Chromatin organisation mediated gene regulatory function of *mrhl* lncRNA at the Sox8 locus

---

A Thesis Submitted for the Degree of

Doctor of Philosophy

by

Bhavana Kayyar

Molecular Biology and Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
(*A Deemed University*)
Bangalore, India

July, 2021

DECLARATION

I hereby declare that the matter embodied in the thesis titled "**Chromatin organisation mediated gene regulatory function by *mrhl* lncRNA at the Sox8 locus**" is an authentic record of research work carried out by me under the guidance of **Prof. M.R.S Rao** at Chromatin Biology Laboratory, Molecular Biology and Genetics Unit (MBGU), Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore, India and that it has not been submitted elsewhere for the award of any degree or diploma.

In keeping with the norm of reporting scientific observations, due acknowledgments have been made whenever work described here has been based on the findings of other investigators. Any omission due to oversight or misjudgement is regretted.

Bhavana Kayyar

Bhavana Kayyar

Bangalore, India
Date: 27.02.2021

CERTIFICATE

I hereby certify that the matter embodied in the thesis titled "**Chromatin organisation mediated gene regulatory function by *mrhl* lncRNA at the Sox8 locus**" has been carried out by **Bhavana Kayyar** under my supervision at Chromatin Biology Laboratory, Molecular Biology and Genetics Unit (MBGU), Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore, India and that it has not been submitted elsewhere for the award of any degree or diploma to any other institution.

Prof. M.R.S Rao

Bangalore, India
Date: 27.02.2021

ACKNOWLEDGEMENTS

I would like to thank Prof. M.R.S Rao for giving me an opportunity to work on lncRNA biology under his guidance. His enthusiasm and excitement to carry out novel and relevant research is infectious and has kept me motivated through challenging times. He has given me the freedom to choose the course of this project and allowed me to grow professionally and personally under his mentorship.

I would like to thank the faculty of MBGU and NSU for the scientific discussions, coursework and constructive criticism over the years which has helped tremendously in shaping the work. I would like to thank Prof. Anuranjan Anand, Dr. Sheeba Vasu and Dr. Purusharth Rajyaguru for their inputs during my MS exam and for the encouragement and support in the subsequent years. I would like to specially thank Prof. Anuranjan Anand for his unwavering support through the years without which I would not have been able to continue my academic journey. I would also like to thank Prof. Kaustuv Sanyal for giving me the opportunity to spend a semester in his lab for my MS rotation. His guidance at the early stages has had an impact on my scientific temperament and approach to research. I would also like to acknowledge Prof. Udaykumar Ranga, Prof. Maneesha Inamdar, Prof. Namita Surolia, Prof. Tapas Kundu, Prof. Hemalatha Balaram, Dr. Ravi Manjithaya, Dr. James Chelliah and Dr. Kushagra Bansal for help, support and guidance throughout. I would also like to thank the central facilities of JNCASR without which it would have been impossible to carry out this work. I would like to thank Dr. Prakash and the animal facility, Suma and Sunil of Confocal facility and Anita of Sanger Sequencing facility.

I would like to specifically thank the co-authors of my publications Anjhana and Utsa who have been instrumental for the completion of this work. Utsa has supported me enthusiastically by carrying out any *in silico* analysis that I have asked of him without and questions and his cooperation was instrumental in the early days of the project and in laying out the hypothesis. Anjhana has very competently shared my work load at the crucial phase of my PhD and has been a very reliable source of support, especially through the challenging times of the pandemic. I would like to specially thank lab members Aditya, Zenia, Aishwaryaa, Roshan, Sangeeta, Shalini, Arun, Subhendu and Sunita for their unconditional support, encouragement and guidance throughout. I'm lucky in having found friends and mentors in them. I would like to thank past members of the lab Nikhil, Vijay, Debosree, Shubhangini, Meenakshi, Mona Lisa, Neha, Dhanur, Raktim, Prathima and Sreenivas for their training and guidance in the initial years of my PhD. I would like to

thank Sharath, Sakthi Veena, Muniraju for helping out with essential tasks in the lab which ensued that my work proceeded without interruptions.

I would like to thank members of MML Neha and Shreyas, who I worked with during my rotation and who trained me during my early days of PhD, Vikas, Gautam, Sundar, Lakshmi Shankar Rai, Lakshmi Sreekumar, Rima, Krishnendu, Radha, Parijat, Asif, Priya, and Jigyasa for the help and the lively lab environment.

I would like to thank Dr. Radu Zabet for giving me the opportunity to work under his guidance at the University of Essex with some of novel techniques. He has provided not only guidance but went out of his way to ensure that I could complete the planned work at Essex. I would also like to thank the faculty at Essex Prof. Phil Mullineaux, Prof. Leo Schalkwyk Dr. Greg Brooke, Dr. Joaquin De Navascues, Dr. Antonio Marco, Dr. Patrick Varga-Weisz, Dr. Andrea Mohr, Dr. Ben Skinner and others for immense support during my visit. All of the faculty were extremely generous with their reagents, funds, time and guidance to make sure that I could complete a significant amount of work. I also thank them for the coursework in R. I would like to thank the technicians, staff and members of the Essex community Mata, Julie Double, Julie Arvidson, Amanda, Tania, Chris, David Knight and others for their incredible support to ensure that I could work uninterrupted late into evenings, though weekends and Christmas to complete my work. Some of them continued to support me long after I returned from Essex. I would like to thank members of the Zabet lab Dr. Keerthi, Patrick, Jareth, Liv, Mila and Romana for making me feel welcome and going out of their way to help me out. I would also like to thank my friends Salma, Mia, Jess and Mariane for helping me out with work and outside the lab.

I would like to thank my friends at JNCASR Veena, Pallabi and Siddharth for their constant support and encouragement. As batchmates, they have been able to relate to the challenges of PhD life and have always pushed me to keep moving forward through the tough times. I would like to thank the student body of MBGU for their support and help. It is impossible to work without constant exchange of instruments, reagents and expertise. It would have been impossible to keep the lab running without their support through the months of lockdown imposed due to the pandemic.

I would like to thank the various facilitates such as Students' Mess, Hostel, Dhanvantari, Admin, Security, Chandraya, Accounts, Academic and others for providing the infrastructure and setup that enables research. I would like to thank JNCASR and EMBO for the fellowships.

Finally a special thanks to my family and friends for their phenomenal support and encouragement. They have been incredibly understanding of my absence from family events and have always encouraged me to stay focused. They have been my pillars of strength though difficult times and have pushed me to complete my graduate studies.

List of Abbreviations

3C	Chromosome conformation capture
3D	Three dimensional
AML	Acute myeloid leukaemia
ATP	Adenosine triphosphate
bp	Basepair
BSA	Bovine serum albumin
cDNA	Complimentary DNA
CDS	Coding sequence
CHE	Chromatin modifying enzymes
ChIA-PET	Chromatin Interaction Analysis with Paired-End Tag
ChIP	Chromatin immunoprecipitation
ChOP	Chromatin Oligo affinity purification
CML	Chronic myeloid leukemia
CpG	Cytosine–guanosine dinucleotides
CRC	Colorectal Cancer
CRE	*Cis* regulatory elements
CSR	Class-switch recombination
DHS	DNase hypersensitive site
DIM	Dimerisation
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EB	Enhancer Blocking
EDTA	Ethylenediaminetetraacetate
EMSA	Electrophoretic Mobility Shift Assay
eRNA	Enhancer RNA
FBS	Fetal bovine serum
HAT	Histone acetyl transferase
HDAC	Histone deacetylase

hnRNP	Heterogeneous nuclear ribonucleoprotein
HRP	Horse radish peroxidase
ICR	Imprinting control region
kb	Kilobase
L	Litre
LB	Luria-Bertani Media
lincRNA	Long intergenic non-coding RNAs
LLPS	Liquid-liquid phase separation
lncRNA	Long non-coding RNA
mESC	Mouse embryonic stem cells
mg	Milligram
miRNA	MicroRNA
ml	Millilitre
mM	Millimolar
mPIC	Mammalian protease inhibitor complex
mRNA	Messenger RNA
NAT	Natural antisense transcripts
ncRNA	Non-coding RNA
NDR	Nucleosome depleted region
NFW	Nuclease free water
ng	Nanogram
NGS	Next generation sequencing
NRE	Negative regulatory element
nt	Nucleotide
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
piRNA	piwi-interacting RNA
qPCR	Quantitative polymerase chain reaction
RBR	RNA binding region
RNA	Ribonucleic acid
RNA PolII	RNA polymerase II
RNAi	RNA interference

RNaseA/H	Ribonuclease A/H
rpm	Rotations per minute
rRNA	Ribosomal RNA
RT-PCR	Real-Time Polymerase Chain Reaction
SDS-PAGE	Sodium dodecyl sulphate- polyacrylamide gel electrophoresis
shRNA	Short hairpin RNA
siRNA	Short interfering RNA
SMH	Somatic hypermutations
smORF	Small Open Reading Frames
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
TAD	Topologically associated domain
TAD (SOX8)	Transactivation domain
TBS	Tris-buffered saline
TF	Transcription factor
TFO	Triplex forming oligo
tRNA	Transfer RNA
TSS	Transcription start site
TTS	Triplex target site
UTR	Untranslated region
UV	Ultraviolet
V	Volts
WRE	Wnt responsive element
XCI	X-chromosome inactivation
µg	Microgram
µl	Microlitre

List of Figures

List of Tables

TABLE OF CONTENTS

Chapter 1

Introduction

Early life on Earth is believed to have evolved around Ribonucleic acids (RNA) as the genetic material before the evolution of Deoxyribonucleic acids (DNA) and proteins according to the RNA world hypothesis proposed in the 1960s. RNAs continue to remain very important biomolecules participating in several fundamental processes in cells today. Crick, Jacob and Monod's Central Dogma of Life, put forward in 1958, categorised RNA as an intermediary between the genetic material DNA and the effector molecule, proteins. While approximately 2% of the genome gives rise to protein-coding RNA molecules, it is now known that genomes are pervasively transcribed to give rise to many non-coding transcripts which perform a myriad of functions within cells.

Initially discovered along with DNA as 'nuclein' in 1869, RNA and DNA were discovered as 'yeast' nucleic acid and 'thymus' nucleic acid respectively, based on the material from which they were isolated. It was not until half a decade later that the two molecules were described to be chemically different. The two were found to be containing different sugars and thus, led to the name Ribose nucleic acid for 'yeast' nucleic acid. RNA was found to be a DNA-like molecule synthesised from a DNA template. Subsequently, the coding function of messenger RNAs (mRNA) was discovered, although the term itself was coined later. The first non-coding RNA discovered in 1955 by Georges Palade was the ribosomal RNA (rRNA), a component of the ribosome in the cytoplasm of cells. Very soon after, the transfer RNA (tRNA) was discovered and was the first non-coding RNA to be characterised. This RNA molecule was predicted by Crick to be an 'adapter' molecule for the translation of RNA into protein. These two classes of non-coding RNAs fall into the category of housekeeping non-coding RNAs because of their integral role in the protein translation process (Jarroux et. al., 2017).

However, RNA comes in more than three flavours and it was subsequently discovered that non-coding RNA (ncRNAs) molecules can also play regulatory roles in cells. The first evidence of regulatory functions of non-coding RNAs was uncovered in the 1980s with the discovery of the bacterial micF transcript in *E.coli*. There were some non-coding transcripts such as 6S RNA, Spot 42 and the eukaryotic 7SK RNA which were discovered earlier but remained functionally uncharacterised till 40 years later. MicF is the bacterial equivalent of

microRNA (miRNA), called sRNA, and functions to regulate the ompF mRNA by complementary binding. This inhibits its expression (Delihas et. al., 2015). The mechanism of action of miRNA has been described in much detail in subsequent pages.

In 1969, Britten and Davidson postulated a Battery model for regulation of gene expression in eukaryotes which explored the possibility of ncRNAs and proteins acting as regulatory intermediaries to convey signals received at sensory genetic elements to receptor elements that affect coding gene transcription. This hypothesis captures the essence of the functions of non-coding RNAs very well. The first long non-coding RNA molecules that were discovered and characterised to have epigenetic regulatory functions were H19 in 1990 and *Xist* in 1992. With the advent of whole-genome sequencing technology, it is now understood that almost the entire mammalian genome is pervasively transcribed and with this, there has been a shift from the mRNA centric view of the transcriptional landscape. Around 70-90% of the genome is estimated to be transcribed based on the discoveries made in ENCODE project and the more recent projects such as the GTEx project and a majority of these transcripts are non protein-coding. This understanding has quashed the historic view of the bulk of the genome containing junk DNA (Kung. J.T.Y et.al., 2013).

1.1 Non-coding RNAs

Based on size and functions, Ribonucleic acids can be categorised into various sub-types. **Figure 1.1** below shows the different categories of RNAs.

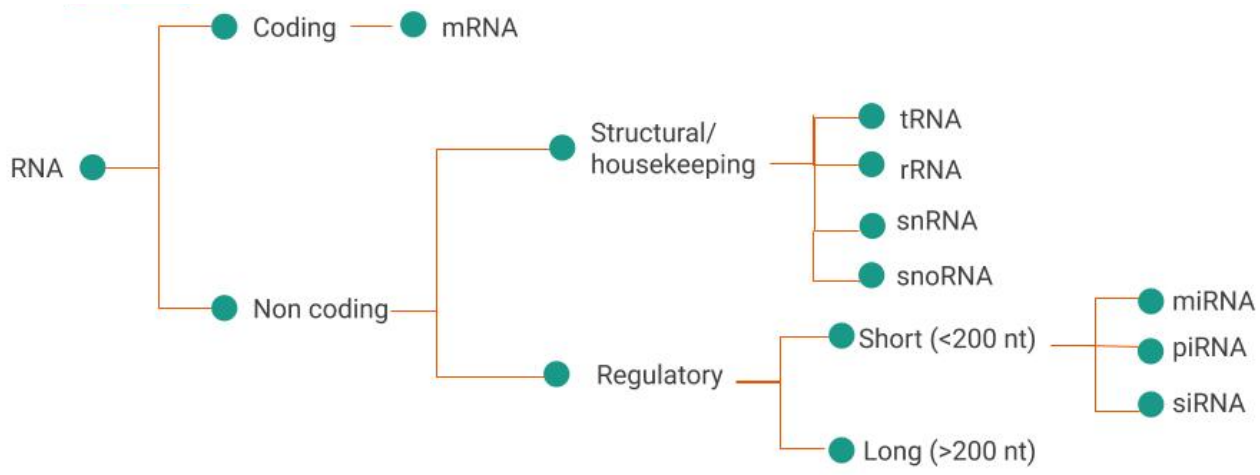


Fig 1.1: Classification of RNA based on their coding potential, function and size. *Based on its protein-coding potential, RNA is categorised as coding and non-coding. Non-coding RNA is further categorised based on function into housekeeping and regulatory. Housekeeping ncRNAs are*

tRNA, rRNA, snRNA and snoRNA. Regulatory ncRNAs are further categorised based on size into short and long ncRNA. Short ncRNAs are of three types - miRNA, piRNA and siRNA.

Based on the protein-coding potential of transcripts, RNA is categorised into coding and non-coding RNA with the messenger RNA (mRNA) being the sole sub-type of coding RNA. mRNA is transcribed from DNA template by RNA polymerase II in the nucleus. The transcripts get exported to the cytoplasm which is the site of protein translation. mRNAs carry the protein code in the form of a sequence of three nucleotide 'codons' which are continuous and non-overlapping.

Non-coding RNA can have either Housekeeping or Regulatory functions. Housekeeping RNAs are expressed in all cells and are of 4 types - transfer RNA (tRNA), ribosomal RNA (rRNA),both of which play important roles in protein translation, small nuclear RNA (snRNA) and small nucleolar (snoRNA) RNA. The characteristically clover-leaf shaped tRNA acts as the adapter between mRNA and the protein that it encodes. The anti-codon loop present in the tRNA recognises the triplet codons on the mRNA by complementary base-pairing. Each tRNA with a specific anti-codon sequence is specifically associated with an amino acid.

Ribosomal RNAs associate with ribosomal proteins to form the ribonucleoprotein complex make up approximately 60% of the mass of ribosomes. Ribosomes are made up of two subunits and in eukaryotes, the small 40S subunit (SSU) of the ribosome contains a single 18S rRNA and the larger subunit contains at least 3 different rRNA components - 5S, 5.8S and 28S. All of these except for 5S rRNA are transcribed as a single precursor in the nucleolus by RNA Polymerase I. 5S rRNA is transcribed independently by RNA polymerase III (Pol III). While initial studies pointed towards rRNA behaving as a scaffold for protein assembly machinery including tRNA binding, it is now known that rRNA can additionally catalyse the peptide bond formation during protein synthesis (Nazar, 2008). Thus, the codons of mRNAs are read by the ribosomes by involving two kinds of housekeeping non-coding RNAs (Feher, J. et al, 2017) and translated into proteins.

1.1.1 Small nuclear RNAs

Small nuclear RNAs (snRNAs) are a group of highly abundant, nuclear localised transcripts of an average length of 150 nucleotides. They can be categorised into two classes based on sequence features and protein co-factors - (A) Sm class of snRNAs comprising of U1, U2, U4, U4atac, U5, U7, U11 and U12 which are RNA polymerase II transcribed, and (B) the Lsm class is made up of U6 and U6atac which are transcribed by RNA polymerase III. The main functions of snRNAs are in splicing to remove introns from pre-mRNAs (Matera et.al., 2007)

1.1.2 Small nucleolar RNAs

Small nucleolar RNAs (snoRNAs) are 60-300 nucleotide long RNAs localised to the nucleolus that associate with various proteins to perform rRNA modification and processing. They are the most abundant group of non-coding RNAs and are categorised into two groups - (A) the C/D box snoRNAs and (B) H/ACA box snoRNAs based on the characteristic sequence motif associated with the two respectively. Many of the snoRNAs are encoded for within introns of genes. Their canonical functions are in 2'-O- ribose methylation and pseudouridylation of rRNAs respectively. In addition to this, some snoRNAs such as SNORD14 and SNORD22 can play a role in pre-rRNA cleavage. SnoRNAs can also have regulatory functions apart from these housekeeping functions within a cell. For instance, they can regulate gene expression at the level of mRNA by regulating alternative splicing. SNORD115 regulates the alternative splicing of at least 5 pre-mRNAs- DPM2, TAF1, RALGPS1, PBRM1, and CRHR1 and along with SNORD116, regulates expression of over 200 genes possibly by regulation of alternative splicing. They can also be precursors of miRNA. SnoRNA–miR-28 is an miRNA like derivative of snoRNA and is significantly upregulated in breast tumours and promotes proliferation of tumour cells. Due to the very diverse roles played by snoRNAs, they are associated with many diseases including cancers (Liang, J. et.al., 2019).

1.1.3 Short non-coding RNAs

Regulatory non-coding RNAs are categorised into small/short and long non-coding RNAs based on the size of the transcripts with anything below 200 nucleotides being categorised

as small. These regulatory RNA molecules are expressed in a tissue specific manner. Small non-coding transcripts are of 3 types- miRNA, piRNA and siRNA. MicroRNAs (miRNA) are molecules of a length of 22 nucleotides. Most often, miRNAs exert their regulatory function by binding to mRNAs predominantly at the 3'UTR and occasionally, also 5'UTR or coding regions and targeting them for degradation, thereby decreasing their steady-state levels. miRNAs are transcribed as pri-miRNA which get processed to a 70 nucleotide long (precursor) pre-miRNA by the microprocessor complex consisting of the proteins DGCR8 and Drosha. The pre-miRNA is exported to the cytoplasm by exportin where the terminal loop is cleaved by Dicer to give rise to the double-stranded miRNA. miRNA can be generated non-canonically independent of Drosha or Dicer activities. Both strands of the miRNA can be loaded into the Argonaute family of proteins present in the RNA-induced silencing complex (RISC) (Scott, M.S. et.al., 2011). A single miRNA can regulate multiple target mRNAs and a single mRNA can be regulated by multiple miRNAs and because of this, miRNAs play an important regulatory role in multiple metabolic and developmental pathways. While short interfering RNAs (siRNAs) are similar to miRNAs by definition and function, they differ in their origin, siRNA being chemically synthesised and introduced into cells to target a specific gene of interest (**figure 1.2**). Examples of miRNA in humans, are miR-17–5p and miR-20a that function as regulators of cell proliferation. They bind to and repress E2F1 mRNA in response to stimulation by c-Myc. E2F1 is a transcription factor whose expression promotes G1-S phase transition in the mammalian cell cycle (O'Donnell et.al., 2005, Carthew, R. W., 2006).

piRNAs are small non-coding RNAs of 26- 30 nucleotides length associated with the Piwi clade of argonaute proteins found in *Drosophila* germ cells. These RNAs were first identified in mouse and rat testes and are highly enriched in germline tissues which express 3 piwi proteins - MIWI, MILI and MIWI2 which are all individually essential for male fertility. They are generated through a Dicer independent mechanism from piRNA clusters and carry 2′-O-Me modification on the 3' end when mature. Based on their expression pattern during development, different types of piRNAs have been identified in mice. Pre-pachytene piRNAs are predominantly present in the germ cells of fetal and newborn mice and account for 95% of know piRNAs. They are enriched for transposon and gene-derived sequences and bind to MILI and MIWI2 whereas pachytene piRNAs originate from distinct intergenic loci and are depleted of repeat sequences. They associate with MILI and MIWI. A very well characterised function of the piwi-piRNA complex is the silencing of transposon

elements in a mechanism similar to the one adopted by miRNA. In murine cells, this is achieved by establishing methylation at CpG DNA at retrotransposon elements. (Weick, E.M. and Miska, E.A, 2014). In addition, piRNAs have also been suspected of regulating protein-coding genes although examples and mechanisms are not yet known.

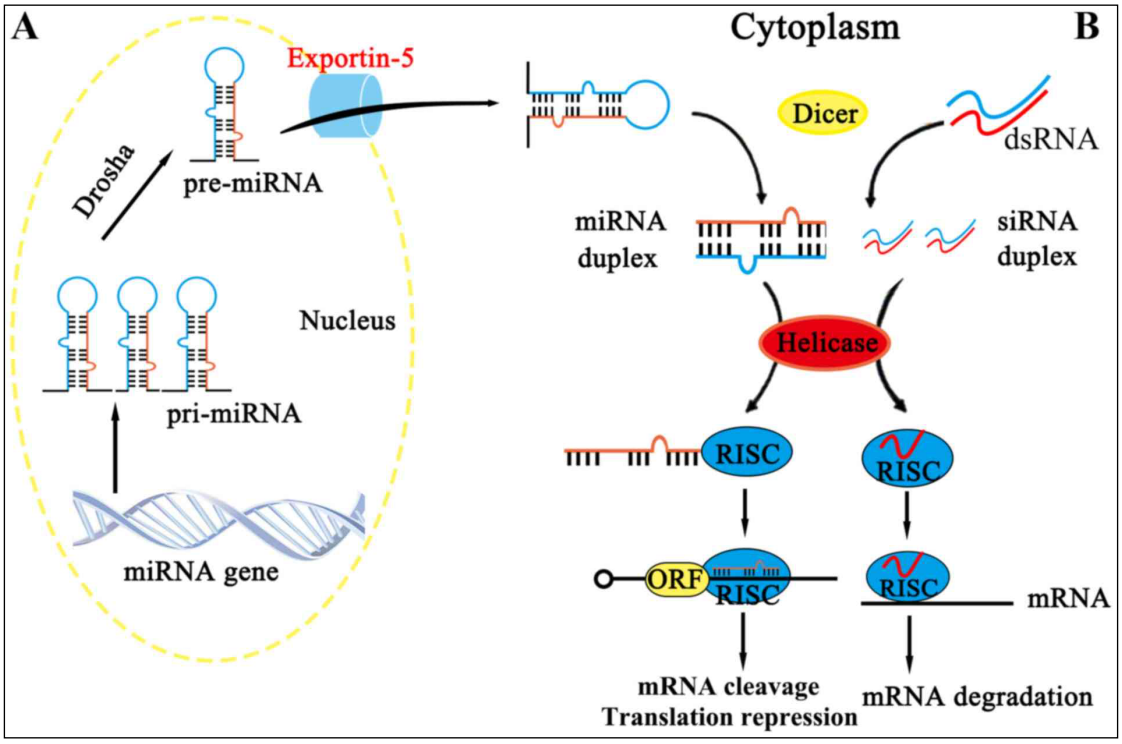


Fig 1.2 : Functional mechanism of miRNA and siRNA *: miRNA is produced from endogenous genes containing hairpin structures of 65–70nt; the hairpin structure is processed by Drosha-DGCR8 complex into pre-miRNA in the nucleus, transferred to the cytoplasm, then processed into miRNA by Dicer, and loaded to protein Argonaute (AGO2) of the RISC complex. (B) siRNA supplied exogenously is cut into a fragment of 21–25nt by the Dicer enzyme and then loaded into the RISC complex. The targeted mRNA is translationally repressed and oftentimes, degraded. Figure has been adapted from Wei J-W et al, 2016 with permission.*

1.1.4 Long non-coding RNAs

Non-coding transcripts larger than the arbitrary size of 200nt are categorised as long non-coding RNAs.These represent the most functionally diverse class of non-coding RNAs and the function of a vast majority of known lncRNAs remain to be characterised. Two of the earliest discovered lncRNAs are H19 and *Xist*, with functional roles in gene imprinting and X chromosome inactivation respectively, paved the way to identifying the role of lncRNAs as epigenetic regulators (Brannan, C. I. et al,1990, Gabory, A. et al, 2009 Brown, C.J et al, 1992, Clemson, C. et al, 1996). Their unexpected abundance came to light during the ENCODE project. Djebali *et. al* observed from the ENCODE project that almost 70% of the

human genome is cumulatively transcribed across different cell types and this discovery has been expanded to many mammals by Carninci *et. al* (Djebali, S. *et al*, 2012, Carninci, P. *et a*l, 2005). Surprisingly, clusters of overlapping non-coding transcripts are more numerous than protein-coding transcript clusters in mice.

The number of functional lncRNAs is debated. It was proposed by Struhl K et al that RNA Pol II can initiate many spurious transcriptional events and hence, many transcripts in the cells could be transcriptional noise (Struhl, K et al, 2007). Further, Ulitsky and Bartel too hypothesised in support of this notion that most of the lncRNAs categorise as transcriptional noise (especially those arising from bidirectional promoters and overlapping transcripts) and are not functional. They proposed that due to varying chromatin cell states across different cell fates, different regions of the chromatin are accessible thereby resulting in a different repertoire of transcripts. In addition, these transcripts are processed due to the underlying mechanisms and chance occurrence of splice sites (Ulitsky, I. and Bartel, D. P. 2013). Guttman et. al, further proposed that the low level of conservation observed in long non-coding transcripts would require that each clade evolve its own set of lncRNAs thereby suggesting that most of the transcripts are transcriptional noise and few have bonafide functions (Guttman, M. et al, 2009). This adds weight to the argument that the lack of evolutionary conservation does indicate the lack of a biological function. However, although the functionality of most lncRNAs lacks evidence, a growing number of lncRNAs are being characterised to have important cellular functions. In contradiction to the above arguments, Mercer T et. al, argued that tight spatial and temporal regulation of lncRNA expression, a characteristic feature of many lncRNAs, suggests biological functionality of lncRNAs (Mercer, T.R et al, 2009).

Long non-coding RNAs share many features with mRNAs. Both classes of RNAs are transcribed by RNA polymerase II and both are generally polyadenlayted and have 5' caps. LncRNA genes exhibit similar epigenetic characteristics as protein-coding genes including similar histone modifications at gene promoters and gene bodies (both are marked by the presence of trimethylation of lysine4 of histone H3 (H3K4me3) at their promoter and trimethylation of lysine36 of histone H3 (H3K36me3) along the length of the transcribed region), DNA methylation, paternal/ maternal effect and post transcriptional modifications (Okazaki, Y et al 2002, Derrien, T. et al, 2012). Some lncRNAs are multi-exonic and undergo co-transcriptional splicing.

LncRNAs differ from mRNAs in having very low sequence conservation phylogenetically. Functionally, too, these transcripts may not be very well conserved. However, conservation at the level of synteny and some secondary structures is observed.

LncRNAs are expressed tissue specifically and at much lower than mRNAs with most of them being restricted to particular developmental contexts. However, the defining feature of long non-coding RNAs is that they do not have significant, functional open reading frames.

1.1.4.1 Classification of lncRNAs

One way to classify long non-coding RNAs is based on their genomic origin as shown in **figure 1.3** below. They can originate from intergenic regions, from within protein-coding genes- either intronic or overlapping in either sense or antisense directions, or from bidirectional promoters. Almost 70%- 80% of all protein-coding sense transcripts are also reported to have antisense transcripts in mice, mostly overlapping at the 5' or 3' ends of the sense transcripts. These are also called Natural Antisense Transcripts (NATs). It has been found that long intergenic non-coding RNAs (lincRNAs) are more conserved than introns and antisense transcripts, more tissue-specifically expressed than protein-coding genes and more stable than intronic lncRNAs (Ma.L. *et.al.*, 2013).

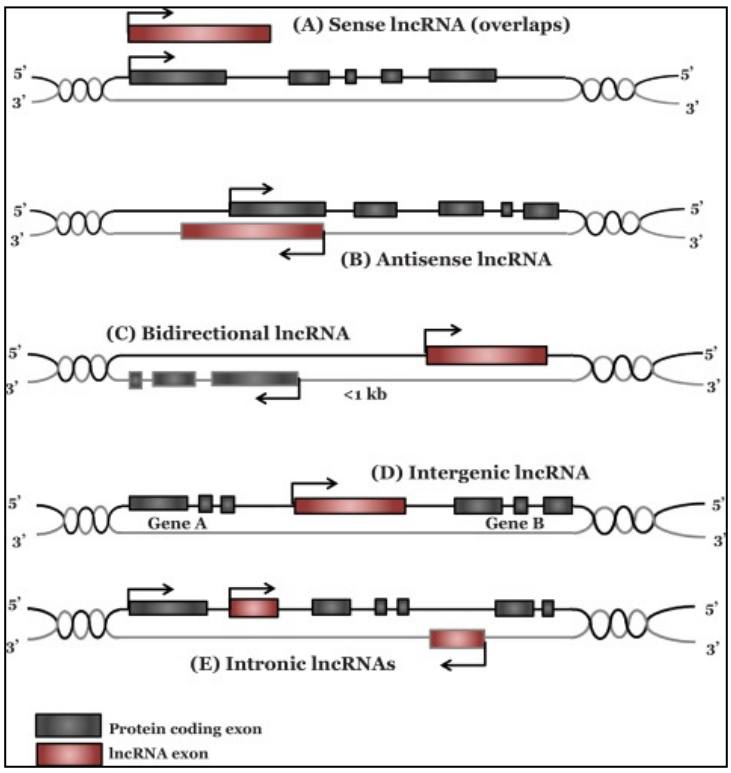


Fig 1.3: Classification of lncRNA based on their genomic origin and orientation with respect to protein-coding genes*. A) Sense-overlapping lncRNAs overlap with one or more introns and/or exons of a protein-coding gene in the sense RNA strand direction B) Antisense lncRNAs originate from the antisense RNA strand of a protein-coding gene. C) Bidirectional lncRNAs are transcribed from the same promoter as a protein-coding gene, but in the opposite direction D) Intergenic lncRNAs are located between protein-coding genes. E) LncRNAs can also be transcribed from within introns of protein-coding genes independent of the host gene. The figure has been adapted with permission from Alahari, S.V et al, 2016.*

Based on their mechanism of action, lncRNAs are classified into 4 different archetypes as shown in **figure 1.4** (Ballantyne, M.D et al, 2016). The expression of most lncRNAs is under significant transcriptional control that is a cumulative outcome of various developmental cues and as a result, is expressed at a specific time and place within an organism. Because of this, lncRNAs can function as molecular signals (**archetype A, figure 1.4**) to mark space, time or developmental stage. Some examples of lncRNAs functioning as signals include *Kcnq1ot1* and *Air* in the imprinting of the Kcnq1 and Igf2r gene clusters respectively. *Kcnq1ot1* and *Air* are transcribed from the paternal alleles and bind to the promoter of the imprinted genes to mediate repressive histone modification by interacting with histone methyltransferases (Braidotti, G. et al, 2004).

A mechanism of negative regulation by lncRNAs is by acting as molecular decoys (**archetype B, figure 1.4**) or 'molecular sinks' and titrating away regulatory molecules from target sites. Examples of lncRNAs operating via this mechanism include the lncRNA TERRA which contacts the telomerase reverse transcriptase (TERT) protein subunit independently of the telomerase template RNA moiety. Telomeric heterochromatin bound TERRA is thought to bind and sequester telomerase to near the telomeric 3'-end while inhibiting its action. As TERRA levels change in a cell cycle-dependent manner, downregulation of TERRA in the S-phase might unleash telomerase and allow extension of the telomeric strand in a cell cycle-dependent manner (Redon, S. et al, 2010, Porro, A. et al, 2010). The lncRNA Gas5 (Growth arrest–specific 5) represses the glucocorticoid receptor through the formation of an RNA motif mimicking the DNA motif of hormone response elements found in the promoter regions of glucocorticoid-responsive genes. Gas5 competes with the promoter DNA for binding to the DNA binding domain of the glucocorticoid receptor and precludes its DNA interaction (Kino, T. et al, 2010).

The third archetype of lncRNA is the guide RNA (**archetype C, figure 1.4**) whereby lncRNA binds to regulatory molecules and directs the localisation of regulatory complex to specific targets. LncRNAs can guide changes in gene expression either in *cis* or in *trans.*

Examples include the lncRNA *Xist* that plays a role in recruiting the Polycomb Repressive Complex to the X-chromosome to be inactivated (Wutz, A. et al., 2002, Sun, B.K. et al, 2006) and ncRNA pRNA that plays a role in recruiting the DNMT3b complex to the rDNA promoter (Schmitz, K.M et al, 2010). The mechanisms of action have been described in detail in subsequent pages.

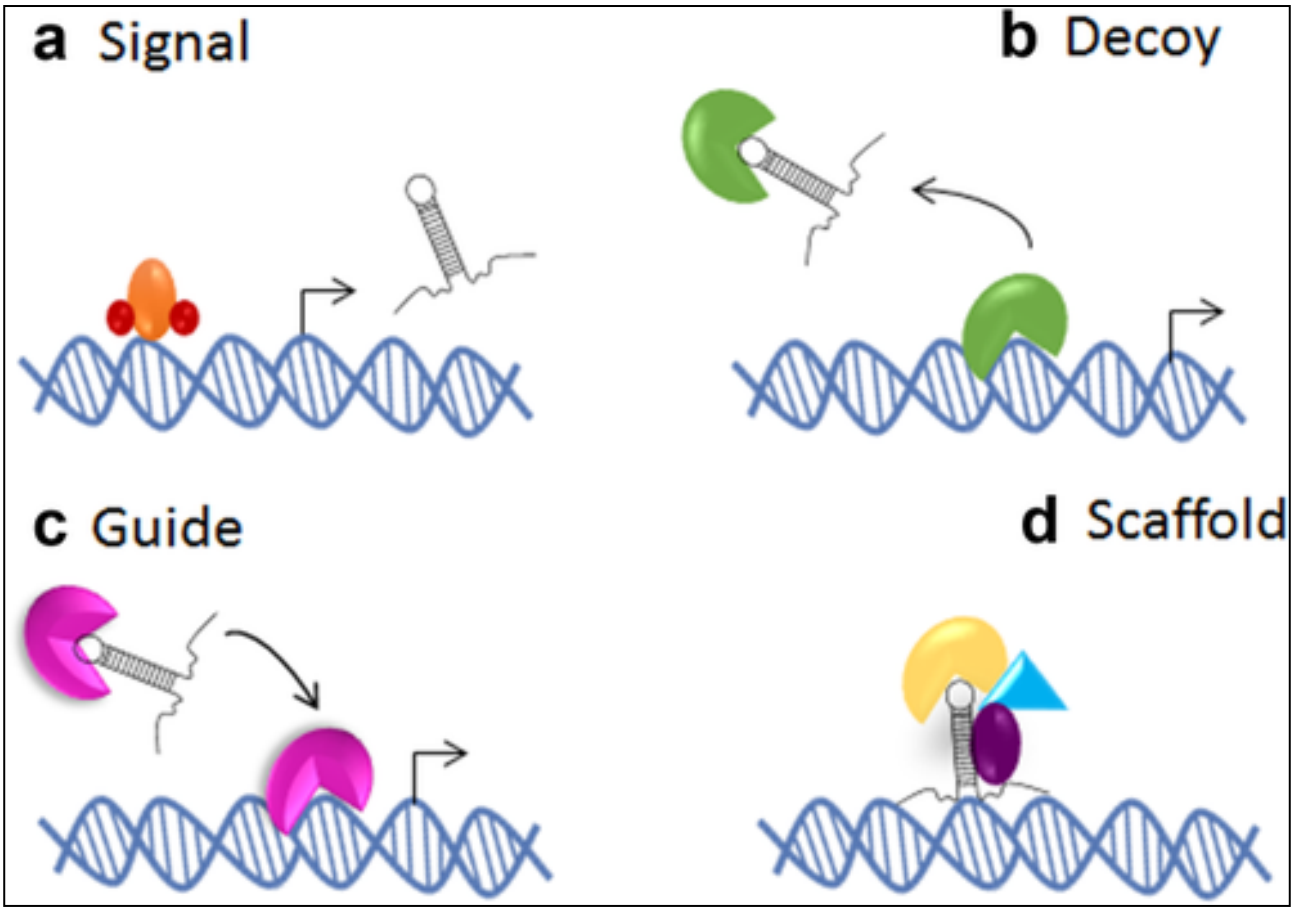


***Fig 1.4**: **Classification of lncRNAs based on the mechanism of action**. Long non-coding RNAs can be classified into archetypes based on their mechanism of function (**a**) lncRNA may act as signals of transcriptional activity (mark of cellular fate) and may signal gene regulation. (**b**) lncRNA may act as sponges for molecules such as miRNA and proteins, thus reducing the bioavailability of the molecule for their target function, thereby altering cellular function. (**c**) lncRNAs can act as guides for chromatin-modifying complexes, thus aiding in their recruitment to DNA and contributing to tissue-specific gene expression. (**d**) lncRNA may act as scaffolds bringing together essential proteins required for gene or cellular regulation. Adapted with permission from Ballantyne, M.D et al, 2016.*

In the fourth archetype, lncRNAs can serve as platforms upon which relevant molecular components of regulatory complexes are assembled. By binding to multiple effector partners at the same time, lncRNAs can bring regulatory molecules together in both time and space. The lncRNA HOTAIR binds the polycomb complex PRC2, which methylates histone H3 on lysine 27 to promote gene repression and also interacts with a second

complex containing LSD1, CoREST, and REST, that demethylates histone H3 on lysine 4 to antagonise gene activation. Thus, multiple chromatin-modifying complexes are bound and targeted by HOTAIR, suggesting that HOTAIR acts as a scaffold and bridges PRC2 and the LSD1/CoREST/REST complex (Tsai, M.C. et al, 2010). Work on the lncRNA ANRIL demonstrates a direct interaction between ANRIL and components from both PRC1 and PRC2 complexes. Binding to ANRIL is important for the functions of both PRC1 and PRC2 proteins, and disruption of either interaction impacts transcriptional repression of the target *INK4b* locus (Kotake, Y. et al, 2011, Yap, K.L. et al, 2010).

It is important to note that most lncRNAs can not be categorised into a single archetype and often perform functions characteristic of multiple archetypes (Wang, C and Chang, H.Y, 2011).

1.1.4.2 Peptide coding lncRNAs

There has been an emerging concept of pervasive translation similar to how genomes undergo pervasive transcription. In contradiction to their defining feature, a notable number of lncRNAs have been found to encode short peptides of the length of less than 100 amino acids which are functionally significant and are called small Open Reading Frames (smORF). Short peptides could be coded for by regions in lncRNAs: the untranslated regions (UTRs) of annotated transcripts either upstream or downstream of the coding sequence (CDS), and even overlapping the CDS of canonical mRNAs. LncRNAs have been discovered to be ribosome-associated by ribosome profiling experiments with close to 82% of mouse lncRNAs appearing to be scanned by ribosome machinery (Ruiz-Orera et. al., 2014). Short peptides have also been discovered and mapped to being translated from lncRNAs in *Drosophila melanogaster*, zebrafish and other mammals. Many of these short peptides are functionally relevant to the cell. For example, SPAR, a polypeptide translated from lncRNA *LINC00961*, inhibits amino acid-mediated mTORC1 activation at the lysosomal membrane by binding to the v-ATPase complex as revealed by knockdown and overexpression studies in mammals. SPAR was discovered by proteomics which was aimed at identifying polypeptides encoded by lncRNAs. The transcript *LINC00961* is conserved across species and is expressed significantly in lung, heart and skeletal muscle (Rion and Ruegg, 2017).

Similarly, DWORF is a 34-amino acid long short-peptide encoded for by a long non-coding RNA in mammals. This peptide appears to compete with negative regulators of the sarco/ endoplasmic reticulum calcium adenosine triphosphatase (SERCA), an ion pump that is a key player in handling calcium in striated muscles (Makarewich, C. A. et al, 2018).

Small peptides play an important regulatory role in diseases such as cancer. The lncRNA *HOXB-AS3* encodes a conserved 53-aa HOXB-AS3 peptide that suppresses colon cancer (CRC) growth (Huang, J. Z. *et al*, 2017).

1.1.4.3 Cytoplasmic Long non-coding RNAs

Around 25% of all expressed lncRNA are detected in the cytoplasm. While less well understood than nuclear lncRNAs, these play a significant role in epigenetic regulation outside the nucleus. LncRNAs and mRNAs share the same mechanism by which they are exported to the cytoplasm. Long and A/U-rich mRNA transcripts with one or only a few exons are dependent on the nuclear RNA export factor 1 (NXF1) pathway for export and many lncRNAs, having fewer exons than mRNAs, exploit this pathway too (Zuckerman, B et al, 2020). A majority of cytoplasmic lncRNAs are associated with polysomal fractions and degradation of some cytoplasmic lncRNAs may be triggered by a translation-dependent mechanism (Carlevaro-Fita, J et al, 2016). Cytoplasmic lncRNAs can be of nuclear origin or be transcribed from mitochondrial DNA and play essential roles in governing cytoplasmic events including mRNA stability, translation and degradation, localisation and recruitment of translational factors and acting as decoys for miRNA and other regulatory proteins present in this cellular compartment. **Figure 1.5** below summarises some of the key events regulated by lncRNA in the cytoplasm. **Table 1.1** below contains examples of cytoplasmic lncRNAs, their partners and their function in the cytoplasm (Noh, J.H. et al, 2018). Interestingly, few lncRNAs of nuclear origin can be detected within the mitochondria. For example, the RNA component of mitochondrial RNA-processing endoribonuclease (*RMRP*) is transcribed in the nuclei and exported to the mitochondria where it is bound to and stabilised by GRSF1.

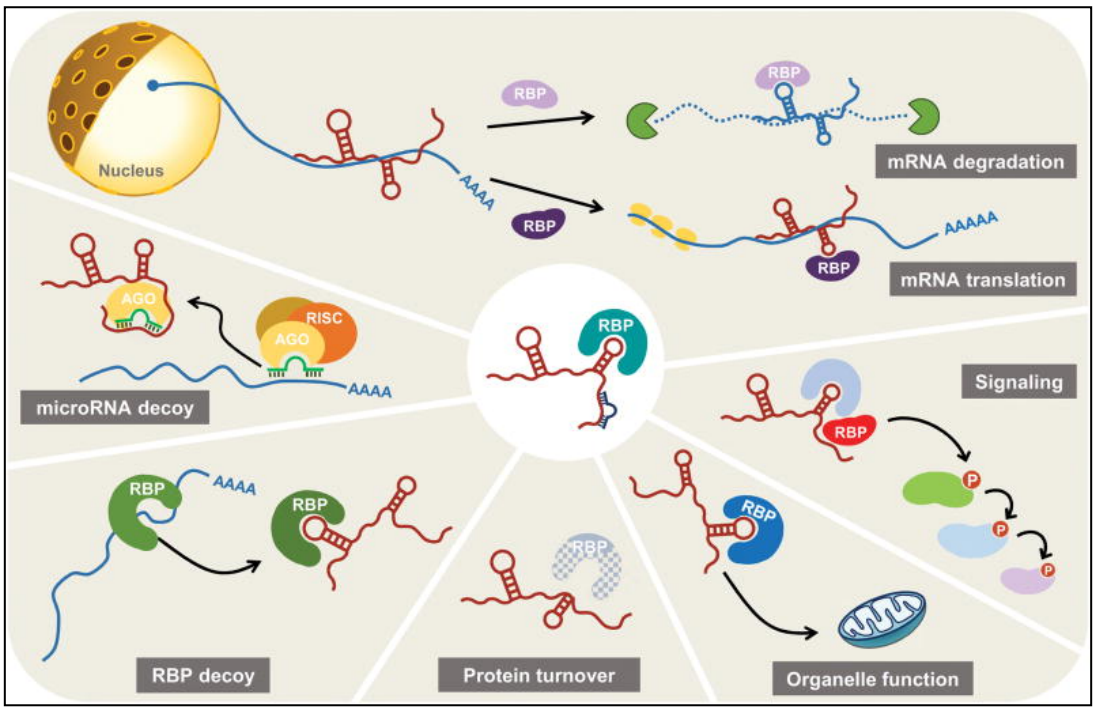


Fig 1.5: Functions of cytoplasmic lncRNAs. *Top- following export to the cytoplasm, lncRNAs can associate with RNA-binding proteins (RBPs) or complementary mRNAs to regulate the stability and/or translation of specific mRNAs. Signalling- the association of RBPs with lncRNAs can activate signalling molecules (e.g., kinases). Organelle function- lncRNAs can mobilise RBPs to cellular organelles where they carry out specific functions. Protein stability- lncRNAs can serve as platforms that facilitate the presentation of specific RBPs to the protein degradation machinery. RBP decoy and microRNA decoy- lncRNAs binding to RBPs and microRNAs can reduce the availability of these factors to mRNAs, thereby modulating mRNA fate. Figure adapted with permission from Noh et al, 2018*

Function	**LncRNA**	**Partner**	**Details of regulation**	**Reference**
mRNA stability and translation	*TINCR*	Stau1	lncRNA target mRNA undergoes STAU1 mediated degradation	Kretz et al., 2013
	lincRNA-p21	HuR	Targets mRNA for RISC mediated degradation	Yoon et al., 2012
Protein stability	*lincRNA-p21*	HIF1A, VHL	Prevents degradation of HIF1A by VHL by binding to both - decoy function	Yang et al., 2014
	HOTAIR	SNUPN1, ATXN1	Scaffold for binding of SNUPN1 and ATXN1 and their ubiquitin ligases leading to their degradation	Yoon et al., 2013

miRNA decoys	*RoR*	miR-145	Sponges miR-145 which otherwise represses production of Oct4, Nanog and Sox2	Loewer et al., 2010; Wang et al., 2013
	H19	let-7	Pro-oncogenic due to sponging of let-7	Kallen et al., 2013
Signaling pathways	*NKILA*	NF-κB	Negative regulation by binding to NF-κB/ IκB complex and masking phosphorylation sites	Liu et al., 2015
	lnC-DC	STAT3	Binds to and promotes phosporylation of STAT3	Wang et al., 2014

Table 1.1: Cytoplasmic lncRNAs, their interacting partners and functions

1.1.4.4 Nuclear Long non-coding RNAs

Nuclear long non-coding RNAs and their mechanism of action have been the subject of extensive study for a few decades now. These lncRNAs can regulate multiple events within the nucleus including regulation of gene expression, nuclear compartmentalisation and organisation thereby driving biological processes. These functions are carried out in association with partner molecules.

LncRNAs often contain sequence motifs through which they can bind to certain nuclear factors, which promote the nuclear localisation and function of the lncRNA. For example, the lncRNA maternally expressed gene 3 (*MEG3*) contains a 356-nucleotide nuclear retention element that associates with U1 snRNP, which retains MEG3 in the nucleus (Azam, S *et al*, 2019). Repeat elements are likely to have roles in the nuclear retention of lncRNAs as well. Using the high-throughput massively parallel RNA assays (MPRNA), Shukla CJ *et al* (Shukla,C.J et al, 2018) uncovered a C-rich sequence that can promote the nuclear retention of lncRNAs. These C-rich sequences from Alu repeats function through their association with the nuclear matrix protein heterogeneous nuclear ribonucleoprotein K (hnRNPK) (Lubelski and Ulitsky, 2019). Examples of other repeat sequences involved in nuclear retention of lncRNAs include the many unique repeats

found in the lncRNA functional intergenic repeating RNA element (FIRRE), ranging in length from 67 to 804 bp, termed repeating RNA domains (RRDs), which establish *FIRRE* chromatin localisation by interacting with hnRNPU (Hacisuleyman, E *et al*, 2016)

1.2 LncRNA in gene regulation

LncRNA loci can regulate target loci through three major principles : (a) either the RNA could be the functional biomolecule that interacts with other components in the cell, such as DNA, RNAs or proteins, (b) a gene regulatory element embedded in the transcription body of a lncRNA gene could direct the activity of the regulatory element through the activity of the lncRNA gene or (c) the process of transcription from the lncRNA locus can influence gene activity (**figure 1.6**). A lncRNA locus can have one of these functions or a combination of them (Ali, T and Grote, P, 2020). The role of the RNA biomolecule in regulating target gene transcription is explored in the following section.

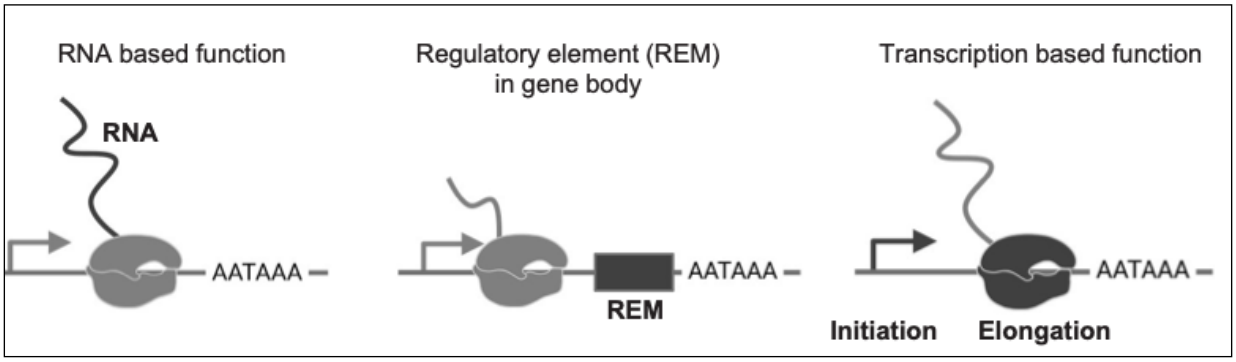


***Fig 1.6**: The three possible functional properties of lncRNA loci - LncRNA loci can exert their regulatory effects on the target loci through three possible mechanisms- the RNA transcript itself, through a regulatory element present in the gene body or through the act of transcription at the lncRNA locus. Figure has been adapted with permission from Ali, T and Grote, P, 2020*

1.3 Protein mediated chromatin interaction

Numerous nuclear lncRNAs localise on chromatin, where they interact with proteins, that facilitate or inhibit their binding and activity at the targeted DNA regions. LncRNAs are known to associate with numerous Chromatin modifying enzymes (CHE) which covalently modify histones or DNA. The binding of lncRNA to CHEs can either be to guide them to aor sequester them away from their target loci to either activate or repress target genes.

