and survival [269-271]
Vsx2	1.55701	Specification and morphogenesis of the sensory retina [272-274]
Dlx3	-1.85345	Neuronal development [275, 276]
Fat2	-2.15291	Migration of epidermal cells and cerebellar development [277]

Appendix Table 2b *Gene list and functions for cluster 6*

Gene Name	Log2(Fold Change)	Function
Trim80	1.87551	Putative uncharacterized protein
Pou3f2	2.45661	Neuronal differentiation [225, 228, 278]
Cwh43	2.54697	Lipid remodeling during GPI-anchor maturation
Tekt5	2.38598	Probable structural component of the sperm flagellum
Csgalnact1	1.86199	Chondroitin chain biosynthesis during cartilage formation
Hcar1	1.91039	Receptor for L-lactate
F2rl3	1.57639	Probable role in platelets activation
Wscd2	-1.92348	WSC domain-containing protein 2
Krt1	3.33992	Probable role in regulation of kinases
Fa2h	1.82857	Alpha hydroxylation of free fatty acids and sphingolipids
Malrd1	2.05934	Involved in transport of FGF19 and regulation of bile synthesis
Abcc9	1.76298	Subunit of ATP sensitive potassium channels
Wfdc1	1.5032	Growth inhibitory activity
Slc7a9	1.89926	Transport of cystine and neutral and dibasic amino acids
Gm4707	1.65698	Uncharacterized protein

Appendix Table 3 *List of differentially expressed TFs*

Gene Symbol	Gene Title	Fold Change
Hoxb7	Homeobox B7	-3.31541
Erg	ETS transcription factor ERG	-3.22962
Bnc1	Basonuclin 1	-2.85009
Nr1h4	Nuclear receptor subfamily 1, group H, member 4	-2.79377
Runx2	Runt related transcription factor 2	-2.50902
Foxp2	Forkhead box P2	-2.49933
Myocd	Myocardin	-2.38788
Atoh1	Atonal bHLH transcription factor 1 [	-2.19323
Foxf2	Forkhead box F2	-2.1155

Tbr1	T-box brain gene 1	-2.09846
Nfia	Nuclear factor I/A	-2.05348
Gcm1	Glial cells missing homolog 1	-2.00664
Tbx19	T-box 19	-1.92391
Hoxc4	Homeobox C4	-1.87857
Smad9	SMAD family member 9	-1.86555
Rarb	Retinoic acid receptor beta	-1.85496
Heyl	Hairy/enhancer-of-split related with YRPW motif-like	-1.8545
Dlx3	Distal-less homeobox 3	-1.85345
Spo11	SPO11 meiotic protein covalently bound to DSB	-1.82591
Eya4	EYA transcriptional coactivator and phosphatase 4	-1.82276
Klf14	Kruppel-like factor 14	-1.77236
Lhx1	LIM homeobox protein 1	-1.77181
Hesx1	Homeobox gene expressed in ES cells	-1.76546
Ikzf3	IKAROS family zinc finger 3	-1.70601
Hand1	Heart and neural crest derivatives expressed 1	-1.69772
Fev	FEV (ETS oncogene family)	-1.69091
Shox2	Short stature homeobox 2	-1.54361
Six3	Sine oculis-related homeobox 3	4.7411
Rb1	RB transcriptional corepressor 1	2.53923
Pou3f2	POU domain, class 3, transcription factor 2	2.45661
Hoxb8	Homeobox B8	2.43747
Esrrg	Estrogen-related receptor gamma	2.27188
Irx6	Iroquois homeobox 6	2.17312
Pdx1	Pancreatic and duodenal homeobox 1	2.14924
Khdrbs2	KH domain containing, RNA binding, signal transduction associated 2	1.9474
Six2	Sine oculis-related homeobox 2	1.80415
Vsx2	Visual system homeobox 2	1.55701
Tbx2	T-box 2	1.50887

Chapter 4

Deciphering lncRNA Mrhl as a molecular player in neuronal lineage development

4.1 Introduction

4.1.1 Vertebrate neuronal development

During gastrulation stages of embryonic development, the ectoderm gives rise to the neural plate which forms the primordium of the central nervous system. The neural plate comprises of a single sheet of neural stem cells referred to as the neuroepithelial cells (NECs). The NECs possess self-renewal capacity and undergo rapid symmetric divisions to expand the neural plate and ultimately lead to the formation of the neural tube. At around E8 of development, neurogenesis sets in and it peaks around E14 (reviewed in [279]). Simultaneously at around E8-E10, the NECs undergo morphological and cellular changes to give rise to the future progenitors of the brain i.e., the radial glia cells (RGCs). The RGCs reside in the ventricular zone (VZ) and subventricular zones (SVZ) of the mammalian brain and are characterized by replacement of the tight junctions of NECs with apically located adherens junctions, expression of neural stem cell, neuronal and astroglial markers and asymmetric divisions. RGCs are multipotential progenitors of the brain generating neurons during embryonic stages and glia in the postnatal stages. A population of RGCs, however, remain quiescent in the SVZ and contribute to adult neurogenesis.

4.1.2 Radial glia cells

In the developing mammalian brain, neural progenitors can be classified as apical progenitors (APs) or basal progenitors (BPs) depending on their site of mitosis on the apical (ventricular zone) or basal surfaces of the brain, respectively (**Fig. 4.1**). RGCs can be either apical RGCs (aRGC) or basal RGCs (bRGC). The aRGCs have their nucleus located in the ventricular or apical zones (interkinetic nuclear migration) and their processes extend until the basal surface which act as guides for migration of newly born neurons to the cortical plate. The aRGCs divide to generate a class of fate-restricted progenitors known as the basal progenitors (BPs) which move basally by a process known as delamination and reside in the SVZ. Intermediate progenitors (IPs) and bRGCs form the two types of BPs. All BPs can undergo either symmetric or asymmetric divisions; whilst the IPs mostly divide symmetrically in a self-consuming manner to generate two neurons thereby contributing to the neuronal output of the developing brain. However IPs can also undergo one or more rounds of proliferative divisions. bRGCs mainly undergo asymmetric divisions to generate a neurogenic IP (nIP) and a bRGC daughter cell. Interestingly, RGCs also divide to generate IPs of astrocytic or oligodendrocytic lineage although gliogenesis sets in majorly after birth wherein the RGCs

have already migrated to the cortical plate and now participate in the generation of different types of glia (reviewed in [279]).

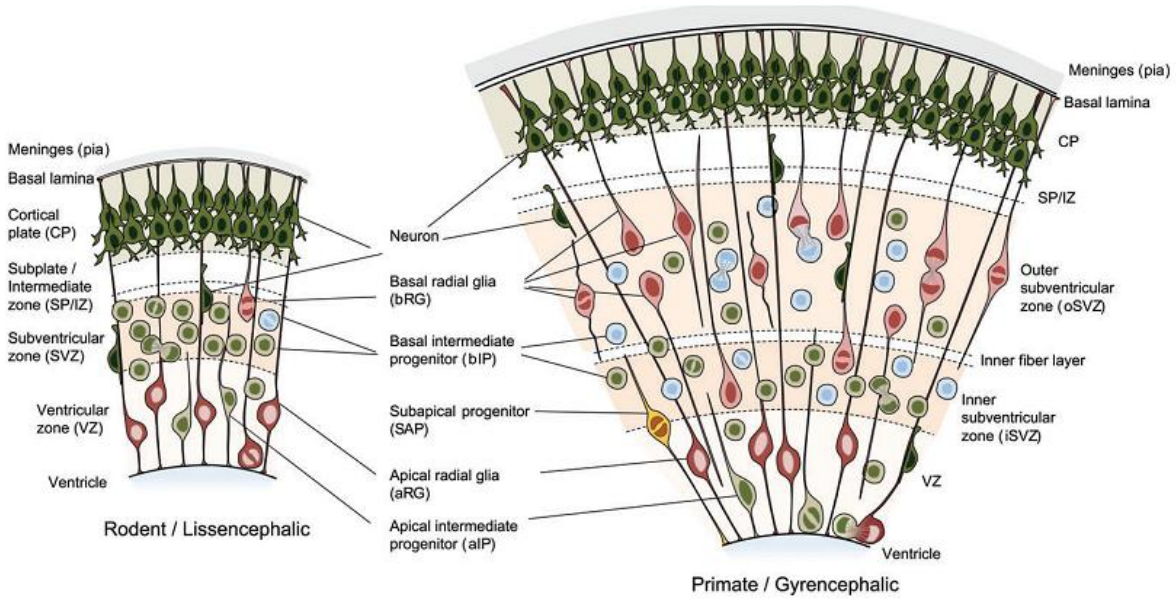


Fig. 4.1 *Types of neural progenitors based on their polarities and layers of the developing brain in rodent (a) and primate species (b). Apical and basal progenitors can be classified based on their ventricular contact, location of mitosis and cell polarity. NECs that generate aRGs and aIPs have not been depicted n the picture. BPs can be IPs or bRGs. IPs can be neurogenic (green) or proliferative (blue). bRGs can again be primary that have either basal processes or both apical and basal processes or secondary that are born in the SVZ and can have apical, basal or both processes. Subapical progenitors (SAPs) have a ventricular contact but undergo mitosis at an abventricular location. Adapted from [280] with permission.*

4.1.3 Pathways and mechanisms in neuronal development and RGCs

I. Transcription factors

One of the major transcription factors responsible for the development of much of the central nervous system in invertebrates and vertebrates is PAX6. Mice mutant for PAX6 display defects in neural stem and progenitor cell proliferation, neurogenesis and patterning of the brain [281-286]. In the developing mouse embryonic brain, PAX6 is known to be strongly expressed in the VZ consisting of NECs and RGCs and down regulated in IPs that have lost their radial processes and are characterized by the expression of TBR2 and cortical pyramidal projection neurons characterized by the expression of TBR1 [254]. Studies by Sansom *et al*

[287] revealed that PAX6 occupies promoters of genes expressed in neural stem and progenitor cells (*Pax6* itself, *Hmga2*, *Cutl1*, *Nr2f2*, *Emx2*, *Sox9*, *Neurog3*, *Tle1*), in BPs (*Eomes*/*Tbr2*, *Neurod1*), in newly born neurons (*Sox4*, *Sox11*) as well as in mature cortical neurons (*Rorb*, *Etv1*) suggesting its widespread regulation of several aspects of neurogenesis in the mouse forebrain. Furthermore, it was shown to bind to genes related to cell cycle and control the expression of *Hes5* and NOTCH ligands indirectly indicating its regulation of self-renewal properties of neural progenitors. Furthermore, the study revealed that decreased levels of PAX6 led to a decrease in the proliferation of cortical stem cells and precocious neurogenesis. Simultaneously, increase in its levels caused generation of more BPs and excessive neurogenesis suggesting that the levels of PAX6 are crucial towards determining the balance between progenitor maintenance and neurogenesis. PAX6 acts in concert with other key transcription factors such as HES1, NGN2 and MASH1/ASCL1 to form a core network regulating the dynamics of neurogenesis and progenitor proliferation in the developing brain. PAX6 has also been implicated in activating genes of the ectodermal and neuronal lineage whilst repressing those of mesodermal/endodermal origin [288].

NGN2 and ASCL1 are two of the crucial proneural transcription factors. Along with NGN1, they are known to be expressed in RGCs in the developing cortex (reviewed in [289]. Although they might not be essential for the specification of RGCs, loss of NGN2 and ASCL1 leads to the formation of disorganized RGCs that differentiate into astrocytes prematurely along with severe defects in neurogenesis and development of the cerebral cortex [290]. NGN2 and ASCL1 have been implied for their roles during neurogenesis. NGN2 and ASCL1 have been shown to induce the generation of differentiated neurons *in vitro* or *in vivo* upon over expression [291, 292]. Furthermore, ASCL1 also is known to reprogram adult fibroblasts into neurons along with BRN2 and MYT1L emphasizing on their roles as master transcription factors of the neurogenesis program [225, 278, 293]. A study of the gene targets of ASCL1 in the embryonic telencephalon by ChIP-chip analysis revealed that whilst ASCL1 binds to the promoters of genes known to arrest cell cycle progression in neural progenitors, it also occupies promoters of genes such as *E2f1* (promotes G1/S transition), its coactivator *Ep400* and one of their targets *Cdca7* [294]. *Tead1*, *Tead2* (promote the proliferation of neural progenitors) and *Id1* (inhibitor of differentiation) were amongst the other genes whose promoters were found to be bound by ASCL1 in the embryonic telencephalon suggesting a novel role for ASCL1 in regulating neural progenitor proliferation. Embryos mutant for *Ascl1* showed a loss of proliferating intermediate

progenitors in the SVZ and decrease in expression of cell cycle facilitative genes although gross defects in proliferation were not observed probably due to compensatory pathways. Thus, a coordinated regulation of multiple targets at specific times by ASCL1 controls the balance between proliferation and differentiation in neural progenitors.

TBR2 and NEUROD1 have also been implicated widely during brain development. TBR2 marks the transition of RGCs to IPs with a concomitant down regulation of PAX6 although the generation of $TBR2^+$ IPs is severely affected in *Pax6* deficient mice [295]. Loss of *Tbr2* was shown to cause a reduction in the number of proliferating cells in SVZ with severe impairment of the formation of upper cortical layer neurons [295, 296]. Interestingly, NGN2 and JMJD3 were shown to associate with TBR2 providing insights into its role in regulating neurogenesis [297]. NEUROD1 is a pioneer transcription factor and has been implicated in binding to inactive neuronal genes and activating their transcription through epigenetic mechanisms and in increasing chromatin accessibility at its target loci inducing the initiation of the neuronal program and participating in neuronal migration [298].

II. Signaling pathways

NOTCH signaling pathway: The NOTCH pathway has been implicated in the regulation of NEC to RGC transition in the embryonic brain at the onset of neurogenesis. In the canonical NOTCH signaling pathway, a transmembrane receptor on a cell interacts with a transmembrane ligand on an adjacent cell leading to the proteolytic cleavage of the NOTCH intracellular domain (NICD) of the transmembrane receptor. The NICD translocates to the nucleus and interacts with RBPJ-κ (in mammals) or Su(H) (in flies) or LAG-1 (in *C. elegans*) to initiate the transcription of NOTCH target genes. Premature activation of the NOTCH pathway was shown to induce RGC characteristics in the mouse forebrain [299]. Furthermore, it has also been shown to inhibit the formation of BPs from RGCs [300]. However, deletion of the NOTCH targets *Hes 1* or *Hes 5* or the effector RBPJ-κ did not lead to defects in the appearance of RGCs although the RGCs were shown to lose their apico-basal polarity and undergo premature neurogenesis [301]. This suggests that the NOTCH pathway is crucial for the proper maintenance and proliferation of the RG progenitors. The NOTCH pathway has also been implicated in the promotion of neural progenitor self-renewal capacity since the HES family of transcription factors inhibit the activity of proneural genes such as *Mash1*, *Math* and *Ngn1*, thereby allowing the maintenance of the progenitor pool [302]. Interestingly, NOTCH pathway also has been shown to promote gliogenesis in the

presence of growth factors such as CNTF, LIF and BMPs implying that a tight control of the pathway is necessary for regulating neural and glial fates in the progenitors (reviewed in [303]).

FGF signaling pathway: The FGF family consists of eighteen secreted proteins that interact with FGF receptors (FGFRs) to initiate signaling via a multitude of downstream molecules such as Ras-MAPK, PI3K-AKT, PLC-ɤ and STAT. Expression of a constitutively active form of FGF2R has been shown to facilitate RGC identity [304]. Amongst the ligands, FGF10 has been observed to be expressed in the VZ at E9.5 and its over expression has been shown to promote the expression of RGC markers. Mutants for FGF10 displayed an extended period of NEC expansion and delayed neurogenesis [305]. FGF signaling has also been implicated in inhibiting RGC to BP transition [306]. Furthermore, FGF signaling is involved in maintaining the proliferative capacity of progenitors in the developing brain cortex by regulating the cell cycle. The cell cycle has been shown to be relatively longer in progenitors dividing to generate post-mitotic neurons as compared to progenitors still undergoing self-renewing divisions [307]. FGF2 has been specifically shown to regulate the expression of CDK1 and down regulate cyclin dependant kinase inhibitor p27 leading to the shortening of the G1 phase hence promoting proliferation and reducing terminal differentiation of progenitors [308].

WNT signaling pathway: The WNT pathway has been shown to perform context-dependant roles in regulating proliferation and differentiation of NECs/RGCs. Studies undertaken towards stabilizing the expression of β-CATENIN in transgenic mice have shown that such mice have enlarged cerebral cortical surface with increased ventricles lined with NECs and a greater proportion of NECs re-entering cell cycle after mitosis [309]. Conditional or topical knockouts of β-CATENIN revealed decreased cell proliferation of cortical precursor cells and premature exit from cell cycle leading to precocious differentiation into neurons, impaired interkinetic nuclear migration, abnormalities in the organization of the neuroepithelium and an increased number of astrocytes in the newborn brains [310, 311]. On the contrary, other studies have proved the involvement of the WNT pathway in generation of BPs from RGCs, in regulating the proliferation of BPs through n-MYC and NGN2 and in facilitating the neuronal differentiation of BPs (reviewed in [289]). Mice mutant for the WNT effector LRP6 also show defects in neurogenesis and a reduction in the size of the cortical plate [312]. In another study by Hirabayashi *et al*, over expression of WNT 7a or a stabilized form of β-CATENIN in mouse cortical neural precursors led to increased neuronal differentiation even

in the presence of Fgf2 whereas a block in the pathway led to impaired neuronal differentiation both *in vitro* and in the mouse brain neocortex [313]. This could be attributed to direct regulation of the proneural gene *Ngn1* by β-CATENIN /TCF complex. Interestingly, at earlier developmental stages, stabilized β-CATENIN did not cause increased neuronal differentiation thereby supporting the context-dependant role of WNT pathway in neural development.

Retinoic acid signaling pathway: In vertebrates, retinol binds to retinol binding protein and transthyretin and enters the target cells through the transmembrane receptor STRA6. Retinol is oxidized to retinal by alcohol dehydrogenases and converted to retinoic acid (RA) by aldehyde dehydrogenases**.** RA is either degraded via oxidation by members of the cytochrome P450 family or transported to the nucleus by cellular retinoic acid binding proteins (CRABPs). The RA binds to a heterodimer of RAR/RXR (retinoic acid receptors) which occupies target genes containing the retinoic acid response element (RARE) leading to the removal of corepressors followed by recruitment of coactivators and gene activation. In the mouse embryonic brain, RA signaling has been predominantly implicated in hindbrain segmentation and patterning. It has also been shown to act in concert with WNTs/FGF to impart dorsal identity and SHH to impart ventral identity in the developing forebrain (reviewed in [314]). In mice mutant for FOXC1 which is required for the production of RA in the meninges, there was a dramatic increase in the population of RGCs and a reduction in the generation of IPs and post-mitotic neurons [315]. In mouse ES cells, it was shown that RA is required to inhibit a neuroectoderm inhibitory pathway supported by NODAL/WNT signaling to facilitate the proliferation of progenitors or their commitment to differentiated neurons in a concentration dependant manner [316]. Furthermore, mice mutant for *Raldh2* (retinaldehyde dehydrogenase 2) and hence defective in RA synthesis exhibit a decrease in the population of SOX2 and SOX1 expressing neuroectodermal progenitor cells with an increased formation of presomitic mesodermal progenitors [317]. These studies have suggested that a balanced RA signaling cue is essential for a development of the embryonic brain.

Thus, a concerted and intertwined action of various signaling pathways and transcription factors are involved in regulating aspects of neural stem cells/ progenitor proliferation and neurogenesis.

4.1.4 LncRNAs in neuronal development

The role of lncRNAs in neuronal development and disorders of the nervous system have been well established. Cataloguing of lncRNAs of the human genome by GENCODE and NONCODE revealed ~40% of them to be expressed in the brain alone. Furthermore, lncRNAs have been shown to be more potent indicators of cell types in the brain than protein-coding mRNAs owing to their tight spatial and temporal regulation of expression and functions (reviewed in [318]).

The roles of lncRNAs such as Dali, Paupar and Pnky in regulating neural progenitor proliferation and/or differentiation and gene regulation have been discussed. LncRNA Rmst is expressed specifically in the brain with the mouse homolog being highly expressed in midbrain dopaminergic neurons [319]. Genome-wide ChIP and chromatin-RNA studies revealed that Rmst (rhabdomyosarcoma 2-associated transcript) interacts with SOX2 to regulate a common set of genes during neurogenesis including *Ascl1, Ngn2, Hey2* and *Dlx1* and Rmst knockdown affected the binding of SOX2 to some of its targets. Interestingly, knockdown of Rmst led to the differentiation of neural progenitors into glia whereas its over expression caused the progenitors to develop into $TUJ1^+$ neurons indicating the importance of Rmst in neuronal differentiation. Whilst lncRNAs Dali, Tuna and Rmst act to regulate the process of differentiation of progenitors into neurons, lncRNAs Pnky and Paupar are involved in mediating mechanisms that maintain the proliferative and self-renewal characteristics of neural stem cells and progenitors. Furthermore, along with lncRNA Pnky, lncRNA Evf2 has been implicated in regulating neuronal development in the *in vivo* scenario in embryonic and/or postnatal brain [320]. Mice harbouring a deletion of Evf2 were shown to undergo an imbalance in the ratio of excitatory to inhibitory neuron balance in the postnatal brain. They also displayed reduced synaptic inhibition characterized by lack of specification of GABAergic interneurons. Evf2 was shown to epigenetically regulate genes such as *Dlx5*, *Dlx6* and *Gad1* that are responsible for the development of the interneuron lineage. LncRNA linc-Brnb1 was also shown to participate in the differentiation of delaminating neural progenitors in the developing brain [321]. Loss of its locus in mice led to a disruption of proper development of the upper cortical neuronal layers and reduction in size of the somatosensory cortex consistent with a loss of BPs that causes precocious migration and differentiation in the lower cortical neuronal layers. Other lncRNAs such as Tug1 (taurine upregulated 1) and Six3os have been shown to regulate more defined cell type specification such as retinal cells (Tug1) and photoreceptors, bipolar cells and Muller glia (Six3os) through

gene regulation at their target loci (reviewed in [318]). LncRNA Bdnf-as (brain derived neurotrophic factor antisense) has been shown to negatively regulate expression of its *cis* gene *Bdnf* [322]. Loss of Bdnf-as led to reduced recruitment of EZH2, a component of the transcriptional repressive PRC2 complex that further led to increased expression of *Bdnf* resulting in neuronal outgrowth, differentiation as well as survival and proliferation *in vitro* and *in vivo*.

Apart from cellular proliferation and fate specification, lncRNAs have also been shown to play roles in synapse formation (reviewed in [318]). LncRNA Bc1 (brain cytoplasmic RNA 1)/Bc200 is one such example which is transported to dendrites in the embryonic and adult nervous systems where it interacts with proteins of the translational machinery to control 48S pre-initiation complex formation and repression of translation in synapses locally to give effect to synaptic turnover and neural plasticity. Knockdown of lncRNA Malat-1 was also shown to lead to a decrease in synaptic density in cultured hippocampal neurons by mediating the recruitment of SR-family related splicing proteins to target genes involved in synaptogenesis.

Several lncRNAs have also been associated with neurodegenerative conditions and brain disorders. A lncRNA Msnp1-as, antisense to its protein-coding gene *Msn* which is involved in the synaptic functioning, was shown to harbour single nucleotide polymorphisms (SNPs) associated with autism spectrum disorder (ASD) [323]. It was observed that all three SNP genotypes strongly correlated with the expression of Msnp1-as and the lncRNA was implicated in binding to *Msn* mRNA and stabilizing it. Although the mechanisms still remain to be delineated, ASD patients exhibit higher levels of both the lncRNA and the protein and the SNPs could be a contributing factor to ASD pathogenesis. LncRNA Gomafu has been implicated in schizophrenia by contributing to aberrant splicing of *Disc1* and *Ebbr4* [324]. Such splicing defects in these genes have been known to be a cause for the pathogenesis of schizophrenia. LncRNA Bace1-as is thought to contribute to the pathogenesis of Alzheimer's disease (AD) [62]. It is expressed antisense to *Bace1* which codes for a transmembrane β-secretase protein and acts to stabilize the mRNA of this gene. By inducing the expression of Bace1, Bace1-as (also over expressed in AD) leads to the increased formation and accumulation of toxic AB-42 peptides, thereby contributing to AD pathogenesis. Interestingly, targeting Bace1-as with small interfering RNAs and reducing its levels has been shown to abrogate AD symptoms in mice models.

Thus, lncRNAs act as important and often crucial players in regulating diverse aspects of neural progenitor proliferation, neuronal differentiation, neural fate specification and brain development. Furthermore, they also possess the potential to be developed as therapeutic targets for various neurodevelopmental disorders. In this context, it becomes essential to address the mechanisms and pathways regulated by lncRNAs in neuronal development.

4.2 Rationale of Current Study

Our studies towards understanding the role of lncRNA Mrhl in mouse embryonic stem cells provided us with insights into its regulation of developmentally relevant transcription factors, cell adhesion and receptor activity related genes and various processes including ion channel and transporter activity, metabolic processes, immune system processes and xenobiotic processes. A recurrent observation amongst the DEG in Mrhl knockdown mESCs and in our system analyses was the perturbation of genes related to the neuronal lineage and nervous system processes (**Appendix tables 2a, 2b and 3**). Amongst the down regulated genes in GO: 0032502, *Atoh1* has been implicated in the specification of subtypes of neural cells, *Tbr1* has been shown to have functions in regulating neuronal identity in the cortex/neocortex of the developing brain, *Unc5d* has roles in cortical cell survival and migration, *Rarb* controls the development of striatonigral projection neurons and cell death and differentiation in the ganglion, *Lhx1* is known to mediate neural retina differentiation, motor neuron migration, maintenance of spinal cord interneurons and Purkinje cell differentiation in the developing cerebellum, *Epha3* has been shown to be involved in differentiation of neural precursor cells, patterning and development of hippocampal axons, *Dlx3* is known to be an important transcription factor in neuronal development and neuron generation and *Fat2* has been implicated in cell adhesion processes along with other FAT cadherins to impart structure to neuronal circuits. The up regulated genes comprised *Nav3* known to have roles in neuron migration and neurite outgrowth, *Dlg2* implicated in the signaling of NMDA receptors (one of the three types of glutamate receptors), *Rb1* known to regulate neuronal differentiation and also neuronal cell growth and survival in human cerebral organoids and *Vsx2* which has important roles in the development of the retina. The gene co-expression analysis revealed *Pou3f2* as one of the major neuronal related genes in Cluster 6 (that exhibited nervous system processes as the functional enrichment), up regulated in expression in Mrhl knockdown in mESCs. During the reprogramming of fibroblasts to neurons, POU3F2 has been shown to be recruited to ASCL1-bound chromatin across the genome to facilitate the generation of induced neurons along with Myt1l [225]. Mice mutant for *Pou3f2* and *Pou3f3* (*Brn2* and

Brn1) showed defects in the proliferation of cortical progenitor cells and the migration of neurons in the upper layers of the cortex [325]. Furthermore, during retinoic acid mediated neuronal differentiation of embryoid bodies, POU3F2 has been observed to directly regulate the survival of neurons and to bind to important genes related to neurogenesis including *Ascl1* and *Ngn1* [228]. Interestingly, in the absence of POU3F2, crucial regulators of neurogenesis such as *Ascl1*, *Neurog1*, *Dcx*, *Dcc*, *Zic1*, *Zic2*, *Zic4*, *Zic5*, *Slit1* and *Uncx* were all down regulated in expression with an up regulation of genes related to mesoderm and endoderm lineages, suggesting POU3F2 to be an important regulator of neuronal differentiation and neuronal fate specification. The TF network interrogation also revealed HEYL, known to be a NOTCH target and for neuronal lineage differentiation, to be a present in the first tier of hierarchy. With these observations and studies in mind, it appears that in mESCs where Mrhl is depleted, a host of crucial neuronal lineage related genes with diverse functions in growth, survival, differentiation, migration and synapse development are perturbed. The roles of lncRNAs in regulating widespread functions in brain development and in neuronal lineage have already been outlined. In this context, we addressed the role of lncRNA Mrhl in neuronal lineage development.

(All materials and methods have been detailed in Chapter 2).

4.3 Results

4.3.1 Mrhl expression is up regulated in mouse embryonic brain stages with predominant expression in neuronal progenitors: To understand the relevance of Mrhl in the context of neuronal lineage and brain development, we first analyzed its expression pattern in embryonic and postnatal brain stages. We isolated embryonic whole brains from stages E10.5 to E18.5 and postnatal whole brains from P0 to P40. qRT-PCR analysis showed that Mrhl is expressed predominantly in the embryonic stages with its expression peaking at stages E12.5 to E18.5 (**Fig. 4.2 a**). Furthermore, there was a drastic down regulation in the levels of Mrhl from stages P0 onwards. Next to know the spatial expression pattern of Mrhl in the embryonic stages of brain development we selected one of the stages, E14.5 for two reasons: firstly, the expression of Mrhl was observed to be significantly high at this stage and secondly, E14 is the stage when NECs have developed into early RGCs and neurogenesis reaches a peak [279]. We harvested brains at E14.5 stage and divided them into fore-, mid- and hindbrain regions (**Fig. 4.2 b**). To confirm the purity of the sections, we performed qRT-PCR analysis for various markers such as FOXG1 for forebrain, EN1 for midbrain and GBX2

for hindbrain [326, 327] and observed that each of the markers were enriched in expression in the corresponding regions of the brain. We also analyzed the expression of PAX6 as a control because of its well established role in the development of RGCs and its defined expression patterns in the fore-, mid- and hindbrains [328]. The analysis showed us that Mrhl is expressed ubiquitously across all regions of the brain at E14.5 stage with no preference of expression across fore-, mid- or hindbrains. At E14.5 stage, neuronal progenitors (RGCs) are in the state of rapid proliferation and generation of specialized progenitors such as BPs which in turn give rise to the newly born neurons. Whilst Mrhl might have specific roles to play across all regions of the developing brain, we focussed on its role in the neuronal progenitors. Towards this end, we first analyzed available ChIP-Seq datasets for RNA Pol II occupancy, H3K4me3 transcription activation histone modification, H3K27me3 transcription repression modification and RNA Pol II-Ser 5p (initiation polymerase) on 1 kb promoter sequence of Mrhl, upstream from the TSS (GSE93009, [329]) in NESTIN^{+} neural progenitor cells isolated from the cortex of E15.5 brains. The analysis revealed that the proximal promoter region of Mrhl (`~100-200 bp up stream from the TSS) possessed distinct enrichments for the occupancy of RNA Pol II, initiating RNA Pol II and H3K4me3 modification (**Fig. 4.2 c**). Hence, the above study showed that Mrhl is actively expressed in the embryonic stages of mouse brain development with an indication of its predominant expression in neural/neuronal progenitors.

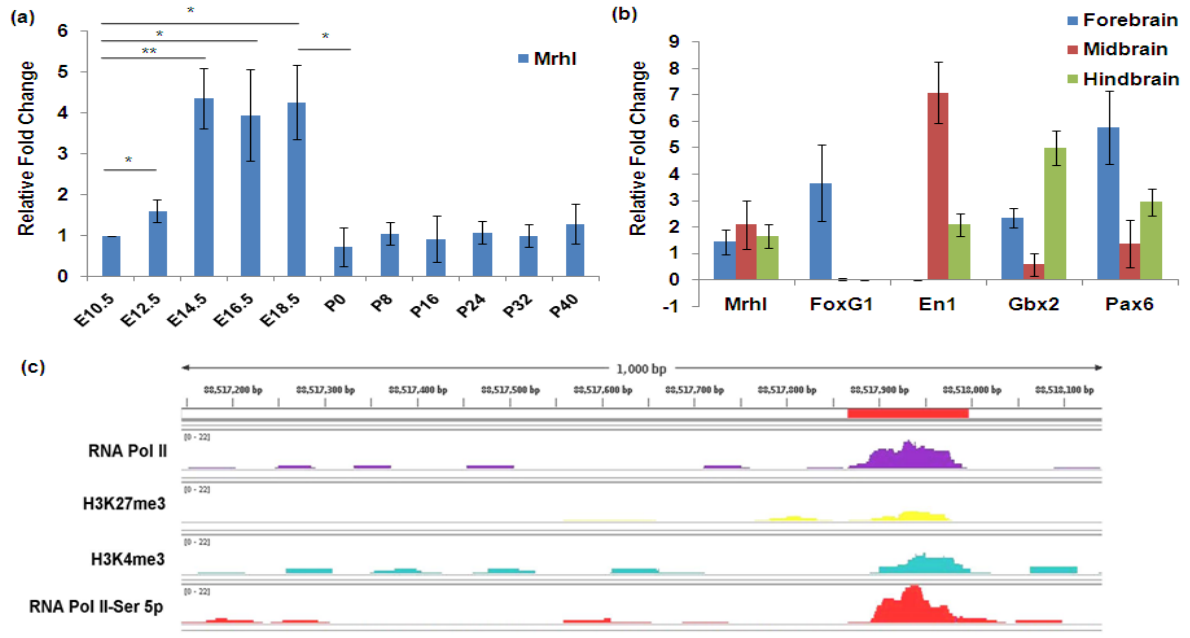


Fig. 4.2 *Mrhl is expressed in neural/ neuronal progenitors in mouse embryonic brain. (a) qRT-PCR analysis for Mrhl expression in various embryonic and postnatal brain stages. (b)*

Mrhl is expressed ubiquitously in fore-, mid- and hindbrain regions of E14.5 brain. (c) ChIP-Seq analysis revealed Mrhl promoter to possess significant enrichment for various transcription activation signatures.

*Error bars indicate standard deviation from three independent experiments. *p<0.05, **p<0.01, ***p<0.001, student's t-test.*

4.3.2 Mrhl shows increased levels of expression in retinoic acid derived neuronal progenitors with concomitant decrease in differentiated neurons: In order to delve deeper into the role of Mrhl in neuronal progenitors, we adopted an *in vitro* retinoic acid mediated differentiation method of mESCs for generating neuronal progenitors. These progenitors have been shown to possess characteristics of $PAX6^+$ RGCs and to generate neurons in the spinal cord upon implantation into the neural tube of chick embryos, emphasizing their properties as progenitor cells [139, 330]. In this context, we observed that Mrhl levels were significantly up regulated in RA treated EBs as compared to vehicle treated EBs (**Fig. 4.3 a**). An analysis for the expression of various markers revealed elevated expression of levels for neural stem cell/progenitor markers such as Nestin and Pax6 along with neuronal markers such as Ascl1 and Tuj1 in retinoic acid treated EBs (RA treated EBs) suggesting these progenitors are committed towards the neuronal lineage (**Fig. 4.3 b**). We validated our observations by performing RNA FISH for Mrhl followed by IF for NESTIN and TUJ1 in embryoid body sections (**Fig. 4.3 c**). Next, we differentiated the progenitors into neurons and analyzed them for the expression of Mrhl and various markers at different time points. Interestingly, Mrhl was observed to be down regulated by ~70-80% from early neuron stages (**Fig. 4.3 e**). It was observed that Ascl1 and Tuj1 expression remained high till 24 hours of neuron formation with Tuj1 being expressed robustly until day 2 and a down regulation for all progenitor markers from 12 hours of early neuron formation (**Fig. 4.3 f**). We validated out observations by RNA FISH and IF staining for progenitors at 0 hours to early neurons at 12 hours to late neurons at 24 hours and day 2 (**Fig. 4.3 g**). Our *in vitro* studies herewith corroborated with our previous *in vivo* analyses demonstrating that Mrhl is expressed predominantly in neuronal progenitors. Furthermore, ChIP-qPCR revealed that H3K4me3 is significantly enriched in the proximal promoter of Mrhl in RA treated EBs as compared to neurons at 24 hours (**Fig. 4.3 d**) illustrating Mrhl is indeed transcriptionally active in the progenitor populations during neuronal development.

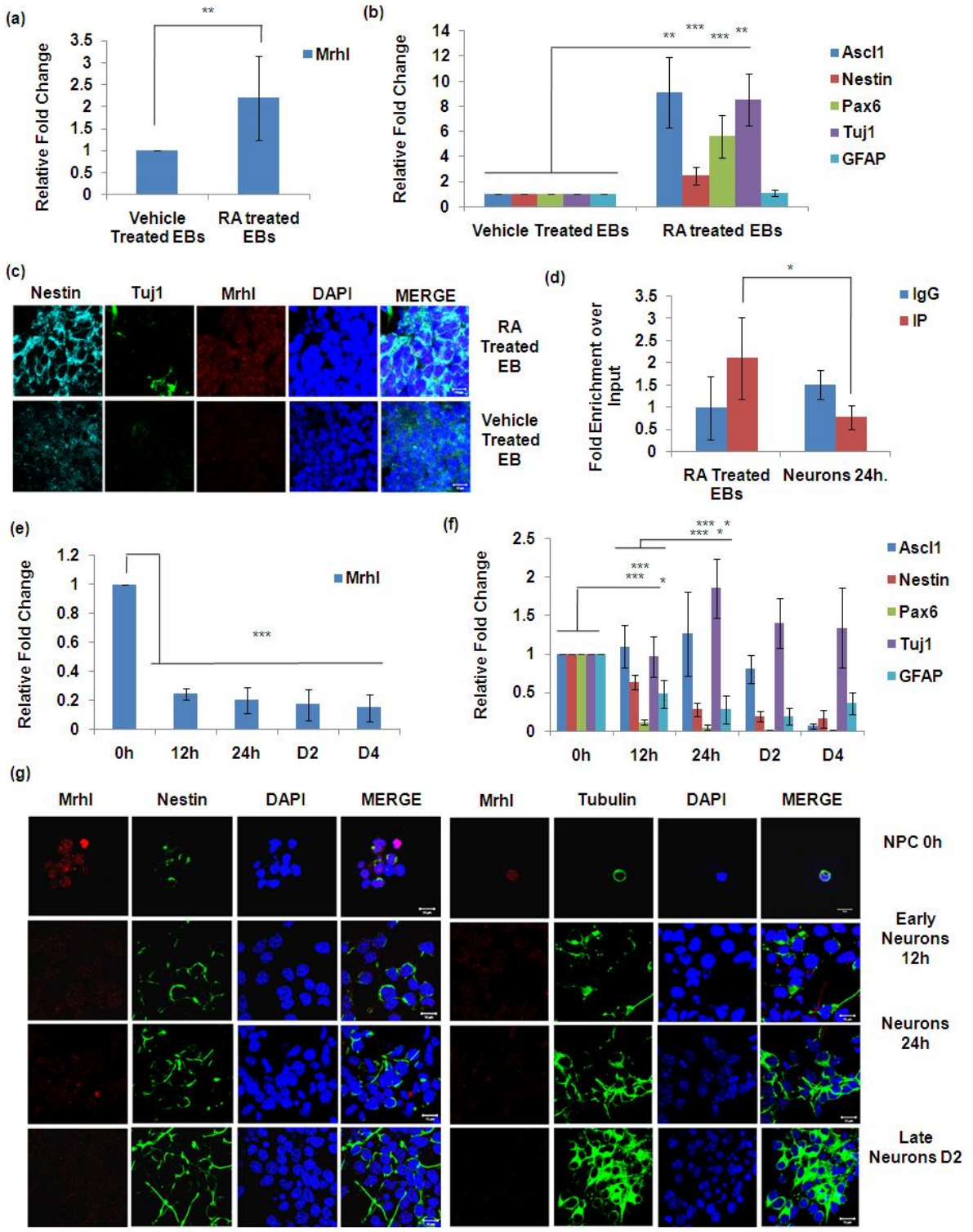


Fig. 4.3 *Mrhl is up regulated in expression in RA derived neuronal/RGC-like progenitors and down regulated in neurons. (a, b) qRT-PCR analysis for Mrhl and associated markers in RA treated EBs. (c) RNA FISH and IF shows increased levels for Mrhl in RA treated EBs. (d) ChIP-qpCR for H3K4me3 at proximal promoter of Mrhl demonstrates its transcriptionally active status in progenitors. (e) qRT-PCR revealed Mrhl to be down regulated in early and*

late neurons. (f) Analysis of various markers to demonstrate neuronal differentiation; (g) RNA FISH and IF validate aforesaid observations.

*Error bars indicate standard deviation from three independent experiments. *p<0.05, **p<0.01, ***p<0.001, student's t-test. Scale bar = 10μm.*

4.3.3 A master transcription factor, PAX6, might be involved in regulating Mrhl in neuronal progenitors: We next sought to understand the underlying mechanisms regulating the expression of Mrhl during neuronal development. Whilst it is important to understand the targets of a lncRNA during cellular differentiation and cell fate specification, it is also essential to address the regulators of lncRNAs themselves. For the former, the *cis* or *trans* methods of action of lncRNAs to regulate their targets in association with RNA binding proteins, transcription factors and chromatin modification complexes have been well established. In case of the latter, it has been shown that the chromatin state and/or methylation status of lncRNAs are important determinants of their expression and hence function. LncRNA Meg3 was shown to be down regulated in hepatocellular carcinoma due to hypermethylation of its promoter probably by the action of DNMTs [331]. Transcription factors that are involved in initiating a genetic program governing a cellular state have also been shown to act as regulators of lncRNAs. In a study by Cawley *et al* to map the binding of three TFs i.e., SP1, c-MYC and p53 across human chromosomes 21 and 22, it was revealed that whilst only 22% of the TF binding sites (TFBS) were present in the 5' region of protein-coding genes, 36% of the TFBS were located within or immediately 3' to such genes [332]. Interestingly, these regions were observed to correlate with non-coding RNAs that responded to common environmental cues and overlapping pairs of coding and non-coding genes were found to be co-regulated by common TFs or signals. In breast cancer cells, lncRNA Jade was found to be up regulated upon a DNA damage signal in a NF-κB dependant fashion [81]. The promoter of Jade harboured five putative binding sites for NF-κB and inhibition of this TF led to suppression of the DNA damage mediated activity of lncRNA Jade. Furthermore, in progenitor cells, lncRNAs have been implicated to act in concert with TFs to dictate cellular states [333]. They have been illustrated to not only repress differentiation programs along with progenitor cell specific TFs or facilitate the binding of TFs to their target genes, they have also been shown to act in positive feedback loops with cell specific TFs to exert their actions. In the context of male germ cell meiotic commitment, Mrhl has been shown to act as a negative regulator of WNT signaling in spermatogonial progenitors whereas during

differentiation, activation of WNT signaling leads the TCF4/β-CATENIN complex to repress Mrhl expression through recruitment of CTBP1 as the co-repressor at its promoter [64, 138]. To gain insights into the probable regulation of Mrhl by TFs in neuronal lineage development, we undertook a prediction based approach wherein we used a combination of GPMiner and JASPAR programs [334, 335] to obtain the predicted TFBS on Mrhl promoter. We analyzed 3 kb upstream of TSS of Mrhl as our input sequence and we observed potential binding sites for a host of TFs known for their roles in the development of various lineages and defining cellular states (**Fig. 4.4 a**). Of noteworthy importance were PAX6 (discussed earlier), TFC4 (WNT effector; discussed earlier), MEIS1 and NFAT due to their established roles in the neuronal lineage. The prediction also depicted the presence of RAR elements on the promoter. Furthermore, we also found a binding site for the NOTCH pathway effector RBPJ-κ near the TSS of Mrhl promoter (**Fig. 4.4 b**). MEIS1 and NFAT have been implicated for their roles in neuronal differentiation, generation of neurons and regulation of axon growth [336, 337]. The importance of retinoic acid signaling mediated through the binding of retinoic acid receptors to RAREs, in neuronal development, has been discussed in a previous section. An analysis of ChIP-Seq data and binding site for retinoic acid receptors [338] did not yield useful information with respect to regulation of Mrhl by RA (data not shown). Keeping these studies in mind, we focussed on the regulation of Mrhl by PAX6, TCF4 or RBPJ-κ given their indispensible roles in progenitor proliferation, their maintenance, commitment to neuronal lineage and brain development.