The binding and recruitment of the Polycomb Repressive Complexes (PRC) by lncRNA has been particularly well described. One of the first lncRNAs shown to be associated with PRC2 was *Xist,* a lncRNA associated with X-chromosome inactivation (XCI) in mammals. *Xist* localises to the same chromosome from which it is transcribed and leads to the establishment and maintenance of repressive histone modifications H3K27me3 by PRC2 in *cis.* It was shown through RNA immunoprecipitation (RIP) and qualitative electrophoretic mobility shift assays (EMSA) to co-precipitate with and bind to the PRC2 proteins EZH2-EED-SUZ12 through an A-repeat (RepA) region. PRC2 was also observed to bind to *Xist*'s antisense transcript *Tsix.* It is probable that *Tsix* prevents RepA-PRC2 action in pre-XCI cells by titrating RepA away from PRC2, by blocking RepA-PRC2 transfer to chromatin, or by preventing PRC2 catalysis (Zhao, J et al, 2008). A more recent study showed that another region downstream of exon1 of *Xist*, termed as XN, binds to and recruits non-canonical PRC1 complex required to initiate H2AK119u1 deposition on the inactive X chromosome (Xi). Additionally, the authors propose that H3K27me3 modification by the PRC2 complex is dependent on H2AK119u1 modification by PRC1 (Almeida, M et al, 2017) contrary to earlier studies which suggested that H2AK119u1 modification by PRC1 to be a downstream event to H3K27me3 modification. **Figure 1.7** below summarises the two models of histone modification catalysed by *Xist* lncRNA

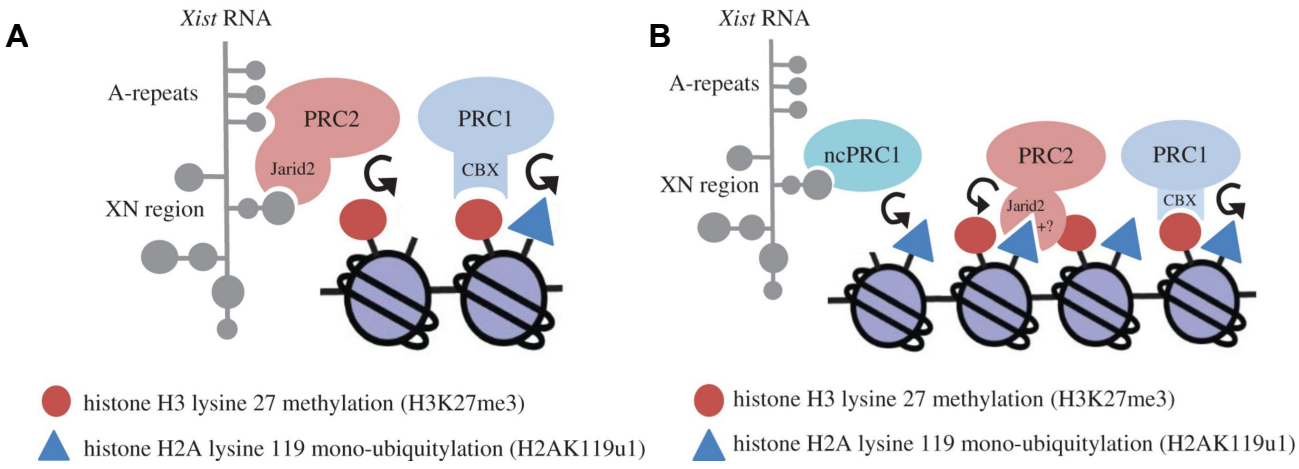


Figure 1.7. Classical (A) and Revised (B) models for Polycomb recruitment by Xist RNA *. (A) Earlier studies proposed direct interaction between core PRC2 subunits and the A-repeat element in Xist RNA. Subsequent studies implicated the Xist XN region and the PRC2 cofactor Jarid2 in initiating PRC2 recruitment which catalyses H3K27me3 on underlying nucleosomes. PRC1 recruitment is indicated as occurring downstream through interaction of the PRC1 subunit CBX and PRC2 mediated H3K27me3. Recruitment of PRC1, in turn, mediates H2AK119u1 deposition. (B) Revised model for Polycomb recruitment by Xist RNA proposes that the Polycomb cascade is initiated by non-canonical PRC1 complexes that are recruited by the Xist XN region. PRC1 mediated H2AK119u1 deposition serves to recruit PRC2 either through recognition by the cofactor, Jarid2 or through an alternative but currently undefined pathway. PRC2 mediated H3K27me3 then*

signals recruitment of canonical PRC1 complexes, further reinforcing H2AK119u1 deposition and Polycomb domain formation. Adapted with permission from Brockdorff, N. 2017.

LncRNA *HOTAIR* is expressed from the HoxC locus in chromosome 12, leading to transcriptional repression of genes at the HOXD locus spanning >40 kb of chromosome 2. *HOTAIR* was shown to be required for PRC2 occupancy and H3K27me trimethylation at target genes within the HOXD locus. Based on RNA immunoprecipitation studies, HOTAIR was shown to bind to and recruit PRC2, *in trans*, to the HoxD locus through direct and specific protein–RNA interactions (Rinn, J.L *et al*, 2007). The lncRNAs *Kcnq1ot1* which is involved in imprinting, Braveheart (*Bvht*), *Air* and *ANRIL* too are a few lncRNAs that follow similar mechanisms and recruit PRC2 to target loci (Davidovich C and Cech T.R, 2015).

LncRNAs can also bind to gene activating chromatin modifying complexes. The WDR5–myeloid/lymphoid or mixed-lineage leukaemia (MLL) histone methyltransferase complex facilitates gene expression through H3K4me3 mark deposition. WDR5 can bind RNA directly. *HOTTIP* lncRNA directly binds WDR5 and thus targets MLL/WDR5 histone H3K4 methylation complexes to the HOXA gene locus, leading to H3K4 trimethylation and HOXA gene transcription (Wang K.C et al, 2011). The enhancer-like lncRNA *NeST* causes all phenotypes conferred by the murine viral susceptibility locus Tmevp3 which harbours the NeST gene including higher lncRNA expression, increased interferon-γ abundance in activated CD8+ T cells, increased Theiler's virus persistence and decreased *Salmonella enterica* pathogenesis. *NeST* RNA binds WDR5 to alter histone H3K4 methylation at the IFN-γ gene locus and regulates IFN-γ gene transcription and susceptibility to viral and bacterial pathogens (Gomez, J.A et al, 2013).

Besides covalent modifications of histones and DNA, lncRNA can also be part of complexes that perform nucleosome remodelling. Based on their mechanisms of action, ATP-dependent remodelling complexes are of 4 classes - switching defective/sucrose nonfermenting (SWI/SNF), imitation switch (ISWI), chromodomain helicase DNA binding (CHD), and inositol requiring 80 (INO80) (Clapier, C.R and Cairns, B.R , 2009).

LncRNAs have been reported to directly bind to subunit of SWI/SNF complexes to either guide the complex to their target or decoy them away from their target. LncRNAs are likely to act as a scaffold for the assembly of the remodelling complexes. Nuclear paraspeckle

assembly transcript 1 (*NEAT1*) is a nuclear-restricted lncRNA that scaffolds the formation of paraspeckles and is misregulated in various human cancers. SWI/SNF complexes are found to be enriched in paraspeckle subdomains and are key components that facilitate the organisation of these subnuclear structures. *NEAT1* directly interacts with the SWI/SNF core unit, BRG1 or BRM, to form the paraspeckle structure, which leads to cell cycle arrest and affects cancer progression. While the nucleosome remodelling activity of SWI/SNF complex is not required for paraspeckle formation, these complexes are likely to act as a part of the structural foundation of paraspeckles. The paraspeckle complex likely decoys the SWI/SNF complex away from their target genomic sites (Kawaguchi, T *et al*, 2015). The lncRNA *MALAT1* guides the SWI/SNF complex to the promoter site of target genes IL-6 and CXCL8 by interacting with the catalytic subunit of the complex, BRG1. Thus, *MALAT1* facilitates NF-κB to induce the expression of these inflammatory factors in hepatocellular carcinoma (Huang et. al., 2018)

The SRCAP remodelling complex from the INO80 family regulates chromatin structure by altering the composition of nucleosomes. In murine embryonic stem cells (mESCs), *lncKdm2b,* a divergent lncRNA for *Kdm2b* gene, interacts with the ATPase subunit of the SRCAP complex. This association increases ATPase activity and promotes complex integrity. This interaction is proposed to facilitate SRCAP to drive expression of the transcription factor *Zbtb3* to maintain mESC pluripotency through upregulation of Nanog (Ye, B. *et al*, 2018).

Targeting of another nucleosome remodeler, the NuRD complex, involves lncRNA interaction as well. The ATPase of the NuRD complex can be one of two related ATPases: CHD3 or CHD4 of the CHD family. The NuRD complex generally operates as a transcriptional repressor by coupling nucleosome sliding with histone deacetylase activity in association with HDAC enzymes. Upon hypotonic stress induction, the NuRD complex is recruited to rRNA genes through an interaction of CHD4 and the lncRNA *PAPAS*. *PAPAS*-dependent NuRD recruitment results in histone H4 acetylation loss and repositioning of nucleosomes at the rDNA promoter to downregulate transcription of rDNA (Zhao, Z. *et al*, 2016, Patty, B.J and Hainer, S.J , 2020).

1.4 Non-protein mediated chromatin interaction

All of the interactions of lncRNA with chromatin discussed above are protein mediated. An alternative means of interaction is a direct interaction of lncRNA with DNA through the formation of hybrid structures - either through the formation of R-loops or DNA:DNA:RNA triple helices.

R-loops are triple-stranded nucleic acid structures consisting of a DNA–RNA hybrid and a displaced single-stranded DNA. Though initially considered a threat to genome stability, the transient nature of R-loops makes them well-suited for regulation. Ginno et al. demonstrated that R-loop formation is involved in gene regulation via its potential to protect DNA from methylation. Most often, R-loops form *in cis* during transcription when a nascent RNA hybridises to the DNA template behind the moving RNA polymerase. Work done in yeast suggests that RNA:DNA hybrids can form *in trans* as well (Ginno, P.A. et al, 2012, Guh, C-Y et al, 2020).

GATA3-AS1 is a lncRNA transcribed antisense to the *GATA3* gene, the master transcription factor for TH2 lineage commitment, and is necessary for efficient *GATA3* transcription. The lncRNA forms an R-loop at *GATA3* gene locus and also binds to components of the MLL methyltransferase recruiting it in *cis* and tethering the chromatin modifier to the chromatin (Gibbons, H.R et al, 2018). Similarly, the *VIM-AS1* lncRNA forms an R-loop at the *VIM* (vimentin) locus and is responsible for an R-loop dependent transcriptional enhancement. The R-loop supports local chromatin de-condensation and enhanced binding of the transcriptional activators of the NF-KB pathway to the VIM promoter (Boque-Sastre, R. *et al*, 2015).

The possibility of a triple helix structure was proposed long ago by Felsenfeld *et al*. (Felsenfeld, G.*et al,* 1957), whereby two pyrimidine strands and one purine strand could interact to form a complex structure. Over the years, the triplex went from being a theoretical possibility to reality with scientific

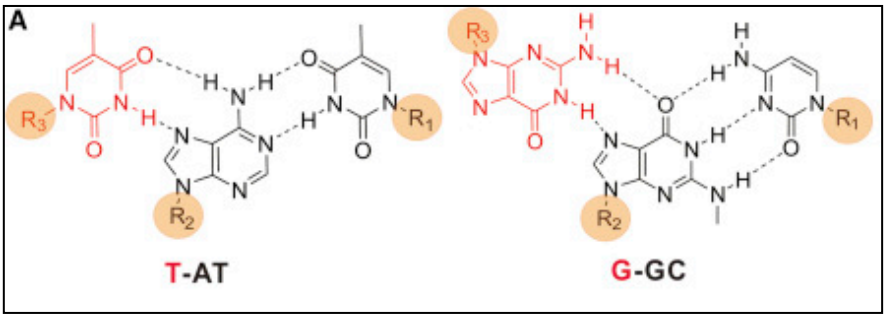


Fig 1.8 : Hoogsteen and Reverse- Hoogsteen base pairing in triple helix formation*. Adapted with permission from Li, Y. et al, 2016*

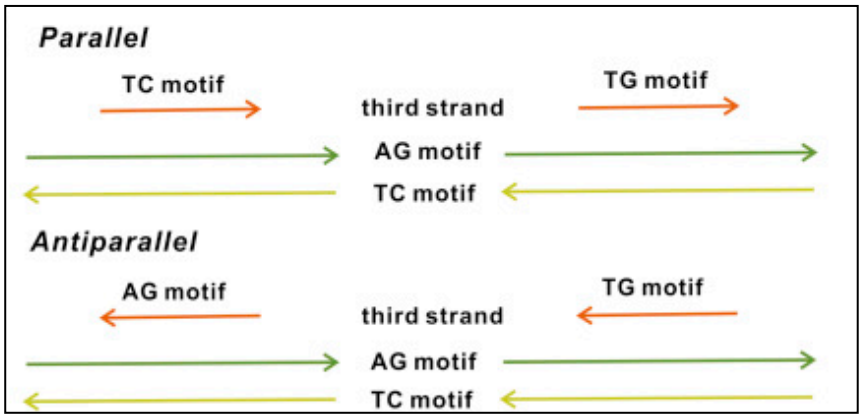


Fig 1.9* : *Parallel and anti parallel triplex. *Arrow indicates 5' - 3' orientation of stand. Adapted with permission from Li, Y. et al, 2016*

evidence to support its existence. In the triplex structure, a third strand can insert into the major groove of the duplex structure with sequence specificity (Radhakrishnan, I. and Patel, D. J., 1994) and each triplet can be formed either by the formation of Hoogsteen or Reverse-Hoogsteen hydrogen bond between the third RNA strand and the Watson-Crick based paired DNA duplex as shown in **figure 1.8**. Based on the nucleotides involved in the base-pairing and their orientation, triplexes can be parallel or anti-parallel with T(U)C motifs preferring to form parallel triplexes and AG motifs preferring antiparallel triplexes as shown in **figure 1.9**.The Hoogsteen hydrogen bond is weaker than the Watson-Crick bonding and requires the presence of multivalent cations such as Mg2+ *in vitro* to neutralise the charge repulsion among three negatively charged nucleic acid strands. Studies suggest that the RNA third strand forms more stable triplexes than its DNA counterparts (Li, Y. *et al*, 2016).

In silico predictions of triplex forming motifs show that these motifs cluster in regulatory regions of the genome, especially within promoters, suggesting a regulatory role for these non-canonical structures (Buske, F.A et al, 2012). Since very specific conditions of pH and ionic concentrations are required *in vitro* for triplex formation, there was scepticism associated with the formation of *in vivo* triplex structures. However, the use of a triplex specific antibody, specifically an antibody with higher affinity for DNA:RNA triplexes over DNA:DNA triplex, for immunofluorescence and immunoprecipitation studies, has strongly supported triplex formation within the nucleus (Ohno, M et al, 2002, Mondal, T. et al, 2015).

It has been suggested that the formation of triple helix is a generic mechanism used by lncRNAs for the sequence-specific targeting of DNA sequences. Additionally, triplexes themselves could act as platforms for the recruitment of some epigenetic regulatory proteins. Triplex formation by lncRNAs can either silence or activate transcription of their

target genes. *pRNA* (promoter- associated RNAs) is a non-coding RNA complementary to the rDNA promoter RNA that originates from an RNA polymerase I (Pol I) promoter located in the intergenic spacer ~2 kb upstream of the pre-rRNA transcription start site (TSS). *pRNA* binds to the promoter of rDNA through the formation of a triplex and this leads to the recruitment of DNMT3b to methylate the DNA at the promoter. The methylated CpG at the rRNA promoter silences rRNA transcription and subsequently leads to the hetero-chromatinisation of this region (Shmitz, K.M. *et al*, 2010) (**figure 1.10**). Similarly, *PARTICLE* lncRNA, which is transcribed from the bidirectional promoter of the MAT2A gene, forms a triplex upstream of the CpG island. The formation of the triplex suppresses MAT2A transcription through the recruitment of G9a methyltransferase and PRC2 chromatin modifier complex (O'Leary, V.B. et al, 2015).

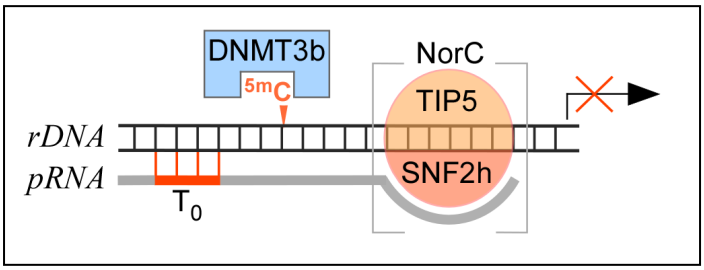


Fig 1.10: Illustration of rDNA gene silencing by the non-coding pRNA *binding through triplex interaction at one end and to the nucleolar remodeling complex (NorC) silencing complex at the opposite end, cytosine methylation catalyzed by DNA cytosine-5-methyltransferase 3b (DNMT3b) recited by the triplex. Adapted with permission from Bacolla, A. et al, 2015*

Another lncRNA *MEG3* binds to and regulates multiple genes involved in the TGFB pathway through the formation of triplex (Mondal, T. et al, 2015). *MEG3* lncRNA interacts with the chromatin-modifying Polycomb Repressive Complex 2 (PRC2) and guides the complex to genomic sites via DNA-RNA triplex formation (Sherpa et al., 2018). More Recently, the lncRNA *CISAL* was found to form DNA:DNA:RNA triplex at the promoter of the BRCA1 gene in carcinomas to inhibit BRCA1 expression by counteracting the binding of the transcription factor GABPA at the promoter (Fan, S. et al, 2020). Similarly, the lncRNA *HITT* associates at the HIF-1α promoter through triplex formation, recruits the chromatin modifier EZH2 to the gene promoter and suppresses it. This repression is relieved under hypoxic conditions found in cancer (Wang, X. et al, 2020). *KHPS1* is an antisense lncRNA transcribed from the bidirectional promoter of the proto-oncogene *SPHK1* and forms a triplex upstream of the promoter. The triplex formation is indispensable for the activation and expression of the poised enhancer of the *SPHK1* gene. The triplex formation recruits E2F1 and p300 and leads to the expression of the eRNA-SPHK1 which is essential for SPHK1 mRNA synthesis (Blank-Giwojna, A. et al, 2019). Interestingly, interchanging the Triplex Forming Region (TFR) between lncRNAs

targets them to the chromatin loci containing the Triplex Target Site (TTS) for the TFR and not the lncRNA's original putative targets (Blank-Giwojna, A. et al, 2019). Most lncRNA:DNA triplex have been reported to from between genomic regions rich in AG motif repeats and AG rich sequences within the RNA. Some of the TFR sequences for lncRNAs have been listed in **table 1.2**. While many more lncRNAs have been reported to interact with chromatin through triplex formation, the biological roles of these triplexes have not been explored.

LncRNA	Triplex forming Motif (5'-3')
pRNA	GUCGACCAGUUGUUCCUUUG
KHPS1	UCCCCCUUUUUUUUUCCUCCU
PARTICLE	AAGGGGGGGGGAA
Meg3	CGGAGAGCAGAGAGGGAGCG
CISAL	CAGCCCCUUACCCACCCCCU
HITT	GAAGGAAGAGAAAGGGG

Table 1.2 : LncRNA involved in triplex formation and the Triplex Forming Motifs harboured within them

1.5 LncRNA in the formation of nuclear bodies

The role of liquid-liquid phase separation (LLPS) in the compartmentalisation of processes within the nucleus is gaining increasing attention. The formation of membrane-less organelles such as stress granules and P-bodies in the cytoplasm and nucleoli and chromatin territories formation within the nucleus has been attributed to LLPS. LLPS occurs when both phases have liquid-like properties, such as high mobility of the individual molecules, and can occur through low-affinity, high-valency interactions between disordered regions of proteins or regions of nucleic acids - the primary components of LLPS are molecules that are capable of multivalent interactions. The key features of LLPS bodies include (a) the cellular bodies are demixed (phase-separated) from the surrounding nucleoplasm or cytoplasm, (b) they can fuse and become a larger droplet, and (c) can break down into smaller droplets under shearing forces. The nucleus is full of proteins that contain intrinsically disordered regions (IDRs) and are capable of phase-separating when clustered locally at a concentration above a threshold. There is a spontaneous aggregation

of IDRs and other interacting partners which leads to the formation of condensates (Hildebrand, E.M and Dekker J, 2020, Statello, L *et al*, 2020).

Many nuclear bodies are built on lncRNAs that act as scaffolding molecules and their formation has been attributed to LLPS. Chujo and Hirose have termed such lncRNA as 'architectural lncRNA' (arcRNAs) and state that lncRNAs categorise as such if 1) they are localised and enriched in a specific nuclear body, and 2) they construct and stabilise the body structure (Chujo, T and Hirose, T, 2017).

The 23kb long mammalian lncRNA Nuclear Paraspeckle Assembly Transcript 1 isoform 2 (*NEAT1_2*) or *NEAT1 long* is one such arcRNA that organises and is key to the formation and functioning of the nuclear paraspeckle. The knockdown of *NEAT1_2* leads to the disintegration of paraspeckles. Paraspeckles are thought to function as molecular sponges to sequester proteins and RNA to modulate their function outside the paraspeckle. They contain more than 60 proteins, including RNA- binding proteins and transcription factors, within them. The expression of *NEAT1_2* is upregulated upon cellular stress and in various cancers. The middle region (8–16.6 kb) of *NEAT1_2* contains two subdomains (12–13 kb and 15.4–16.6 kb) responsible for recruiting the core paraspeckle proteins NONO and SFPQ to initiate the assembly of paraspeckles. This region of *NEAT1_2* possibly initiates LLPS by associating with NONO and SFPQ. The Prion-like domains (PLDs) of paraspeckle proteins such as FUS and RBM14 are additionally required to construct paraspeckles. Paraspeckle formation is initiated at the site of *NEAT1_2* transcription - another key feature of arcRNAs and nearly 50 molecules of NEAT1_2 are required for the formation of each paraspeckle. The formation of a paraspeckle is summarised in **figure 1.11** below (Yamazaki, T *et al*, 2018, Nakagawa, S et al, 2018, Statello, L *et al*, 2020).

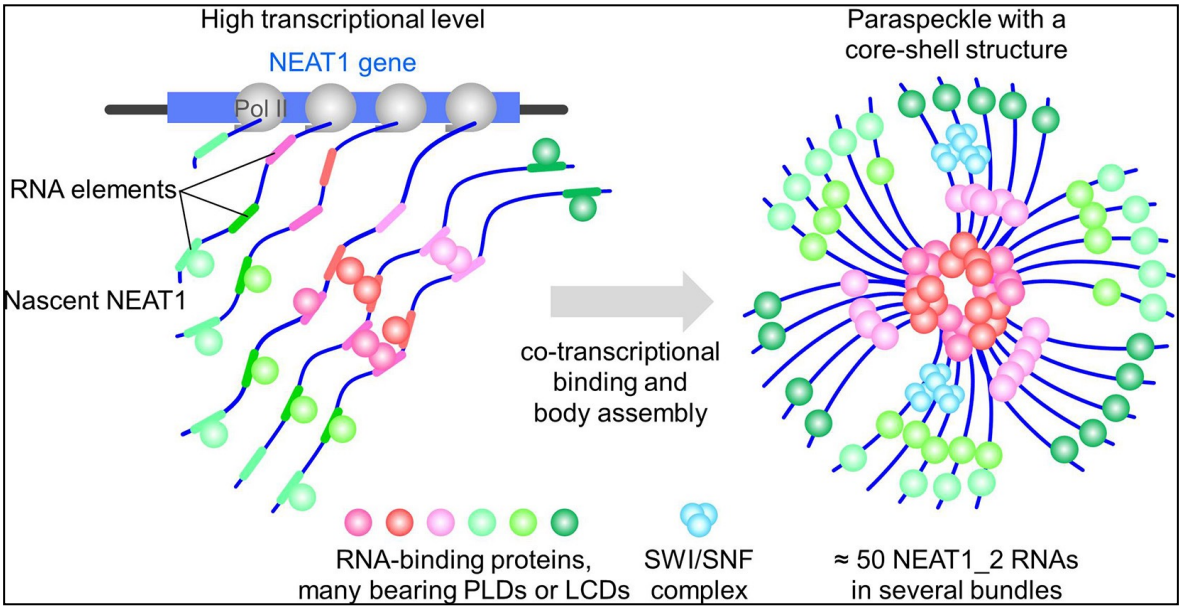


Fig 1.11: NEAT1_2 in paraspeckle assembly via LLPS - *High levels of NEAT1_2 transcription during cellular stress is coupled with the binding of specific paraspeckle proteins (PSPs) to initiate Paraspeckle assembly at NEAT1_2 transcription sites. PSPs stabilise NEAT1_2 and/or promote interactions with other PSPs and SWI/SNF complexes. Each paraspeckle contains approximately 50 NEAT1_2 molecules that are organised in several bundles, forming a core-shell spheroidal structure. Figure adapted with permission from Chujo, T and Hirose, T, 2017*

Interestingly, proteins of the DExD-box helicase family have been demonstrated to undergo LLPS to compartmentalise various RNA processing events in a rare example of cell organisation that is conserved from prokaryotes to humans. They further regulate the flux of RNA molecules in and out of these phase separated condensates (Hondele, M et al, 2019).

1.6 LncRNA in nuclear organisation and architecture

The essential role of RNA in nuclear organisation was shown nearly 3 decades ago by Nickerson J A *et al* who showed that upon inhibition on RNA synthesis by Actinomycin D or RNase A mediated RNA degradation, there was a decay of chromatin and nuclear matrix organisation (Nickerson, J. A. *et al*, 1989). Recent studies to investigate roles of specific lncRNAs have shown that lncRNA-protein complexes play important roles in mediating inter- and intra-chromosomal contacts, gene regulation and chromatin compartmentalisation. Transcription can be regulated by 3-dimensional genome organisation by bringing distal regions of the genome in close contact through the formation of loop domains. These interactions, particularly between elements like enhancers and promoters or other regulatory elements, can lead to context dependent

gene activation or repression. Many proteins in mammals have been implicated in enabling physical contact between different genomic elements including RNA polymerase II, the mediator complex and transcription factors such as GATA1 and KLF1. But one protein that has been receiving much attention for its ability to mediate chromatin contacts in recent times is CTCF (Zhang, K. *et al,* 2016).

1.7 CTCF

The 11 Zn-finger domain containing CTCF is a DNA-binding protein and has been implicated in a myriad of regulatory functions within the nucleus. It was initially characterised as a transcriptional factor capable of both transcriptional activation and repression. It is the main insulator protein in vertebrates which means that CTCF can interfere with enhancer-promoter communication or buffer genes from chromosomal position effects caused by heterochromatin spreading. But subsequently, CTCF has been implicated in the control of cell proliferation and apoptosis, chromatin domain insulation, X-chromosome inactivation, prevention of oligonucleotide repeat expansion, and other chromatin processes. The binding of CTCF relative to gene regulatory elements has provided insights into its probable functions. The binding of CTCF at enhancer elements in a cell-type specific manner suggests that it could be involved in regulating lineage-specific transcription. Binding of CTCF has also been observed at promoters, insulators and boundary elements. CpG is present within the binding consensus for CTCF and so, CTCF binding is sensitive to DNA methylation. CTCF Target Sites tend to be conserved in evolution and occupancy is largely invariant across different cell types. The multi-functionality of CTCF is based on its ability to bind a wide range of diverse DNA sequences as well as to interact with cofactor proteins through the combinatorial use of its 11 zinc fingers (ZFs) (Kim, S. et al., 2015).

Partner proteins of CTCF also determine the function of CTCF at specific loci. Some of the reported protein partners of CTCF are Ying-Yang 1, cohesin, DDX5, CHD8 and Kaiso (Ong C-T and Corces V.J., 2014). **Figure 1.12** summarises the known interacting proteins of CTCF.

While enhancer blocking and insulation by acting as a chromatin barrier are the classical functions of CTCF, efforts to understand these roles in depth revealed that CTCF is capable of bringing in contact distant genomic regions, specifically regulatory regions, for

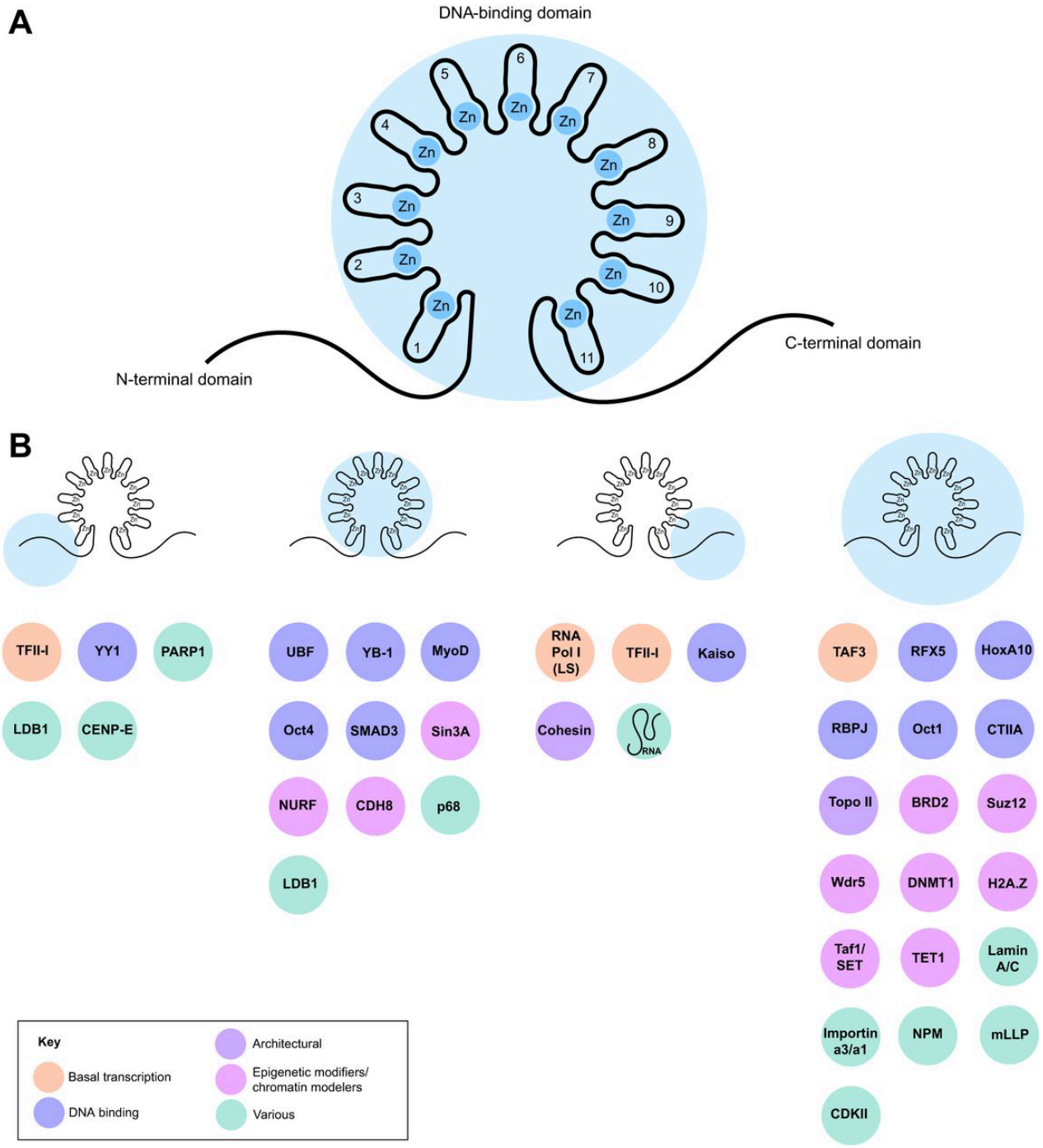


Fig. 1.12: Protein interacting partner of CTCF and the functions associated with them*. (A) The three major domains of CTCF: the N-terminal domain, the central zinc-finger domain containing Zn-fingers 1-11 and the C-terminal domain. (B) A variety of CTCF-interacting proteins are known to bind to specific domains of CTCF. While a handful of proteins interact with the N- and C-terminal domains of CTCF, multiple proteins interact with the zinc-finger domain. Additional proteins have been shown to interact with CTCF, although their binding has not been mapped to specific domains. Adapted with permission from Arzate- Meijia et al, 2018*

various functional outcomes and in turn, affect higher order chromatin organisation.

1.8 Cis-regulatory elements (CREs) in regulating gene expression

1.8.1 Silencers

Silencer elements, which act to inhibit gene transcription, have been defined in few genes including the cellular oncogene c-myc. These negative gene regulatory elements can contact the promoters through looping interactions and repress gene expression or through the binding of a multitude of Transcription Factors that repress transcription. As with activating sequences such as enhancers, some silencer elements are constitutively active whilst others display cell type specific activity. In addition to gene promoters, silencers, enhancers and insulators create a complex array of *cis*-regulatory elements (CREs). In Eukaryotes, silencers can be of two types- classical silencer elements and negative regulatory elements (NREs), also known as passive silencers. Classical silencer elements are position independent elements that can regulate transcription whereas NREs are position-dependent elements that direct passive repression. NREs are recognised in a large number of gene promoters, as well as in introns, exons and various flanking sequences. NREs can either physically inhibit the interaction of Transcription Factors with their specific DNA-binding sites, or interfere with specific signals which control various transcriptional events, such as splicing sites and 5' polyadenylation signals or affect transcriptional elongation (Ogbourne, S. and Antalis, T.M, 1998).

Enhancer-blocking (EB) elements, which prevent the action of an enhancer on a promoter when placed between the two, are a type of *cis*-acting NRE. Petrykowska et al showed the presence of multiple such Enhancer blocking silencers in the CFTR locus (Petrykowska, H. M, *et. al*, 2008). They further suggest that CTCF, the only protein that is known to bind barrier elements in mammals, is likely to bind to and aid in the functioning of EB elements due to the increased prevalence of CTCF motifs within these regions. It is likely that CTCF recruits other auxiliary proteins to EB regions.

One of the earliest and best-studied examples is the involvement of CTCF in enhancer blocking at the *H19/Igf2* locus. The capacity of CTCF–DNA complexes to form loops via protein dimerisation was also originally described for the *H19-IFG2* imprinted locus. Both the *H19* gene and the *Igf2* gene share a common endodermal enhancer. Through its

insulator function, CTCF ensures the expression of either *H19* or *Igf2* gene from each allele in mammals. More specifically, between the enhancer and the *Igf2* genes lies an Imprinting Control Region (ICR). CTCF binds to the ICR in a methylation sensitive manner. When bound at the ICR, CTCF silences the *Igf2* gene. The enhancer contacts the promoter of the *H19* gene and results in *H19* expression. DNA methylation of the ICR on the paternal allele prevents CTCF binding at the region and thereby leads to the expression of the *Igf2* gene. The insulating function of CTCF is dependent on protein cofactors such as the cohesin complex and the DEAD-box helicase DDX5/p68 (Yao, H. *et al*, 2010). Additionally, work done using circular chromosome conformation capture indicates that the *H19* ICR form various intra- and inter-chromosomal contacts, including with *Igf2* DMR1 locus, which are dependent on CTCF binding within the ICR (Burke, L.J *et al*, 2005). The regulation of the locus has been summarised in **figure 1.13**.

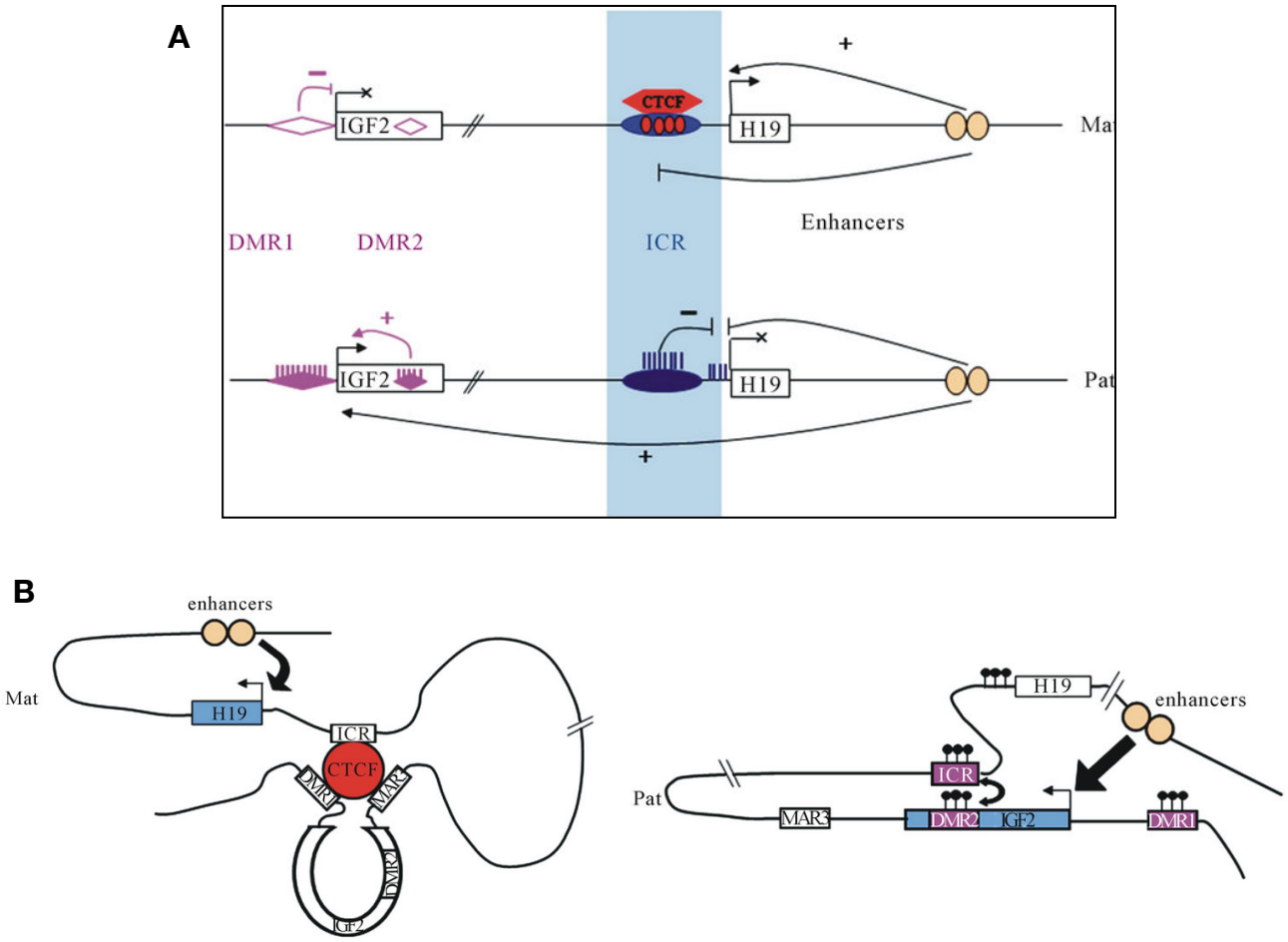


Fig 1.13: Regulation of H19/Igf2 Imprinted region by CTCF. *(A) Schematic representation of the H19/ Igf2 locus and the associated regulatory elements on the maternal and paternal alleles. (B) Looping interactions that regulate the H19/Igf2 locus with the involvement of CTCF - On the paternal chromosome, the H19 gene and the adjacent DMR are methylated preventing the binding of the insulator CTCF thus allowing the enhancer to access to the Igf2 gene resulting in its*

expression. In the absence of DMR methylation on the maternal chromosome, CTCF binds and prevents the enhancer from contacting Igf2, effectively silencing the gene. Instead, enhancer acts upon H19 gene resulting in its expression. Adapted with permission from Vennin C et al, 2013

The ability of CTCF to mediate DNA loops has been confirmed genome-wide approaches, solidifying the key role of CTCF in the organisation of chromatin architecture. CTCF-mediated chromatin loops were shown to connect enhancers with promoters, to insulate promoters from enhancers, to mediate imprinting of mammalian genes, to control V(D)J recombination, and to organise the major histocompatibility complex (MHC) class II genes. Because of CTCF's role in mediating chromatin contacts, it is considered an architectural protein (Ong, C-T. and Corces, V.J., 2014).

It has been reported that CTCF can bind to RNA molecules as well as other proteins. An analysis of CTCF associated RNA molecules via UV-crosslinking and immunoprecipitation followed by high-throughput sequencing (CLIP-seq) carried out in mouse embryonic stem cells revealed that CTCF can interact with a multitude of transcripts both protein-coding mRNA and lncRNA (Kung, J.T et al, 2015). Saldana-Meyer R et al, characterised the RNA binding Region (RBR) within CTCF. Disruption of the RBR abrogates chromatin binding ability of CTCF (Saldana-Meyer, R. et al, 2019).

At the H19/Igf2 locus, CTCF and cohesin binding is dependent on the presence of lncRNA steroid receptor RNA activator (*SRA*) to ensure proper expression of the genes from the two alleles. Depletion of *SRA* reduces CTCF-mediated insulator activity at the *IGF2/H19* imprinting control region and increased IGF2 expression (Yao,H *et al*, 2010). The lncRNA *Firre*, too, associates with CTCF and cohesin to position in inactive X chromosome at the nucleolus to maintain the repressive H3K27me3 modification (Yang, F. et al., 2015).

1.8.2 Enhancers

Enhancers are regulatory elements that control cell-type-specific spatiotemporal gene expression programmes by engaging in physical contacts with their cognate gene promoters. Based on putative enhancer mapping using cofactor binding and histone marks in mammalian cells, enhancers by far outnumber genes in mammalian genomes. In many cases, enhancers are located at great genomic distances from the target genes they control (in some cases hundreds of kilobases) and thus contact their cognate genes through long-range chromosome interactions. Studies have shown that enhancer–

promoter interactions are established concomitantly with gene expression, without being able to disentangle whether enhancer–promoter contacts are the cause or the effect of gene activation. Active enhancers are often transcribed to produce long non-coding transcripts called as enhancer RNAs (eRNAs) (Ong C-T and Corces V.J., 2011).

The lncRNA CCAT1-L is transcribed from a super-enhancer located 550kb upstream of the MYC gene locus. Upon transcription, CCAT1-L brings the enhancer, from which it is transcribed, in contact with the promoter of MYC gene to regulate its expression in a CTCF-dependant manner. In fact, the lncRNA physically associates with CTCF to mediate the enhancer-promoter loop formation (Xiang, J.F et al, 2014).

It has been proposed that, in addition to CTCF, eRNAs also interact with the Mediator complex to exert their activity. The depletion of Mediator abrogates DNA bending and subsequent enhancer-promoter loops resulting in diminished transcription of eRNA target genes This is supported by evidence from disease-causing mutations in Mediator subunit MED12, which cause the reduced ability to associate with eRNAs. Cohesin, known to form rings to connect two segments of DNA, has also been shown to stabilise eRNA-induced enhancer-promoter interactions and, therefore, influence target gene transcription (Shibayama, Y. et al, 2014, Schoenfelder, S. and Fraser, P., 2019).

1.9 Some CTCF partner proteins

1.9.1 BORIS

Brother Of the Regulator of Imprinting Sites (BORIS), also known as CTCFL, is a germ cell specific variant paralogy of CTCF that recognise the same DNA sequences. Both CTCF and BORIS are co-expressed in germ cells. They share an almost identical DNA binding domain. BORIS is also aberrantly expressed in a wide range of cancers. It was proposed that CTCF and BORIS compete for DNA binding at target sequences due to both having an almost identical DNA biding domain. But, Pugacheva et al demonstrated that BORIS occupies one-third of all the CTCF Target Sites and “sidestepped” the remaining two-thirds of them. The occupancy by BORIS *in vivo* could be determined by the DNA sequence and the site’s architecture (Pugacheva, E.M et al, 2015).

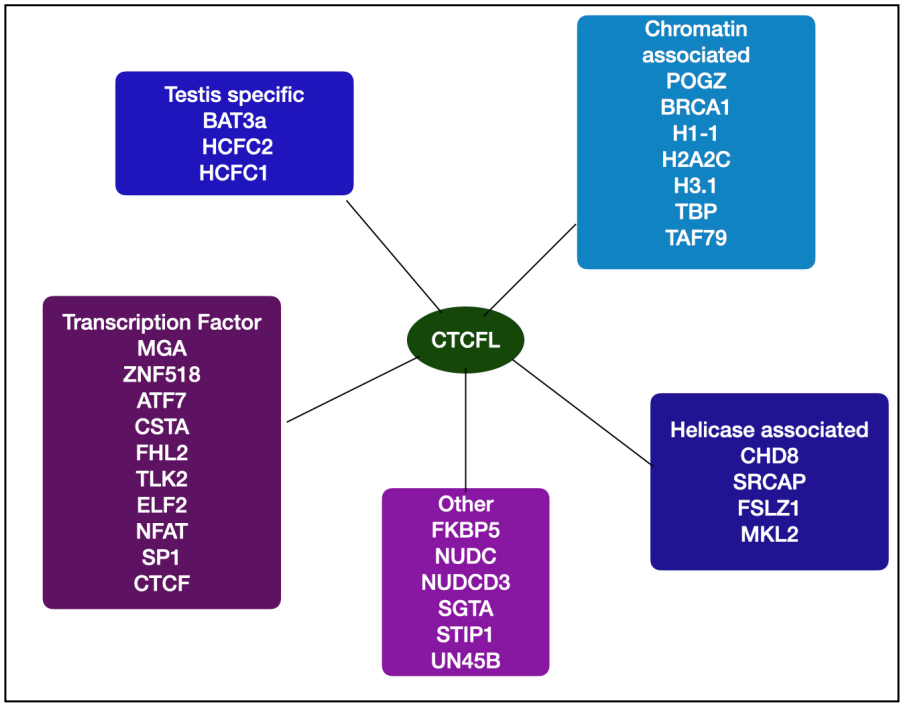


Fig 1.14: Protein interactions of CTCFL. *Protein interacting partners of CTCF L. CTCFL interaction network differs substantially in size from that of CTCF. Information adapted from Jabbari, K. et al, 2018, de Necochea-Campion, R. et al, 2011*

Though the two proteins have an almost identical DNA biding domain, due to their completely distinct amino and carboxyl termini, differences in biological functions between the two proteins are expected. An analysis of the protein interaction networks for both CTCF and BORIS shows that CTCFL interaction network differs substantially in size and in proteins from that of CTCF. While 52 interacting protein partners have been identified for CTCF, only 19 first-tier interacting partners of CTCFL have been identified. The dissimilarity of the protein-interaction networks indicates functional divergence between CTCF and BORIS. From this analysis, it can be seen that BORIS, unlike CTCF, does not interact with proteins such as cohesin or YY1 which are required for the architectural function of CTCF (Jabbari, K. et al, 2018). Additionally, a recent study showed that BORIS was unable to anchor cohesin to CTCF binding sites unlike CTCF (Pugacheva, E.M. et al, 2020). To date, established BORIS functions are limited to the transcriptional activation or repression of some germline and cancer-related genes.

1.9.2 Cohesin

Cohesin is a one of the structural maintenance of chromosomes (SMC) complexes that was initially identified for it role in holding together sister chromatids. The complex is

composed of two members of the structural maintenance of chromosomes (SMC) family of proteins—Smc1 and Smc3—and two additional subunits known as Rad21/Scc1 and SA/ Scc3 and is conserved in eukaryotes (Guillou E et al, 2010). The structure of the complex is depicted in **figure 1.15**. Topological entrapment of DNA is one of the remarkable features of cohesin that enables its involvement in various functions. SMC complexes can tether together two regions of DNA, either within a single DNA molecule or between DNA molecules

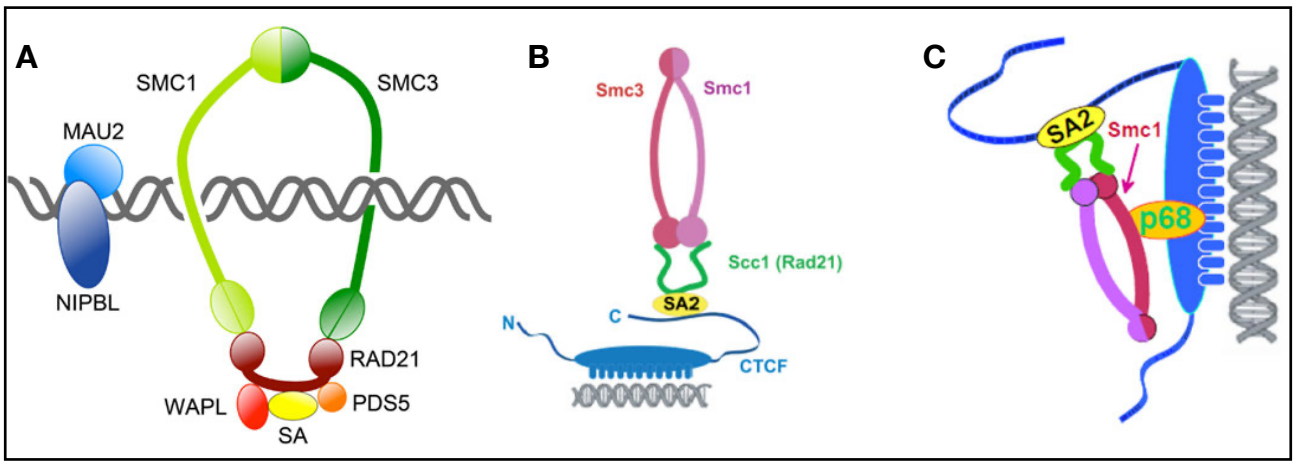


Fig 1.15: Structure and interaction of the cohesin complex. (A) *SMC complexes form ring-like structures. The cohesin complex consists of four core subunits: SMC1, SMC3, RAD21, and SA.These subunits form a large ring capable of topologically encircling DNA strands. The NIPBL/ MAU2 dimer loads cohesin onto DNA between the hinge domains of the SMC subunits of cohesin which may serve as an entry gate for DNA, whereas WAPL/PDS5 release cohesin from chromosomes by opening the SMC3-RAD21 interface. Image A has been adapted with permission from Horsfield, J.A. et al, 2012* ***(B)*** *The cohesin complex interacts with CTCF through the SA2 subunit* ***(C)*** *p68 bridges the interaction between SMC1 subunit of the cohesin complex and CTCF. Images B and C have been taken with permission from Giles, K. E. et al, 2010*

Several studies have implied that cohesin can serve as a basement of chromatin loops. Hi-C studies have reported that cohesin is located on the bases of loop domains and is required for the formation of the domain structure throughout the genome. In fact, degradation of Rad21 eliminates loop domains of hundreds kilobase in size and also affects higher-order genome compartmentalisation. Interestingly, cohesin complex harbouring different variant of SA subunit: either SA1 or SA2 seems to regulate different chromatin structures. Cohesin-SA1 preferentially contributes to TADs boundaries together with CTCF, while cohesin-SA2 promotes cell-type-specific enhancer-promoter contacts in a manner independent of CTCF (Nishiyama, T., 2019). Recently, it was demonstrated that CTCF interacted through its N-terminal region with and retained cohesin at nearly 95% of CTCF bound sites by stalling cohesin translocation on chromatin through a roadblock

mechanism. RNA-binding by the first CTCF Zinc finger may contribute to the formation of such a cohesin-blocking structure by creating additional steric constraints for cohesin ring sliding (Pugacheva, E.M. et al, 2020).

1.9.3 YY1

Another zinc-finger containing protein which is capable of enabling looping contacts is Ying-Yang1 (YY1). Biochemical studies have indicated that the zinc fingers of YY1 may interact with the N-terminus of CTCF, suggesting that YY1 could anchor loops via homodimerisation or heterodimerisation mechanisms. Additionally, the REPO domain of YY1 can interact with cohesin. YY1 and CTCF share many features: both are ubiquitously expressed, essential, zinc-coordinating proteins that bind hypo-methylated DNA sequences, and facilitate loop formation through the formation of homo- or heterodimers. The two proteins differ in that YY1 preferentially occupies active enhancers and promoters, while CTCF preferentially occupies sites distal from these regulatory elements that tend to form larger loops and participate in insulation. Similar to CTCF, YY1 has also been reported to bind to RNA and this interaction stabilises homodimers of YY1. Studying YY1 occupancy in Neuronal Progenitor Cells (NPCs) revealed that YY1-mediated looping interactions were nested within larger constitutive interactions anchored by constitutively occupied CTCF sites. YY1 binding is strongly enriched at promoters and enhancers in several mouse and human cell types. ChIA-PET and HiChIP experiments with YY1 identified it as a candidate to mediate structural interactions between promoters and enhancers. Specially, in NPCs it was observed that CTCF mediated loops gave way to YY1 enabled lineage specific promoter-enhancer contacts upon differentiation (Beagan, J.A .et al, 2017, Weintraub, A.S. et al, 2017).

1.10 Genome-Wide chromatin organisation and lncRNA

Chromatin inside the nucleus is efficiently packaged and organised in 3D space to allow for regulated expression, replication, recombination and damage repair. It has long been known that chromosomes occupy distinct positions within the nucleus. The 3D genome is organised into hierarchical layers, which have been postulated to represent structural and functional building blocks of genome organisation and is illustrated in **figure 1.16**. The arrangement of individual DNA bases defines local organisation and underlies DNA duplex

bendability and meltability. Interactions ranging from hundreds to a few thousand base pairs is defined by factors that bend and wrap DNA as well as the DNA supercoiling state and are the intermediate scale of genome organisation. The global scale of genome organisation ranges from many thousands of base pairs to entire chromosomes and is defined by how chromosomes are folded into loop domains and are organised spatially in the nucleus. Technological advances such as high-resolution microscopy and genome-wide contact mapping through Hi-C has allowed us to further understand the hierarchical organisation of nuclear chromatin. But more importantly, it has helped to bridge the gap between chromatin architecture and gene regulation suggesting that the two are interdependent. Mapping of genome wide contacts based using Hi-C data shows that the genome is segregated into two compartments within the nucleus: the transcriptionally active compartment called A compartments and the transcriptionally inactive B compartments (Rowley M.J and Corces V., 2018). Regions within A compartment preferentially interact with regions in other A compartments and the same is observed for regions within B compartments. Recent observations and analyses strongly suggest that one process contributing to compartmentalisation is LLPS. With decreasing sequencing costs making higher depth of sequencing possible, we now have access to chromosome contact maps at resolutions as low as 1kb (Rao S.S.P et al., 2014) with maps of 40kb resolution becoming routine. Using these maps to understand the directionality of chromatin interactions, it was discovered that chromatin form distinct Topologically Associated Domains or TADs.

TADs are regions within A or B compartments of a chromosome, of an average size of 1 MB in mammalians cells, which show preferential chromatin contacts within themselves rather than with other regions of the genome. Neighbouring TADs are separated by boundaries that are marked by the presence of CTCF binding and the presence of highly transcribed genes which are generally housekeeping genes. Boundaries of a single TAD appear to be demarcated by convergent CTCF binding sites. TADs contain within them smaller sub-TADs or contact domains which have a higher interaction frequency. It has been proposed that TAD boundaries impose a spatial insulation that plays a pivotal role in gene expression control by mediating or facilitating intra-domain enhancer–promoter contacts, while inhibiting cross-boundary communication between regulatory elements to prevent aberrant gene activation. Insulated neighbourhoods, loops formed between two convergent CTCF sites that are co-bound by cohesin (which together form the insulated

neighbourhood anchor), have been postulated to function as structural units of gene expression control (Schoenfelder, S. and Fraser, P., 2019).

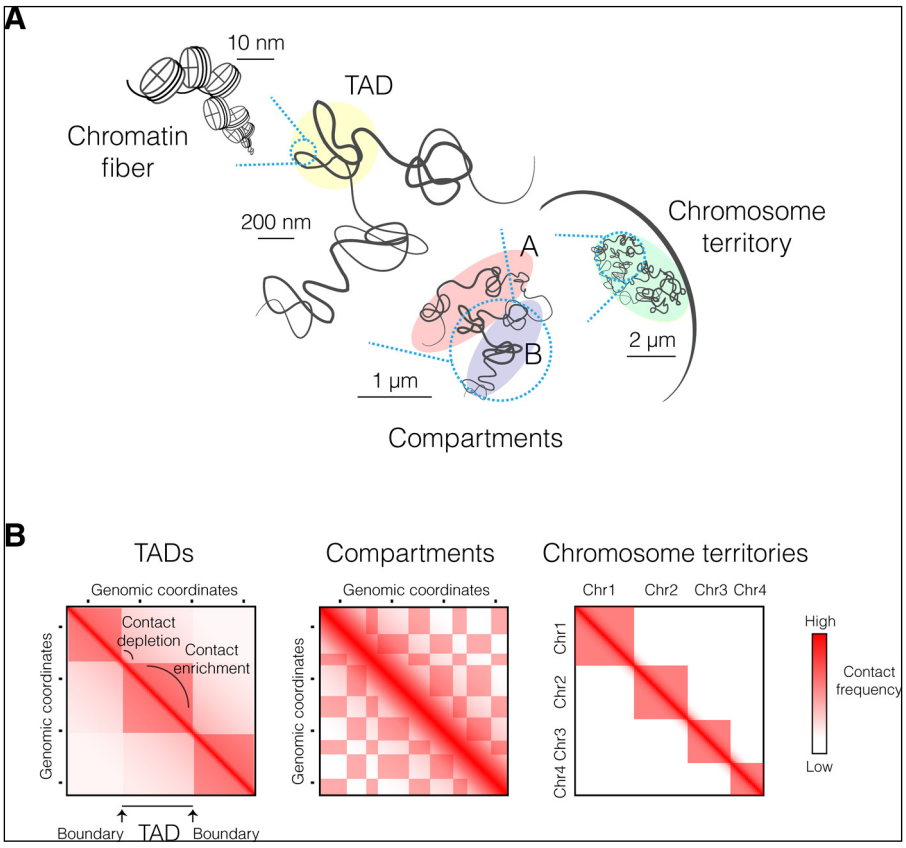


Fig 1.16: Hierarchical organisation of chromatin within the nucleus *(**A**) Schematic view of chromosome folding inside the nucleus. The ~11-nm chromatin fiber at the level of DNA histone associations is the finest layer of chromatin folding. Chromatin is packed at different nucleosome densities depending on the transcriptional status of the genes in the region and folds at the submegabase scale into higher-order domains with preferential internal interactions, referred to as TADs. At the chromosomal scale, chromatin is segregated into active "A" and repressed "B" compartments, and preferentially contacts chromatin regions of the same epigenetic states. Individual chromosomes occupy their own chromosome territories. (**B**) Schematic representation of Hi-C maps at different genomic scales. Genomic coordinates are indicated on both axes and the contact frequency between regions is represented by a color code. At the submegabase scale, TADs appear as squares along the diagonal enriched in interactions, separated by contact depletion zones delimited by TAD boundaries. At the chromosomal scale, chromatin long-range interactions form a characteristic plaid pattern of mutually excluded A and B compartments. Intrachromosomal interactions are overrepresented compared to interchromosomal contacts, consistent with the formation of individual chromosome territories. Image has been adapted from Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Sci Adv. 2019 Apr 10;5(4):eaaw1668. Reprinted with permission from AAAS.*

On a genome-wide scale, ChIA-PET experiments have helped to map inter- and intra-chromosomal interactions mediated by CTCF. Results from ChIA-PET studies have suggested that CTCF configures the genome into distinct chromatin domains that exhibit

unique epigenetic states establishing functional expression domains (Handoko, L et al, 2012). LncRNAs seem to have a role to play in the organisation and maintenance of higher order structures such as TADs.

HOTTIP is known to coordinate transcription of the 5′ tip of HOXA genes. Altered TAD might result in inappropriate promoter/enhancer interactions to alter transcription of oncogenes or tumor suppressors. *HOTTIP* acts as an epigenetic regulator to define oncogenic *HOXA* topologically associated domain (TAD) and drive *HOXA* associated leukemic transcription program. *HOTTIP* regulates a fraction of CTCF binding sites (CBSs) in the AML genome by directly interacting with CTCF and its binding motifs. Depletion of *HOTTIP* lncRNA impairs CTCF defined TADs in the Wnt target gene loci and reduces Wnt target gene expression (Luo, H. et al, 2019, *Cancer Cell*, Luo, H et al, 2019, *Blood*).

During XCI, the inactivated X chromosome globally loses A/ B-compartmentalisation and local TADs, and becomes partitioned into two large TADs that are hinged by a DNA region containing the DXZ4 macrosatellite. The boundary of the TADs are enriched for CTCF and the lncRNA *Firre.* Though the deletion of *Firre* has no effect on TAD boundary, CTCF binding reduces and the ectopic chromosomal interactions across the boundary increases suggesting a role for *Firre* in recruiting architectural proteins to the border (Barutcu, A. R. et al, 2018).

TADs also play important regulatory roles in Class-switch recombination (CSR) and somatic hypermutations (SHM)- two processes essential for antibody diversifications. Intra-TAD interactions during B cell development regulate antibody diversification. A ncRNA, ncRNA-*CSRIgA*, has been characterised to play important roles in controlling TAD interaction dynamics, recruitment of genome organisation regulatory factors at critical sites, and antibody gene diversification mechanisms. It does so by regulating IgA CSR within $TAD^{lncCSRIgA}$, the TAD that harbours the genes for IgA and lncRNA-CSRIgA in humans. Studies performed on the lncRNA-*CSRIgA* transcript assign it the role of recruitment of cohesin subunit SMC3 protein, FACT subunit SUPT16H, and PARP1 at the pivotal $CTCF^{lncCSRIgA}$ binding site of $TAD^{lncCSRIgA}$ (Rothschild, G et al, 2020).

1.11 Gene regulation by regulatory elements within the lncRNA transcription unit

Gene regulatory elements such as enhancers can be present within the body of lncRNA genes and the activity in this locus can influence the action of the regulatory elements. Unlike enhancer lncRNAs that are involved in bringing together the enhancer (locus from which they are transcribed) and the target promoters, the regulatory elements present within some genes are not dependent on the transcripts originating from the locus to facilitate the interaction (Ali, T and Grote P, 2020).

ThymoD lncRNA is transcribed from the enhancer for Bcl11b and has CTCF binding sites sent within the ThymoD gene body. Under the silenced conditions, the ThymoD gene is present close to the nuclear lamina. ThymoD transcription promotes demethylation at CTCF binding sites enabling CTCF binding. This activates cohesin-dependent looping to reposition the Bcl11b enhancer from the lamina to the nuclear interior to juxtapose the Bcl11b enhancer and promoter into a single-loop domain. This activation is lost when the transcription of ThymoD is blocked by insertion of a polyA signal after exon two and before the CTCF-binding site and, consequently, the CTCF-binding site is methylated. The ThymoD RNA itself is dispensable (Isoda, T et al, 2017).

In a similar mechanism, several antisense transcripts regulate their *cis* gene at the Proto-cadherin alpha (Pcdha) cluster. This cluster produces three distinct variants from three alternative TSSs to achieve stochastic expression of splice variants. The first exon of each of these variants also codes for an antisense lncRNA transcript each, named Pcdha-as, and the expression of the lncRNAs precedes the expression of the protein-coding gene. The lncRNA positively regulates the most distal gene's expression. Mechanistically, the expression of the Pcdha-as transcripts leads to the demethylation of a CTCF-binding site in the region upstream of the Pcdha gene, thereby allowing for a stable loop formation with a distal enhancer region (Ali, T and Grote P, 2020).

1.12 Gene regulation by the act of transcription of lncRNA

The act of transcriptional initiation or elongation of a lncRNA can have an essential role in regulating protein-coding genes in the vicinity.

For example, the lncRNA AIRN negatively regulates Igf2r through transcription. The TSS of the lncRNA AIRN (antisense Igf2r RNA non-coding) is located in the second intron of the Igf2r protein-coding gene, and AIRN is transcribed antisense to Ifg2r. The transcription of AIRN negatively regulates Igfr2 and when transcription of AIRN is blocked by a polyA insertion before the promoter of Igf2r, this negative regulation is abolished. This does not happen if the polyA is inserted after the promotor of Igf2r. Thus, the transcription of AIRN, and not the RNA product itself, is important for the regulation of the Igfr2 gene (Latos, P.A et al, 2012, Ali, T and Grote P, 2020). AIRN falls under the category of macro lncRNAs, a subclass of lncRNAs that show RNA hallmarks such as inefficient splicing, extreme length, high repeat content, lack of conservation and a short half-life. These features can be indicators that the lncRNA product is less important than the act of transcription (Guenzl, P. M., & Barlow, D. P. , 2012)

Transcriptional initiation is important for the *PVT1* lncRNA locus that causes the activation of the Myc oncogene. The *PVT1* locus encodes several transcripts with alternative start sites. *PVT1* promoter has the function of a DNA boundary element, blocking *MYC* oncogene from accessing cell-type-specific enhancers. The promoter of *PVT1* inhibits and limits *MYC* expression in *cis*. Transcriptional initiation of *PVT1* is important for this shielding of *MYC* gene (Cho, S. et al, 2018).

1.13 *Mrhl* long non-coding RNA

Meiotic Recombination Hotspot Locus (*Mrhl*) RNA is a 2.4kb long, mono-exonic, polyadenylated lncRNA discovered in our lab. It does not harbour a significant ORF when analysed bioinformatically and does not code for any peptide in *In vitro* coupled transcription and translation assay. In adult mice, this lncRNA shows tissue specific expression in spleen, liver, kidney and testis but not in the brain, heart or muscle cells. It is transcribed by RNA polymerase II from an independent transcriptional unit within the 15th intron of the mouse phkb gene in chromosome 8 and is syntenically conserved in humans (Ganesan, G and Rao, M.R.S, 2008, Nishant, K.T. et al, 2004, Fatima, R. et al 2019).

At 5.5 kb, human mrhl or *hmrhl* is larger than its mouse counterpart having acquired different repeat elements during evolution. Similar to *mrhl*, it shows varied levels of expression across tissues. It is highly expressed in pancreas, spleen and expressed at low

levels in the brain and skeletal muscles. The two share 65% identity over a stretch of 1223bp as shown in **Figure 1.17** (Fatima, R. et al 2019).

Fig 1.17: BLAST result aligning the sequences of mrhl and hmrhl*. The two share identity of 65% over a region of ~1.2 kb.*

This transcript has been extensively characterised in the mouse spermatogonial cells (Nishant, K.T. et al, 2004, Ganesan, G and Rao, M.R.S, 2008, Arun, G. et al, 2012, Akhade, V.S et al, 2014, Akhade, V.S et al, 2016, Kataruka, S. et al, 2017) and recently the function of *mrhl* in mouse embryonic stem cells and of the human homolog *hmrhl* have been explored (Pal, D et al, 2021, Choudhury, S.R et al, 2020). In the B-type spermatogonial cells (Gc1-spg cell line), the 2.4kb transcript is nuclear localised, shows distinct and punctate signal, and co-localises with Drosha and Nucleolin. The 2.4kb transcript gets processed to an 80-nucleotide (nt) intermediate RNA by the Drosha machinery both within the nucleus and in a cell-free system *in vitro* . Although Dicer can further process the 80-nt intermediate RNA to 22nt fragment *in vitro*, the mature 22-nt

miRNA derived from *mrhl* RNA is not detected *in vivo* in testicluar cells although it is still plausible that under certain physiological or pathological conditions or in certain tissues, this 22-nt species may be generated *in vivo* (Ganesan, G and Rao, M.R.S, 2008).

Mrhl is chromatin-associated within the nucleus which is indicative of a nuclear regulatory role for it. In an effort to understand its biological function within the nucleus, siRNA-mediated silencing of *mrhl* RNA in mouse spermatogonial Gc1-Spg cells was carried out and an expression array analysis was performed to study the genes that are perturbed upon *mrhl* RNA silencing. A weighted gene co-expression network was generated from the perturbed gene list showed that many of the important transcription factors and genes involved in the Wnt signalling pathway, including TCF4 were represented. TCF4, in association with beta-catenin, regulates the expression of Wnt target loci by binding at the Wnt Responsive Elements (WRE) in the target gene promoters modulating their expression by recruiting co-activators or co-repressors.

The Wnt effector molecule beta-catenin translocates into the nucleus under *mrhl* RNA downregulated conditions. Observations from experiments to identify the protein interacting partners of *mrhl* indicate that the DEAD-box helicase p68/Ddx5 interacts with *mrhl* RNA not only in the spermatogonial cells but also in liver, kidney and spleen. Ddx5/ p68 is a known RNA binding protein whose cytoplasmic translocation in its tyrosine-phosphorylated form is shown to stabilise cytoplasmic beta-catenin. p68 is essential in translocating beta-catenin to the nucleus upon *mrhl* RNA downregulation and so, plays an important role in the regulation of Wnt signalling. The role of *mrhl* RNA may be to serve as a scaffold for p68 protein binding and keeping protein in a de-phosphorylated state, thereby retaining it in the nucleus. Thus, *mrhl* RNA functions as a negative regulators of Wnt signalling in spermatogonial cells (Arun, G. et al, 2012) . This mechanism has been summarised in **figure 1.18**.

Mrhl lncRNA itself is downregulated upon induction of Wnt signalling pathway with Wnt3a ligand. The downregulation of *mrhl* is mediated by the occupancy of β-catenin at the TCF4 binding site present in the upstream promoter region of the gene and is dependent upon the recruitment of the co-repressor Ctbp1 at this promoter region. Interestingly, markers of meiotic commitment including Stra8, c-Kit and Hspa2 are upregulated following Wnt3a ligand treatment of Gc1-Spg cells and the dowregulation of *mrhl* RNA upon Wnt3a ligand

treatment is essential for the upregulation of these marker gene expression (Akhade, V.S et al, 2016).

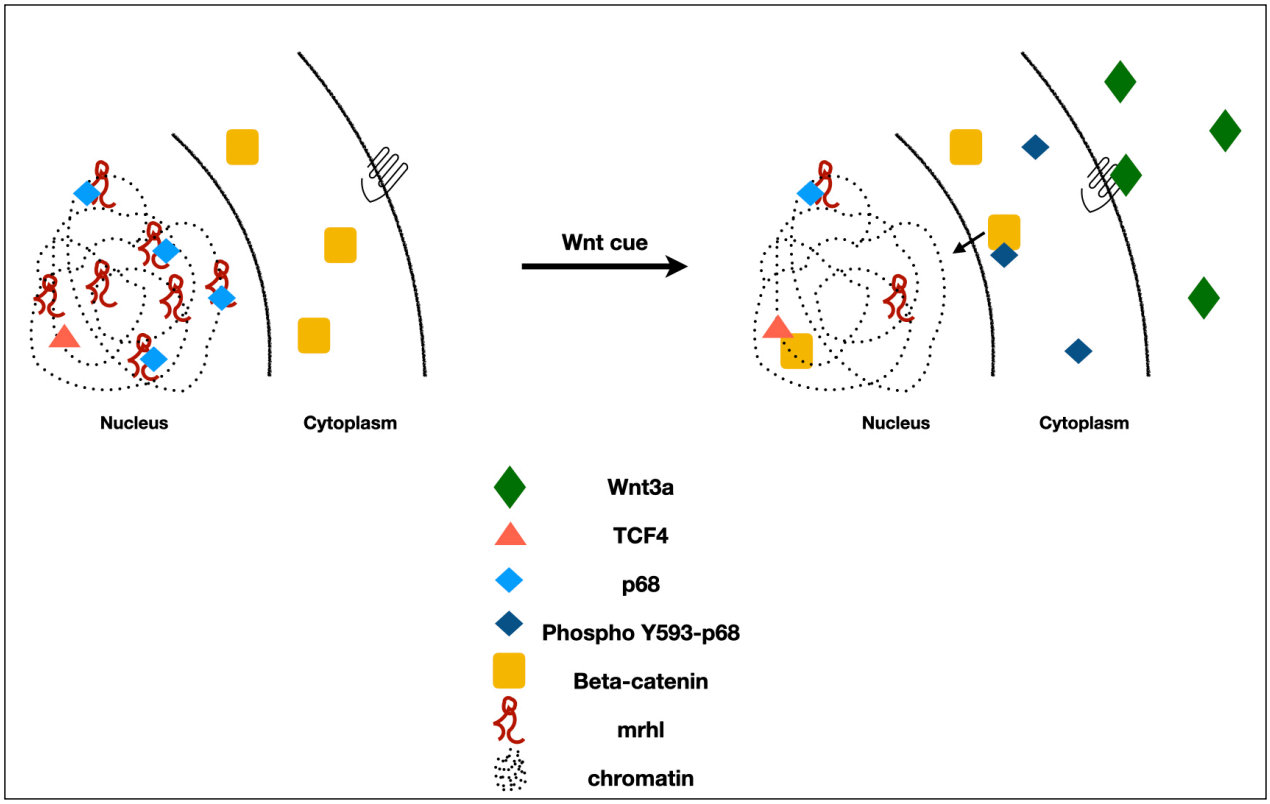


Fig 1.18: Regulatory function of mrhl lncRNA inside spermatogonial cells. *Mrhl lncRNA participates in the nuclear retention of p68 when expressed. Under these conditions, beta-catenin in localised to the cytoplasm and nuclear TCF4 is bound at Wnt responsive element present in promoter of the target genes. Wnt signalling induction results in mrhl downregulation and p68 shuttles between the nucleus and cytoplasm to aid in the nuclear translocation of beta-catenin. Beta-catenin joins TCF4 at WRE and regulates expression of target genes.*

Since *mrhl* is a chromatin bound RNA whose silencing perturbs the expression of many genes, genome wide chromatin occupancy of *mrhl* RNA was investigated to understand the broader role of *mrhl* RNA in gene regulation in spermatogonial cells by the Chromatin Oligo affinity precipitation (ChOP) technique. From ChOP-sequencing experiment, around 1370 loci were found to be associated with *mrhl*. 37 genes were found to overlap between the ChOP and expression array datasets and were termed as Genes Regulated by the Physical Association of *Mrhl* (GRPAM). These are listed in **table 1.3** below. The presence of p68 is essential for *mrhl* binding to and regulating a majority of GRPAM (Akhade, V.S et al, 2014). **Figure 1.19** summarises the regulatory role of *mrhl* in mouse spermatogonial cells

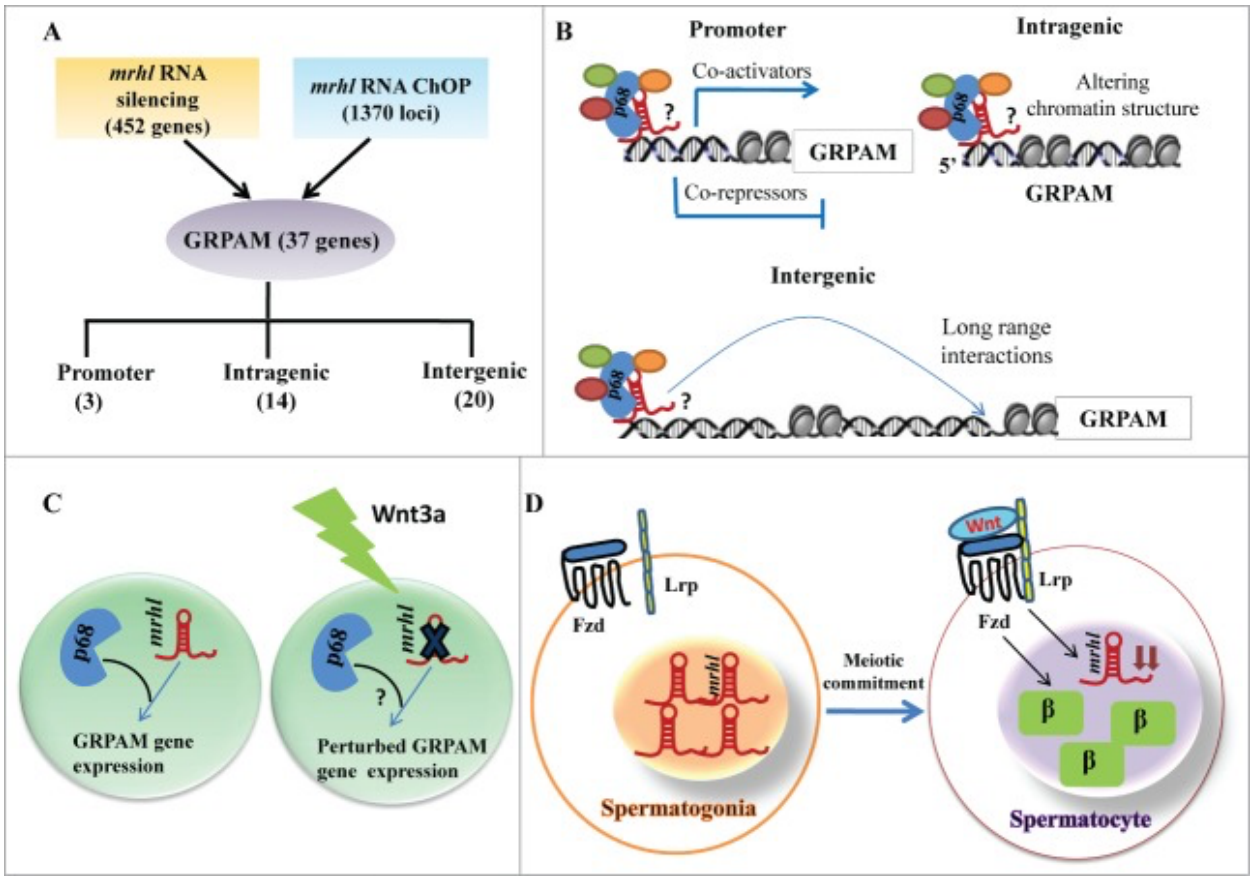


***Fig 1.19**: Chromatin occupancy of mrhl RNA and regulation of gene expression in spermatogenesis. (**A**) Genes common to microarray data and mrhl RNA ChOP - 37 GRPAM genes and their classification based on location of ChOP sequence reads with number of genes in each category mentioned in brackets (B) Hypothetical mechanisms of gene regulation by mrhl RNA-p68 complex at GRPAM loci. Regulation of promoter class of GRPAM can be through recruitment of co-activators or co-repressors while mrhl binding at intragenic and intergenic sites can regulate GRPAM through alteration of chromatin structure and long range interactions respectively. (**C**) Downregulation of mrhl RNA and perturbation of GRPAM gene expression in Gc1-Spg spermatogonial cell line upon Wnt3a treatment. (**D**) Inverse correlation between Wnt activated state and mrhl RNA expression levels between spermatogonia (higher expression of mrhl RNA) and differentiated spermatocytes (Wnt activation and decreased levels of mrhl RNA). β: β catenin, Fzd: Frizzled, Lrp: Low-density lipoprotein receptor-related protein. Figure adapted with permission from Akhade, V.S et al, 2014*

Gene	*Mrhl* ChOP site relative to gene	Regulatory effect of *mrhl* silencing
Lrba	Promoter	Downregulated
H28	Promoter	Upregulated
Sox8	**Promoter**	**Upregulated**
Odz4	Genic	Downregulated
Lamb3	Genic	Downregulated
Nrxn1	Genic	Downregulated

Gene	*Mrhl* ChOP site relative to gene	Regulatory effect of *mrhl* silencing
Kcnq5	Genic	Downregulated
Rab40b	Genic	Downregulated
Ssx2ip	Genic	Downregulated
Myo18b	Genic	Downregulated
Il1rapl1	Genic	Downregulated
Spag16	Genic	Upregulated
Gabrg2	Genic	Upregulated
Palm	Genic	Upregulated
Ksr1	Genic	Upregulated
Bach2	Genic	Upregulated
Stox2	Genic	Upregulated
Tsc22d1	51.5 kb upstream	Downregulated
Spam1	27 kb upstream	Downregulated
Ostm1	22.2 kb upstream	Downregulated
Mageb16	122.5 kb upstream	Downregulated
Grik2	440 kb upstream	Downregualted
Kcnh7	276 kb upstream	Downregualted
Adamts20	696 kb downstream	Downregulated
Sla2	0.54 kb downstream	Downregulated
Thbs4	25.5 kb downstream	Downregulated
Ppargc1a	183.8 kb downstream	Downregulated
Serpinb8	0.7 Mb upstream	Downregulated
Mrpl32	75.75 kb downstream	Downregulated
Prickle1	63.4 kb upstream	Upregulated
Znrf3	132.3 kb upstream	Upregulated
Mael	33 kb downstream	Upregulated
Rarg	11.1 kb upstream	Upregulated
Npepps	14.7 kb downstream	Upregulated
Cdh9	180.2 kb upstream	Upregulated
Zfp455	24.7 kb downstream	Upregulated

Gene	*Mrhl* ChOP site relative to gene	Regulatory effect of *mrhl* silencing
Hhipl2	581 kb downstream	Upregulated

Table 1.3: List of GRPAM, the location of mrhl ChOP site with respect to the genes and the regulatory effect of mrhl silencing on the genes.

Mrhl lncRNA in mESC, too, is chromatin associated. It binds to more genomic loci in mESC when compared to Gc1-spg cells but this interaction is independent of p68. The silencing of *mrhl* perturbs the expression of a large number of genes, especially those involved in the regulation of lineage commitment and differentiation (Pal, D et al, 2021). A similar mechanism of action is observed for *hmrhl* in the CML cell line K562. While a number of cancer related genes, transcription factors and phenotypes including invasion and cell migrations properties are regulated by *hmrhl*, a very interesting CML specific function of the *hmrhl* locus is the acquired enhancer property. It is likely that *hmrhl* behaves as an enhancer lncRNA to drive the expression of the host gene phkb in CML cells (Choudhury, S.R et al, 2020, Fatima, R. et al 2019).

Of the 37 GRPAM in mouse spermatogonial cells, *Mrhl* regulates 3 genes by binding to their promoters- H28 (interferon-induced protein 44 like), Lrba (LPS Responsive Beige-Like Anchor Protein) and the transcriptional factor Sox8 (Akhade, V.S et al, 2014).

1.14 Sox8 Transcription Factor

The Sox group of transcription factors play important roles in pre- and post-natal development in animals. They play critical roles in cell fate and differentiation decisions in many cell lineages. All Sox proteins share a highly conserved high-mobility-group (HMG) box domain that was originally identified in the sex determining gene SRY located in the Y chromosome. The canonical high-mobility-group domain is characteristic of chromatin-associated proteins. The high-mobility-group box domain, which is believed to have derived from the canonical high-mobilty-group domain, binds DNA in the minor groove. The affinity and specificity of its DNA binding is increased by interacting with other transcription factors. HMG domain proteins are unique in their ability to alter the conformation of DNA by bending it and to increase its protein accessibility and plasticity. They thereby facilitate the formation of functionally active complexes of transcription factors on gene enhancer sequences or enhanceosomes. A few Sox proteins such as

SOX2 and SOX9 have been shown to engage at regions of condensed chromatin and are considered pioneer factors. The HMG box domain alone performs the functions of DNA binding, DNA bending and protein interactions (Lefebvre, V. et al, 2007).

There are a total of 20 Sox proteins that are divided into 9 groups (**Figure 1.20**). Proteins within the same group share a high degree of identity both within and outside the HMG box, whereas Sox proteins from different groups share partial identity in the HMG box domain and none outside this domain.

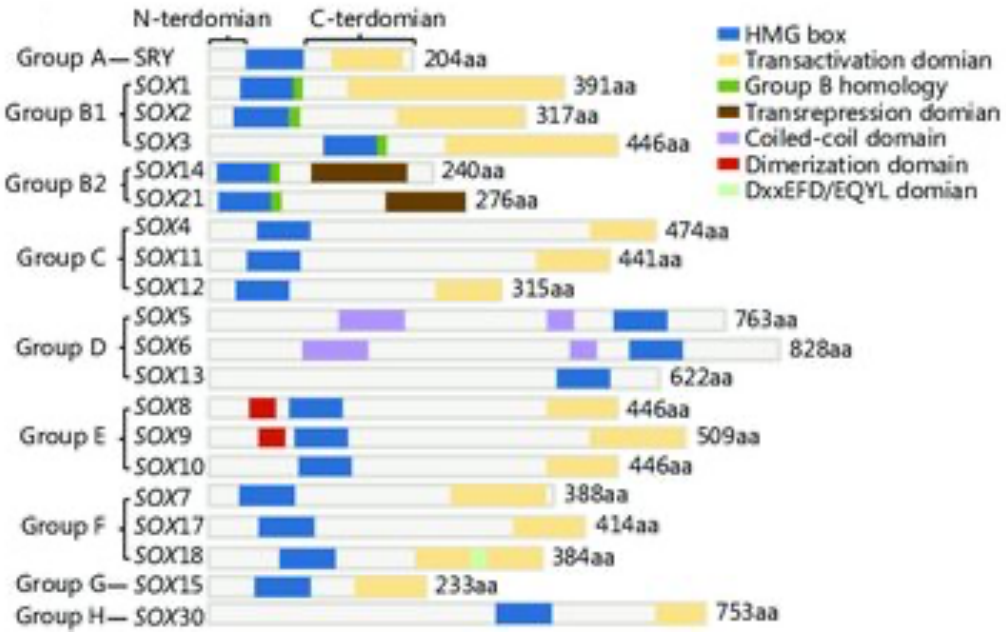


Fig 1.20: *Schematic representation of the structures of the 20 known SOX proteins and their classification into groups. The functional domains are indicated along with the length of SOX proteins. Groups and representative protein members are indicated to the left. N-terminal and C-terminal domains of SRY are depicted at the top. The sizes in amino acids (aa) of the various SOX proteins are indicated to the right of the schematic. Image has been adapted with permission from Hu, J. et al, 2019.*

The SoxE groups is made up of 3 Sox proteins- Sox8, Sox9 and Sox10. SoxE factors possess a 40 amino acid dimerisation (DIM) domain upstream of the HMG domain. Their dimerisation (homo- or hetero- with other SoxE members) relies upon the presence of a palindromic DNA binding sequence. SoxE factors additionally possess both a C-terminal transactivation domain (TAD) and a second transactivation domain in the middle of the protein (TAM, or K2 domain) as shown in **figure 1.20**. Sox transcriptional activity can depend upon whether the transcriptional partner is an activator or repressor (Schock, E.N and LaBonne, C., 2020).

Sox8 has been implicated in regulating different cell fates including oligodendrocyte specification and terminal maturation, inhibition of osteoblast differentiation and initiation of formation and maintenance of neural crest to name a few. However, its important role in testis development makes it a prime candidate to be studied among the 3 genes that *mrhl* regulates by binding at the promoter (Schock, E. N and LaBonne, 2020, Turnescu, T. et al, 2018, Lefebvre, V., 2019).

Sox8 is expressed in the developing testis, with expression starting at E12.5, shortly after Sox9 expression is upregulated (Schepers, G., et al, 2003). Studies on Sox8 null mice indicates that Sox8 is essential for the maintenance of male fertility beyond the first wave of spermatogenesis. It was found that ablation of SOX8 resulted in progressive degeneration of the seminiferous epithelium through perturbed physical interactions between Sertoli cells and the developing germ cells (**figure 1.21**) (O'Bryan, M. K., et al, 2008).

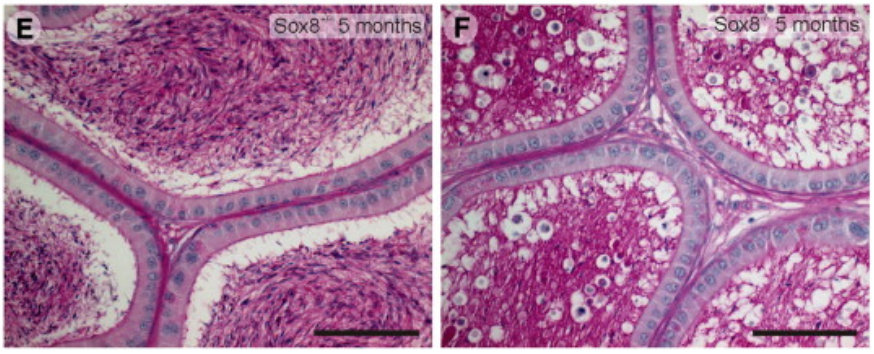


Fig 1.21: Effect of Sox8 knockout on spermatogenesis in mice- *Epidydimal section of Sox8+/+ and Sox8-/- mice showing relative absence of sperm cells in knockout animals. Adapted with permission from O' Bryan MK, Koopman, P. 2008*

While Sox9 plays a non-redundant role in sex determination, partial knockout of Sox9 before E11.5 showed a strong XY sex reversal phenotype only in Sox8 null background mice suggesting that Sox8 reinforces the function of Sox9 in testis formation (Chaboissier, M. C, et al, 2004). Additionally, sertoli cell-specific conditional inactivation of Sox9 in Sox8 null background mice two days after the sex determination stage leads to downregulation of the testis-promoting Dmrt1 gene with upregulation of the ovarian-specific genes Wnt4, Rspo1 and Foxl2. Testes with these Sertoli cells undergo testis-to-ovary genetic reprogramming and Sertoli-to-granulosa cell transdifferentiation. Expression of both these SoxE proteins protects the adult mouse testis from complete degeneration (Barrionuevo, F.

J, et al, 2016). Moreover, SOX8 alone can compensate for the loss of SOX9 for Sertoli cell differentiation during female-to-male sex reversal (Richardson, N. et al, 2020).

In the Gc1-spg cells, the *mrhl* downregulation induced meiotic commitment is dependent on Sox8. The Wnt induced upregulation of the meiotic and pre-meiotic markers Lhx8, Stra8, c-Kit, Mtl5 and Hspa2 is abrogated upon Sox8 knockdown upon Wnt induction (**Figure 1.22A**). Moreover, binding sites for Sox8 are present in the promoters of these genes (**Figure 1.22B**). Stra8 is the master regulator of a transcription program leading to meiotic commitment and progression (Feng, C. W, et al, 2014) and it is likely that Sox8 regulates the expression of Stra8 and is thereby involved in regulating meiotic commitment of the cells (**Figure 1.22C**) (Kataruka, S. et al, 2017).

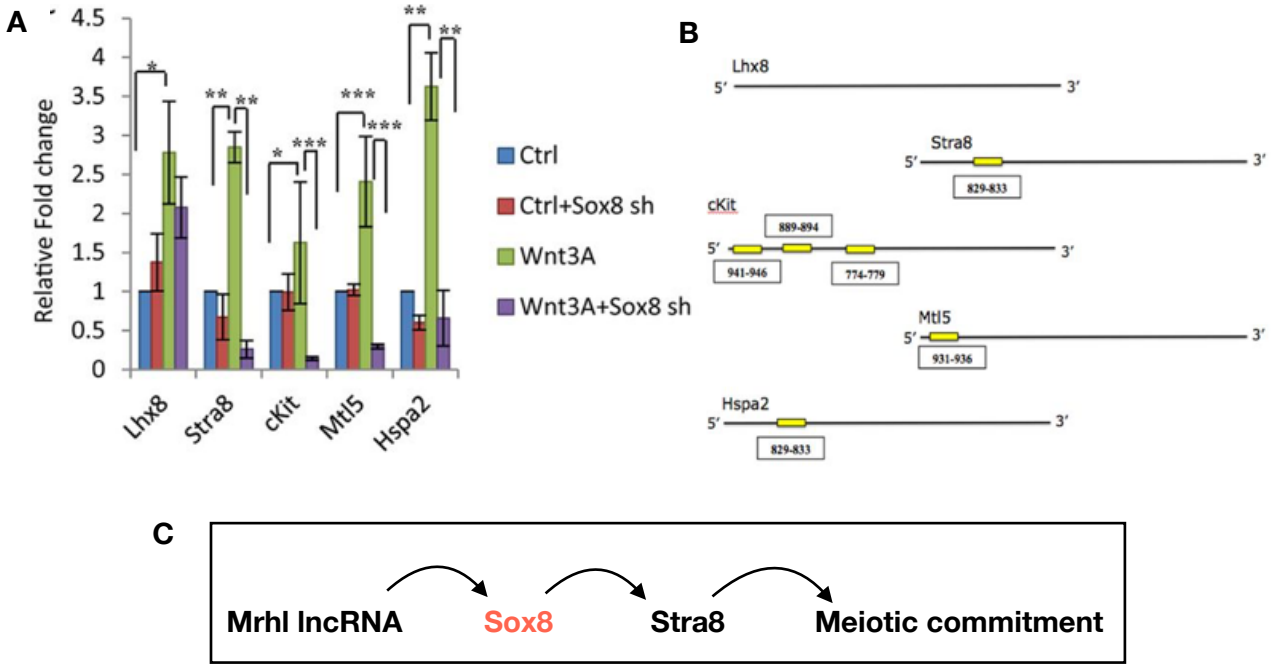


Fig.1.22 - Meiotic commitment of Gc1-spg cells in response to mrhl downregulation is Sox8 dependent - *(A) Upregulation of meiotic and pre-meiotic markers is dependent on Sox8 (B) Sox8 binding site is present within the promoter for these markers. Figures A and B have been adapted from Kataruka, S. et al, 2017 (C) The probable regulatory cascade in Gc1-spg cells*

1.15 Chromatin Dynamics at the Sox8 promoter

At the bidirectional promoter of Sox8, *mrhl* binds around 140bp upstream of the Transcriptional Start Site (TSS) in a p68-dependent manner to maintain Sox8 gene in the transcriptionally repressed state. In the presence of *mrhl*, the Mad-max repressive complex along with the co-repressors Sin3a and HDAC are also bound very close to the *mrhl* binding site. The repressive histone modification H3K27me3 is present at the Sox8

locus and TCF4 is bound at the WRE present in the promoter around 800bp upstream of the TSS. Substantial changes in chromatin dynamics at the Sox8 promoter are observed upon activation of the Wnt signalling pathway. As *mrhl* levels reduce, the lncRNA is evicted from the promoter and the Mad-Max repressive complex is replaced by the Myc-Max activating complex. The histone acetyltransferase Pcaf is recruited to the same binding site. The Wnt effector molecule Beta-catenin binds to the WRE with TCF4. Along with a significant decrease in H3K27me3 histone modification, a concomitant increase in the H3K4me3 and H3K9ac histone modifications associated with transcriptional activity is observed at the Sox8 promoter and Sox8 is actively transcribed. This mechanism is illustrated in **Figure 1.23**. The spermatogonial cells commit to meiosis upon receiving the Wnt cue as evidenced by the increase in the expression levels of pre-meiotic and meiotic marker genes such as Stra8, Scp3, Dmc1, and Spo11 and this commitment is Sox8 dependent (Kataruka, S. et al, 2017).

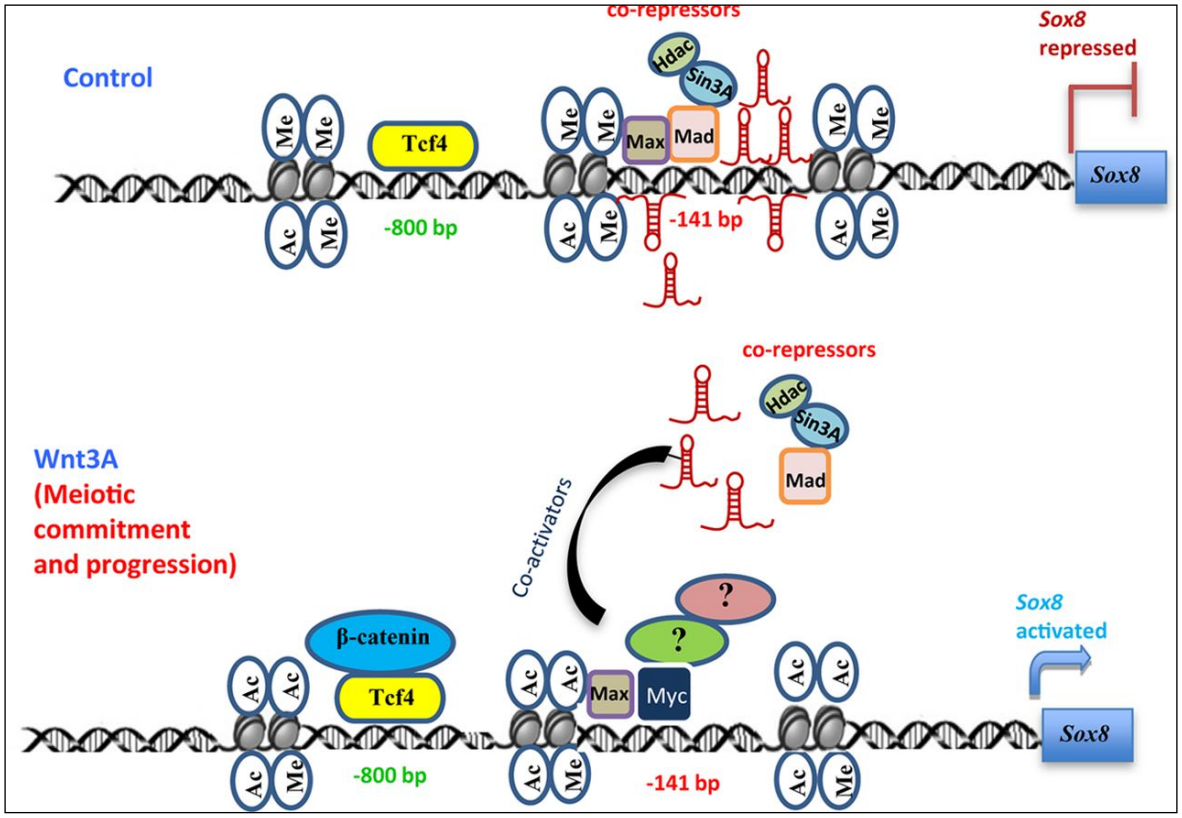


Fig 1.23: Chromatin dynamics at the Sox8 promoter *- Mrhl binds to the promoter of Sox8 to maintain it in the transcriptionally repressed state by recruiting the Mad-Max complex along with the co-repressors Hdac and Sin3a to the promoter. H3K27me3 levels are high and H3K9ac levels are low in this state. Upon activation of Wnt signalling pathway, mrhl gets downregulated and is no longer bound at the Sox8 promoter. The Mad-max complex is replaced by the Myc-Max activating complex. Beta-catenin binds at the WRE present in the distal promoter. H3K27me3 repressive histone mar levels reduce significantly and is accompanied by a concomitant increase in H3K4me3 and H3K9ac levels which are associated with active transcription. Adapted from Kataruka, S. et al, 2017.*

1.16 AIMS AND SCOPE OF CURRENT STUDY

As discussed previously, characterisation of the functional role of *mrhl* in the spermatogonial cells revealed that multiple genes are disrupted upon *mrhl* silencing and while a subset of them are upregulated, the rest are downregulated. It is likely that different regulatory protein complexes are recruited by *mrhl* to the target sites to impart these varied regulatory effects. Additionally, *mrhl* binding sites identified from ChOP-sequencing appear to be distributed across the promoter, gene body or intergenic sites.

Mrhl lncRNA is also expressed in the mouse embryonic stem cells and has wide-spread chromatin occupancy in mouse embryonic stem cells with *mrhl* binding at many more chromatin loci in this system. Upon silencing of *mrhl* in mESC, the perturbed transcriptome is vastly different from those in spermatogonial cells with only 25 genes being common to both systems. Genes involved in the regulation of various cellular processes are differentially regulated mESC. Specifically, a number of genes involved in cell adhesion/ receptor activity and lineage-specific TFs are misregulated upon *mrhl* silencing in the ES cells (Pal, D et al, 2021). This can be attributed to different regions of the chromatin being accessible at two different developmental stages to facilitate cell-type specific transcriptional programs.

Also as discussed previously, the human homolog *hmrhl* is highly upregulated in leukaemia, specifically in CML. *Hmrhl* does not interact with p68 in CML cell lines and it has no relationship with the Wnt signalling pathway. However, *hmrhl* is chromatin bound and binds to multiple loci in CML cell line K562. Knockdown of *hmrhl* in CML lines perturbs the expression of important transcription factors such as KLF12,KLF2, MAFA, STAT4, ASCL2, KLF4, BATF and TP63, which have all been linked to leukemia and lymphomas in various studies, and other genes such as PDGRFβ, PRDM16, PTPRK and ZIC1, which are all involved in signalling pathways with fundamental role in development, cell growth, proliferation and migration (Choudhury, S.R et al, 2020). The syntenic conservation of lncRNA *mrhl* (Fatima, R. et al 2019) suggests a functional association that has been maintained across evolution. While the lncRNA regulates cell type-specific genes and relevant transcriptional programs, it is likely that there is a common mechanism by which the genes are regulated. Since the commonality does not lie in the signalling pathway or the protein interacting partner the mechanism is elusive.