WNT pathway (TCF4/ β-CATENIN): Since the role of WNT signaling is highly context-dependant, we first addressed its status in neuronal progenitors and differentiated neurons. We performed IF for β-CATENIN on sections of vehicle treated and RA treated EBs and observed that β-CATENIN remained membrane localized and it did not exhibit nuclear translocation upon RA treatment in neuronal progenitors (**Fig. 4.5 a**). We concluded from these data that in RA derived neuronal progenitors, WNT pathway is not activated and hence is not involved in regulating Mrhl. Furthermore, since we have already established that in RA treated EBs/neuronal progenitors, Mrhl levels are high, this observation was in accordance with our earlier studies in male germ cell meiotic commitment wherein WNT has been shown to negatively regulate Mrhl.

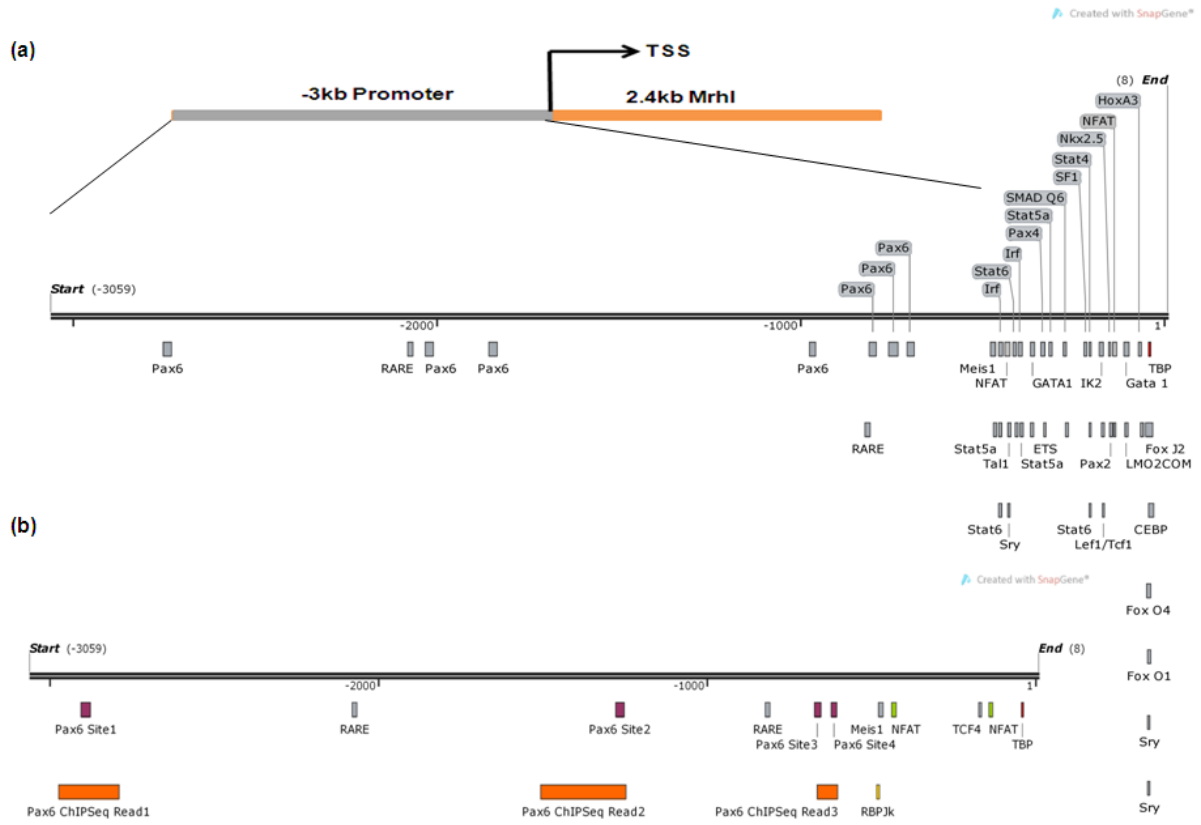


Fig. 4.4 *Promoter analysis of Mrhl for potential TFBS using GPMiner and JASPAR programs. 3 kb of Mrhl sequence upstream from the TSS was used as the input. (a) Depiction of all predicted TFBS. (b) Depiction of neuronal lineage related predicted TFBS and PAX6 ChIP-Seq reads and sites as found by analysis (described later in text).*

Next, we observed the status of β-CATENIN in early and late differentiated neurons. We hypothesized that down regulation of Mrhl in neurons might be due to the activation of WNT signaling as increased WNT pathway activity has been implicated in facilitating neuronal differentiation [312, 313]. We observed that from early neurons (8 hours to 16 hours) to later neuron generation (24 hours), β-CATENIN was predominantly localized to the membrane with no detectable nuclear translocation (**Fig. 4.5 b**). From these experiments, we concluded that WNT pathway is not involved in regulating Mrhl in the context of RA induced neuronal differentiation.

NOTCH Pathway (RBPJ-κ): Our promoter analysis revealed the presence of a potential binding site for RBPJ-κ close to the TSS and the role of NOTCH signaling in the specification and proliferation of RG progenitors has been discussed earlier. To address probable regulation of Mrhl by the NOTCH pathway, we first tested the status of NOTCH

activation in RA treated EBs. IF analysis revealed a predominant nuclear localization of NICD in RA derived neuronal progenitors (**Fig. 4.5 c).** To further delineate the probable regulation, we subjected RA treated EBs to the NOTCH pathway inhibitor DAPT and scored for the expression levels of its target Hes5 along with Mrhl. We observed that in RA+DAPT treated EBs, Hes5 was significantly down regulated implying inhibition of the NOTCH pathway but at the same time we did not observe any changes in the levels of Mrhl (**Fig. 4.5 d**). Interestingly, Pax6 and Tuj1 levels were also not altered upon NOTCH inhibition suggesting that NOTCH signaling might be required but dispensable in the context of RA derived neuronal progenitors. Our studies hereby revealed that NOTCH is not involved in regulating Mrhl in RA derived neuronal progenitors.

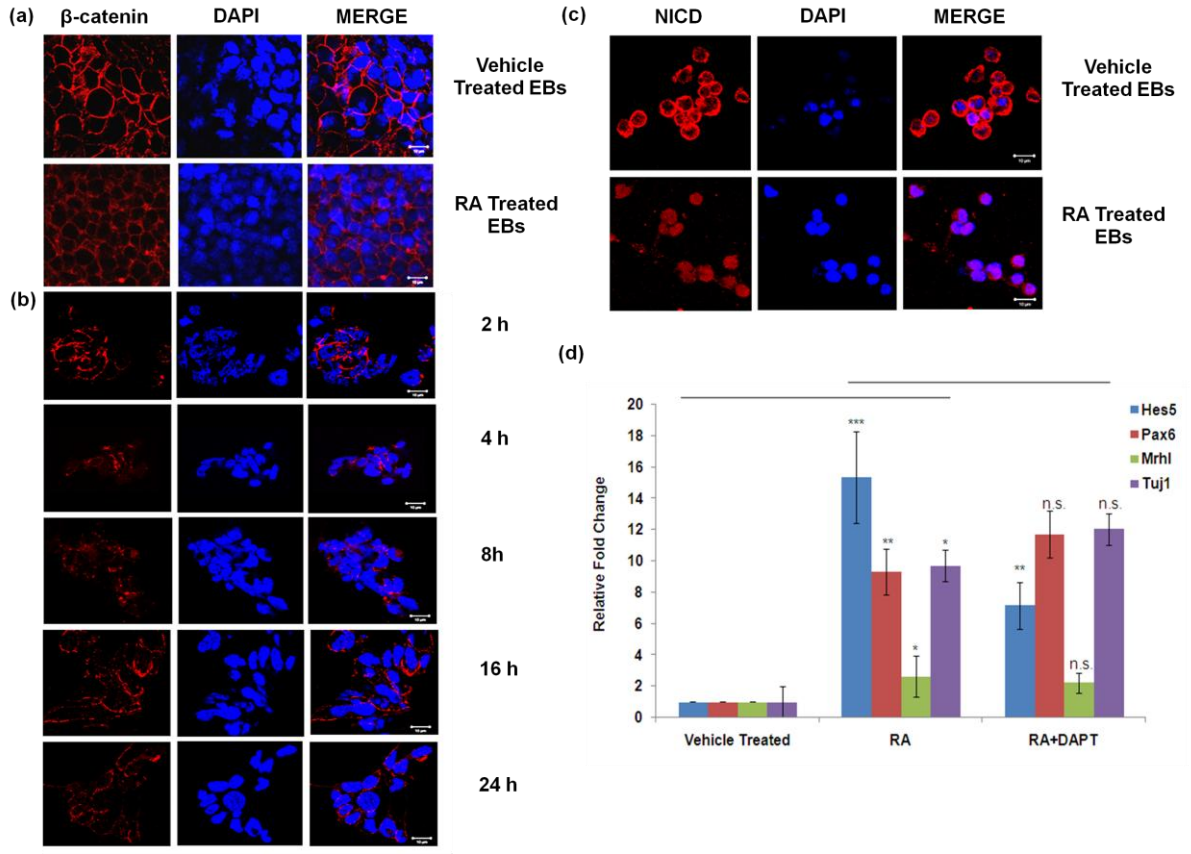


Fig. 4.5 *WNT and NOTCH pathways are not involved in Mrhl regulation during neuronal lineage development. (a) β-CATENIN did not show any nuclear translocation in vehicle treated versus RA treated EB sections. (b) β-CATENIN localization status remained same in early and late differentiated neurons. (c) NOTCH signaling is activated in RA derived neuronal progenitors as seen by NICD nuclear localization. (d) Inhibition of NOTCH pathway by DAPT does not affect Mrhl expression.*

*Error bars indicate standard deviation from three independent experiments. *p<0.05, **p<0.01, ***p<0.001, student's t-test. Scale bar = 10µm.*

PAX6: To understand the role of PAX6 in the regulation of Mrhl in neuronal progenitors, we first performed RNA FISH/IF analyses for Mrhl and PAX6 in E14.5 brain sections. We observed that in the VZ/SVZ, PAX6 and Mrhl were detected at high levels and they displayed robust co-expression in the nuclei (**Fig. 4.6 a**). Neither Mrhl nor PAX6 showed nuclear expression in the cortical plate (CP). We next analyzed available ChIP-Seq datasets for PAX6 and histone modifications (GSE 66961, [339]) in E12.5 mouse embryonic forebrain to correlate our promoter predictions with *in vivo* data. We observed that in the 3 kb upstream region of Mrhl, there were significant enrichments for PAX6 occupancy at five candidate sites (as have been indicated by red highlights in the scale bar, **Fig. 4.6 b**). Next, we mapped these reads for the presence of PAX6 binding motifs using FIMO. In a study by Xie and Cvekl, using SELEX (systematic evolution of ligands by exponential enrichment), PAX6 was shown to bind to nine novel motifs on the DNA with only one of the motifs being similar to the known consensus [153]. The isoform PAX6(5A) which has a 14 a.a. insertion in the PAI domain was also tested for binding to these sites. The study revealed that both PAX6 and PAX6(5A) bind with varying affinities to these motifs and emphasized on the heterogeneity and biological context dependant regulation of targets by PAX6/ PAX6(5A). FIMO analysis [340] of the Mrhl promoter for these nine motifs revealed the presence of multiple motifs in four out of the five candidate sites which we have termed as Sites 1, 2, 3 and 4 for further reference (**Fig. 4.6 c-e**). We observed that the site 1 harboured two overlapping motifs i.e., motif 1-1 and 2-1; site 2 also comprised two overlapping motifs i.e., 1-3 and 2-1 whereas sites 3 and 4 were situated in close proximity to each other and both harboured motif 4-1.

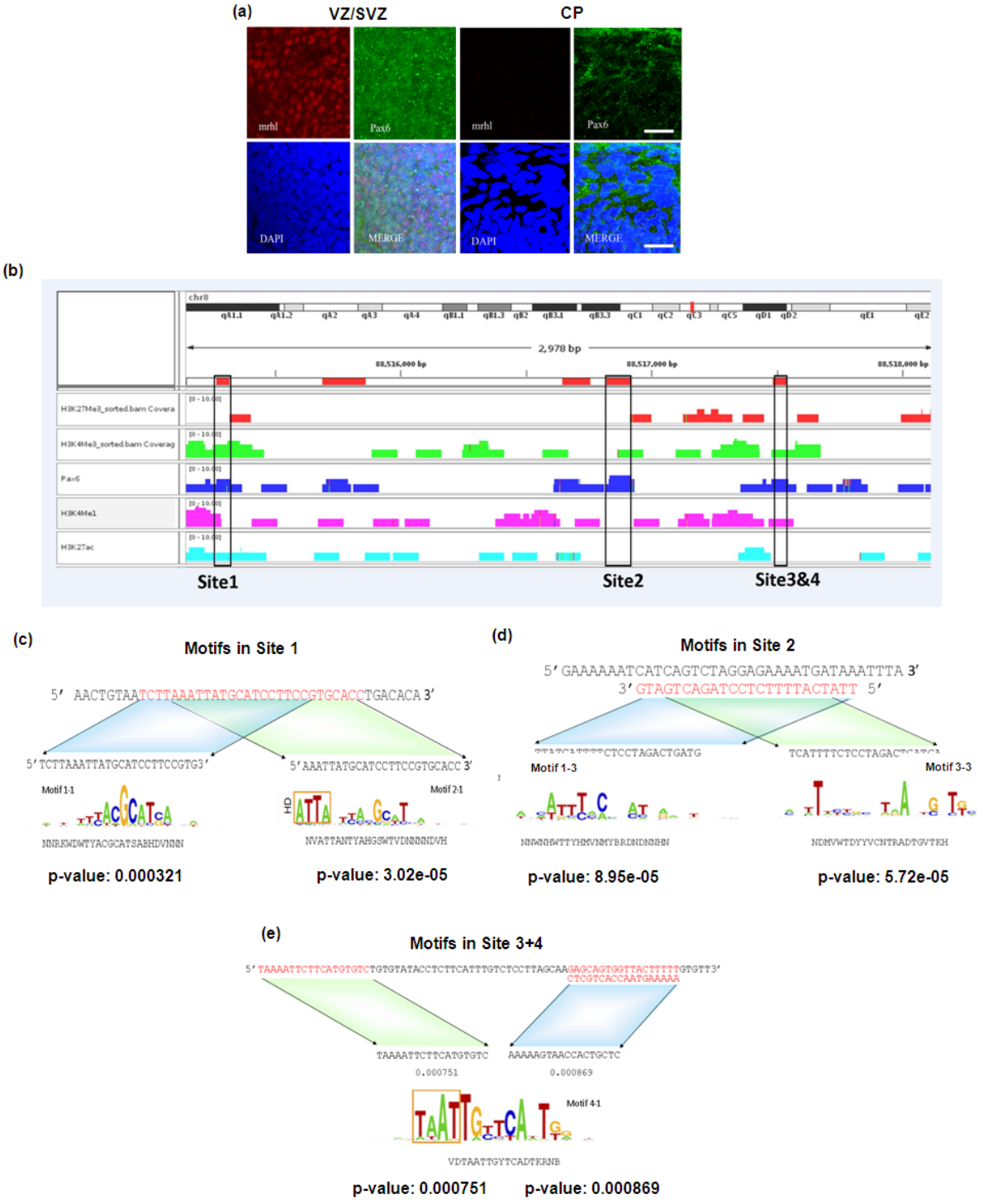


Fig. 4.6 *Correlation between PAX6 and Mrhl. (a) PAX6 and Mrhl display co-expression in VZ/SVZ regions of embryonic brain. (b) ChIP-Seq analysis for PAX6 and histone modifications on Mrhl 3 kb promoter sequence upstream of its TSS. (c-e) FIMO analysis revealed presence of various PAX6 motifs on Mrhl promoter. Scale bar = 20µm.*

We next carried out validation for sites 1 to 4 through multiple approaches to understand the functional regulation of Mrhl by PAX6 in neuronal progenitors. We isolated chromatin from E14.5 mouse embryonic whole brains and RA treated EBs/ neuronal progenitors and performed ChIP for PAX6. qRT-PCR analysis showed that sites 1 and sites 3/4 were occupied both in the *in vivo* and *in vitro* scenarios whereas site 2 was occupied only in the *in vivo* context (**Fig. 4.7 a-c**). This can be explained through the fact that TFs or proteins associating with PAX6 in the brain and in RA derived neuronal progenitors might be similar but not identical and *in vivo* additional factors might mediate/stabilize the binding of PAX6 to site 2. We next performed EMSA with purified mouse full-length PAX6/PAX(5A) proteins (**Fig. 4.7 d**) to analyze if PAX6 was binding directly or its binding was being mediated through other factors to these sites on the Mrhl promoter. We observed that PAX6 bound with strong affinity even at low protein concentrations to site 1. In site 2, we detected only very weak binding whereas sites 3 and 4 did not show any direct binding by PAX6 (**Fig. 4.7 e**). This led us to hypothesize that sites 2, 3 and 4 might be instead or additionally be bound by PAX6(5A) since it has already been demonstrated that motif 1-3 in site 2 and motif 4-1 in sites 3 and 4 are bound by PAX6(5A) with moderately high affinity [153]. We also tested PAX6(5A) binding to site 1 since motif 2-1 present at this site was shown to bind the 5A isoform. Furthermore, our hypothesis was supported by the study of Pinson *et al* in which it was shown that the ratio of Pax6:Pax6(5a) shifts from 8:1 to 3:1 between E12.5 and E14.5 of brain development [341]. However, we did not observe one to one binding for PAX6(5A) in any of the sites on the Mrhl promoter (**Fig. 4.7 f**). We concluded from these studies that site 1 is bound by PAX6 *in vivo* in the developing embryonic brain and *in vitro* in RA derived neuronal progenitors whereas PAX6/PAX6(5A) binding to sites 2, 3 and 4 could be mediated or stabilized by additional TFs or mediators. Evidence for other factors interacting and co-occupying target sites has been provided in lens wherein SOX2 has been shown to either bind to the same target promoters as PAX6 in neural progenitors [288] or form a co-DNA binding partner complex during lens development to activate enhancers of target genes [342].

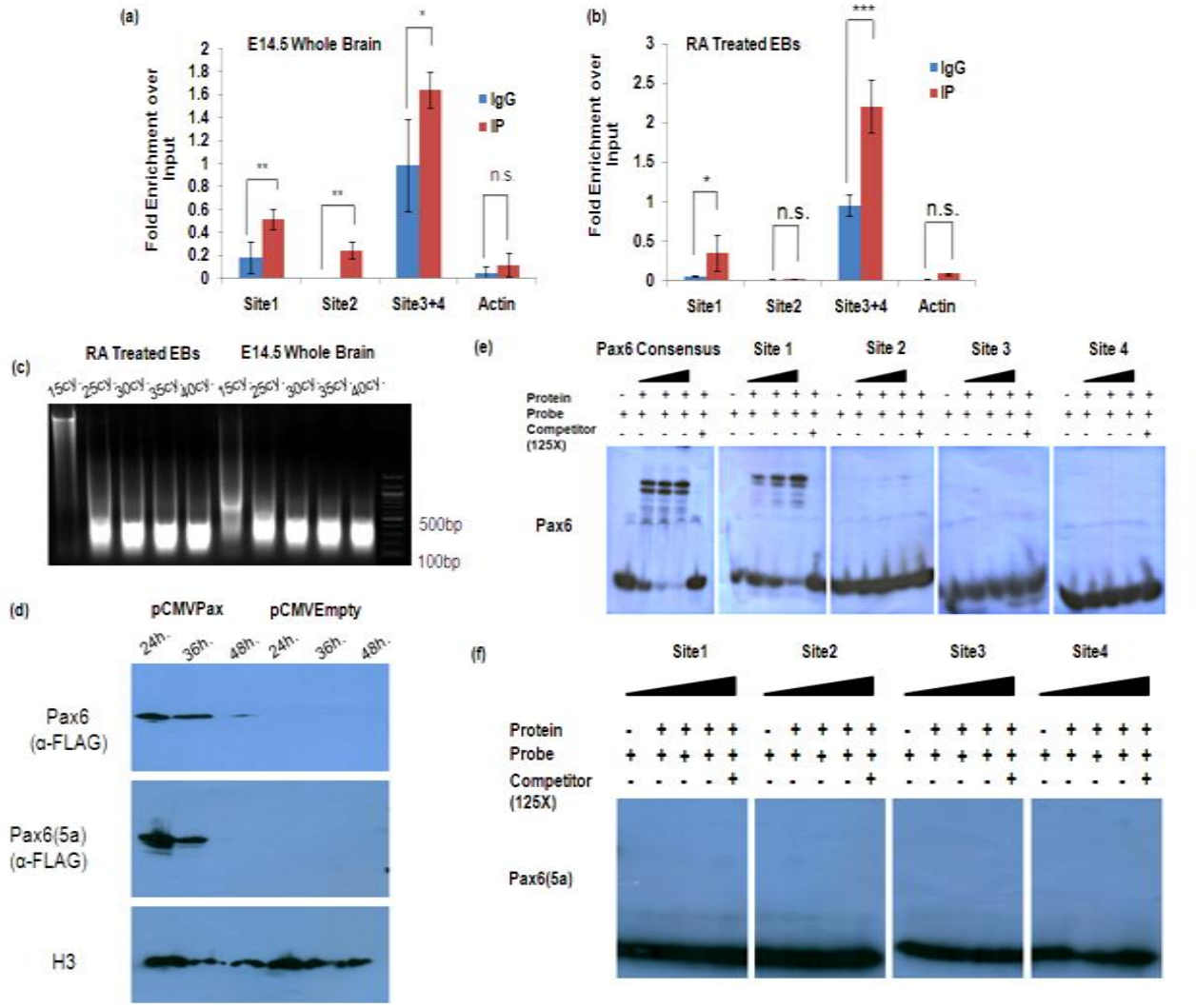


Fig. 4.7 *PAX6 physically interacts with Mrhl at its promoter. (a, b) ChIP-qPCR analysis revealed PAX6 occupancy at various sites on Mrhl promoter both in vivo and in vitro. (c) Sonication pattern for chromatin isolated from E14.5 whole brain or RA derived neuronal progenitors. (d) Western blot to check purified PAX6 and PAX6(5A) proteins. (e) EMSA for PAX6 consensus and Mrhl sites 1-4 with PAX6. (f) EMSA for Mrhl sites 1-4 with PAX6(5A).*

*Error bars indicate standard deviation from three independent experiments. *p<0.05, **p<0.01, ***p<0.001, student's t-test.*

We then performed luciferase reporter assays to delineate the functionality of PAX6/PAX6(5A) binding on the Mrhl promoter. We generated three different Mrhl promoter construct in pGL4 .10 vector. The first construct harboured sites 3 and 4; the second construct harboured sites 2, 3 and 4 and the third construct consisted of all the four sites. We then transfected P19EC cells with each of these constructs alone or in conjunction with mouse

PAX6 or PAX6(5A) expression vectors. We chose P19EC cells for our luciferase assays since it has been already proven to lack endogenous expression of PAX6 [153]. We observed that sites 3 and 4 did not elicit any luciferase signal either in the presence of PAX6 or PAX6(5A) (**Fig. 4.8 a**). In the presence of site 2, however, a significant luciferase signal was observed both in the presence of PAX6 and PAX6(5A) (**Fig. 4.8 b**). Lastly, in the presence of the site 1, we observed a highly consistent luciferase reporter activity induction by both the isoforms (**Fig. 4.8 c**). The results obtained in terms of regulation of Mrhl by PAX6 in neuronal progenitors have been summarized (**Fig. 4.9 a)** and a model for the same has been proposed based on the same (**Fig. 4.9 b**). We infer from our observations that the regulation by PAX6 of sites 3 and 4 on Mrhl promoter might be highly dependent on the context and pool of TFs/mediators present in the biological environment under investigation or, the occupancy (as revealed by ChIP-qPCR) in the *in vivo* or *in vitro* systems might be redundant and not functionally relevant. The latter theory would be supported by our reporter activity. Regulation at site 2 seems to be again governed by the associated TFs/proteins since along with physical occupancy we demonstrate a one to one weak binding by PAX6 and reporter activity by both PAX6 and PAX6(5A). Furthermore, the occupancy of site 2 only in the embryonic brain reinstates the context-dependant regulation and highly heterogeneous activity of PAX6 mediated regulation of target sites. Site 1, situated most upstream in the Mrhl promoter, stands as the functionally most relevant site of regulation of Mrhl by PAX6 in neuronal progenitors as revealed by our ChIP-qPCR, EMSA and luciferase assays. Even though the motif is present at this site (motif 2-1) for the binding of PAX6(5A), critical residues might be important for dictating a stable one to one interaction whereas such interactions might be guided and/or stabilized in association with other factors in cells. Furthermore, the histone modification profile for H3K27me3, H3K4me3, H3K27ac and H3K4me1 (**Fig. 4.6 b**) supports the fact that Site 1 is probably the functionally most relevant one since it has a very high enrichment for H3K4me3 transcription activation histone mark and is immediately adjacent to a site enriched for the enhancer mark H3K4me1. Studies on the epigenetic regulation of promoters and enhancers by PAX6 has revealed that PAX6 immunoprecipitates in lens contain the chromatin modifiers MLL1, MLL2 and SET1A which are histone methyl transferases and CBP and EP300 which are histone acetyl transferases [343]. Also, knockdown of PAX6 led to decrease in the abundance of H3K4me1 mark in distal enhancers and H3K4me3 mark in promoters at target occupied sites. The site 1 on Mrhl promoter can be classified as a distal promoter element due to an indicative enrichment of H3K4me3 or as a proximal enhancer since it is very adjacent to a region harbouring

H3K4me1 modification. Although mechanisms pertaining to PAX6 mediated regulation of its targets in the embryonic brain remains to be further delved into, it would be interesting to study if PAX6 regulates Mrhl at the epigenetic level. Our studies herewith have indicated that a master transcription factor, PAX6 is probably involved in regulating a lncRNA such as Mrhl during neuronal lineage development and proves to be a step forward towards understanding how lncRNAs integrate into the molecular circuitry in the context of progenitor cells, their maintenance/differentiation and fate specification.

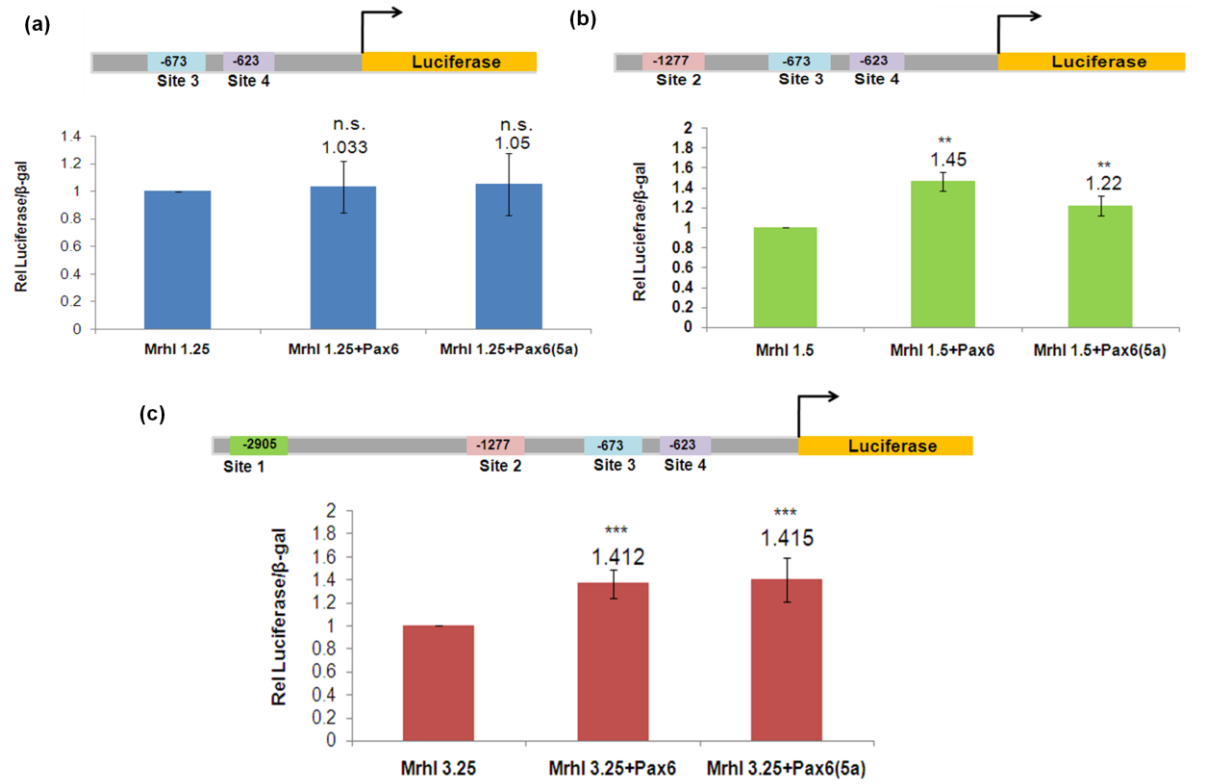


Fig. 4.8 *Luciferase reporter assays for Mrhl promoter (a) sites 3, 4, (b) sites 2-4 and (c) sites 1-4 with PAX6 and PAX6(5A). Promoter activity without any protein was kept as the control.*

*Error bars indicate standard deviation from three independent experiments. *p<0.05, **p<0.01, ***p<0.001, student's t-test.*

(a)

Motif	Site on Mrhl Promoter	ChIP-qPCR	EMSA	Luciferase Assay
1-1 & 2-1	1 (-2905bp)	+ in vivo & in vitro	Strong binding for Pax6	+ for both Pax6 & Pax6(5a)
1-3 & 3-3	2 (-1277bp)	+ in vivo - In vitro	Weak Binding for Pax6	+ for both Pax6 & Pax6(5a)
4-1	3+4 (-673 -623bp)	+ in vivo & in vitro	-	-

(b)

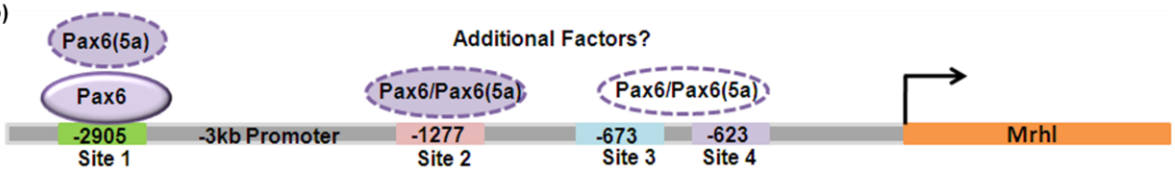


Fig. 4.9 *Illustration of probable PAX6-Mrhl dynamics in neuronal progenitors. (a) Summary of results obtained from ChIP-qPCR, EMSA and luciferase reporter assay. (b) Proposed model for potential regulation of Mrhl by PAX6 during neuronal lineage development.*

4.4 Summary: The study discussed in this chapter was aimed at deciphering the role of lncRNA Mrhl in the context of neuronal lineage development. Our transcriptome studies on mESCs unravelled the potential role of Mrhl in regulating developmental processes, especially genes and processes related to various aspects of neuronal development such as survival of progenitors, neuronal cell fate specification, differentiation of progenitors into neurons, neuronal migration and axon development and synapse formation. Hence we adopted both *in vivo* and *in vitro* approaches to further delineate the functional relevance of Mrhl during neuronal lineage development. We performed our initial studies and analysis in mouse embryonic brains which provided us with an indication of a potential role of Mrhl in neural/neuronal progenitors. Our *in vitro* retinoic acid mediated neuronal commitment of mESCs corroborated our *in vivo* observations. We next addressed the regulators of Mrhl in neuronal progenitors and our studies revealed a master neuronal progenitor/proneural TF, PAX6 to be a probable regulator of lncRNA Mrhl. While additional studies are being targeted towards understanding if Mrhl is necessary for the maintenance of these progenitors and/or their differentiation and whether its down regulation in neurons is a cause or a consequence, the current studies emphasize on how lncRNAs integrate into existing molecular circuits of progenitor cells and are themselves regulated by major TFs or signaling pathways during development and/or differentiation of progenitors.

The major findings and implications of this study are as follows:

- LncRNA Mrhl shows temporal specific expression during mouse embryonic brain development with its expression peaking at stages E14.5-E18.5. It is expressed ubiquitously across fore-, mid- and hindbrain regions of E14.5 brain. Available ChIP-Seq data analysis indicated its predominant expression in neural/neuronal progenitors.
- Retinoic acid treatment of EBs generated neuronal lineage committed progenitors equivalent to RGCs. Mrhl is up regulated in expression in these progenitors with a drastic down regulation in expression in early and late differentiated neurons.
- A promoter prediction analysis to understand the potential regulator(s) of Mrhl in neuronal progenitors revealed predicted binding sites for diverse TFs. TCF4, RBPJ-κ and PAX6 appeared as the interesting candidates. IF studies for β-CTAENIN which is the coactivator of TCF4 in the WNT pathway showed that WNT signaling is not activated in RA derived neuronal progenitors/neurons. Further studies showed that NOTCH pathway is activated (nuclear translocation of NICD, the coactivator of RBPJ-κ) but inhibition of the pathway did not affect Mrhl expression in RA derived neuronal progenitors.
- Analysis of available ChIP-Seq data suggested PAX6 as a probable TF involved in regulation of Mrhl. FIMO analysis revealed four sites on 3 kb sequence of Mrhl, upstream of its TSS, to possess motifs for PAX6. Validations by ChIP-qPCR for physical occupancy, EMSA for one to one binding and luciferase reporter assays for functional occupancy revealed site 1 (situated at -2905 bp) on Mrhl promoter to be functionally most relevant. The study also suggested that site 2 (situated at -1277 bp) is probably a functional one too with the proximal sites 3 and 4 (situated at -673 and -623 bp respectively) being redundant or context-dependant. Furthermore, there could be a possibility of regulation of these sites by PAX6 or its isoform PAX6(5A) or both in neuronal progenitors.

4.5 Future Perspectives

- An inducible knockdown line for PAX6 in mESCs has been generated to induce PAX6 depletion in RA treated EBs and confirm the regulation of Mrhl by PAX6.
- The mechanisms behind PAX6 mediated regulation of Mrhl would be interesting to address. Site 1 on Mrhl promoter is enriched for H3K4me3 histone mark and is also immediately adjacent to a region exhibiting H3K4me1 enhancer modification. PAX6

is known to interact with MLL group of methyl transferases in lens and it would be interesting to study if PAX6 exerts epigenetic regulation of lncRNA Mrhl.

- The mechanism of down regulation of Mrhl in mature differentiated neurons and whether this phenomenon is necessary for the formation of mature neurons or its just a consequence also needs to be addressed.
- It would be further interesting to study the targets of Mrhl in neuronal progenitors to gain insights into the molecular and biological pathways that it regulates in these progenitors. In this regard, the generation of an inducible knockdown line for Mrhl in mESCs which can be subjected to RA mediated neuronal commitment followed by transcriptome studies in Mrhl knockdown conditions in neuronal progenitors is being undertaken.

Chapter 5

Gaining insights into the cellular/biological phenotypes upon knockdown/knockout of lncRNA Mrhl

5.1 Introduction

5.1.1 RNAi-mediated loss-of-function approaches

Traditional approaches to understand the function of a gene was based on isolation of mutants of interest followed by identification of the genes/proteins responsible for the phenotype under study. Such genetic screens are the approaches in forward genetics. In reverse genetics, the starting point is the gene under study itself which is made to undergo some mutations or is depleted from the system using definitive approaches to understand the cellular phenotypes in absence of a functional form of the gene. The forward genetics approach took scientists decade to fish out functional genes from a vast pool of the same. The discovery of RNA interference by Fire and Mello revolutionized the study of functions for a gene of interest. A gene product/RNA could now be easily targeted with synthetically generated siRNAs or shRNAs and the resulting phenotypes in the cellular context be addressed to assign the gene a function.

A vast number of lncRNAs have been characterized functionally based on siRNA or shRNA-mediated loss-of-function studies. In one of the earlier studies by Mohamed *et al*, siRNA mediated knockdown of lncRNA AK028326 (Gomafu/Miat) led to the reduction in expression levels of the pluripotency factors Oct4 and Nanog along with a significant increase in the levels of mesodermal and trophectodermal markers [199]. Interestingly, the lncRNA was itself regulated by Oct4. The loss-of function studies not only revealed that lncRNA AK028326 was important in the maintenance of embryonic stem cells but also demonstrated a probable positive auto-feedback loop between Oct4 and the lncRNA. Knockdown of lncRNA St8Sia3 in human ES cells and iPSCs led to a decrease in the growth of the cells associated with elevated apoptosis [101]. The lncRNA was named as Linc-Ror or regulator of reprogramming for its essential roles in the maintenance of iPSCs.

Stable knockdown lines for lncRNA Braveheart in mESCs generated by shRNA transduction, showed no change in morphology of the cells, expression of pluripotency markers, cell cycle kinetics or apoptosis levels although Braveheart depleted ES cells showed a marked decrease in generation of cardiac tissue and lower levels of expression for cardiac troponin, establishing its importance and specificity in cardiac lineage development [31]. In the context of endoderm specification, lncRNA Deanr1 (definitive endoderm associated lncRNA1) was implicated to be functionally relevant through similar approaches. shRNA mediated generation of stable lines for lncRNA Deanr1 in human ES cells revealed impaired

differentiation of the cells towards endoderm lineage with lower number of CD184/CD117$^+$ cells (definitive endoderm markers) and lower expression of endodermal marker genes such as FoxA2 and Sox17 thereby demonstrating the role of lncRNA Deanr1 in the specification of the endoderm lineage [344].

LncRNA Pnky was shown to be expressed in the VZ/SVZ region of adult mouse brain. Knockdown of Pnky in neural stem cell (NSC) monolayer cultures showed the cells to be positive for NESTIN, SOX2 and GFAP. However, when subjected to neuronal differentiation, the knockdown cells showed an aberrant increase in the number of TUJ1$^+$ neurons as compared to control cells. Further analysis of the knockdown cells revealed that they possessed 2.74 times more number of neuroblasts than control cells and underwent a marked increase in number of divisions of the progenitors before differentiating into neuroblasts, although the length of the cell cycle remained constant. These phenotype studies for Pnky knockdown showed that lack of expression of Pnky led to increased cell divisions of progenitors, enhanced differentiation into neurobalsts and hence an increase in terminal neuronal differentiation. In the mouse embryonic brain, electroporation of VZ cells at E13.5 with Pnky shRNA construct caused a decrease in SOX2$^+$ NSCs and an increase in SATB2$^+$ young neurons. The *in vitro* and *in vivo* loss-of-function studies for lncRNA Pnky implicated the importance of Pnky in maintenance of SOX2$^+$ NSCs, regulating the proportion of neurons in embryonic and adult brains and in controlling the phenomenon of neuronal differentiation of NSCs [204].

In a study by Ulitsky *et al* [345], RNAi for several lncRNAs in zebrafish with the help of morpholino antisense oligos revealed their importance in the development of various aspects of the nervous system. Knockdown of lncRNA Cyrano in zebrafish embryos led to formation of small eyes and heads, short curly tails (probable function of the lncRNA in notochord), defects in neural tube opening, loss of NEUROD$^+$ neurons in the retina and tectum and enlarged nasal placodes. LincRNA Megamind was also phenotypically characterized using morpholinos in this study. Morphants had an expanded midbrain ventricle, contraction of the forebrain ventricle, smaller heads and eyes by 48 hpf and loss of NEUROD$^+$ neurons in the retina and tectum. The RNAi induced loss-of-function studies helped understand the importance of lncRNAs in embryonic and brain development. Interestingly, these lncRNAs displayed syntenic conservation with deep conservation of short stretches of sequences across mammals and supplementing morphants with the mouse or human mature lncRNAs resulted in rescue phenotype.

5.1.2 Gene knockout/genome editing mediated loss-of-function approaches

Approaches for lncRNA knockouts: The area of interrogating functions of lncRNAs through knockouts or genome editing has taken a huge leap from Cre/LoxP mediated gene deletions to genome editing through CRISPR. Knockouts of lncRNAs can be generated through any of the approaches as mentioned below (reviewed in [346]):

- Total deletion of the gene body of the lncRNA
- Promoter deletion of the lncRNA
- Insertion of transcription termination sequences into an exon of the lncRNA

In the following section, the challenges for these strategies and related considerations for generating lncRNA knockouts have been discussed.

Challenges and considerations for lncRNA knockouts: One of the major challenges with lncRNAs as compared to protein-coding genes is that in case of the latter, coding mutations and indels often lead to alterations in the functional properties of the protein and hence the phenotype. However, in case of lncRNAs, such small-scale mutations or small stretches of insertions and deletions might not result in a drastic change in their functional properties since lncRNAs require lower degree of sequence conservation and depend more on domains of structure for their functions. Hence, to understand the functional relevance of lncRNAs through genome editing approaches, it often becomes essential to perform deletions of large stretches such as the promoter region of the lncRNA or the gene body itself. This poses a major challenge because many lncRNAs regulate their neighbouring genes in *cis* through sequence-based chromatin looping mechanisms or by acting as enhancer RNAs. Hence, such large-scale deletions for lncRNAs possess the capability to alter the functions of the neighbouring genes or host genes (in case of intronic lncRNAs). This turns out to be a very important consideration whilst generating lncRNA knockout or genome edited models because it is crucial not to disrupt the parent gene function to attribute visible phenotypes to the lncRNA deletion under study. Often rescue experiments are inevitable to obtain such insights.

A second consideration arises when in the case of certain lncRNAs, no visible phenotypes are observed in knockout animal models although their essentiality *in vitro* under cell culture conditions has been well established. This might be due to functional redundancy or compensatory mechanisms of action operating in the animal during development. Since

lncRNAs are mostly involved in fine regulation of their target loci, *in vivo* the scenario might be very different with other molecular players taking over the action of the lncRNAs under study. LncRNA Malat-1 has been extensively characterized with respect to its function in nuclear speckles that are rich in pre-mRNA splicing regulators and splicing factors. Malat-1 has been shown to regulate synaptogenesis in cultured hippocampal neurons [347], its lack causes dysregulation of splicing events [67, 348] and it has also been implicated in regulating migration in several human cell lines through controlling motility-related genes [184, 349]. However, knockout mice of *Malat-1* generated through insertion of poly (A) signal 69 bp downstream from its TSS were shown to be viable and fertile with no apparent dysregulation in the formation of nuclear speckles [350]. In an independent study, a ~3 kb deletion from the 5' end of *Malat-1* was performed to generate the knockout mice [351]. This study also revealed similar results where *Malat-1* knockout mice did not show any changes in the formation of nuclear speckles or global splicing events. A closer inspection showed no changes in gene expression at the genome-wide level with only local genes, situated in *cis* with respect to *Malat-1*, being perturbed in expression. Interestingly, Neat1 was shown to be up regulated in expression in various Malat-1 knockout tissues, suggesting a possible compensatory mechanism.

For the knockout of lncRNA Hotair, two independent studies were carried out. In the study by Li *et al* [352], the entire gene body of *Hotair* was deleted through homologous recombination using the Cre/LoxP methodology. *Hotair* knockout mice were viable and fertile but were shown to undergo malformations in the spine and metacarpal-carpal bones with global de-repression of genes located in the *HoxD* locus including *Hoxd10*, *Hoxd11* and *Hoxd13*. Several other imprinted genes in the *Dlk1-Meg3* and *Igf2-H19* loci were also seen to undergo de-repression in the knockout mice which was in corroboration with in vitro studies wherein Hotair has been shown to interact with PRC2 and LSD1 chromatin modifier complexes to impart repressive histone modifications at target genes. In the study by Amandio *et al* [353], however, a full-length deletion of *Hotair* did not lead to any skeletal malformations or gene de-repression in the knockout animals. The apparent discrepancy could be attributed to the fact that in the first study, the knockout was generated in the C57BL/6 genetic background whereas in the second study, the knockout was generated in CBAxBL6 genetic background. Furthermore, in the former case, the global gene analysis was performed in tail tip fibroblasts from newborn mice whereas in the latter study, it was performed in E12.5 embryo fragments [72]. These studies on Hotair knockout mice

emphasize the stage-specific temporal and spatial regulatory mechanisms of action of lncRNAs.

Many lncRNAs have been shown to be essential for the development and viability of animals through knockout mice models. A study towards generating multiple lincRNA knockout mice models for 18 lincRNAs revealed peri- and post-natal lethality for three of the lincRNAs namely, Fendrr, Peril and Mdgt [321]. For *Peril* and *Mdgt*, incomplete penetrance and growth defects were observed for the mice that survived. *Brn1b* knockout mice displayed defects in the formation of the cortex, specifically in the generation of upper layer II-IV neurons in the neocortex. LincRNA Brn1b was shown to be expressed in the neural progenitors of the ventral and dorsal telencephalon using galactosidase assays since the knockout strategy was performed to replace the *Brn1b* locus with a lacZ cassette. At E15.5, it was shown to be expressed in the neural progenitors of the VZ/SVZ in the developing cerebral cortex with its expression being restricted to the upper layer neurons at E18.5. Postnatally, it was shown to be expressed in the primary somatosensory and the visual cortices at P7. The knockout model for lincRNA Brn1b served as an excellent model for studying the role of this lincRNA in brain development. Apart from its role in the generation of upper layer neurons, it was also shown to be involved in the formation of intermediate progenitors (or IPs) in the dorsal cerebral cortex with no effect on the generation of apical progenitors (or APs) in the VZ. But at the molecular level, RNA-Seq analysis revealed reduction in the expression levels of its neighbouring *Brn1* gene and up regulation of the *Brn1a* locus which shares a bidirectional promoter with *Brn1b* suggesting *cis* regulatory mechanisms for this lincRNA. As mentioned earlier, suitable rescue experiments would further aid in attributing such phenotypes to lincRNA Brn1b alone and not to the perturbation of its *cis* loci.

For *Fendrr* knockout mice, knockout via lacZ replacement of the loci led to perinatal lethality with defects in the respiratory system along with esophagus and gut development. Interestingly, its neighbouring gene *Foxf1a* or any genes within +/- 1 Mb of the *Fendrr* locus was not seen to be perturbed in expression, establishing the importance of this lincRNA in development. However, a different strategy to generate *Fendrr* knockout mice by replacement of the first exon with multiple transcription termination sequences resulted in lethality at E13.75 due to defects in the heart and body wall [30]. Furthermore, in the latter study, rescue experiments with a single allele of *Fendrr* in the mutant embryos containing the terminated locus showed that *Foxf1* expression level was unchanged in rescued E9.5 embryos, the embryos were phenotypically normal until E17.5 while at E18.5 rescue was

observed in majority of the embryos thereby, attributing the knockout phenotypes solely to *Fendrr*. Although in both the cases, the genetic background of the mice was same, the strategy used to generate the knockouts could have led to the differences in phenotypes, emphasizing the challenges for generating lncRNA knockouts.

LncRNA Tsx is expressed from the X-inactivation center that is replete with various lncRNAs contributing towards the phenomenon of dosage compensation [214]. It was also found to be expressed very highly in the female brain with 10-100 times higher expression in the male gonads as compared to the male brain. A knockout strategy to delete the first exon and a portion of the upstream regulatory sequence at the *Tsx* locus resulted in viable and fertile knockout mice. -/- females when crossed with -/Y males or wild-type males displayed slightly skewed sex ratio with a preference towards generating female pups along with a lower mean litter. In case of males, although there were no observable differences in fertility, mutant males had a much smaller testis size with elevated apoptosis levels at 7days, 14 days and 2 months with the apoptosis rate being the most notable at 1 year. The observations suggested a loss of germ cells as the animals proceeded through the first meiosis. However, when -/Y males were crossed to wild-type females, litter size and sex ratios were not skewed leaving such observations unexplainable. It was also shown that mutant mice had less fear and enhanced short-term hippocampal memory. Furthermore, ES cells harbouring Tsx deletion displayed severe defects in growth, differentiation and elevated apoptosis. Thus, lncRNA Tsx knockout studies revealed multiple roles for this lncRNA in stem cells, male germ cell development and in learning and behaviour although the defects in stem cells observed in culture were not translated *in vivo*, suggesting compensatory mechanisms.

Based on the above studies on lncRNA knockouts, the following summary can be concluded:

- LncRNAs can have well-established functions *in vitro* but they might not be recapitulated *in vivo* due to compensatory mechanisms as is in the case of Malat-1.
- LncRNAs knockout phenotype can be highly dependent on the genetic background of the mice like in the case of Hotair.
- Knockout of certain lncRNAs that function through *cis*-regulation of their target genes or that share promoters or regulatory sequences with adjacent genes can lead to dysregulation of their neighbouring genes making it difficult to attribute the knockout phenotypes solely to the lncRNAs as in the case of Brn1b. In such cases, additional rescue experiments need to be performed to address the issue.

- The targeting strategy can lead to notable differences in resultant phenotypes in knockout mice as is the case for lncRNA Fendrr.
- Knockout studies, however, can provide essential insights into the roles of lncRNAs in regulating developmental and behavioural aspects as is in the case of lncRNAs, Peril, Mdgt, Fendrr and Tsx, to mention a few.

5.2 Rationale of Current Study

Our studies herewith on lncRNA Mrhl in mESCs and neuronal lineage development have provided us with potential roles for the lncRNA in regulating developmental processes. In mESCs, knockdown followed by transcriptome and systems studies revealed developmental processes consisting of transcription factors, cell adhesion and receptor pathway related genes from various lineages to be differentially expressed. Gene clustering and TF interaction studies also revealed developmental related or lineage-specific networks to be functionally enriched. A significant proportion of such processes and DEG belonged to the neuronal lineage. Following this observation, a probe into the relevance of Mrhl in neuronal lineage and brain development showed its predominance of expression in neural/neuronal progenitors of the mouse embryonic brain or *in vitro* derived RGC-like progenitors with a drastic lowering of its expression levels in the postnatal brains or differentiated neurons *in vitro*. The potential regulation of Mrhl by a neuronal progenitor master transcription factor PAX6 further established the integration of lncRNAs functionally into existing circuits of cells. With these molecular studies in mind, we aimed at understanding the resultant cellular phenotypes upon its knockdown in mESCs, its role in regulating differentiation given the dysregulation of major lineage-specific genes upon its transient knockdown in mESCs, how it affects the neuronal commitment program and whether it is necessary for the formation and/or maintenance of neuronal progenitors. Furthermore, we also wanted to study its role in the *in vivo* scenario through generation of a knockout mouse model in the context of its well-established roles in fine-tuning meiotic germ cell commitment and its potential roles in developmental and differentiation in mESCs and in neuronal lineage development.

(All materials and methods have been detailed in Chapter 2).

5.3 Results

5.3.1 A stable knockdown line for lncRNA Mrhl in mESCs does not show changes in pluripotent characteristics: We generated a stable knockdown line for Mrhl in mESCs using shRNA 4 (sh.4) construct in the pLKO.1 vector harbouring puromycin cassette as the eukaryotic selection marker. mESCs were transduced with viral particles containing either sh.4 or scrambled shRNA (scr.) and positive colonies were selected with puromycin for a week. qRT-PCR analyses showed that there was an average down regulation of ~50% in the Mrhl sh.4 mESCs as compared to scrambled control cells (**Fig. 5.1 a**). There was also no significant change observed in the levels of the pluripotency markers Oct4, Sox2 and Nanog (**Fig. 5.1 b**). This was further confirmed at the protein level by western blotting and IF studies for OCT4 (**Fig. 5.1 c, d**). Next we performed alkaline phosphatase assay to further confirm the 'stemness' of the stable knockdown cells and observed that there was no quantifiable difference between Mrhl sh.4 and scrambled control cells in alkaline phosphatise activity (**Fig. 5.1 e**). We also checked for changes in adhesion properties of the knockdown cells by observing the status of F-ACTIN (phalloidin staining) and E-CADHERIN. E-CADHERIN is known to be an important cell signaling modulator for pluripotency in ES cells and changes in F-ACTIN levels would be indicative of changes in mESC colony morphology or cell-cell or cell-matrix interactions and hence changes in ES cell characteristics [354-356]. IF studies followed by quantification for F-ACTIN (**Fig. 5.1 f**) and IF studies for E-CADHERIN (**Fig. 5.1 g**) revealed no significant changes in theIR expression or organization in knockdown cells as compared to scrambled control cells. Thus, from these studies we conclude that stable knockdown of Mrhl in mESCs does not lead to a change in the pluripotent status or ES cell characteristics. These observations are in concordance with our RNA-Seq studies in Mrhl transient knockdown cells where we did not observe any changes in pluripotent factors/pathways.

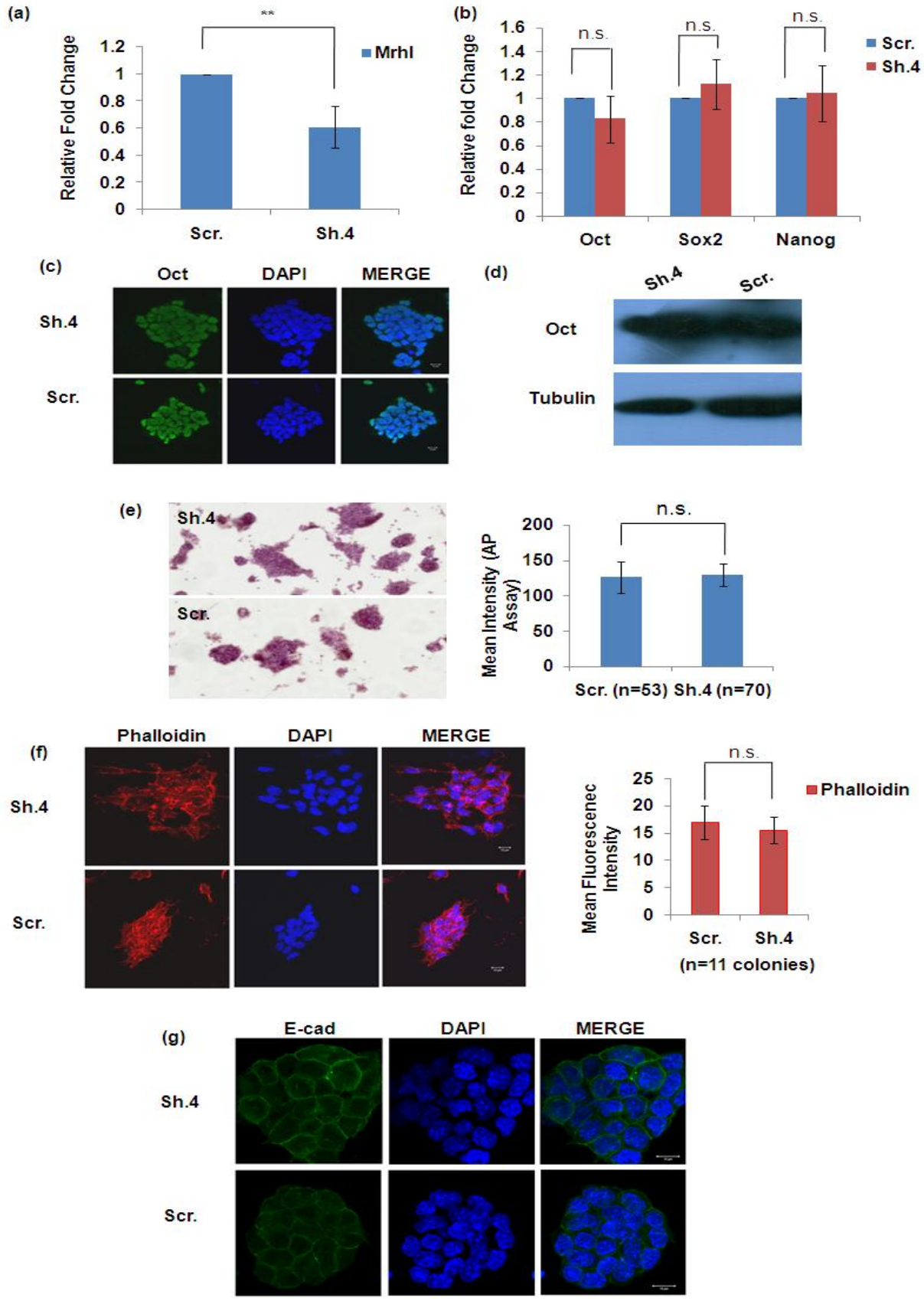


Fig. 5.1 *Stable knockdown of Mrhl in mESCs does not affect its pluripotent status. (a) qRT-pCR validation for knockdown status of Mrhl. (b) Levels of pluripotency markers in Mrhl sh.4*

stable cells as compared to scrambled control. (c, d) Validation of OCT4 expression through IF and WB respectively. (e) Alkaline phosphatase assay did not reveal changes in pluripotent status of knockdown cells. (f, g) F-actin (phalloidin) and E-cadherin staining revealed no changes in cell adhesion properties in knockdown cells.

*Error bars indicate standard deviation from three independent experiments. *p<0.05, **p<0.01, ***p<0.001, student's t-test. Scale bar = 10μm.*

5.3.2 EB differentiation of Mrhl knockdown cells showed aberrance in expression of lineage markers: Our RNA-Seq studies revealed that knockdown of Mrhl causes differential expression of various lineage-specific and differentiation related genes. At the same time, pluripotency markers do not undergo any changes in expression which is counter-intuitive. Furthermore, our earlier data suggest that EB differentiation leads to an up regulation in Mrhl levels at days 4 and 6. A plausible hypothesis in this regard would be, under conditions of Mrhl knockdown, mESCs develop from a naive state to a primed state which would be more responsive to appropriate differentiation signals. Hence, in order to understand the implications of these observations, we subjected Mrhl sh.4 stable knockdown cells to general differentiation via the formation of EBs and compared the development of various lineages with respect to scrambled control cells. A preliminary observation of the EBs under the microscope revealed a distinct difference in the size probably pertaining to differences in growth rate (**Fig. 5.2 a**). Till day 4 of differentiation, the EBs appeared similar in size. At days 6 and 8, Mrhl sh.4 EBs appeared to be larger in size as compared to the scrambled control counterparts. Our earlier transcriptome studies in Mrhl transient knockdown mESCs and analysis of the PANTHER process GO: 0032502 showed dysregulation of genes belonging to ectodermal (neuronal lineage) and mesodermal (cardiac, skeletal and hematopeotic) lineages in majority. Based on these observations, we performed a qRT-PCR analysis for the relevant markers of the aforementioned lineages in the EBs at different time points and found that Mrhl knockdown EBs expressed ectodermal, mesodermal and neuroectodermal markers (Fgf5, T and Nestin respectively) at higher levels in the earlier days. However, in the later days of differentiation i.e. at days 6 and 8, there was a rapid decrease in the expression of these markers in knockdown EBs in comparison to control EBs (**Fig. 5.2 b-d**). These observations reveal aberrance in the specification of early germ layers during differentiation of mESCs that are depleted of Mrhl with premature specification of ectoderm, mesoderm and neuroectoderm layers during the initial stages and enhanced proliferation in the later stages of differentiation. The latter phenomenon of an increased

proliferation rate might be a cause of premature differentiation and formation of the ectodermal, mesodermal and neuroectodermal progenitors which then undergo rapid cell divisions leading to an apparent increase in the size of the knockdown EBs whereas in the control EBs, there is a more balanced and timely specification of these lineages along with a controlled rate of cell division. These observations are however preliminary and further experiments to confirm them are necessary to understand in detail the effects of knockdown of Mrhl in mESCs in the context of general differentiation and specific differentiation cues such as RA mediated neuronal differentiation. The effect on the formation of endodermal lineage needs to be addressed too. In purview of *in vivo* development of embryos, Mrhl was found to be up regulated during the stages of organogenesis in brain (ectoderm), heart, kidney (mesoderm), liver and lung (endoderm) (**section 3.3.1, Fig. 3.4**). A probable hypothesis in this regard would be that Mrhl is required during the early gastrulation stages of embryonic development [since EBs at days 2 and 4 recapitulate such events (**section 3.3.2**)] to ensure a balanced specification of the three lineages and thereafter during organogenesis probably in the formation and/or maintenance of tissue-specific progenitors.

5.3.3 Mrhl knockout mice are viable and fertile: In order to further understand the role of Mrhl in cellular differentiation and development, we generated knockout mice for Mrhl in C57BL/6 background. The full-length 2.4 kb exon of Mrhl was targeted with a vector containing Neomycin expression cassette for positive selection and DTA expression cassette to rule out random recombinants (**Fig. 5.3 a**). Five of the ES cell clones were then confirmed for recombination by Southern blotting (**Fig. 5.3 b**)**.** A semi-quantitative PCR was further performed on these recombinant ES cell clones to confirm that the expression of the parent gene *Phkb* is not perturbed (**Fig. 5.3 c**). Clone 4G3 was selected for transplantation into the foster mother and generation of the knockout mice. The knockout mice were bred and genotyped for at least three generations (**Fig. 5.3 d**). It was observed that the mice are viable, fertile and produce normal litter size. A qRT-PCR analysis for the expression of Mrhl and Phkb in P27 testes confirmed the status of Mrhl knockout across wild-type, heterozygous and homozygous mice with no change in Phkb expression (**Fig. 5.3 e**). In lieu of lncRNAs and their knockouts as discussed earlier, phenotypes observed *in vitro* might not be recapitulated *in vivo* due to compensatory mechanisms. However, histological examinations of various tissues, especially the brain and testes remain to be evaluated to understand the role of Mrhl in lineage specification and cell fate determination in tissues during development.

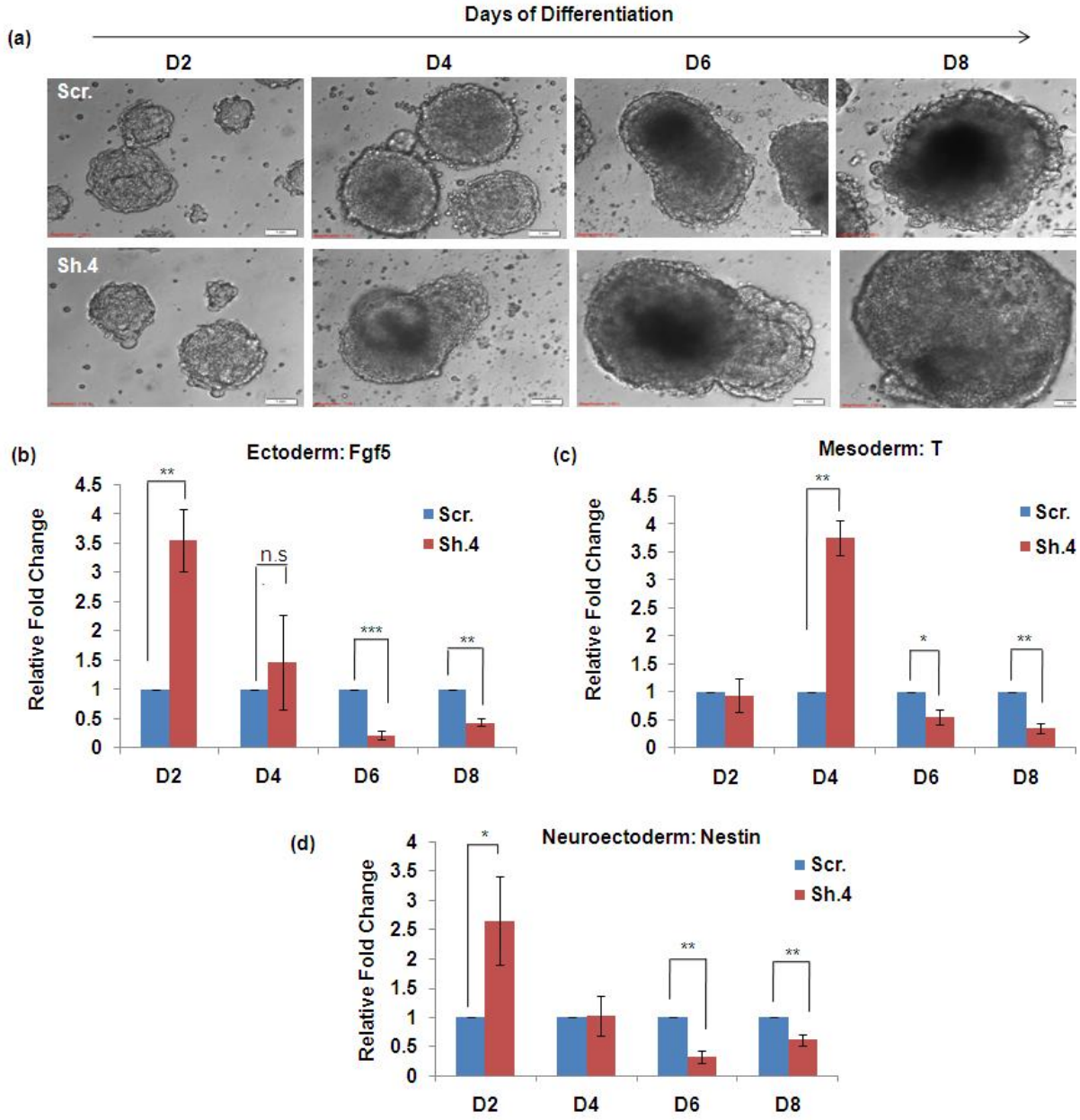


Fig. 5.2 *Aberrance in proliferation and differentiation of EBs derived from Mrhl sh.4 cells. (a) Brightfield microscopy indicated larger size for Mrhl sh.4 EBs as compared to scrambled control. (b-d) qRT-PCR analysis of ectodermal, mesodermal and neuroectodermal markers showed altered differentiation properties for Mrhl knockdown cells.*

*Error bars indicate standard deviation from one representative experiment in biological triplicates. *p<0.05, **p<0.01, ***p<0.001, student's t-test. Scale bar = 1mm.*

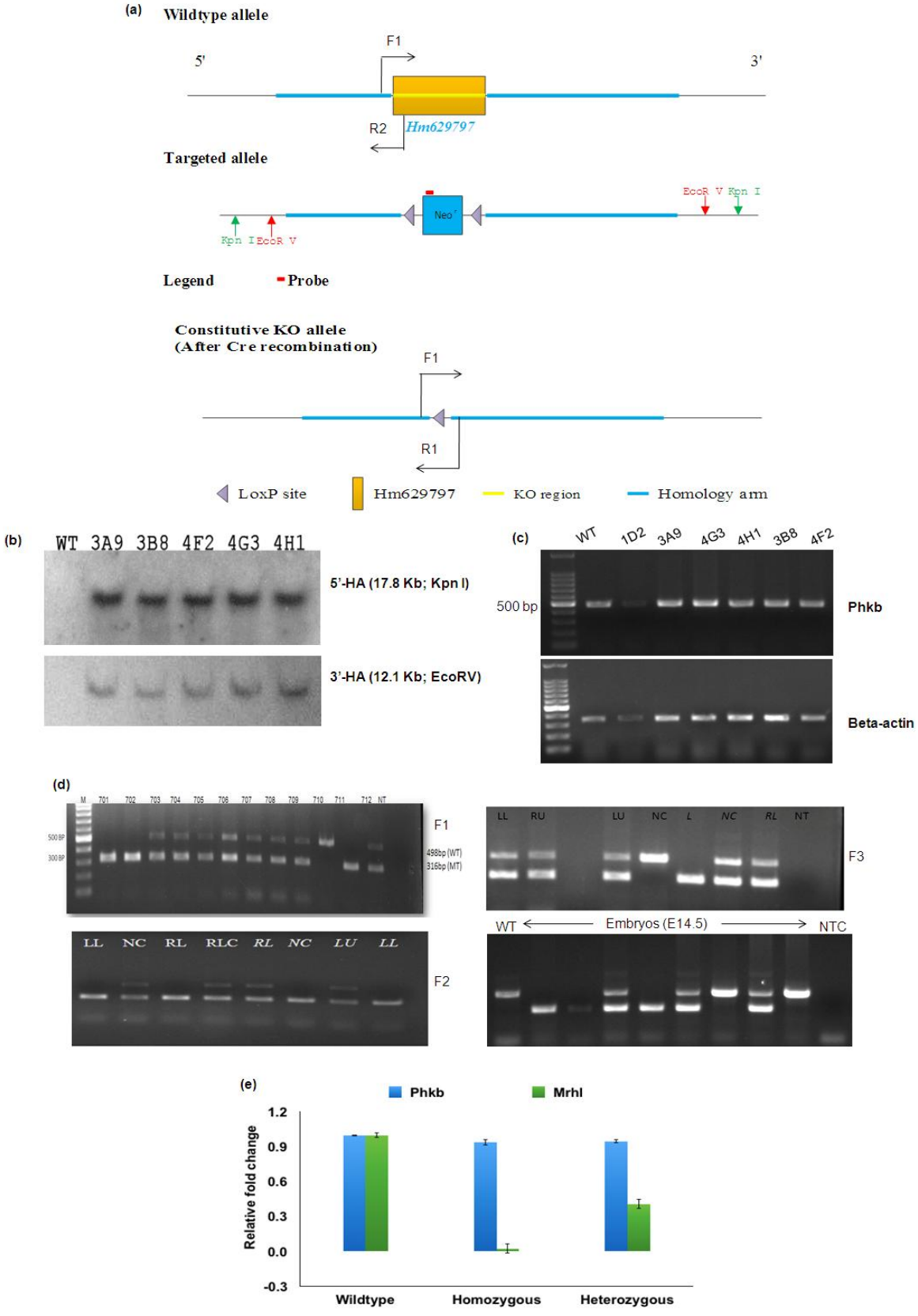


Fig. 5.3 *Knockout strategy and preliminary characterization of Mrhl knockout mice. (a) Knockout strategy for Mrhl exon deletion. (b) Confirmation of targeted ES clones by*

Southern blotting. (c) PCR confirmation of Phkb expression status in targeted ES clones. Clone 1D2 was not confirmed by Southern but had been confirmed for recombination by PCR. (d) Genotyping for generations F1 to F3 and E14.5 embryos. (e) Validation of expression levels of Phkb and Mrhl across wild-type, heterozygous and homozygous mice.

WT: wild-type; MT: mutant/knockout; NTC: no template control

5.4 Summary: In this study, we have aimed at understanding the functional relevance of Mrhl in mESCs and neuronal lineage development through knockdown/knockout approaches. We have generated a stable knockdown line for Mrhl in mESCs and knockout mouse line. The stable knockdown cells showed no changes in the status of pluripotency. However, in the context of providing a differentiation cue, they showed abnormalities in the specification of various lineages. The knockout mice have been characterized and further studies need to be performed.

The major findings and implications of the current study are as follows:

- Stable knockdown cells for Mrhl in mESCs showed an average down regulation of~50%. qRT-PCR and WB analysis for pluripotency markers and alkaline phosphatase assay did not show any significant changes in the pluripotent status of the cells. IF studies revealed no observable changes in cell adhesion properties for the knockdown cells as compared to scrambled control cells. These observations are in agreement with our RNA-Seq analysis upon transient knockdown of Mrhl wherein pluripotency related genes were not perturbed.
- Upon providing a general differentiation cue to the knockdown cells through the formation of EBs, it was observed that there is premature specification of ectodermal, mesodermal and neuroectodermal lineages during the early stages of differentiation and a lower expression level of the corresponding markers at later stages. At the later stages, there was also observed an increased size of the Mrhl sh. EBs as compared to the scrambled control. In perspective of our studies on Mrhl in mESCs during differentiation, in neuronal and male germ cell progenitors and its expression in embryonic organs at various stages, we hypothesize that Mrhl might have a role in a balanced specification of the lineages during early stages of differentiation whereas during the later stages it might be required further for the proper differentiation programs to be activated leading upto cell fate specification in tissues.

- Mrhl knockout mice are viable, fertile and produce normal sized litters. 2.4 kb deletion of the Mrhl exon did not lead to a change in the expression level of its parent gene Phkb.

5.5 Future Perspectives:

- The role of Mrhl in neuronal differentiation needs to be addressed by subjecting the Mrhl sh.4 stable knockdown cells to RA mediated differentiation to understand whether Mrhl is involved in the formation and/or maintenance of such progenitors.
- Histological examination of various tissues in the knockout mice model of Mrhl, especially, in the brain and the testes have to be performed to decipher the necessity of Mrhl in cell fate specification along with differentiation and development programs.

Chapter 6

Discussion and Summary

The completion of the Human Genome Project followed by cataloguing and annotation of function elements in the mammalian genome through ENCODE (in humans) [12] and FANTOM (in mouse) [11] revolutionized the concept of pervasive transcription [13]and functional relevance of non-coding RNAs in mammalian physiology. The advent of high-throughput transcriptome and interactome studies coupled with biochemical approaches established a class of these non coding RNAs i.e. the long non-coding RNAs as essential components of fine-tuning molecular and signaling pathways in vertebrate and invertebrate biological phenomena from development to disease. In this context, our laboratory has extensively studied lncRNA Mrhl or meiotic recombination hotspot locus RNA, discovered in the process of characterizing the recombination locus in chromosome 8 in mouse [134]. The 2.4 kb lncRNA was found to be polyadenylated, nuclear-localized, exhibit tissue-specific expression and possess an independent transcription unit although it is located within the 15th intron of a parent housekeeping gene *Phkb*. Subsequent studies have revealed it to be involved in negatively regulating the WNT signaling pathway in association with a protein partner p68 in spermatogonial progenitor cells [64]. Interestingly, Mrhl itself was found to be negatively regulated by the WNT pathway at the epigenetic level, through the recruitment of the co-repressor CTBP1 at its promoter and establishment of repressive histone modifications during meiotic commitment of these progenitors [138]. Furthermore, Mrhl was found to regulate an important transcription factor SOX8, involved in spermatogenesis, through association at the chromatin level [137]. These studies delineate a fantastic negative feedback mechanism operating within Mrhl and a developmentally important context-dependant pathway such as the WNT pathway in spermatogonial progenitors along with emphasizing the regulation of lineage-specific transcription factors such as SOX8 in the context of an essential phenomenon like male germ cell meiotic commitment.

A predominant characteristic of lncRNAs is the high tissue-specific and often temporal expression as well as context dependant regulatory mechanisms and interaction partners in diverse physiological and biological phenomena. LncRNA Tuna has been shown to participate in the regulation of core pluripotency factors and maintenance of pluripotency of mouse ES cells [201]. It was also demonstrated to be necessary for their neural differentiation. It was found to form a complex with three RNA binding proteins, PTBP1, hn-RNPK and NCL and interact with the chromatin of target genes such as *Nanog*, *Sox2* and *Fgf4* in association with hn-RNPK. SOX2 has been implicated as an important TF in ES cells and neural progenitors and Tuna and SOX2 were found to co-regulate a common set of genes

involved in development, differentiation, proliferation and neurogenesis. LncRNA Tsx has also been shown to be involved in the maintenance of mouse ES cells, pachytene spermatocytes in testes and cognition and behaviour in mice [214], implying the wide functional implications of lncRNAs in mammalian physiology. Similarly, lncRNA Gomafu has been implicated in pluripotency maintenance in mouse ES cells [199] along with playing roles in the specification of osteogenic [111] and oligodendrocyte lineages [200]. Our studies were hence directed towards deciphering lncRNA Mrhl and its role in cell fate determination and lineage specification in other tissues, given its already established role in regulating the WNT pathway and SOX8 during male germ cell meiotic commitment. Towards this end, we chose mouse ES cells (mESCs) as our first model system of study leading upto exploring its role in neuronal differentiation and obtaining an understanding its relevance in the cellular/biological contexts.

Transcriptome studies in mESCs revealed roles for Mrhl in regulating developmental and differentiation related programs

mESCs derived from the inner cell mass of blastocysts stage embryos have served as a model system for addressing development related queries. The nuclear localization and moderate expression stability of Mrhl along with a lack of regulation of Mrhl by the WNT signaling/p68 cascade in mESCs prompted us to delineate its role in this cellular context (**Figs. 3.5 a, e and 3.6 b**). Transient knockdown followed by detailed systems analyses of the transcriptome sequencing data revealed interesting aspects about gene regulation in mESCs by Mrhl. 1143 genes were found to be dysregulated upon Mrhl knockdown cells with 729 being down regulated, 414 being up regulated and with the differentially expressed genes (DEG) belonging to diverse biological processes and molecular functions (**Fig. 3.7 d, e**) Statistical enrichment analysis showed developmental processes to be one amongst the over represented functional processes with the majority of the perturbed genes belonging to this functional class (**Table 1**). A curation of these genes led to the observation that ~20% of them belonged to the neuronal lineage (**Tables 2a, 2b**). In fact, ectoderm development was also found to be a significantly over represented functional process in the above analysis, the ectoderm being the origin for all nervous system related cell types. Gene clustering and co-expression analysis to understand the statistically correlated genes regulating specific functions in mESCs in Mrhl depleted conditions led to the identification of nervous system processes as one of the functionally enriched gene clusters (**Fig. 3.8 a**). ATOH1 and POU3F2 were found to be two of the most important neuronal related transcription factors with the

highest perturbation in expression along with TBR1, DLX3, LHX1 and VSX2, all of which have established roles in neurogenesis, neuronal progenitors and brain development (**Appendix tables 2a and 2b**). ATOH1 and POU3F2 have extensive protein-protein interaction networks with other transcription factors involved in neuronal related pathways as was observed from a STRING interrogation. These observations suggested that Mrhl might have a role to play in the development of the neuronal lineage. Other perturbed genes such as *Nav3*, *Unc5d*, *Rarb*, *Epha3*, *Dlg2*, *Rb1* and *Fat2* have also been established to function in the regulation of neuronal migration, synapse formation and signaling (**Appendix table 2a**).

The other over represented processes in the statistical enrichment analysis pertained to cell-cell signaling, ion transport, synaptic transmission, response to endogenous stimuli, anion transport, and neuromuscular synaptic transmission (**Table 1**). Gene clustering also revealed metabolic processes, ion channel and transporter activity, immune system processes and xenobiotic processes as the functionally enriched clusters (**Fig. 3.8 a**). Whilst we have focussed only on developmental processes for our study herewith, it would be intriguing to understand the role of Mrhl in regulation of other processes that recur in the systems analyses such as ion transport. Further, analysis of the developmental processes showed important lineage determining genes belonging to cardiac, hematopoietic, skeletal and pancreatic lineages to be perturbed in expression along with epidermal and axis specification genes (**Tables 2a and 2b**). We addressed the role of Mrhl in neuronal lineage development since a significant proportion of the perturbed genes belonged to this process. However, the possibility of Mrhl regulating multiple lineage specification programs in response to the appropriate differentiation cues looks promising. In fact, its roles in spermatogonial progenitor cells have been well established. In this context, it would be interesting to explore Mrhl as a molecular player in cellular differentiation, fate specification and development.

A further interesting inference from this study was the context dependant and versatile role of Mrhl in regulating cellular processes. A comparison of the DEG from mESCs and GC1-Spg spermatognial progenitor model systems revealed a widely different perturbed expression dataset with only 25 genes in common (**Fig. 3.7 g**). A gene ontology of these common genes showed important processes such as signal transduction, transmembrane transport, regulation of ion transport and multicellular organismal processes to be represented. This implied that during cellular differentiation and development, in addition to cell or tissue specific targets, Mrhl might be involved in regulating some key common pathways or genes across cell types.

Since transcription factor networks have been established to determine cellular and disease states, a construction of a TF matrix from the DEG in Mrhl knockdown mESCs showed that dysregulated TFs majorly categorized into developmental processes in gene ontology analysis (**Fig. 3.8 d**). RUNX2 and ERG known for their roles in the skeletal and vascular lineages respectively were found to have the maximum number of binding sites for the other TFs (**Fig. 3.8 e, f**), posing them as probable master TFs in the matrix. Furthermore, a novel TF hierarchy was generated based on the predicted matrix and known interactions from the STRING database that might be perturbed in mESCs under conditions of Mrhl depletion (**Fig. 3.9**). RUNX2 was at the top of this hierarchy with HEYL and HOXB7 being the common TFs between the predicted matrix and the hierarchy. Interestingly, all of these four TFs are not only involved in determining specific lineages or patterning [247, 248, 250, 251], they were also seen to be down regulated in expression upon knockdown of Mrhl in mESCs. This analysis indicates the existence of a regulatory TF network probably governed by Mrhl that might function to impart specific cellular identities to mESCs in the presence of various environmental cues or signals.

Our transcriptome studies and systems analyses in mESCs upon lncRNA Mrhl knockdown thus suggests that when Mrhl is depleted from mESCs, specific lineage determining genes and development related processes are majorly affected along with ion transport, transmembrane transport and metabolic processes with no perturbation of pluripotency related genes or pathways. Amongst all the development related processes, neuronal lineage development and nervous system processes appear to be over represented. Mrhl in mESCs probably acts to define a cellular state which is primed for responding to the appropriate differentiation cues and development into one or the other specific lineage.

Studies on Mrhl in neuronal lineage development revealed a probable role in neuronal progenitors and PAX6 mediated regulation of Mrhl

Based on our studies in mESCs and Mrhl, we probed into the functional relevance of this lncRNA in the context of neuronal lineage development. We performed both *in vivo* studies in mouse embryonic and postnatal brains as well as *in vitro* studies in radial glia like neuronal progenitors derived from RA treatment of embryoid bodies and observed a predominance of expression of Mrhl in the progenitor populations of embryonic brains and *in vitro* culture (**Figs. 4.2 a and 4.3 a, c**). ChIP studies confirmed its transcriptionally active state in the neuronal progenitors (**Figs. 4.2 c and 4.3 d**) with FISH experiments demonstrating its nuclear

localization in this population of cells. An interesting observation herewith was that Mrhl is expressed ubiquitously across the fore-, mid- and hindbrain regions at E14.5 stage (**Fig. 4.2 b**) suggesting that it might have roles to play in cell fate specification and/or differentiation related processes in the embryonic brain aside from the progenitor populations of the VZ/SVZ of the forebrain which however, is beyond the scope of this study since characteristics of progenitor cells in other regions of the brain still remain to be defined.

In order to understand the probable mechanisms behind the regulation of Mrhl in neuronal progenitors, we performed a promoter TF binding site prediction analysis which revealed a key observation that the promoter of Mrhl is replete with binding sites for a host of lineage-specific and development related TFs (**Fig. 4.4**). This is relevant to our transcriptome studies in mESCs since developmental processes were the most perturbed gene set and it suggests a mechanism wherein Mrhl not only acts to regulate but in turn is regulated by TFs that specify cell states and commitment programs. With respect to the neuronal lineage development, PAX6 [281-288], TCF4 (WNT pathway [309-313]) and RBPJ-κ (NOTCH pathway [299-303]) were the interesting candidates for regulating Mrhl in neuronal progenitors. In this regard, we demonstrated that WNT signaling pathway, for which TCF4 is the transcriptional effector, is not activated in RA derived neuronal progenitors (**Fig. 4.5 a**) although earlier reports have shown WNT to be involved in neural precursors and/or their differentiation into neurons. Furthermore, down regulation of Mrhl in neurons did not lead to an activation of the WNT pathway (**Fig. 4.5 b**), unlike in spermatogonial progenitors, suggesting context-dependant regulation of Mrhl in cellular differentiation. NOTCH pathway, with RBPJ-κ as the effector of target gene transcription was observed to be activated in the progenitors but not involved in the regulation of Mrhl (**Fig. 4.5 c, d**). Finally, through a series of bioinformatics, co-expression, ChIP-qPCR, EMSA and luciferase reporter studies we demonstrated that PAX6 might be involved in regulating Mrhl in neuronal progenitors through multiple sites of action (**Figs. 4.6, 4.7, 4.8**).

The upstream most site (Site 1, -2905 bp) appeared to be occupied both *in vivo* and *in vitro*, bound directly by PAX6 and was also observed to be functionally relevant as proved by reporter assays. The second site (Site 2, -1277 bp) was occupied only *in vivo* but not *in vitro*, weakly bound by PAX6 and was seen to be functionally important as well. The sites most proximal to the TSS (Sites 3 and 4, -673 and -623 bp) might be more redundant because it was found to be occupied both *in vivo* and *in vitro* but EMSA and reporter assays did not reveal any one to one binding or luciferase signal. An important conclusion from this study is

that a master transcription factor, conserved across vertebrates and invertebrates and involved in controlling and mediating neuronal progenitor and differentiation programs also regulates a lncRNA. Furthermore it is interesting to note the highly context-dependent regulation of lncRNA Mrhl by PAX6 during neuronal commitment (**Fig. 4.9**). Whilst Site 1 might be the major site of action of PAX6 on Mrhl promoter, Site 2 might be slightly ambiguous due to differences in occupancy in the *in vivo* and *in vitro* scenarios. This discrepancy can however be explained by the difference in the pool of TFs or mediators that associate with PAX6 in the two conditions. Also, given the position of Site 1 and the enrichment of H3K4me3 and H3K4me1 modifications, it would be interesting to study if PAX6 regulates this site on Mrhl as a distal promoter or a proximal enhancer element. PAX6 has in fact, been shown to be associated with various H3K4-specific DNA methyltransferases [343] to mediate gene regulation at the enhancer and promoter elements of target loci. Site 3 can again be regulated in a context-dependant manner but seems more likely to be redundant because of a lack of luciferase activity. Another interesting perspective is the probable regulation of Mrhl mediated by PAX6 and its isoform PAX6(5A) in conjunction or independent of each other. Reports have demonstrated that the two isoforms are not only expressed but also act in synergism in different ratios to regulate their target genes [341, 357]. This can indeed explain the luciferase activity in our reporter assays despite lack of one to one binding of PAX6(5A) on any of the sites on Mrhl promoter.

Our studies on Mrhl in the development of the neuronal lineage showed that Mrhl expression is temporally regulated during brain development in mouse and it exhibits predominant expression in neuronal progenitors. With regards to the regulators of Mrhl in neuronal progenitors, we observed that WNT and NOTCH pathways are not involved in regulating Mrhl although the WNT pathway has been shown to function in a negative feedback loop with Mrhl in spermatogonial progenitors and Mrhl promoter harbours the binding site for both TCF4 and RBPJ-κ. A major neuronal progenitor master transcription factor PAX6 was demonstrated to probably regulate Mrhl through multiple sites of action, highly dependent on the associative pool of factors and the cellular contexts.

Our studies on Mrhl in mouse embryonic stem cells and neuronal lineage development have been summarized in **Fig. 6.1**.

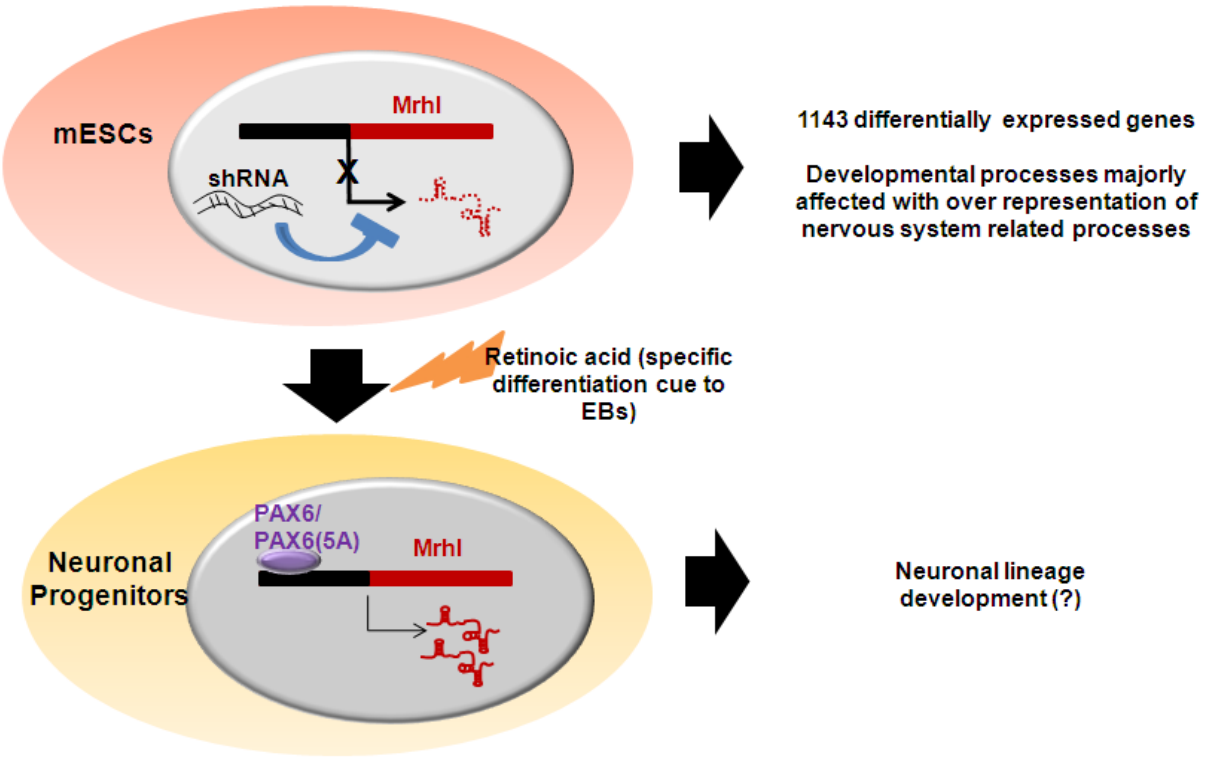


Fig. 6.1 *Summary of our findings and studies on the role of lncRNA Mrhl in mESCs and neuronal lineage development. Mrhl depletion in mESCs affected expression of important lineage specific and development related genes and processes. An over represented class of genes was found to belong to nervous system processes and neuronal lineage with functions in neuronal cell survival, migration, differentiation, synapse and brain development. Mrhl was found to be predominantly expressed in neuronal progenitor populations, its transcriptional active state being probably regulated by a master transcription factor PAX6.*

Knockdown and knockout studies identify potential roles for Mrhl in lineage specification in mESCs and development

Our studies on lncRNA Mrhl in mESCs and neuronal lineage development provided us with key insights into its role in regulating differentiation and development related genes, with a probable role in neuronal progenitor cell populations. To further understand its functional relevance, we aimed at performing phenotypic studies. The characterization of a stable knockdown mESC line for Mrhl revealed no changes in the expression of pluripotency associated genes or cell adhesion properties (**Fig. 5.1**), demonstrating correlation with our transcriptome and systems studies. Furthermore, in accordance with our hypothesis that Mrhl primes the cells to be responsive to appropriate differentiation signals, EB differentiation of

knockdown cells showed aberrance in the formation of ectoderm, mesoderm and early neuroectoderm lineages (**Fig. 5.2**). Preliminary analysis revealed premature specification of these lineages in the early stages of differentiation accompanied by an indicative increase in the proliferation rate of the EBs at later stages.