In silico analysis has suggested a good probability for the formation of DNA:DNA:RNA triplex in tethering the lncRNA to their target sites in both mESC and in CML cells (Pal, D et al, 2021, Choudhury, S.R et al, 2020). It has been suggested that the specificity of lncRNA binding is conferred by the triplex forming site within the lncRNA. We wanted to investigate this probability in the spermatogonial cells at the Sox8 locus as a part of our efforts to find a common mechanism of regulation. Locus specific studies not only enable to investigate the possibility of the nucleotides being involved in the triplex formation but also the functional relevance of such binding.

While the chromatin dynamics at the Sox8 promoter has been well characterised in the conditions corresponding to transcriptional repression (in the presence of *mrhl*) and active transcription (Wnt activated, *mrhl* downregulated conditions), it is likely that additional regulatory events are at play in the locus at regions other than the promoter. After all, the process of transcription involves intricately regulated cohort of events and multiple regulatory elements in addition to the promoter are involved. Additionally, while 3 out of the 37 GRPAM loci at regulated by *mrhl* binding at the gene promoters, the binding of *mrhl* at regions other than the promoter contributes to transcriptional regulation at the remaining 34 GRPAM which indicates that events at these non-promoter genomic elements contribute equally to the process of transcription. Intergenic binding of *mrhl* to regulate GRPAM is also indicative of a probable regulation through the mediation of long-range interactions. The essentiality of *mrhl* and p68 in the transcriptional repression of Sox8 gene is reminiscent of the regulation of the H19/Igf2 imprinted locus involving the lncRNA Receptor RNA Activator (SRA) and p68 as essential molecular players. CTCF binds to the Imprinting Control Region present between the shared enhancer and the *Igf2* gene in a methylation sensitive manner on the maternal allele. When bound at the ICR, CTCF silences the *Igf2* gene. The enhancer contacts the promoter of the *H19* gene and results in *H19* expression. DNA methylation of the ICR on the paternal allele prevents CTCF binding at the region and thereby leads to the expression of the *H19* gene. The insulating function of CTCF is dependent on protein cofactors such as the cohesin complex, the DEAD-box helicase DDX5/p68. CTCF and cohesin binding is dependent on the presence of lncRNA SRA. At the ICR on the maternal allele, CTCF, cohesin and p68 bind as a complex. Within this complex, p68 interacts with both CTCF and cohesin. The interaction with CTCF and the ability of the complex to maintain insulation is dependent on the lncRNA SRA.

p68/SRA binding is also important for the recruitment of the cohesin complex to this locus. Depletion of SRA reduces CTCF-mediated insulator activity at the *IGF2/H19* imprinting control region and results in decreased H19 expression and increased IGF2 expression and this mechanism is illustrated in **Figure 1.24**. Moreover, p68 binding and CTCF/cohesin binding coincide at multiple loci in human cells (Yao,H *et al*, 2010).

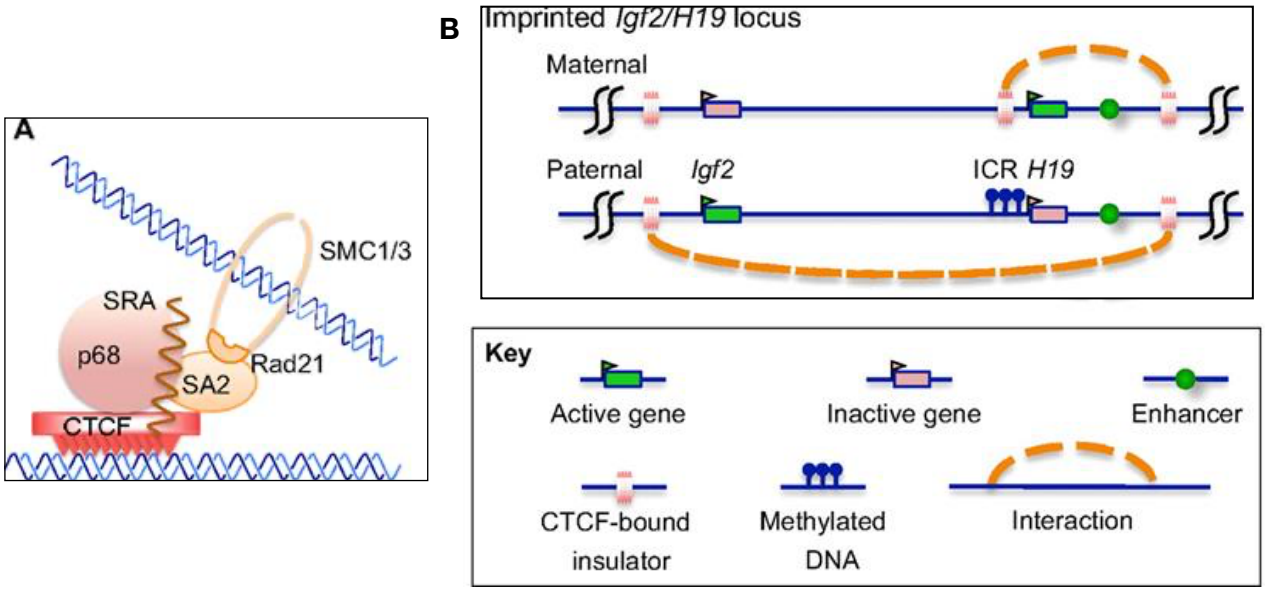


Fig. 1.24 Imprinting at the H19/Igf2 locus by CTCF, cohesin, p68 and SRA lncRNA *(**A**) The CTCF-cohesin complex is stabilised by the RNA helicase p68 and the RNA molecule SRA (brown wavy line), which are required for insulation **(B)** In mammals, only the maternally derived allele shows H19 gene activity induced by the downstream enhancer. On the paternally derived allele, the DNA of the imprinted control region (ICR) is methylated and prevents CTCF binding such that the insulator function at this site is abrogated. In this situation, the downstream insulator contacts the insulator upstream of the Igf2 gene and positions the enhancer next to the Igf2 promoter, which activates the Igf2 gene. Only some of the long-range interactions are shown in this simplified depiction of the locus. The image has been adapted with permission from Herold, M et al, 2012.*

It is likely that the regulation of Sox8 by mrhl lncRNA too could be along the lines of the mechanism seen at the H19/Igf2 imprinted locus. Keeping this in mind, the aim of this study is as follows:

Investigating the role of *mrhl* lncRNA in chromatin organisation and orchestrating chromatin dynamics, specifically at the mouse Sox8 genomic locus.

Chapter 2

Materials and Methods

2.1 Chemicals and Reagents

All chemical reagents and organic solvents were of AR grade and were purchased from Sigma-Aldrich.

2.2 Bacterial Culture

E.coli strains XL1-Blue or DH5α have been used for plasmid amplification except for BAC plasmids. Bacterial cultures were grown by inoculating a single colony in Luria Bertani (LB) Broth (10g bactotryptone, 5g yeast extract and 10g NaCl in 1L of water) for 12-16 hours at 37°C with shaking at 180rpm. For selection where appropriate, ampicillin or chloramphenicol was added to the culturing LB media at 100 µg/ml final concentration or 25 µg/ml final concentration respectively.

Bacterial colonies were grown on LB agar plates by streaking bacteria from Glycerol stocks or transformation mix on LB agar plates with appropriate antibiotic selection and incubating for upto 16 hours at 37°C in a static incubator.

For long term storage, glycerol stocks of bacteria were prepared by mixing overnight grown bacterial culture with 50% v/v glycerol. Stocks were stored at -80°.

2.3 Bacterial Transformation

Competent cells were prepared by modifying the Transformation and Storage Solution (TSS) method as stated by Chung *et a*l (Chung, C. T *et al*, 1989). The bacterial strain was inoculated in LB media and grown overnight at 37°C. Using 1% of the overnight primary culture as inoculum, a secondary culture was grown at 37°C until it reached an optical density of 0.3 to 0.4. The culture was chilled on ice and a bacterial cell pellet was obtained by centrifugation at 1300xg for 10 minutes at 4ºC. The cell pellet was resuspended in 1/10th of the culture volume of TSS (Sucrose – 8gm**,** 0.5M EDTA – 10ml, 1M Tris-HCl (pH 8.0) – 1ml, Triton X 100 – 0.5ml in total of 100ml volume). Aliquots of 150µl of competent cells were stored at -80ºC.

For transformation, competent cells were mixed with the desired DNA (approximately 50ng of DNA) and incubated on ice for 30 minutes. The cells were subjected to heat shock at 42ºC for 90 seconds and immediately chilled on ice for 5 minutes. 800µl of LB medium was added for recovery of the cells and incubated at 37°C for 1 hour. The cells were then pelleted and resuspended in 100µl of LB and plated on LB agar containing appropriate antibiotic.

2.4 BAC plasmid preparation by Alkaline Lysis

Colonies were inoculated in 10ml LB broth containing chloramphenicol to a final concentration of 25 µg/ml and grown overnight at 37 ºC at 180rpm. The cells were pelleted at 3000Xg for 10 minutes and media was aspirated to leave the pellet dry. Pellet was resuspended in 400 µl of ice cold buffer P1 (50 mM Tris-HCl, pH 8.0, 10 mMEDTA, 100µg/mL of RNase A). To this, 400 µL of buffer P2 (0.2M NaOH, 1% SDS) was added and tubes were inverted to mix. 400 µL of buffer P3 (3.0 M potassium acetate, pH 5.5) was added and inverted to mix. Tubes were incubated in ice for 10 minutes. Lysate was centrifuged at maximum speed for 5 minutes at 4ºC. Plasmid was recovered by isopropanol precipitation after extraction with phenol:chloroform. The air-dried pellet was dissolved in nuclease-free water. BAC plasmids were procured from BACPAC resources, California

2.5 SDS-PAGE

Proteins were resolved on 8%, 10% or 12% polyacrylamide gels containing sodium dodecyl sulphate depending on the molecular size of the proteins being investigated (Sambrook and Russell, 2006).

The composition of the electrophoresis buffer was 25mM Tris, 250mM glycine pH 8.3, 0.1% SDS. The composition of the sample buffer (5X) was 250mM Tris-Cl pH 6.8, 5% beta mercaptoethanol, 10% SDS, 0.02% bromophenol blue, 30% glycerol. 100V was used for electrophoresis for 3 hours.

2.6 Western Blotting

The proteins resolved by SDS-PAGE were transferred on to nitrocellulose membrane using Amersham Biosciences' semidry transfer unit with the transfer buffer (7.5g of glycine, 1.65g of Tris base, 1ml of 10% SDS, 100ml of methanol made upto 500ml with water). The transfer was set up at 2.5mA/ cm^2 of blot area. After transfer, the membrane was stained with 0.1% Ponceau stain (0.1g of Ponceau, 5ml of acetic acid made upto 100ml using

water) to check the efficiency of transfer of proteins. The membrane was blocked using 5% skimmed milk in 1x PBS (137 mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4 ,1.8 mM KH_2PO_4) or 3% BSA in 1X TBS (20mM Tris base, 150mM NaCl) for 1 hour at room temperature. The membrane was then incubated with the primary antibody of appropriate titre in 1% skimmed milk in 0.05% PBST or 1% BSA in 0.3% TBST overnight at 4ºC as recommended by the antibody manufacturer. The membrane was washed with 0.05% PBST or 0.3% TBST multiple times for 10 minutes each. Subsequently the membrane was incubated with the secondary antibody (Goat-anti rabbit HRP conjugate from GeNei Labs, Catalog number: HP03 or protein A HRP conjugate from GeNei Labs, Catalog number: HPO8) in 1% skimmed milk in 0.05%PBST or 1% BSA in 1XTBS containing 0.3% Tween-20 at a dilution of 1:5000 for 1 hour at room temperature. This was followed by multiple washes of 10 minutes each using 0.05% PBST or 0.3% TBST. Detection was carried out using the Biorad ECL-analysis system.

2.7 Real-Time PCR

50ng -100ng of DNA/ cDNA template was used per reaction. SYBR green from TaKaRa was used for the reactions. 20µl reactions were set up using the required primers in a Biorad CFX-96 machine.

2.8 Culturing of mammalian cells

Mammalian cell-line Gc1-spg (derived from B-type mouse spermatogonial cells) (American Type Culture Collection, cat no: CRL-2053) was cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS and 10 U/ml of penicillin and 0.1mg/ml of streptomycin (Complete media). The cells were grown at 37^0C at 5% CO_2 until 95% confluence was reached. For subculturing, cells were detached from culture dishes by trypsinisation (0.25% trypsin and 0.02% EDTA) at 37^0C for 3 minutes. Trypsin was neutralised with complete medium and cells were pelleted at 900Xg for 4 minutes. Cell pellet was resuspended in complete medium and seeded in required number of culture dishes at a splitting ratio of 1:10.

Control L-cells (American Type Culture Collection, Catalog number:CRL-2648 or L-Wnt3a (American Type Culture Collection, Catalog number: CRL- 2647) cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS and 10 U/ml of penicillin and 0.1mg/ml of streptomycin (Complete media) supplemented with G418 at a final concentration of 40 µg/ml. Incubation conditions were similar to that used for Gc1-spg cells. For subculturing, cells were detached from culture dishes by trypsinisation (0.25%

trypsin and 0.02% EDTA) at 37^{0}C for 1 minute. Trypsin was neutralised in the presence of complete medium and cells were pelleted at 900Xg for 4 minutes. Cell pellet was resuspended in complete medium and seeded in required number of culture dishes at a splitting ratio of 1:10.

2.9 Preparation of Conditioned medium

Control L-cells or L-Wnt3a cells were seeded at 60% confluence in Complete media supplemented with G418 (Gibco, Catalog number: 10131035) as per manufacturer's instructions. Cells were incubated for 4 days at 37^{0}C at 5% CO_2. The supernatant media was collected and replaced with fresh media and incubated for 3 more days. The supernatant media was collected and the two batches of media were mixed. This conditioned media was centrifuged at 3000Xg for 5 minutes to remove cell debris. Then the media was filter sterilised and stored at 4^{0}C till use.

Gc1-spg cells were seeded in a culture dish in complete media and allowed to grow for 24 hours. Then, the media was replaced with conditioned media and grown for 72 hours with conditioned media replacement every 24 hours. The cells were harvested at the end of 72 hours and used for downstream experiments.

2.10 Freezing of mammalian cells

Cells grown to confluence were washed with 1X PBS and trypsinised as described previously. The cells were pelleted by centrifugation at 900Xg for 4 minutes. The pellet was resuspended in freezing medium (95% complete medium +5% DMSO) and immediately transferred to cryovials. The vials were stored at -80^{0}C in Cell Freeing containers designed to achieve a rate of cooling very close to -1°C/minute for two days before being transferred to the vapour phase of liquid nitrogen.

2.11 Thawing of mammalian cells

Frozen stocks were thawed in a water bath at 37^{0}C for 2-5 minutes. The thawed stocks were transferred immediately into 5ml of complete medium to dilute the DMSO. Cells were pelleted by centrifugation at 900Xg for 4 minutes. The pellet was gently resuspended in complete medium and seeded in culture dishes.

2.12 Generation of inducible silencing cell line

Lentiviral shRNA plasmids targeting either *Mrhl* lncRNA (custom made) or control non-targeting plasmids (Catalog number: SHC332) (Sigma-Aldrich) cloned into the backbone

vector pLKO-puro-3XLacO by Sigma-Aldrich were prepared using the Macherey-Nagel NucleoSpin plasmid kit as per manufacturer's instructions.

Gc1-spg cells were seeded at 50% confluence. 24 hours later, cells were transfected with either *mrhl*-targeting or non-target plasmids at a concentration of 1.5µg of plasmid/ ml of culture media using Lipofectamine 2000 as per manufacturer's instructions. 48 hours post transfection, cells were subjected to antibiotic selection using puromycin at a concentration of 3µg/ml for 72 hours with a media change every 24 hours. At the end of 72hours, antibiotic was withdrawn and the cells were allowed to grow confluent in complete media. Stocks were made of these shRNA integrated cell lines.

2.13 Induction of silencing in cell line

Cells were seeded for induction at 20% confluence. 24 hours later, *mrhl* or non-target silencing were induced by adding 0.5mM IPTG and 2.5µg/ ml of puromycin in complete media. Silencing was induced for a total of 4 days or 5 days with media change every 24 hours. Cells at the end of silencing induction were used for downstream application.

2.14 Mammalian genomic DNA isolation

Cell Pellets were washed twice with 1X PBS and resuspended in 5 volumes of digestion buffer (50mM Tris-Cl pH8.0, 100mM EDTA pH 8.0, 100mM NaCl, 1%SDS) along with Proteinase K to a final concentration of 0.5mg/ml. The reaction was incubated overnight at 55°C with gentle shaking. RNase A (200µg) was added to the reaction and the reaction was incubated at 37°C for 1 hour. Phenol:chlorofrom:isoamylalcohol extraction was performed to eliminate the protein and DNA was ethanol precipitated. After washing the pellet with 70% ethanol, the pellet was air-dried and resuspended in nuclease-free water.

2.15 Preparation of P7 and P21 testicular cells for Chromatin immunoprecipitation

The testicular samples prepared were harvested from BALB/C mice in the ages groups of 7dpp and 21dpp. The animals were sacrificed and the testes were decapsulated in PBS (pH 7.4) on ice by carefully removing the tunica albuginea. The seminiferous tubules were subsequently released into PBS (pH 7.4) and were chopped into small pieces. This was then subjected to mechanical homogenisation on ice to procure a homogenised single-cell suspension of the sample which was then processed for various assays.

2.16 Chromatin Immunoprecipitation

Around 2-3X10^6 cells were used per ChIP. To crosslink, cells harvested by trypsinisation were resuspended in 5ml of 1X PBS and Formaldehyde was added to a final concentration of 1%. The cells were incubated for 10 minutes with shaking at room temperature. The formaldehyde was quenched by adding glycine to a final concentration of 0.25M and incubating with shaking at room temperature for 5 minutes. Cells were pelted by centrifugation for 4 minutes at 900Xg at 4°C. Cell pellet was resuspended in 200 µL of SDS Lysis Buffer (1% SDS, 10 mM EDTA, 50 mM Tris, pH 8.1, 1X mPIC (mammalian protease inhibitor cocktail from Sigma-Aldrich, cat. no: 04693132001), 0.2µM PMSF (phenylmethylsulfonylfluoride) and incubated for 10 minutes on ice. The lysate was sonicated using the Diagenode Bioruptor for 35 cycles on High, 30 seconds On/ 30 seconds Off cycles to enrich chromatin in the size range of 200-500 basepairs. The lysate was centrifuged at 14,000Xg to sediment out the debris. The supernatant was diluted 10 times in ChIP 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,1X mPIC, 0.2µM PMSF) and split into two aliquots, one each for antibody binding and isotype control binding. 5%-10% of the lysate was kept aside as input control for the experiment. To each of the tubes, around 5-10µg of specific antibody or equivalent amount of isotype control antibody was added and incubated overnight at 4°C with rotation. 35µl of equilibrated Protein A Dynabeads (Thermo Fisher Scientific, Catalog number: 10001) were added to the tubes to capture the antibody and incubated at 4°C for 3 hours with rotation. The beads were captured using a magnetic rack (Thermo Fisher Scientific) and washed two times with low-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), once with high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), once with Lithium Chloride wash buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholic acid (sodium salt), 1 mM EDTA, 10 mM Tris, pH 8.1) and twice with T.E buffer for 3 minutes each with rotation at 4°C. For western blotting, the beads were boiled with Laemmli buffer and loaded into SDS-Polyacrylamide gel.

For Real-Time (RT) PCR, DNA was eluted from the beads using freshly prepared elution buffer (1%SDS, 0.1M $NaHCO_3$) for 30 minutes with rotation at room temperature. DNA was subjected to proteinase K treatment and reverse-crosslinking was done overnight at 65°C in the presence of NaCl to a final concentration to 0.2M. ChIP DNA was precipitated by ethanol precipitation after phenol:chloroform purification and resuspended in nuclease

free water. This DNA was used as template for RT-PCR. ChIP experiments were carried out in biological triplicates and RT-PCR was set up in duplicates

The data from ChIP was plotted as % Input and was calculated as follows:

Relative occupancy = % of input X 2^ [Ct(Input) - Ct (ChIP)]

The antibodies used in this study are listed in **table 2.1** below.

Antibodies	Host	Manufacturer	Catalog number
CTCF	Rabbit	AbCam	ab188408
Rad21	Rabbit	AbCam	ab992
p68	Rabbit	Cusabio	CSB-PA003685
YY1	Rabbit	Diagenode	C15410345
H3K4me1	Rabbit	Diagenode	C15410037
H3K27ac	Rabbit	Diagenode	C15410174
Ezh2	Rabbit	Diagenode	C15410039
Isotype control	Rabbit	GeNei	1620480101730

Table 2.1 - List of antibodies used in the study

2.17 Electrophoretic Mobility Shift Assay (EMSA)

2.17.1 Probe labeling

Single stranded DNA oligonucleotides (Sigma-Aldrich) was radio labelled by assembling the following reaction.

10X PNK buffer - To final of 1X concentration
DNA oligo - 40pmoles
T4 PNK - 20 Units
ATP [γ-32P] - 3µCi

The components were thoroughly mixed and incubated at 37°C for 30 minutes and then heat inactivated at 65°C for 20 minutes. DNA was ethanol precipitated after a phenol:chloroform:isoamlyalcohol clean-up. The oligo was mixed with equimolar concentration of its complementary unlabelled oligo and annealed by heating to 95°C for 10 minutes and slowly cooling to room temperature in annealing buffer (10mM Tris-Cl pH 8.0, 20mM NaCl). DNA:RNA hybrid duplex was also prepared similarly by heating

complementary oligonucleotides to 95°C for 10 minutes and slowly cooling to room temperature in annealing buffer.

2.17.2 Probe purification

Sephadex G-50 slurry was prepared by dissolving 1g of powder in 16ml of nuclease-free water and allowing it to swell overnight at 4°C. G-50 column was prepared by packing glass wool at the tip of a 1ml syringe and then packing the rest of it with the swollen beads. Water was removed from the column by centrifugation and then the annealed probes were loaded into the column. Probes were collected by centrifugation and radiocount of the probe was measured using scintillation counter.

2.18 Triplex EMSA

The protocol from Mondal, T. et al (Mondal, T. *et al*, 2015) was followed. DNA probes equivalent to 10,000 cpm were used per reaction. RNA oligonucelotides (Sigma-Aldrich) were heated to 70°C for 10 minutes and immediately chilled on ice to remove any secondary structure. Triplex binding reactions were assembled in 10µl volume with DNA probes, RNA oligonucleotides in molar ratios varying from 0.5 to 50 times of DNA probes, 10% glycerol and binding buffer (10mM Tris-Cl pH 7.5, 25mM NaCl, 10mM $MgCl_2$) and incubated for 6 hours at room temperature. RNase H and RNase A were added to the control tubes after incubation and incubated at 37°C for 30 minutes. Binding reactions were mixed with EMSA loading dye and loaded into a pre-chilled 20% Native Polyacrylamide gel for electrophoresis. Electrophoresis was carried out overnight at 4°C using pre-chilled running buffer (1X TBE, 8mM $MgCl_2$) at 50V and then for 24 hours at 100V. The gel was dried at 80°C using a vacuum drier and exposed to X-ray film for 48 hours before developing the film.

2.19 Circular Dichroism Spectroscopy

The protocol from Mondal, T. et al (Mondal, T. et al, 2015) was followed. 2.5µM each of complementary DNA and RNA (1:1:1 ratio) oligonucleotides were heated to 80°C for 5 minutes in 200µl of 1X triplex forming buffer (10mM Tris-Cl pH 7.5, 25mM NaCl, 10mM $MgCl_2$) and slowly cooled to room temperature. Following this, the reaction was incubated for 8 hours at room temperature. Control reaction of dsDNA only and RNA only were treated similarly. The reactions were stored overnight at 4 degrees. The spectra were recorded on Jasco 500A machine set to 5°C with settings from 200nm to 320nm at a scanning speed of 10nm/minute. The recorded spectra were baseline corrected with

spectrum of buffer alone. Spectra were plotted as average of 4 trials as molar ellipticity. Control artificial spectra were generated by summation of the individual spectra recorded for the dsDNA only and RNA only control oligonucleotides.

2.20 In nucleus triplex pulldown assay

The protocol from Mondal, T. et al (Mondal, T. et al, 2015) was followed. Nuclei were prepared by resuspending cells in 5 times cell pellet volumes of 1X nuclei isolation buffer (40mM Tris-Cl pH 7.5, 20mM $MgCl_2$, 4%Triton X-100, 1.28M sucrose) and incubated for 20 minutes on ice without pipetting and harvested by centrifugation at 2,500Xg for 5 minutes at 4°C. The nuclei were washed twice with ice cold 1X PBS and resuspended in 100µl of 10µM RNA oligo (either control or test) containing 1X forming triplex buffer pipetting. RNA oligonucleotides were synthesised (Sigma-Aldrich) with biotin modification on the 3' end and psoralen modification on the 5' end. The reaction was incubated to allow for oligo binding at 30°C at 650rpm for 2.5 hours. UV cross-linking was carried out at 365nm for 10 minutes at room temperature to facilitate the covalent crosslinks of psoralen to the DNA. Nuclei were lysed and chromatin was sheared using the sonicator for 10 minutes on high setting with 30 sec on/ 30sec off cycle. Debris was removed by centrifugation at 17,000Xg for 10 minutes at 4°C. 50µl of equilibrated streptavidin-magnabeads (Thermo Fisher Scientific, Catalog number: 65001) were added to the lysate and the reaction was incubated at 30°C for 1 hour with rotation at 650 rpm. After 3 washes with 1X triplex forming buffer at room temperature, the beads were resuspended in 400µl of DNA elution buffer (1% SDS, 0.05M $NaHCO_3$) containing 2µl of RNase A (10mg/ml stock) and incubated at 37°C for 30 minutes. Proteinase K treatment was carried out at 65°C for 45 minutes. NaCl was added to a final concentration of 200mM. Ethanol precipitation was carried to precipitate DNA out after phenol:chloroform purification and the DNA was used to set up real-time PCR

2.21 Methylation Assay

For the assay, the EpiTect II DNA Methylation Enzyme Kit (Catalog number. 335452) and EpiTect Methyl II PCR Primer Assay for Mouse Sox8 (CpG Island 104707) (Catalog number. : EPMM104707-1A) were used as per manufacturer's instructions. In brief, genomic DNA was isolated form Control and Wnt induced Gc1-spg cells or testicular cells from the testis of 7-day old or 21-day old mice using DNeasy Blood and Tissue Kit from Qiagen (Catalog number. 69504). Genomic DNA was subjected to enzymatic digestion

using the EpiTect II DNA Methylation Enzyme Kit which contains enzymes which are methylation sensitive and methylation dependent. The digested DNA was used as template for RT-PCR using the primers supplied in EpiTect Methyl II PCR Primer Assay for mouse Sox8. Methylation percentage was calculated using the algorithm spreadsheet supplied with the kit.

2.22 Chromosome conformation capture assay

BAC plasmids containing the Sox8 locus (RP23- 70O24) and the control Ercc3 locus (RP23-148C24) (BACPAC resources, California) were purified by Alkaline Lysis method. The plasmids were mixed in equimolar ratio and 20µg of mixed plasmid was subjected to restriction digestion with the enzyme Sau3AI (NEB, Catalog number: R0169) overnight at 37°C. DNA was precipitated by ethanol precipitation method and resuspended in ligation master mix (1X NEB Ligation buffer, 0.8% Triton X-100, 0.5X BSA, 1600U of T4 DNA ligase (NEB, Catalog number: M0202). The reaction was incubated at 21 °C for 4 hours. Ligated DNA was precipitated by ethanol precipitation method and resuspended in Nuclease Free Water (NFW). The library was used as template for PCR.

1X 10^7 cells were cross linked using 1% formaldehyde at room temperature for 10 minutes. Formaldehyde was quenched using glycine to a final concentration of 0.125M at room temperature for 5 minutes. After 2 washes with 1X PBS, nuclei were isolated by resuspending cells in 10 times cell pellet volume of cell lysis buffer (10mM Tris-Cl pH8.0, 1.5mM MgCl2, 10mM KCl,0.5mM DTT, 0.05% NP-40, 1X mPIC and 0.2µM PMSF) and incubating on ice for 30 minutes. All buffers were made with RNase-free water. Nuclei were pelleted by centrifugation and washed once with cell lysis buffer. Nuclei were permeabilised using 0.7% SDS made with NFW and at incubating at 62°C for 15 minutes. Triton X-100 was added to a final concentration of 10%. 50µl of 10X DpnII buffer and 375U of DpnII (NEB, Catalog number: R0543) restriction enzyme was aded and the reaction was incubated overnight at 37°C with shaking at 900rpm. The reaction was heat inactivated at 62°C for 20 minutes. In-situ ligation was carried out by adding ligation master mix (1X NEB Ligation buffer, 0.8% Triton X-100, 0.5X BSA, 1600U of T4 DNA ligase (NEB, Catalog number: M0202) and incubating at 21 °C for 4 hours. Nuclei were pelleted by centrifugation and resuspended in 200 µL of SDS Lysis Buffer (1% SDS, 10 mM EDTA, 50 mM Tris, pH 8.1, 1X mPIC, 0.2µM PMSF) and incubated for 10 minutes on ice. The lysate and was subjected to proteinase K treatment and reverse-crosslinking overnight at 65°C in the presence of NaCl to a final concentration to 0.2M. DNA was precipitated by ethanol

precipitation method and the library was then used as template for PCR. Agarose gel images were quantified with ImageJ (Schneider, S.A et al, 2012) software and the relative frequency of interaction was calculated as described by Naumova et al (Naumova, N. et al, 2012)

2.23 HiChIP

2.23.1 HiC library preparation

The protocol by Mumbach et. al, was followed for HiChIP (Mumbach, M.R et al, 2016). 1X 10^7 cells were cross linked using 1% formaldehyde at room temperature for 10 minutes. Formaldehyde was quenched using Glycine to a final concentration of 0.125M at room temperature for 5 minutes. After 2 washes with 1X PBS, nuclei were isolated by resuspending cells in 10 times cell pellet volume of cell lysis buffer (10mM Tris-Cl pH8.0, 1.5mM MgCl2, 10mM KCl,0.5mM DTT, 0.05% NP-40, 1X mPIC and 0.2µM PMSF) and incubating on ice for 30 minutes. All buffers were made with RNase-free water. Nuclei were pelleted by centrifugation and washed once with cell lysis buffer. Nuclei were permeabilised using 0.7% SDS made with NFW at incubating at 62°C for 15 minutes. SDS was quenched by the addition of Triton X-100 to a final concentration of 10%. 50µl of 10X DpnII buffer and 375U of DpnII restriction enzyme was aded and the reaction was incubated for 1 hour at 37°C with shaking at 900rpm. The reaction was heat inactivated at 62°C for 20 minutes. Biotin was incorporated to the DNA ends by addition of biotin master mix (3µM each of dCTP, dGTP,dTTP and Biotin 14-dATP, 50U of DNA polymerase I, Large (Klenow)) and incubating the reaction for 37°C for 1 hour with rotation at 500rpm. In-situ ligation was carried out by adding ligation master mix (1X NEB Ligation buffer, 0.8% Triton X-100, 0.5X BSA, 1600U of T4 DNA ligase buffer) and incubating at 21 °C for 4 hours.

2.23.2 Chromatin Immunoprecipitation

Nuclei were pelleted by centrifugation and resuspended in 200 µL of SDS Lysis Buffer (1% SDS, 10 mM EDTA, 50 mM Tris, pH 8.1, 1X mPIC, 0.2µM PMSF) and incubated for 10 minutes on ice. The lysate was sonicated using the Diagenode Bioruptor for 20 cycles on High, 30 seconds On/ 30 seconds Off cycles to enrich chromatin in the size range of 200-500 basepairs. The lysate was centrifuged at 14,000Xg to sediment out the debris. The supernatant was taken and diluted 10 times in ChIP 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, 1X mPIC, 0.2µM PMSF). 5%-10% of the lysate was kept aside as input control for the experiment. To each of the tubes, 15l of anti-CTCF antibody (Cell Signaling Technology, Catalog number :

3418) was added as per manufacturer's instructions and incubated overnight at 4°C with rotation. 35µl of equilibrated Protein A Dynabeads were added to the tubes to capture the antibody and incubated at 4°C for 3 hours with rotation. The beads were captured using a magnetic rack (Thermo Fisher Scientific, Catalog number : CS15000) and washed once each with low-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), Lithium chloride wash buffer (0.25 M LiCl, 1% IGEPAL-CA630, 1% deoxycholic acid (sodium salt), 1 mM EDTA, 10 mM Tris, pH 8.1) and twice with T.E buffer for 3 minutes each with rotation at 4°C. DNA was eluted from the beads using freshly prepared elution buffer (1%SDS, 0.1M $NaHCO_3$) for 30 minutes with rotation at room temperature and was subjected to proteinase K treatment and reverse crosslinking overnight at 65°C in the presence of NaCl to a final concentration to 0.2M. Sample was purified using Zymo ChIP DNA Clean and concentrator kit kit and eluted in 40µl of NFW.

2.23.3 Biotin pulldown and Illumina sequencing sample preparation

30 µl of 10 mg/ml Dynabeads MyOne Streptavidin T1 (Thermo Fisher Scientific , Catalog number: 65601) beads were washed with 80 µl of 1× Tween Washing Buffer (1× TWB: 5mM Tris-HCl (pH 7.5); 0.5mM EDTA; 1M NaCl; 0.05% Tween 20). Beads were resuspended in 60 µl of 2X Binding Buffer (10mM Tris-HCl pH 7.5, 1mM EDTA, 2M NaCl) and added to the samples. The mixture was incubated for 15 minutes with rotation to bind biotinylated DNA to the streptavidin beads. Beads were washed twice by adding 120 µl of 1× TWB and heating the tubes at 55°C for 2 min with mixing. The beads were washed with 20µl 1× NEB T4 DNA ligase buffer. Beads were resuspended in master mix (85 µl of 1× NEB T4 DNA ligase buffer, 5 µl of 10 mM dNTP mix, 5 µl of 10 U/µl NEB T4 PNK, 4 µl of 3 U/µl NEB T4 DNA polymerase and 5U of NEB DNA polymerase I, Large (Klenow) Fragment) and incubated for 30 minutes at room temperature to repair ends of sheared DNA and remove biotin. Following 2 washes with 1X TWB as previously described, beads were resuspended in 6µl of EP buffer and 54µl of MilliQ from NEB Illumina prep kit. Adapter ligation and sequencing primer addition was performed using the NEBNext Ultra DNA Library prep kit as described by the manufacturer. After PCR for primer addition, fragments in the size range of 300bp-700bp were size selected by using the AMPure XP beads as per manufacturer's instructions. The sequencing library resuspended in NFW and QC was performed using Qubit and Bioanalyser and was submitted to Novogene (UK)

Company Limited. Paired end sequencing with read length of 150bp was performed and samples were sequenced to a depth of 150 Million reads per samples on Illumina platform.

2.24 RNA-sequencing

Total RNA was isolated from $1X10^6$ cells using Trizol from TaKaRa according to manufacturer's instructions. Quality of the RNA was checked on agarose gel and total RNA was submitted to Nucleome Informatics Private Limited (Hyderabad, India) for library preparation and RNA sequencing. Initial QC was performed using Qubit and sequencing library was prepared using the NEB Next ultra II mRNA Library Prep Kit. QC for libraries was performed using Bioanalyser. The library was paired-end sequenced with read length of 150bp on Illumina NovaSeq 6000 to a depth of 40 Million reads per samples.

2.25 ChIP-sequencing

Chromatin immunoprecipitation was performed as previously described using either antibody against CTCF from Cell Signalling technology (Catalog number 3418) or against YY1. After reverse cross linking and proteinase K treatment, samples were purified using the Zymo ChIP DNA Clean and concentrator kit. Samples were submitted to Nucleome Informatics Private Limited (Hyderabad, India) for library preparation and ChIP sequencing. Sequencing library was prepared using the NEBNext Ultra DNA Library prep kit and QC was performed using Bioanalyzer. Paired end sequencing with read length of 150bp was performed and samples were sequenced to a depth of 30 Million reads per samples on Illumina NovaSeq 6000.

2.26 Systems analysis

2.26.1 Triplexator prediction

To identify all putative triplexes that can form between *mrhl* lncRNA (2.4knt) and Sox8 promoter (+500bp to -1500bp from TSS) , analysis was run with default parameters as specified by Buske F.A. et al (Buske, F.A. et al, 2012) to identify TFO-TTS pairs in single-strand and duplex sequences. Some of the important parameters defined are that the Triplex forming oligo (TFO from RNA) and triplex target site (TTS from DNA) pair must be at least 15 bps in length, have at most 20% errors, tolerate up to 2 consecutive errors, require a guanine ratio of at least 20% and use the purine motif only.

2.26.2 ChIP-sequencing data analysis

Raw FASTQ files were downloaded from the NCBI GEO repository, and were re-analysed to generate the aligned files for the visualisation of regions of interest. All aligned files were aligned to mouse genome (mm10) using Bowtie2 (Langmead, B. and Salzberg ,S. 2012) and then sorted, indexed, made free from PCR duplicates using Samtools (Li, H. et al, 2009). Aligned files were loaded in the IGV genome browser to visualise the enrichment of peaks at the regions of interest. Peak calling was done with the MACS2 (Feng, J. et al, 2012) pipeline.

2.26.3 RNA-sequencing data analysis

Raw FASTQ files were downloaded from the NCBI GEO repository, and were re-analysed with the TopHat (Kim, D. et al, 2013) and Cufflinks (Trapnell, C. et al, 2010) pipeline. Aligned files were loaded on the IGV (Robinson, J.T et al, 2011) genome browser to visualise the gene expression enrichment.

A list of all publicly available datasets used for analysis has been given below in **table 2.2**.

ChIP-seq datasets	GEO Accession number
mESC CTCF	GSM723015
mESC SMC1	GSM560341
mESC RNA Pol II	GSM723019
mESC H3K4me3	GSM723017
mESC Input	GSM723020
Mouse brain cortex CTCF	GSM722631
Mouse brain cortex SMC1	GSM1838869
Mouse brain cortex RNA PolII	GSM722634
Mouse brain cortex H3K4me3	GSM722633
Mouse brain cortex input	GSM722635
RNA-seq datasets	
mESC	GSM723776
Adult Brain cortex	GSE96684

Table 2.2 - List of publicly available datasets analysed

A list of all the primers and oligonucleotides used in the study is given below in table 2.3

	Forward (5' - 3')	Reverse (5' - 3')
Primers for 3C		
Sox C	CCAAGTGCAGCTAGGAGTCTCTC	
Sox test 1	AGCACCTGCGACACGGCATC	
Sox test 2	CTGGGAGCAGTACCTGCCAGAGG	
Sox test 3	GGCAGAAGTTTGGATATCCAGAAGC	
Sox test 4	GCCTGCCTCTGTCTACGCTTGG	
Sox test 5	CCAGTGCTTGAAACTCAATGGATGG	
Sox test 6	TCTCTCTGCTCGCCCTCATCC	
Sox test 7	CTGCAATCCCAGCACTGGAG	
Ercc3 1	CTGACCCTCAGCCTGTTAGAGC	
Ercc3 2	ACCAGTCTTGCCTTGTGTCAGC	
ChIP RT primers		
Sox8 promoter	AGAGGGCTAAGGGTGACTGACT	GTTTGGTTGCAATAGCGGATTC
Sox8 exon3	GATAACCTCGCTGCTGAGCTCGG	CTGGTGTCACCCACCAGCTCC
Sox8 enhancer 8kb	CCGCTATCCAGATCACCAGG	CTGCTGAGTGACCGATGAGAC
Sox8 enhancer 16kb	GCCTCAGGACTCACATCTGGC	TGTGGGTCCTTGCCAGGAGC
Sox8 triplex region	CCTTAATGGTGACCTTATTCTATTCTAG	CCTTTCTTGGCAGGTAATGG
Actin promoter N.C.	CGCTCACTCACCGGCCTC	GTCCGGGCCTCGATGCTG
Gene desert N.C.	TGGCTGTCCTGGCCTGC	GGCAGCCTATGCAGCATTCAATG
Quantification primers		
Mrhl	TGAGGACCATGGCTGGACTCT	AGATGCAGTTTCCAATGTCCAAAT
Beta-actin	AGGTCATCACTATTGGCAACG	TACTCCTGCTTGCTGATCCAC
Phkb	AAGCCCAGCAATGAGGACTC	AGCACCCACCACACTAACAC
Sox8	TGCTGAGCTCTGCGTTATGGAG	GTCTGGTGCCTATGCCTGTGC
EMSA oligonucleotides		

	Forward (5’ - 3’)	Reverse (5’ - 3’)
Positive control DNA	AGAGAGAGGGAGAGAG	CTCTCTCCCTCTCTCT
TTS1	GGGAGGGAGACAGAGAGG	CCTCTCTGTCTCCCTCCC
TTS2	GGAAGAGGGAGGGAGA	TCTCCCTCCCTTCTTCC
TTS3	AGACAGAGAGGGA	TCCCTCTCTGTCT
TTS4	AGCAGGAAGCAGG	CCTGCTTCCTGCT
TTS5	AGAGGGAAGAGGG	CCCTCTTCCCTCT
TTS Negative control	ACCACGTGGGCCAGGCGC	GCGCCTGGCCCACGTGGT
TTS mutant	GGAACACCCACCCAGA	TCTGGGTGGGTGTTCC
Positive control RNA	CGGAGAGCAGAGAGGGAGCG	
TFO1	UGAGAGAGAGAGAUGG	
TFO2	AGAAGAAGGAAGAC	
Negative control RNA	CUUAUACUGCAUAAAU	
In nucleus pulldown oligo		
TFO1	psoralen - UGAGAGAGAGAGAUGG - biotin	
NC TFO	psoralen - CUUAUACUGCAUAAAU- biotin	

Table 2.3 - List of primers and oligonucleotides used in the study

Chapter 3

Results

3.1 Triplex binding site detection

Many chromatin bound lncRNAs associate with chromatin through the formation of the triple helix. *Mrhl,* too, is a chromatin associated lncRNA and it is possible that this 2.4kb long lncRNA binds at the Sox8 locus through triple helix in addition to its p68- dependent binding at the Sox8 promoter. Most lncRNAs form triplexes at regions rich in the polypurine AG repeat motifs. To explore whether triple helix formation between *mrhl* lncRNA and the Sox8 locus was a possibility, a primary sequence analysis was done manually at the Sox8 locus in the region upto 2kb upstream of the TSS to look for regions rich in AG repeat motifs. At around 1.2 kb upstream of the TSS, a region harbouring nearly 50 AG-repeats was discovered and was considered as a candidate region for triplex formation **(Figure 3.1)**.

Fig 3.1 - Homopurine (AG) repeats 1.2kb upstream of Sox8 TSS highlighted by the red box

Next, triplex formation prediction was carried out *in silico* using the tool Triplexator (Buske F.A et al, 2012) in order to understand the propensity for triplex formation by *mrhl* at the Sox8 locus. This computational framework works by identifying the triplex-forming oligonucleotides (TFO) in single-stranded sequences, identifying the triplex target sites (TTS) in double-stranded sequences able to accommodate a third strand and assessing the compatibility of potential TFO/TTS pairs. Full length of *mrhl* lncRNA and the Sox8 promoter region of -2kb to +500 bp from TSS were the inputs for Triplexator analysis and a list of probable TFO/TTS pairs were obtained as an output of the prediction software. The complete list is given in **table 3.1.**

TRIPLEXATOR PREDICTION									
TFO start	TFO end	TTS start	TTS end	Score	Error-rate	Errors	Motif	strand	orientation
90	100	1416	1426	8	0.2	o3t7	R	-	A
390	400	302	312	8	0.2	t2t6	R	-	A
390	400	948	958	8	0.2	t4d5	R	-	A
391	404	1224	1237	11	0.15	d6o10	R	-	A
389	400	1224	1235	10	0.091	d6	R	-	A
389	404	1226	1241	12	0.2	d4t9o12	R	-	A
389	400	1228	1239	9	0.18	d2t7	R	-	A
390	401	1231	1242	9	0.18	t4t8	R	-	A
389	404	1232	1247	12	0.2	t3t7o12	R	-	A
389	399	1234	1244	8	0.2	t1t5	R	-	A
389	401	1245	1257	10	0.17	t1d8	R	-	A
389	399	1247	1257	9	0.1	d6	R	-	A
391	401	1352	1362	8	0.2	t7d8	R	-	A
389	399	1352	1362	8	0.2	t7d8	R	-	A
365	375	302	312	8	0.2	o3t7	R	-	A
365	375	1103	1113	8	0.2	o3d5	R	-	A
365	375	1233	1243	8	0.2	o3t7	R	-	A
710	720	1235	1245	8	0.2	o3t4	R	-	A
709	720	1416	1427	9	0.18	o4t7	R	-	A
696	706	1246	1256	8	0.2	o3b7	R	-	A
906	916	949	959	8	0.2	t3b4	R	-	A
906	916	1226	1236	9	0.1	b4	R	-	A
1373	1383	302	312	8	0.2	t3t7	R	-	A
1378	1388	1105	1115	8	0.2	d3o5	R	-	A
1373	1383	1103	1113	8	0.2	t3d5	R	-	A
1377	1387	1237	1247	8	0.2	t3o6	R	-	A
1373	1383	1233	1243	8	0.2	t3t7	R	-	A
1455	1466	948	959	9	0.18	b5t9	R	-	A

1452	1463	1228	1239	9	0.18	d2o8	R	-	A
1456	1466	1236	1246	9	0.1	o4	R	-	A
1452	1462	1352	1362	8	0.2	t2b8	R	-	A
2030	2041	302	313	9	0.18	o3t6	R	-	A
2030	2040	1233	1243	8	0.2	o3t6	R	-	A
2362	2374	1106	1118	10	0.17	d2d9	R	-	A
2362	2374	1239	1251	10	0.17	t1t8	R	-	A

Table 3.1 : Results from Triplexator predictor software - Complete list of potential TFO/TTS pairs within mrhl lncRNA and Sox8 promoter

The score is indicative of the number of nucleotides within the defined length of 15bps which can participate in Hoogsteen/ Reverse- Hoogsteen base pairing in a given TFO/TTS pair. All predictions with a score of 10 and above were considered as potential triplex forming pairs for downstream experiments, a criterion used previously for the lncRNA Meg3 (Mondal T et al, 2015). The shortlisted TFO/TTS pairs are listed in **table 3.2** below. While most of the target site predictions were clustered around the AG repeat rich region of the Sox8 promoter, two regions of *mrhl* lncRNA were predicted to have triplex forming ability, one mapping to the 5' end of the RNA (region between 389 and 404 nts) and one mapping to the 3' end (region between 2362 and 2374 nts)

Shortlisted predictions									
TFO start	TFO end	TTS start	TTS end	Score	Error-rate	Errors	Motif	strand	orientation
391	404	1224	1237	11	0.15	d6o10	R	-	A
389	400	1224	1235	10	0.091	d6	R	-	A
389	404	1226	1241	12	0.2	d4t9o12	R	-	A
389	404	1232	1247	12	0.2	t3t7o12	R	-	A
389	401	1245	1257	10	0.17	t1d8	R	-	A
2362	2374	1106	1118	10	0.17	d2d9	R	-	A
2362	2374	1239	1251	10	0.17	t1t8	R	-	A

Table 3.2 : Shortlisted TFO/TTS Triplexator predictions of score 10 and above

3.2 Experimental validation of triplex formation *in vitro*

To check if any of the *in silico* predicted TTS/TFO pairs were capable of forming triple helix, Electrophoretic Mobility Shift Assay (EMSA) was performed for the shortlisted TFO/TTS pairs. The Triplex target sites were found to be clustered between the regions -1224 and -1257 of the Sox8 promoter. This 33nt region was split between 3 DNA oligos for EMSA experiments. Table 2.3 below outlines the oligonucleotides used for EMSA. In addition to the test oligonucleotides arising from the Triplexator predictions, a positive control TFO/TTS pair from *Meg3* lncRNA and its target TGFBR1 gene (triplex formation at this locus has been characterised by Mondal, T. *et al*, 2015) and a negative control TFO/TTS pair taken from within *mrhl* lncRNA and Sox8 promoter respectively with no AG bias were included for the EMSA.

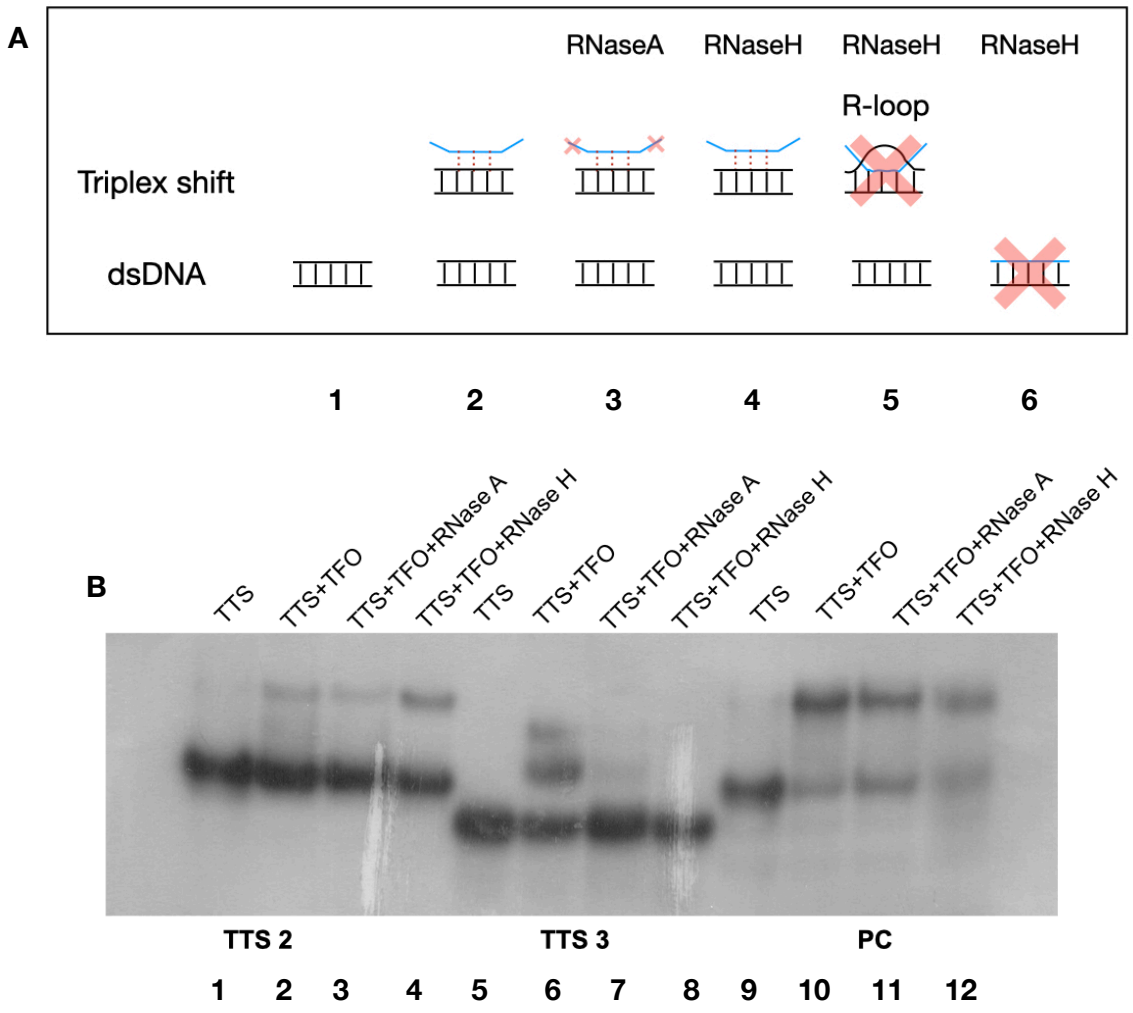
TTS	Sequence	vs	TFO	Sequence
Positive control	AGAGAGAGGGAGAGAG TCT CTC TCC CTC TCTC	-	**Positive control**	CGGAGAGCAGAGAGGGAGCG
TTS1	GGGAGGGAGACAGAGAGG CCC TCC CTCTGTC TCTCC	-	**TFO1**	UGAGAGAGAGAGAUGG
TTS2	GGAAGAGGGAGGGAGA CCT TCT CCCTCC CTCT			
TTS3	AGACAGAGAGGGA TCT GTCT CTCCCT			
TTS4	AGCAGGAAGCAGG TCG TCCTTCGTCC	-	**TFO2**	AGAAGAAGGAAGAC
TTS5	AGAGGGAAGAGGG TCT CCCTTCTCCC			
TTS Mutant	GGAACACCCACCCAGA CCTTGTGGGTGGGTCT	-	**TFO1**	UGAGAGAGAGAGAUGG
Negative control	ACCACGTGGGCCAGGCGC TGGTGCACCCGGTCCGCG	-	**Negative control**	CUUAUACUGCAUAAAU

Table 3.3 - Oligonucleotides used for Triplex EMSA

Using the experimental conditions which worked well for the positive control TFO/TTS pair (**Figure 3.2 B** lanes 9-12), EMSA was performed for the test TFO/TTS pairs. Of the 5 tested combinations, 2 pairs - TFO1/TTS2 and TFO1/TTS3, showed a shift in the mobility (**Figure 3.2 B** Compare lanes 1 and 2 and 5 and 6). Two additional experimental controls that were included were RNase A and RNase H digestions. RNase H selectively degrades

DNA:RNA hybrids and is included to ensure that the shift in mobility is due to the formation of triplex and not a hybrid duplex/ R-loop (**Fig 3.2 A** lane 5,6) while RNase A degrades single-stranded RNA. A shift due to the formation of triplex would be resistant to digestion with both of these enzymes (**Fig 3.2 A** lane 3, 4). Of the two pairs that showed a shift in the mobility, only TFO1/TTS2 mobility shift was resistant to RNase A and RNase H digestion and was considered as shift due to triplex formation (**Fig 3.2 B TTS2 lane 4**). To ensure enzyme activity of RNase H, a DNA:RNA hybrid dimer was subjected to RNase H digestion (**Figure 3.2 C**).

To ensure that the mobility shift observed in TFO1/TTS2 pair was indeed due to the participation of the AG motif, a mutant DNA oligo was synthesised with all G nucleotides changed to C nucleotides. There was no shift in mobility observed when this oligonucleotide probe was used for EMSA in combination with RNA TFO1 (**Figure 3.2 D**). The negative control TFO/TTS pair also showed no shift in mobility under the experimental conditions (**Figure 3.2 E**)



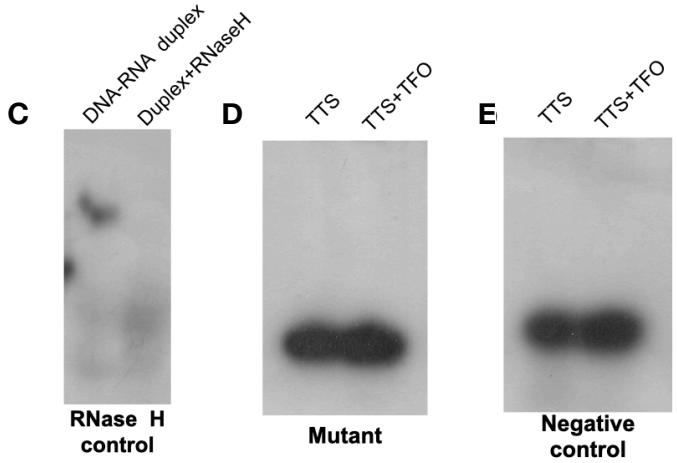


Fig 3.2 - Electrophoretic mobility shift assay to detect triplex formation in vitro. (A) *The schematic depicts the possibilities and expected results of the EMSA. The duplex (lane 1) appears as a lower band in the lane and an upward shifted band is expected when triplex forms (lane 2). The triplex band is resistant to RNase A digestion but overhanging single stranded regions of the RNA oligo could be cleaved by the enzyme (lane 3) resulting in either no difference in the observed shifted band, the appearance of a smaller sized triplex band or reduction in the band intensity of the triplex band. The triplex band is resistant to RNase H digestion (lane 4) unless the RNA forms an R-loop structure with the dsDNA oligo (lane 5). DNA:RNA duplex either on its own (lane 6) or in an R-loop (lane 5) is degraded by RNase H.* ***(B)*** *Shift in mobility is detected in TFO1/TTS2 and TFO1/TTS3 pairs and in the positive control (PC) samples. The shift in mobility in PC and TFO1/ TTS2 are resistant to RNase H enzyme digestion while the mobility shift in TTS3/TFO1 is lost upon RNase H enzyme digestion.* ***(C)*** *RNA:DNA duplex degrades with RNase H enzyme digestion.* ***(D)*** *When the G nucleotides of TTS2 are changed to C nucleotides in the mutant oligonucleotide, there is a loss in mobility shift* ***(E)*** *Negative control TFO/TTS pair does not show shift in mobility.*

Based on the results from EMSA, TFO1/TTS2 was considered as a potential triplex forming TFO/TTS pair and was taken forward for subsequent experiments. To further confirm triplex formation *in vitro* between the pair, Circular Dichroism spectra for the TFO-TTS test and TTS/ negative control TFO (NC TFO) pairs were recorded. Spectra were recorded for TFO1/TTS2 or NC TFO/TTS2 pairs in addition to spectra for TTS2 only, TFO1 only and NC TFO only. The spectrum recorded for the test oligos TFO1/TTS2 showed characteristics of a triplex (Mondal, T. *et al*, 2015) (**Figure 3.3 B**) with two minima- a strong minima at 210nm and an additional one at 240nm (whereas the spectrum for the control oligo NC TFO/TTS2 pair showed a strong minima at around 240nm (**Figure 3.3 A**). As an additional control, artificial spectrum was generated by plotting the sum of the individual spectra for either TTS2 and TFO1 pair (**Figure 3.3 D** in red) or TTS2 and NC TFO pair (**Figure 3.3 C** in red). When the artificial spectrum was overlaid with the spectrum obtained for the triplex reaction, an overlap in the spectra for the control oligo NC TFO + TTS2 (**Figure 3.3 C**) was observed. But such an overlap was not seen for the TFO1 containing spectra (**Figure 3.3 D**) indicating that the spectrum for the reaction containing NC TFO was due to the contributions of the individual components and not because of the

formation of any secondary structure whereas that the spectrum recorder for TFO1/TTS2 was indeed due to the formation of a triple helix.

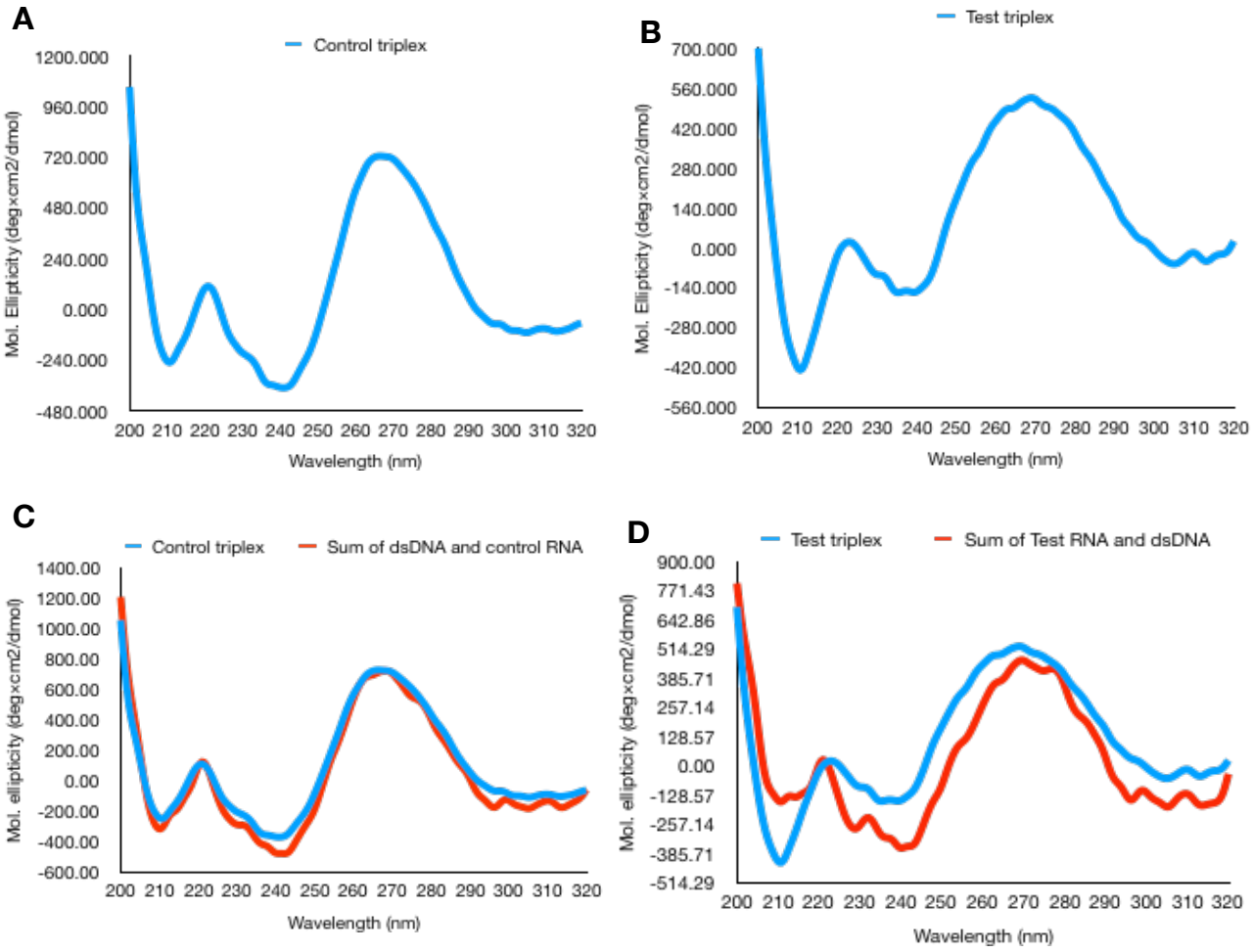


Fig 3.3 - Circular dichroism spectra for triplex oligonucleotides. ***(A)*** *CD spectrum recorded for the control oligonucleotide containing NC TFO/TTS2 pair,* ***(B)*** *CD spectrum recorded for/TFO1/ TTS2 pair,* ***(C)*** *Artificial spectrum generated by summation of individual CD spectra recorded for TTS2 only and NC TFO only (in red) overlaid with spectrum of triplex reaction for the oligonucleotides,* ***(D)*** *Artificial spectrum generated by summation of individual CD spectra recorded for TTS2 only and TFO1 only (in red) overlaid with spectrum of triplex reaction for the oligonucleotides. Plots are an average of 4 independently recorded spectra.*

3.3 Experimental validation of triplex formation *in vivo*

In an attempt to understand if the triplex helix forms within the cell at the Sox8 locus, we resorted to triplex affinity pulldown assay in the nucleus of the mouse B-type spermatogonial cell line Gc1-spg. Briefly, TFO1 or NC TFO RNA oligonucleotides with biotin modification on the 3' end and a psoralen modification on the 5' end was incubated with the nuclei of cells. The psoralen intercalates with DNA in the presence of UV light and helps to cross-link the RNA oligonucleotide to its site of interaction on the chromatin. Biotin affinity pulldown was carried out to enrich for chromatin regions interacting with the

oligonucleotides and their enrichment was quantified by qPCR. As a negative control, the enrichment for a region within the promoter of beta-actin gene was scored for in addition to the region harbouring TTS2 at the Sox8 promoter. RNase H enzyme digestion was performed to ensure that the interaction between the oligonucleotide and chromatin was due to triple helix formation and not DNA:RNA duplex. The data was plotted as fold enrichment of genomic regions with test oligo over control oligo. The results for the pulldown assay are given in **figure 3.4**.

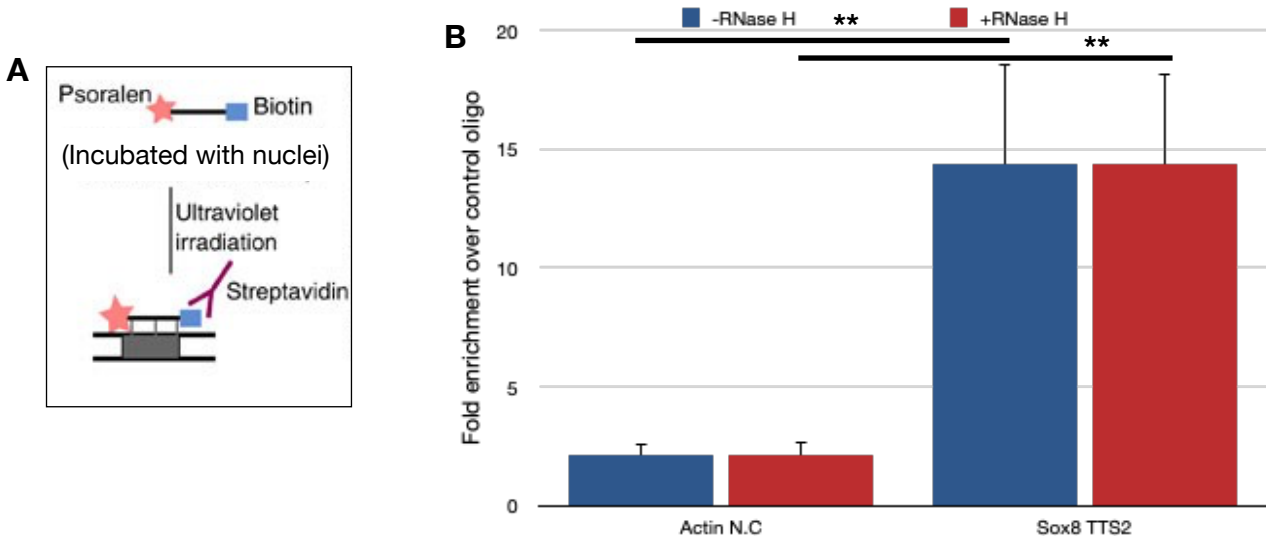


Fig 3.4 : Triplex pulldown assay from Gc1-spg cell nuclei. ***(A)*** *Psoralen and biotin labeled RNA oligonucleotides are incubated with the nuclei. UV irradiation forms crosslink between psoralen, a photoactivable moiety, and the chromatin and Biotin modification helps in affinity pulldown. Figure has been adapted with permission from Mondal, T. et al, 2015* ***(B)*** *Almost 15 fold enrichment of the Sox8 promoter locus was obtained in chromatin associated with TFO1 as compared to NC TFO. This enrichment was specific to the Sox locus and was not seen at the negative control locus from within the beta-actin promoter. The enrichment did not decrease upon treatment with RNase H enzyme. Data in the graph has been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (two-tailed Student's t test)*

Sox8 promoter region was significantly enriched (~15 fold) in the TFO1 pulldown fraction when compared to NC TFO associated chromatin fraction. Such enrichment was not observed at the negative control beta-actin promoter region. This enrichment was unaffected by RNAseH digestion indicating that the RNA oligo interacts at the Sox8 promoter within the cells through the formation of a DNA:DNA:RNA triplex.

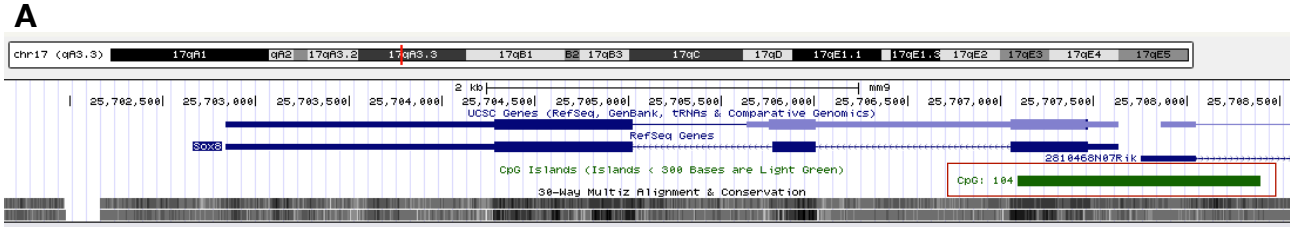
Based on these experiments, we conclude that it is very likely that *mrhl* lncRNA interacts with the Sox8 locus at around -1200 bp from TSS through the formation of a triplex in addition to the p68- dependent interaction at -140 bp.

3.4 Epigenetic Mechanism of gene repression at the Sox8 locus

One of the modes of epigenetic gene repression is through the methylation of DNA. DNA methylation in the promoter region of genes leads to gene inactivation and silencing. In mammals, DNA methylation is found primarily at cytosine–guanosine dinucleotides (CpGs). CpG sites occur at lower than expected frequencies throughout the mammalian genome but are found more frequently at stretches of DNA called CpG islands, typically found in or near promoter regions of genes (Baylin, S. 2005. A CpG island of 1.3kb encompassing the TSS and promoter of Sox8 was observed in UCSC Genome Browser (**Figure 3.5 A**). We wanted to see if there was any correlation between the methylation status of this CpG island and Sox8 expression status. The possibility of DNA methylation regulating Sox8 gene expression was particularly of interest since triplex formation by pRNA (promoter-associated non-coding RNA that is complementary to the promoter of rDNA) at the rDNA locus acts as a platform for the recruitment of DNMT3b which methylates the DNA at the locus to repress transcription (Shmitz, K.M. *et al*, 2010). We wanted to investigate the methylation status as the first step to understand if triplex formation by *mrhl* lncRNA at the Sox8 locus could also be recruiting DNMT3b to the locus.

We expected to see a concomitant decrease in CpG methylation with activation of Sox8 expression if this was the epigenetic mechanism of gene repression. The methylation status was probed in Gc1-spg cells cultured in control media and in Wnt conditioned media. As shown previously, *mrhl* is downregulated and Sox8 expression is activated upon Wnt induction in Gc1-spg cells (Arun, G. et al, 2012). The methylation status was also probed in the mice of testis of ages P7 and P21. The testis from mice of these two age groups closely resemble control Gc1-spg cells and Wnt induced Gc1-spg cells respectively and have been used to confirm the biological relevance of key observations involving *mrhl* lncRNA previously (Akhade, V.S et al, 2015). However, no significant reduction in methylation levels was observed upon Sox8 transcriptional activation in either the cells or in the testes (**Figure 3.5 C**). In fact, a slight increase in the methylation levels was observed. Based on these results, we concluded that Sox8 transcriptional repression is not due to DNA methylation at the promoter.

A



B **CpG Island Info**

Position: chr17:25707089-25708396
Band: 17qA3.3
Genomic Size: 1308
View DNA for this feature (mm9/Mouse)
Size: 1308
CpG count: 104
C count plus G count: 859
Percentage CpG: 15.9%
Percentage C or G: 65.7%
Ratio of observed to expected CpG: 0.74

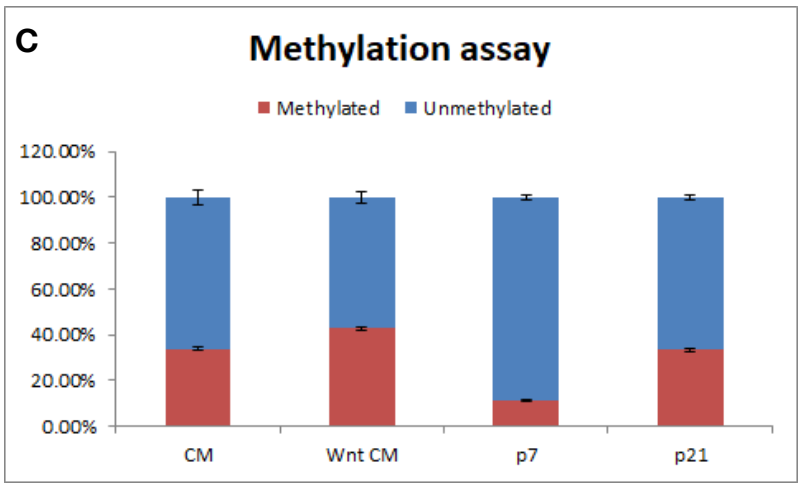


Fig. 3.5 - CpG island at Sox8 promoter and its methylation status. (A) *CpG island number 104 found at the Sox8 locus is shown in dark green.* ***(B)*** *Information about the CpG island* ***(C)*** *Methylation assay performed to score for % methylation of Sox8 promoter in Control (CM) and Wnt treated (Wnt CM) media and P7 and P21 testes showed no decrease in methylation levels upon Sox8 transcriptional activation. In fact, an increase in the % methylation was observed. Data in the graph has been plotted as mean ± S.D., N=3.*

A characteristic feature of a silenced gene is the association of the repressive histone modification H3K27me3. From previous work done on this locus, it is known that H3K27me3 levels are relatively high at the promoter in the Sox8 repressed state when *mrhl* is bound at the promoter (Kataruka, S. et al, 2017). The levels of this histone modification reduces upon the transcriptional activation of Sox8 (**Figure 3.6 A**). The Polycomb Repressive Complex 2 (PRC2) is responsible for this histone modification. This multicomponent complex has 4 core subunits - Suz12, Ezh1/2, Eed and RbAp46/48 in addition to several auxiliary subunits (Margueron, R., & Reinberg, D., 2011). The enzymatic subunit Ezh2 has been reported to interact with multiple lncRNAs (Wang, Y. *et al*, 2018) and it has been proposed that lncRNA confers target specificity to PRC2.

We performed ChIP for the Ezh2 subunit of PRC2 in control and Wnt induced Gc1-spg cells to check for its presence at the locus. Ezh2 occupancy at the Sox8 locus was observed in the control conditions when Sox8 is repressed. A reduction in Ezh2 occupancy was noted upon Wnt induced Sox8 activation (**Figure 3.6 D**). We concluded that PRC2 is one of the complexes involved is transcriptional regulation of the Sox8 gene. While lncRNAs such as Xist interact with PRC2 through a secondary structure within the RNA (Wutz, A. et al., 2002), triplex formation by lncRNAs PARTICLE and Meg3 recruit PRC2 to target loci through triplex formation (O'Leary, V.B. et al, 2015, Sherpa et al., 2018). *Mrhl* lncRNA is likely to be involved in the recruitment of PRC2 either through triplex formation or independent of it.

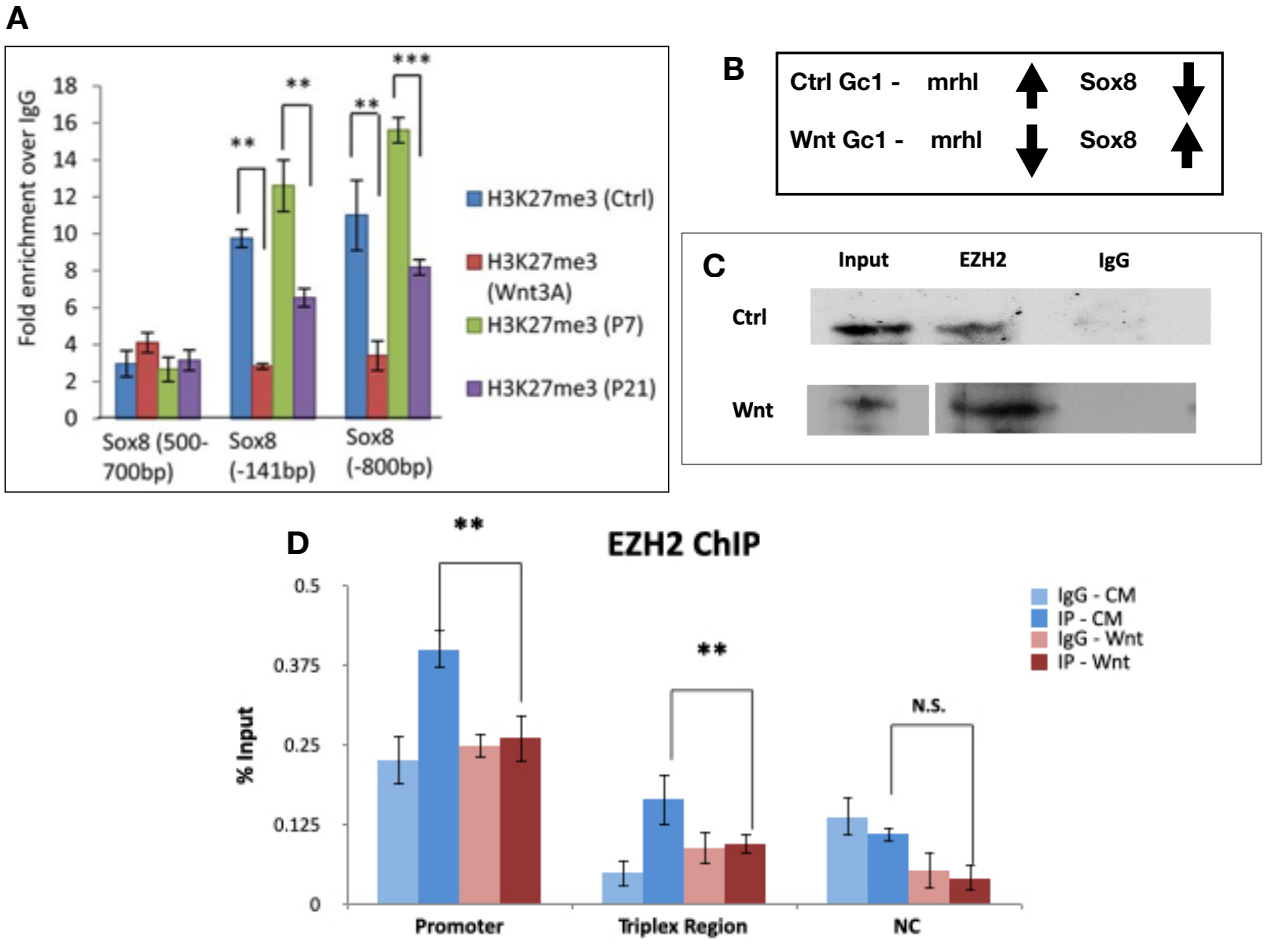


Fig 3.6 - PRC2 complex at the Sox8 locus - (A) *Higher levels of H3K27me3 mark are observed at the Sox8 locus in conditions of transcriptional repression and the levels of this modification reduce when transcription is active. Image has been taken from Kataruka, S. et al, 2017. (B) Schematic outlining the expression status of mrhl lncRNA and Sox8 in Gc1-spg without and with Wnt activation. (C) ChIP western blotting for Ezh2 subunit of PRC2 in Gc1-spg cells under the two conditions shows enrichment of the protein in IP reaction over the isotype control reaction. (D) ChIP-qPCR for Ezh2 subunit of PRC2 shows occupancy of the protein at the Sox8 locus, both at the promoter close to mrhl ChOP site and at the triplex target site. Difference in occupancy of Ezh2 at the negative control site was found to be non-significant. Data in the graph has been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (Student's t test)*

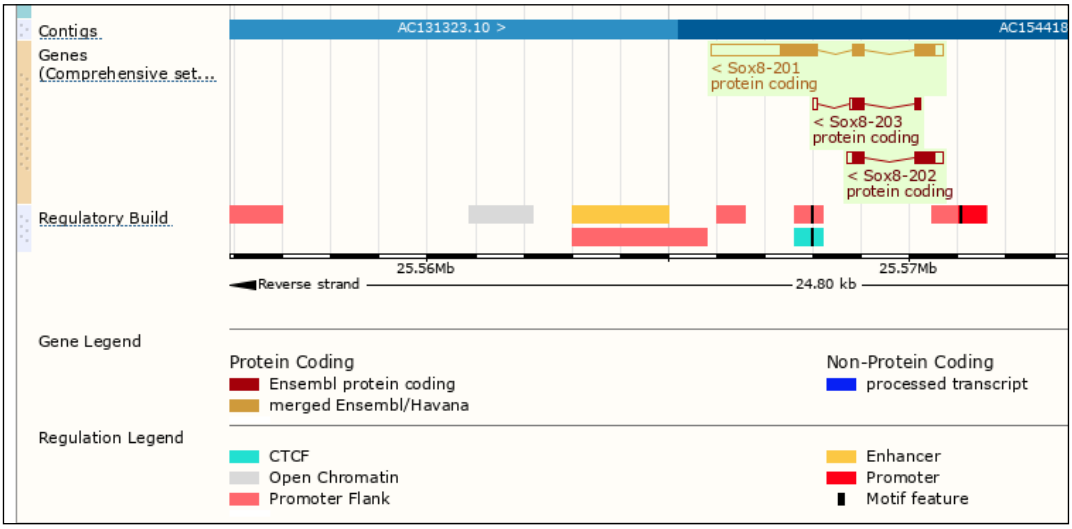
3.5 Involvement of the architectural proteins CTCF and cohesin in Sox8 regulation.

The transcriptional repression of Sox8 by *mrhl* lncRNA and its protein partner p68 was reminiscent of gene repression at the H19/Igf2 imprinted locus by SRA lncRNA and p68. As previously discussed, two additional players are involved in regulating the alleles specific expression of genes at this imprinted locus, namely, CTCF and cohesin (**Fig 1.13**).

We explored the Sox8 locus in ENSEMBL database as a first step to look for the possible involvement of CTCF and cohesin in gene regulation. Sure enough, we found a CTCF binding site within exon 3 of the Sox8 gene (**Figure 3.7. A**). It was seen that cohesin

subunits Rad21 and SMC3 too bound very close to the CTCF binding site (**Figure 3.7 B**). Moreover, the binding of CTCF at this site was only observed in tissues where Sox8 was not expressed (kidney, liver and spleen). In fact, the tissues which showed CTCF occupancy correlated with the tissues with *mrhl* lncRNA expression (kidney and liver but not in brain and heart) (**Figure 3.7 C**).

A



B

Cell type	Evidence type	Feature name	Location
kidney:a8w	DNase1 & TFBS	CTCF	17:25567774-25568178
liver:a8w	DNase1 & TFBS	CTCF	17:25567769-25568410
liver:a8w	Hists & Pols	H3K4me1	17:25568060-25568320
MEL	DNase1 & TFBS	CTCF	17:25567743-25568224
MEL	DNase1 & TFBS	SMC3	17:25567746-25568244
MEL	DNase1 & TFBS	Rad21	17:25567747-25568223

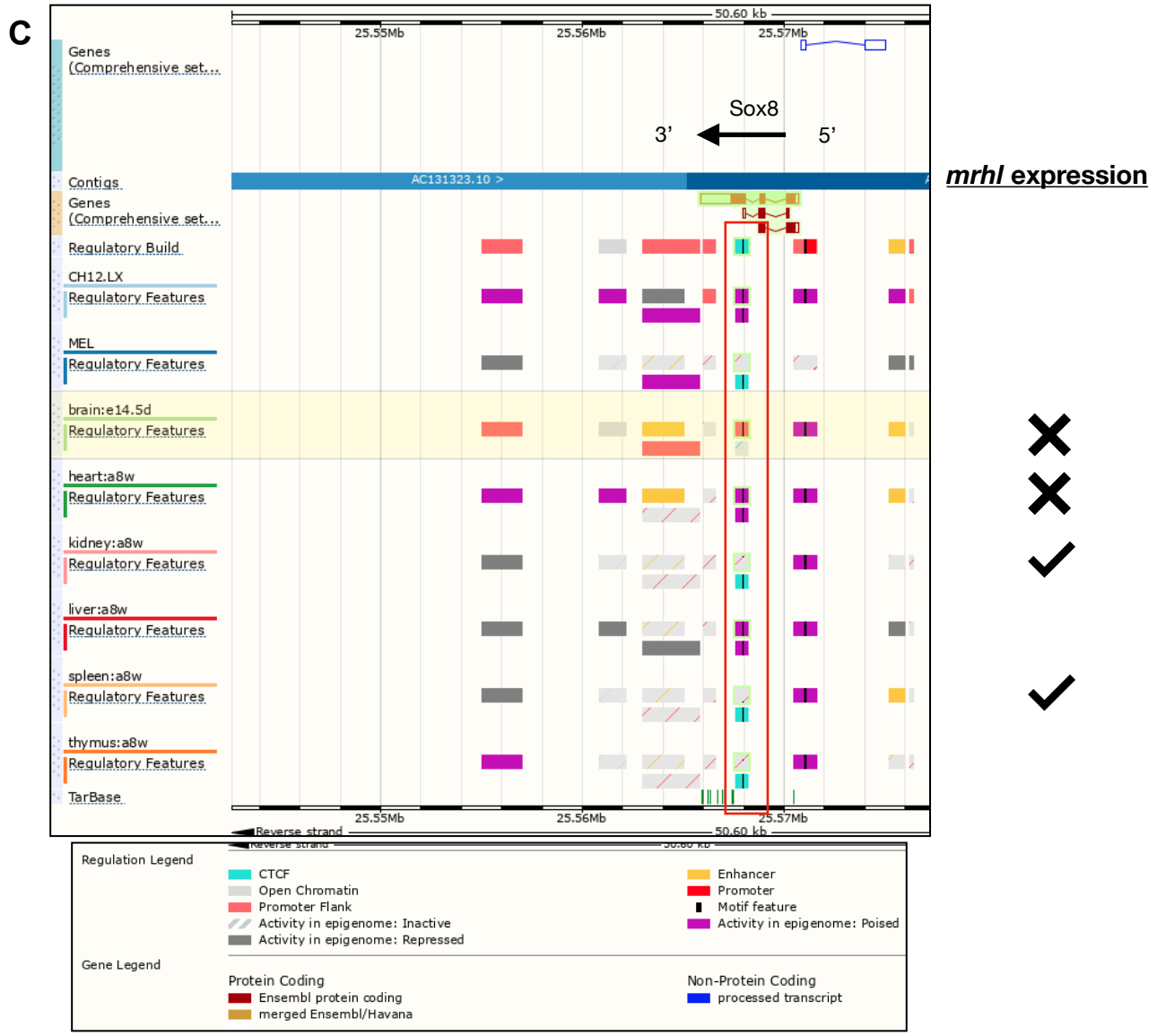
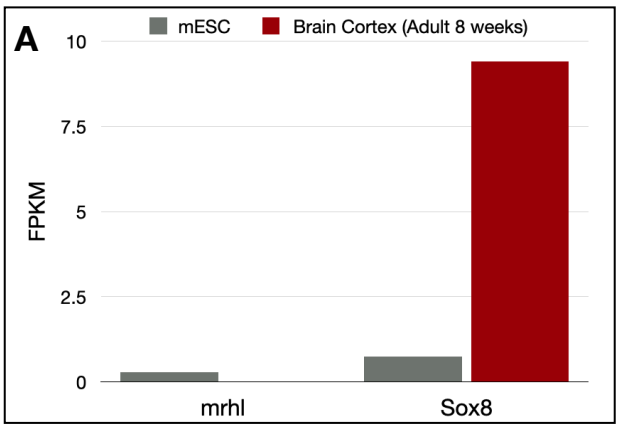


Fig 3.7: CTCF and cohesin at Sox8 locus. (A) *CTCF binding site (in cyan) is present in exon3 of Sox8.* ***(B)*** *Cohesin subunits SMC3 and Rad21 bind close to the CTCF binding site at the Sox8 exon3 only in those tissues in which CTCF is bound.* ***(C)*** *Tissues show correlation between CTCF binding (highlighted in red box) at the cognate binding site and expression of mrhl lncRNA .CTCF is not bound at its cognate binding site in brain and heart tissues in which mrhl is not expressed, but is bound in kidney and spleen tissues in which mrhl is expressed. The expression status of mrhl is indicated to the right side as either tick marks (expressed) or cross marks (not expressed). CTCF binding correlates with lack of activity in the Sox8 gene locus indicative of inverse correlation between CTCF binding and Sox8 expression.*

To understand the occupancy of these two proteins at the Sox8 locus relative to the transcriptional expression state of the gene, we wanted to analyse publicly available ChIP-seq datasets. However, ChIP-seq datasets for all proteins and histone modifications of our interest were not available for spermatogonial cells or mouse testes of the appropriate age groups. Since the CTCF occupancy at the Sox8 cognate binding site correlated with *mrhl* lncRNA expression and CTCF occupancy negatively correlated with Sox8 expression (**Figure 3.7 B,C**), we performed analysis of ChIP-seq datasets in two surrogate systems -

the mouse embryonic stem cell system corresponding to *mrhl* expressed, Sox8 repressed state and the mouse adult brain cortex corresponding to *mrhl* repressed, Sox8 expressed condition. These two systems were specifically chosen since *mrhl* lncRNA expression status in them has been explored by members of our group previously (Pal, D. et al, 2020 and Pal, D. Ph.D Thesis).

First, the inverse correlation between the expression levels of *mrhl* lncRNA and Sox8 was confirmed from publicly available RNA-seq data (**Figure 3.8 A**), ChIP-Seq datasets were analysed to look at the occupancy of CTCF and cohesin (SMC1 subunit) and RNA pol II and the histone modification H3K4me3 datasets indicative of transcriptional status of the gene. The Sox8 locus was visualised using Integrated Genome Browser (**Figure 3.8 B**). In agreement with information available in ENSEMBL, CTCF and cohesin occupancy was observed at the binding site within exon 3 of Sox8 gene in mESC (highlighted in black), a tissue in which *mrhl* is expressed and Sox8 expression level is low (relatively low levels of H3K4me3 and RNA PolII occupancy at the TSS). CTCF and cohesin occupancy was notably reduced in mouse adult brain cortex **(Figure 3.8 C)**, a tissue in which *mrhl* lncRNA levels are low and Sox8 is expressed (higher levels of H3K4me3 and RNA PolII occupancy). Interestingly, an additional peak for both CTCF and cohesin (SMC1) was observed at the Sox8 promoter (highlighted in red), very close to the *mrhl* binding site where the presence of a CTCF binding site is not indicated in the ENSEMBL database. Occupancy of the protein partners p68 has been previously observed at the promoter of Sox8 and the binding of *mrhl* at this site is dependent on p68 (Akhade, V.S et al, 2015).



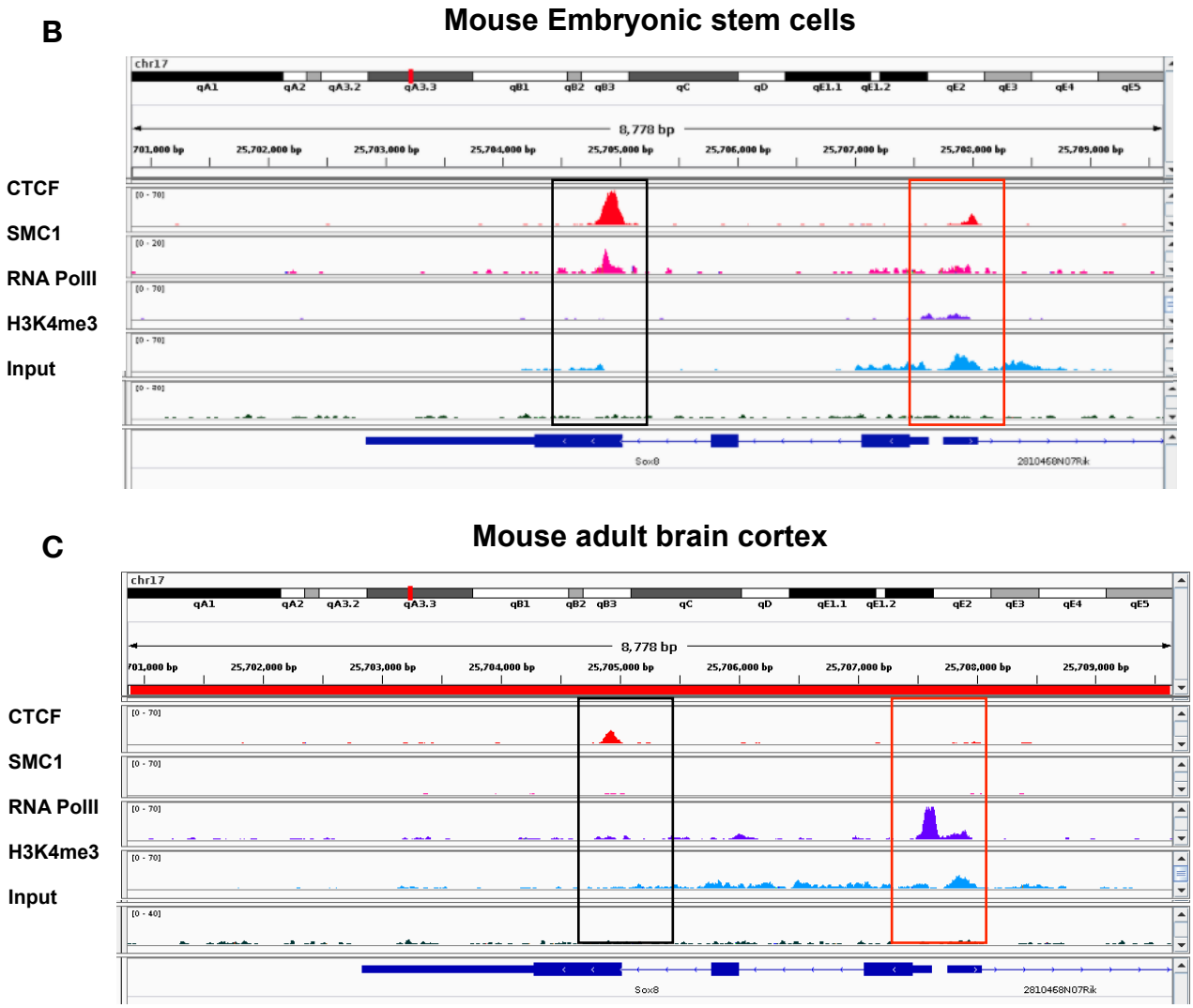
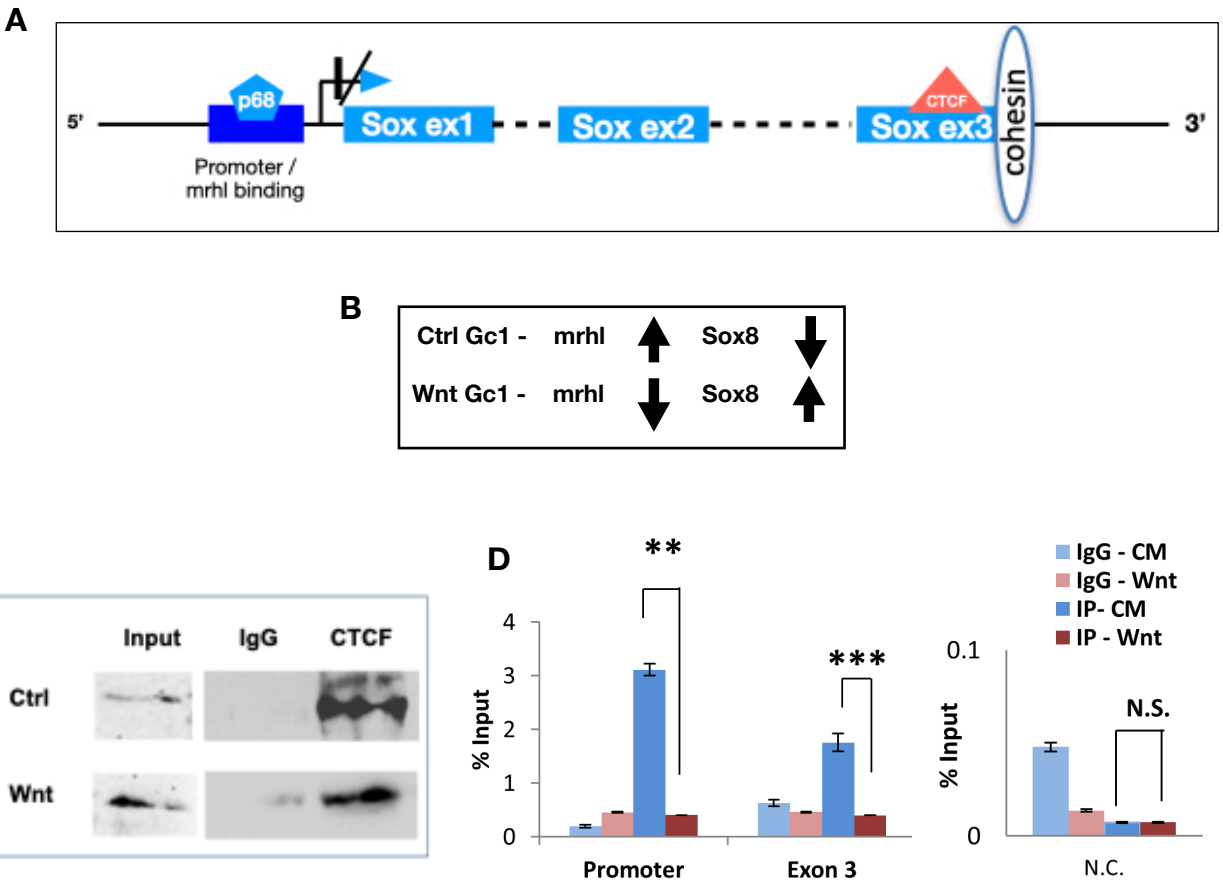


Fig 3.8 - CTCF and cohesin bind at Sox8 locus . (A) *Analysis of publicly available RNA-seq data shows that mrhl transcript level is low and Sox8 level is high in the mouse adult brain cortex and the inverse is true in the embryonic stem cell system - the two surrogate systems in which all further ChIP- seq data analysis has been carried out.* ***(B)*** *CTCF and cohesin bind within exon 3 of Sox8 when mrhl is expressed in mESC (highlighted by black box). Sox8 transcription is at basal levels as evidenced by low levels of PolII occupancy and H3K4me3 histone modification which are indicative of transcriptional activity. Additional peak for CTCF and cohesin occupancy is observed at the promoter of Sox8 very close to the mrhl binding site (Highlighted in red) (C) There is reduced CTCF occupancy and no observable cohesin binding at the Sox8 locus when Sox8 is transcriptionally active in adult brain cortex when compared to mESC. The secondary peaks for CTCF and cohesin at the promoter are no longer observed at the Sox8 promoter (highlighted in red)*

3.6 Experimental validation of architectural protein binding at Sox8 locus

To confirm that CTCF and cohesin are bound at the Sox8 locus in the mouse spermatogonial cells, ChIP-qPCR was performed for CTCF and Rad21 subunit of cohesin in control and Wnt induced Gc1-spg cells. Due to the unavailability of CTCF or cohesin

ChIP-seq data in spermatogonial cells at the onset of this work, no positive control could be maintained for these experiments at the qPCR level. However, western blotting (**Figure 3.9 C, E**) was performed to ensure that the ChIP was working by looking for enrichment of specific protein in the IP reaction over isotype control reaction. This experimental setup has been followed for all ChIP experiments. A negative control locus from within a gene-desert region in chromosome 3 was included for qPCR to ensure that there was no false positive signal. The results from the experiment correlated with the results from the ChIP-seq datasets. In Gc1-spg cells under control condition, where Sox8 is maintained in the repressed condition by *mrhl* lncRNA, CTCF binds to exon 3 and an additional occupancy signal at the Sox8 promoter close to *mrhl* binding site is also observed. Upon induction of the Wnt signalling pathway with Wnt3a cue, when *mrhl* lncRNA is downregulated and Sox8 transcription is activated (**Figure 3.9 B**), there is a significant reduction in CTCF and Rad21 occupancy at exon 3 and the promoter (**Figure 3.9 D,F**). Additionally, ChIP-qPCR for p68, whose binding at the Sox8 promoter has been shown previously, was performed and both the promoter and exon 3 regions of Sox8 showed enrichment for p68 binding in the control condition and a significant reduction in the Wnt induced condition (**Figure 3.9 H)**.