To extrapolate the phenotypic changes at the organismal level, we generated a knockout mice model targeting a deletion of the full-length 2.4 kb single exon of Mrhl (**Fig. 5.3**). The mice were observed to be viable with normal litter size and no change in expression levels of Phkb.

Our preliminary loss-of-function studies coupled with our earlier studies for Mrhl suggest that it might be important for regulating the balanced specification of lineages during early differentiation corresponding to the gastrulation stages *in vivo* whereas in the later stages, it might be further required for the formation and/or maintenance of progenitor populations corresponding to organogenesis *in vivo* (**Fig. 6.2**).

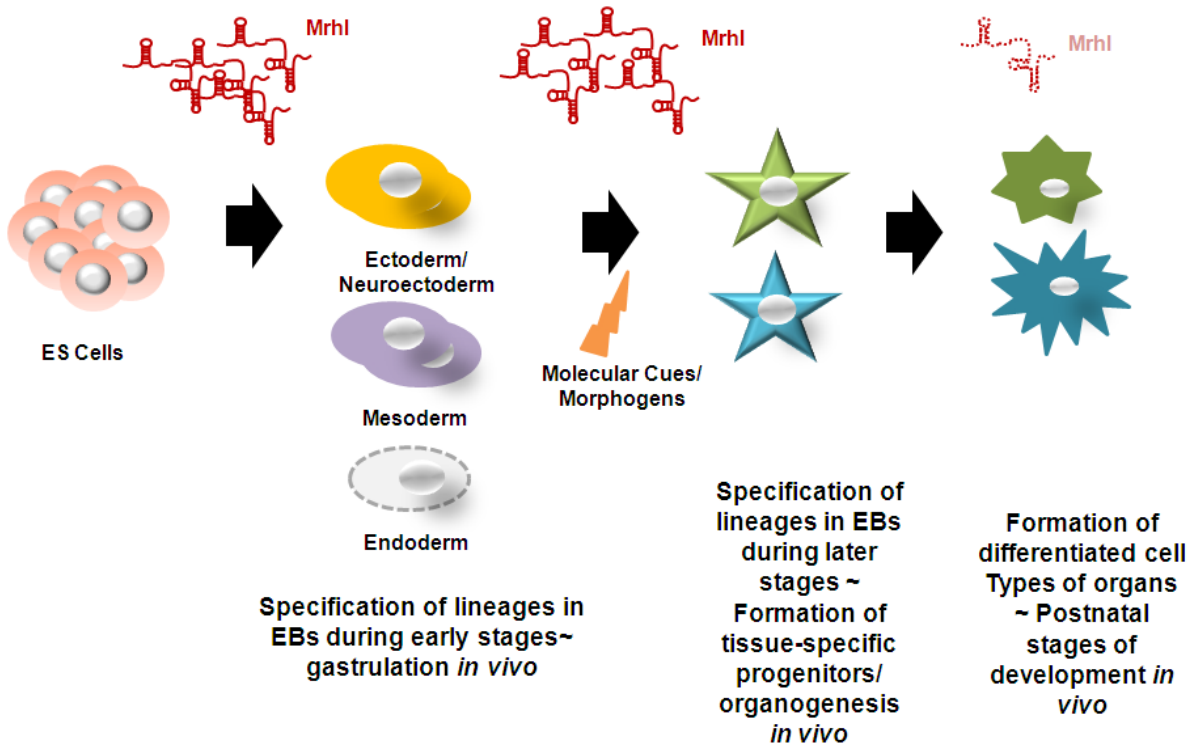


Fig. 6.2 *A suggestive model for the role of lncRNA Mrhl in development. Our studies have revealed that Mrhl is expressed in mESCs and up regulated during general EB differentiation and RA mediated neuronal differentiation. Depletion of mrhl led to dysregulation of lineage-specific genes and aberrance in EB differentiation. A hypothesis in this regard is that Mrhl is required for the balanced specification of lineages during early stages of EB differentiation*

corresponding to gastrulation in vivo. During later stages of EB differentiation corresponding to the formation of tissue-specific progenitors and organogenesis in vivo, Mrhl might be functionally required as well. This has been indicative from our studies in the context of male germ cell and neuronal commitment. In differentiated mature cell types corresponding to postnatal stages of organ development, Mrhl expression is drastically reduced, emphasizing its role in early mouse embryonic development.

Future Perspectives

In purview of our studies on lncRNA Mrhl in mouse embryonic stem cells, we showed that genes belonging to varied developmental aspects such as lineage specification, patterning, signaling, cell adhesion and receptor activity are majorly dysregulated. In this respect, the mechanism of target gene regulation by Mrhl in mESCs needs to be addressed. LncRNAs have been shown to regulate their targets through diverse mechanisms, a predominant one for nuclear lncRNAs being chromatin-mediated regulation. In the context of male germ cell meiotic commitment, Mrhl has already been shown to associate and directly regulate key genes such as *Sox8* through association with the chromatin [117]. ChOP-Seq studies for Mrhl in mESCs are being performed in this respect, owing to its nuclear localization and chromatin association as revealed through our initial biochemical studies, to understand the mechanism of target gene regulation. Furthermore, our studies on the differentiation and proliferation status of Mrhl knockdown embryoid bodies have to be further corroborated with additional studies to fully understand the role of Mrhl in regulating aspects of early lineage specification events. In the context of neuronal lineage development, it would be crucial to address the role of Mrhl in neuronal progenitors per se through loss-of-function approaches coupled with transcriptome studies. Furthermore, the PAX6 mediated regulation of Mrhl in these progenitors followed by their concomitant down regulation of expression in neurons gives an indication of an autofeedback loop that might be operating between Mrhl and PAX6 in this context. Additionally, our studies on the status of transcription activation or enhancer histone modifications of Mrhl in neuronal progenitors and in cultured neurons indicate an epigenetic regulation and it would be interesting to explore the mechanisms behind the same. Further studies on Mrhl knockout mice can be directed towards understanding its role in cell fate specification, lineage development, organogenesis and development.

Stem cells and their applications in regenerative medicine have been under exploration since their discovery and have currently been established to be useful for transplantations, tissue

engineering and drug studies. Stem cells can be classified into ESCs (embryonic stem cells), TSPSCs (tissue-specific progenitor stem cells), mesenchymal stem cells (MSCs), umbilical cord stem cells (UCSCs), bone marrow stem cells (BMSCs) and induced pluripotent stem cells (iPSCs) based on their regenerative medicine applications [358]. ESCs can be used for the treatment of spinal cord injury patients [359], differentiated into cone cells for treatment of age-related macular degeneration [360], differentiated into cardiomyocytes for cardiac tissue injury repair [361], programmed to generate insulin secreting β-cells for type II diabetes [362, 363] or be used to derive hepatic progenitors for drug screening studies (reviewed in [358]) and chondrocytes for the treatment of cartilage defects [364]. Although several considerations remain to be addressed in terms of host versus graft rejection and obtaining pure populations of functional progenitors, stem cell based therapy is one of the most promising fronts in regenerative medicine. In the most recent development, human parthenogenetic neural stem cells have been successfully transplanted into the striatum and substantia niagra of the brain of Parkinson's disease patients and they have shown improved cognitive abilities, decrease in impulsive and compulsive behaviours with no adverse effects such as the formation of tumors or cysts or inflammation within six months of receiving their first dose of stem cells. With these perspectives in mind, it becomes imperative to address the molecular players and pathways that operate for the optimal functioning of stem cells and their derived progenitors. It has been only three decades since lncRNAs have been discovered as crucial in regulating, modulating and mediating major pathways and processes and they hold potential to be used as targets for therapeutic purposes in the field of regenerative medicine. The plethora of lncRNAs that are involved in maintaining ESCs or iPSCs or that are involved in the generation, proliferation and maintenance of TSPSCs have been discussed herewith widely. Our studies on lncRNA Mrhl in regulating differentiation and development programs in mESCs or in playing potential roles in neural progenitor specification (or spermatogonial progenitors) is a small step towards characterizing lncRNAs as essential components of the circuits of stem cells such as ESCs and TSPSCs with a wider aim of developing stem cell based therapies.

Annexure 1 *List of DEG*

Down Regulated:

Gene name	Gene type	Gene description	Fold_change
D130043K22Rik	protein_coding	RIKEN cDNA D130043K22 gene [Source:MGI Symbol;Acc:MGI:3036268]	-4.80699
Snord17	snoRNA	small nucleolar RNA, C/D box 17 [Source:MGI Symbol;Acc:MGI:3819523]	-4.78565
4932431P20Rik	transcribed_unitary_pseudogene	RIKEN cDNA 4932431P20 gene [Source:MGI Symbol;Acc:MGI:2149781]	-4.31578
A430010J10Rik	processed_transcript	RIKEN cDNA A430010J10 gene [Source:MGI Symbol;Acc:MGI:2442501]	-4.13893
Gm24407	snRNA	predicted gene, 24407 [Source:MGI Symbol;Acc:MGI:5454184]	-4.11711
Cfap61	protein_coding	cilia and flagella associated protein 61 [Source:MGI Symbol;Acc:MGI:1926024]	-4.00258
Gm16425	processed_pseudogene	predicted gene 16425 [Source:MGI Symbol;Acc:MGI:3643338]	-3.97683
Gm43913	lincRNA	predicted gene, 43913 [Source:MGI Symbol;Acc:MGI:5690305]	-3.94961
1810062O18Rik	antisense	RIKEN cDNA 1810062O18 gene [Source:MGI Symbol;Acc:MGI:1922852]	-3.91181
Gm5451	processed_pseudogene	predicted gene 5451 [Source:MGI Symbol;Acc:MGI:3643790]	-3.7904
Abcb9	protein_coding	ATP-binding cassette, sub-family B (MDR/TAP), member 9 [Source:MGI Symbol;Acc:MGI:1861729]	-3.66662
Rbms3	protein_coding	RNA binding motif, single stranded interacting protein [Source:MGI Symbol;Acc:MGI:2444477]	-3.66038
Pink1	protein_coding	PTEN induced putative kinase 1 [Source:MGI Symbol;Acc:MGI:1916193]	-3.6133
Slc17a6	protein_coding	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 6 [Source:MGI Symbol;Acc:MGI:2156052]	-3.53357
Scn4b	protein_coding	sodium channel, type IV, beta [Source:MGI Symbol;Acc:MGI:2687406]	-3.51964
Sis	protein_coding	sucrase isomaltase (alpha-glucosidase) [Source:MGI Symbol;Acc:MGI:1917233]	-3.50696
Grm8	protein_coding	glutamate receptor, metabotropic 8 [Source:MGI Symbol;Acc:MGI:1351345]	-3.49297
Zscan4f	protein_coding	zinc finger and SCAN domain containing 4F [Source:MGI Symbol;Acc:MGI:3708485]	-3.47278
Pde6g	protein_coding	phosphodiesterase 6G, cGMP-specific, rod, gamma [Source:MGI Symbol;Acc:MGI:97526]	-3.46445
Gm37943	TEC	predicted gene, 37943 [Source:MGI Symbol;Acc:MGI:5611171]	-3.4498
Gm24265	snRNA	predicted gene, 24265 [Source:MGI Symbol;Acc:MGI:5454042]	-3.43904
Gm8093	transcribed_processed_pseudogene	predicted gene 8093 [Source:MGI Symbol;Acc:MGI:3642975]	-3.39953
Stum	protein_coding	mechanosensory transduction mediator [Source:MGI Symbol;Acc:MGI:2138735]	-3.39884
Mcoln2	protein_coding	mucolipin 2 [Source:MGI Symbol;Acc:MGI:1915529]	-3.3944
Sp9	protein_coding	trans-acting transcription factor 9 [Source:MGI Symbol;Acc:MGI:3574660]	-3.32375
Hoxb7	protein_coding	homeobox B7 [Source:MGI Symbol;Acc:MGI:96188]	-3.31541
Abcg4	protein_coding	ATP binding cassette subfamily G member 4 [Source:MGI Symbol;Acc:MGI:1890594]	-3.29901
Bean1	protein_coding	brain expressed, associated with Nedd4, 1 [Source:MGI Symbol;Acc:MGI:1929597]	-3.28908
Hepacam2	protein_coding	HEPACAM family member 2 [Source:MGI Symbol;Acc:MGI:2141520]	-3.2777
Lrfn5	protein_coding	leucine rich repeat and fibronectin type III domain containing 5 [Source:MGI Symbol;Acc:MGI:2144814]	-3.25557
Gm28536	antisense	predicted gene 28536 [Source:MGI Symbol;Acc:MGI:5579242]	-3.24408
Erg	protein_coding	ETS transcription factor [Source:MGI Symbol;Acc:MGI:95415]	-3.22962
Ryr3	protein_coding	ryanodine receptor 3 [Source:MGI Symbol;Acc:MGI:99684]	-3.20811
Gm11870	processed_pseudogene	predicted gene 11870 [Source:MGI Symbol;Acc:MGI:3650285]	-3.20331
Abca14	protein_coding	ATP-binding cassette, sub-family A (ABC1), member 14 [Source:MGI Symbol;Acc:MGI:2388708]	-3.17359
Gm16046	lincRNA	predicted gene 16046 [Source:MGI Symbol;Acc:MGI:3802093]	-3.15637
Actl10	protein_coding	actin-like 10 [Source:MGI Symbol;Acc:MGI:1917612]	-3.13901
Cyp2c40	protein_coding	cytochrome P450, family 2, subfamily c, polypeptide 40 [Source:MGI Symbol;Acc:MGI:1306815]	-3.08613
Gm15873	antisense	predicted gene 15873 [Source:MGI Symbol;Acc:MGI:3802055]	-3.06625
C6	protein_coding	complement component 6 [Source:MGI Symbol;Acc:MGI:88233]	-3.06253
Pgm5	protein_coding	phosphoglucomutase 5 [Source:MGI Symbol;Acc:MGI:1925668]	-3.03277
Fyb2	protein_coding	FYN binding protein 2 [Source:MGI Symbol;Acc:MGI:2685466]	-3.01786
Gm42729	TEC	predicted gene 42729 [Source:MGI Symbol;Acc:MGI:5662866]	-3.01754
Tspan32os	antisense	tetraspanin 32, opposite strand [Source:MGI Symbol;Acc:MGI:3705185]	-3.01706
Foxl2os	processed_transcript	forkhead box L2, opposite strand [Source:MGI Symbol;Acc:MGI:3614944]	-3.01628
1700003G18Rik	processed_transcript	RIKEN cDNA 1700003G18 gene [Source:MGI Symbol;Acc:MGI:1916600]	-3.00463
Ccdc184	protein_coding	coiled-coil domain containing 184 [Source:MGI Symbol;Acc:MGI:2146066]	-2.99971
Tnr	protein_coding	tenascin R [Source:MGI Symbol;Acc:MGI:99516]	-2.96716
Cpne4	protein_coding	copine IV [Source:MGI Symbol;Acc:MGI:1921270]	-2.94687
Epha3	protein_coding	Eph receptor A3 [Source:MGI Symbol;Acc:MGI:99612]	-2.93209
Tlr7	protein_coding	toll-like receptor 7 [Source:MGI Symbol;Acc:MGI:2176882]	-2.91691

Gm25939	snRNA	predicted gene, 25939 [Source:MGI Symbol;Acc:MGI:5455716]	-2.90942
BC051142	protein_coding	cDNA sequence BC051142 [Source:MGI Symbol;Acc:MGI:3039565]	-2.90069
Gm15774	antisense	predicted gene 15774 [Source:MGI Symbol;Acc:MGI:3783215]	-2.90019
1700109K24Rik	antisense	RIKEN cDNA 1700109K24 gene [Source:MGI Symbol;Acc:MGI:1921553]	-2.89963
Slc38a8	protein_coding	solute carrier family 38, member 8 [Source:MGI Symbol;Acc:MGI:2685433]	-2.89925
Gm25777	snoRNA	predicted gene, 25777 [Source:MGI Symbol;Acc:MGI:5455554]	-2.8943
Gm25777	snoRNA	predicted gene, 25777 [Source:MGI Symbol;Acc:MGI:5455554]	-2.8943
Duxbl3	protein_coding	double homeobox B-like 3 [Source:MGI Symbol;Acc:MGI:3710520]	-2.89332
Htra3	protein_coding	HtrA serine peptidase 3 [Source:MGI Symbol;Acc:MGI:1925808]	-2.89302
Rnu3a	snoRNA	U3A small nuclear RNA [Source:MGI Symbol;Acc:MGI:97977]	-2.88681
Rnu3a	snoRNA	U3A small nuclear RNA [Source:MGI Symbol;Acc:MGI:97977]	-2.88681
Insrr	protein_coding	insulin receptor-related receptor [Source:MGI Symbol;Acc:MGI:1346037]	-2.87654
Gm20005	lincRNA	predicted gene, 20005 [Source:MGI Symbol;Acc:MGI:5012190]	-2.87422
Lair1	protein_coding	leukocyte-associated Ig-like receptor 1 [Source:MGI Symbol;Acc:MGI:105492]	-2.86801
Rgs8	protein_coding	regulator of G-protein signaling 8 [Source:MGI Symbol;Acc:MGI:108408]	-2.86483
Gm29543	lincRNA	predicted gene 29543 [Source:MGI Symbol;Acc:MGI:5580249]	-2.85719
Cpz	protein_coding	carboxypeptidase Z [Source:MGI Symbol;Acc:MGI:88487]	-2.85698
Col15a1	protein_coding	collagen, type XV, alpha 1 [Source:MGI Symbol;Acc:MGI:88449]	-2.85378
Bnc1	protein_coding	basonuclin 1 [Source:MGI Symbol;Acc:MGI:1097164]	-2.85009
Gm21182	processed_pseudogene	predicted gene, 21182 [Source:MGI Symbol;Acc:MGI:5434537]	-2.84915
Tdrd6	protein_coding	tudor domain containing 6 [Source:MGI Symbol;Acc:MGI:2679727]	-2.83925
Gm15910	antisense	predicted gene 15910 [Source:MGI Symbol;Acc:MGI:3802158]	-2.83259
Gm3383	protein_coding	predicted gene 3383 [Source:MGI Symbol;Acc:MGI:3781561]	-2.83071
Gm20625	lincRNA	predicted gene 20625 [Source:MGI Symbol;Acc:MGI:5313072]	-2.81822
Gm960	protein_coding	predicted gene 960 [Source:MGI Symbol;Acc:MGI:2685806]	-2.8054
Cpq	protein_coding	carboxypeptidase Q [Source:MGI Symbol;Acc:MGI:1889205]	-2.80107
Ghrl	protein_coding	ghrelin [Source:MGI Symbol;Acc:MGI:1930008]	-2.79729
Nr1h4	protein_coding	nuclear receptor subfamily 1, group H, member 4 [Source:MGI Symbol;Acc:MGI:1352464]	-2.79377
Phf24	protein_coding	PHD finger protein 24 [Source:MGI Symbol;Acc:MGI:2140712]	-2.78547
Krt15	protein_coding	keratin 15 [Source:MGI Symbol;Acc:MGI:96689]	-2.77852
Lamc3	protein_coding	laminin gamma 3 [Source:MGI Symbol;Acc:MGI:1344394]	-2.77708
Bambi-ps1	processed_pseudogene	BMP and activin membrane-bound inhibitor, pseudogene (Xenopus laevis) [Source:MGI Symbol;Acc:MGI:1932402]	-2.77487
Pabpc6	protein_coding	poly(A) binding protein, cytoplasmic 6 [Source:MGI Symbol;Acc:MGI:1914793]	-2.76589
Col24a1	protein_coding	collagen, type XXIV, alpha 1 [Source:MGI Symbol;Acc:MGI:1918605]	-2.7645
B3galt2	protein_coding	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2 [Source:MGI Symbol;Acc:MGI:1349461]	-2.75656
Rab44	protein_coding	RAB44, member RAS oncogene family [Source:MGI Symbol;Acc:MGI:3045302]	-2.75279
Gm7452	transcribed_processed_pseudogene	predicted pseudogene 7452 [Source:MGI Symbol;Acc:MGI:3645151]	-2.74542
Sspo	protein_coding	SCO-spondin [Source:MGI Symbol;Acc:MGI:2674311]	-2.74148
Gm26724	lincRNA	predicted gene, 26724 [Source:MGI Symbol;Acc:MGI:5477218]	-2.73266
4632428C04Rik	lincRNA	RIKEN cDNA 4632428C04 gene [Source:MGI Symbol;Acc:MGI:1921600]	-2.73097
Gm42725	TEC	predicted gene 42725 [Source:MGI Symbol;Acc:MGI:5662862]	-2.72824
Ckm	protein_coding	creatine kinase, muscle [Source:MGI Symbol;Acc:MGI:88413]	-2.72429
Frmpd3	protein_coding	FERM and PDZ domain containing 3 [Source:MGI Symbol;Acc:MGI:3646547]	-2.72259
Gm11991	processed_pseudogene	predicted gene 11991 [Source:MGI Symbol;Acc:MGI:3651121]	-2.7215
Tspo2	protein_coding	translocator protein 2 [Source:MGI Symbol;Acc:MGI:1917276]	-2.71599
Tspo2	protein_coding	translocator protein 2 [Source:MGI Symbol;Acc:MGI:1917276]	-2.71599
Gm16796	antisense	predicted gene, 16796 [Source:MGI Symbol;Acc:MGI:4439720]	-2.71587
Mypopos	antisense	Myb-related transcription factor, partner of profilin, opposite strand [Source:MGI Symbol;Acc:MGI:3826523]	-2.70946
Cdv3-ps	processed_pseudogene	Cdv3 retrotransposed pseudogene [Source:MGI Symbol;Acc:MGI:3646628]	-2.70735
Gm36756	antisense	predicted gene, 36756 [Source:MGI Symbol;Acc:MGI:5595915]	-2.70064
Gm22513	snRNA	predicted gene, 22513 [Source:MGI Symbol;Acc:MGI:5452290]	-2.68756
AI839979	lincRNA	expressed sequence AI839979 [Source:MGI Symbol;Acc:MGI:2140975]	-2.67804
Eif4e1b	protein_coding	eukaryotic translation initiation factor 4E family member 1B [Source:MGI Symbol;Acc:MGI:2685119]	-2.678
Qrich2	protein_coding	glutamine rich 2 [Source:MGI Symbol;Acc:MGI:2684912]	-2.67361
Akap3	protein_coding	A kinase (PRKA) anchor protein 3 [Source:MGI Symbol;Acc:MGI:1341149]	-2.67343
Lrrc74b	protein_coding	leucine rich repeat containing 74B [Source:MGI Symbol;Acc:MGI:1921935]	-2.66628
A930017M01Rik	transcribed_processed_pseudogene	RIKEN cDNA A930017M01 gene [Source:MGI Symbol;Acc:MGI:2685151]	-2.66564
Hoxb3os	antisense	homeobox B3 and homeobox B2, opposite strand [Source:MGI Symbol;Acc:MGI:103211]	-2.66295
Gm6266	processed_pseudogene	predicted gene 6266 [Source:MGI Symbol;Acc:MGI:3646310]	-2.65745
Gm45632	TEC	predicted gene 45632 [Source:MGI Symbol;Acc:MGI:5791468]	-2.65226
Gm30692	lincRNA	predicted gene, 30692 [Source:MGI Symbol;Acc:MGI:5589851]	-2.64584
Gm12798	lincRNA	predicted gene 12798 [Source:MGI Symbol;Acc:MGI:3651483]	-2.63416
Gm11766	lincRNA	predicted gene 11766 [Source:MGI Symbol;Acc:MGI:3650673]	-2.62633
Gm22973	snRNA	predicted gene, 22973 [Source:MGI Symbol;Acc:MGI:5452750]	-2.62535
Hpx	protein_coding	hemopexin [Source:MGI Symbol;Acc:MGI:105112]	-2.61467
Gm9913	protein_coding	predicted gene 9913 [Source:MGI Symbol;Acc:MGI:3642395]	-2.60726
4933427G23Rik	transcribed_processed_pseudogene	RIKEN cDNA 4933427G23 gene [Source:MGI Symbol;Acc:MGI:3036235]	-2.60674
Slc16a7	protein_coding	solute carrier family 16 (monocarboxylic acid transporters), member 7	-2.60115

		[Source:MGI Symbol;Acc:MGI:1330284]	
Angptl7	protein_coding	angiopoietin-like 7 [Source:MGI Symbol;Acc:MGI:3605801]	-2.60103
H3f3aos	antisense	H3 histone, family 3A, opposite strand [Source:MGI Symbol;Acc:MGI:3802006]	-2.59996
Ankk1	protein_coding	ankyrin repeat and kinase domain containing 1 [Source:MGI Symbol;Acc:MGI:3045301]	-2.58214
2210008F06Rik	lincRNA	RIKEN cDNA 2210008F06 gene [Source:MGI Symbol;Acc:MGI:1917307]	-2.58163
Gm1866	processed_pseudogene	predicted gene 1866 [Source:MGI Symbol;Acc:MGI:3037724]	-2.58043
Gm42997	lincRNA	predicted gene 42997 [Source:MGI Symbol;Acc:MGI:5663134]	-2.57943
Gm12480	lincRNA	predicted gene 12480 [Source:MGI Symbol;Acc:MGI:3651952]	-2.57809
Gm28791	sense_intronic	predicted gene 28791 [Source:MGI Symbol;Acc:MGI:5579497]	-2.57541
Gabrr1	protein_coding	gamma-aminobutyric acid (GABA) C receptor, subunit rho 1 [Source:MGI Symbol;Acc:MGI:95625]	-2.57091
Gm2238	TEC	predicted gene 2238 [Source:MGI Symbol;Acc:MGI:3780408]	-2.54852
Dnah7c	protein_coding	dynein, axonemal, heavy chain 7C [Source:MGI Symbol;Acc:MGI:3639762]	-2.54087
Gm43412	TEC	predicted gene 43412 [Source:MGI Symbol;Acc:MGI:5663549]	-2.53913
Gm4884	protein_coding	predicted gene 4884 [Source:MGI Symbol;Acc:MGI:3649090]	-2.53054
Adad1	protein_coding	adenosine deaminase domain containing 1 (testis specific) [Source:MGI Symbol;Acc:MGI:103258]	-2.51935
Cd6	protein_coding	CD6 antigen [Source:MGI Symbol;Acc:MGI:103566]	-2.51738
Gm7389	processed_pseudogene	predicted gene 7389 [Source:MGI Symbol;Acc:MGI:3645153]	-2.51336
Gm17907	processed_pseudogene	predicted gene, 17907 [Source:MGI Symbol;Acc:MGI:5010092]	-2.51296
Gm42556	antisense	predicted gene 42556 [Source:MGI Symbol;Acc:MGI:5662693]	-2.51058
Gm35206	antisense	predicted gene, 35206 [Source:MGI Symbol;Acc:MGI:5594365]	-2.51013
Runx2	protein_coding	runt related transcription factor 2 [Source:MGI Symbol;Acc:MGI:99829]	-2.50902
1110003F10Rik	TEC	RIKEN cDNA 1110003F10 gene [Source:MGI Symbol;Acc:MGI:1915726]	-2.50354
Foxp2	protein_coding	forkhead box P2 [Source:MGI Symbol;Acc:MGI:2148705]	-2.49933
Nxf7	protein_coding	nuclear RNA export factor 7 [Source:MGI Symbol;Acc:MGI:2159343]	-2.49721
Gm43713	antisense	predicted gene 43713 [Source:MGI Symbol;Acc:MGI:5663850]	-2.49344
Gm38357	TEC	predicted gene, 38357 [Source:MGI Symbol;Acc:MGI:5611585]	-2.49201
Abcc6	protein_coding	ATP-binding cassette, sub-family C (CFTR/MRP), member 6 [Source:MGI Symbol;Acc:MGI:1351634]	-2.48291
1700095A21Rik	antisense	RIKEN cDNA 1700095A21 gene [Source:MGI Symbol;Acc:MGI:1914774]	-2.45727
Gm7451	processed_pseudogene	predicted gene 7451 [Source:MGI Symbol;Acc:MGI:3645148]	-2.45437
Gm38047	TEC	predicted gene, 38047 [Source:MGI Symbol;Acc:MGI:5611275]	-2.45233
Cckbr	protein_coding	cholecystokinin B receptor [Source:MGI Symbol;Acc:MGI:99479]	-2.44985
Gm6277	antisense	predicted gene 6277 [Source:MGI Symbol;Acc:MGI:3779583]	-2.44864
Gm19144	processed_pseudogene	predicted gene, 19144 [Source:MGI Symbol;Acc:MGI:5011329]	-2.44167
Naip5	protein_coding	NLR family, apoptosis inhibitory protein 5 [Source:MGI Symbol;Acc:MGI:1298220]	-2.43802
Gm37062	sense_intronic	predicted gene, 37062 [Source:MGI Symbol;Acc:MGI:5610290]	-2.43481
Kif19a	protein_coding	kinesin family member 19A [Source:MGI Symbol;Acc:MGI:2447024]	-2.43258
Ppef1	protein_coding	protein phosphatase with EF hand calcium-binding domain 1 [Source:MGI Symbol;Acc:MGI:1097157]	-2.42731
Kcnc1	protein_coding	potassium voltage gated channel, Shaw-related subfamily, member 1 [Source:MGI Symbol;Acc:MGI:96667]	-2.42119
4930500M09Rik	antisense	RIKEN cDNA 4930500M09 gene [Source:MGI Symbol;Acc:MGI:3630173]	-2.41948
A730015C16Rik	protein_coding	RIKEN cDNA A730015C16 gene [Source:MGI Symbol;Acc:MGI:3704235]	-2.41225
Spag8	protein_coding	sperm associated antigen 8 [Source:MGI Symbol;Acc:MGI:3056295]	-2.41093
Gm10417	TEC	predicted gene 10417 [Source:MGI Symbol;Acc:MGI:3642677]	-2.40171
Mccc1os	antisense	methylcrotonoyl-Coenzyme A carboxylase 1 (alpha), opposite strand [Source:MGI Symbol;Acc:MGI:3590669]	-2.39953
Gm10570	protein_coding	predicted gene 10570 [Source:MGI Symbol;Acc:MGI:3642427]	-2.39831
Proz	protein_coding	protein Z, vitamin K-dependent plasma glycoprotein [Source:MGI Symbol;Acc:MGI:1860488]	-2.39682
Vmn2r-ps103	unprocessed_pseudogene	vomeronasal 2, receptor, pseudogene 103 [Source:MGI Symbol;Acc:MGI:3761522]	-2.39635
Vmn2r27	protein_coding	vomeronasal 2, receptor27 [Source:MGI Symbol;Acc:MGI:3761517]	-2.39271
Boll	protein_coding	boule homolog, RNA binding protein [Source:MGI Symbol;Acc:MGI:1922638]	-2.39002
Gm26574	antisense	predicted gene, 26574 [Source:MGI Symbol;Acc:MGI:5477068]	-2.39
Myocd	protein_coding	myocardin [Source:MGI Symbol;Acc:MGI:2137495]	-2.38788
Lsmem2	lincRNA	leucine-rich single-pass membrane protein 2 [Source:MGI Symbol;Acc:MGI:3612240]	-2.387
BC107364	protein_coding	cDNA sequence BC107364 [Source:MGI Symbol;Acc:MGI:3618860]	-2.38695
Gm12359	antisense	predicted gene 12359 [Source:MGI Symbol;Acc:MGI:3702436]	-2.38647
Gm43593	TEC	predicted gene 43593 [Source:MGI Symbol;Acc:MGI:5663730]	-2.38552
Gm44699	sense_intronic	predicted gene 44699 [Source:MGI Symbol;Acc:MGI:5753275]	-2.3787
Gm13552	processed_pseudogene	predicted gene 13552 [Source:MGI Symbol;Acc:MGI:3650819]	-2.37736
Olfr1369-ps1	protein_coding	olfactory receptor 1369, pseudogene 1 [Source:MGI Symbol;Acc:MGI:3031203]	-2.37714
1700029M20Rik	lincRNA	RIKEN cDNA 1700029M20 gene [Source:MGI Symbol;Acc:MGI:1913597]	-2.3687
Zdhhc22	protein_coding	zinc finger, DHHC-type containing 22 [Source:MGI Symbol;Acc:MGI:2685108]	-2.36241
Col17a1	protein_coding	collagen, type XVII, alpha 1 [Source:MGI Symbol;Acc:MGI:88450]	-2.36238
Gm6104	processed_pseudogene	predicted gene 6104 [Source:MGI Symbol;Acc:MGI:3648587]	-2.35864
Gm27000	lincRNA	predicted gene, 27000 [Source:MGI Symbol;Acc:MGI:5504115]	-2.35845

Gm38359	TEC	predicted gene, 38359 [Source:MGI Symbol;Acc:MGI:5611587]	-2.35843
Rai2	protein_coding	retinoic acid induced 2 [Source:MGI Symbol;Acc:MGI:1344378]	-2.35572
Fbxo40	protein_coding	F-box protein 40 [Source:MGI Symbol;Acc:MGI:2443753]	-2.35492
Fxyd4	protein_coding	FXYD domain-containing ion transport regulator 4 [Source:MGI Symbol;Acc:MGI:1889005]	-2.3487
Slc2a9	protein_coding	solute carrier family 2 (facilitated glucose transporter), member 9 [Source:MGI Symbol;Acc:MGI:2152844]	-2.34323
Tbxas1	protein_coding	thromboxane A synthase 1, platelet [Source:MGI Symbol;Acc:MGI:98497]	-2.34242
Olfr1494	protein_coding	olfactory receptor 1494 [Source:MGI Symbol;Acc:MGI:3031328]	-2.33735
Pcdhb20	protein_coding	protocadherin beta 20 [Source:MGI Symbol;Acc:MGI:2136758]	-2.33435
Dusp15	protein_coding	dual specificity phosphatase-like 15 [Source:MGI Symbol;Acc:MGI:1934928]	-2.33266
9530052E02Rik	processed_transcript	RIKEN cDNA 9530052E02 gene [Source:MGI Symbol;Acc:MGI:3588277]	-2.33236
Cct8l1	protein_coding	chaperonin containing TCP1, subunit 8 (theta)-like 1 [Source:MGI Symbol;Acc:MGI:2685289]	-2.33154
Gm10752	TEC	predicted gene 10752 [Source:MGI Symbol;Acc:MGI:3642645]	-2.33006
Gm18905	processed_pseudogene	predicted gene, 18905 [Source:MGI Symbol;Acc:MGI:5011090]	-2.32833
Gm12590	lincRNA	predicted gene 12590 [Source:MGI Symbol;Acc:MGI:3650370]	-2.325
Gm17971	processed_pseudogene	predicted gene, 17971 [Source:MGI Symbol;Acc:MGI:5010156]	-2.32173
Crb1	protein_coding	crumbs family member 1, photoreceptor morphogenesis associated [Source:MGI Symbol;Acc:MGI:2136343]	-2.31806
Gm15758	processed_transcript	predicted gene 15758 [Source:MGI Symbol;Acc:MGI:3783201]	-2.31529
Plscr4	protein_coding	phospholipid scramblase 4 [Source:MGI Symbol;Acc:MGI:2143267]	-2.30728
Gm7902	processed_pseudogene	predicted gene 7902 [Source:MGI Symbol;Acc:MGI:3645041]	-2.29878
Ocstamp	protein_coding	osteoclast stimulatory transmembrane protein [Source:MGI Symbol;Acc:MGI:1921864]	-2.29515
Gm38120	TEC	predicted gene, 38120 [Source:MGI Symbol;Acc:MGI:5611348]	-2.28463
Gm45558	lincRNA	predicted gene 45558 [Source:MGI Symbol;Acc:MGI:5791394]	-2.28367
Gm37516	TEC	predicted gene, 37516 [Source:MGI Symbol;Acc:MGI:5610744]	-2.28164
Gm26876	lincRNA	predicted gene, 26876 [Source:MGI Symbol;Acc:MGI:5477370]	-2.27573
Gm29480	antisense	predicted gene 29480 [Source:MGI Symbol;Acc:MGI:5580186]	-2.27405
Gm22993	misc_RNA	predicted gene, 22993 [Source:MGI Symbol;Acc:MGI:5452770]	-2.27132
AA413626	processed_pseudogene	expressed sequence AA413626 [Source:MGI Symbol;Acc:MGI:3035311]	-2.26764
Gm37074	TEC	predicted gene, 37074 [Source:MGI Symbol;Acc:MGI:5610302]	-2.26635
Gm20554	lincRNA	predicted gene, 20554 [Source:MGI Symbol;Acc:MGI:5295661]	-2.26346
Gm7867	processed_pseudogene	predicted gene 7867 [Source:MGI Symbol;Acc:MGI:3644627]	-2.26149
Gm11953	processed_pseudogene	predicted gene 11953 [Source:MGI Symbol;Acc:MGI:3649564]	-2.25604
Zp3r	protein_coding	zona pellucida 3 receptor [Source:MGI Symbol;Acc:MGI:104965]	-2.25387
Gm14288	protein_coding	predicted gene 14288 [Source:MGI Symbol;Acc:MGI:3706570]	-2.25271
BC030499	protein_coding	cDNA sequence BC030499 [Source:MGI Symbol;Acc:MGI:2652869]	-2.2519
Gm43331	sense_intronic	predicted gene 43331 [Source:MGI Symbol;Acc:MGI:5663468]	-2.25185
Gm44017	sense_intronic	predicted gene, 44017 [Source:MGI Symbol;Acc:MGI:5690409]	-2.2491
Gm5464	protein_coding	predicted gene 5464 [Source:MGI Symbol;Acc:MGI:3643060]	-2.24897
Gm12696	processed_pseudogene	predicted gene 12696 [Source:MGI Symbol;Acc:MGI:3649288]	-2.2446
Ndnf	protein_coding	neuron-derived neurotrophic factor [Source:MGI Symbol;Acc:MGI:1915419]	-2.24445
Gm8038	processed_pseudogene	predicted gene 8038 [Source:MGI Symbol;Acc:MGI:3644148]	-2.24214
Rn7sk	misc_RNA	RNA, 7SK, nuclear [Source:MGI Symbol;Acc:MGI:103186]	-2.2407
Cpxm2	protein_coding	carboxypeptidase X 2 (M14 family) [Source:MGI Symbol;Acc:MGI:1926006]	-2.23726
Gm26857	lincRNA	predicted gene, 26857 [Source:MGI Symbol;Acc:MGI:5477351]	-2.23091
Ces2e	protein_coding	carboxylesterase 2E [Source:MGI Symbol;Acc:MGI:2443170]	-2.22927
Gm45120	TEC	predicted gene 45120 [Source:MGI Symbol;Acc:MGI:5753696]	-2.22849
Nostrin	protein_coding	nitric oxide synthase trafficker [Source:MGI Symbol;Acc:MGI:3606242]	-2.22443
Gm10857	lincRNA	predicted gene 10857 [Source:MGI Symbol;Acc:MGI:3641733]	-2.22202
Efcab9	protein_coding	EF-hand calcium binding domain 9 [Source:MGI Symbol;Acc:MGI:1916556]	-2.22068
Hrh2	protein_coding	histamine receptor H2 [Source:MGI Symbol;Acc:MGI:108482]	-2.21776
Gypa	protein_coding	glycophorin A [Source:MGI Symbol;Acc:MGI:95880]	-2.21257
Gm42648	TEC	predicted gene 42648 [Source:MGI Symbol;Acc:MGI:5662785]	-2.20919
Gm1947	processed_pseudogene	predicted pseudogene 1947 [Source:MGI Symbol;Acc:MGI:3037805]	-2.20498
Chrnd	protein_coding	cholinergic receptor, nicotinic, delta polypeptide [Source:MGI Symbol;Acc:MGI:87893]	-2.2038
Gm26558	protein_coding	predicted gene, 26558 [Source:MGI Symbol;Acc:MGI:5477052]	-2.20038
Kcp	protein_coding	kielin/chordin-like protein [Source:MGI Symbol;Acc:MGI:2141640]	-2.19756
4930444F02Rik	lincRNA	RIKEN cDNA 4930444F02 gene [Source:MGI Symbol;Acc:MGI:1921218]	-2.19461
Atoh1	protein_coding	atonal bHLH transcription factor 1 [Source:MGI Symbol;Acc:MGI:104654]	-2.19323
Olfr772	protein_coding	olfactory receptor 772 [Source:MGI Symbol;Acc:MGI:3030606]	-2.191
Gm5822	processed_pseudogene	predicted gene 5822 [Source:MGI Symbol;Acc:MGI:3648836]	-2.18888
Gm14402	unprocessed_pseudogene	predicted gene 14402 [Source:MGI Symbol;Acc:MGI:3649812]	-2.18527
Gm18737	processed_pseudogene	predicted gene, 18737 [Source:MGI Symbol;Acc:MGI:5010922]	-2.18488
Gm16161	antisense	predicted gene 16161 [Source:MGI Symbol;Acc:MGI:3801899]	-2.18415
Sv2b	protein_coding	synaptic vesicle glycoprotein 2 b [Source:MGI Symbol;Acc:MGI:1927338]	-2.18197
Gm43336	TEC	predicted gene 43336 [Source:MGI Symbol;Acc:MGI:5663473]	-2.18191
Gm19045	processed_pseudogene	predicted gene, 19045 [Source:MGI Symbol;Acc:MGI:5011230]	-2.17949
L3mbtl1	protein_coding	L3MBTL1 histone methyl-lysine binding protein [Source:MGI Symbol;Acc:MGI:2676663]	-2.17889
Atp2c2	protein_coding	ATPase, Ca++ transporting, type 2C, member 2 [Source:MGI Symbol;Acc:MGI:1916297]	-2.17872

Entpd3	protein_coding	ectonucleoside triphosphate diphosphohydrolase 3 [Source:MGI Symbol;Acc:MGI:1321386]	-2.17085
Npc1l1	protein_coding	NPC1 like intracellular cholesterol transporter 1 [Source:MGI Symbol;Acc:MGI:2685089]	-2.1687
Yipf7	protein_coding	Yip1 domain family, member 7 [Source:MGI Symbol;Acc:MGI:1922831]	-2.16838
Gm18959	processed_pseudogene	predicted gene, 18959 [Source:MGI Symbol;Acc:MGI:5011144]	-2.16702
4633401B06Rik	TEC	RIKEN cDNA 4633401B06 gene [Source:MGI Symbol;Acc:MGI:1918078]	-2.16483
Gm38944	lincRNA	predicted gene, 38944 [Source:MGI Symbol;Acc:MGI:5621829]	-2.16266
Gm8822	processed_pseudogene	predicted gene 8822 [Source:MGI Symbol;Acc:MGI:3648640]	-2.16218
Gm42842	TEC	predicted gene 42842 [Source:MGI Symbol;Acc:MGI:5662979]	-2.15854
2900078I11Rik	TEC	RIKEN cDNA 2900078I11 gene [Source:MGI Symbol;Acc:MGI:1920254]	-2.15739
Adgrg6	protein_coding	adhesion G protein-coupled receptor G6 [Source:MGI Symbol;Acc:MGI:1916151]	-2.15627
Cryba1	protein_coding	crystallin, beta A1 [Source:MGI Symbol;Acc:MGI:88518]	-2.15528
Gm43258	sense_intronic	predicted gene 43258 [Source:MGI Symbol;Acc:MGI:5663395]	-2.1551
Fat2	protein_coding	FAT atypical cadherin 2 [Source:MGI Symbol;Acc:MGI:2685369]	-2.15291
Gm21542	processed_pseudogene	predicted gene, 21542 [Source:MGI Symbol;Acc:MGI:5434897]	-2.15228
Gm37558	TEC	predicted gene, 37558 [Source:MGI Symbol;Acc:MGI:5610786]	-2.14997
Gm45740	lincRNA	predicted gene 45740 [Source:MGI Symbol;Acc:MGI:5804855]	-2.14275
Gm16128	processed_pseudogene	predicted gene 16128 [Source:MGI Symbol;Acc:MGI:3801948]	-2.13996
2010109I03Rik	protein_coding	RIKEN cDNA 2010109I03 gene [Source:MGI Symbol;Acc:MGI:1914288]	-2.13819
Casc1	protein_coding	cancer susceptibility candidate 1 [Source:MGI Symbol;Acc:MGI:2444480]	-2.13431
Gm45413	TEC	predicted gene 45413 [Source:MGI Symbol;Acc:MGI:5791249]	-2.13322
Nckap5los	antisense	NCK-associated protein 5-like, opposite strand [Source:MGI Symbol;Acc:MGI:1918162]	-2.13006
Gm43628	TEC	predicted gene 43628 [Source:MGI Symbol;Acc:MGI:5663765]	-2.12665
5430400D12Rik	bidirectional_promoter_lncRNA	RIKEN cDNA 5430400D12 gene [Source:MGI Symbol;Acc:MGI:1918589]	-2.12531
Tnf	protein_coding	tumor necrosis factor [Source:MGI Symbol;Acc:MGI:104798]	-2.12012
Spata9	protein_coding	spermatogenesis associated 9 [Source:MGI Symbol;Acc:MGI:1922821]	-2.11983
Ajap1	protein_coding	adherens junction associated protein 1 [Source:MGI Symbol;Acc:MGI:2685419]	-2.11753
Ccdc69	protein_coding	coiled-coil domain containing 69 [Source:MGI Symbol;Acc:MGI:1196234]	-2.11654
Slc8a3	protein_coding	solute carrier family 8 (sodium/calcium exchanger), member 3 [Source:MGI Symbol;Acc:MGI:107976]	-2.11587
Foxf2	protein_coding	forkhead box F2 [Source:MGI Symbol;Acc:MGI:1347479]	-2.1155
Dcstamp	protein_coding	dendrocyte expressed seven transmembrane protein [Source:MGI Symbol;Acc:MGI:1923016]	-2.1122
Gm3239	protein_coding	predicted gene 3239 [Source:MGI Symbol;Acc:MGI:3781417]	-2.11151
Gm32391	lincRNA	predicted gene, 32391 [Source:MGI Symbol;Acc:MGI:5591550]	-2.10904
Dscam	protein_coding	DS cell adhesion molecule [Source:MGI Symbol;Acc:MGI:1196281]	-2.10382
Gm43071	TEC	predicted gene 43071 [Source:MGI Symbol;Acc:MGI:5663208]	-2.10181
Tbr1	protein_coding	T-box brain gene 1 [Source:MGI Symbol;Acc:MGI:107404]	-2.09846
Gm15760	transcribed_processed_pseudogene	predicted gene 15760 [Source:MGI Symbol;Acc:MGI:3642834]	-2.097
B230119M05Rik	processed_transcript	RIKEN cDNA B230119M05 gene [Source:MGI Symbol;Acc:MGI:1925233]	-2.09636
Prss32	protein_coding	protease, serine 32 [Source:MGI Symbol;Acc:MGI:1917064]	-2.09489
Gm6133	protein_coding	predicted gene 6133 [Source:MGI Symbol;Acc:MGI:3647029]	-2.09134
Gm43335	TEC	predicted gene 43335 [Source:MGI Symbol;Acc:MGI:5663472]	-2.08587
Gm44199	antisense	predicted gene, 44199 [Source:MGI Symbol;Acc:MGI:5690591]	-2.08433
Mill2	protein_coding	MHC I like leukocyte 2 [Source:MGI Symbol;Acc:MGI:2179989]	-2.07542
Sytl4	protein_coding	synaptotagmin-like 4 [Source:MGI Symbol;Acc:MGI:1351606]	-2.06757
E030022I16Rik	TEC	RIKEN cDNA E030022I16 gene [Source:MGI Symbol;Acc:MGI:2444690]	-2.06617
Gm22806	snoRNA	predicted gene, 22806 [Source:MGI Symbol;Acc:MGI:5452583]	-2.06235
Gm21962	protein_coding	predicted gene, 21962 [Source:MGI Symbol;Acc:MGI:5439431]	-2.06173
Gm10139	protein_coding	predicted gene 10139 [Source:MGI Symbol;Acc:MGI:3642815]	-2.06059
2210408F21Rik	processed_transcript	RIKEN cDNA 2210408F21 gene [Source:MGI Symbol;Acc:MGI:1920902]	-2.0602
M5C1000I18Rik	lincRNA	RIKEN cDNA M5C1000I18 gene [Source:MGI Symbol;Acc:MGI:3642038]	-2.0553
Nfia	protein_coding	nuclear factor I/A [Source:MGI Symbol;Acc:MGI:108056]	-2.05348
Gm45493	TEC	predicted gene 45493 [Source:MGI Symbol;Acc:MGI:5791329]	-2.05138
Fibcd1	protein_coding	fibrinogen C domain containing 1 [Source:MGI Symbol;Acc:MGI:2138953]	-2.05121
Kctd19	protein_coding	potassium channel tetramerisation domain containing 19 [Source:MGI Symbol;Acc:MGI:3045294]	-2.05001
Gm13523	antisense	predicted gene 13523 [Source:MGI Symbol;Acc:MGI:3651426]	-2.04891
A830073O21Rik	TEC	RIKEN cDNA A830073O21 gene [Source:MGI Symbol;Acc:MGI:2443692]	-2.04755
Gm13061	processed_transcript	predicted gene 13061 [Source:MGI Symbol;Acc:MGI:3650259]	-2.04423
Vmn1r58	protein_coding	vomeronasal 1 receptor 58 [Source:MGI Symbol;Acc:MGI:3033473]	-2.04337
Gm16907	lincRNA	predicted gene, 16907 [Source:MGI Symbol;Acc:MGI:4439831]	-2.04232
Ttc6	protein_coding	tetratricopeptide repeat domain 6 [Source:MGI Symbol;Acc:MGI:2684915]	-2.04157
Gm37570	TEC	predicted gene, 37570 [Source:MGI Symbol;Acc:MGI:5610798]	-2.0406
B9d1os	antisense	B9 protein domain 1, opposite strand [Source:MGI Symbol;Acc:MGI:1920466]	-2.0389
Olfr655	protein_coding	olfactory receptor 655 [Source:MGI Symbol;Acc:MGI:3030489]	-2.03514
Gm26546	lincRNA	predicted gene, 26546 [Source:MGI Symbol;Acc:MGI:5477040]	-2.03229
Tmem235	protein_coding	transmembrane protein 235 [Source:MGI Symbol;Acc:MGI:3651706]	-2.02884
Cbln2	protein_coding	cerebellin 2 precursor protein [Source:MGI Symbol;Acc:MGI:88282]	-2.02833
1700034H15Rik	antisense	RIKEN cDNA 1700034H15 gene [Source:MGI Symbol;Acc:MGI:1921515]	-2.02829
1700034H15Rik	antisense	RIKEN cDNA 1700034H15 gene [Source:MGI Symbol;Acc:MGI:1921515]	-2.02829

Vipr1	protein_coding	vasoactive intestinal peptide receptor 1 [Source:MGI Symbol;Acc:MGI:109272]	-2.0256
Gm23639	snoRNA	predicted gene, 23639 [Source:MGI Symbol;Acc:MGI:5453416]	-2.02547
Gm26603	lincRNA	predicted gene, 26603 [Source:MGI Symbol;Acc:MGI:5477097]	-2.02304
Gm43654	TEC	predicted gene 43654 [Source:MGI Symbol;Acc:MGI:5663791]	-2.017
Tcea1-ps1	processed_pseudogene	transcription elongation factor A (SII) 1, pseudogene 1 [Source:MGI Symbol;Acc:MGI:1861432]	-2.01525
Opn4	protein_coding	opsin 4 (melanopsin) [Source:MGI Symbol;Acc:MGI:1353425]	-2.00944
Gm11218	processed_pseudogene	predicted gene 11218 [Source:MGI Symbol;Acc:MGI:3651636]	-2.00932
Rps12-ps24	processed_pseudogene	ribosomal protein S12, pseudogene 24 [Source:MGI Symbol;Acc:MGI:3642518]	-2.00746
Gcm1	protein_coding	glial cells missing homolog 1 [Source:MGI Symbol;Acc:MGI:108045]	-2.00664
Cdh7	protein_coding	cadherin 7, type 2 [Source:MGI Symbol;Acc:MGI:2442792]	-2.00347
Gbgt1	protein_coding	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1 [Source:MGI Symbol;Acc:MGI:2449143]	-1.9964
Hist1h3i	protein_coding	histone cluster 1, H3i [Source:MGI Symbol;Acc:MGI:2448350]	-1.99623
Spocd1	protein_coding	SPOC domain containing 1 [Source:MGI Symbol;Acc:MGI:3652045]	-1.99595
Gm44321	TEC	predicted gene, 44321 [Source:MGI Symbol;Acc:MGI:5690713]	-1.99306
Dok5	protein_coding	docking protein 5 [Source:MGI Symbol;Acc:MGI:1924079]	-1.99087
Pyroxd2	protein_coding	pyridine nucleotide-disulphide oxidoreductase domain 2 [Source:MGI Symbol;Acc:MGI:1921830]	-1.98812
C2cd6	protein_coding	C2 calcium dependent domain containing 6 [Source:MGI Symbol;Acc:MGI:1920713]	-1.98649
Pla2r1	protein_coding	phospholipase A2 receptor 1 [Source:MGI Symbol;Acc:MGI:102468]	-1.98465
Aqp7	protein_coding	aquaporin 7 [Source:MGI Symbol;Acc:MGI:1314647]	-1.97981
Pcsk5	protein_coding	proprotein convertase subtilisin/kexin type 5 [Source:MGI Symbol;Acc:MGI:97515]	-1.97484
Higd1b	protein_coding	HIG1 domain family, member 1B [Source:MGI Symbol;Acc:MGI:1922939]	-1.97471
Gm9081	processed_pseudogene	predicted gene 9081 [Source:MGI Symbol;Acc:MGI:3644454]	-1.96821
Gm8825	processed_pseudogene	predicted gene 8825 [Source:MGI Symbol;Acc:MGI:3647098]	-1.96782
Gm20628	antisense	predicted gene 20628 [Source:MGI Symbol;Acc:MGI:5313075]	-1.96088
Gm43293	lincRNA	predicted gene 43293 [Source:MGI Symbol;Acc:MGI:5663430]	-1.96063
Mst1	protein_coding	macrophage stimulating 1 (hepatocyte growth factor-like) [Source:MGI Symbol;Acc:MGI:96080]	-1.95824
Gm37568	processed_pseudogene	predicted gene, 37568 [Source:MGI Symbol;Acc:MGI:5610796]	-1.9577
Birc7	protein_coding	baculoviral IAP repeat-containing 7 (livin) [Source:MGI Symbol;Acc:MGI:2676458]	-1.95604
Tmc5	protein_coding	transmembrane channel-like gene family 5 [Source:MGI Symbol;Acc:MGI:1921674]	-1.95509
Gm12403	processed_pseudogene	predicted gene 12403 [Source:MGI Symbol;Acc:MGI:3650109]	-1.95343
Gm41231	lincRNA	predicted gene, 41231 [Source:MGI Symbol;Acc:MGI:5624116]	-1.95327
Gdf10	protein_coding	growth differentiation factor 10 [Source:MGI Symbol;Acc:MGI:95684]	-1.95133
Tfr2	protein_coding	transferrin receptor 2 [Source:MGI Symbol;Acc:MGI:1354956]	-1.95131
Efcab8	protein_coding	EF-hand calcium binding domain 8 [Source:MGI Symbol;Acc:MGI:3644206]	-1.95077
Slc38a5	protein_coding	solute carrier family 38, member 5 [Source:MGI Symbol;Acc:MGI:2148066]	-1.946
Gm10772	protein_coding	predicted gene 10772 [Source:MGI Symbol;Acc:MGI:3704404]	-1.94579
A430073D23Rik	lincRNA	RIKEN cDNA A430073D23 gene [Source:MGI Symbol;Acc:MGI:4361019]	-1.94358
Gm9762	processed_pseudogene	predicted pseudogene 9762 [Source:MGI Symbol;Acc:MGI:3704220]	-1.93845
Uts2r	protein_coding	urotensin 2 receptor [Source:MGI Symbol;Acc:MGI:2183450]	-1.93029
Gm5687	processed_pseudogene	predicted gene 5687 [Source:MGI Symbol;Acc:MGI:3644274]	-1.93022
Tbx19	protein_coding	T-box 19 [Source:MGI Symbol;Acc:MGI:1891158]	-1.92391
Wscd2	protein_coding	WSC domain containing 2 [Source:MGI Symbol;Acc:MGI:2445030]	-1.92348
Gcnt7	protein_coding	glucosaminyl (N-acetyl) transferase family member 7 [Source:MGI Symbol;Acc:MGI:3606143]	-1.92272
Chrna3	protein_coding	cholinergic receptor, nicotinic, alpha polypeptide 3 [Source:MGI Symbol;Acc:MGI:87887]	-1.92004
Rpl15-ps3	processed_pseudogene	ribosomal protein L15, pseudogene 3 [Source:MGI Symbol;Acc:MGI:3782952]	-1.91989
Olfr922	protein_coding	olfactory receptor 922 [Source:MGI Symbol;Acc:MGI:3030756]	-1.91962
Unc5d	protein_coding	unc-5 netrin receptor D [Source:MGI Symbol;Acc:MGI:2389364]	-1.9187
Gm26781	antisense	predicted gene, 26781 [Source:MGI Symbol;Acc:MGI:5477275]	-1.91863
Gm10263	processed_pseudogene	predicted gene 10263 [Source:MGI Symbol;Acc:MGI:3642825]	-1.91451
Chst8	protein_coding	carbohydrate (N-acetylgalactosamine 4-0) sulfotransferase 8 [Source:MGI Symbol;Acc:MGI:1916197]	-1.90953
Gm25360	snRNA	predicted gene, 25360 [Source:MGI Symbol;Acc:MGI:5455137]	-1.90893
Gm9761	processed_pseudogene	predicted gene 9761 [Source:MGI Symbol;Acc:MGI:3708640]	-1.90886
Gm16167	processed_pseudogene	predicted gene 16167 [Source:MGI Symbol;Acc:MGI:3802049]	-1.90737
Gm14044	processed_pseudogene	predicted gene 14044 [Source:MGI Symbol;Acc:MGI:3650452]	-1.90704
4933430L12Rik	antisense	RIKEN cDNA 4933430L12 gene [Source:MGI Symbol;Acc:MGI:1918501]	-1.90537
Slc6a14	protein_coding	solute carrier family 6 (neurotransmitter transporter), member 14 [Source:MGI Symbol;Acc:MGI:1890216]	-1.90468
Gm10432	processed_transcript	predicted gene 10432 [Source:MGI Symbol;Acc:MGI:3642011]	-1.90143
Gm16410	processed_pseudogene	predicted gene 16410 [Source:MGI Symbol;Acc:MGI:3648382]	-1.9006
Apoo-ps	processed_pseudogene	apolipoprotein O, pseudogene [Source:MGI Symbol;Acc:MGI:3649039]	-1.90018
Pdgfd	protein_coding	platelet-derived growth factor, D polypeptide [Source:MGI Symbol;Acc:MGI:1919035]	-1.89977
Gm29202	antisense	predicted gene 29202 [Source:MGI Symbol;Acc:MGI:5579908]	-1.8981

Gm13357	processed_pseudogene	predicted gene 13357 [Source:MGI Symbol;Acc:MGI:3651263]	-1.89702
Gm15267	processed_pseudogene	predicted gene 15267 [Source:MGI Symbol;Acc:MGI:3705447]	-1.89374
Avpr1a	protein_coding	arginine vasopressin receptor 1A [Source:MGI Symbol;Acc:MGI:1859216]	-1.88935
Gm4032	transcribed_processed_pseudogene	predicted gene 4032 [Source:MGI Symbol;Acc:MGI:3782206]	-1.88923
Gm38080	processed_pseudogene	predicted gene, 38080 [Source:MGI Symbol;Acc:MGI:5611308]	-1.88899
Gm6855	processed_pseudogene	predicted gene 6855 [Source:MGI Symbol;Acc:MGI:3648447]	-1.88884
Tcp11x2	protein_coding	t-complex 11 family, X-linked 2 [Source:MGI Symbol;Acc:MGI:1919091]	-1.88853
Gm45183	TEC	predicted gene 45183 [Source:MGI Symbol;Acc:MGI:5753759]	-1.88796
Gm11365	antisense	predicted gene 11365 [Source:MGI Symbol;Acc:MGI:3652070]	-1.88725
Ms4a1	protein_coding	membrane-spanning 4-domains, subfamily A, member 1 [Source:MGI Symbol;Acc:MGI:88321]	-1.88473
Gm43072	TEC	predicted gene 43072 [Source:MGI Symbol;Acc:MGI:5663209]	-1.88268
Pde4b	protein_coding	phosphodiesterase 4B, cAMP specific [Source:MGI Symbol;Acc:MGI:99557]	-1.88248
Bnip3l-ps	transcribed_processed_pseudogene	BCL2/adenovirus E1B interacting protein 3-like, pseudogene [Source:MGI Symbol;Acc:MGI:3642435]	-1.88128
4930519P11Rik	protein_coding	RIKEN cDNA 4930519P11 gene [Source:MGI Symbol;Acc:MGI:1921971]	-1.8802
Gm23442	snoRNA	predicted gene, 23442 [Source:MGI Symbol;Acc:MGI:5453219]	-1.88002
Hoxc4	protein_coding	homeobox C4 [Source:MGI Symbol;Acc:MGI:96195]	-1.87857
Gm44045	TEC	predicted gene, 44045 [Source:MGI Symbol;Acc:MGI:5690437]	-1.8784
Gm27006	processed_pseudogene	predicted gene, 27006 [Source:MGI Symbol;Acc:MGI:5504121]	-1.87825
Hrg	protein_coding	histidine-rich glycoprotein [Source:MGI Symbol;Acc:MGI:2146636]	-1.87223
Col26a1	protein_coding	collagen, type XXVI, alpha 1 [Source:MGI Symbol;Acc:MGI:2155345]	-1.87196
Gm6556	processed_transcript	predicted gene 6556 [Source:MGI Symbol;Acc:MGI:3647197]	-1.87143
Cetn4	protein_coding	centrin 4 [Source:MGI Symbol;Acc:MGI:2677454]	-1.87089
A530040E14Rik	antisense	RIKEN cDNA A530040E14 gene [Source:MGI Symbol;Acc:MGI:3612703]	-1.8693
Gm38122	TEC	predicted gene, 38122 [Source:MGI Symbol;Acc:MGI:5611350]	-1.86833
Tex16	protein_coding	testis expressed gene 16 [Source:MGI Symbol;Acc:MGI:1890545]	-1.86812
Smad9	protein_coding	SMAD family member 9 [Source:MGI Symbol;Acc:MGI:1859993]	-1.86555
Gm8290	processed_pseudogene	predicted gene 8290 [Source:MGI Symbol;Acc:MGI:3648461]	-1.86513
9130024F11Rik	processed_transcript	RIKEN cDNA 9130024F11 gene [Source:MGI Symbol;Acc:MGI:1926150]	-1.86466
Slc22a3	protein_coding	solute carrier family 22 (organic cation transporter), member 3 [Source:MGI Symbol;Acc:MGI:1333817]	-1.86385
Olfml3	protein_coding	olfactomedin-like 3 [Source:MGI Symbol;Acc:MGI:1914877]	-1.86368
Thbd	protein_coding	thrombomodulin [Source:MGI Symbol;Acc:MGI:98736]	-1.86355
Nol4	protein_coding	nucleolar protein 4 [Source:MGI Symbol;Acc:MGI:2441684]	-1.86315
Gm45053	TEC	predicted gene 45053 [Source:MGI Symbol;Acc:MGI:5753629]	-1.86212
Gm26823	lincRNA	predicted gene, 26823 [Source:MGI Symbol;Acc:MGI:5477317]	-1.86015
A230028O05Rik	lincRNA	RIKEN cDNA A230028O05 gene [Source:MGI Symbol;Acc:MGI:2442126]	-1.86013
Cldn11	protein_coding	claudin 11 [Source:MGI Symbol;Acc:MGI:106925]	-1.85883
Olfr398	protein_coding	olfactory receptor 398 [Source:MGI Symbol;Acc:MGI:3030232]	-1.85873
Dnah10	protein_coding	dynein, axonemal, heavy chain 10 [Source:MGI Symbol;Acc:MGI:1860299]	-1.85817
Gm43759	TEC	predicted gene 43759 [Source:MGI Symbol;Acc:MGI:5663896]	-1.85732
Rarb	protein_coding	retinoic acid receptor, beta [Source:MGI Symbol;Acc:MGI:97857]	-1.85496
Heyl	protein_coding	hairy/enhancer-of-split related with YRPW motif-like [Source:MGI Symbol;Acc:MGI:1860511]	-1.8545
Dlx3	protein_coding	distal-less homeobox 3 [Source:MGI Symbol;Acc:MGI:94903]	-1.85345
Gm9826	processed_pseudogene	predicted gene 9826 [Source:MGI Symbol;Acc:MGI:3642725]	-1.85248
Lrrc55	protein_coding	leucine rich repeat containing 55 [Source:MGI Symbol;Acc:MGI:2685197]	-1.85095
Gm15976	antisense	predicted gene 15976 [Source:MGI Symbol;Acc:MGI:3801898]	-1.85084
Fas	protein_coding	Fas (TNF receptor superfamily member 6) [Source:MGI Symbol;Acc:MGI:95484]	-1.8475
Gm28085	antisense	predicted gene 28085 [Source:MGI Symbol;Acc:MGI:5578791]	-1.84706
Gm10425	antisense	predicted gene 10425 [Source:MGI Symbol;Acc:MGI:3642471]	-1.84662
A630095N17Rik	protein_coding	RIKEN cDNA A630095N17 gene [Source:MGI Symbol;Acc:MGI:2686470]	-1.84565
Wnk4	protein_coding	WNK lysine deficient protein kinase 4 [Source:MGI Symbol;Acc:MGI:1917097]	-1.84448
Pik3ap1	protein_coding	phosphoinositide-3-kinase adaptor protein 1 [Source:MGI Symbol;Acc:MGI:1933177]	-1.84308
Crybb3	protein_coding	crystallin, beta B3 [Source:MGI Symbol;Acc:MGI:102717]	-1.84073
Rxfp2	protein_coding	relaxin/insulin-like family peptide receptor 2 [Source:MGI Symbol;Acc:MGI:2153463]	-1.84008
Mylk3	protein_coding	myosin light chain kinase 3 [Source:MGI Symbol;Acc:MGI:2443063]	-1.83506
Krt12	protein_coding	keratin 12 [Source:MGI Symbol;Acc:MGI:96687]	-1.83444
Gm26711	lincRNA	predicted gene, 26711 [Source:MGI Symbol;Acc:MGI:5477205]	-1.83384
Usp18	protein_coding	ubiquitin specific peptidase 18 [Source:MGI Symbol;Acc:MGI:1344364]	-1.83247
Col22a1	protein_coding	collagen, type XXII, alpha 1 [Source:MGI Symbol;Acc:MGI:1916950]	-1.83229
Tmem200b	protein_coding	transmembrane protein 200B [Source:MGI Symbol;Acc:MGI:3646343]	-1.82811
Gm43482	TEC	predicted gene 43482 [Source:MGI Symbol;Acc:MGI:5663619]	-1.82774
Ppp1r3g	protein_coding	protein phosphatase 1, regulatory subunit 3G [Source:MGI Symbol;Acc:MGI:1923737]	-1.82695
2510009E07Rik	protein_coding	RIKEN cDNA 2510009E07 gene [Source:MGI Symbol;Acc:MGI:1919440]	-1.82656
Gm12198	processed_transcript	predicted gene 12198 [Source:MGI Symbol;Acc:MGI:3650407]	-1.82644
Spo11	protein_coding	SPO11 meiotic protein covalently bound to DSB [Source:MGI Symbol;Acc:MGI:1349669]	-1.82591
Matn4	protein_coding	matrilin 4 [Source:MGI Symbol;Acc:MGI:1328314]	-1.82386
Eya4	protein_coding	EYA transcriptional coactivator and phosphatase 4 [Source:MGI	-1.82276

		Symbol;Acc:MGI:1337104]	
Apol10b	protein_coding	apolipoprotein L 10B [Source:MGI Symbol;Acc:MGI:3043522]	-1.82224
Gm44043	TEC	predicted gene, 44043 [Source:MGI Symbol;Acc:MGI:5690435]	-1.8221
Cysrt1	protein_coding	cysteine rich tail 1 [Source:MGI Symbol;Acc:MGI:1915109]	-1.82183
Gm16170	antisense	predicted gene 16170 [Source:MGI Symbol;Acc:MGI:3805548]	-1.81857
Ropn1l	protein_coding	ropporin 1-like [Source:MGI Symbol;Acc:MGI:2182357]	-1.81801
Gm7160	lincRNA	predicted gene 7160 [Source:MGI Symbol;Acc:MGI:3646078]	-1.8146
Gm36862	processed_transcript	predicted gene, 36862 [Source:MGI Symbol;Acc:MGI:5596021]	-1.81259
Gm39363	lincRNA	predicted gene, 39363 [Source:MGI Symbol;Acc:MGI:5622248]	-1.81071
B230317F23Rik	lincRNA	RIKEN cDNA B230317F23 gene [Source:MGI Symbol;Acc:MGI:2443920]	-1.80932
Gm21917	antisense	predicted gene, 21917 [Source:MGI Symbol;Acc:MGI:5434081]	-1.80829
Gm12404	processed_transcript	predicted gene 12404 [Source:MGI Symbol;Acc:MGI:3649601]	-1.80793
Ccdc81	protein_coding	coiled-coil domain containing 81 [Source:MGI Symbol;Acc:MGI:1918134]	-1.80777
Gm45041	TEC	predicted gene 45041 [Source:MGI Symbol;Acc:MGI:5753617]	-1.80316
Gm42530	lincRNA	predicted gene 42530 [Source:MGI Symbol;Acc:MGI:5662667]	-1.80193
9430025C20Rik	pseudogene	RIKEN cDNA 9430025C20 gene [Source:MGI Symbol;Acc:MGI:3045277]	-1.8011
Gm10167	processed_pseudogene	predicted pseudogene 10167 [Source:MGI Symbol;Acc:MGI:3704263]	-1.80097
Snord118	snoRNA	small nucleolar RNA, C/D box 118 [Source:MGI Symbol;Acc:MGI:3819519]	-1.79882
Gm45528	TEC	predicted gene 45528 [Source:MGI Symbol;Acc:MGI:5791364]	-1.79868
Gm37935	TEC	predicted gene, 37935 [Source:MGI Symbol;Acc:MGI:5611163]	-1.79834
Gm43360	TEC	predicted gene 43360 [Source:MGI Symbol;Acc:MGI:5663497]	-1.79514
Gm7118	processed_pseudogene	predicted gene 7118 [Source:MGI Symbol;Acc:MGI:3648119]	-1.79307
Hmx2	protein_coding	H6 homeobox 2 [Source:MGI Symbol;Acc:MGI:107159]	-1.79247
4930483J18Rik	processed_transcript	RIKEN cDNA 4930483J18 gene [Source:MGI Symbol;Acc:MGI:1914888]	-1.7917
Cxcr6	protein_coding	chemokine (C-X-C motif) receptor 6 [Source:MGI Symbol;Acc:MGI:1934582]	-1.79159
4933425D22Rik	sense_intronic	RIKEN cDNA 4933425D22 gene [Source:MGI Symbol;Acc:MGI:1914009]	-1.7903
C79798	antisense	expressed sequence C79798 [Source:MGI Symbol;Acc:MGI:2139494]	-1.78687
Gm26734	lincRNA	predicted gene, 26734 [Source:MGI Symbol;Acc:MGI:5477228]	-1.78558
4930518I15Rik	protein_coding	RIKEN cDNA 4930518I15 gene [Source:MGI Symbol;Acc:MGI:1921954]	-1.78413
C130060C02Rik	lincRNA	RIKEN cDNA C130060C02 gene [Source:MGI Symbol;Acc:MGI:3041192]	-1.78368
Gm16121	antisense	predicted gene 16121 [Source:MGI Symbol;Acc:MGI:3802131]	-1.78352
Pnliprp2	protein_coding	pancreatic lipase-related protein 2 [Source:MGI Symbol;Acc:MGI:1336202]	-1.7813
4933416C03Rik	protein_coding	RIKEN cDNA 4933416C03 gene [Source:MGI Symbol;Acc:MGI:3588269]	-1.77732
Tmem119	protein_coding	transmembrane protein 119 [Source:MGI Symbol;Acc:MGI:2385228]	-1.77369
Gm10785	antisense	predicted gene 10785 [Source:MGI Symbol;Acc:MGI:3642149]	-1.77354
Klf14	protein_coding	Kruppel-like factor 14 [Source:MGI Symbol;Acc:MGI:3577024]	-1.77236
Lhx1	protein_coding	LIM homeobox protein 1 [Source:MGI Symbol;Acc:MGI:99783]	-1.77181
B230303A05Rik	transcribed_processed_pseudogene	RIKEN cDNA B230303A05 gene [Source:MGI Symbol;Acc:MGI:3646895]	-1.7702
Gm32816	processed_transcript	predicted gene, 32816 [Source:MGI Symbol;Acc:MGI:5591975]	-1.76969
Gm40977	lincRNA	predicted gene, 40977 [Source:MGI Symbol;Acc:MGI:5623862]	-1.76936
Gm14026	processed_pseudogene	predicted gene 14026 [Source:MGI Symbol;Acc:MGI:3649712]	-1.76828
Has1	protein_coding	hyaluronan synthase 1 [Source:MGI Symbol;Acc:MGI:106590]	-1.76737
Ankub1	protein_coding	ankrin repeat and ubiquitin domain containing 1 [Source:MGI Symbol;Acc:MGI:2685256]	-1.76714
Gm43813	processed_transcript	predicted gene 43813 [Source:MGI Symbol;Acc:MGI:5663950]	-1.76611
Gm13142	processed_pseudogene	predicted gene 13142 [Source:MGI Symbol;Acc:MGI:3651745]	-1.76591
Hesx1	protein_coding	homeobox gene expressed in ES cells [Source:MGI Symbol;Acc:MGI:96071]	-1.76546
Usp17la	protein_coding	ubiquitin specific peptidase 17-like A [Source:MGI Symbol;Acc:MGI:107699]	-1.76342
Gm12419	processed_pseudogene	predicted gene 12419 [Source:MGI Symbol;Acc:MGI:3650581]	-1.76282
1700056N10Rik	lincRNA	RIKEN cDNA 1700056N10 gene [Source:MGI Symbol;Acc:MGI:1920642]	-1.76111
B630019K06Rik	protein_coding	RIKEN cDNA B630019K06 gene [Source:MGI Symbol;Acc:MGI:2147918]	-1.76027
Kcnmb1	protein_coding	potassium large conductance calcium-activated channel, subfamily M, beta member 1 [Source:MGI Symbol;Acc:MGI:1334203]	-1.7599
Fam181a	protein_coding	family with sequence similarity 181, member A [Source:MGI Symbol;Acc:MGI:3647570]	-1.75917
Rnf150	protein_coding	ring finger protein 150 [Source:MGI Symbol;Acc:MGI:2443860]	-1.75841
Gm25735	misc_RNA	predicted gene, 25735 [Source:MGI Symbol;Acc:MGI:5455512]	-1.75787
Zmynd12	protein_coding	zinc finger, MYND domain containing 12 [Source:MGI Symbol;Acc:MGI:2140259]	-1.75587
AC122821.1	processed_pseudogene	ribosomal protein L37 (Rpl37) pseudogene	-1.75363
Gm26953	lincRNA	predicted gene, 26953 [Source:MGI Symbol;Acc:MGI:5504068]	-1.75262
Mup20	protein_coding	major urinary protein 20 [Source:MGI Symbol;Acc:MGI:3651981]	-1.7518
4732463B04Rik	antisense	RIKEN cDNA 4732463B04 gene [Source:MGI Symbol;Acc:MGI:3642483]	-1.75171
Rab9b	protein_coding	RAB9B, member RAS oncogene family [Source:MGI Symbol;Acc:MGI:2442454]	-1.75118
Col6a4	protein_coding	collagen, type VI, alpha 4 [Source:MGI Symbol;Acc:MGI:1915803]	-1.7509
Gm13642	processed_pseudogene	predicted gene 13642 [Source:MGI Symbol;Acc:MGI:3651865]	-1.74685
Gm37983	TEC	predicted gene, 37983 [Source:MGI Symbol;Acc:MGI:5611211]	-1.74541
G430049J08Rik	protein_coding	RIKEN cDNA G430049J08 gene [Source:MGI Symbol;Acc:MGI:3645649]	-1.74436
Gm17968	processed_pseudogene	predicted gene, 17968 [Source:MGI Symbol;Acc:MGI:5010153]	-1.7422
Gm43982	TEC	predicted gene, 43982 [Source:MGI Symbol;Acc:MGI:5690374]	-1.74196
Gm9269	unprocessed_pseudogene	predicted gene 9269 [Source:MGI Symbol;Acc:MGI:3646451]	-1.74086
Tmed6	protein_coding	transmembrane p24 trafficking protein 6 [Source:MGI Symbol;Acc:MGI:1913519]	-1.74012
Gm3149	protein_coding	predicted gene 3149 [Source:MGI Symbol;Acc:MGI:3781328]	-1.73811

5930403N24Rik	processed_transcript	RIKEN cDNA 5930403N24 gene [Source:MGI Symbol;Acc:MGI:2444171]	-1.73486
Gm15535	sense_intronic	predicted gene 15535 [Source:MGI Symbol;Acc:MGI:3782983]	-1.73422
Gdap1	protein_coding	ganglioside-induced differentiation-associated-protein 1 [Source:MGI Symbol;Acc:MGI:1338002]	-1.73402
P2rx6	protein_coding	purinergic receptor P2X, ligand-gated ion channel, 6 [Source:MGI Symbol;Acc:MGI:1337113]	-1.73235
Mdm4-ps	processed_pseudogene	transformed mouse 3T3 cell double minute 4, pseudogene [Source:MGI Symbol;Acc:MGI:2136991]	-1.73215
D930020B18Rik	protein_coding	RIKEN cDNA D930020B18 gene [Source:MGI Symbol;Acc:MGI:2442001]	-1.7319
Sost	protein_coding	sclerostin [Source:MGI Symbol;Acc:MGI:1921749]	-1.73093
Rgs5	protein_coding	regulator of G-protein signaling 5 [Source:MGI Symbol;Acc:MGI:1098434]	-1.73057
2310043P16Rik	TEC	RIKEN cDNA 2310043P16 gene [Source:MGI Symbol;Acc:MGI:1917385]	-1.73036
Gm15481	processed_pseudogene	predicted gene 15481 [Source:MGI Symbol;Acc:MGI:3705452]	-1.72986
A930001A20Rik	lincRNA	RIKEN cDNA A930001A20 gene [Source:MGI Symbol;Acc:MGI:1924377]	-1.72585
Synpr	protein_coding	synaptoporin [Source:MGI Symbol;Acc:MGI:1919253]	-1.72495
Slc25a2	protein_coding	solute carrier family 25 (mitochondrial carrier, ornithine transporter) member 2 [Source:MGI Symbol;Acc:MGI:2137907]	-1.72361
Gm12345	processed_pseudogene	predicted gene 12345 [Source:MGI Symbol;Acc:MGI:3649862]	-1.72132
Gm37373	TEC	predicted gene, 37373 [Source:MGI Symbol;Acc:MGI:5610601]	-1.71826
Gm15484	processed_pseudogene	predicted gene 15484 [Source:MGI Symbol;Acc:MGI:3705726]	-1.71808
Gm44830	TEC	predicted gene 44830 [Source:MGI Symbol;Acc:MGI:5753406]	-1.71801
Gm42463	sense_intronic	predicted gene 42463 [Source:MGI Symbol;Acc:MGI:5662600]	-1.71716
Gm45309	sense_intronic	predicted gene 45309 [Source:MGI Symbol;Acc:MGI:5791145]	-1.71709
Hist1h2ab	protein_coding	histone cluster 1, H2ab [Source:MGI Symbol;Acc:MGI:2448306]	-1.71208
Gm37733	TEC	predicted gene, 37733 [Source:MGI Symbol;Acc:MGI:5610961]	-1.71186
Gm6525	protein_coding	predicted pseudogene 6525 [Source:MGI Symbol;Acc:MGI:3648757]	-1.71144
Gm5526	processed_pseudogene	predicted pseudogene 5526 [Source:MGI Symbol;Acc:MGI:3643066]	-1.70947
Tgtp2	protein_coding	T cell specific GTPase 2 [Source:MGI Symbol;Acc:MGI:3710083]	-1.70786
Smim6	protein_coding	small integral membrane protein 6 [Source:MGI Symbol;Acc:MGI:1915778]	-1.70748
Gm13886	processed_pseudogene	predicted gene 13886 [Source:MGI Symbol;Acc:MGI:3649359]	-1.70627
A230107N01Rik	antisense	RIKEN cDNA A230107N01 gene [Source:MGI Symbol;Acc:MGI:2445042]	-1.70616
Gm12856	processed_pseudogene	predicted gene 12856 [Source:MGI Symbol;Acc:MGI:3649664]	-1.70614
Ikzf3	protein_coding	IKAROS family zinc finger 3 [Source:MGI Symbol;Acc:MGI:1342542]	-1.70601
Aadacl2	protein_coding	arylacetamide deacetylase like 2 [Source:MGI Symbol;Acc:MGI:3646333]	-1.70523
Cyp2c55	protein_coding	cytochrome P450, family 2, subfamily c, polypeptide 55 [Source:MGI Symbol;Acc:MGI:1919332]	-1.70498
Gm44027	TEC	predicted gene, 44027 [Source:MGI Symbol;Acc:MGI:5690419]	-1.70323
Gm13245	processed_pseudogene	predicted gene 13245 [Source:MGI Symbol;Acc:MGI:3649656]	-1.70205
Rgs7	protein_coding	regulator of G protein signaling 7 [Source:MGI Symbol;Acc:MGI:1346089]	-1.69865
Gm37696	antisense	predicted gene, 37696 [Source:MGI Symbol;Acc:MGI:5610924]	-1.69853
Gm5086	lincRNA	predicted gene 5086 [Source:MGI Symbol;Acc:MGI:3644417]	-1.69788
Hand1	protein_coding	heart and neural crest derivatives expressed 1 [Source:MGI Symbol;Acc:MGI:103577]	-1.69772
Dapk2	protein_coding	death-associated protein kinase 2 [Source:MGI Symbol;Acc:MGI:1341297]	-1.69679
Gm7299	processed_pseudogene	predicted gene 7299 [Source:MGI Symbol;Acc:MGI:3648716]	-1.69677
Gm10812	pseudogene	predicted gene 10812 [Source:MGI Symbol;Acc:MGI:3708535]	-1.69548
Gm42693	TEC	predicted gene 42693 [Source:MGI Symbol;Acc:MGI:5662830]	-1.69534
Gm6204	processed_pseudogene	predicted gene 6204 [Source:MGI Symbol;Acc:MGI:3649038]	-1.69483
Gm44249	TEC	predicted gene, 44249 [Source:MGI Symbol;Acc:MGI:5690641]	-1.69464
Platr31	antisense	pluripotency associated transcript 31 [Source:MGI Symbol;Acc:MGI:5477339]	-1.69434
Gm45399	TEC	predicted gene 45399 [Source:MGI Symbol;Acc:MGI:5791235]	-1.69419
Luzp4	protein_coding	leucine zipper protein 4 [Source:MGI Symbol;Acc:MGI:3708816]	-1.6941
Naalad2	protein_coding	N-acetylated alpha-linked acidic dipeptidase 2 [Source:MGI Symbol;Acc:MGI:1919810]	-1.69406
Gm26755	lincRNA	predicted gene, 26755 [Source:MGI Symbol;Acc:MGI:5477249]	-1.6933
BC001981	antisense	cDNA sequence BC001981 [Source:MGI Symbol;Acc:MGI:3704411]	-1.69284
Gm38384	TEC	predicted gene, 38384 [Source:MGI Symbol;Acc:MGI:5611612]	-1.69259
Gm6055	processed_pseudogene	predicted gene 6055 [Source:MGI Symbol;Acc:MGI:3648620]	-1.69183
B3gat1	protein_coding	beta-1,3-glucuronyltransferase 1 (glucuronosyltransferase P) [Source:MGI Symbol;Acc:MGI:1924148]	-1.69167
Fev	protein_coding	FEV (ETS oncogene family) [Source:MGI Symbol;Acc:MGI:2449712]	-1.69091
Gm28379	antisense	predicted gene 28379 [Source:MGI Symbol;Acc:MGI:5579085]	-1.69074
Gm43460	TEC	predicted gene 43460 [Source:MGI Symbol;Acc:MGI:5663597]	-1.69008
Fam83a	protein_coding	family with sequence similarity 83, member A [Source:MGI Symbol;Acc:MGI:2447773]	-1.68949
Fam83a	protein_coding	family with sequence similarity 83, member A [Source:MGI Symbol;Acc:MGI:2447773]	-1.68949
Mageb4	protein_coding	melanoma antigen, family B, 4 [Source:MGI Symbol;Acc:MGI:2148568]	-1.68842
Rad51ap2	protein_coding	RAD51 associated protein 2 [Source:MGI Symbol;Acc:MGI:3644580]	-1.68729
Col3a1	protein_coding	collagen, type III, alpha 1 [Source:MGI Symbol;Acc:MGI:88453]	-1.68687
Slc28a3	protein_coding	solute carrier family 28 (sodium-coupled nucleoside transporter), member 3 [Source:MGI Symbol;Acc:MGI:2137361]	-1.68594
Lrrc32	protein_coding	leucine rich repeat containing 32 [Source:MGI Symbol;Acc:MGI:93882]	-1.68589
Gm43389	TEC	predicted gene 43389 [Source:MGI Symbol;Acc:MGI:5663526]	-1.68432
Gm15163	processed_pseudogene	predicted gene 15163 [Source:MGI Symbol;Acc:MGI:3642430]	-1.68388