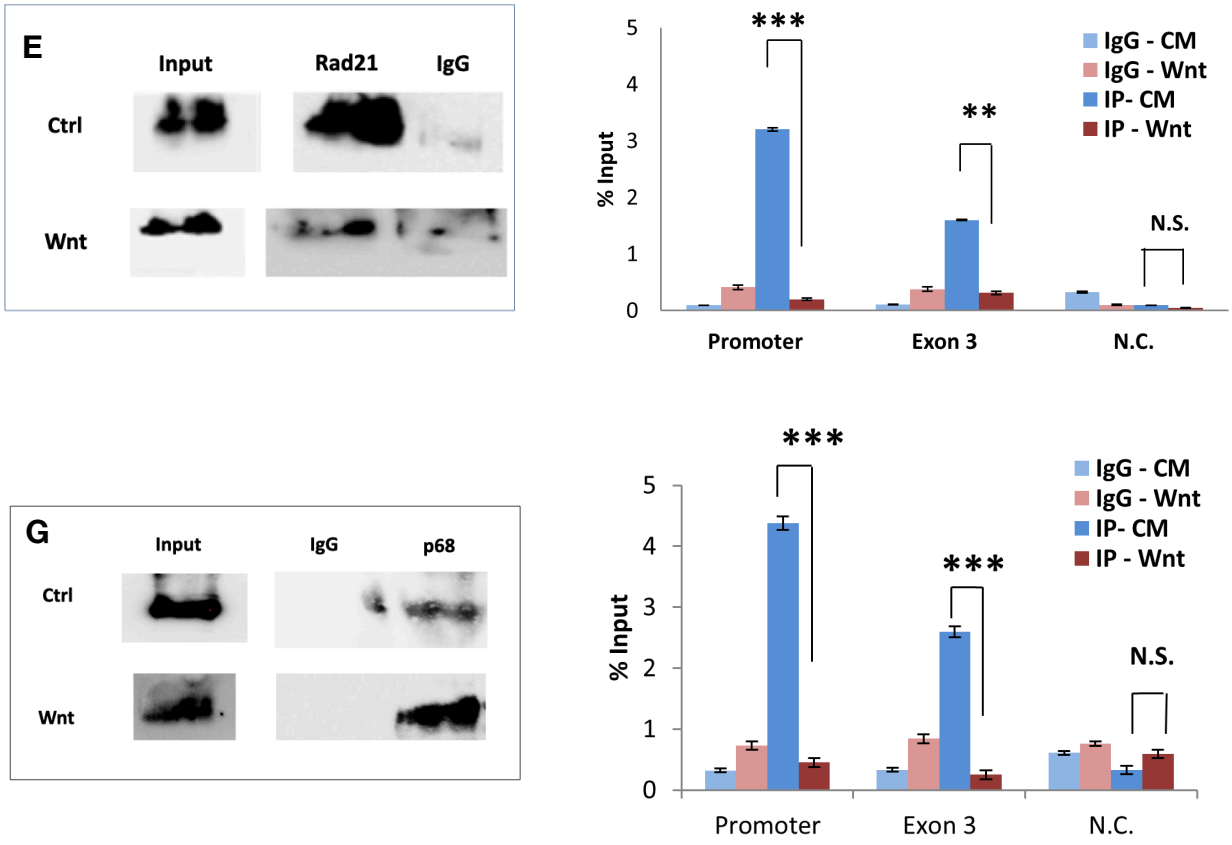
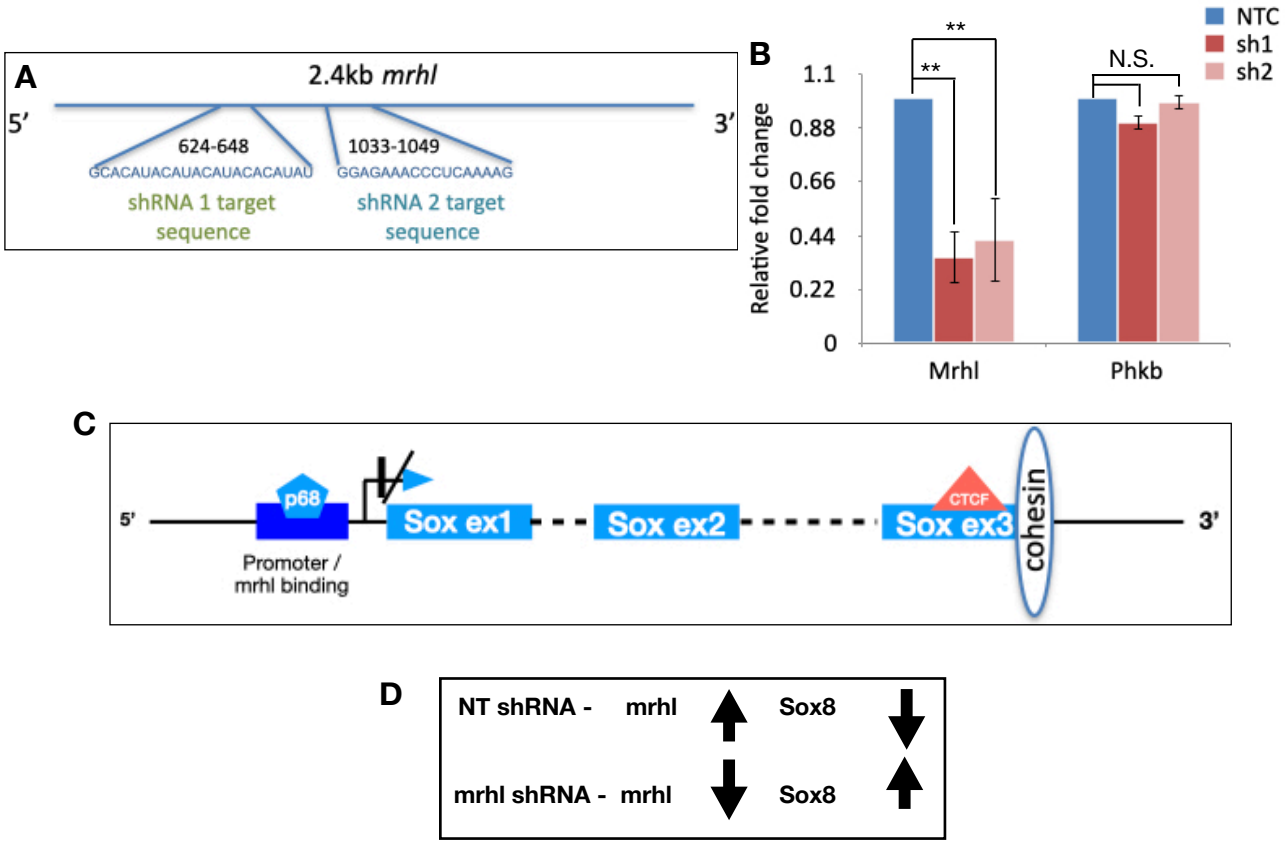


Fig 3.9 - CTCF, cohesin and p68 bind at the Sox8 locus in mouse spermatogonial cells. (A) *Schematic representation of Sox8 locus showing the ChIP-seq validated binding sites for CTCF and cohesin at exon 3 of Sox8 and ChIP validated binding site for p68 close to the mrhl binding site at the promoter of Sox8.* ***(B)*** *Schematic depicting the inverse correlation between mrhl and Sox8 expression in control and Wnt induced Gc1-spg cells.* ***(C)*** *Western blotting for CTCF ChIP performed in control and Wnt induced Gc1-spg cells.* ***(D)*** *Results for ChIP-qPCR for CTCF show binding of CTCF at both the promoter and exon3 of Sox8 gene and significantly reduced occupancy upon induction of the Wnt signalling pathway* ***(E)*** *Western blotting for Rad21 subunit ChIP of cohesin performed in control and Wnt indeed Gc1-spg cells.* ***(F)*** *Results for ChIP-qPCR for Rad21 subunit of cohesin show binding of Rad21 at both the promoter and exon3 of Sox8 gene and this occupancy reduces significantly upon induction of the Wnt signalling pathway.* ***(G)*** *Western blotting for p68 ChIP was performed in control and Wnt induced Gc1-spg cells* ***(H)*** *p68, whose occupancy at the Sox8 promoter has been previously validated by ChIP-qPCR, also shows occupancy at exon 3 of Sox8 gene when repressed. The occupancy at both loci reduces significantly upon Wnt induced downregulation of mrhl. Data in the graphs have been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (Student's t test)*

To understand if this differential occupancy of the three proteins was dependent on *mrhl* lncRNA or a downstream effect the Wnt signalling pathway activation independent of the lncRNA, inducible silencing cell lines for *mrhl* were generated in Gc1-spg cells. Either a non-target control shRNA or two different shRNAs targeting two regions of *mrhl* lncRNA (**Figure 3.10 A**) were integrated into Gc1-spg cells and stable integrants were selected for. We were able to silence *mrhl* transcript levels by around 70% and 65% with the two shRNAs respectively when compared to the non-target control shRNA. Since *mrhl* is

transcribed from within the 15th intron of the phkb gene, albeit from an independent transcription unit, we looked at the transcript levels of phkb as well to ascertain that RNAi was specific to *mrhl* only and not targeting phkb. The perturbation in phkb transcript level, the host gene of *mrhl*, was negligible (**Figure 3.10 B**). These cells were used for ChIP-qPCR experiments for CTCF, Rad21 and p68 proteins. The same occupancy trend as observed in Control and Wnt induced cells were observed in non-target control shRNA integrated cells and *mrhl* targeting shRNA integrated cells respectively. Both the promoter and exon3 showed association with CTCF, Rad21 (**Figure 3.10 E, F**) and p68 (**Figure 3.10 G**) in the non-target control cells and this association decreased significantly in *mrhl* knockdown cells. This indicated *mrhl* dependent binding of not only p68 but also of CTCF and Rad21 at the locus. The slight difference in the silencing efficiencies between mrhl targeting shRNA1 and shRNA2 was reflected as a slight difference in the reduction of occupancy of the three proteins further validating that the occupancy of these proteins at the locus was *mrhl* dependent. The reduction was slightly higher in cells in which *mrhl* was targeted with shRNA1 when compared to shRNA2 (**Figure 3.10 E,F and G** - difference between the last two bars labeled *mrhl* sh1 and *mrhl* sh2).



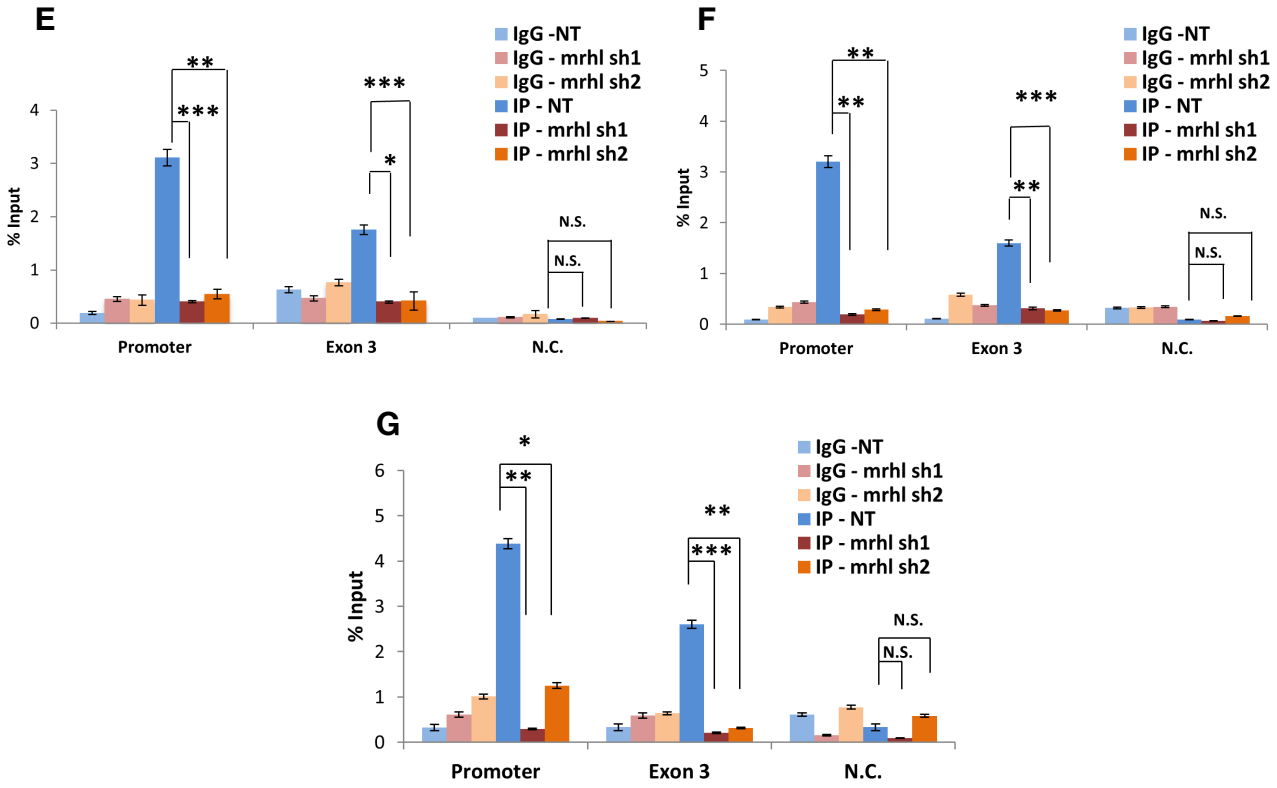


Fig. 3.10 -CTCF, cohesin and p68 bind at the Sox8 locus in the presence of mrhl lncRNA ***(A)****Schematic representation of the two shRNA targeting sequences within mrhl lncRNA, one targeting nucleotides between positions 624 and 648 and the other between 1033 and 1049.* ***(B)*** *Mrhl silencing efficiency of the two shRNA upon induction with IPTG for 4 days.* ***(C)*** *Schematic representation of Sox8 locus showing the ChIP-seq validated binding sites for CTCF and cohesin at exon 3 of Sox8 and ChIP validated binding site for p68 close to the mrhl binding site at the promoter of Sox8.* ***(D)*** *Inverse correlation between mrhl and Sox8 expression in control and mrhl shRNA integrated Gc1-spg cells.* ***(E)*** *Results for ChIP-qPCR for CTCF show significant reduction in occupancy of CTCF at both the promoter and exon 3 of Sox8 upon induction of silencing of mrhl with both shRNA construct 1 and shRNA construct 2.* ***(F)*** *Results for ChIP-qPCR for Rad21 subunit of cohesin show significant reduction in occupancy of Rad21 at both the promoter and exon 3 of Sox8 upon induction of silencing of mrhl with both shRNA construct 1 and shRNA construct 2. Figure legend next to (F) is common to both (E) and (F)* ***(G)*** *The occupancy of p68 at exon3 and promoter of Sox8 significantly reduces upon shRNA induced silencing of mrhl as seen in the results from ChIP-qPCR. Data in the graphs have been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (Student's t test)*

The occupancy of these three proteins at the Sox8 locus in mice testes of ages P7 and P21 was also investigated by ChIP-qPCR so as to understand the biological relevance of these proteins in regulating Sox8 expression. The occupancy trend of the proteins observed in the experiments performed using cells was also seen in experiments with mice testes. Occupancy of CTCF (**Figure 3.11 B,C**), Rad21 of cohesin (**Figure 3.11 D,E**) and p68 (**Figure 3.11 F,G**) was observed at both the promoter and exon 3 of Sox8 in P7 mice testis, in which *mrhl* is expressed and Sox8 is transcriptionally inactive but not in the

testes of 21 day old mice, in which *mrhl* transcript levels are low and Sox8 is transcriptionally active (**Figure 3.11 A**).

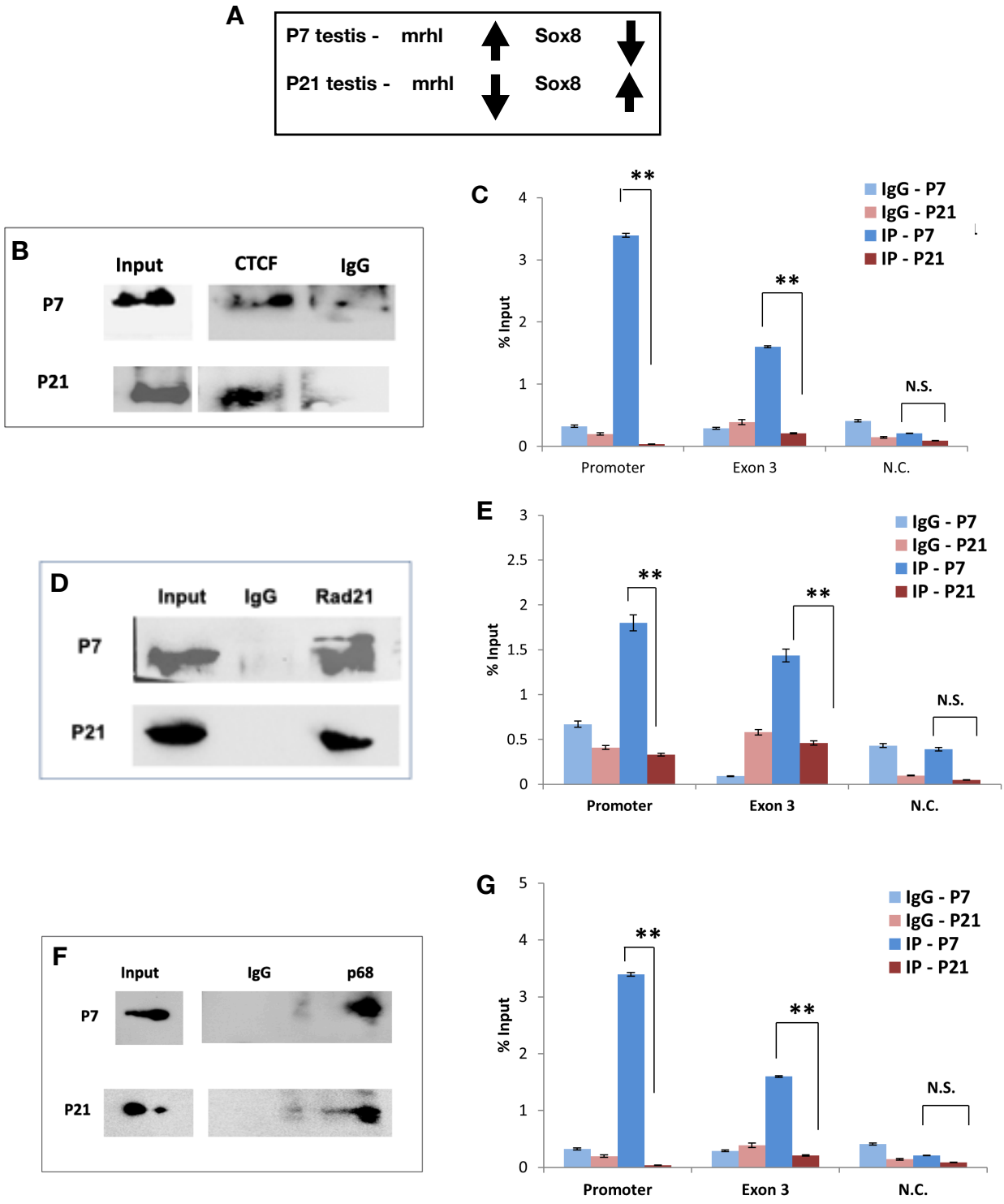


Fig 3.11: CTCF, cohesin and p68 bind to the sox8 locus in mouse testes (A) *Schematic representation of the expression status of mrhl and Sox8 in mice testes of P7 and P21 age groups. The testes of 7-day old mice have higher expression of mrhl lncRNA and lower expression of Sox8 as compared to the testes of 21-day old mice.* ***(B)*** *Western blot for CTCF ChIP performed with P7 and P21 mice testes* ***(C)*** *Occupancy of CTCF is observed at the promoter and exon 3 of Sox8 locus in P7 mice testes and a significant reduction in this occupancy is observed in P21 mice testes* ***(D)*** *Western blot for Rad21 ChIP performed with P7 and P21 mice testes* ***(E)*** *Occupancy of Rad21 subunit of cohesin is observed at the promoter and exon 3 of Sox8 locus in P7 mice testes*

and a significant reduction in this occupancy is observed in P21 mice testes ***(F)*** *ChIP Weston blotting for p68 in testes of 7-day old and 21-day old mice (G) Occupancy of p68, already known to bind at the promoter as seen in previous ChIP-qPCR experiments, is observed at exon 3 of Sox8 locus as well in P7 mice testes and a significant reduction in this occupancy is observed in P21 mice testes. Data in the graphs have been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (Student's t test)*

From the ChIP-seq analysis and ChIP-qPCR experiments, it was observed that ChIP signal for proteins bound at exon3 is seen at the promoter and vice-versa. We reasoned that the peak for these proteins was being observed at both these loci because the two regions were being brought in contact with each other through the formation of a chromatin loop. Based on various reports of protein- protein interactions, it is known that PRC2, CTCF, cohesin, p68, HDAC, Sin3a, and MAD-Max all interact with each other a form a complex (**Figure 3.12 A**). Combining the information from literature with our observations, it is likely that the promoter and exon 3 of Sox8 locus are being brought into contact with each other by the protein complex comprising of the afore mentioned proteins through a chromatin loop (**Figure 3.12 B**).

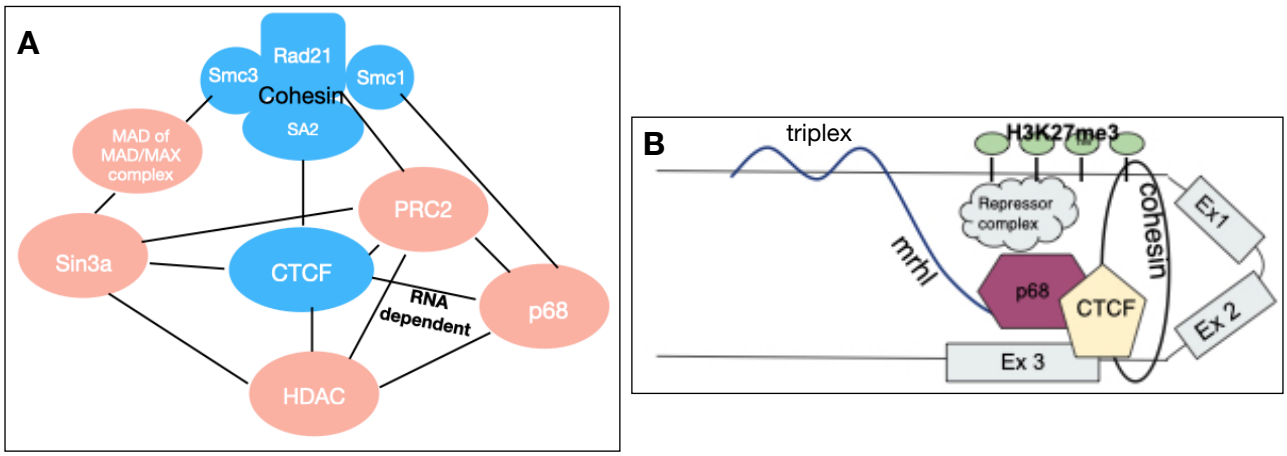


Fig. 3.12: Protein-protein interactions of proteins present at theSox8 locus (A) *Simplified depiction of protein-protein interactions of some of the proteins found at the Sox8 locus in the repressed state. The proteins in blue shapes are occupants of exon3 of Sox8 while the proteins in pink ovals are occupants of promoter of Sox8* ***(B)*** *The protein complex is likely to bring the promoter and exon 3 of Sox8 in contact with each other through chromatin looping.*

CTCF mediated loops require partner proteins to be bound at the two anchor points. The formation of loops are mediated by the dimerisation of CTCF with the partner protein, most often by homo-dimerisation, and the loop is stabilised by cohesin (Fudenberg, G., *et al*, 2017). The presence of CTCF at the promoter was observed through ChIP-qPCR. We wanted to further investigate if the occupancy signal observed for CTCF at the promoter was seen as on outcome of contact with exon 3 or if it could be because of CTCF binding

at the promoter as well. To this end, the presence of a CTCF binding site at the promoter of the gene was looked for using the prediction tool present in CTCFBSDB 2.0. The DNA sequences of the promoter and exon 3 of Sox8 were used to look for the presence of the binding site. The tool utilises multiple position weight matrices (PWM) generated based on CTCF motifs reported by different research groups to scan the sequences. Any hit with a score of above 3 is considered as a potential binding site. However, each PWM predicts the presence of a single binding site in the sequence of the highest score and therefore, it is likely that the input sequences contain more than one CTCF binding site. From the list of predictions, it was observed that the presence of CTCF binding site was predicted not only within exon 3 but also in the promoter DNA sequence (**Figure 3.13 A**). Based on the results of this prediction and results of the CTCF ChIP-qPCR experiment performed at the promoter and exon 3 of Sox8, it is very likely that CTCF molecules bind at the promoter and exon 3 of Sox8 and homo-dimerise, bringing the two sequences in close proximity. This contact is stabilised by the cohesin complex and *mrhl* lncRNA and p68 are essential members of this complex (**Figure 3.13 B**).

A

Motif PWM	Motif Sequence	Input Sequence Name	Motif Length	Motif Orientation	Score
EMBL_M1	CGCCGCCTAGTGGA	exon3	14	-	12.427
EMBL_M1	GGTCACCTGGTGGC	promoter	14	-	10.1556
EMBL_M2	GGAACAGCA	exon3	9	+	11.4118
EMBL_M2	GTCACTGCC	promoter	9	-	6.0656
MIT_LM2	TGTCCACTAGGCGGCGCCC	exon3	19	+	7.22188
MIT_LM2	GAGCCACCAGGTGACCCTG	promoter	19	+	5.51758
MIT_LM7	TGTCCACTAGGCGGCGCCCT	exon3	20	+	11.2928
MIT_LM7	GAGCCACCAGGTGACCCTGG	promoter	20	+	9.84716

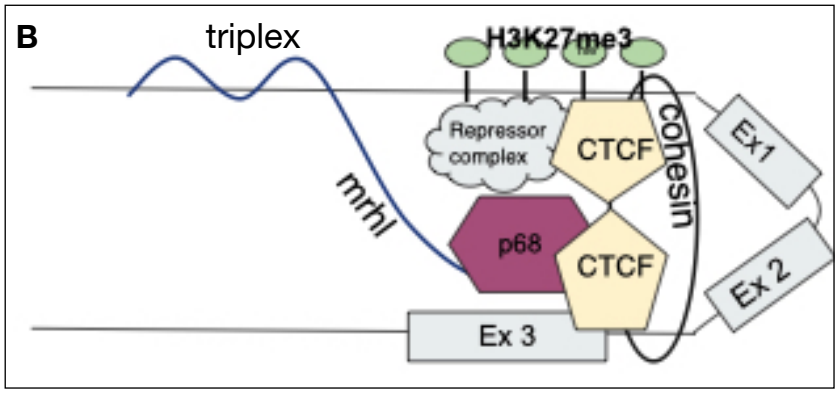


Fig 3.13: CTCF binding site at the Sox8 locus (A) *Predicted CTCF binding site from CTCFBSDB 2.0 database based in the DNA sequences of Sox8 promoter and exon 3. All results with a score higher than 3 have been listed in the table.* ***(B)*** *Model depicting the probable repressive chromatin*

loop established at the Sox8 locus - CTCF molecules bind to both the promoter and exon 3 and homodimerise, bringing the two regions in close proximity. The contact is stabilised by the cohesin complex. p68 and mrhl lncRNA are essential members of this complex. In addition, mrhl lncRNA interacts with the locus though triplex formation upstream of the Sox8 TSS.

3.7 Enhancers elements in the vicinity of Sox8

Negative regulatory elements that contact the promoter and lead to the repression of the gene are termed as ‘silencers’ or insulators and very often, silencers block the promoter from contacting enhancers. The presence of enhancer elements in the vicinity of Sox8 can be observed in ENSEMBL both upstream and downstream of the gene. Moreover, activity in the enhancer elements were seen to correlate with the transcriptional activity of the Sox8 gene (**Figure 3.14**). In the brain and heart tissues, CTCF was not bound at the cognate binding site and Sox8 was in the poised state. The enhancers present both upstream and downstream of the gene were active whereas in the kidney and thymus, CTCF was bound (indicative of Sox8 being transcriptionally repressed) and the enhancers were not active. This is indicative that the elements have potential to be enhancers for the Sox8 gene.

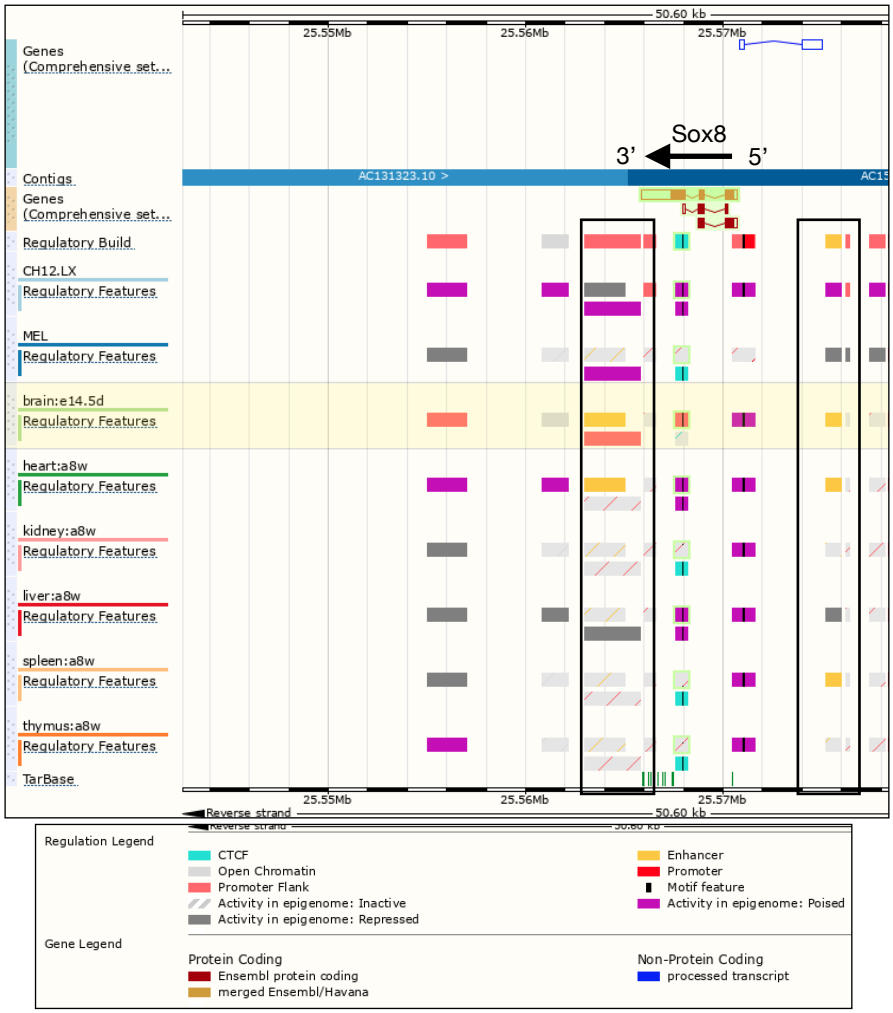


Fig. 3.14: Regulatory elements around the Sox8 locus in ENSEMBL database. *The enhancer elements on either side of the Sox8 gene are highlighted in the black boxes. The enhancer element present upstream and downstream of Sox8 show activity corresponding to transcriptional activity of the Sox8 gene.*

Sox9, a paralog of Sox8, is regulated by tissue specific enhancers with at least 5 different regions behaving as enhancers e in different tissues. In the testis, Sox9 is regulated by a 3.2 kb testis specific enhancer of *Sox9* (TES) which includes a core 1.4 kb element, TESCO. TES/TESCO deletion experiments indicate the presence of additional enhancers which act redundantly, especially in the early stages of development (Mead, T.J. et al, 2013, Gonen, N et al, 2017). The presence of multiple tissue specific enhancers has also been reported for Sox10, the other member of the SoxE group of proteins (Werner, T., et al, 2007). A study identified 7 different evolutionarily conserved elements from chicken to mice and humans in a 220kb genomic interval in the vicinity of the Sox8 gene with potential to act as enhancers for Sox8 (**Figure 3.15 A**). Of these 7, 4 of the elements present upstream of the Sox8 gene (E1 to E4) acted as enhancers in different embryonic tissues. None of the elements behaved as enhancer for cells of the embryonic gonad (Guth, S.I et al, 2010). However, the possibility that these elements could behave as enhancers later in development or in the postnatal stages exists. The same study also observed that the promoter of Sox8 alone could not drive robust transcription in most tissues in which Sox8 is expressed. Another study aimed at identifying regulatory elements in cells of the gonad identified a region downstream of the Sox8 gene as an NDR (nucleosome-depleted region) that arises *de novo*, specifically at E13.5 stage in male mice (**Figure 3.15 B**). Since the NDR is not present at E10.5, it is likely that any enhancer activity of this region, if present, was missed in the previous study by Guth et. al, which studied enhancer activity of the conserved elements between the ages of E10.5 and E12.5. Additionally, the authors also found a second NDR further downstream which was found in both male and female gonads at the age of E10.5 but only in the cells of the male gonads at age E13.5. Since the NDRs in this study were collectively enriched for motifs of Transcription Factor (TF) binding, specifically TFs that promote sex-specific cell differentiation, the authors opined that the NDRs are likely to be regulatory elements that promote divergence of sex- specific gonadal cells from their common progenitors (Garcia-Moreno, S. A et al, 2019).

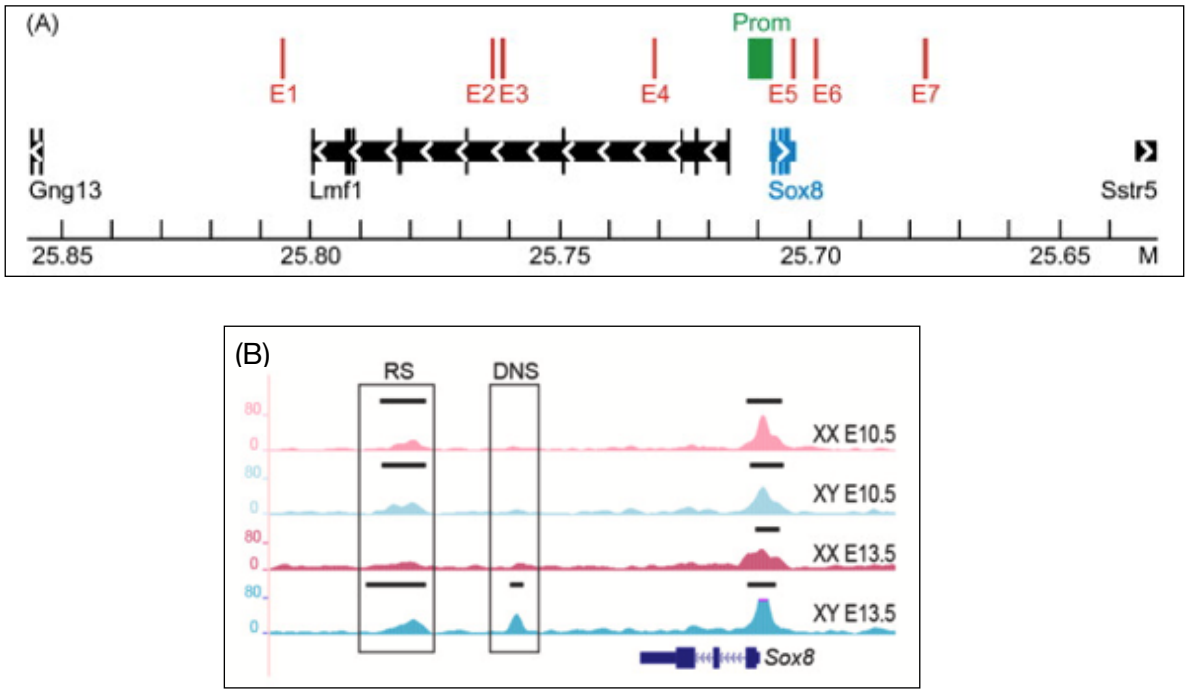
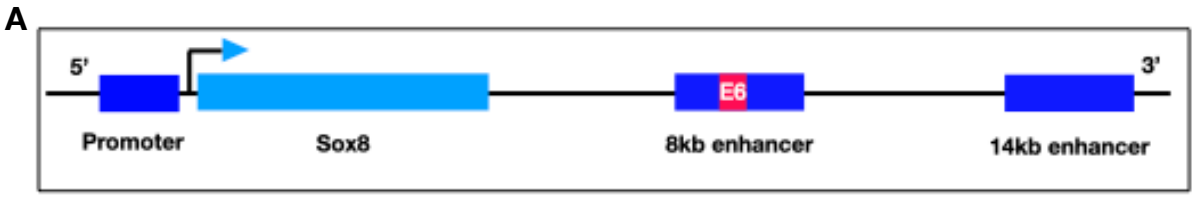


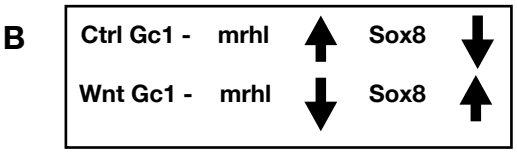
Fig 3.15: Enhancers of Sox8 (A) *The 7 evolutionarily conserved elements (labelled E1 to E7 in red) around the Sox8 gene (promoter indicated in green). Image has been adapted with permission from Guth, S.I et al, 2010* ***(B)*** *De novo NDR in cells of the male gonad at E13.5 which has the propensity to be the enhancer for Sox8. Image has been adapted with permission from Garcia-Moreno, S. A et al, 2019.*

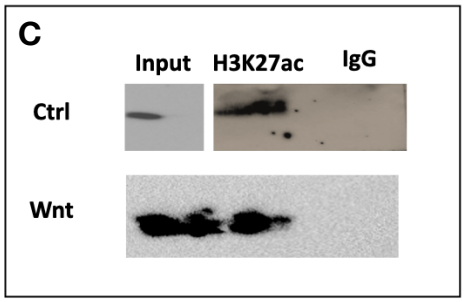
3.8 Experimental validation of Enhancer elements

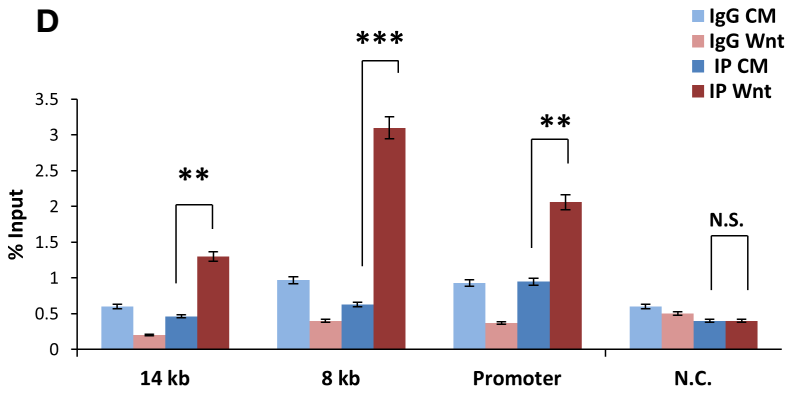
Based on the information available in literature on the enhancers for Sox8 and the elements highlighted in ENSEMBL database as potential enhancers, two downstream elements - a 3.2kb long element around 8kb downstream of Sox8 TSS and another 2.1kb long element 14kb downstream were identified as a potential enhancer elements to be studied further in the mouse spermatogonial cells. The DNA segment 8kb downstream harbours within it, the evolutionarily conserved element E6 from the previously mentioned study by Guth, S.I et al (Guth, S.I et al, 2010). To further explore the potential of these sequences to behave as an enhancers, chromatin immunoprecipitation for the histone modifications H3K4me1 indicative of a poised enhancer and H3K27ac indicative of active enhancer were performed under conditions in which Sox8 was transcriptionally repressed or active. In Gc1-spg cells under control conditions, there was no enrichment for either one of the histone modifications at either one of the putative enhancer elements. Upon activation of Wnt signalling with the Wnt3a cue in Gc1-spg cells, a significant increase in both these modifications was observed at both the putative enhancer elements (**Figure 3.16 D,F**). Additionally, an increase in these modifications was also observed at the

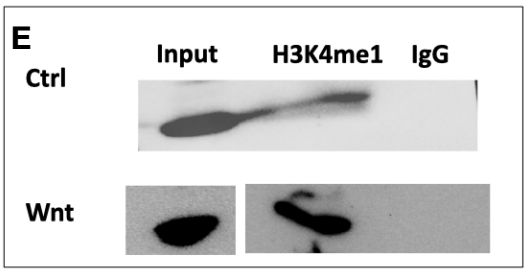
promoter of the gene (**Figure 3.16 D,F**). H3K27ac modification is observed at both active promoters and enhancers (Paauw, N. D. et. al, 2018). Similarly, H3K4me1 modification also has been reported at promoter proximal regions of genes (Bae, S., & Lesch, B. J., 2020). Following a similar trend, higher levels of both H3K27ac and H3K4me1 histone modifications were also observed in *mrhl* knockdown cells when compared to non-target control cells (**Figure 3.16 H, I**) at both the putative enhancer elements along with at the gene promoter. The increase in the histone modification in *mrhl* knockdown cells was lesser than that observed upon Wnt signalling activation at the promoter and enhancer elements 8kb downstream. The increase at the enhancer element 14kb downstream was comparable in *mrhl* knockdown cells and Wnt activated Gc1-spg cells.

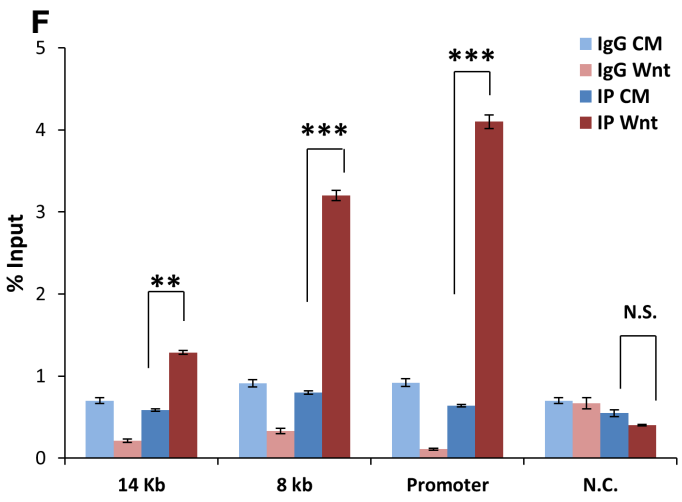












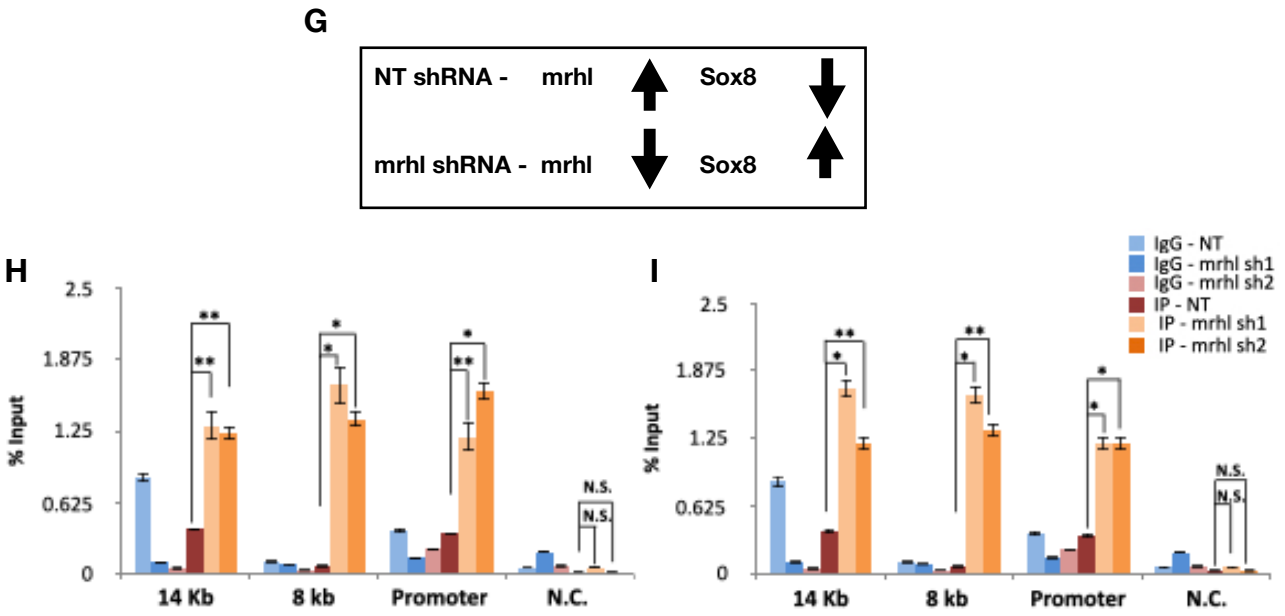


Fig 3.16: Enhancer histone modifications at the Sox8 locus ***(A)*** *Schematic representing the two putative enhancer elements relative to Sox8* ***(B)*** *Schematic indicating the inverse correlation between mrhl and Sox8 expression in control and Wnt treated Gc1-spg cells.* ***(C)*** *ChIP western blotting for H3K27ac in control and Wnt treated Gc1-spg cells shows enrichment of protein in IP reaction over isotype control* ***(D)*** *Results from ChIP-qPCR experiment for H3K27ac shows a significant increase in the levels of this modification with Wnt signalling induced Sox8 transcriptional activation at the promoter and both the enhancer elements.* ***(E)*** *ChIP western blotting for H3K4me1 in control and Wnt treated Gc1-spg cells shows enrichment of protein in IP reaction over isotype control* ***(F)*** *Results from ChIP-qPCR experiment for H3K4me1 shows a significant increase in the levels of this modification with Wnt signalling induced Sox8 transcriptional activation at the promoter and both the enhancer elements.* ***(G)*** *Schematic representing the inverse correlation between mrhl and Sox8 expression in control and mrhl shRNA integrated Gc1-spg cells.* ***(H)*** *Results from ChIP-qPCR experiment for H3K27ac shows a significant increase in the levels of this modification with Sox8 transcriptional activation at the promoter and both the enhancer elements in mrhl knockdown cells when compared to cells containing non-target control shRNA* ***(I)*** *Results from ChIP-qPCR experiment for H3K27ac shows a significant increase in the levels of this modification with Sox8 transcriptional activation at the promoter and both the enhancer elements in mrhl knockdown cells when compared to cells containing non-target control shRNA Figure legend next to (I) is common to both (H) and (I). Data in the graphs have been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (Student's t test)*

The increase in levels of enhancer histone modifications was also validated in mice testes of P7 and P21 age groups. A significant increase of both H3K4me1 and H3K27ac levels was observed in the testes of 21 day old mice, in which Sox8 is actively transcribed as compared to 7 day old mice (**Figure 3.17 D,F**) at both of the putative elements and at the gene promoter in agreement with the trend observed in the experiments performed with cells.

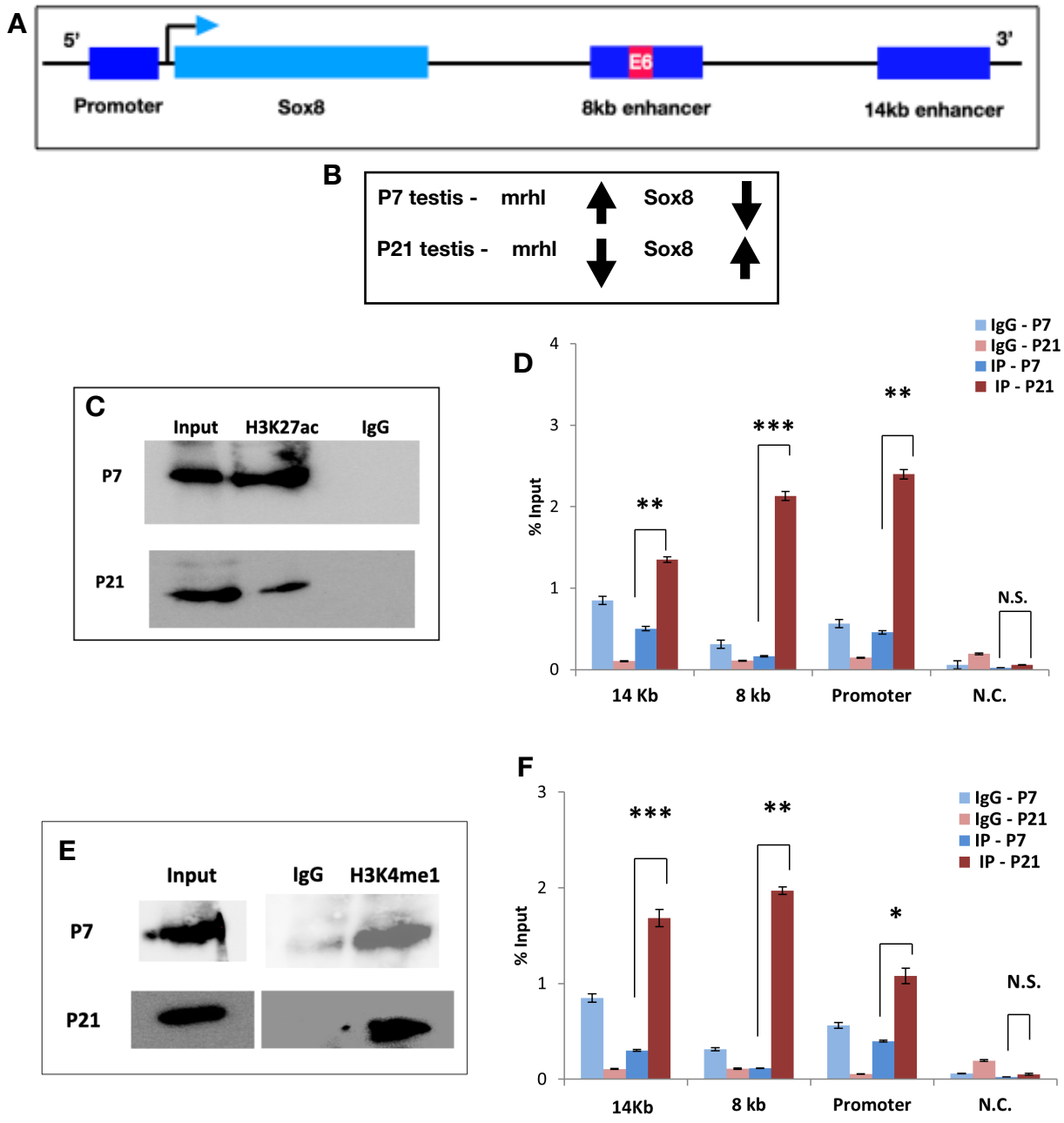
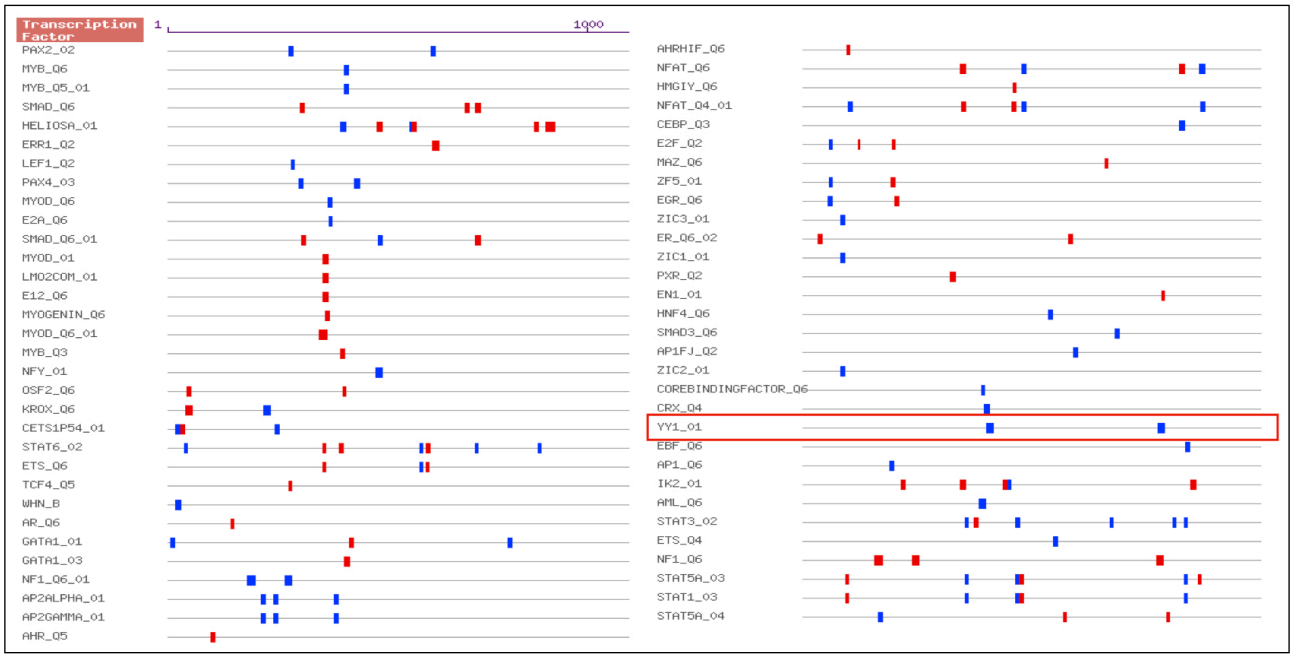


Fig 3.17: Enhancer modifications at the Sox8 locus in mouse testes (A) *Schematic representing the two putative enhancer elements relative to Sox8* ***(B)*** *Schematic depicting the inverse correlation between mrhl and Sox8 expression in P7 and P21 mice testes.* ***(C)*** *ChIP western blotting for H3K27ac in P7 and P21 mice testes shows enrichment of protein in IP reaction over isotype control* ***(D)*** *Results from ChIP-qPCR experiment for H3K27ac shows a significant increase in the levels of this modification in P21 mice testes when compared to P7 mice testes at the promoter and both the enhancer elements.* ***(E)*** *ChIP western blotting for H3K4me1 in P7 and P21 mice testes shows enrichment of protein in IP reaction over isotype control* ***(F)*** *Results from ChIP-qPCR experiment for H3K4me1 shows a significant increase in the levels of this modification in P21 mice testes when compared to P7 mice testes at the promoter and both the enhancer elements. Data in the graphs have been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (Student's t test)*

3.9 Presence of the transcription factor YY1 at Sox8 locus

The DNA sequence of the Sox8 promoter was scanned using the online tools Gene Promoter Miner to look for the presence of binding sites of other CTCF interacting proteins that could be of relevance to the regulation of Sox8. The results from the prediction tool indicated to binding sites for the CTCF interacting transcription factor YY1 among other TFs (**Figure 3.18 A**). YY1 is of particular interest since it has been implicated in mediating activating chromatin loops by bringing together promoters and enhancers (Weintraub, A.S et al, 2017). A switch from loops mediated by CTCF to those mediated by YY1 has been observed during neural lineage commitment and differentiation from progenitor cells (Beagan, J.A et al, 2017). We wanted to investigate if YY1 could be playing a role in the regulation of Sox8 transcription and to this end, we performed Chromatin immunoprecipitation for YY1 under the two conditions where Sox8 is transcriptionally repressed and activated. In Gc1-spg cell under control conditions, no enrichment for YY1 was observed at either the promoter or the enhancer elements. When Wnt cue is provided, an increase in the occupancy of YY1 is observed at both the enhancer elements along with an increase in occupancy at the gene promoter also (**Figure 3.18 F**). Similarly, an increased occupancy of YY1 is observed at the promoter and both the enhancers upon *mrhl* knockdown mediated upregulation of Sox8 when compared to non-target control (**Figure 3.18 G**), albeit to lower levels than in Wnt induced cells. The occupancy of YY1 at the Sox8 locus was also investigated in mice testes and the results were in agreement with the results from the experiments done with cells (**Figure 3.18 H**).