Gm26448	snoRNA	predicted gene, 26448 [Source:MGI Symbol;Acc:MGI:5456225]	-1.68348
Cpb1	protein_coding	carboxypeptidase B1 (tissue) [Source:MGI Symbol;Acc:MGI:1923953]	-1.68295
Gm37850	antisense	predicted gene, 37850 [Source:MGI Symbol;Acc:MGI:5611078]	-1.68241
Gm4353	protein_coding	predicted gene 4353 [Source:MGI Symbol;Acc:MGI:3782538]	-1.68169
Gm2423	processed_pseudogene	predicted gene 2423 [Source:MGI Symbol;Acc:MGI:3805957]	-1.68159
Tnfaip2	protein_coding	tumor necrosis factor, alpha-induced protein 2 [Source:MGI Symbol;Acc:MGI:104960]	-1.67968
Iqsec3	protein_coding	IQ motif and Sec7 domain 3 [Source:MGI Symbol;Acc:MGI:2677208]	-1.67931
Myo15	protein_coding	myosin XV [Source:MGI Symbol;Acc:MGI:1261811]	-1.67908
Ppp1r32	protein_coding	protein phosphatase 1, regulatory subunit 32 [Source:MGI Symbol;Acc:MGI:1915002]	-1.6783
Gm45487	TEC	predicted gene 45487 [Source:MGI Symbol;Acc:MGI:5791323]	-1.67693
BC030343	lincRNA	cDNA sequence BC030343 [Source:MGI Symbol;Acc:MGI:2679267]	-1.67626
Gm28988	antisense	predicted gene 28988 [Source:MGI Symbol;Acc:MGI:5579694]	-1.67518
Cabyr	protein_coding	calcium-binding tyrosine-(Y)-phosphorylation regulated (fibrousheathin 2) [Source:MGI Symbol;Acc:MGI:1918382]	-1.67508
Gm37069	TEC	predicted gene, 37069 [Source:MGI Symbol;Acc:MGI:5610297]	-1.67439
Gm15366	processed_pseudogene	predicted gene 15366 [Source:MGI Symbol;Acc:MGI:3709611]	-1.67345
Cd28	protein_coding	CD28 antigen [Source:NCBI gene;Acc:12487]	-1.67338
Cd28	protein_coding	CD28 antigen [Source:MGI Symbol;Acc:MGI:88327]	-1.67338
Mapk10	protein_coding	mitogen-activated protein kinase 10 [Source:MGI Symbol;Acc:MGI:1346863]	-1.67274
Cnnm4	protein_coding	cyclin M4 [Source:MGI Symbol;Acc:MGI:2151060]	-1.67186
Abcd2	protein_coding	ATP-binding cassette, sub-family D (ALD), member 2 [Source:MGI Symbol;Acc:MGI:1349467]	-1.6708
Gm10010	lincRNA	predicted gene 10010 [Source:MGI Symbol;Acc:MGI:3641978]	-1.67037
Lhx9	protein_coding	LIM homeobox protein 9 [Source:MGI Symbol;Acc:MGI:1316721]	-1.66969
Rab40b	protein_coding	Rab40B, member RAS oncogene family [Source:MGI Symbol;Acc:MGI:2183451]	-1.6677
Zc3h12d	protein_coding	zinc finger CCCH type containing 12D [Source:MGI Symbol;Acc:MGI:3045313]	-1.66719
Krt8-ps	processed_pseudogene	keratin 8, pseudogene [Source:MGI Symbol;Acc:MGI:3779503]	-1.66658
Rpl31-ps13	processed_pseudogene	ribosomal protein L31, pseudogene 13 [Source:MGI Symbol;Acc:MGI:3809036]	-1.66636
Gm44567	TEC	predicted gene 44567 [Source:MGI Symbol;Acc:MGI:5753143]	-1.66627
Gm8994	protein_coding	predicted gene 8994 [Source:MGI Symbol;Acc:MGI:3644226]	-1.66588
Lrp8os3	bidirectional_promoter_lncRNA	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor, opposite strand 3 [Source:MGI Symbol;Acc:MGI:3651133]	-1.66344
Gm42841	TEC	predicted gene 42841 [Source:MGI Symbol;Acc:MGI:5662978]	-1.66306
Ppp1r2-ps4	processed_pseudogene	protein phosphatase 1, regulatory (inhibitor) subunit 2, pseudogene 4 [Source:MGI Symbol;Acc:MGI:3645199]	-1.66232
Spata18	protein_coding	spermatogenesis associated 18 [Source:MGI Symbol;Acc:MGI:1920722]	-1.66205
2900060N12Rik	bidirectional_promoter_lncRNA	RIKEN cDNA 2900060N12 gene [Source:MGI Symbol;Acc:MGI:1920281]	-1.66162
Gm15546	processed_pseudogene	predicted gene 15546 [Source:MGI Symbol;Acc:MGI:3782995]	-1.66036
Gm42879	TEC	predicted gene 42879 [Source:MGI Symbol;Acc:MGI:5663016]	-1.66014
A330069E16Rik	lincRNA	RIKEN cDNA A330069E16 gene [Source:MGI Symbol;Acc:MGI:3583899]	-1.6593
Sqor	protein_coding	sulfide quinone oxidoreductase [Source:MGI Symbol;Acc:MGI:1929899]	-1.65929
Gm42875	antisense	predicted gene 42875 [Source:MGI Symbol;Acc:MGI:5663012]	-1.65908
Gm7890	processed_pseudogene	predicted gene 7890 [Source:MGI Symbol;Acc:MGI:3645227]	-1.65905
Slc24a5	protein_coding	solute carrier family 24, member 5 [Source:MGI Symbol;Acc:MGI:2677271]	-1.65841
Pcdh8	protein_coding	protocadherin 8 [Source:MGI Symbol;Acc:MGI:1306800]	-1.65716
Rgs4	protein_coding	regulator of G-protein signaling 4 [Source:MGI Symbol;Acc:MGI:108409]	-1.65598
Chrng	protein_coding	cholinergic receptor, nicotinic, gamma polypeptide [Source:MGI Symbol;Acc:MGI:87895]	-1.65551
Gm5694	processed_pseudogene	predicted gene 5694 [Source:MGI Symbol;Acc:MGI:3645764]	-1.65308
Tmem117	protein_coding	transmembrane protein 117 [Source:MGI Symbol;Acc:MGI:2444580]	-1.65107
Gm11743	processed_pseudogene	predicted gene 11743 [Source:MGI Symbol;Acc:MGI:3651178]	-1.65028
Grem2	protein_coding	gremlin 2, DAN family BMP antagonist [Source:MGI Symbol;Acc:MGI:1344367]	-1.64958
Tmem200c	protein_coding	transmembrane protein 200C [Source:MGI Symbol;Acc:MGI:3646281]	-1.64953
Gm43307	TEC	predicted gene 43307 [Source:MGI Symbol;Acc:MGI:5663444]	-1.64859
H2-T24	protein_coding	histocompatibility 2, T region locus 24 [Source:MGI Symbol;Acc:MGI:95958]	-1.64433
Aadac	protein_coding	arylacetamide deacetylase [Source:MGI Symbol;Acc:MGI:1915008]	-1.64257
Gm28941	sense_intronic	predicted gene 28941 [Source:MGI Symbol;Acc:MGI:5579647]	-1.64201
Gm45425	lincRNA	predicted gene 45425 [Source:MGI Symbol;Acc:MGI:5791261]	-1.64066
9530059O14Rik	lincRNA	RIKEN cDNA 9530059O14 gene [Source:MGI Symbol;Acc:MGI:2442421]	-1.64054
Gm11725	antisense	predicted gene 11725 [Source:MGI Symbol;Acc:MGI:3649385]	-1.64
5033404E19Rik	transcribed_processed_pseudogene	RIKEN cDNA 5033404E19 gene [Source:MGI Symbol;Acc:MGI:2149700]	-1.63863
Pde1a	protein_coding	phosphodiesterase 1A, calmodulin-dependent [Source:MGI Symbol;Acc:MGI:1201792]	-1.63745
Lca5l	protein_coding	Leber congenital amaurosis 5-like [Source:MGI Symbol;Acc:MGI:3041157]	-1.63548
Epn3	protein_coding	epsin 3 [Source:MGI Symbol;Acc:MGI:1919139]	-1.63532
Gm7145	protein_coding	predicted gene 7145 [Source:MGI Symbol;Acc:MGI:3648947]	-1.63503
Ptpre	protein_coding	protein tyrosine phosphatase, receptor type, E [Source:MGI Symbol;Acc:MGI:97813]	-1.63338
Wdr66	protein_coding	WD repeat domain 66 [Source:MGI Symbol;Acc:MGI:1918495]	-1.63318

Gm17250	lincRNA	predicted gene, 17250 [Source:MGI Symbol;Acc:MGI:4936884]	-1.63303
AL732506.1	lincRNA	RNA component of mitochondrial RNAase P	-1.63297
Col12a1	protein_coding	collagen, type XII, alpha 1 [Source:MGI Symbol;Acc:MGI:88448]	-1.63238
Misp	protein_coding	mitotic spindle positioning [Source:MGI Symbol;Acc:MGI:1926156]	-1.63222
Fat4	protein_coding	FAT atypical cadherin 4 [Source:MGI Symbol;Acc:MGI:3045256]	-1.62977
Gm14861	processed_pseudogene	predicted gene 14861 [Source:MGI Symbol;Acc:MGI:3802070]	-1.62775
Itpr2	protein_coding	inositol 1,4,5-triphosphate receptor 2 [Source:MGI Symbol;Acc:MGI:99418]	-1.62394
4930578M07Rik	processed_transcript	RIKEN cDNA 4930578M07 gene [Source:MGI Symbol;Acc:MGI:1923171]	-1.62363
Gm6395	processed_pseudogene	predicted gene 6395 [Source:MGI Symbol;Acc:MGI:3645295]	-1.62191
A230087F16Rik	antisense	RIKEN cDNA A230087F16 gene [Source:MGI Symbol;Acc:MGI:2444635]	-1.61834
Gm12350	processed_pseudogene	predicted gene 12350 [Source:MGI Symbol;Acc:MGI:3649478]	-1.61659
Gm44951	TEC	predicted gene 44951 [Source:MGI Symbol;Acc:MGI:5753527]	-1.61583
Bfsp1	protein_coding	beaded filament structural protein 1, in lens-CP94 [Source:MGI Symbol;Acc:MGI:101770]	-1.61439
Ankrd65	protein_coding	ankyrin repeat domain 65 [Source:MGI Symbol;Acc:MGI:2685285]	-1.61287
Gm5842	processed_pseudogene	predicted gene 5842 [Source:MGI Symbol;Acc:MGI:3645251]	-1.61282
Gm37777	TEC	predicted gene, 37777 [Source:MGI Symbol;Acc:MGI:5611005]	-1.61192
B930036N10Rik	antisense	RIKEN cDNA B930036N10 gene [Source:MGI Symbol;Acc:MGI:3702496]	-1.61088
Rps2-ps13	processed_pseudogene	ribosomal protein S2, pseudogene 13 [Source:MGI Symbol;Acc:MGI:3705640]	-1.60963
Gabrb2	protein_coding	gamma-aminobutyric acid (GABA) A receptor, subunit beta 2 [Source:MGI Symbol;Acc:MGI:95620]	-1.60543
Gm20949	processed_pseudogene	predicted gene, 20949 [Source:MGI Symbol;Acc:MGI:5434304]	-1.60369
Gckr	protein_coding	glucokinase regulatory protein [Source:MGI Symbol;Acc:MGI:1096345]	-1.60307
Trim9	protein_coding	tripartite motif-containing 9 [Source:MGI Symbol;Acc:MGI:2137354]	-1.60208
Gltpd2	protein_coding	glycolipid transfer protein domain containing 2 [Source:MGI Symbol;Acc:MGI:2444527]	-1.60182
Clec12a	protein_coding	C-type lectin domain family 12, member a [Source:MGI Symbol;Acc:MGI:3040968]	-1.60161
Gm12657	processed_pseudogene	predicted gene 12657 [Source:MGI Symbol;Acc:MGI:3651714]	-1.59735
Gm12397	processed_pseudogene	predicted gene 12397 [Source:MGI Symbol;Acc:MGI:3649202]	-1.59315
Gpr150	protein_coding	G protein-coupled receptor 150 [Source:MGI Symbol;Acc:MGI:2441872]	-1.58971
Gm27038	processed_pseudogene	predicted gene, 27038 [Source:MGI Symbol;Acc:MGI:5504153]	-1.58516
Kif26b	protein_coding	kinesin family member 26B [Source:MGI Symbol;Acc:MGI:2447076]	-1.57933
Pzp	protein_coding	PZP, alpha-2-macroglobulin like [Source:MGI Symbol;Acc:MGI:87854]	-1.57807
Serpina7	protein_coding	serine (or cysteine) peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 7 [Source:MGI Symbol;Acc:MGI:3041197]	-1.57784
Gm38380	TEC	predicted gene, 38380 [Source:MGI Symbol;Acc:MGI:5611608]	-1.57522
Sp6	protein_coding	trans-acting transcription factor 6 [Source:MGI Symbol;Acc:MGI:1932575]	-1.57151
Slc26a1	protein_coding	solute carrier family 26 (sulfate transporter), member 1 [Source:MGI Symbol;Acc:MGI:2385894]	-1.56868
Gm12816	processed_pseudogene	predicted gene 12816 [Source:MGI Symbol;Acc:MGI:3649628]	-1.56735
Gm7129	processed_pseudogene	predicted gene 7129 [Source:MGI Symbol;Acc:MGI:3648820]	-1.56697
Tg	protein_coding	thyroglobulin [Source:MGI Symbol;Acc:MGI:98733]	-1.56522
Gm44553	TEC	predicted gene 44553 [Source:MGI Symbol;Acc:MGI:5753129]	-1.56366
Gm28809	antisense	predicted gene 28809 [Source:MGI Symbol;Acc:MGI:5579515]	-1.5632
4933431C10Rik	TEC	RIKEN cDNA 4933431C10 gene [Source:MGI Symbol;Acc:MGI:1918572]	-1.56221
Gm37764	TEC	predicted gene, 37764 [Source:MGI Symbol;Acc:MGI:5610992]	-1.56207
Plk-ps1	processed_pseudogene	polo like kinase, pseudogene 1 [Source:MGI Symbol;Acc:MGI:103247]	-1.55928
Gm40709	bidirectional_promoter_lncRNA	predicted gene, 40709 [Source:MGI Symbol;Acc:MGI:5623594]	-1.55896
Gm43498	processed_pseudogene	predicted gene 43498 [Source:MGI Symbol;Acc:MGI:5663635]	-1.55819
Gm37885	antisense	predicted gene, 37885 [Source:MGI Symbol;Acc:MGI:5611113]	-1.55666
Gm10657	TEC	predicted gene 10657 [Source:MGI Symbol;Acc:MGI:3642259]	-1.55613
Tac1	protein_coding	tachykinin 1 [Source:MGI Symbol;Acc:MGI:98474]	-1.55606
Gm38243	TEC	predicted gene, 38243 [Source:MGI Symbol;Acc:MGI:5611471]	-1.55523
Gm37423	processed_transcript	predicted gene, 37423 [Source:MGI Symbol;Acc:MGI:5610651]	-1.55464
Gm14048	processed_pseudogene	predicted gene 14048 [Source:MGI Symbol;Acc:MGI:3650660]	-1.5542
Gm37968	TEC	predicted gene, 37968 [Source:MGI Symbol;Acc:MGI:5611196]	-1.55385
1700126G02Rik	antisense	RIKEN cDNA 1700126G02 gene [Source:MGI Symbol;Acc:MGI:1923898]	-1.5535
Gm44792	TEC	predicted gene 44792 [Source:MGI Symbol;Acc:MGI:5753368]	-1.55348
Gm37174	TEC	predicted gene, 37174 [Source:MGI Symbol;Acc:MGI:5610402]	-1.54944
Snhg7os	antisense	small nucleolar RNA host gene 7, opposite strand [Source:MGI Symbol;Acc:MGI:3045374]	-1.54883
Gm43203	TEC	predicted gene 43203 [Source:MGI Symbol;Acc:MGI:5663340]	-1.54833
Gm13502	processed_pseudogene	predicted gene 13502 [Source:MGI Symbol;Acc:MGI:3649894]	-1.54754
Gm38372	TEC	predicted gene, 38372 [Source:MGI Symbol;Acc:MGI:5611600]	-1.54594
Terc	misc_RNA	telomerase RNA component [Source:MGI Symbol;Acc:MGI:109558]	-1.54507
Gm15587	antisense	predicted gene 15587 [Source:MGI Symbol;Acc:MGI:3783035]	-1.54455
Shox2	protein_coding	short stature homeobox 2 [Source:MGI Symbol;Acc:MGI:1201673]	-1.54361
Tmem56	protein_coding	transmembrane protein 56 [Source:MGI Symbol;Acc:MGI:1923195]	-1.54348
Gm26568	antisense	predicted gene, 26568 [Source:MGI Symbol;Acc:MGI:5477062]	-1.53905
Gm16580	processed_pseudogene	predicted gene 16580 [Source:MGI Symbol;Acc:MGI:4415000]	-1.53895
Itga2	protein_coding	integrin alpha 2 [Source:MGI Symbol;Acc:MGI:96600]	-1.53566
Gm37464	TEC	predicted gene, 37464 [Source:MGI Symbol;Acc:MGI:5610692]	-1.53554
BB014433	protein_coding	expressed sequence BB014433 [Source:MGI Symbol;Acc:MGI:2142823]	-1.53534
Gm37900	TEC	predicted gene, 37900 [Source:MGI Symbol;Acc:MGI:5611128]	-1.5342

Gm6206	processed_pseudogene	predicted pseudogene 6206 [Source:MGI Symbol;Acc:MGI:3644890]	-1.53314
Lrp1b	protein_coding	low density lipoprotein-related protein 1B (deleted in tumors) [Source:MGI Symbol;Acc:MGI:2151136]	-1.52993
Plin4	protein_coding	perilipin 4 [Source:MGI Symbol;Acc:MGI:1929709]	-1.52916
Nipal1	protein_coding	NIPA-like domain containing 1 [Source:MGI Symbol;Acc:MGI:1917951]	-1.52667
BC106179	processed_transcript	cDNA sequence BC106179 [Source:MGI Symbol;Acc:MGI:3702726]	-1.52251
Acer1	protein_coding	alkaline ceramidase 1 [Source:MGI Symbol;Acc:MGI:2181962]	-1.52245
Gm17137	transcribed_processed_pseudogene	predicted gene 17137 [Source:MGI Symbol;Acc:MGI:4937964]	-1.52234
Gm28438	unprocessed_pseudogene	predicted gene 28438 [Source:MGI Symbol;Acc:MGI:5579144]	-1.52085
Slc10a1	protein_coding	solute carrier family 10 (sodium/bile acid cotransporter family), member 1 [Source:MGI Symbol;Acc:MGI:97379]	-1.51985
Gm38375	TEC	predicted gene, 38375 [Source:MGI Symbol;Acc:MGI:5611603]	-1.51983
4930512J16Rik	lincRNA	RIKEN cDNA 4930512J16 gene [Source:MGI Symbol;Acc:MGI:1922381]	-1.51838
Gm10382	protein_coding	predicted gene 10382 [Source:MGI Symbol;Acc:MGI:3647829]	-1.51639
Gm42585	TEC	predicted gene 42585 [Source:MGI Symbol;Acc:MGI:5662722]	-1.51563
Adcy10	protein_coding	adenylate cyclase 10 [Source:MGI Symbol;Acc:MGI:2660854]	-1.51484
Gm21964	protein_coding	predicted gene, 21964 [Source:MGI Symbol;Acc:MGI:5439433]	-1.51344
Gm26945	lincRNA	predicted gene, 26945 [Source:MGI Symbol;Acc:MGI:5504060]	-1.51336
Gm26945	lincRNA	predicted gene, 26945 [Source:MGI Symbol;Acc:MGI:5504060]	-1.51336
Hsf3	protein_coding	heat shock transcription factor 3 [Source:MGI Symbol;Acc:MGI:3045337]	-1.51277
Gm28363	protein_coding	predicted gene 28363 [Source:MGI Symbol;Acc:MGI:5579069]	-1.5123
4930433N12Rik	processed_transcript	RIKEN cDNA 4930433N12 gene [Source:MGI Symbol;Acc:MGI:2149746]	-1.50887
Snord15a	snoRNA	small nucleolar RNA, C/D box 15A [Source:MGI Symbol;Acc:MGI:3645887]	-1.50778
Th	protein_coding	tyrosine hydroxylase [Source:MGI Symbol;Acc:MGI:98735]	-1.50775
Gm43009	TEC	predicted gene 43009 [Source:MGI Symbol;Acc:MGI:5663146]	-1.50685
Syndig1l	protein_coding	synapse differentiation inducing 1 like [Source:MGI Symbol;Acc:MGI:2685107]	-1.50388
Gm37979	TEC	predicted gene, 37979 [Source:MGI Symbol;Acc:MGI:5611207]	-1.50325
1700020L24Rik	protein_coding	RIKEN cDNA 1700020L24 gene [Source:MGI Symbol;Acc:MGI:1913580]	-1.50027

Up Regulated:

Gene name	Gene type	Gene description	Foldchange
C920009B18Rik	transcribed_unprocessed_pseudogene	RIKEN cDNA C920009B18 gene [Source:MGI Symbol;Acc:MGI:3583961]	9.52011
Gm27403	misc_RNA	predicted gene, 27403 [Source:MGI Symbol;Acc:MGI:5530785]	7.3341
Six3	protein_coding	sine oculis-related homeobox 3 [Source:MGI Symbol;Acc:MGI:102764]	4.7411
Ush2a	protein_coding	usherin [Source:MGI Symbol;Acc:MGI:1341292]	4.59849
Sycp2l	protein_coding	synaptonemal complex protein 2-like [Source:MGI Symbol;Acc:MGI:2685114]	4.31866
Slc24a2	protein_coding	solute carrier family 24 (sodium/potassium/calcium exchanger), member 2 [Source:MGI Symbol;Acc:MGI:1923626]	3.9423
Smpd5	protein_coding	sphingomyelin phosphodiesterase 5 [Source:MGI Symbol;Acc:MGI:3709877]	3.8133
Pcdh11x	protein_coding	protocadherin 11 X-linked [Source:MGI Symbol;Acc:MGI:2442849]	3.52243
Ankef1	protein_coding	ankyrin repeat and EF-hand domain containing 1 [Source:MGI Symbol;Acc:MGI:2441685]	3.45407
Gm9962	antisense	predicted gene 9962 [Source:MGI Symbol;Acc:MGI:3642695]	3.37566
Krt1	protein_coding	keratin 1 [Source:MGI Symbol;Acc:MGI:96698]	3.33992
A830036E02Rik	antisense	RIKEN cDNA A830036E02 gene [Source:MGI Symbol;Acc:MGI:3686876]	3.31073
Saa3	protein_coding	serum amyloid A 3 [Source:MGI Symbol;Acc:MGI:98223]	3.27767
Dlg2	protein_coding	discs large MAGUK scaffold protein 2 [Source:MGI Symbol;Acc:MGI:1344351]	3.23586
Nav3	protein_coding	neuron navigator 3 [Source:MGI Symbol;Acc:MGI:2183703]	3.12058
Cldn2	protein_coding	claudin 2 [Source:MGI Symbol;Acc:MGI:1276110]	3.08405
Mc5r	protein_coding	melanocortin 5 receptor [Source:MGI Symbol;Acc:MGI:99420]	3.07678
Cd247	protein_coding	CD247 antigen [Source:MGI Symbol;Acc:MGI:88334]	3.07538
AW822252	transcribed_unprocessed_pseudogene	expressed sequence AW822252 [Source:MGI Symbol;Acc:MGI:2148030]	3.06159
Slc24a3	protein_coding	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3 [Source:MGI Symbol;Acc:MGI:2137513]	3.05864
Scg5	protein_coding	secretogranin V [Source:MGI Symbol;Acc:MGI:98289]	3.04641
Klhl33	protein_coding	kelch-like 33 [Source:MGI Symbol;Acc:MGI:3644593]	3.04409
Nkain2	protein_coding	Na+/K+ transporting ATPase interacting 2 [Source:MGI Symbol;Acc:MGI:1923447]	3.04233
Gm9597	transcribed_unprocessed_pseudogene	predicted gene 9597 [Source:MGI Symbol;Acc:MGI:3780005]	3.03877
Otop2	protein_coding	otopetrin 2 [Source:MGI Symbol;Acc:MGI:2388365]	3.01943
Vwa3a	protein_coding	von Willebrand factor A domain containing 3A [Source:MGI Symbol;Acc:MGI:3041229]	2.99598
Lrrc19	protein_coding	leucine rich repeat containing 19 [Source:MGI Symbol;Acc:MGI:2140219]	2.99545
D930036K23Rik	TEC	RIKEN cDNA D930036K23 gene [Source:MGI Symbol;Acc:MGI:2442341]	2.9842
D630044L22Rik	protein_coding	RIKEN cDNA gene D630044L22 gene [Source:MGI Symbol;Acc:MGI:3709661]	2.93555
Slc26a5	protein_coding	solute carrier family 26, member 5 [Source:MGI Symbol;Acc:MGI:1933154]	2.92693
9630013K17Rik	antisense	RIKEN cDNA 9630013K17 gene [Source:MGI Symbol;Acc:MGI:1926133]	2.90466
Gm7446	processed_pseudogene	predicted gene 7446 [Source:MGI Symbol;Acc:MGI:3644947]	2.86935
Slc4a10	protein_coding	solute carrier family 4, sodium bicarbonate cotransporter-like, member 10 [Source:MGI Symbol;Acc:MGI:2150150]	2.85806
AV039307	processed_transcript	expressed sequence AV039307 [Source:MGI Symbol;Acc:MGI:2139243]	2.85636
Apoa1	protein_coding	apolipoprotein A-I [Source:MGI Symbol;Acc:MGI:88049]	2.80639
Pkhd1	protein_coding	polycystic kidney and hepatic disease 1 [Source:MGI Symbol;Acc:MGI:2155808]	2.76875
Gm281	protein_coding	predicted gene 281 [Source:MGI Symbol;Acc:MGI:2685127]	2.76271
Haglr	bidirectional_promoter_lncRNA	Hoxd antisense growth associated long non-coding RNA [Source:MGI Symbol;Acc:MGI:3026978]	2.7433
Gm8326	processed_pseudogene	predicted gene 8326 [Source:MGI Symbol;Acc:MGI:3644839]	2.74285
Dnah12	protein_coding	dynein, axonemal, heavy chain 12 [Source:MGI Symbol;Acc:MGI:107720]	2.72786
Dnah12	protein_coding	dynein, axonemal, heavy chain 12 [Source:MGI Symbol;Acc:MGI:107720]	2.72786
2310001H17Rik	lincRNA	RIKEN cDNA 2310001H17 gene [Source:MGI Symbol;Acc:MGI:1923682]	2.68082
Erich2	protein_coding	glutamate rich 2 [Source:MGI Symbol;Acc:MGI:1913998]	2.60759
Gm16470	processed_pseudogene	predicted pseudogene 16470 [Source:MGI Symbol;Acc:MGI:3645079]	2.60607
4930579K19Rik	processed_transcript	RIKEN cDNA 4930579K19 gene [Source:MGI Symbol;Acc:MGI:1923131]	2.59846
Cd19	protein_coding	CD19 antigen [Source:MGI Symbol;Acc:MGI:88319]	2.58118
Gm15270	processed_transcript	predicted gene 15270 [Source:MGI Symbol;Acc:MGI:3705149]	2.5783
Gm5075	processed_pseudogene	predicted gene 5075 [Source:MGI Symbol;Acc:MGI:3645483]	2.55217
Gm43815	TEC	predicted gene 43815 [Source:MGI Symbol;Acc:MGI:5663952]	2.54861
Cwh43	protein_coding	cell wall biogenesis 43 C-terminal homolog [Source:MGI Symbol;Acc:MGI:2444131]	2.54697
Rb1	protein_coding	RB transcriptional corepressor 1 [Source:MGI Symbol;Acc:MGI:97874]	2.53923
Arhgef33	protein_coding	Rho guanine nucleotide exchange factor (GEF) 33 [Source:MGI Symbol;Acc:MGI:2685787]	2.52912
Adamts17	protein_coding	a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 17 [Source:MGI Symbol;Acc:MGI:3588195]	2.52046
2700068H02Rik	antisense	RIKEN cDNA 2700068H02 gene [Source:MGI Symbol;Acc:MGI:1919813]	2.51687
Rps4x-ps	processed_pseudogene	Rps4x retrotransposed pseudogene [Source:MGI Symbol;Acc:MGI:3644569]	2.50993
Plin1	protein_coding	perilipin 1 [Source:MGI Symbol;Acc:MGI:1890505]	2.48386
Gm12579	processed_pseudogene	predicted gene 12579 [Source:MGI Symbol;Acc:MGI:3651711]	2.4821
Gm13436	processed_pseudogene	predicted gene 13436 [Source:MGI Symbol;Acc:MGI:3651005]	2.48086

Duox2	protein_coding	dual oxidase 2 [Source:MGI Symbol;Acc:MGI:3036280]	2.47652
Gm17090	antisense	predicted gene 17090 [Source:MGI Symbol;Acc:MGI:4937917]	2.47572
Aox3	protein_coding	aldehyde oxidase 3 [Source:MGI Symbol;Acc:MGI:1918974]	2.4699
Hal	protein_coding	histidine ammonia lyase [Source:MGI Symbol;Acc:MGI:96010]	2.46857
Rpl31-ps9	processed_pseudogene	ribosomal protein L31, pseudogene 9 [Source:MGI Symbol;Acc:MGI:3704195]	2.46762
Dll4	protein_coding	delta like canonical Notch ligand 4 [Source:MGI Symbol;Acc:MGI:1859388]	2.46391
Mid2	protein_coding	midline 2 [Source:MGI Symbol;Acc:MGI:1344333]	2.45801
Pou3f2	protein_coding	POU domain, class 3, transcription factor 2 [Source:MGI Symbol;Acc:MGI:101895]	2.45661
Hoxb8	protein_coding	homeobox B8 [Source:MGI Symbol;Acc:MGI:96189]	2.43747
Gm13160	processed_pseudogene	predicted gene 13160 [Source:MGI Symbol;Acc:MGI:3649984]	2.42307
Zpbp	protein_coding	zona pellucida binding protein [Source:MGI Symbol;Acc:MGI:1855701]	2.4109
Gm10169	processed_pseudogene	predicted pseudogene 10169 [Source:MGI Symbol;Acc:MGI:3642393]	2.40088
Sardh	protein_coding	sarcosine dehydrogenase [Source:MGI Symbol;Acc:MGI:2183102]	2.39889
Sntg2	protein_coding	syntrophin, gamma 2 [Source:MGI Symbol;Acc:MGI:1919541]	2.38685
Tekt5	protein_coding	tektin 5 [Source:MGI Symbol;Acc:MGI:1917676]	2.38598
Gm14698	protein_coding	predicted gene 14698 [Source:MGI Symbol;Acc:MGI:3709613]	2.353
Hs6st3	protein_coding	heparan sulfate 6-O-sulfotransferase 3 [Source:MGI Symbol;Acc:MGI:1354960]	2.34349
Gpr12	protein_coding	G-protein coupled receptor 12 [Source:MGI Symbol;Acc:MGI:101909]	2.3385
Tlr5	protein_coding	toll-like receptor 5 [Source:MGI Symbol;Acc:MGI:1858171]	2.33031
Rab11a	protein_coding	RAB11A, member RAS oncogene family [Source:MGI Symbol;Acc:MGI:1858202]	2.32757
Esrrg	protein_coding	estrogen-related receptor gamma [Source:MGI Symbol;Acc:MGI:1347056]	2.27188
Corin	protein_coding	corin [Source:MGI Symbol;Acc:MGI:1349451]	2.27059
Gm44021	TEC	predicted gene, 44021 [Source:MGI Symbol;Acc:MGI:5690413]	2.26343
Gm29237	antisense	predicted gene 29237 [Source:MGI Symbol;Acc:MGI:5579943]	2.25938
Alox8	protein_coding	arachidonate 8-lipoxygenase [Source:MGI Symbol;Acc:MGI:1098228]	2.2551
Gm42521	TEC	predicted gene 42521 [Source:MGI Symbol;Acc:MGI:5662658]	2.24636
Gm12454	antisense	predicted gene 12454 [Source:MGI Symbol;Acc:MGI:3649695]	2.24236
Gm8818	processed_pseudogene	predicted pseudogene 8818 [Source:MGI Symbol;Acc:MGI:3645903]	2.2373
Gm11772	antisense	predicted gene 11772 [Source:MGI Symbol;Acc:MGI:3649468]	2.23391
Gm7889	unprocessed_pseudogene	predicted gene 7889 [Source:MGI Symbol;Acc:MGI:3648247]	2.23136
Sostdc1	protein_coding	sclerostin domain containing 1 [Source:MGI Symbol;Acc:MGI:1913292]	2.23033
Ccdc183	protein_coding	coiled-coil domain containing 183 [Source:MGI Symbol;Acc:MGI:1924308]	2.22962
Gm13653	processed_pseudogene	predicted gene 13653 [Source:MGI Symbol;Acc:MGI:3649358]	2.22948
Slc23a4	protein_coding	solute carrier family 23 member 4 [Source:MGI Symbol;Acc:MGI:1917272]	2.22632
Cmtm5	protein_coding	CKLF-like MARVEL transmembrane domain containing 5 [Source:MGI Symbol;Acc:MGI:2447164]	2.21405
Lst1	protein_coding	leukocyte specific transcript 1 [Source:MGI Symbol;Acc:MGI:1096324]	2.20903
Sidt1	protein_coding	SID1 transmembrane family, member 1 [Source:MGI Symbol;Acc:MGI:2443155]	2.20859
Gm10337	protein_coding	predicted gene 10337 [Source:MGI Symbol;Acc:MGI:3704459]	2.20717
Chrne	protein_coding	cholinergic receptor, nicotinic, epsilon polypeptide [Source:MGI Symbol;Acc:MGI:87894]	2.20323
Gm17745	lincRNA	predicted gene, 17745 [Source:MGI Symbol;Acc:MGI:5009823]	2.20157
Stxbp5l	protein_coding	syntaxin binding protein 5-like [Source:MGI Symbol;Acc:MGI:2443815]	2.19438
Gm15708	sense_intronic	predicted gene 15708 [Source:MGI Symbol;Acc:MGI:3783149]	2.19181
Gm15894	antisense	predicted gene 15894 [Source:MGI Symbol;Acc:MGI:3801859]	2.18692
Gm5251	processed_pseudogene	predicted gene 5251 [Source:MGI Symbol;Acc:MGI:3647666]	2.18141
Gm3272	processed_pseudogene	predicted pseudogene 3272 [Source:MGI Symbol;Acc:MGI:3781450]	2.17718
Irx6	protein_coding	Iroquois homeobox 6 [Source:MGI Symbol;Acc:MGI:1927642]	2.17312
Gm13056	lincRNA	predicted gene 13056 [Source:MGI Symbol;Acc:MGI:3650725]	2.17307
Gm4045	processed_pseudogene	predicted gene 4045 [Source:MGI Symbol;Acc:MGI:3782220]	2.16759
1700001G11Rik	processed_transcript	RIKEN cDNA 1700001G11 gene [Source:MGI Symbol;Acc:MGI:1916553]	2.16683
Spaca1	protein_coding	sperm acrosome associated 1 [Source:MGI Symbol;Acc:MGI:1914902]	2.16592
Slc10a4	protein_coding	solute carrier family 10 (sodium/bile acid cotransporter family), member 4 [Source:MGI Symbol;Acc:MGI:3606480]	2.16304
Gm6611	processed_pseudogene	predicted gene 6611 [Source:MGI Symbol;Acc:MGI:3645568]	2.15809
Stab1	protein_coding	stabilin 1 [Source:MGI Symbol;Acc:MGI:2178742]	2.1546
Pdx1	protein_coding	pancreatic and duodenal homeobox 1 [Source:MGI Symbol;Acc:MGI:102851]	2.14924
Milr1	protein_coding	mast cell immunoglobulin like receptor 1 [Source:MGI Symbol;Acc:MGI:2685731]	2.14367
Slc22a20	protein_coding	solute carrier family 22 (organic anion transporter), member 20 [Source:MGI Symbol;Acc:MGI:2685809]	2.14083
Astl	protein_coding	astacin-like metalloendopeptidase (M12 family) [Source:MGI Symbol;Acc:MGI:3046414]	2.14023
Mybpc3	protein_coding	myosin binding protein C, cardiac [Source:MGI Symbol;Acc:MGI:102844]	2.13581
Mesp1	protein_coding	mesoderm posterior 1 [Source:MGI Symbol;Acc:MGI:107785]	2.12023
Gm5160	protein_coding	predicted gene 5160 [Source:MGI Symbol;Acc:MGI:3648528]	2.09783
Gm14052	processed_pseudogene	predicted gene 14052 [Source:MGI Symbol;Acc:MGI:3650009]	2.09498
1700001G17Rik	TEC	RIKEN cDNA 1700001G17 gene [Source:MGI Symbol;Acc:MGI:1914753]	2.09056
Gm11649	transcribed_processed_pseudogene	predicted gene 11649 [Source:MGI Symbol;Acc:MGI:3757863]	2.08529
Gm28453	transcribed_processed_pseudogene	predicted gene 28453 [Source:MGI Symbol;Acc:MGI:5579159]	2.07855

Ceacam20	protein_coding	carcinoembryonic antigen-related cell adhesion molecule 20 [Source:MGI Symbol;Acc:MGI:1918851]	2.07469
Prodh2	protein_coding	proline dehydrogenase (oxidase) 2 [Source:MGI Symbol;Acc:MGI:1929093]	2.07388
Gm17216	lincRNA	predicted gene 17216 [Source:MGI Symbol;Acc:MGI:4938043]	2.06392
Malrd1	protein_coding	MAM and LDL receptor class A domain containing 1 [Source:MGI Symbol;Acc:MGI:1928271]	2.05934
Usp17lc	protein_coding	ubiquitin specific peptidase 17-like C [Source:MGI Symbol;Acc:MGI:107698]	2.05671
Gm12091	processed_pseudogene	predicted gene 12091 [Source:MGI Symbol;Acc:MGI:3650089]	2.05431
Gm12166	protein_coding	predicted gene 12166 [Source:MGI Symbol;Acc:MGI:3650635]	2.0456
Pnlip	protein_coding	pancreatic lipase [Source:MGI Symbol;Acc:MGI:97722]	2.038
Pkd1l2	protein_coding	polycystic kidney disease 1 like 2 [Source:MGI Symbol;Acc:MGI:2664668]	2.03705
Gm6673	unprocessed_pseudogene	predicted gene 6673 [Source:MGI Symbol;Acc:MGI:3645556]	2.03506
Gm26811	antisense	predicted gene, 26811 [Source:MGI Symbol;Acc:MGI:5477305]	2.02898
Gm5576	processed_pseudogene	predicted pseudogene 5576 [Source:MGI Symbol;Acc:MGI:3644832]	2.02733
Rhcg	protein_coding	Rhesus blood group-associated C glycoprotein [Source:MGI Symbol;Acc:MGI:1888517]	2.01965
A730036I17Rik	lincRNA	RIKEN cDNA A730036I17 gene [Source:MGI Symbol;Acc:MGI:3041182]	2.01174
Akr1c21	protein_coding	aldo-keto reductase family 1, member C21 [Source:MGI Symbol;Acc:MGI:1924587]	2.01032
Wdr27	protein_coding	WD repeat domain 27 [Source:MGI Symbol;Acc:MGI:1918932]	1.9987
Gm44942	processed_pseudogene	predicted gene 44942 [Source:MGI Symbol;Acc:MGI:5753518]	1.99851
Tex26	protein_coding	testis expressed 26 [Source:MGI Symbol;Acc:MGI:1923110]	1.99832
Gm5611	processed_pseudogene	predicted gene 5611 [Source:MGI Symbol;Acc:MGI:3647121]	1.99777
Il17rc	protein_coding	interleukin 17 receptor C [Source:MGI Symbol;Acc:MGI:2159336]	1.99724
Tuba8	protein_coding	tubulin, alpha 8 [Source:MGI Symbol;Acc:MGI:1858225]	1.99701
5330417C22Rik	protein_coding	RIKEN cDNA 5330417C22 gene [Source:MGI Symbol;Acc:MGI:1923930]	1.99628
Gm44501	protein_coding	predicted readthrough transcript, 44501 [Source:MGI Symbol;Acc:MGI:5529083]	1.99606
Gm10327	processed_pseudogene	predicted pseudogene 10327 [Source:MGI Symbol;Acc:MGI:3704356]	1.99565
Gm44430	TEC	predicted gene, 44430 [Source:MGI Symbol;Acc:MGI:5690822]	1.9932
Arhgap27os1	antisense	Rho GTPase activating protein 27, opposite strand 1 [Source:MGI Symbol;Acc:MGI:3705097]	1.993
Col6a6	protein_coding	collagen, type VI, alpha 6 [Source:MGI Symbol;Acc:MGI:2444259]	1.99216
Pnoc	protein_coding	prepronociceptin [Source:MGI Symbol;Acc:MGI:105308]	1.9909
4930519A11Rik	lincRNA	RIKEN cDNA 4930519A11 gene [Source:MGI Symbol;Acc:MGI:1921946]	1.99016
Gm11517	processed_pseudogene	predicted gene 11517 [Source:MGI Symbol;Acc:MGI:3650064]	1.98674
Kcnv2	protein_coding	potassium channel, subfamily V, member 2 [Source:MGI Symbol;Acc:MGI:2670981]	1.98495
Gm26931	processed_pseudogene	predicted gene, 26931 [Source:MGI Symbol;Acc:MGI:5504046]	1.98428
Gm43766	TEC	predicted gene 43766 [Source:MGI Symbol;Acc:MGI:5663903]	1.98396
Olfr506	protein_coding	olfactory receptor 506 [Source:MGI Symbol;Acc:MGI:3030340]	1.98281
Pacrg	protein_coding	PARK2 co-regulated [Source:MGI Symbol;Acc:MGI:1916560]	1.98267
Gm8773	protein_coding	predicted gene 8773 [Source:MGI Symbol;Acc:MGI:3646213]	1.97986
Slc13a4	protein_coding	solute carrier family 13 (sodium/sulfate symporters), member 4 [Source:MGI Symbol;Acc:MGI:2442367]	1.97879
Gm12617	processed_pseudogene	predicted gene 12617 [Source:MGI Symbol;Acc:MGI:3649880]	1.97689
Gm6983	processed_pseudogene	predicted gene 6983 [Source:MGI Symbol;Acc:MGI:3646709]	1.97661
Fam183b	protein_coding	family with sequence similarity 183, member B [Source:MGI Symbol;Acc:MGI:1922679]	1.9766
Myot	protein_coding	myotilin [Source:MGI Symbol;Acc:MGI:1889800]	1.97644
Pigz	protein_coding	phosphatidylinositol glycan anchor biosynthesis, class Z [Source:MGI Symbol;Acc:MGI:2443822]	1.97507
Gm4959	processed_pseudogene	predicted gene 4959 [Source:MGI Symbol;Acc:MGI:3647044]	1.97363
Gm11688	processed_pseudogene	predicted gene 11688 [Source:MGI Symbol;Acc:MGI:3650868]	1.9736
Gm12424	processed_pseudogene	predicted gene 12424 [Source:MGI Symbol;Acc:MGI:3650991]	1.97323
Tmem132c	protein_coding	transmembrane protein 132C [Source:MGI Symbol;Acc:MGI:2443061]	1.97186
Crisp1	protein_coding	cysteine-rich secretory protein 1 [Source:MGI Symbol;Acc:MGI:102553]	1.97165
Gm15547	processed_pseudogene	predicted gene 15547 [Source:MGI Symbol;Acc:MGI:3782996]	1.9684
Gm14429	unprocessed_pseudogene	predicted gene 14429 [Source:MGI Symbol;Acc:MGI:3649510]	1.96692
I830134H01Rik	lincRNA	RIKEN cDNA I830134H01 gene [Source:MGI Symbol;Acc:MGI:3588250]	1.96645
Catsper2	protein_coding	cation channel, sperm associated 2 [Source:MGI Symbol;Acc:MGI:2387404]	1.96487
Hsbp1l1	protein_coding	heat shock factor binding protein 1-like 1 [Source:MGI Symbol;Acc:MGI:1913505]	1.96446
Ssmem1	protein_coding	serine-rich single-pass membrane protein 1 [Source:MGI Symbol;Acc:MGI:1922897]	1.96431
Slc27a5	protein_coding	solute carrier family 27 (fatty acid transporter), member 5 [Source:MGI Symbol;Acc:MGI:1347100]	1.9639
Klk1	protein_coding	kallikrein 1 [Source:MGI Symbol;Acc:MGI:102850]	1.96336
Elobl	protein_coding	elongin B-like [Source:MGI Symbol;Acc:MGI:1860403]	1.96288
Gm27239	antisense	predicted gene 27239 [Source:MGI Symbol;Acc:MGI:5521082]	1.9627
Gm5575	processed_pseudogene	predicted gene 5575 [Source:MGI Symbol;Acc:MGI:3647974]	1.96084
Hes5	protein_coding	hes family bHLH transcription factor 5 [Source:MGI Symbol;Acc:MGI:104876]	1.96057
Gm14321	lincRNA	predicted gene 14321 [Source:MGI Symbol;Acc:MGI:3701951]	1.9577
Lrrc36	protein_coding	leucine rich repeat containing 36 [Source:MGI Symbol;Acc:MGI:2448585]	1.95745
Klb	protein_coding	klotho beta [Source:MGI Symbol;Acc:MGI:1932466]	1.95706

Htr6	protein_coding	5-hydroxytryptamine (serotonin) receptor 6 [Source:MGI Symbol;Acc:MGI:1196627]	1.95681
Ggn	protein_coding	gametogenetin [Source:MGI Symbol;Acc:MGI:2181461]	1.95548
Scn4a	protein_coding	sodium channel, voltage-gated, type IV, alpha [Source:MGI Symbol;Acc:MGI:98250]	1.95509
Sybu	protein_coding	syntabulin (syntaxin-interacting) [Source:MGI Symbol;Acc:MGI:2442392]	1.95465
Gm5566	processed_pseudogene	predicted pseudogene 5566 [Source:MGI Symbol;Acc:MGI:3648428]	1.95449
Ggt5	protein_coding	gamma-glutamyltransferase 5 [Source:MGI Symbol;Acc:MGI:1346063]	1.95259
Slc4a1	protein_coding	solute carrier family 4 (anion exchanger), member 1 [Source:MGI Symbol;Acc:MGI:109393]	1.95217
D030055H07Rik	antisense	RIKEN cDNA D030055H07 gene [Source:MGI Symbol;Acc:MGI:3605036]	1.95162
Hoxa9	protein_coding	homeobox A9 [Source:MGI Symbol;Acc:MGI:96180]	1.95082
Khdrbs2	protein_coding	KH domain containing, RNA binding, signal transduction associated 2 [Source:MGI Symbol;Acc:MGI:2159649]	1.9474
Ly6g6d	protein_coding	lymphocyte antigen 6 complex, locus G6D [Source:MGI Symbol;Acc:MGI:2148931]	1.94669
Bmper	protein_coding	BMP-binding endothelial regulator [Source:MGI Symbol;Acc:MGI:1920480]	1.94657
Platr29	lincRNA	pluripotency associated transcript 29 [Source:MGI Symbol;Acc:MGI:3650001]	1.94628
Hspb9	protein_coding	heat shock protein, alpha-crystallin-related, B9 [Source:MGI Symbol;Acc:MGI:1922732]	1.94621
Gm31693	lincRNA	predicted gene, 31693 [Source:MGI Symbol;Acc:MGI:5590852]	1.94521
Gm11851	lincRNA	predicted gene 11851 [Source:MGI Symbol;Acc:MGI:3652014]	1.94399
Gm16239	antisense	predicted gene 16239 [Source:MGI Symbol;Acc:MGI:3801746]	1.9439
Gm11533	processed_pseudogene	predicted gene 11533 [Source:MGI Symbol;Acc:MGI:3651874]	1.94341
Gm8522	processed_pseudogene	predicted gene 8522 [Source:MGI Symbol;Acc:MGI:3645035]	1.94239
Gm16259	processed_pseudogene	predicted gene 16259 [Source:MGI Symbol;Acc:MGI:3826542]	1.94221
4930451G09Rik	transcribed_unitary_pseudogene	RIKEN cDNA 4930451G09 gene [Source:MGI Symbol;Acc:MGI:1921934]	1.94106
Gm13171	processed_pseudogene	predicted gene 13171 [Source:MGI Symbol;Acc:MGI:3651089]	1.94014
Gm15097	protein_coding	predicted gene 15097 [Source:MGI Symbol;Acc:MGI:3710639]	1.93983
Gm15190	processed_pseudogene	predicted gene 15190 [Source:MGI Symbol;Acc:MGI:3705630]	1.93948
Hist1h2br	protein_coding	histone cluster 1 H2br [Source:MGI Symbol;Acc:MGI:3710645]	1.93929
Gm42903	lincRNA	predicted gene 42903 [Source:MGI Symbol;Acc:MGI:5663040]	1.93928
Gm32486	lincRNA	predicted gene, 32486 [Source:MGI Symbol;Acc:MGI:5591645]	1.93767
Gm32486	lincRNA	predicted gene, 32486 [Source:MGI Symbol;Acc:MGI:5591645]	1.93767
B630019A10Rik	antisense	RIKEN cDNA B630019A10 gene [Source:MGI Symbol;Acc:MGI:2443371]	1.9372
Gm5522	processed_pseudogene	predicted gene 5522 [Source:MGI Symbol;Acc:MGI:3649087]	1.93694
Abcb5	protein_coding	ATP-binding cassette, sub-family B (MDR/TAP), member 5 [Source:MGI Symbol;Acc:MGI:1924956]	1.93492
Olfr420	protein_coding	olfactory receptor 420 [Source:MGI Symbol;Acc:MGI:3030254]	1.93457
Gm38560	bidirectional_promoter_lncRNA	predicted gene, 38560 [Source:MGI Symbol;Acc:MGI:5621445]	1.93427
9530062K07Rik	bidirectional_promoter_lncRNA	RIKEN cDNA 9530062K07 gene [Source:MGI Symbol;Acc:MGI:1925989]	1.93407
Gm15697	unprocessed_pseudogene	predicted gene 15697 [Source:MGI Symbol;Acc:MGI:3783138]	1.93335
Gm22204	misc_RNA	predicted gene, 22204 [Source:MGI Symbol;Acc:MGI:5451981]	1.93115
Gm5857	processed_pseudogene	predicted gene 5857 [Source:MGI Symbol;Acc:MGI:3644956]	1.93036
Muc16	protein_coding	mucin 16 [Source:MGI Symbol;Acc:MGI:1920982]	1.92765
9530085L11Rik	TEC	RIKEN cDNA 9530085L11 gene [Source:MGI Symbol;Acc:MGI:2443889]	1.92631
Gm30934	antisense	predicted gene, 30934 [Source:MGI Symbol;Acc:MGI:5590093]	1.92433
Gm33148	lincRNA	predicted gene, 33148 [Source:MGI Symbol;Acc:MGI:5592307]	1.92407
Ttc30a2	protein_coding	tetratricopeptide repeat domain 30A2 [Source:MGI Symbol;Acc:MGI:3700200]	1.92398
Gm12052	lincRNA	predicted gene 12052 [Source:MGI Symbol;Acc:MGI:3652305]	1.92306
Gm30893	antisense	predicted gene, 30893 [Source:MGI Symbol;Acc:MGI:5590052]	1.92305
Smim5	protein_coding	small integral membrane protein 5 [Source:MGI Symbol;Acc:MGI:1913778]	1.92216
Gm12058	processed_pseudogene	predicted gene 12058 [Source:MGI Symbol;Acc:MGI:3652312]	1.92144
Gm16312	antisense	predicted gene 16312 [Source:MGI Symbol;Acc:MGI:3826559]	1.91144
Gm11423	antisense	predicted gene 11423 [Source:MGI Symbol;Acc:MGI:3651335]	1.91052
Hcar1	protein_coding	hydrocarboxylic acid receptor 1 [Source:MGI Symbol;Acc:MGI:2441671]	1.91039
Gm45827	lincRNA	predicted gene 45827 [Source:MGI Symbol;Acc:MGI:5804942]	1.90841
Tmem254a	protein_coding	transmembrane protein 254a [Source:MGI Symbol;Acc:MGI:1196450]	1.90837
Rnf212b	protein_coding	ring finger protein 212B [Source:MGI Symbol;Acc:MGI:5589964]	1.90505
Gm14373	processed_pseudogene	predicted gene 14373 [Source:MGI Symbol;Acc:MGI:3651800]	1.90499
Unc93a	protein_coding	unc-93 homolog A [Source:MGI Symbol;Acc:MGI:1933250]	1.90434
Slc7a9	protein_coding	solute carrier family 7 (cationic amino acid transporter, y+ system), member 9 [Source:MGI Symbol;Acc:MGI:1353656]	1.89926
Cfap161	protein_coding	cilia and flagella associated protein 161 [Source:MGI Symbol;Acc:MGI:1922806]	1.89868
9130019P16Rik	processed_transcript	RIKEN cDNA 9130019P16 gene [Source:MGI Symbol;Acc:MGI:1918824]	1.8983
Tmod4	protein_coding	tropomodulin 4 [Source:MGI Symbol;Acc:MGI:1355285]	1.89531
Smc1b	protein_coding	structural maintenance of chromosomes 1B [Source:MGI Symbol;Acc:MGI:2154049]	1.89474
Apoc2	protein_coding	apolipoprotein C-II [Source:MGI Symbol;Acc:MGI:88054]	1.89157
Ifi27l2a	protein_coding	interferon, alpha-inducible protein 27 like 2A [Source:MGI Symbol;Acc:MGI:1924183]	1.89083
Spi1	protein_coding	spleen focus forming virus (SFFV) proviral integration oncogene [Source:MGI Symbol;Acc:MGI:98282]	1.88953

Vmn1r197	protein_coding	vomeronasal 1 receptor 197 [Source:MGI Symbol;Acc:MGI:2159686]	1.88692
Gm38307	TEC	predicted gene, 38307 [Source:MGI Symbol;Acc:MGI:5611535]	1.88597
Kcnn3	protein_coding	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3 [Source:MGI Symbol;Acc:MGI:2153183]	1.88592
Gm15843	processed_pseudogene	predicted gene 15843 [Source:MGI Symbol;Acc:MGI:3801845]	1.88516
E230014E18Rik	lincRNA	RIKEN cDNA E230014E18 gene [Source:MGI Symbol;Acc:MGI:3045388]	1.88336
Gm20531	antisense	predicted gene 20531 [Source:MGI Symbol;Acc:MGI:5141996]	1.88326
Gm5422	transcribed_processed_pseudogene	predicted pseudogene 5422 [Source:MGI Symbol;Acc:MGI:3643411]	1.88101
A430110C17Rik	TEC	RIKEN cDNA A430110C17 gene [Source:MGI Symbol;Acc:MGI:2444994]	1.87881
Trim80	protein_coding	tripartite motif-containing 80 [Source:MGI Symbol;Acc:MGI:3588186]	1.87551
Akr1c13	protein_coding	aldo-keto reductase family 1, member C13 [Source:MGI Symbol;Acc:MGI:1351662]	1.87081
Nap1l5	protein_coding	nucleosome assembly protein 1-like 5 [Source:MGI Symbol;Acc:MGI:1923555]	1.86934
Magix	protein_coding	MAGI family member, X-linked [Source:MGI Symbol;Acc:MGI:1859644]	1.86907
Gm26525	lincRNA	predicted gene, 26525 [Source:MGI Symbol;Acc:MGI:5477019]	1.867
Gm11729	lincRNA	predicted gene 11729 [Source:MGI Symbol;Acc:MGI:3650942]	1.86595
Csgalnact1	protein_coding	chondroitin sulfate N-acetylgalactosaminyltransferase 1 [Source:MGI Symbol;Acc:MGI:2442354]	1.86199
Gtf2ird2	protein_coding	GTF2I repeat domain containing 2 [Source:MGI Symbol;Acc:MGI:2149780]	1.85834
Gm38948	lincRNA	predicted gene, 38948 [Source:MGI Symbol;Acc:MGI:5621833]	1.85782
Krt79	protein_coding	keratin 79 [Source:MGI Symbol;Acc:MGI:2385030]	1.85732
P2ry12	protein_coding	purinergic receptor P2Y, G-protein coupled 12 [Source:MGI Symbol;Acc:MGI:1918089]	1.85582
Gm13216	processed_pseudogene	predicted gene 13216 [Source:MGI Symbol;Acc:MGI:3651709]	1.85474
Rbx1-ps	processed_pseudogene	ring-box 1, pseudogene [Source:MGI Symbol;Acc:MGI:3710517]	1.85439
Gm45848	lincRNA	predicted gene 45848 [Source:MGI Symbol;Acc:MGI:5804963]	1.85289
Rpl3-ps2	processed_pseudogene	ribosomal protein L3, pseudogene 2 [Source:MGI Symbol;Acc:MGI:3645345]	1.84911
Cyp24a1	protein_coding	cytochrome P450, family 24, subfamily a, polypeptide 1 [Source:MGI Symbol;Acc:MGI:88593]	1.84672
Gm9824	processed_pseudogene	predicted pseudogene 9824 [Source:MGI Symbol;Acc:MGI:3641964]	1.84366
Gm19531	antisense	predicted gene, 19531 [Source:MGI Symbol;Acc:MGI:5011716]	1.84101
Gm15846	processed_pseudogene	predicted gene 15846 [Source:MGI Symbol;Acc:MGI:3801824]	1.83399
Gm11832	antisense	predicted gene 11832 [Source:MGI Symbol;Acc:MGI:3650310]	1.83058
Gm6457	processed_pseudogene	predicted pseudogene 6457 [Source:MGI Symbol;Acc:MGI:3648519]	1.82948
Fa2h	protein_coding	fatty acid 2-hydroxylase [Source:MGI Symbol;Acc:MGI:2443327]	1.82857
Lims2	protein_coding	LIM and senescent cell antigen like domains 2 [Source:MGI Symbol;Acc:MGI:2385067]	1.82709
Cbln3	protein_coding	cerebellin 3 precursor protein [Source:MGI Symbol;Acc:MGI:1889286]	1.82688
Gm10157	processed_pseudogene	predicted gene 10157 [Source:MGI Symbol;Acc:MGI:3642264]	1.81345
4833418N02Rik	lincRNA	RIKEN cDNA 4833418N02 gene [Source:MGI Symbol;Acc:MGI:1921847]	1.80943
Tex16	protein_coding	testis expressed gene 16 [Source:MGI Symbol;Acc:MGI:1890545]	1.80778
Gm44421	lincRNA	predicted gene, 44421 [Source:MGI Symbol;Acc:MGI:5690813]	1.80694
Gm6274	processed_pseudogene	predicted gene 6274 [Source:MGI Symbol;Acc:MGI:3645367]	1.80576
Six2	protein_coding	sine oculis-related homeobox 2 [Source:MGI Symbol;Acc:MGI:102778]	1.80415
Gm14407	processed_pseudogene	predicted gene 14407 [Source:MGI Symbol;Acc:MGI:3649809]	1.79704
Sult2a4	protein_coding	sulfotransferase family 2A, dehydroepiandrosterone (DHEA)-preferring, member 4 [Source:MGI Symbol;Acc:MGI:3645854]	1.79287
Gad1os	antisense	glutamate decarboxylase 1, opposite strand [Source:MGI Symbol;Acc:MGI:1916236]	1.79243
Gm5069	transcribed_processed_pseudogene	predicted pseudogene 5069 [Source:MGI Symbol;Acc:MGI:3644516]	1.79024
Gm42869	sense_intronic	predicted gene 42869 [Source:MGI Symbol;Acc:MGI:5663006]	1.78939
Gm45551	unprocessed_pseudogene	predicted gene 45551 [Source:MGI Symbol;Acc:MGI:5791387]	1.78514
Trim43a	protein_coding	tripartite motif-containing 43A [Source:MGI Symbol;Acc:MGI:3645218]	1.78134
Gm29100	antisense	predicted gene 29100 [Source:MGI Symbol;Acc:MGI:5579806]	1.77466
Gm26884	lincRNA	predicted gene, 26884 [Source:MGI Symbol;Acc:MGI:5477378]	1.7736
Gm7461	TEC	predicted gene 7461 [Source:MGI Symbol;Acc:MGI:3645338]	1.77154
Slco1b2	protein_coding	solute carrier organic anion transporter family, member 1b2 [Source:MGI Symbol;Acc:MGI:1351899]	1.76735
Gm14286	antisense	predicted gene 14286 [Source:MGI Symbol;Acc:MGI:3650190]	1.76692
Gm15884	antisense	predicted gene 15884 [Source:MGI Symbol;Acc:MGI:3801897]	1.76354
Abcc9	protein_coding	ATP-binding cassette, sub-family C (CFTR/MRP), member 9 [Source:MGI Symbol;Acc:MGI:1352630]	1.76298
Gm42692	TEC	predicted gene 42692 [Source:MGI Symbol;Acc:MGI:5662829]	1.7625
Gm21399	processed_pseudogene	predicted gene, 21399 [Source:MGI Symbol;Acc:MGI:5434754]	1.75991
Gm37855	lincRNA	predicted gene, 37855 [Source:MGI Symbol;Acc:MGI:5611083]	1.74967
AC225888.1	pseudogene		1.74779
Olfr1156	protein_coding	olfactory receptor 1156 [Source:MGI Symbol;Acc:MGI:3030990]	1.74779
Gm45088	TEC	predicted gene 45088 [Source:MGI Symbol;Acc:MGI:5753664]	1.74557
Gm31532	processed_pseudogene	predicted gene, 31532 [Source:MGI Symbol;Acc:MGI:5590691]	1.74555
Ddx60	protein_coding	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 [Source:MGI Symbol;Acc:MGI:2384570]	1.741
Cdh23	protein_coding	cadherin 23 (otocadherin) [Source:MGI Symbol;Acc:MGI:1890219]	1.73977
Krt72	protein_coding	keratin 72 [Source:MGI Symbol;Acc:MGI:2146034]	1.7343
Grik2	protein_coding	glutamate receptor, ionotropic, kainate 2 (beta 2) [Source:MGI	1.72609

		Symbol;Acc:MGI:95815]	
Dpep2	protein_coding	dipeptidase 2 [Source:MGI Symbol;Acc:MGI:2442042]	1.72019
Dpep2	protein_coding	dipeptidase 2 [Source:MGI Symbol;Acc:MGI:2442042]	1.72019
Gm45591	sense_intronic	predicted gene 45591 [Source:MGI Symbol;Acc:MGI:5791427]	1.71664
Gm45235	processed_transcript	predicted gene 45235 [Source:MGI Symbol;Acc:MGI:5753811]	1.71578
Gm26860	lincRNA	predicted gene, 26860 [Source:MGI Symbol;Acc:MGI:5477354]	1.71466
Gm26555	antisense	predicted gene, 26555 [Source:MGI Symbol;Acc:MGI:5477049]	1.71024
Gm29724	antisense	predicted gene, 29724 [Source:MGI Symbol;Acc:MGI:5588883]	1.70835
Mir124-2hg	lincRNA	Mir124-2 host gene (non-protein coding) [Source:MGI Symbol;Acc:MGI:1917691]	1.70238
Slc6a11	protein_coding	solute carrier family 6 (neurotransmitter transporter, GABA), member 11 [Source:MGI Symbol;Acc:MGI:95630]	1.70003
Gm35208	lincRNA	predicted gene, 35208 [Source:MGI Symbol;Acc:MGI:5594367]	1.69813
Rpl15-ps5	processed_pseudogene	ribosomal protein L15, pseudogene 5 [Source:MGI Symbol;Acc:MGI:5010232]	1.69235
4930412M03Rik	lincRNA	RIKEN cDNA 4930412M03 gene [Source:MGI Symbol;Acc:MGI:2443570]	1.68756
Tnk2os	antisense	tyrosine kinase, non-receptor 2, opposite strand [Source:MGI Symbol;Acc:MGI:3642140]	1.68629
Il17f	protein_coding	interleukin 17F [Source:MGI Symbol;Acc:MGI:2676631]	1.68344
Gm14398	processed_pseudogene	predicted gene 14398 [Source:MGI Symbol;Acc:MGI:3650079]	1.68187
Gm10284	processed_pseudogene	predicted pseudogene 10284 [Source:MGI Symbol;Acc:MGI:3704311]	1.68061
Sowaha	protein_coding	sosondowah ankyrin repeat domain family member A [Source:MGI Symbol;Acc:MGI:2687280]	1.67344
Gm43633	TEC	predicted gene 43633 [Source:MGI Symbol;Acc:MGI:5663770]	1.66303
Adcy4	protein_coding	adenylate cyclase 4 [Source:MGI Symbol;Acc:MGI:99674]	1.65902
Gm4707	protein_coding	predicted gene 4707 [Source:MGI Symbol;Acc:MGI:3782887]	1.65698
Gm18562	unprocessed_pseudogene	predicted gene, 18562 [Source:MGI Symbol;Acc:MGI:5010747]	1.6557
AV356131	bidirectional_promoter_lncRNA	expressed sequence AV356131 [Source:MGI Symbol;Acc:MGI:2142184]	1.65423
Grin1	protein_coding	glutamate receptor, ionotropic, NMDA1 (zeta 1) [Source:MGI Symbol;Acc:MGI:95819]	1.64781
Gm26642	antisense	predicted gene, 26642 [Source:MGI Symbol;Acc:MGI:5477136]	1.64452
C430014B12Rik	processed_transcript	RIKEN cDNA C430014B12 gene [Source:MGI Symbol;Acc:MGI:2444820]	1.64163
Gm16253	protein_coding	predicted gene 16253 [Source:MGI Symbol;Acc:MGI:3826531]	1.63977
Gm9354	processed_pseudogene	predicted gene 9354 [Source:MGI Symbol;Acc:MGI:3647598]	1.63944
Gm38979	antisense	predicted gene, 38979 [Source:MGI Symbol;Acc:MGI:5621864]	1.63491
Gm6129	processed_pseudogene	predicted gene 6129 [Source:MGI Symbol;Acc:MGI:3645100]	1.6284
Rps8-ps1	processed_pseudogene	ribosomal protein S8, pseudogene 1 [Source:MGI Symbol;Acc:MGI:3647445]	1.62572
2210406O10Rik	lincRNA	RIKEN cDNA 2210406O10 gene [Source:MGI Symbol;Acc:MGI:1923960]	1.62444
Abhd15	protein_coding	abhydrolase domain containing 15 [Source:MGI Symbol;Acc:MGI:1914727]	1.62191
Gm36595	lincRNA	predicted gene, 36595 [Source:MGI Symbol;Acc:MGI:5595754]	1.6216
Gpr35	protein_coding	G protein-coupled receptor 35 [Source:MGI Symbol;Acc:MGI:1929509]	1.61258
Htr4	protein_coding	5 hydroxytryptamine (serotonin) receptor 4 [Source:MGI Symbol;Acc:MGI:109246]	1.61038
Calr4	protein_coding	calreticulin 4 [Source:MGI Symbol;Acc:MGI:2140435]	1.60936
Gm9234	processed_pseudogene	predicted pseudogene 9234 [Source:MGI Symbol;Acc:MGI:3648545]	1.60575
Nlrp9c	protein_coding	NLR family, pyrin domain containing 9C [Source:MGI Symbol;Acc:MGI:3028627]	1.60513
Upb1	protein_coding	ureidopropionase, beta [Source:MGI Symbol;Acc:MGI:2143535]	1.6049
Gm37233	lincRNA	predicted gene, 37233 [Source:MGI Symbol;Acc:MGI:5610461]	1.60452
Tubb4b-ps1	processed_pseudogene	tubulin, beta 4B class IVB, pseudogene 1 [Source:MGI Symbol;Acc:MGI:3642800]	1.6043
Gm26616	TEC	predicted gene, 26616 [Source:MGI Symbol;Acc:MGI:5477110]	1.60425
Gm15607	processed_pseudogene	predicted gene 15607 [Source:MGI Symbol;Acc:MGI:3783054]	1.60372
Gdf6	protein_coding	growth differentiation factor 6 [Source:MGI Symbol;Acc:MGI:95689]	1.60351
4930451E10Rik	antisense	RIKEN cDNA 4930451E10 gene [Source:MGI Symbol;Acc:MGI:1925361]	1.59881
Nrg2	protein_coding	neuregulin 2 [Source:MGI Symbol;Acc:MGI:1098246]	1.59667
Tmem132d	protein_coding	transmembrane protein 132D [Source:MGI Symbol;Acc:MGI:3044963]	1.59004
Fam71e1	protein_coding	family with sequence similarity 71, member E1 [Source:MGI Symbol;Acc:MGI:1922788]	1.5864
Bpifb5	protein_coding	BPI fold containing family B, member 5 [Source:MGI Symbol;Acc:MGI:2385160]	1.58232
2900027M19Rik	TEC	RIKEN cDNA 2900027M19 gene [Source:MGI Symbol;Acc:MGI:1920130]	1.58025
F2rl3	protein_coding	coagulation factor II (thrombin) receptor-like 3 [Source:MGI Symbol;Acc:MGI:1298207]	1.57639
Card11	protein_coding	caspase recruitment domain family, member 11 [Source:MGI Symbol;Acc:MGI:1916978]	1.57332
Gm19972	TEC	predicted gene, 19972 [Source:MGI Symbol;Acc:MGI:5012157]	1.57286
Gm45354	TEC	predicted gene 45354 [Source:MGI Symbol;Acc:MGI:5791190]	1.57208
Mybphl	protein_coding	myosin binding protein H-like [Source:MGI Symbol;Acc:MGI:1916003]	1.56835
Myo3a	protein_coding	myosin IIIA [Source:MGI Symbol;Acc:MGI:2183924]	1.56664
Gm11362	processed_pseudogene	predicted gene 11362 [Source:MGI Symbol;Acc:MGI:3651391]	1.56122
Gm38218	TEC	predicted gene, 38218 [Source:MGI Symbol;Acc:MGI:5611446]	1.56076
Iigp1	protein_coding	interferon inducible GTPase 1 [Source:MGI Symbol;Acc:MGI:1926259]	1.56018
Gm5559	processed_pseudogene	predicted gene 5559 [Source:MGI Symbol;Acc:MGI:3779498]	1.55729
Vsx2	protein_coding	visual system homeobox 2 [Source:MGI Symbol;Acc:MGI:88401]	1.55701

Pld5	protein_coding	phospholipase D family, member 5 [Source:MGI Symbol;Acc:MGI:2442056]	1.55522
Gm17210	antisense	predicted gene 17210 [Source:MGI Symbol;Acc:MGI:4938037]	1.54948
Gm38099	TEC	predicted gene, 38099 [Source:MGI Symbol;Acc:MGI:5611327]	1.54924
Six3os1	processed_transcript	SIX homeobox 3, opposite strand 1 [Source:MGI Symbol;Acc:MGI:1925118]	1.5481
Gm8098	processed_transcript	predicted gene 8098 [Source:MGI Symbol;Acc:MGI:3645749]	1.54553
Gm45174	TEC	predicted gene 45174 [Source:MGI Symbol;Acc:MGI:5753750]	1.54509
Gm7420	transcribed_processed_pseudogene	predicted gene 7420 [Source:MGI Symbol;Acc:MGI:3647909]	1.54332
Cped1	protein_coding	cadherin-like and PC-esterase domain containing 1 [Source:MGI Symbol;Acc:MGI:2444814]	1.53954
Thrsp	protein_coding	thyroid hormone responsive [Source:MGI Symbol;Acc:MGI:109126]	1.53853
Tnfaip8l2	protein_coding	tumor necrosis factor, alpha-induced protein 8-like 2 [Source:MGI Symbol;Acc:MGI:1917019]	1.5361
Gsto2	protein_coding	glutathione S-transferase omega 2 [Source:MGI Symbol;Acc:MGI:1915464]	1.53513
Gm45598	transcribed_unprocessed_pseudogene	predicted gene 45598 [Source:MGI Symbol;Acc:MGI:5791434]	1.53466
Eif5al3-ps	processed_pseudogene	eukaryotic translation initiation factor 5A-like 3, pseudogene [Source:MGI Symbol;Acc:MGI:3643585]	1.53362
Npm3-ps1	processed_pseudogene	nucleoplasmin 3, pseudogene 1 [Source:MGI Symbol;Acc:MGI:894683]	1.53352
Acpp	protein_coding	acid phosphatase, prostate [Source:MGI Symbol;Acc:MGI:1928480]	1.53196
Gm11397	protein_coding	predicted gene 11397 [Source:MGI Symbol;Acc:MGI:3709608]	1.53102
Gm6641	processed_pseudogene	predicted gene 6641 [Source:MGI Symbol;Acc:MGI:3645310]	1.52928
Gm26943	processed_pseudogene	predicted gene, 26943 [Source:MGI Symbol;Acc:MGI:5504058]	1.52735
Npffr1	protein_coding	neuropeptide FF receptor 1 [Source:MGI Symbol;Acc:MGI:2685082]	1.5254
Crb1-ps	processed_pseudogene	crumbs family member 1, photoreceptor morphogenesis associated, pseudogene [Source:MGI Symbol;Acc:MGI:5645790]	1.52423
Rps12-ps10	processed_pseudogene	ribosomal protein S12, pseudogene 10 [Source:MGI Symbol;Acc:MGI:3649508]	1.52336
1700086D15Rik	protein_coding	RIKEN cDNA 1700086D15 gene [Source:MGI Symbol;Acc:MGI:1921532]	1.52325
1700028I16Rik	lincRNA	RIKEN cDNA 1700028I16 gene [Source:MGI Symbol;Acc:MGI:1917253]	1.52319
Gm13785	processed_pseudogene	predicted gene 13785 [Source:MGI Symbol;Acc:MGI:3649781]	1.52266
Mybpc1	protein_coding	myosin binding protein C, slow-type [Source:MGI Symbol;Acc:MGI:1336213]	1.52215
Serpina3k	protein_coding	serine (or cysteine) peptidase inhibitor, clade A, member 3K [Source:MGI Symbol;Acc:MGI:98377]	1.52058
Gm40614	lincRNA	predicted gene, 40614 [Source:MGI Symbol;Acc:MGI:5623499]	1.52026
Gm16133	antisense	predicted gene 16133 [Source:MGI Symbol;Acc:MGI:3802155]	1.51724
Gm43769	TEC	predicted gene 43769 [Source:MGI Symbol;Acc:MGI:5663906]	1.51719
Gm9967	processed_transcript	predicted gene 9967 [Source:MGI Symbol;Acc:MGI:3704300]	1.51699
Tnfrsf14	protein_coding	tumor necrosis factor receptor superfamily, member 14 (herpesvirus entry mediator) [Source:MGI Symbol;Acc:MGI:2675303]	1.51544
4632404M16Rik	sense_intronic	RIKEN cDNA 4632404M16 gene [Source:MGI Symbol;Acc:MGI:1921598]	1.51417
Gm9899	processed_transcript	predicted gene 9899 [Source:MGI Symbol;Acc:MGI:3708711]	1.51023
Gm11827	lincRNA	predicted gene 11827 [Source:MGI Symbol;Acc:MGI:3649801]	1.50968
Rnase12	protein_coding	ribonuclease, RNase A family, 12 (non-active) [Source:MGI Symbol;Acc:MGI:3528588]	1.50937
Tbx2	protein_coding	T-box 2 [Source:MGI Symbol;Acc:MGI:98494]	1.50887
5430403N17Rik	lincRNA	RIKEN cDNA 5430403N17 gene [Source:MGI Symbol;Acc:MGI:5439418]	1.50776
Gm16285	processed_pseudogene	predicted gene 16285 [Source:MGI Symbol;Acc:MGI:3826537]	1.50765
Flt3	protein_coding	FMS-like tyrosine kinase 3 [Source:MGI Symbol;Acc:MGI:95559]	1.50741
Gm42829	sense_intronic	predicted gene 42829 [Source:MGI Symbol;Acc:MGI:5662966]	1.50702
Gm13362	antisense	predicted gene 13362 [Source:MGI Symbol;Acc:MGI:3652150]	1.50606
Cyp21a1	protein_coding	cytochrome P450, family 21, subfamily a, polypeptide 1 [Source:MGI Symbol;Acc:MGI:88591]	1.50551
Cdhr5	protein_coding	cadherin-related family member 5 [Source:MGI Symbol;Acc:MGI:1919290]	1.50407
Wfdc1	protein_coding	WAP four-disulfide core domain 1 [Source:MGI Symbol;Acc:MGI:1915116]	1.5032

References

1. Dahm, R., *Friedrich Miescher and the discovery of DNA.* Dev Biol, 2005. **278**(2): p. 274-88.
2. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.
3. Wells, W.A., *Ribosomes, or the particles of Palade*. J Cell Biol. 2005 Jan 3;168(1):12. doi: 10.1083/jcb1681fta3.
4. Hoagland, M.B., et al., *A soluble ribonucleic acid intermediate in protein synthesis.* J Biol Chem, 1958. **231**(1): p. 241-57.
5. Coffin, J.M. and H. Fan, *The Discovery of Reverse Transcriptase.* Annu Rev Virol, 2016. **3**(1): p. 29-51.
6. Guo, F., A.R. Gooding, and T.R. Cech, *Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site.* Mol Cell, 2004. **16**(3): p. 351-62.
7. Baer, M.F., et al., *The recognition by RNase P of precursor tRNAs.* J Biol Chem, 1988. **263**(5): p. 2344-51.
8. Robertson, H.D., S. Altman, and J.D. Smith, *Purification and properties of a specific Escherichia coli ribonuclease which cleaves a tyrosine transfer ribonucleic acid presursor.* J Biol Chem, 1972. **247**(16): p. 5243-51.
9. Robertson, M.P. and G.F. Joyce, *The origins of the RNA world.* Cold Spring Harb Perspect Biol, 2012. **4**(5).
10. Sen, G.L. and H.M. Blau, *A brief history of RNAi: the silence of the genes.* Faseb j, 2006. **20**(9): p. 1293-9.
11. de Hoon, M., J.W. Shin, and P. Carninci, *Paradigm shifts in genomics through the FANTOM projects.* Mamm Genome, 2015. **26**(9-10): p. 391-402.
12. *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
13. Dinger, M.E., et al., *Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications.* Brief Funct Genomic Proteomic, 2009. **8**(6): p. 407-23.
14. Kim, V.N., *Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes.* Genes Dev, 2006. **20**(15): p. 1993-7.
15. Ishizu, H., H. Siomi, and M.C. Siomi, *Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines*. Genes Dev, 2012. **26**(21): p. 2361-73.
16. Jensen, T.H., A. Jacquier, and D. Libri, *Dealing with pervasive transcription.* Mol Cell, 2013. **52**(4): p. 473-84.
17. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res, 2012. **22**(9): p. 1760-74.
18. Bartolomei, M.S., S. Zemel, and S.M. Tilghman, *Parental imprinting of the mouse H19 gene.* Nature, 1991. **351**(6322): p. 153-5.
19. Brown, C.J., et al., *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus.* Cell, 1992. **71**(3): p. 527-42.
20. Jandura, A. and H.M. Krause, *The New RNA World: Growing Evidence for Long Noncoding RNA Functionality.* Trends Genet, 2017. **33**(10): p. 665-676.
21. Ward, M., et al., *Conservation and tissue-specific transcription patterns of long noncoding RNAs.* J Hum Transcr, 2015. **1**(1): p. 2-9.
22. Kapusta, A. and C. Feschotte, *Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications.* Trends Genet, 2014. **30**(10): p. 439-52.
23. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.* Nature, 2009. **458**(7235): p. 223-7.
24. Kino, T., et al., *Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor.* Sci Signal, 2010. **3**(107): p. ra8.
25. Maenner, S., et al., *2-D structure of the A region of Xist RNA and its implication for PRC2 association.* PLoS Biol, 2010. **8**(1): p. e1000276.
26. Wutz, A., T.P. Rasmussen, and R. Jaenisch, *Chromosomal silencing and localization are mediated by different domains of Xist RNA.* Nat Genet, 2002. **30**(2): p. 167-74.
27. Gutschner, T., M. Hammerle, and S. Diederichs, *MALAT1 -- a paradigm for long noncoding RNA function in cancer.* J Mol Med (Berl), 2013. **91**(7): p. 791-801.

28. Ard, R., R.C. Allshire, and S. Marquardt, *Emerging Properties and Functional Consequences of Noncoding Transcription.* Genetics, 2017. **207**(2): p. 357-367.
29. Sati, S., et al., *Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci.* Nucleic Acids Res, 2012. **40**(20): p. 10018-31.
30. Grote, P., et al., *The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse.* Dev Cell, 2013. **24**(2): p. 206-14.
31. Klattenhoff, C.A., et al., *Braveheart, a long noncoding RNA required for cardiovascular lineage commitment.* Cell, 2013. **152**(3): p. 570-83.
32. Cesana, M., et al., *A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA.* Cell, 2011. **147**(2): p. 358-69.
33. Sun, M. and W.L. Kraus, *From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease.* Endocr Rev, 2015. **36**(1): p. 25-64.
34. Kanduri, C., *Kcnq1ot1: a chromatin regulatory RNA.* Semin Cell Dev Biol, 2011. **22**(4): p. 343-50.
35. Kaneko, S., et al., *Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA.* Genes Dev, 2010. **24**(23): p. 2615-20.
36. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.* Science, 2008. **322**(5902): p. 750-6.
37. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.
38. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes.* Science, 2010. **329**(5992): p. 689-93.
39. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.* Cell, 2010. **142**(3): p. 409-19.
40. Barsotti, A.M., et al., *p53-Dependent induction of PVT1 and miR-1204.* J Biol Chem, 2012. **287**(4): p. 2509-19.
41. Werner, M.S., et al., *Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription.* Nat Struct Mol Biol, 2017. **24**(7): p. 596-603.
42. Brown, J.A., et al., *Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix.* Nat Struct Mol Biol, 2014. **21**(7): p. 633-40.
43. Keller, C., et al., *Noncoding RNAs prevent spreading of a repressive histone mark.* Nat Struct Mol Biol, 2013. **20**(11): p. 1340.
44. Mondal, T., et al., *MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures.* Nat Commun, 2015. **6**: p. 7743.
45. Grote, P. and B.G. Herrmann, *The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis.* RNA Biol, 2013. **10**(10): p. 1579-85.
46. Carpenter, S., et al., *A long noncoding RNA mediates both activation and repression of immune response genes.* Science, 2013. **341**(6147): p. 789-92.
47. Han, P., et al., *A long noncoding RNA protects the heart from pathological hypertrophy.* Nature, 2014. **514**(7520): p. 102-106.
48. Berghoff, E.G., et al., *Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes.* Development, 2013. **140**(21): p. 4407-16.
49. Cajigas, I., et al., *Evf2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling.* Development, 2015. **142**(15): p. 2641-52.
50. Kim, T.K., M. Hemberg, and J.M. Gray, *Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers.* Cold Spring Harb Perspect Biol, 2015. **7**(1): p. a018622.
51. Ling, J., et al., *The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter.* J Mol Biol, 2005. **350**(5): p. 883-96.
52. Kim, T.K., et al., *Widespread transcription at neuronal activity-regulated enhancers.* Nature, 2010. **465**(7295): p. 182-7.
53. NE, I.I., et al., *Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes.* Nat Commun, 2014. **5**: p. 3979.

54. Miao, Y., et al., *Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function.* Nat Commun, 2018. **9**(1): p. 292.
55. Sleutels, F., R. Zwart, and D.P. Barlow, *The non-coding Air RNA is required for silencing autosomal imprinted genes.* Nature, 2002. **415**(6873): p. 810-3.
56. Wutz, A., et al., *Imprinted expression of the Igf2r gene depends on an intronic CpG island.* Nature, 1997. **389**(6652): p. 745-9.
57. Nagano, T., et al., *The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin.* Science, 2008. **322**(5908): p. 1717-20.
58. Lu, M.H., et al., *Long noncoding RNA BC032469, a novel competing endogenous RNA, upregulates hTERT expression by sponging miR-1207-5p and promotes proliferation in gastric cancer.* Oncogene, 2016. **35**(27): p. 3524-34.
59. Wu, X.S., et al., *LncRNA-PAGBC acts as a microRNA sponge and promotes gallbladder tumorigenesis.* EMBO Rep, 2017. **18**(10): p. 1837-1853.
60. Gong, C. and L.E. Maquat, *lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements.* Nature, 2011. **470**(7333): p. 284-8.
61. Wang, J., C. Gong, and L.E. Maquat, *Control of myogenesis by rodent SINE-containing lncRNAs.* Genes Dev, 2013. **27**(7): p. 793-804.
62. Faghihi, M.A., et al., *Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase.* Nat Med, 2008. **14**(7): p. 723-30.
63. Tochitani, S. and Y. Hayashizaki, *Nkx2.2 antisense RNA overexpression enhanced oligodendrocytic differentiation.* Biochem Biophys Res Commun, 2008. **372**(4): p. 691-6.
64. Arun, G., et al., *mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells.* Mol Cell Biol, 2012. **32**(15): p. 3140-52.
65. Marchese, F.P., et al., *A Long Noncoding RNA Regulates Sister Chromatid Cohesion.* Mol Cell, 2016. **63**(3): p. 397-407.
66. Lee, S., et al., *Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins.* Cell, 2016. **164**(1-2): p. 69-80.
67. Tripathi, V., et al., *The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation.* Mol Cell, 2010. **39**(6): p. 925-38.
68. Sasaki, Y.T., et al., *MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles.* Proc Natl Acad Sci U S A, 2009. **106**(8): p. 2525-30.
69. Souquere, S., et al., *Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies.* Mol Biol Cell, 2010. **21**(22): p. 4020-7.
70. Hacisuleyman, E., et al., *Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre.* Nat Struct Mol Biol, 2014. **21**(2): p. 198-206.
71. Barutcu, A.R., et al., *A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus.* Nat Commun, 2018. **9**(1): p. 1444.
72. Pan, Y., et al., *The Emerging Roles of Long Noncoding RNA ROR (lincRNA-ROR) and its Possible Mechanisms in Human Cancers.* Cell Physiol Biochem, 2016. **40**(1-2): p. 219-229.
73. Morlando, M., M. Ballarino, and A. Fatica, *Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia.* Front Med (Lausanne), 2015. **2**: p. 23.
74. Oskar Marin-Bejar, M.H., *Long noncoding RNAs : from identification to functions and mechanisms*. 2015, DovePress. p. 257-274.
75. Hung, K.H., Y. Wang, and J.C. Zhao, *Regulation of Mammalian Gene Dosage by Long Noncoding RNAs.* Biomolecules, 2013. **3**(1): p. 124-42.
76. McHugh, C.A., et al., *The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.* Nature, 2015. **521**(7551): p. 232-6.
77. Monnier, P., et al., *H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1.* Proc Natl Acad Sci U S A, 2013. **110**(51): p. 20693-8.
78. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.
79. Wang, X., et al., *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription.* Nature, 2008. **454**(7200): p. 126-30.

80. Wan, G., et al., *Long non-coding RNA ANRIL (CDKN2B-AS) is induced by the ATM-E2F1 signaling pathway.* Cell Signal, 2013. **25**(5): p. 1086-95.
81. Wan, G., et al., *A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation.* Embo j, 2013. **32**(21): p. 2833-47.
82. Hung, T., et al., *Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters.* Nat Genet, 2011. **43**(7): p. 621-9.
83. Kotake, Y., et al., *Long Non-coding RNA, PANDA, Contributes to the Stabilization of p53 Tumor Suppressor Protein.* Anticancer Res, 2016. **36**(4): p. 1605-11.
84. Chaudhary, R., et al., *Prosurvival long noncoding RNA PINCR regulates a subset of p53 targets in human colorectal cancer cells by binding to Matrin 3.* Elife, 2017. **6**.
85. Rosic, S., F. Kohler, and S. Erhardt, *Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division.* J Cell Biol, 2014. **207**(3): p. 335-49.
86. Bouzinba-Segard, H., A. Guais, and C. Francastel, *Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8709-14.
87. Blower, M.D., *Centromeric Transcription Regulates Aurora-B Localization and Activation.* Cell Rep, 2016. **15**(8): p. 1624-33.
88. Wong, L.H., et al., *Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere.* Genome Res, 2007. **17**(8): p. 1146-60.
89. Krawczyk, M. and B.M. Emerson, *p50-associated COX-2 extragenic RNA (PACER) activates COX-2 gene expression by occluding repressive NF-κB complexes.* Elife, 2014. **3**: p. e01776.
90. Li, Z., et al., *The long noncoding RNA THRIL regulates TNFα expression through its interaction with hnRNPL.* Proc Natl Acad Sci U S A, 2014. **111**(3): p. 1002-7.
91. Zhang, Q., et al., *NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression.* MBio, 2013. **4**(1): p. e00596-12.
92. Imam, H., et al., *The lncRNA NRON modulates HIV-1 replication in a NFAT-dependent manner and is differentially regulated by early and late viral proteins.* Sci Rep, 2015. **5**: p. 8639.
93. Evans, J.R., F.Y. Feng, and A.M. Chinnaiyan, *The bright side of dark matter: lncRNAs in cancer.* J Clin Invest, 2016. **126**(8): p. 2775-82.
94. Prensner, J.R., et al., *The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex.* Nat Genet, 2013. **45**(11): p. 1392-8.
95. Kotake, Y., et al., *Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene.* Oncogene, 2011. **30**(16): p. 1956-62.
96. Yap, K.L., et al., *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a.* Mol Cell, 2010. **38**(5): p. 662-74.
97. Huarte, M., *The emerging role of lncRNAs in cancer.* Nat Med, 2015. **21**(11): p. 1253-61.
98. Pickard, M.R. and G.T. Williams, *The hormone response element mimic sequence of GAS5 lncRNA is sufficient to induce apoptosis in breast cancer cells.* Oncotarget, 2016. **7**(9): p. 10104-16.
99. Arun, G., et al., *Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss.* Genes Dev, 2016. **30**(1): p. 34-51.
100. Gibb, E.A., et al., *Human cancer long non-coding RNA transcriptomes.* PLoS One, 2011. **6**(10): p. e25915.
101. Loewer, S., et al., *Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells.* Nat Genet, 2010. **42**(12): p. 1113-7.
102. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.* Genes Dev, 2011. **25**(18): p. 1915-27.
103. Fort, A., et al., *Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance.* Nat Genet, 2014. **46**(6): p. 558-66.
104. Ohnishi, Y., et al., *Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages.* Nat Cell Biol, 2014. **16**(1): p. 27-37.
105. Xu, C., et al., *Long non-coding RNA GAS5 controls human embryonic stem cell self-renewal by maintaining NODAL signalling.* Nat Commun, 2016. **7**: p. 13287.