A



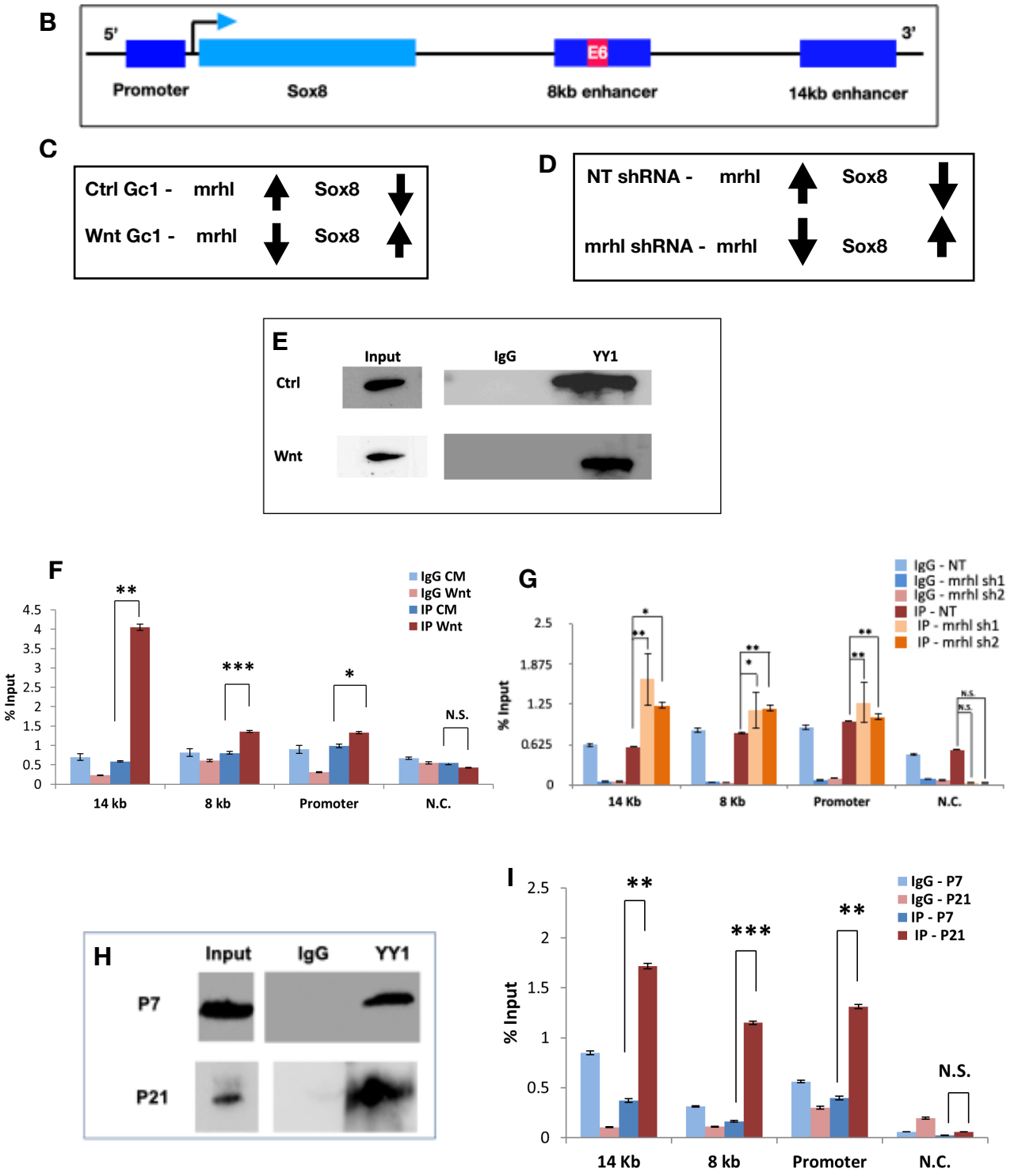


Fig. 3.18: YY1 is involved in the activation of Sox8. ***(A)*** *Results from the Gene promoter miner tool to scan the promoter of Sox8 for presence of binding sites for transcription factors. YY1 has two binding sites in the promoter and been highlighted with the red box.* ***(B)*** *Schematic representing the two putative enhancer elements relative to Sox8* ***(C)*** *Inverse correlation between mrhl and Sox8 expression in control and Wnt treated Gc1-spg cells.* ***(D)*** *Schematic depicting the inverse correlation between mrhl and Sox8 expression in control and mrhl shRNA integrated Gc1-spg cells.* ***(E)*** *ChIP western blotting for YY1 performed in Gc1-spg cells under control and Wnt induced conditions shows the enrichment of YY1 in IP reactions over the isotype control reactions.* ***(F)*** *ChIP for YY1 performed in Gc1-spg cells shows the increase in occupancy of YY1 upon Wnt induction when compared to control conditions at both the enhancer elements and at the promoter.* ***(G)*** *ChIP for YY1 performed in Gc1-spg cells shows the increase in occupancy of YY1 upon mrhl*

*lncRNA knockdown with shRNA when compared to non-target control shRNA at both the enhancer elements and at the promoter.**(H)** ChIP western blotting for YY1 performed in P7 and P21 mice testes shows the enrichment of YY1 in IP reactions over the isotype control reactions.**(I)** ChIP-qPCR for YY1 in mice testes. Data in the graphs have been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (two-tailed Student's t test)*

Based on the results of the ChIP experiments, we hypothesised that the repressive CTCF mediated loop gives way to a YY1 mediated activating chromatin loop upon *mrhl* downregulation at the Sox8 locus (**Figure 3.19**) which brings the gene promoter in contact with an enhancer.

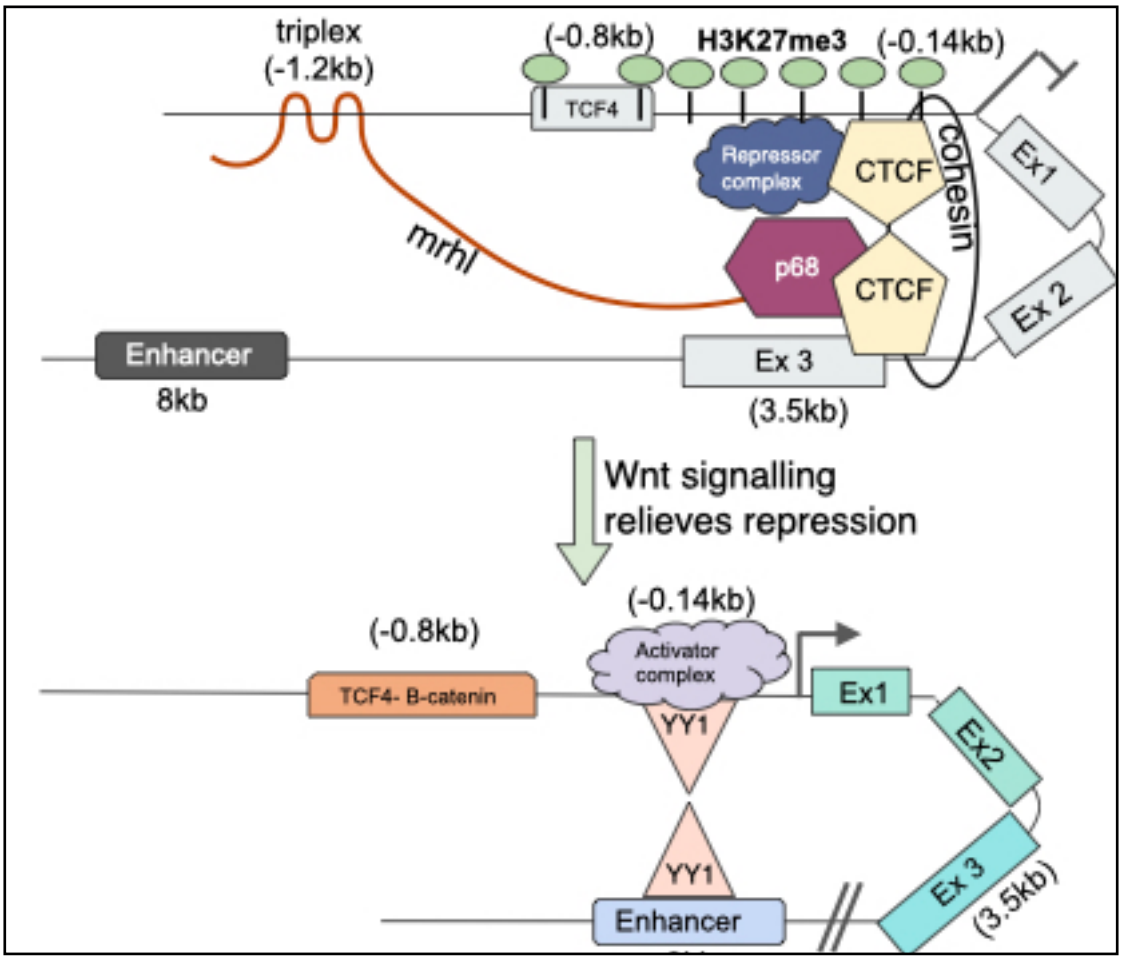


***Fig. 3.19: Figure depicting the proposed chromatin organisation mediated gene regulation at the Sox8 locus.** In the B-type spermatogonial cells, mrhl bound at the Sox8 locus through triplex-mediated and protein-mediated chromatin interactions maintains it in the repressed condition by associating with CTCF and cohesin and bringing the promoter in contact with exon3 though chromatin looping. Upon activation of the Wnt signalling pathway, mrhl levels reduce and the transcription of Sox8 is activated. The repressive promoter- exon 3 interaction gives way to an activating enhancer - promoter contact which is mediated by the architectural protein YY1.*

3.10 Investigating the formation at differential chromatin loops at the Sox8 locus

Chromatin Conformation Capture (3C) experiment was carried out to explore the loop formation proposed above. Key steps of the experimental work flow and design have been shown in **figure 3.20 A** below. In brief, cross linked nuclei of cells are permeabilised and subjected to restriction digestion with either a 6-bp cutter or 4-bp cutter enzyme decided based on the size of the DNA segment within which interaction is being probed. The experimental resolution with 4-bp cutters are of a few hundred basepairs. Since the proposed loops at the Sox8 locus are within a span of 5kb in the repressed state and 10kb in the activated state, the 4-bp cutter enzyme DpnII was used. The restriction digested chromatin within the nuclei are then ligated and purified and used as template for PCR. Fragments of DNA that are closer in proximity - either neighbouring fragments in the genome or those brought close together through chromatin loops, will have a higher frequency of ligation than distal DNA fragments. Primers for PCR are designed in the same direction within multiple fragments between the two points which are suspected to interact close to the restriction sites (Naumova, N et al, 2012). This ensures that amplification in PCR from only fragments that have successfully undergone restriction digestion and ligation in the correct orientation and not from genomic DNA (**Figure 3.20 B**). When looping interactions are being investigated and compared across different conditions/ cell types, a control locus, such as the locus of a housekeeping gene, whose looping status remains constant in included. Ercc3 locus, which codes for the Excision Repair 3 protein, is commonly considered as the control locus for 3C experiments and was included in our experiments as well (Dekker, J. 2006). Another control that was included for the 3C experiment were BAC plasmids containing the Sox8 and Ercc3 loci that were subjected to the same experimental procedure as the cells. The template generated this way acts as a control for primer efficiency in the experiment (Dekker, J. 2006). An anchor/ viewpoint is chosen from within one of the two regions of DNA suspected to participate in chromosomal contact. Results from the 3C experiment are plotted as relative frequency of ligation or interaction as a distance of genomic distance. For linear DNA, the relative interaction frequency reduces as a function of distance. The presence of interaction between the anchor fragment and any region of DNA is marked by the presence of a local peak in the interaction frequencies (**Fig. 3.20 C**)

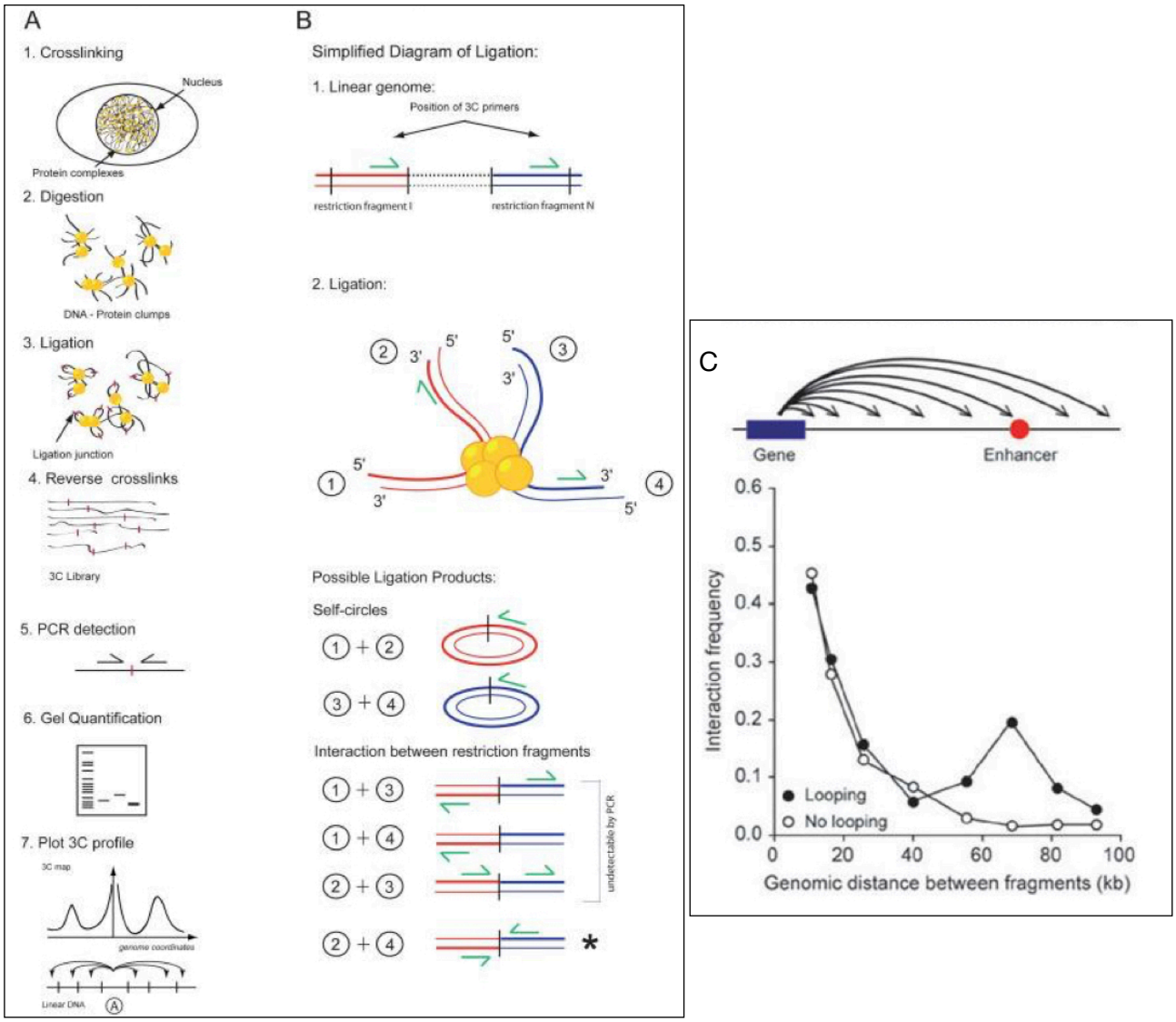


Fig 3.20: Chromosome conformation capture assay (A) *An illustration of the 3C method. Genomic DNA is crosslinked (1), capturing 3D interactions inside the cell. After removal of cell membrane, the chromatin is digested inside permeabilised nuclei (2). The DNA ends are then ligated together (3) creating DNA junctions representing the proximity of restriction fragments in the fixed sample. After ligation, the crosslinks are reversed (4) and 3C template is purified. Finally, the ligation products are detected (5) using PCR-based methods. After quantification (6) the results are plotted as a 3C profile (7), revealing interactions between anchor (labeled "A") and all other fragments in the genomic regions.* ***(B)*** *Possible outcome of ligation reaction between two restriction fragments. There are many restriction fragments contained within one complex. To further understand the ligation step, a simplified image shows a view where one complex contains only two restriction fragments -red and blue. The 5' and 3' ends are indicated for each strand. Each digested end has been numbered 1–4. Also indicated are the locations where 3C primers have been designed. All the primers are unidirectional, located near the restriction site. There are six possible ligation products that result from this molecule. Only one of them results in a detectable product - that is when end 2 and end 4 are ligated to each other. Image has been adapted with permission from Naumova, N et al, 2012* ***(C)*** *A theoretical example of a typical 3C analysis to detect interactions between a gene (blue rectangle) and a regulatory element (red circle). Open circles indicate predicted interactions between a gene and sites located up to 100 kb away in the absence of specific looping contacts. These interactions reflect random collisions and are characterised by a decreasing frequency of interaction as a function of distance. Filled circles illustrate the expected pattern of interactions in the presence of a specific interaction between the*

gene and a distant enhancer. The presence of the specific interaction is apparent by the local peak in interaction frequencies. Image has been adapted with permission from Dekker, J. 2006.

3C was performed with Gc1-spg cells in which *mrhl* had been targeted through RNAi using a pool of both *mrhl* targeting shRNA. The silencing efficiency was found to be close to 60% (**Figure 3.21 A**). Cells with non-targeting control shRNA were used as control. *Mrhl* knockdown cells were used as we were interested in identifying the differential chromatin loops mediated by the lncRNA only and not those which were occurring as a downstream effect of the activation of signalling pathways. Activation of the Wnt signalling cascade initiates a different Wnt transcriptional program in target cells (Logan, C. Y., & Nusse, R.,2004) and changes in chromatin looping in the region surrounding the Sox8 locus can be expected as a result.

Nuclear permeabilisation, restriction digestion and ligation conditions were standardised so as to obtain the arose gel pattern shown in **Figure 3.21 B**. Undigested chromatin was seen as a crisp high molecular weight band, DpnII digested chromatin was seen as a smear enriched in size of below 3kb along with a discrete band above 10kb as expected due to the absence of recognition sites in the mouse pericentromeric repeats (Krijger, P. et al, 2020). A clear upward shift of the smear was observed after ligation. Such contact libraries from non-target control and *mrhl* knockdown cells were used as template for PCR. PCR was set up using the primer designed within the promoter as the constant primer as we were interested in identifying the contacts of the promoter with the various regulatory elements under the two different transcriptional states.

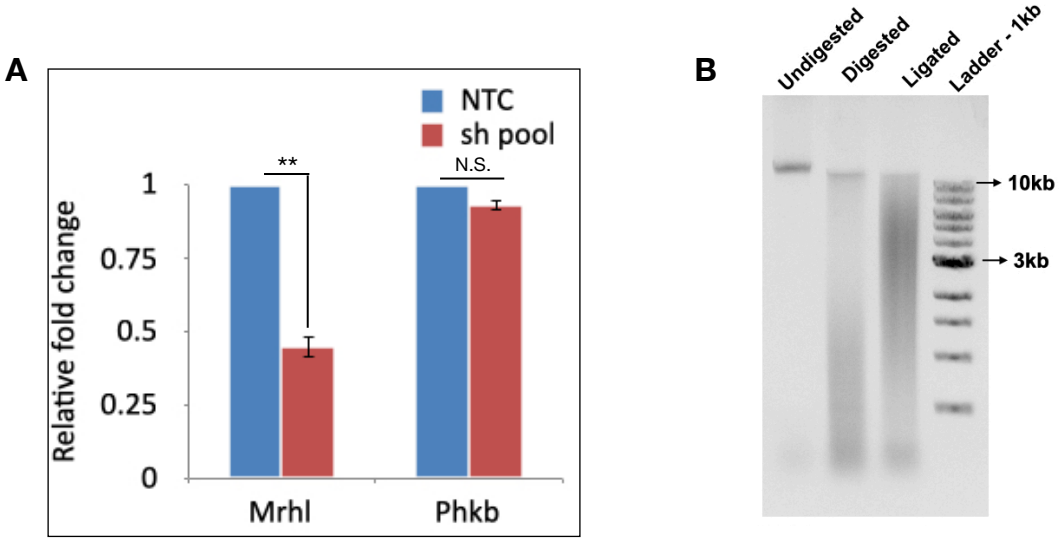


Fig 3.21 - 3C in Gc1-spg cells (A) *Mrhl was targeted for silencing using a pool of both shRNA and silencing efficiency of 60% was obtained in these cells.* ***(B)*** *Agarose gel image for chromatin*

*isolated at different steps of the 3C protocol shows a high molecular weight band before restriction digestion, a smear with majority of chromatin below 3kb post digestion with DpnII along with the presence of a discrete high molecular weight band and an upward shift of the smear post ligation. Data in the graph has been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (two-tailed Student's t test)*

The results from the 3C experiments were in support of our hypothesised differential chromatin contacts under the two different transcriptional states. Under control conditions, a peak in ligation frequency corresponding to promoter- exon 3 silencer contact was observed. In the *mrhl* knockdown cells, promoter-exon 3 contact was no longer observed but an increase contact frequency between promoter and enhancer element present 8kb downstream was observed (**Figure 3.22**). These two peaks were indicative of differential contacts in the cells between the two conditions. An additional smaller peak in the relative ligation frequency was also observed in cells under both conditions for promoter and enhancer element present 12kb downstream of TSS. This could potentially be a contact that is constant under the two different transcriptional states of Sox8. Alternatively, while it is expected that the ligation frequency reduces as a function of distance, small perturbations in this trend are observed in experiments (Dekker, J., 2002, Bhattacharya, A. et al, 2012). The additional peak could be one such perturbation. Since this was not a differential contact between the two transcriptional states, we didn't focus further efforts towards understanding the peak in more detail.

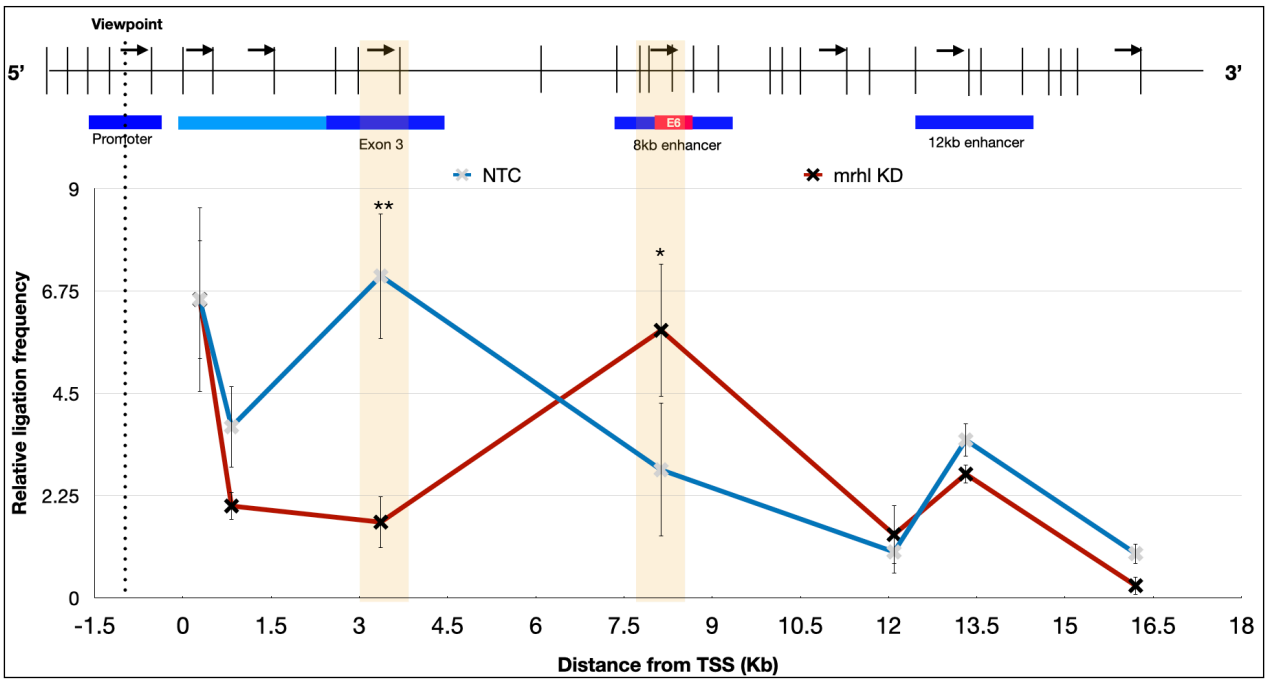


Fig 3.22 - Chromosome Conformation Capture for Sox8 locus *- 3C was performed in Gc1-spg cells with non- target control shRNA (blue line in the graph) or mrhl targeting shRNA (Red line in the graph). The schematic on top shows the positions of the DpnII sites (vertical bars) in the Sox8 locus relative to the genomic organisation. The black arrows indicate the 3C primer binding sites*

*and their directionality. The viewpoint primer is within the promoter of Sox8 (indicated with dotted line). The relative ligation frequency is plotted on the y-axis and the distance from TSS in kb is plotted on the X-axis. In the control cells (blue line), a peak in the relative ligation frequency is observed between promoter and exon 3 (highlighted with shaded box). In mrhl knockdown cells, the peak at exon 3 is no longer observed but a peak at enhancer 8kb downstream of TSS (highlighted with shaded box) is observed indicative of interaction of the promoter with this segments of DNA. Data in the graph has been plotted as mean ± S.D. , N=3. *** P ≤ 0.0005, ** P≤ 0.005, * P≤ 0.05, N.S - Not significant (two-tailed Student's t test)*

Chapter 4

Ongoing work and Future Perspectives

As seen in the previous section , At the Sox8 locus, different chromatin loops correspond to the different transcriptional states and drive transcription. Different gene regulatory elements are brought together in 3D space as a result of this looping switch. More importantly, key architectural proteins CTCF, cohesin and YY1 play key roles in enabling contact between these chromatin segments. The master regulator of this switch, however, is the lncRNA *mrhl*.

A recent study identified hundreds of evolutionarily conserved lncRNAs which are syntenically conserved. These lncRNA show tissue specific expression and tend to be genomically associated with coding genes involved in developmental and differentiation processes relevant to that particular tissue, very often regulating developmentally important transcription factors. Additionally, they harbour binding sites for Zinc-finger contacting proteins such as CTCF and YY1 and their binding significantly overlaps with TAD boundaries. Termed as tapRNAs for topological anchor point RNAs because RNAs in this set are linked to chromatin organisation structures, overlapping binding sites for the CTCF chromatin organiser and located at chromatin loop anchor points and borders of topologically associating domains (TADs), 73% of them show high conservation in patches of sequence between human and mouse. Oftentimes, tapRNAs and their neighbouring genes are co-expressed in a tissue-specific manner. Many of the tapRNAs and associated genes are over expressed in different cancer lines. LncRNAs, specifically those possessing the characteristics of tapRNAs, are emerging as important actors in the regulation of nuclear architecture and have been implicated in chromatin looping (Amaral, P.P, et al, 2018).

The characteristics of *mrhl* RNA in mouse and humans cells can be compared to many of the defined features of tapRNAs. *Mrhl* is syntentically conserved and shares partial sequence homology in mouse and humans (Fatima, R. et al 2019). The lncRNA is expressed tissue specifically and is genomically associated with genes involved in tissue-specific processes. *Mrhl* lncRNA regulates key developmental transcription factors including Sox8 in spermatogonial cell, POU3F2, Runx2 and Foxp2 in mESC, and TAL1

and TP53 in K562 cells of CML origin. Additionally, the lncRNA plays a role in cancer development with it being significantly overexpressed in K562 cells, regulating the cancer phenotypes of cell invasion, migration and proliferation and is co-expressed with the host phkb gene (Akhade, V.S. et al, 2015, Pal, D. et al, 2021, Roy, S.R et al, 2020).

Since *mrhl* RNA is involved in mediating a looping switch at the Sox8 in association with the zinc-finger protein CTCF, and it possesses most of the characteristics of tapRNAs, we were curious to see if *mrhl* is involved in defining global chromatin architecture in the mouse spermatogonial cells.

4.1 CTCF binding close to GRPAM

As a first step towards understanding if CTCF and cohesin could be involved along with *mrhl* in the regulation of target genes, the presence of binding of these two proteins in the vicinity of GRPAM was searched for in ENSEMBL database. Sure enough, we could find indication of CTCF and cohesin binding in the vicinity of at least 6 genes and has been summarised in table 3.1 below. While some of the binding sites were very close to the *mrhl* ChOP sites (Sox8, Lrba, Mael and Rab40b), all of them were present very close to or within the target gene. This is a good indicator of the potential regulatory role for CTCF and cohesin in expression of *mrhl* target genes.

Genes	Proteins bound	Distance from mrhl ChOP site	Distance of CTCF from gene
Sox8	CTCF, Rad21, SMC3	3.1kb	Intragenic
Lrba	CTCF, Rad21, SMC3	2.2kb	3kb
Mael	CTCF, Rad21, SMC3	1.2kb	Intragenic
Rab40b	CTCF, Rad21, SMC3	729bp	Intragenic
Rarg	CTCF, Rad21, SMC3	17.9kb	Intragenic
Kcnq5	CTCF, Rad21, SMC3	19.6kb	Intragenic

Table 4.1 - Presence of CTCF and cohesin binding close to a subset of GRPAM, the distance of the binding site from the mrhl ChOP site and the distance of the binding site with respect to target gene.

4.2 CTCF HiChIP

HiChIP was performed in control Gc1-spg cells and cells in which a pool of both mrhl targeting shRNA had been integrated. The success of ChIP for CTCF was evaluated by performing Western blotting with the input (post-ligation sonicated) and ChIP samples (**Figures 4.1 (A)**). Preliminary analysis of the CTCF HiChIP data shows a modest reduction in the number of inter-chromosomal contacts with a concomitant increase in intra-chromosomal contacts (**Figure 4.1(B)**) .

During meiotic commitment and progression, the inter-chromosomal contact to intra-chromosomal contact ratio decreases. This is due to a decrease in inter-chromosomal contact and an increase in intra-chromosomal contacts (Vara, C. et al., 2019). Studies from our group have reported that the downregulation of *mrhl* results in meiotic commitment of mouse spermatogonial cells (Akhade, V.S. et al., 2016). The preliminary analysis of results from the HiChIP experiment too are indicative of the knockdown of *mrhl* resulting in changes in chromosomal contacts characteristic of meiotic progression as noted by Vara, C. et al mentioned above.

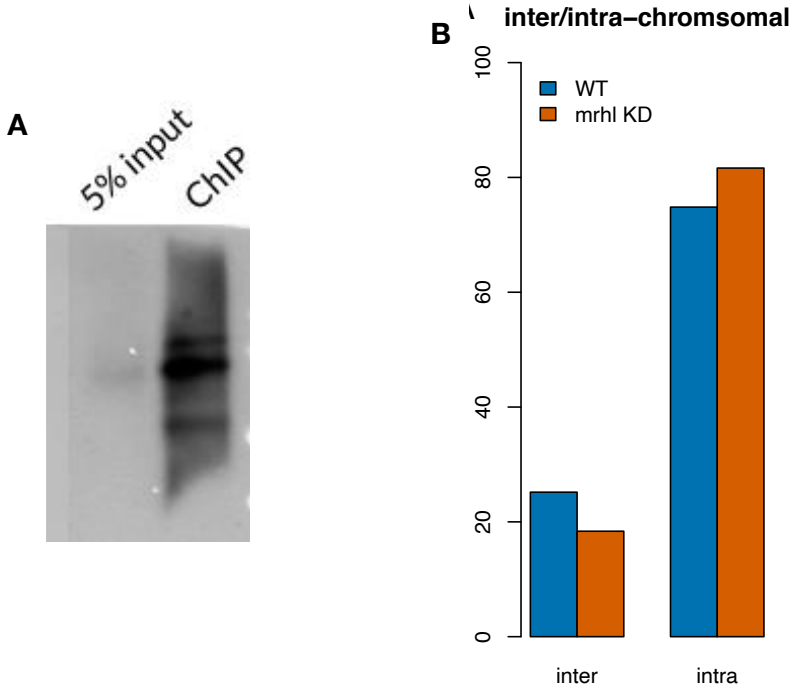


Fig 4.1 - HiChIP in mouse spermatogonial cells *- (A) Western blotting performed to confirm the pulldown of CTCF in the HiChIP workflow shows enrichment of CTCF protein over input. (B) Graph showing the inter- and intra- chromosomal contacts identified in the HiCHIP experiment performed in cells without and with mrhl knockdown shows a modest change in the contacts.*

4.3 CTCF and YY1 ChIP-sequencing and Transcriptome sequencing

ChIP sequencing for both CTCF and YY1 proteins were carried out in cells under the same conditions as those used for the HiChIP experiments. The CTCF ChIP-sequencing data will help in defining the viewpoints for HiChIP. An antibody specific to CTCF from Cell Signalling Technology (CST) that does not cross-react with CTCFL or BORIS has been used for HiChIP and ChIP-sequencing experiments. As information was not available about the specificity of the AbCam CTCF antibody used for the Sox8 locus specific studies, we validated CTCF occupancy at the Sox8 locus in the ChIP samples prepared with the CST CTCF antibody (**Figure 4.2 (B)**) to confirm that CTCF and not CTCFL is bound.

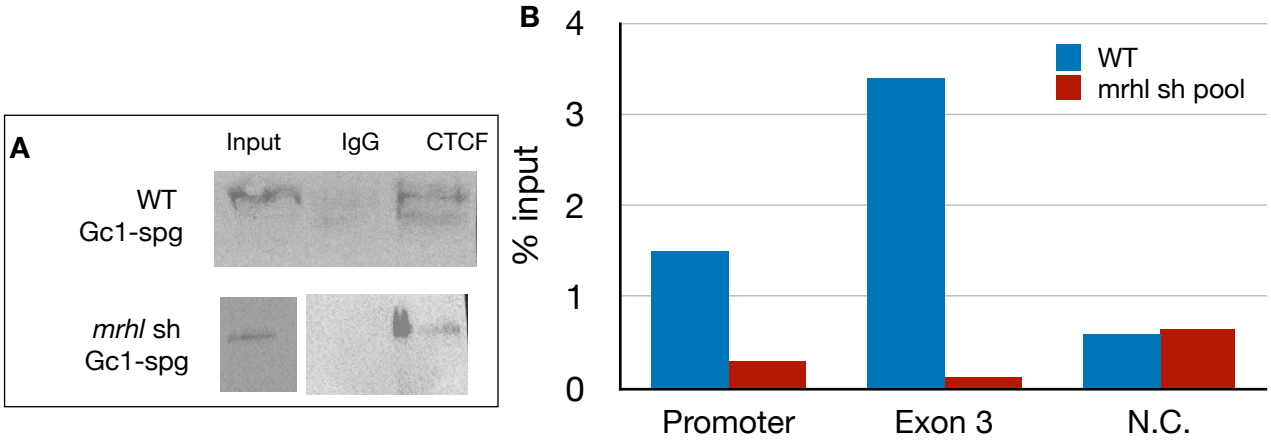


Fig 4.2 - CTCF ChIP samples prepared for sequencing *- (A) ChIP western blotting for CTCF in control and mrhl knockdown Gc1-spg cells shows enrichment of CTCF in IP reactions over isotype control (B) ChIP RT-PCR was performed for promoter and exon 3 of Sox8 locus and the negative control region using the samples (biological duplicates) prepared for sequencing and enrichment of CTCF was observed at the Sox8 locus under control conditions but not upon mrhl knockdown.*

Since the downregulation of *mrhl* leads to looping switch from CTCF-mediated to YY1 mediated at the Sox8 locus, we would like to investigate if this trend is observed at multiple loci and we hope to get an indication of such a switch by comparing the results of YY1 ChIP-sequencing with the HiChIP data.

Finally, we hope to understand if any of the differential loops identified in the presence and knockdown of mrhl contribute functionally to gene expression by comparing the HiChIP data to results from transcriptome sequencing carried out in control and *mrhl* knockdown cells.

The ChIP sequencing data is currently being generated. Preliminary pre-processing of RNA sequencing data shows that many genes involved in meiotic progression are perturbed upon *mrhl* knockdown, again validating this previously generated data from microarray analysis. However, much work remains to be done before conclusive observations can be made from the generated data. We hope to gain valuable insights from analysing these datasets regarding the involvement of *mrhl* lncRNA in regulating gene expression by modulating the 3D organisation of chromatin at multiple loci.

Chapter 5

Discussion

Since the discovery of the first lncRNAs, Xist and H19, the role of lncRNAs in gene regulation has been the focus of a large cohort of studies. With multiple research groups focussing on dissecting the role of lncRNAs, we now know that these lncRNAs recruit multiple protein complexes to target loci to impart their epigenetic changes. Additionally, our understanding of how these lncRNAs themselves interact with target chromatin loci has increased considerably.

5.1 *Mrhl* lncRNA and Sox8

Mrhl lncRNA, a 2.4kb nuclear lncRNA in mouse, is a negative regulator of the Wnt signalling pathway (Arun, G. et al, 2012). It has been shown to bind to and regulate genes at multiple loci within the genome of the mouse B-type spermatogonial cells in association with its partner protein p68. The binding of *mrhl* can be classified based on its proximity to the target gene as promoter- , intragenic- or intergenic -binding (Akhade, V.S. et al, 2015). Sox8 is one among 3 genes which is regulated by *mrhl* binding at the gene promoter.

Sox8 is a developmentally important transcription factor that is critical for the maintenance of adult male fertility. Sox8 knockout mice become progressively infertile because of age-related degeneration of spermatogenesis (O'Bryan, M.K. et al., 2008). The Sertoli specific deletion of Sox9, another essential transcription factor involved in sex determination and maintenance of mammalian testes, in Sox8 null embryonic mice results in failure to achieve the first wave of spermatogenesis (Barrionuevo, F. et al., 2009). The deletion of both Sox8 and Sox8 in adult sertoli cells results in testis to ovarian genetic reprogramming (Barrionuevo, F, et al., 2016) while the presence of Sox8 alone is sufficient for ovarian to testicular genetic reprogramming in the absence of R-spondin1 (Richardson, N. et al., 2020) . Most of these studies have focussed on understanding the role of Sox8 in Sertoli cells in mammalian testes. Studies from our group were the first to not only report the expression of Sox8 in spermatogenic cells but also explore the potential role of this transcription factor in meiotic commitment. Early studies from our group showed that *mrhl* lncRNA bound to the promoter of Sox8 in a p68-dependent manner (Akhade, V.S. et al.,

2015) and the latest study explored the dynamics at the promoter of Sox8 brought about by *mrhl* RNA (Kataruka, S. et al., 2017). The study demonstrated that the Mad-Max transcription factors along with the co-repressors Sin3a and HDAC1 were bound at the Sox8 promoter close to the *mrhl* binding site in the spermatogonial cells in the presence of *mrhl*. This was associated with high levels of the repressive histone modification H3K27me3 and low levels of the histone modifications H3K4me9 and H3K9ac associated with active transcription. There was a concomitant activation of Sox8 expression with downregulation of *mrhl.* Associated changes at the promoter of Sox8 included the Mad-Max transcription factors being replaced by the Myc-Max transcription factors and increased levels of H3K4me3 and H3K9ac histone modifications. The levels of H3K27me3 modification were found to decrease. Simultaneously, beta-catenin was found to bind at the WRE present at the promoter (Kataruka, S. et al., 2017).

Activation of the Wnt signalling cascade in Gc1-spg cells resulted in their meiotic commitment marked by the increase in the levels of pre-meiotic and meiotic markers (Akhade, V.S. et al., 2016). This meiotic commitment was found to be Sox8 dependent. Additionally, binding sites for Sox8 transcription factor were found in the promoters of some of the pre-meiotic and meiotic markers. Stra8 was among these markers (Kataruka, S. et al., 2017). Stra8 is a key regulator of meiosis during spermatogenesis (Feng, C. W, et al, 2014). It is likely that Sox8 regulates meiotic commitment via the master regulator of meiosis, Stra8 (**figure 5.1**). In this context, it is interesting to study the regulation of Sox8 by *mrhl* lncRNA in further detail.

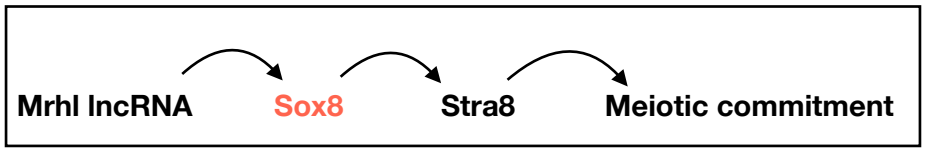


Fig 5.1: Proposed regulatory cascade in spermatogonial cells

5.2 Triple helix in gene regulation

In the current study, we have explored deeper into understanding the role of *mrhl* in modulating the chromatin dynamics at the Sox8 locus. Firstly, we have elucidated through a combination of approaches, Electrophoretic Mobility Shift Assay and Circular Dichoism spectroscopy i*n vitro* and interaction studies *in vivo,* that *mrhl* interacts at the Sox8 locus not only at the *mrhl* binding site at the promoter in a p68-dependent manner but also

directly with the chromatin through the formation of a DNA:DNA:RNA triplex, a mechanism of interaction that is common to many chromatin bound lncRNAs such as Meg3, PARTICLE, HOTAIR and KHPS1 (Mondal, T. et al., 2015, O'Leary, V.B. et al., 2015, Kalwa, M. et al., 2016, Blank-Giwojna, A. et al., 2019)

At the Sox8 locus, a region at the 5' end of *mrhl* is responsible for triplex formation. It is interesting to note the *mrhl* lncRNA harbours multiple potential triplex forming regions within it. The predictions from the *in silico* prediction tool Triplexator for the Sox8 locus had suggested two different regions to have triplex forming potential - one towards the 5' end and another towards the 3' end. While the sequence towards the 5' end performed well in the *in vitro* studies, it is possible that the sequence towards the 3' end forms triplex as well in a different context. The predictions from Triplexator using different genomic regions as input, or instance Pou3f2, Runx2 or FoxP2 (Pal, D et al, 2021), show that other regions within *mrhl* lncRNA, too, have the potential to from triplex **(Figure 5.2)**. It is likely that *mrhl* interacts at multiple other loci through the formation of DNA:DNA:RNA triplex through different TFOs present within it in a context dependent manner.

***Fig 5.2: The various predicted triplex forming sites within mrhl lncRNA**. The regions predicted to form triplex in mESC relevant transcription factors are highlighted in red and the regions predicted to participate in triple formation at the Sox8 locus in black.*

We were interested in further identifying other protein complexes that could be involved in epigenetic gene repression of Sox8. Many genes are repressed through methylation of CpG rich DNA at their promoters. The presence of a 1.3kb long CpG island encompassing the promoter of Sox8 suggested that a probable mechanism of gene repression could be through the methylation of this CpG island. Moreover, triplex formation by the ncRNA pRNA has been previously reported to act as a platform for the recruitment of the DNA methyltransferase DNMT3b at the rDNA promoter which goes on to methylate the DNA at the gene promoter, thereby repressing transcription. We wanted to explore if triplex formation by *mrhl* too could be recruiting DNA methyltransferases similarly. However, no reduction in methylation levels is observed experimentally corresponding to Sox8 activation in either the mouse spermatogonial cell line Gc1-spg upon Wnt induction or in 21-day old mouse testes when compared to 7-day old mouse testes. This suggests that

DNA methylation is not the mechanism of epigenetic repression of Sox8 and that it is unlikely that the DNA:DNA:RNA triplex recruits DNA methyltransferases to the Sox8 locus.

5.3 PRC2 in gene regulation

The presence of high levels of H3K27me3 repressive histone mark in the Sox8 transcriptional repressed state (Kataruka, S. et al, 2017) was indicative of the presence of PRC2 at the Sox8 locus. PRC2 is the multi-protein complex responsible for catalysing the methylation of H3K27. We confirm the presence of this complex by ChIP for the Ezh2 subunit of PRC2. The occupancy of Ezh2 is no longer observed upon Sox8 transcriptional activation when the levels of H3K27me3 are relatively low. A common feature of the mammalian PRC2-binding region is the presence of CpG islands (CGIs) and PRC2 binding is enriched at CpG-rich regions which are adjacent to the TSS of silenced genes. Multiple studies indicate that high-density DNA methylation seems to be mutually exclusive with PRC2 since most of the CGIs or CG-rich regions occupied by PRC2 are hypomethylated (Yang, Y. and Li, G., 2020). The trend observed at the Sox8 locus is in agreement with these studies. The promoter of Sox8 is situated within a CpG island and the levels of methylation are lower in the Sox8 transcriptionally repressed state than in the active state.

Another factor influencing PRC2 binding to target loci is its interaction with RNA molecules. With multiple studies uncovering that lncRNAs bind to and recruit PRC2 to target loci, it is now believed that lncRNA interaction could be one of the mechanisms by which PRC2 gains target specificity. Multiple lncRNAs such as PARTICLE and Meg3 recruit PRC2 to target loci through triplex formation (Mondal, T. et al., 2015, O'Leary, V.B. et al., 2016). Different subunits of PRC2 recognise and bind to different secondary structures/ DNA sequences through which they get targeted specifically to genomic loci. For instance, unmethylated GCG trinucleotide motif showing an unwound DNA helix (compared to canonical B-DNA) can specifically recruit PRC2-MTF2 while the Suz12 subunit has been reported to bind to the two-hairpin motif present in RNA molecules. PRC2 subunit JARID2 preferentially binds to GC rich DNA sequences (Yang, Y. and Li, G., 2020). At the same time, other reports suggest that PRC2 could also bind to RNA molecules promiscuously (Davidovich, C. et al, 2014). At the Sox8 locus, multiple possible modes of recruitment of PRC2 exist, namely, the presence of a hypomethylated CpG island, the presence of *mrhl* lncRNA and also the formation of triplex by *mrhl* lncRNA. It would be interesting to

determine which one or a combination of these elements are required for PRC2 recruitment in the future.

5.4 CTCF and cohesin in gene regulation

Using a combinatorial approach of utilising information available in ENSEMBL database, analysis of ChIP-seq data and ChIP qPCR experiments, we show the *mrhl* dependent occupancy of CTCF and cohesin at the Sox8 locus. CTCF is also known as the master weaver of the genome (Phillips, J.E and Corces, V.J, 2009) due to its widespread role in mediating inter- and intra-chromosomal contacts in mammals. In a manner reminiscent of the repressive complex present on the maternal allele of the H19/Igf2 imprinting complex that ensure that H19 gene is expressed from the maternal allele (Yao, H. et al., 2010). CTCF and cohesin are found associated at the Sox8 locus in a *mrhl*- dependant manner along with the DEAD-box RNA helicase p68.

CTCFL is a testis-specific paralog of CTCF that recognises the same DNA motif. CTCFL is expressed only transiently in pre-meiotic male germ cells together with CTCF and the two paralogs compete for binding at a subset of the CTCF binding sites (Nishana, M. et al, 2020) . CTCFL functions as a transcription factor and does not have a role in chromatin organisation like CTCF as it can not anchor cohesin to chromatin like CTCF can (Pugacheva, E. et al, 2020). ChIP qPCR using CTCF specific antibody in the spermatogonial cells performed by us further confirms that CTCF and not CTCFL is bound at the Sox8 locus. It has been previously demonstrated that lncRNA SRA and p68 are essential components of the repressive complex at the H19/Igf2 imprinted locus as they bind to both CTCF and cohesin and stabilise their interaction. The deletion of either one of the components resulting in the destabilisation of the complex resulting in reduced insulation by CTCF and increase in Igf2 gene expression from the maternal allele (Yao, H. et al, 2010). At the Sox8 locus, the knockdown of *mrhl* lncRNA results in the eviction of CTCF and cohesin. *Mrhl* binding itself is p68 dependent. It is likely that *mrhl* and p68 perform a role similar to SRA and p68 in stabilising the CTCF-cohesin complex.