106. Seila, A.C., et al., *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.
107. Mattick, J.S., *The genetic signatures of noncoding RNAs.* PLoS Genet, 2009. **5**(4): p. e1000459.
108. Chen, L.L. and G.G. Carmichael, *Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA.* Mol Cell, 2009. **35**(4): p. 467-78.
109. Redrup, L., et al., *The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing.* Development, 2009. **136**(4): p. 525-30.
110. Covarrubias, S., et al., *CRISPR/Cas-based screening of long non-coding RNAs (lncRNAs) in macrophages with an NF-kappaB reporter.* J Biol Chem, 2017. **292**(51): p. 20911-20920.
111. Jin, C., et al., *Long non-coding RNA MIAT knockdown promotes osteogenic differentiation of human adipose-derived stem cells.* Cell Biol Int, 2017. **41**(1): p. 33-41.
112. Joung, J., et al., *Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood.* Nature, 2017. **548**(7667): p. 343-346.
113. Guil, S., et al., *Intronic RNAs mediate EZH2 regulation of epigenetic targets.* Nat Struct Mol Biol, 2012. **19**(7): p. 664-70.
114. Li, Z., et al., *Exon-intron circular RNAs regulate transcription in the nucleus.* Nat Struct Mol Biol, 2015. **22**(3): p. 256-64.
115. Kaneko, S., et al., *PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells.* Nat Struct Mol Biol, 2013. **20**(11): p. 1258-64.
116. Chu, C., et al., *Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions.* Mol Cell, 2011. **44**(4): p. 667-78.
117. Akhade, V.S., et al., *Genome wide chromatin occupancy of mrhl RNA and its role in gene regulation in mouse spermatogonial cells.* RNA Biol, 2014. **11**(10): p. 1262-79.
118. Chu, C., et al., *Systematic discovery of Xist RNA binding proteins.* Cell, 2015. **161**(2): p. 404-16.
119. Engreitz, J.M., et al., *The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome.* Science, 2013. **341**(6147): p. 1237973.
120. Simon, M.D., et al., *The genomic binding sites of a noncoding RNA.* Proc Natl Acad Sci U S A, 2011. **108**(51): p. 20497-502.
121. Simon, M.D., et al., *High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation.* Nature, 2013. **504**(7480): p. 465-469.
122. Engreitz, J.M., et al., *RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites.* Cell, 2014. **159**(1): p. 188-199.
123. Kudla, G., et al., *Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast.* Proc Natl Acad Sci U S A, 2011. **108**(24): p. 10010-5.
124. Helwak, A., et al., *Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding.* Cell, 2013. **153**(3): p. 654-65.
125. Watts, J.M., et al., *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-6.
126. Underwood, J.G., et al., *FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing.* Nat Methods, 2010. **7**(12): p. 995-1001.
127. Kertesz, M., et al., *Genome-wide measurement of RNA secondary structure in yeast.* Nature, 2010. **467**(7311): p. 103-7.
128. Wan, Y., et al., *Landscape and variation of RNA secondary structure across the human transcriptome.* Nature, 2014. **505**(7485): p. 706-9.
129. Guttman, M., et al., *Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins.* Cell, 2013. **154**(1): p. 240-51.
130. Carrieri, C., et al., *Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat.* Nature, 2012. **491**(7424): p. 454-7.
131. Hupe, M., et al., *Evaluation of TRAP-sequencing technology with a versatile conditional mouse model.* Nucleic Acids Res, 2014. **42**(2): p. e14.
132. Kashi, K., et al., *Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome.* Biochim Biophys Acta, 2016. **1859**(1): p. 3-15.

133. Kopp, F. and J.T. Mendell, *Functional Classification and Experimental Dissection of Long Noncoding RNAs.* Cell, 2018. **172**(3): p. 393-407.
134. Nishant, K.T., H. Ravishankar, and M.R. Rao, *Characterization of a mouse recombination hot spot locus encoding a novel non-protein-coding RNA.* Mol Cell Biol, 2004. **24**(12): p. 5620-34.
135. Zerbino, D.R., et al., *Ensembl 2018.* Nucleic Acids Res, 2018. **46**(D1): p. D754-d761.
136. Ganesan, G. and S.M. Rao, *A novel noncoding RNA processed by Drosha is restricted to nucleus in mouse.* Rna, 2008. **14**(7): p. 1399-410.
137. Kataruka, S., et al., *Mrhl Long Noncoding RNA Mediates Meiotic Commitment of Mouse Spermatogonial Cells by Regulating Sox8 Expression.* Mol Cell Biol, 2017. **37**(14).
138. Akhade, V.S., et al., *Mechanism of Wnt signaling induced down regulation of mrhl long non-coding RNA in mouse spermatogonial cells.* Nucleic Acids Res, 2016. **44**(1): p. 387-401.
139. Bibel, M., et al., *Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells.* Nat Protoc, 2007. **2**(5): p. 1034-43.
140. Pijnappel, W.W.M.P., M.P.A. Baltissen, and H.T.M. Timmers, *Protocol for lentiviral knock down in mouse ES cells.* 2013.
141. de Planell-Saguer, M., M.C. Rodicio, and Z. Mourelatos, *Rapid in situ codetection of noncoding RNAs and proteins in cells and formalin-fixed paraffin-embedded tissue sections without protease treatment.* Nat Protoc, 2010. **5**(6): p. 1061-73.
142. Cotney, J.L. and J.P. Noonan, *Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing.* Cold Spring Harb Protoc, 2015. **2015**(2): p. 191-9.
143. Uchil, P.D., A. Nagarajan, and P. Kumar, *Assay for beta-Galactosidase in Extracts of Mammalian Cells.* Cold Spring Harb Protoc, 2017. **2017**(10): p. pdb.prot095778.
144. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nat Protoc, 2012. **7**(3): p. 562-78.
145. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function.* Genome Res, 2003. **13**(9): p. 2129-41.
146. de Hoon, M.J., et al., *Open source clustering software.* Bioinformatics, 2004. **20**(9): p. 1453-4.
147. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
148. Warde-Farley, D., et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W214-20.
149. Mathelier, A., et al., *JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles.* Nucleic Acids Res, 2014. **42**(Database issue): p. D142-7.
150. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.
151. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-8.
152. Jensen, L.J., et al., *STRING 8--a global view on proteins and their functional interactions in 630 organisms.* Nucleic Acids Res, 2009. **37**(Database issue): p. D412-6.
153. Xie, Q. and A. Cvekl, *The orchestration of mammalian tissue morphogenesis through a series of coherent feed-forward loops.* J Biol Chem, 2011. **286**(50): p. 43259-71.
154. Grant, C.E., T.L. Bailey, and W.S. Noble, *FIMO: scanning for occurrences of a given motif.* Bioinformatics, 2011. **27**(7): p. 1017-8.
155. Takaoka, K. and H. Hamada, *Cell fate decisions and axis determination in the early mouse embryo.* Development, 2012. **139**(1): p. 3-14.
156. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.
157. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. **78**(12): p. 7634-8.

158. Huang, G., et al., *Molecular basis of embryonic stem cell self-renewal: from signaling pathways to pluripotency network.* Cell Mol Life Sci, 2015. **72**(9): p. 1741-57.
159. Thomson, J.A., et al., *Isolation of a primate embryonic stem cell line.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7844-8.
160. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
161. Okita, K. and S. Yamanaka, *Intracellular signaling pathways regulating pluripotency of embryonic stem cells.* Curr Stem Cell Res Ther, 2006. **1**(1): p. 103-11.
162. Chappell, J. and S. Dalton, *Roles for MYC in the establishment and maintenance of pluripotency.* Cold Spring Harb Perspect Med, 2013. **3**(12): p. a014381.
163. Varlakhanova, N.V., et al., *myc maintains embryonic stem cell pluripotency and self-renewal.* Differentiation, 2010. **80**(1): p. 9-19.
164. Scognamiglio, R., et al., *Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause.* Cell, 2016. **164**(4): p. 668-80.
165. Zhang, P., et al., *Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression.* J Biol Chem, 2010. **285**(12): p. 9180-9.
166. Niwa, H., et al., *Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3.* Genes Dev, 1998. **12**(13): p. 2048-60.
167. Matsuda, T., et al., *STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells.* Embo j, 1999. **18**(15): p. 4261-9.
168. Ying, Q.L., et al., *The ground state of embryonic stem cell self-renewal.* Nature, 2008. **453**(7194): p. 519-23.
169. Chen, Y., K. Blair, and A. Smith, *Robust self-renewal of rat embryonic stem cells requires fine-tuning of glycogen synthase kinase-3 inhibition.* Stem Cell Reports, 2013. **1**(3): p. 209-17.
170. Yang, S.H., et al., *A genome-wide RNAi screen reveals MAP kinase phosphatases as key ERK pathway regulators during embryonic stem cell differentiation.* PLoS Genet, 2012. **8**(12): p. e1003112.
171. Huang, G., et al., *STAT3 phosphorylation at tyrosine 705 and serine 727 differentially regulates mouse ESC fates.* Stem Cells, 2014. **32**(5): p. 1149-60.
172. Jirmanova, L., et al., *Differential contributions of ERK and PI3-kinase to the regulation of cyclin D1 expression and to the control of the G1/S transition in mouse embryonic stem cells.* Oncogene, 2002. **21**(36): p. 5515-28.
173. Sun, H., et al., *PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway.* Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6199-204.
174. Keller, G., *Embryonic stem cell differentiation: emergence of a new era in biology and medicine.* Genes Dev, 2005. **19**(10): p. 1129-55.
175. Kim, M., et al., *Regulation of mouse embryonic stem cell neural differentiation by retinoic acid.* Dev Biol, 2009. **328**(2): p. 456-71.
176. Wongpaiboonwattana, W. and M.P. Stavridis, *Neural differentiation of mouse embryonic stem cells in serum-free monolayer culture.* J Vis Exp, 2015(99): p. e52823.
177. Jing, Y., et al., *In vitro differentiation of mouse embryonic stem cells into neurons of the dorsal forebrain.* Cell Mol Neurobiol, 2011. **31**(5): p. 715-27.
178. Tischfield, D.J. and S.A. Anderson, *Differentiation of Mouse Embryonic Stem Cells into Cortical Interneuron Precursors.* J Vis Exp, 2017(130).
179. Wichterle, H., et al., *Directed differentiation of embryonic stem cells into motor neurons.* Cell, 2002. **110**(3): p. 385-97.
180. Kawasaki, H., et al., *Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity.* Neuron, 2000. **28**(1): p. 31-40.
181. Murry, C.E. and G. Keller, *Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development.* Cell, 2008. **132**(4): p. 661-80.
182. Coraux, C., et al., *Reconstituted skin from murine embryonic stem cells.* Curr Biol, 2003. **13**(10): p. 849-53.

183. Kubo, A., et al., *Development of definitive endoderm from embryonic stem cells in culture.* Development, 2004. **131**(7): p. 1651-62.
184. Tano, K., et al., *MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes.* FEBS Lett, 2010. **584**(22): p. 4575-80.
185. Choi, D., et al., *In vitro differentiation of mouse embryonic stem cells: enrichment of endodermal cells in the embryoid body.* Stem Cells, 2005. **23**(6): p. 817-27.
186. Ninomiya, N., et al., *BMP signaling regulates the differentiation of mouse embryonic stem cells into lung epithelial cell lineages.* In Vitro Cell Dev Biol Anim, 2013. **49**(3): p. 230-7.
187. Olsen, A.L., D.L. Stachura, and M.J. Weiss, *Designer blood: creating hematopoietic lineages from embryonic stem cells.* Blood, 2006. **107**(4): p. 1265-75.
188. Fauzi, I., N. Panoskaltsis, and A. Mantalaris, *In Vitro Differentiation of Embryonic Stem Cells into Hematopoietic Lineage: Towards Erythroid Progenitor's Production.* Methods Mol Biol, 2016. **1341**: p. 217-34.
189. Guo, X.M., et al., *Creation of engineered cardiac tissue in vitro from mouse embryonic stem cells.* Circulation, 2006. **113**(18): p. 2229-37.
190. Kattman, S.J., et al., *Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines.* Cell Stem Cell, 2011. **8**(2): p. 228-40.
191. Matsuura, K., et al., *Creation of mouse embryonic stem cell-derived cardiac cell sheets.* Biomaterials, 2011. **32**(30): p. 7355-62.
192. Toyooka, Y., et al., *Embryonic stem cells can form germ cells in vitro.* Proc Natl Acad Sci U S A, 2003. **100**(20): p. 11457-62.
193. Kimura, T., et al., *Induction of primordial germ cell-like cells from mouse embryonic stem cells by ERK signal inhibition.* Stem Cells, 2014. **32**(10): p. 2668-78.
194. Ge, W., et al., *In vitro differentiation of germ cells from stem cells: a comparison between primordial germ cells and in vitro derived primordial germ cell-like cells.* Cell Death Dis, 2015. **6**(10): p. e1906.
195. Hayashi, K., et al., *Offspring from oocytes derived from in vitro primordial germ cell-like cells in mice.* Science, 2012. **338**(6109): p. 971-5.
196. Hayashi, K., et al., *Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells.* Cell, 2011. **146**(4): p. 519-32.
197. Guttman, M., et al., *lincRNAs act in the circuitry controlling pluripotency and differentiation.* Nature, 2011. **477**(7364): p. 295-300.
198. Dinger, M.E., et al., *Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation.* Genome Res, 2008. **18**(9): p. 1433-45.
199. Sheik Mohamed, J., et al., *Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells.* Rna, 2010. **16**(2): p. 324-37.
200. Mercer, T.R., et al., *Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation.* BMC Neurosci, 2010. **11**: p. 14.
201. Lin, N., et al., *An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment.* Mol Cell, 2014. **53**(6): p. 1005-19.
202. Wang, Y., et al., *Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal.* Dev Cell, 2013. **25**(1): p. 69-80.
203. Vance, K.W., et al., *The long non-coding RNA Paupar regulates the expression of both local and distal genes.* Embo j, 2014. **33**(4): p. 296-311.
204. Ramos, A.D., et al., *The long noncoding RNA Pnky regulates neuronal differentiation of embryonic and postnatal neural stem cells.* Cell Stem Cell, 2015. **16**(4): p. 439-447.
205. Chalei, V., et al., *The long non-coding RNA Dali is an epigenetic regulator of neural differentiation.* Elife, 2014. **3**: p. e04530.
206. Hu, W., et al., *Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation.* Genes Dev, 2011. **25**(24): p. 2573-8.
207. Wagner, L.A., et al., *EGO, a novel, noncoding RNA gene, regulates eosinophil granule protein transcript expression.* Blood, 2007. **109**(12): p. 5191-8.

208. Zhang, X., et al., *A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster.* Blood, 2009. **113**(11): p. 2526-34.
209. Li, L., et al., *A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells.* Cell Death Dis, 2016. **7**(3): p. e2140.
210. Mueller, A.C., et al., *MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis.* Mol Cell Biol, 2015. **35**(3): p. 498-513.
211. Zhu, M., et al., *Lnc-mg is a long non-coding RNA that promotes myogenesis.* Nat Commun, 2017. **8**: p. 14718.
212. Kretz, M., et al., *Suppression of progenitor differentiation requires the long noncoding RNA ANCR.* Genes Dev, 2012. **26**(4): p. 338-43.
213. Kretz, M., et al., *Control of somatic tissue differentiation by the long non-coding RNA TINCR.* Nature, 2013. **493**(7431): p. 231-5.
214. Anguera, M.C., et al., *Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain.* PLoS Genet, 2011. **7**(9): p. e1002248.
215. Wang, J., T. Sinha, and A. Wynshaw-Boris, *Wnt signaling in mammalian development: lessons from mouse genetics.* Cold Spring Harb Perspect Biol, 2012. **4**(5).
216. Muñoz-Descalzo, S., A.K. Hadjantonakis, and A.M. Arias, *Wnt/ß-catenin signalling and the dynamics of fate decisions in early mouse embryos and embryonic stem (ES) cells.* Semin Cell Dev Biol, 2015. **47-48**: p. 101-9.
217. Steinhart, Z. and S. Angers, *Wnt signaling in development and tissue homeostasis.* Development, 2018. **145**(11).
218. Desbaillets, I., et al., *Embryoid bodies: an in vitro model of mouse embryogenesis.* Exp Physiol, 2000. **85**(6): p. 645-51.
219. Koike, M., et al., *Characterization of embryoid bodies of mouse embryonic stem cells formed under various culture conditions and estimation of differentiation status of such bodies.* J Biosci Bioeng, 2007. **104**(4): p. 294-9.
220. Leahy, A., et al., *Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation.* J Exp Zool, 1999. **284**(1): p. 67-81.
221. van Dam, S., et al., *Gene co-expression analysis for functional classification and gene-disease predictions.* Brief Bioinform, 2018. **19**(4): p. 575-592.
222. Seyfried, N.T., et al., *A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease.* Cell Syst, 2017. **4**(1): p. 60-72.e4.
223. Lai, H.C., et al., *In vivo neuronal subtype-specific targets of Atoh1 (Math1) in dorsal spinal cord.* J Neurosci, 2011. **31**(30): p. 10859-71.
224. Mulvaney, J. and A. Dabdoub, *Atoh1, an essential transcription factor in neurogenesis and intestinal and inner ear development: function, regulation, and context dependency.* J Assoc Res Otolaryngol, 2012. **13**(3): p. 281-93.
225. Wapinski, O.L., et al., *Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons.* Cell, 2013. **155**(3): p. 621-35.
226. Graham, V., et al., *SOX2 functions to maintain neural progenitor identity.* Neuron, 2003. **39**(5): p. 749-65.
227. Inoue, T., et al., *Zic1 and Zic3 regulate medial forebrain development through expansion of neuronal progenitors.* J Neurosci, 2007. **27**(20): p. 5461-73.
228. Urban, S., et al., *A Brn2-Zic1 axis specifies the neuronal fate of retinoic-acid-treated embryonic stem cells.* J Cell Sci, 2015. **128**(13): p. 2303-18.
229. Sankar, S., et al., *Gene regulatory networks in neural cell fate acquisition from genome-wide chromatin association of Geminin and Zic1.* Sci Rep, 2016. **6**: p. 37412.
230. Kuwahara, A., et al., *Tcf3 represses Wnt-β-catenin signaling and maintains neural stem cell population during neocortical development.* PLoS One, 2014. **9**(5): p. e94408.
231. Mesman, S. and M.P. Smidt, *Tcf12 Is Involved in Early Cell-Fate Determination and Subset Specification of Midbrain Dopamine Neurons.* Front Mol Neurosci, 2017. **10**: p. 353.
232. Uittenbogaard, M. and A. Chiaramello, *Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis.* Brain Res Gene Expr Patterns, 2002. **1**(2): p. 115-21.

233. Curto, G.G., et al., *Pax6 is essential for the maintenance and multi-lineage differentiation of neural stem cells, and for neuronal incorporation into the adult olfactory bulb.* Stem Cells Dev, 2014. **23**(23): p. 2813-30.
234. Georgala, P.A., C.B. Carr, and D.J. Price, *The role of Pax6 in forebrain development.* Dev Neurobiol, 2011. **71**(8): p. 690-709.
235. Bedogni, F., et al., *Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex.* Proc Natl Acad Sci U S A, 2010. **107**(29): p. 13129-34.
236. McKenna, W.L., et al., *Tbr1 and Fezf2 regulate alternate corticofugal neuronal identities during neocortical development.* J Neurosci, 2011. **31**(2): p. 549-64.
237. Milet, C., et al., *Pax3 and Zic1 drive induction and differentiation of multipotent, migratory, and functional neural crest in Xenopus embryos.* Proc Natl Acad Sci U S A, 2013. **110**(14): p. 5528-33.
238. Sato, T., N. Sasai, and Y. Sasai, *Neural crest determination by co-activation of Pax3 and Zic1 genes in Xenopus ectoderm.* Development, 2005. **132**(10): p. 2355-63.
239. Ferri, A.L., et al., *Foxa1 and Foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner.* Development, 2007. **134**(15): p. 2761-9.
240. Nefzger, C.M., J.M. Haynes, and C.W. Pouton, *Directed expression of Gata2, Mash1, and Foxa2 synergize to induce the serotonergic neuron phenotype during in vitro differentiation of embryonic stem cells.* Stem Cells, 2011. **29**(6): p. 928-39.
241. Pristerà, A., et al., *Transcription factors FOXA1 and FOXA2 maintain dopaminergic neuronal properties and control feeding behavior in adult mice.* Proc Natl Acad Sci U S A, 2015. **112**(35): p. E4929-38.
242. Schreiber, J., et al., *Redundancy of class III POU proteins in the oligodendrocyte lineage.* J Biol Chem, 1997. **272**(51): p. 32286-93.
243. Wilkinson, A.C., H. Nakauchi, and B. Göttgens, *Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity.* Cell Syst, 2017. **5**(4): p. 319-331.
244. Blais, A. and B.D. Dynlacht, *Constructing transcriptional regulatory networks.* Genes Dev, 2005. **19**(13): p. 1499-511.
245. Dunn, S.J., et al., *Defining an essential transcription factor program for naïve pluripotency.* Science, 2014. **344**(6188): p. 1156-1160.
246. Goode, D.K., et al., *Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation.* Dev Cell, 2016. **36**(5): p. 572-87.
247. Komori, T., *Roles of Runx2 in Skeletal Development.* Adv Exp Med Biol, 2017. **962**: p. 83-93.
248. Shah, A.V., G.M. Birdsey, and A.M. Randi, *Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG.* Vascul Pharmacol, 2016. **86**: p. 3-13.
249. Jalali, A., et al., *HeyL promotes neuronal differentiation of neural progenitor cells.* J Neurosci Res, 2011. **89**(3): p. 299-309.
250. Bhatlekar, S., J.Z. Fields, and B.M. Boman, *Role of HOX Genes in Stem Cell Differentiation and Cancer.* Stem Cells Int, 2018. **2018**: p. 3569493.
251. Seifert, A., et al., *Role of Hox genes in stem cell differentiation.* World J Stem Cells, 2015. **7**(3): p. 583-95.
252. van de Willige, D., C.C. Hoogenraad, and A. Akhmanova, *Microtubule plus-end tracking proteins in neuronal development.* Cell Mol Life Sci, 2016. **73**(10): p. 2053-77.
253. Huang, W.H., et al., *Atoh1 governs the migration of postmitotic neurons that shape respiratory effectiveness at birth and chemoresponsiveness in adulthood.* Neuron, 2012. **75**(5): p. 799-809.
254. Englund, C., et al., *Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex.* J Neurosci, 2005. **25**(1): p. 247-51.
255. Hevner, R.F., et al., *Tbr1 regulates differentiation of the preplate and layer 6.* Neuron, 2001. **29**(2): p. 353-66.
256. Cooper, J.A., *Molecules and mechanisms that regulate multipolar migration in the intermediate zone.* Front Cell Neurosci, 2014. **8**: p. 386.

257. Takemoto, M., et al., *Laminar and areal expression of unc5d and its role in cortical cell survival.* Cereb Cortex, 2011. **21**(8): p. 1925-34.
258. Koszinowski, S., et al., *RARβ regulates neuronal cell death and differentiation in the avian ciliary ganglion.* Dev Neurobiol, 2015. **75**(11): p. 1204-18.
259. Rataj-Baniowska, M., et al., *Retinoic Acid Receptor beta Controls Development of Striatonigral Projection Neurons through FGF-Dependent and Meis1-Dependent Mechanisms.* J Neurosci, 2015. **35**(43): p. 14467-75.
260. Inoue, J., et al., *The expression of LIM-homeobox genes, Lhx1 and Lhx5, in the forebrain is essential for neural retina differentiation.* Dev Growth Differ, 2013. **55**(7): p. 668-75.
261. Palmesino, E., et al., *Foxp1 and lhx1 coordinate motor neuron migration with axon trajectory choice by gating Reelin signalling.* PLoS Biol, 2010. **8**(8): p. e1000446.
262. Pillai, A., et al., *Lhx1 and Lhx5 maintain the inhibitory-neurotransmitter status of interneurons in the dorsal spinal cord.* Development, 2007. **134**(2): p. 357-66.
263. Zhao, Y., et al., *LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum.* Proc Natl Acad Sci U S A, 2007. **104**(32): p. 13182-6.
264. Aoki, M., T. Yamashita, and M. Tohyama, *EphA receptors direct the differentiation of mammalian neural precursor cells through a mitogen-activated protein kinase-dependent pathway.* J Biol Chem, 2004. **279**(31): p. 32643-50.
265. Drescher, U., *The Eph family in the patterning of neural development.* Curr Biol, 1997. **7**(12): p. R799-807.
266. Gao, P.P., et al., *Ephrin-dependent growth and pruning of hippocampal axons.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 4073-7.
267. Yue, Y., et al., *Mistargeting hippocampal axons by expression of a truncated Eph receptor.* Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10777-82.
268. Rasmussen, A.H., H.B. Rasmussen, and A. Silahtaroglu, *The DLGAP family: neuronal expression, function and role in brain disorders.* Mol Brain, 2017. **10**(1): p. 43.
269. Chen, D., et al., *Rb-mediated neuronal differentiation through cell-cycle-independent regulation of E2f3a.* PLoS Biol, 2007. **5**(7): p. e179.
270. Ferguson, K.L., et al., *The Rb-CDK4/6 signaling pathway is critical in neural precursor cell cycle regulation.* J Biol Chem, 2000. **275**(43): p. 33593-600.
271. Matsui, T., et al., *Retinoblastoma protein controls growth, survival and neuronal migration in human cerebral organoids.* Development, 2017. **144**(6): p. 1025-1034.
272. Sigulinsky, C.L., et al., *Vsx2/Chx10 ensures the correct timing and magnitude of Hedgehog signaling in the mouse retina.* Dev Biol, 2008. **317**(2): p. 560-75.
273. Vitorino, M., et al., *Vsx2 in the zebrafish retina: restricted lineages through derepression.* Neural Dev, 2009. **4**: p. 14.
274. Zou, C. and E.M. Levine, *Vsx2 controls eye organogenesis and retinal progenitor identity via homeodomain and non-homeodomain residues required for high affinity DNA binding.* PLoS Genet, 2012. **8**(9): p. e1002924.
275. Wapinski, O.L., et al., *Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons.* Cell Rep, 2017. **20**(13): p. 3236-3247.
276. Zhu, H. and A.J. Bendall, *Dlx3 is expressed in the ventral forebrain of chicken embryos: implications for the evolution of the Dlx gene family.* Int J Dev Biol, 2006. **50**(1): p. 71-5.
277. Avilés, E.C. and L.V. Goodrich, *Configuring a robust nervous system with Fat cadherins.* Semin Cell Dev Biol, 2017. **69**: p. 91-101.
278. Vierbuchen, T., et al., *Direct conversion of fibroblasts to functional neurons by defined factors.* Nature, 2010. **463**(7284): p. 1035-41.
279. Yao, B. and P. Jin, *Unlocking epigenetic codes in neurogenesis.* Genes Dev, 2014. **28**(12): p. 1253-71.
280. Florio, M. and W.B. Huttner, *Neural progenitors, neurogenesis and the evolution of the neocortex.* Development, 2014. **141**(11): p. 2182-94.
281. Estivill-Torrus, G., et al., *Pax6 is required to regulate the cell cycle and the rate of progression from symmetrical to asymmetrical division in mammalian cortical progenitors.* Development, 2002. **129**(2): p. 455-66.

282. Marquardt, T., et al., *Pax6 is required for the multipotent state of retinal progenitor cells.* Cell, 2001. **105**(1): p. 43-55.
283. Quinn, J.C., et al., *Pax6 controls cerebral cortical cell number by regulating exit from the cell cycle and specifies cortical cell identity by a cell autonomous mechanism.* Dev Biol, 2007. **302**(1): p. 50-65.
284. Stoykova, A., et al., *Forebrain patterning defects in Small eye mutant mice.* Development, 1996. **122**(11): p. 3453-65.
285. Stoykova, A., et al., *Pax6 modulates the dorsoventral patterning of the mammalian telencephalon.* J Neurosci, 2000. **20**(21): p. 8042-50.
286. Yun, K., S. Potter, and J.L. Rubenstein, *Gsh2 and Pax6 play complementary roles in dorsoventral patterning of the mammalian telencephalon.* Development, 2001. **128**(2): p. 193-205.
287. Sansom, S.N., et al., *The level of the transcription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis.* PLoS Genet, 2009. **5**(6): p. e1000511.
288. Thakurela, S., et al., *Mapping gene regulatory circuitry of Pax6 during neurogenesis.* Cell Discov, 2016. **2**: p. 15045.
289. Martynoga, B., D. Drechsel, and F. Guillemot, *Molecular control of neurogenesis: a view from the mammalian cerebral cortex.* Cold Spring Harb Perspect Biol, 2012. **4**(10).
290. Nieto, M., et al., *Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors.* Neuron, 2001. **29**(2): p. 401-13.
291. Mizuguchi, R., et al., *Combinatorial roles of olig2 and neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons.* Neuron, 2001. **31**(5): p. 757-71.
292. Nakada, Y., et al., *Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification.* Development, 2004. **131**(6): p. 1319-30.
293. Chanda, S., et al., *Generation of induced neuronal cells by the single reprogramming factor ASCL1.* Stem Cell Reports, 2014. **3**(2): p. 282-96.
294. Castro, D.S., et al., *A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets.* Genes Dev, 2011. **25**(9): p. 930-45.
295. Arnold, S.J., et al., *The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone.* Genes Dev, 2008. **22**(18): p. 2479-84.
296. Sessa, A., et al., *Tbr2 directs conversion of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis in the developing neocortex.* Neuron, 2008. **60**(1): p. 56-69.
297. Sessa, A., et al., *The Tbr2 Molecular Network Controls Cortical Neuronal Differentiation Through Complementary Genetic and Epigenetic Pathways.* Cereb Cortex, 2017. **27**(6): p. 3378-3396.
298. Pataskar, A., et al., *NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program.* Embo j, 2016. **35**(1): p. 24-45.
299. Gaiano, N., J.S. Nye, and G. Fishell, *Radial glial identity is promoted by Notch1 signaling in the murine forebrain.* Neuron, 2000. **26**(2): p. 395-404.
300. Mizutani, K., et al., *Differential Notch signalling distinguishes neural stem cells from intermediate progenitors.* Nature, 2007. **449**(7160): p. 351-5.
301. Imayoshi, I., et al., *Essential roles of Notch signaling in maintenance of neural stem cells in developing and adult brains.* J Neurosci, 2010. **30**(9): p. 3489-98.
302. Sakamoto, M., et al., *The basic helix-loop-helix genes Hesr1/Hey1 and Hesr2/Hey2 regulate maintenance of neural precursor cells in the brain.* J Biol Chem, 2003. **278**(45): p. 44808-15.
303. Zhou, Z.D., et al., *Notch as a molecular switch in neural stem cells.* IUBMB Life, 2010. **62**(8): p. 618-23.
304. Yoon, K., et al., *Fibroblast growth factor receptor signaling promotes radial glial identity and interacts with Notch1 signaling in telencephalic progenitors.* J Neurosci, 2004. **24**(43): p. 9497-506.

305. Sahara, S. and D.D. O'Leary, *Fgf10 regulates transition period of cortical stem cell differentiation to radial glia controlling generation of neurons and basal progenitors.* Neuron, 2009. **63**(1): p. 48-62.
306. Kang, W., et al., *The transition from radial glial to intermediate progenitor cell is inhibited by FGF signaling during corticogenesis.* J Neurosci, 2009. **29**(46): p. 14571-80.
307. Calegari, F., et al., *Selective lengthening of the cell cycle in the neurogenic subpopulation of neural progenitor cells during mouse brain development.* J Neurosci, 2005. **25**(28): p. 6533-8.
308. Lukaszewicz, A., et al., *Contrasting effects of basic fibroblast growth factor and neurotrophin 3 on cell cycle kinetics of mouse cortical stem cells.* J Neurosci, 2002. **22**(15): p. 6610-22.
309. Chenn, A. and C.A. Walsh, *Regulation of cerebral cortical size by control of cell cycle exit in neural precursors.* Science, 2002. **297**(5580): p. 365-9.
310. Machon, O., et al., *Role of beta-catenin in the developing cortical and hippocampal neuroepithelium.* Neuroscience, 2003. **122**(1): p. 129-43.
311. Woodhead, G.J., et al., *Cell-autonomous beta-catenin signaling regulates cortical precursor proliferation.* J Neurosci, 2006. **26**(48): p. 12620-30.
312. Zhou, C.J., et al., *Neuronal production and precursor proliferation defects in the neocortex of mice with loss of function in the canonical Wnt signaling pathway.* Neuroscience, 2006. **142**(4): p. 1119-31.
313. Hirabayashi, Y., et al., *The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells.* Development, 2004. **131**(12): p. 2791-801.
314. Rhinn, M. and P. Dolle, *Retinoic acid signalling during development.* Development, 2012. **139**(5): p. 843-58.
315. Siegenthaler, J.A., et al., *Retinoic acid from the meninges regulates cortical neuron generation.* Cell, 2009. **139**(3): p. 597-609.
316. Engberg, N., et al., *Retinoic acid synthesis promotes development of neural progenitors from mouse embryonic stem cells by suppressing endogenous, Wnt-dependent nodal signaling.* Stem Cells, 2010. **28**(9): p. 1498-509.
317. Cunningham, T.J., et al., *Retinoic Acid Activity in Undifferentiated Neural Progenitors Is Sufficient to Fulfill Its Role in Restricting Fgf8 Expression for Somitogenesis.* PLoS One, 2015. **10**(9): p. e0137894.
318. Briggs, J.A., et al., *Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution.* Neuron, 2015. **88**(5): p. 861-877.
319. Ng, S.Y., et al., *The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis.* Mol Cell, 2013. **51**(3): p. 349-59.
320. Bond, A.M., et al., *Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry.* Nat Neurosci, 2009. **12**(8): p. 1020-7.
321. Sauvageau, M., et al., *Multiple knockout mouse models reveal lincRNAs are required for life and brain development.* Elife, 2013. **2**: p. e01749.
322. Modarresi, F., et al., *Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation.* Nat Biotechnol, 2012. **30**(5): p. 453-9.
323. Kerin, T., et al., *A noncoding RNA antisense to moesin at 5p14.1 in autism.* Sci Transl Med, 2012. **4**(128): p. 128ra40.
324. Barry, G., *Integrating the roles of long and small non-coding RNA in brain function and disease.* Mol Psychiatry, 2014. **19**(4): p. 410-6.
325. Sugitani, Y., et al., *Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons.* Genes Dev, 2002. **16**(14): p. 1760-5.
326. Kirkeby, A., et al., *Generation of regionally specified neural progenitors and functional neurons from human embryonic stem cells under defined conditions.* Cell Rep, 2012. **1**(6): p. 703-14.
327. Su, Z., et al., *Antagonism between the transcription factors NANOG and OTX2 specifies rostral or caudal cell fate during neural patterning transition.* J Biol Chem, 2018. **293**(12): p. 4445-4455.
328. Osumi B., K.T., *The role of transcription factor Pax6 in brain development and evolution. evidence and hypothesis.* Springer, Tokyo, 2013: p. 43-61

329. Liu, J., et al., *Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development.* Cell Rep, 2017. **20**(6): p. 1307-1318.
330. Plachta, N., et al., *Developmental potential of defined neural progenitors derived from mouse embryonic stem cells.* Development, 2004. **131**(21): p. 5449-56.
331. Braconi, C., et al., *microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer.* Oncogene, 2011. **30**(47): p. 4750-6.
332. Cawley, S., et al., *Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs.* Cell, 2004. **116**(4): p. 499-509.
333. Perry, R.B. and I. Ulitsky, *The functions of long noncoding RNAs in development and stem cells.* Development, 2016. **143**(21): p. 3882-3894.
334. Lee, T.Y., et al., *GPMiner: an integrated system for mining combinatorial cis-regulatory elements in mammalian gene group.* BMC Genomics, 2012. **13 Suppl 1**(Suppl 1): p. S3.
335. Mazzone, A., et al., *Identification and characterization of a novel promoter for the human ANO1 gene regulated by the transcription factor signal transducer and activator of transcription 6 (STAT6).* Faseb j, 2015. **29**(1): p. 152-63.
336. Barber, B.A., et al., *Dynamic expression of MEIS1 homeoprotein in E14.5 forebrain and differentiated forebrain-derived neural stem cells.* Ann Anat, 2013. **195**(5): p. 431-40.
337. Nguyen, T. and S. Di Giovanni, *NFAT signaling in neural development and axon growth.* Int J Dev Neurosci, 2008. **26**(2): p. 141-5.
338. Moutier, E., et al., *Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology.* J Biol Chem, 2012. **287**(31): p. 26328-41.
339. Sun, J., et al., *Identification of in vivo DNA-binding mechanisms of Pax6 and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development.* Nucleic Acids Res, 2015. **43**(14): p. 6827-46.
340. Xie, Q., et al., *Pax6 interactions with chromatin and identification of its novel direct target genes in lens and forebrain.* PLoS One, 2013. **8**(1): p. e54507.
341. Pinson, J., et al., *Regulation of the Pax6 : Pax6(5a) mRNA ratio in the developing mammalian brain.* BMC Dev Biol, 2005. **5**: p. 13.
342. Kamachi, Y., et al., *Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development.* Genes Dev, 2001. **15**(10): p. 1272-86.
343. Sun, J., et al., *Pax6 associates with H3K4-specific histone methyltransferases Mll1, Mll2, and Set1a and regulates H3K4 methylation at promoters and enhancers.* Epigenetics Chromatin, 2016. **9**(1): p. 37.
344. Jiang, W., et al., *The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression.* Cell Rep, 2015. **11**(1): p. 137-48.
345. Ulitsky, I., et al., *Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution.* Cell, 2011. **147**(7): p. 1537-50.
346. Li, L. and H.Y. Chang, *Physiological roles of long noncoding RNAs: insight from knockout mice.* Trends Cell Biol, 2014. **24**(10): p. 594-602.
347. Bernard, D., et al., *A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression.* Embo j, 2010. **29**(18): p. 3082-93.
348. Lin, R., et al., *Control of RNA processing by a large non-coding RNA over-expressed in carcinomas.* FEBS Lett, 2011. **585**(4): p. 671-6.
349. Tseng, J.J., et al., *Metastasis associated lung adenocarcinoma transcript 1 is up-regulated in placenta previa increta/percreta and strongly associated with trophoblast-like cell invasion in vitro.* Mol Hum Reprod, 2009. **15**(11): p. 725-31.
350. Nakagawa, S., et al., *Malat1 is not an essential component of nuclear speckles in mice.* Rna, 2012. **18**(8): p. 1487-99.
351. Zhang, B., et al., *The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult.* Cell Rep, 2012. **2**(1): p. 111-23.
352. Li, L., et al., *Targeted disruption of Hotair leads to homeotic transformation and gene derepression.* Cell Rep, 2013. **5**(1): p. 3-12.

353. Amândio, A.R., et al., *Hotair Is Dispensible for Mouse Development.* PLoS Genet, 2016. **12**(12): p. e1006232.
354. Schratt, G., et al., *Serum response factor is crucial for actin cytoskeletal organization and focal adhesion assembly in embryonic stem cells.* J Cell Biol, 2002. **156**(4): p. 737-50.
355. Smith, K.N., et al., *Long Noncoding RNA Moderates MicroRNA Activity to Maintain Self-Renewal in Embryonic Stem Cells.* Stem Cell Reports, 2017. **9**(1): p. 108-121.
356. Zeira, E., et al., *The knockdown of H19lncRNA reveals its regulatory role in pluripotency and tumorigenesis of human embryonic carcinoma cells.* Oncotarget, 2015. **6**(33): p. 34691-703.
357. Chauhan, B.K., et al., *Functional interactions between alternatively spliced forms of Pax6 in crystallin gene regulation and in haploinsufficiency.* Nucleic Acids Res, 2004. **32**(5): p. 1696-709.
358. Mahla, R.S., *Stem Cells Applications in Regenerative Medicine and Disease Therapeutics.* Int J Cell Biol, 2016. **2016**: p. 6940283.
359. Shroff, G. and R. Gupta, *Human embryonic stem cells in the treatment of patients with spinal cord injury.* Ann Neurosci, 2015. **22**(4): p. 208-16.
360. Zhou, S., et al., *Differentiation of human embryonic stem cells into cone photoreceptors through simultaneous inhibition of BMP, TGFbeta and Wnt signaling.* Development, 2015. **142**(19): p. 3294-306.
361. Shiba, Y., et al., *Human ES-cell-derived cardiomyocytes electrically couple and suppress arrhythmias in injured hearts.* Nature, 2012. **489**(7415): p. 322-5.
362. Bruin, J.E., et al., *Treating diet-induced diabetes and obesity with human embryonic stem cell-derived pancreatic progenitor cells and antidiabetic drugs.* Stem Cell Reports, 2015. **4**(4): p. 605-20.
363. Salguero-Aranda, C., et al., *Differentiation of Mouse Embryonic Stem Cells toward Functional Pancreatic beta-Cell Surrogates through Epigenetic Regulation of Pdx1 by Nitric Oxide.* Cell Transplant, 2016. **25**(10): p. 1879-1892.
364. Cheng, A., et al., *Cartilage repair using human embryonic stem cell-derived chondroprogenitors.* Stem Cells Transl Med, 2014. **3**(11): p. 1287-94.

List of Publications

Pal D, Neha CV, Bhaduri U, Zenia, Subbulakshmi C, Rao MRS. *LncRNA Mrhl orchestrates differentiation programs in mouse embryonic stem cells through chromatin mediated regulation*. Manuscript under preparation 2019.

Pal D, Iyer DP, Bhaduri U, McGreal R, Cvekl A, Rao MRS. *Identification of Pax6 as the major transcriptional regulator of lncRNA Mrhl in neuronal progenitors*. Manuscript under preparation 2019.

Book chapters:

Pal D, Rao M.R.S. *Long noncoding RNAs in pluripotency of stem cells and cell fate specification*. In: Rao M. (eds) Long Non Coding RNA Biology. Adv Exp Med Biol. 2017; **1008** (Springer, Singapore).

Akhade V.S, **Pal D**, Kanduri C. *Long noncoding RNA: genome organization and mechanism of action*. In: Rao M. (eds) Long Non Coding RNA Biology. Adv Exp Med Biol. 2017; **1008** (Springer, Singapore).

Review:

Fatima R, Akhade VS, **Pal D**, Rao SM. *Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets*. Mol Cell Ther. 2015; **3** (5).