The widely accepted loop-extrusion model of chromatin loop formation proposes that the cohesin protein complex slides along chromatin forming a growing loop until it meets two CTCF molecules bound with convergent orientation. This prevents cohesin from sliding further. Preliminary *in silico* analysis suggests the presence of CTCF binding sites both at the promoter and within exon3 of Sox8. The prediction software utilised for this study utilises PWMs based on the most widely found CTCF binding sites present in the

mammalian genome and does not exhaustively predict the presence of all CTCF binding sites present within the sequence. Moreover, heterogeneity is observed in CTCF binding motifs. In each species, the CTCF binding profile is composed of substantial numbers of both deeply conserved and evolutionarily recent sites. CTCF binding sites at TAD boundaries are highly conserved across species while evolutionarily recent sites play role in modulating gene regulation (Kentepozidou, E. et al, 2020). Further, cell-type specific CTCF bound sites have also been reported to have a varied binding motif as compared to constitutively bound sites (Essien, K. et al, 2009). Additionally, clusters of closely located CTCF binding sites help stabilize cohesin and are located significantly closer to TSSs (Kentepozidou, E. et al, 2020). We hope to be able to conclusively address questions about the presence of CTCF binding sites at the Sox8 locus, the number of binding sites and their orientations in depth utilising the data from our CTCF ChIP-seq performed in the mouse spermatogonial cells (ongoing work).

DNA methylation at the DMR regulating the imprinting at the H19/Igf2 locus prevents the binding of CTCF to its cognate binding site within the DMR. The slight increase in the methylation at the CpG island of the Sox8 promoter upon its transcriptional activation may possibly serve the same purpose.

5.4.1 CTCF associated MAZ protein

The presence of binding site for the Myc-associated protein MAZ has been reported close to CTCF binding sites in the genome. MAZ stabilises CTCF binding and similar to CTCF, MAZ too binds at insulator elements and prevents enhancer - promoter contact. Additionally, CTCF/MAZ double sites are more effective at sequestering cohesin than sites occupied only by CTCF (Xiao, T. *et al*, 2021). Interestingly, data available in ENSEMBL shows MAZ binding within exon 3 of Sox8 close to the CTCF binding site **(Figure 5.3)** while GPminer predicts the presence of MAZ binding site at the promoter of Sox8 (Figure 3.18 A).

MEL cell line	DNase1 & TFBS	MAZ
MEL cell line	DNase1 & TFBS	MEF2A
MEL cell line	DNase1 & TFBS	MXI1
MEL cell line	DNase1 & TFBS	MYB
MEL cell line	DNase1 & TFBS	MYC
MEL cell line	DNase1 & TFBS	Max
MEL cell line	DNase1 & TFBS	NELFe
MEL cell line	DNase1 & TFBS	NRF1
MEL cell line	Hists & Pols	PolII
MEL cell line	DNase1 & TFBS	RCOR1
MEL cell line	DNase1 & TFBS	Rad21
MEL cell line	DNase1 & TFBS	SIN3A
MEL cell line	DNase1 & TFBS	SMC3

Fig 5.3: *ENSEMBL data shows the occupancy of MAZ (highlighted in red) in the vicinity of CTCF binding site at the Sox8 locus.*

CTCF binding at the Sox8 locus is observed in those tissues in which *mrhl* is expressed such as the kidney and and not in those tissues in which *mrhl* is not expressed such as the brain and heart. Thus, *mrhl* is possibly involved in the regulation of Sox8 by associating with cohesin and p68 in multiple tissues and this mode of regulation is not restricted to only spermatogonial cells.

5.5 Silencer elements in gene regulation

In addition to promoters, silencers/insulators and enhancers together make up *cis*-regulatory elements (CREs) of a gene. Silencer elements are harder to identify and characterise than enhancers since there are no distinctive features associated with these CREs. H3K27me3 mark enrichment, apart from characterising transcriptional repression, has been found to be enriched within silencer elements. Most H3K27me3^{+} silencer elements are DNase I Hypersensitive and have binding sites for ubiquitous repressors such as CTCF, SMAD group of proteins and tissue specific TFBS such as STAT family of proteins. In mammalian cells, silencer elements have an elevated TF binding of CTCF and its working partners SMC3 and RAD21 and repressor proteins such as PAX5 and RUNX3 (Huang, D. et al, 2019). The authors of the same study also found that almost 75% of H3K27me3-DHS coincided with active histone modifications such as H3K4me1 and H3K27ac. The element within exon 3 has many of these characteristics. From experiments, we known that there is occupancy of CTCF and cohesin within this genomic region. Additionally, information available in ENSEMBL database suggests that this element is DNase hypersensitive and shows the presence of both H3K4me1 and K3K27me3. The results of the 3C experiment further indicate that this element contacts the gene promoter in the transcriptionally repressed state. Taken together, all these evidences support the ‘silencer’ function of exon 3 of Sox8.

The 3D genome browser allows one to visualise the results from multiple chromatin interaction datasets such as HiC, ChIA-PET and hiChIP performed in human and mouse cells (Wang, Y. et al., 2018) . Additionally, the browser also gives information about the predicted linkage of DNase hypersensitive regulatory elements to their target genes based on various parameters. While none of the available datasets had the resolution (minimum

resolution of 5kb) to visualise the interaction between the Sox8 promoter and exon 3 (a distance of 3.5 - 4kb), DHS link map predicts the interaction between these two elements (shown in **figure 5.4** below). Not only that, the map also predicts an interaction between the promoter and a region upstream, presumably the triplex forming region. This suggests that under the transcriptionally repressed state, the promoter not only contacts the silencer element present downstream but also the TTS present upstream. Due to limitations inherent to the experimental design, we could not query for this interaction in the 3C experiments.

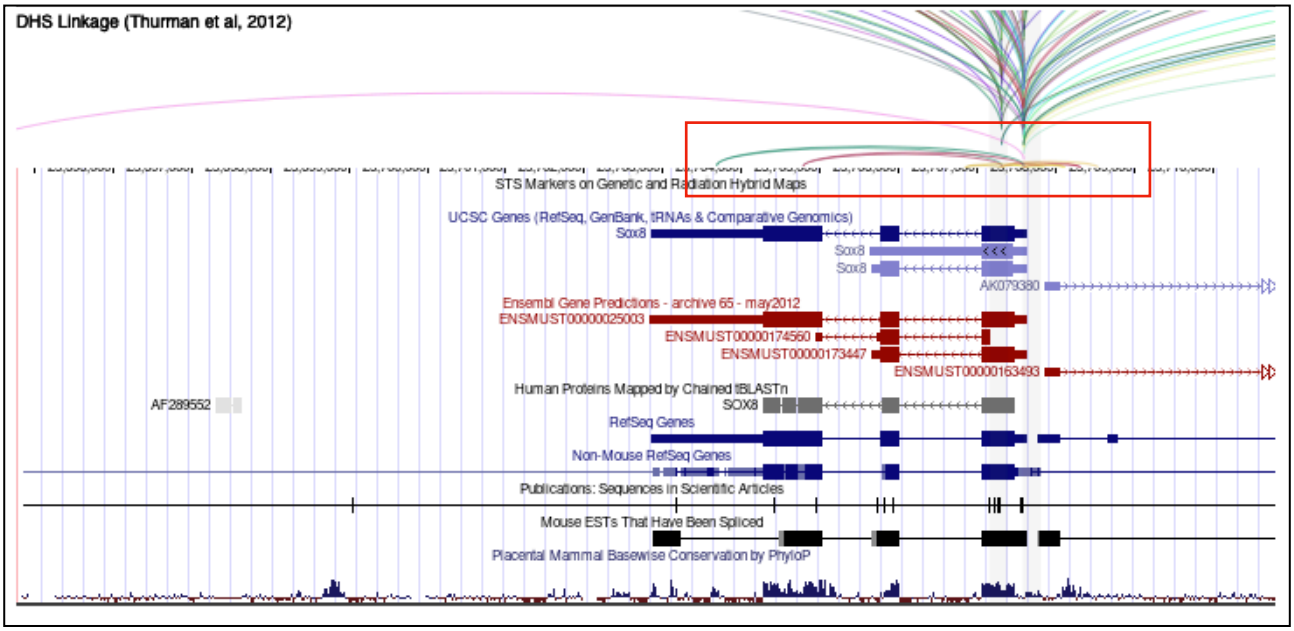


Fig 5.4: DHS linkage map for Sox8 locus from 3D Genome Browser *(highlighted in red box) - DHS linkage map shows predicted linkage between the promoter and exon 3 of Sox8. Additionally, there is predicted linkage between promoter and a region upstream, presumably the triplex forming region.*

5.6 Enhancer elements in gene regulation

Thus far, the promoter and insulator elements of Sox8 have been discussed in detail. This leaves out the third CRE - the enhancer.

Based on evidence from literature, we have identified two putative enhancers for Sox8 in spermatogonial cells located downstream of the gene. We observe activity at both these enhancer elements upon *mrhl* knockdown as evidenced from enhancer specific histone modification ChIP experiments. Only one of these two regions, the enhancer present 8kb downstream, contacts the promoter of Sox8 as observed from the 3C experiment and is likely to drive the expression of Sox8 in meiotically committed spermatogonia. This

enhancer harbours within it one of the evolutionarily conserved element, E6, identified in the study by Guth, et al, (Guth, S.I et al, 2010) as a putative enhancer. However, this does not mean that the other enhancer element has no role to play in regulating Sox8 expression or that the E6 harbouring enhancer is the sole enhancer regulating the expression of Sox8.

Sox9 is regulated by multiple tissue specific enhancers and in the testis, is regulated by multiple enhancers acting synergistically. The upstream regulatory region of *Sox9* is very complex and all the putative enhancers (33 of them) identified in the upstream regulatory region of Sox9 and indicated with the enhancers that drive Sox9 expression in the testis highlighted in blue in **figure 5.5** below (Gonen, N et al, 2018). It is possible that multiple enhancers regulate the expression of Sox8 too in a similar manner. For instance, presence of a breakpoint between the conserved elements E1 and E2 negatively impacted testis determination resulting in 46 XY, disorder in sex determination (DSD) in humans (Portnoi, M.F. et al, 2018) indicating that the enhancer elements present upstream of the gene too could be playing a role in regulating Sox8 expression in gonadal cells.

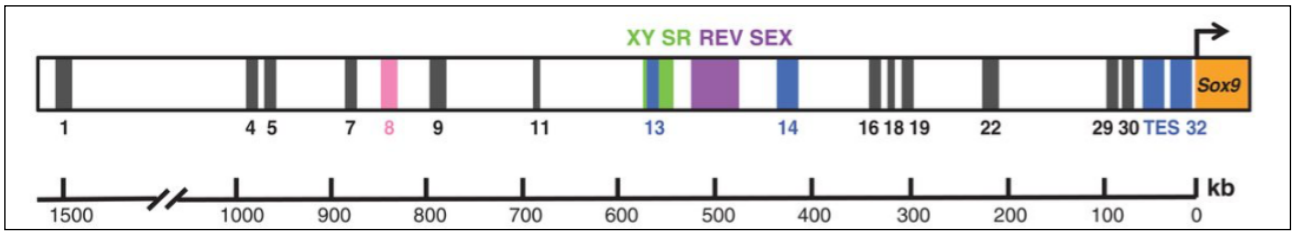


Fig 5.5: Putative enhancer elements present in the regulatory region upstream of Sox9 *- The enhancer elements that drove testis specific expression have been highlighted in blue. These 4 elements are likely to regulate Sox9 expression synergistically. Image taken from Gonen N, Futtner CR, Wood S, Garcia-Moreno SA, Salamone IM, Samson SC, Sekido R, Poulat F, Maatouk DM, Lovell-Badge R. Sex reversal following deletion of a single distal enhancer of Sox9. Science. 2018 Jun 29;360(6396):1469-1473. Reprinted with permission from AAAS.*

The studies based on which the two downstream elements were chosen were all performed in embryonic gonads of specific age groups only and thus, gives us no information on the regulatory elements in action post-natally. Our study has focussed on characterising the regulatory elements of Sox8 in a genomic region of 25-30kb only. Extensive characterisation including genomic deletion of the regulatory elements in a larger region is required to identify and better understand the possible interplay between various enhancers in regulating Sox8 expression. Such characterisation is beyond the

scope of the current study. Further, the possibility of *trans*- interactions regulating Sox8 expression has not been explored.

The regulatory role of the E6 harbouring enhancer in somatic cells is not known. According to ENSEMBL, this enhancer region shows activity in cells of the heart and brain also. However, our analysis of enhancer histone modification ChIP-seq data of the mouse adult brain cortex did not show an enrichment of with H3K4me1 or H3K27ac at the E6 enhancer region (data not shown). Despite that, the possibility that E6 harbouring enhancer may not be a testis-specific enhancer still exists.

5.7 YY1 in gene regulation

Of all the regulatory proteins identified at the Sox8 locus, YY1 is the only one with contradictory functional roles. YY1 can act both as a transcriptional activator or transcriptional repressor in a context dependent manner and because of this reason, is named as Ying-Yang1. YY1 can interact with the PRC2 complex and HDACs to mediate transcriptional repression while it can activate transcription by interacting with HATs (Verheul, T.C.J., et al, 2020). Myc-Max dimer is a reported transcriptional activator dimer (Grandori, C. et al, 2000). However, YY1 associates with the Myc-Max transcriptional regulators and mediate transcriptional repression at the *α*3 gene in osteosarcoma cells (Nigris, F. d., et al, 2006). Other reports suggest that YY1 can only interact with Myc and not the Myc-Max dimer and that YY1- Myc can activate gene expression (Vernon, E.J and Gaston, K. 2000).

Even as an architectural protein, YY1 can mediate the formation of chromatin loops which can either have gene repressive or activating outcomes. YY1 dimerises with CTCF to mediate chromatin loop formation to repress E6 and E7 oncogenes of the human papillomavirus genome in infected cells (Pentland, I. et al, 2018). At the same time, YY1 binds to promoter-proximal elements and active enhancers and forms dimers that facilitate the interaction of these DNA elements (Weintraub, A. et al, 2017). Thus YY1 at the Sox8 locus was a wildcard that could be involved in either function. However, the results from the ChIP experiments clearly indicated the association of YY1 at the regulatory elements only in the Sox8 transcriptionally active state. Further, the occupancy of YY1 at the promoter and enhancer elements suggested a role for it in facilitating the interaction of the enhancer with the promoter and this has been validated by chromosome conformation capture.

Mrhl lncRNA is a 2.4kb long transcript that has multiple functional domains within it. At the Sox8 locus, a region from the 5' end of the lncRNA participates in triplex formation. Results from previous work suggest that a region towards the 3' end of *mrhl* is involved in its interaction with p68 (Kayyar, B. M.S Thesis). The gene regulatory function is an outcome of the combinatorial function of all the different domains of *mrhl*. Further, many lncRNAs involved in 3D genome organisation, specifically tapRNAs, have been reported to interact directly with CTCF. *Mrhl,* too, could be a CTCF interacting RNA. *Mrhl,* then, can be categorised as scaffold lncRNA which functions to bring together multiple regulatory proteins at the target locus.

The mechanism of silencing at the Sox8 locus by *mrhl* lncRNA via triplex formation, PRC2 recruitment, and the involvement of CTCF, cohesin and p68 fits into the growing theme of gene silencing mechanism by lncRNAs. Stating that *mrhl* associates with this protein complex to mediate the formation of a repressive loop is a simplistic view of events. Taking into account the very large size of the repressive complex made up of CTCF, cohesin complex, Sin3a, HDAC1, Mad-Max transcription factor dimer, p68, PRC2 and *mrhl*, it would be more realistic to state that *mrhl* creates a repressive environment around the Sox8 locus.

5.8 Bidirectional promoter of Sox8

Another interesting aspect of the Sox8 locus is its bidirectional promoter. The lncRNA Cerox1 is transcribed from the bidirectional promoter antisense to Sox8. Cerox1 has been recently characterised as a cytoplasmic lncRNA that plays a role as a post transcriptional regulator of mitochondrial complex I catalytic activity by competing with complex I transcripts for the binding of miR-488-3p (Sirey, T.M., et al, 2019).

Looking at the tissue specific expression patterns of Sox8 and the antisense lncRNA Cerox1 from NCBI, the two genes are co-expressed (**Figure 4.6**). We were interested to see if the two genes are co-regulated by *mrhl.* This was particularly of interest since the Sox8 TTS falls within the intron of Cerox1. To this end, we scored for the expression level of Cerox1 upon Wnt signalling activation in Gc1-spg cells and in *mrhl* knockdown cells. While we observed an upregulation for Cerox1 upon Wnt activation, the same trend could not be recapitulated in *mrhl* knockdown cells. Due to this, we did not continue the pursuit.

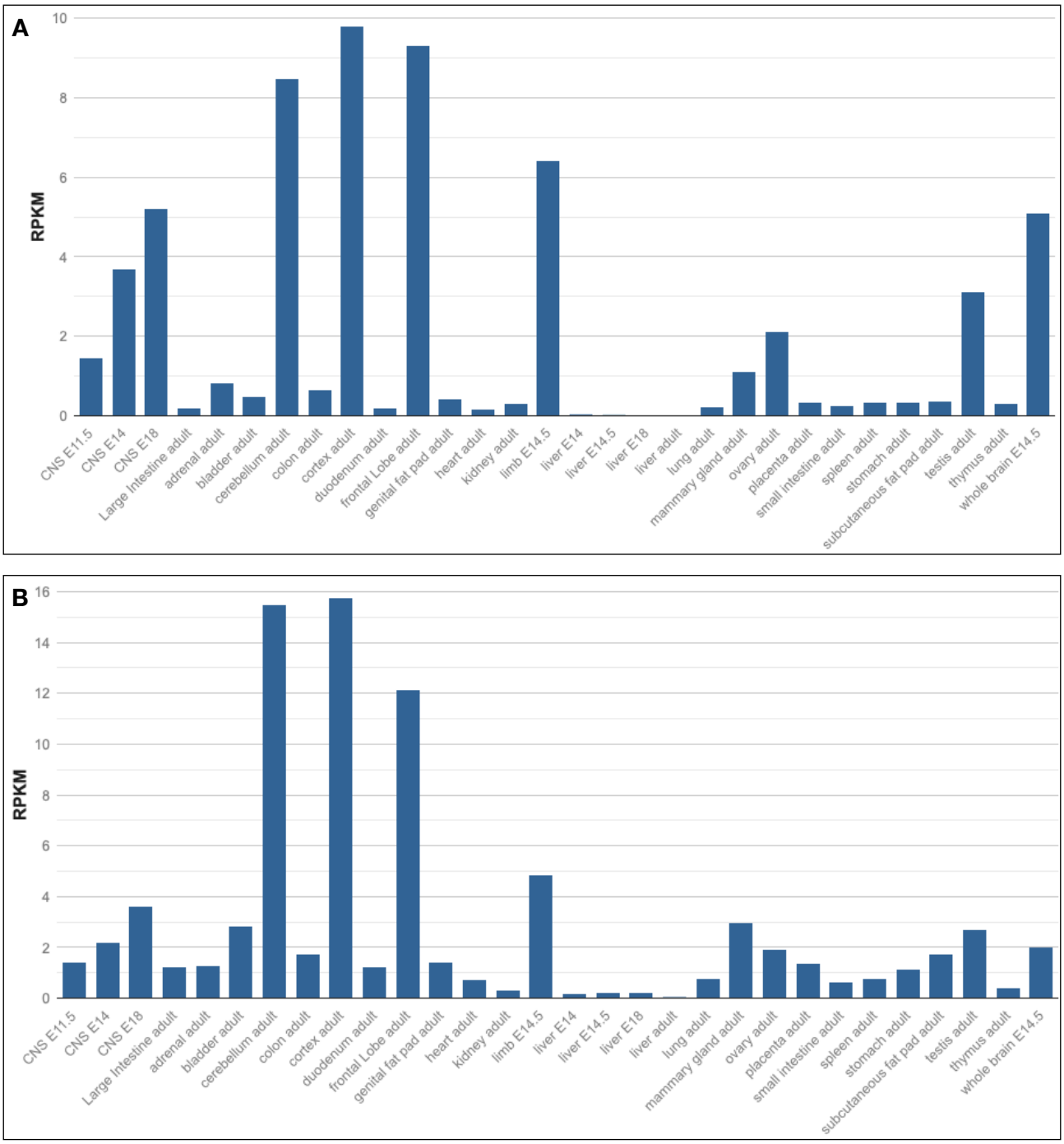
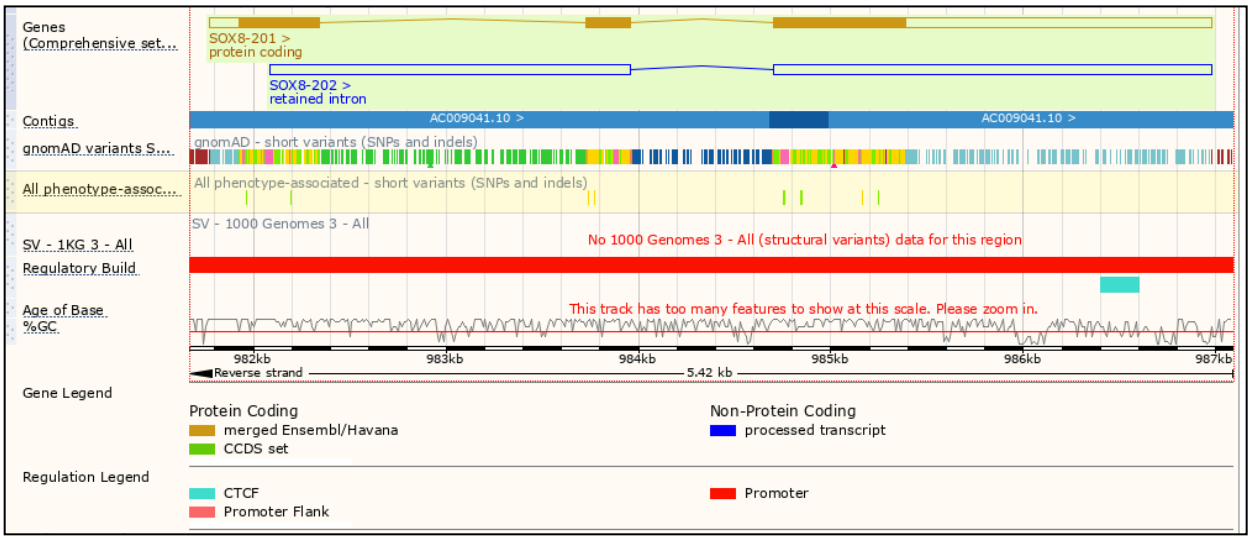


***Fig 5.6** : Tissue specific expression of (A) Sox8 and (B) Cerox1 from NCBI is indicative of co-expression of the two genes*

In the recent update of ENSEMBL, the human Sox8 locus can be seen to not only contain a conserved CTCF binding site in exon 3 (**Figure 5.7 (A)**) but also has binding of many of the regulatory protein induing PRC2, CTCF and cohesin, Sin3a, YY1 and Myc-Max transcription factors keeping in line with our current study on the mouse Sox8 locus (**figure 5.7(B)**). It is interesting to note that YY1 appears to be bound in the same transcriptional context as cohesin binding in the human Sox8 locus in contrast to what is observed in the

mouse Sox8 locus. While the data is indicative of a conserved regulatory mechanism, the involvement of a lncRNA in the regulation of Sox8 in humans remains to be seen.

A



B

SK-N.	DNase1 & TFBS	Rad21
SK-N.	DNase1 & TFBS	SIN3A
SK-N.	DNase1 & TFBS	TAF1
SK-N.	DNase1 & TFBS	TEAD4
SK-N.	DNase1 & TFBS	Tcf12
SK-N.	DNase1 & TFBS	USF1
SK-N.	DNase1 & TFBS	USF2
SK-N.	DNase1 & TFBS	Yy1
NPC_2	DNase1 & TFBS	CTCF
NPC_2	DNase1 & TFBS	DNase1
NPC_2	DNase1 & TFBS	EZH2
MCF-7	DNase1 & TFBS	HDAC2
MCF-7	DNase1 & TFBS	HES1
MCF-7	DNase1 & TFBS	HSF1
MCF-7	DNase1 & TFBS	JUN
MCF-7	DNase1 & TFBS	Jund
MCF-7	DNase1 & TFBS	MAFK
MCF-7	DNase1 & TFBS	MNT
MCF-7	DNase1 & TFBS	MYC
MCF-7	DNase1 & TFBS	Max

Fig 5.7 - The human Sox8 locus from ENSEMBL - (A) *CTCF binding site within exon 3 is conserved in human (B) Many of the proteins regulating mouse Sox8 expression are also seen at the human Sox8 locus including cohesin (Rad21), Sin3A, PRC2 (Ezh2), CTCF, YY1, Myc and Max.*

It is evident from the reports in literature, the deluge of ongoing research in the field and the aforementioned results of the current study that gene regulation in higher eukaryotes is a very complex phenomenon involving an intricate interplay of genomic elements and epigenetic events. It is by no means the outcome of a single or even a handful of regulatory molecules. While our understanding of the series of events contributing to gene regulation has increased substantially in the past few decades, it is evident that there are many gaps yet to be filled before we have a complete picture of gene regulation. With an increase in interdisciplinary research, many previously unexplored concepts are being explored towards gaining a comprehensive picture. For instance, the concept of liquid-liquid phase separation in regulating biological processes is nascent but it is already clear that it is an important determinant in the fate of genes and, as an extension, the fate of the cell. These avenues have not been explored for *mrhl* lncRNA. We believe that with advances in our understanding of gene regulation, many more roles of *mrhl* lncRNA and lncRNAs as a collective will come to light.

Chapter 6

Summary and Conclusions

To summarise the results of the current study, *mrhl* lncRNA is involved in the regulation of the mouse Sox8 gene by modulation the chromatin dynamics at the locus. The lncRNA interacts directly with the chromatin at the Sox8 locus through the formation of a DNA:DNA:RNA triplex. The presence of *mrhl* lncRNA at the locus correlates with the occupancy of PRC2 which is responsible for the H3K27me3 histone modification. Additionally, m*rhl* mediates the formation of a repressive chromatin loop. This loop brings the gene promoter in contact with the silencer element present within Sox8 exon 3. CTCF binds at both the loop anchor points i.e, promoter and silencer, and dimerises to bring them in contact with each other and thereby mediates loop formation. The loop is stabilised by the cohesin complex. *Mrhl* does not bind at the locus in the absence of p68 and CTCF occupancy is not observed in the absence of *mrhl*. Therefore, *mrhl* and p68 are both essential components of this repressive complex. Upon *mrhl* downregulation, enhancer elements downstream of the Sox8 gene are activated. YY1 occupancy is observed at Sox8 promoter and at the enhancer elements. Further, promoter-silencer contact gives way to a promoter-enhancer contact that is mediated by YY1 dimerisation. The enhancer that contacts the promoter harbours within it an evolutionarily conserved element previously suspected to have enhancer activity. This looping switch is accompanied by the repressive Mad-Max transcription factors, PRC2 and the co-repressors Sin3a and HDAC1 being replaced by the activating Myc-Max transcription factors along with the transcriptional co-activator Pcaf. There is a concomitant decrease of the repressive histone modification H3K27me3 and increase in H3K9ac and H3K4me3 activating histone modifications. The chromatin dynamics mediated by *mrhl* identified by the current study are summarised in **figure 6.1** below.

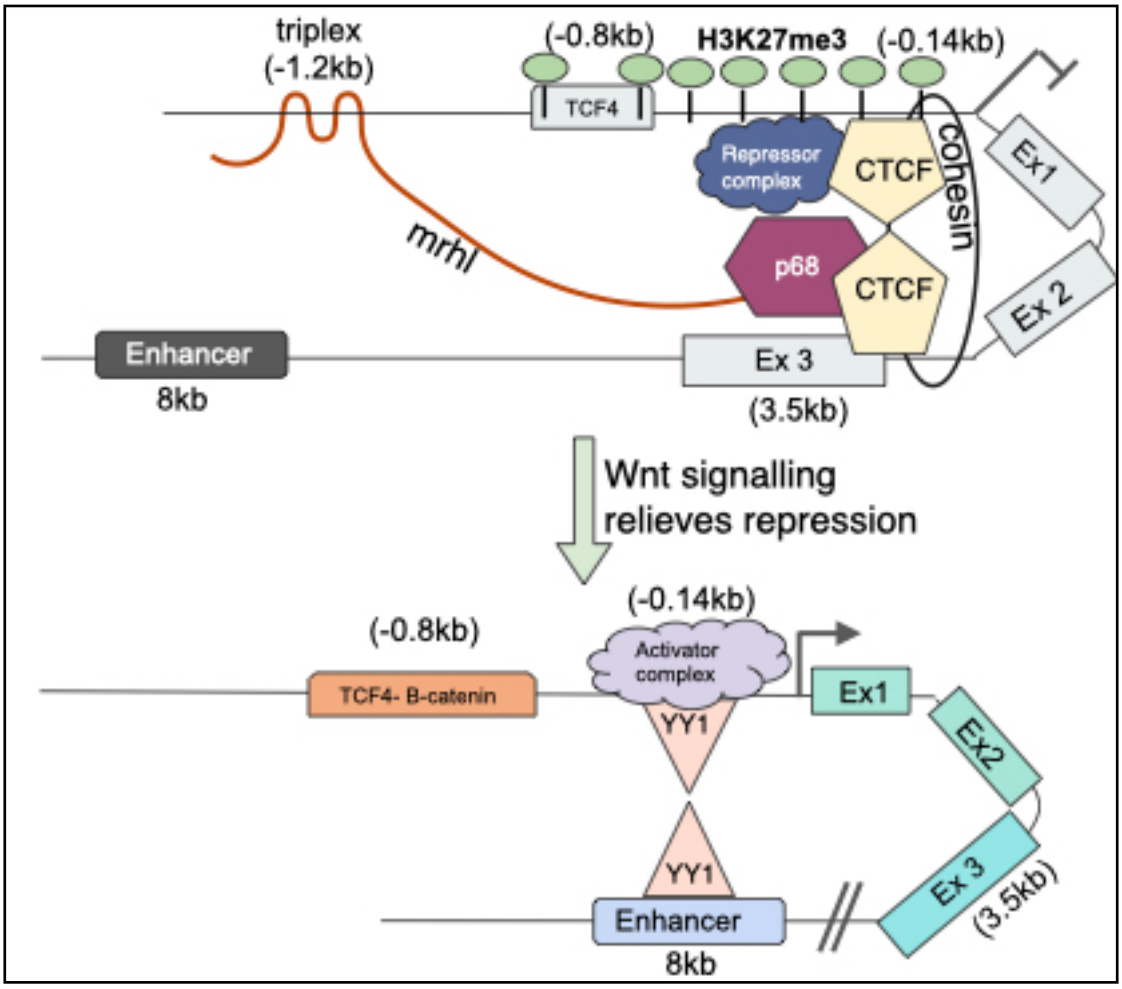


Fig 6.1: Regulatory events at the Sox8 locus mediated by mrhl lncRNA

The preliminary results from the genome wide studies validate the previous observations made from work carried out by our group. The preliminary data from HiChIP show a decrease in inter-chromosomal contacts and an increase in intra-chromosomal contact upon knockdown of *mrhl* lncRNA. This change in contacts is characteristic of meiotic progression in germ cells (Vara, C. et al, 2019). Similarly, pre-processed data from RNA-seq from *mrhl* downregulated cells too shows perturbation of genes involved in regulating meiosis and meiosis relevant cellular events upon *mrhl* knockdown. This is in agreement with work from our group which has showed that *mrhl* downregulation results in the meiotic commitment of the spermatogonial cells. Complete analysis of the generated NGS data will provide valuable insights into the genome wide chromatin organisation mediated gene regulatory function of *mrhl* lncRNA in the context of meiotic commitment of mouse spermatogonial cells.

References

1. Akhade, Vijay Suresh, Gayatri Arun, Sainitin Donakonda, and Manchanahalli R. Satyanarayana Rao. “Genome Wide Chromatin Occupancy of Mrhl RNA and Its Role in Gene Regulation in Mouse Spermatogonial Cells.” RNA Biology 11, no. 10 (2014): 1262–79. https://doi.org/10.1080/15476286.2014.996070.
2. Akhade, Vijay Suresh, Shrinivas Nivrutti Dighe, Shubhangini Kataruka, and Manchanahalli R. Satyanarayana Rao. “Mechanism of Wnt Signaling Induced down Regulation of Mrhl Long Non-Coding RNA in Mouse Spermatogonial Cells.” Nucleic Acids Research 44, no. 1 (January 8, 2016): 387–401. https://doi.org/10.1093/nar/gkv1023.
3. Alahari, S. V., S. C. Eastlack, and S. K. Alahari. “Role of Long Noncoding RNAs in Neoplasia: Special Emphasis on Prostate Cancer.” International Review of Cell and Molecular Biology 324 (2016): 229–54. https://doi.org/10.1016/bs.ircmb.2016.01.004.
4. Ali, Tamer, and Phillip Grote. “Beyond the RNA-Dependent Function of LncRNA Genes.” ELife 9 (October 23, 2020): e60583. https://doi.org/10.7554/eLife.60583.
5. Almeida, Mafalda, Greta Pintacuda, Osamu Masui, Yoko Koseki, Michal Gdula, Andrea Cerase, David Brown, et al. “PCGF3/5-PRC1 Initiates Polycomb Recruitment in X Chromosome Inactivation.” Science (New York, N.Y.) 356, no. 6342 (June 9, 2017): 1081–84. https://doi.org/10.1126/science.aal2512.
6. Amaral, Paulo P., Tommaso Leonardi, Namshik Han, Emmanuelle Viré, Dennis K. Gascoigne, Raúl Arias-Carrasco, Magdalena Büscher, et al. “Genomic Positional Conservation Identifies Topological Anchor Point RNAs Linked to Developmental Loci.” Genome Biology 19, no. 1 (March 15, 2018): 32. https://doi.org/10.1186/s13059-018-1405-5.
7. Arun, Gayatri, Vijay Suresh Akhade, Sainitin Donakonda, and Manchanahalli R. Satyanarayana Rao. “Mrhl RNA, a Long Noncoding RNA, Negatively Regulates Wnt Signaling through Its Protein Partner Ddx5/P68 in Mouse Spermatogonial Cells.” Molecular and Cellular Biology 32, no. 15 (August 2012): 3140–52. https://doi.org/10.1128/MCB.00006-12.
8. Arzate-Mejía, Rodrigo G., Félix Recillas-Targa, and Victor G. Corces. “Developing in 3D: The Role of CTCF in Cell Differentiation.” Development (Cambridge, England) 145, no. 6 (March 22, 2018): dev137729. https://doi.org/10.1242/dev.137729.
9. Azam, Sikandar, Shuai Hou, Baohui Zhu, Weijie Wang, Tian Hao, Xiangxue Bu, Misbah Khan, and Haixin Lei. “Nuclear Retention Element Recruits U1 SnRNP Components to Restrain Spliced LncRNAs in the Nucleus.” RNA Biology 16, no. 8 (August 2019): 1001–9. https://doi.org/10.1080/15476286.2019.1620061.
10. Bacolla, Albino, Guliang Wang, and Karen M. Vasquez. “New Perspectives on DNA and RNA Triplexes As Effectors of Biological Activity.” PLoS Genetics 11, no. 12 (December 2015): e1005696. https://doi.org/10.1371/journal.pgen.1005696.
11. Bae, Sunhee, and Bluma J. Lesch. “H3K4me1 Distribution Predicts Transcription State and Poising at Promoters.” Frontiers in Cell and Developmental Biology 8 (2020): 289. https://doi.org/10.3389/fcell.2020.00289.

12. Ballantyne, Md, Ra McDonald, and Ah Baker. “LncRNA/MicroRNA Interactions in the Vasculature.” Clinical Pharmacology & Therapeutics 99, no. 5 (May 2016): 494–501. https://doi.org/10.1002/cpt.355.
13. Barrionuevo, Francisco, Ina Georg, Harry Scherthan, Charlotte Lécureuil, Florian Guillou, Michael Wegner, and Gerd Scherer. “Testis Cord Differentiation after the Sex Determination Stage Is Independent of Sox9 but Fails in the Combined Absence of Sox9 and Sox8.” Developmental Biology 327, no. 2 (March 15, 2009): 301–12. https://doi.org/10.1016/j.ydbio.2008.12.011.
14. Barrionuevo, Francisco J., Alicia Hurtado, Gwang-Jin Kim, Francisca M. Real, Mohammed Bakkali, Janel L. Kopp, Maike Sander, Gerd Scherer, Miguel Burgos, and Rafael Jiménez. “Sox9 and Sox8 Protect the Adult Testis from Male-to-Female Genetic Reprogramming and Complete Degeneration.” ELife 5 (June 21, 2016): e15635. https://doi.org/10.7554/eLife.15635.
15. Barutcu, A. Rasim, Philipp G. Maass, Jordan P. Lewandowski, Catherine L. Weiner, and John L. Rinn. “A TAD Boundary Is Preserved upon Deletion of the CTCF-Rich Firre Locus.” Nature Communications 9, no. 1 (April 13, 2018): 1444. https://doi.org/10.1038/s41467-018-03614-0.
16. Baylin, Stephen B., and Peter A. Jones. “Epigenetic Determinants of Cancer.” Cold Spring Harbor Perspectives in Biology 8, no. 9 (September 1, 2016): a019505. https://doi.org/10.1101/cshperspect.a019505.
17. Beagan, Jonathan A., Michael T. Duong, Katelyn R. Titus, Linda Zhou, Zhendong Cao, Jingjing Ma, Caroline V. Lachanski, Daniel R. Gillis, and Jennifer E. Phillips-Cremins. “YY1 and CTCF Orchestrate a 3D Chromatin Looping Switch during Early Neural Lineage Commitment.” Genome Research 27, no. 7 (July 2017): 1139–52. https://doi.org/10.1101/gr.215160.116.
18. Bhattacharya, Anandi, Chih-Yu Chen, Sara Ho, and Jennifer A. Mitchell. “Upstream Distal Regulatory Elements Contact the Lmo2 Promoter in Mouse Erythroid Cells.” PloS One 7, no. 12 (2012): e52880. https://doi.org/10.1371/journal.pone.0052880.
19. Blank-Giwojna, Alena, Anna Postepska-Igielska, and Ingrid Grummt. “LncRNA KHPS1 Activates a Poised Enhancer by Triplex-Dependent Recruitment of Epigenomic Regulators.” Cell Reports 26, no. 11 (March 12, 2019): 2904-2915.e4. https://doi.org/10.1016/j.celrep.2019.02.059.
20. Boque-Sastre, Raquel, Marta Soler, Cristina Oliveira-Mateos, Anna Portela, Catia Moutinho, Sergi Sayols, Alberto Villanueva, Manel Esteller, and Sonia Guil. “Head-to-Head Antisense Transcription and R-Loop Formation Promotes Transcriptional Activation.” Proceedings of the National Academy of Sciences of the United States of America 112, no. 18 (May 5, 2015): 5785–90. https://doi.org/10.1073/pnas.1421197112.
21. Braidotti, G., T. Baubec, F. Pauler, C. Seidl, O. Smrzka, S. Stricker, I. Yotova, and D.P. Barlow. “The Air Noncoding RNA: An Imprinted Cis-Silencing Transcript.” Cold Spring Harbor Symposia on Quantitative Biology 69, no. 0 (January 1, 2004): 55–66. https://doi.org/10.1101/sqb.2004.69.55.

22. Brannan, C. I., E. C. Dees, R. S. Ingram, and S. M. Tilghman. “The Product of the H19 Gene May Function as an RNA.” Molecular and Cellular Biology 10, no. 1 (January 1990): 28–36. https://doi.org/10.1128/mcb.10.1.28-36.1990.
23. Brockdorff, Neil. “Polycomb Complexes in X Chromosome Inactivation.” Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 372, no. 1733 (November 5, 2017): 20170021. https://doi.org/10.1098/rstb.2017.0021.
24. Brower, Jeffrey V., Chae Ho Lim, Marda Jorgensen, S. Paul Oh, and Naohiro Terada. “Adenine Nucleotide Translocase 4 Deficiency Leads to Early Meiotic Arrest of Murine Male Germ Cells.” Reproduction (Cambridge, England) 138, no. 3 (September 2009): 463–70. https://doi.org/10.1530/REP-09-0201.
25. Brown, C. J., B. D. Hendrich, J. L. Rupert, R. G. Lafrenière, Y. Xing, J. Lawrence, and H. F. Willard. “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 71, no. 3 (October 30, 1992): 527–42. https://doi.org/10.1016/0092-8674(92)90520-m.
26. Burke, Les J., Ru Zhang, Marek Bartkuhn, Vijay K. Tiwari, Gholamreza Tavoosidana, Sreenivasulu Kurukuti, Christine Weth, et al. “CTCF Binding and Higher Order Chromatin Structure of the H19 Locus Are Maintained in Mitotic Chromatin.” The EMBO Journal 24, no. 18 (September 21, 2005): 3291–3300. https://doi.org/10.1038/sj.emboj.7600793.
27. Buske, Fabian A., Denis C. Bauer, John S. Mattick, and Timothy L. Bailey. “Triplexator: Detecting Nucleic Acid Triple Helices in Genomic and Transcriptomic Data.” Genome Research 22, no. 7 (July 2012): 1372–81. https://doi.org/10.1101/gr.130237.111.
28. Carlevaro-Fita, Joana, Anisa Rahim, Roderic Guigó, Leah A. Vardy, and Rory Johnson. “Cytoplasmic Long Noncoding RNAs Are Frequently Bound to and Degraded at Ribosomes in Human Cells.” RNA (New York, N.Y.) 22, no. 6 (June 2016): 867–82. https://doi.org/10.1261/rna.053561.115.
29. Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, N. Maeda, R. Oyama, et al. “The Transcriptional Landscape of the Mammalian Genome.” Science (New York, N.Y.) 309, no. 5740 (September 2, 2005): 1559–63. https://doi.org/10.1126/science.1112014.
30. Carthew, Richard W. “Gene Regulation by MicroRNAs.” Current Opinion in Genetics & Development 16, no. 2 (April 2006): 203–8. https://doi.org/10.1016/j.gde.2006.02.012.
31. Chaboissier, Marie-Christine, Akio Kobayashi, Valerie I. P. Vidal, Susanne Lützkendorf, Henk J. G. van de Kant, Michael Wegner, Dirk G. de Rooij, Richard R. Behringer, and Andreas Schedl. “Functional Analysis of Sox8 and Sox9 during Sex Determination in the Mouse.” Development 131, no. 9 (May 1, 2004): 1891–1901. https://doi.org/10.1242/dev.01087.
32. Chatterjee, Sumantra, and Nadav Ahituv. “Gene Regulatory Elements, Major Drivers of Human Disease.” Annual Review of Genomics and Human Genetics 18 (August 31, 2017): 45–63. https://doi.org/10.1146/annurev-genom-091416-035537.
33. Cho, Seung Woo, Jin Xu, Ruping Sun, Maxwell R. Mumbach, Ava C. Carter, Y. Grace Chen, Kathryn E. Yost, et al. “Promoter of LncRNA Gene PVT1 Is a Tumor-Suppressor

DNA Boundary Element." Cell 173, no. 6 (May 31, 2018): 1398-1412.e22. https://doi.org/10.1016/j.cell.2018.03.068.
34. Choudhury, Subhendu Roy, Sangeeta Dutta, Utsa Bhaduri, and Manchanahalli R Satyanarayana Rao. "LncRNA Hmrhl Regulates Expression of Cancer Related Genes in Chronic Myelogenous Leukemia through Chromatin Association." Preprint. Cancer Biology, September 18, 2020. https://doi.org/10.1101/2020.09.17.301770.
35. Chujo, Takeshi, and Tetsuro Hirose. "Nuclear Bodies Built on Architectural Long Noncoding RNAs: Unifying Principles of Their Construction and Function." Molecules and Cells 40, no. 12 (December 31, 2017): 889–96. https://doi.org/10.14348/molcells.2017.0263.
36. Chung, C. T., S. L. Niemela, and R. H. Miller. "One-Step Preparation of Competent Escherichia Coli: Transformation and Storage of Bacterial Cells in the Same Solution." Proceedings of the National Academy of Sciences of the United States of America 86, no. 7 (April 1989): 2172–75. https://doi.org/10.1073/pnas.86.7.2172.
37. Clapier, Cedric R., and Bradley R. Cairns. "The Biology of Chromatin Remodeling Complexes." Annual Review of Biochemistry 78 (2009): 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223.
38. Clemson, C. M., J. A. McNeil, H. F. Willard, and J. B. Lawrence. "XIST RNA Paints the Inactive X Chromosome at Interphase: Evidence for a Novel RNA Involved in Nuclear/Chromosome Structure." The Journal of Cell Biology 132, no. 3 (February 1996): 259–75. https://doi.org/10.1083/jcb.132.3.259.
39. Davidovich, Chen, and Thomas R. Cech. "The Recruitment of Chromatin Modifiers by Long Noncoding RNAs: Lessons from PRC2." RNA (New York, N.Y.) 21, no. 12 (December 2015): 2007–22. https://doi.org/10.1261/rna.053918.115.
40. Davidovich, Chen, Leon Zheng, Karen J. Goodrich, and Thomas R. Cech. "Promiscuous RNA Binding by Polycomb Repressive Complex 2." Nature Structural & Molecular Biology 20, no. 11 (November 2013): 1250–57. https://doi.org/10.1038/nsmb.2679.
41. Dekker, J. "Capturing Chromosome Conformation." Science 295, no. 5558 (February 15, 2002): 1306–11. https://doi.org/10.1126/science.1067799.
42. Dekker, Job. "The Three 'C' s of Chromosome Conformation Capture: Controls, Controls, Controls." Nature Methods 3, no. 1 (January 2006): 17–21. https://doi.org/10.1038/nmeth823.
43. Delihas, Nicholas. "Discovery and Characterization of the First Non-Coding RNA That Regulates Gene Expression, MicF RNA: A Historical Perspective." World Journal of Biological Chemistry 6, no. 4 (November 26, 2015): 272–80. https://doi.org/10.4331/wjbc.v6.i4.272.
44. Derrien, Thomas, Rory Johnson, Giovanni Bussotti, Andrea Tanzer, Sarah Djebali, Hagen Tilgner, Gregory Guernec, et al. "The GENCODE v7 Catalog of Human Long Noncoding RNAs: Analysis of Their Gene Structure, Evolution, and Expression." Genome Research 22, no. 9 (September 2012): 1775–89. https://doi.org/10.1101/gr.132159.111.

45. Djebali, Sarah, Carrie A. Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, et al. “Landscape of Transcription in Human Cells.” Nature 489, no. 7414 (September 6, 2012): 101–8. https://doi.org/10.1038/nature11233.
46. Essien, Kobby, Sebastien Vigneau, Sofia Apreleva, Larry N Singh, Marisa S Bartolomei, and Sridhar Hannenhalli. “CTCF Binding Site Classes Exhibit Distinct Evolutionary, Genomic, Epigenomic and Transcriptomic Features.” Genome Biology 10, no. 11 (2009): R131. https://doi.org/10.1186/gb-2009-10-11-r131.
47. Fan, Song, Tian Tian, Xiaobin Lv, Xinyuan Lei, Zhaohui Yang, Mo Liu, Faya Liang, et al. “LncRNA CISAL Inhibits BRCA1 Transcription by Forming a Tertiary Structure at Its Promoter.” IScience 23, no. 2 (February 21, 2020): 100835. https://doi.org/10.1016/j.isci.2020.100835.
48. Fatima, Roshan, Subhendu Roy Choudhury, Divya T R, Utsa Bhaduri, and M. R. S. Rao. “A Novel Enhancer RNA, Hmrhl, Positively Regulates Its Host Gene, Phkb, in Chronic Myelogenous Leukemia.” Non-Coding RNA Research 4, no. 3 (September 2019): 96–108. https://doi.org/10.1016/j.ncrna.2019.08.001.
49. Feher, Joseph. “DNA and Protein Synthesis.” In Quantitative Human Physiology, 120–29. Elsevier, 2017. https://doi.org/10.1016/B978-0-12-800883-6.00011-2.
50. Felsenfeld, G., and A. Rich. “Studies on the Formation of Two- and Three-Stranded Polyribonucleotides.” Biochimica Et Biophysica Acta 26, no. 3 (December 1957): 457–68. https://doi.org/10.1016/0006-3002(57)90091-4.
51. Feng, Chun-Wei, Josephine Bowles, and Peter Koopman. “Control of Mammalian Germ Cell Entry into Meiosis.” Molecular and Cellular Endocrinology 382, no. 1 (January 25, 2014): 488–97. https://doi.org/10.1016/j.mce.2013.09.026.
52. Feng, Jianxing, Tao Liu, Bo Qin, Yong Zhang, and Xiaole Shirley Liu. “Identifying ChIP-Seq Enrichment Using MACS.” Nature Protocols 7, no. 9 (September 2012): 1728–40. https://doi.org/10.1038/nprot.2012.101.
53. Fudenberg, Geoffrey, Nezar Abdennur, Maxim Imakaev, Anton Goloborodko, and Leonid A. Mirny. “Emerging Evidence of Chromosome Folding by Loop Extrusion.” Cold Spring Harbor Symposia on Quantitative Biology 82 (2017): 45–55. https://doi.org/10.1101/sqb.2017.82.034710.
54. Gabory, Anne, Marie-Anne Ripoche, Anne Le Digarcher, Françoise Watrin, Ahmed Ziyyat, Thierry Forné, Hélène Jammes, et al. “H19 Acts as a Trans Regulator of the Imprinted Gene Network Controlling Growth in Mice.” Development (Cambridge, England) 136, no. 20 (October 2009): 3413–21. https://doi.org/10.1242/dev.036061.
55. Ganesan, Gayatri, and Satyanarayana M. R. Rao. “A Novel Noncoding RNA Processed by Drosha Is Restricted to Nucleus in Mouse.” RNA (New York, N.Y.) 14, no. 7 (July 2008): 1399–1410. https://doi.org/10.1261/rna.838308.
56. Garcia-Moreno, S. Alexandra, Christopher R. Futtner, Isabella M. Salamone, Nitzan Gonen, Robin Lovell-Badge, and Danielle M. Maatouk. “Gonadal Supporting Cells Acquire Sex-Specific Chromatin Landscapes during Mammalian Sex Determination.” Developmental Biology 446, no. 2 (February 15, 2019): 168–79. https://doi.org/10.1016/j.ydbio.2018.12.023.

57. Gibbons, Hunter R., Guzel Shaginurova, Laura C. Kim, Nathaniel Chapman, Charles F. Spurlock, and Thomas M. Aune. “Divergent LncRNA GATA3-AS1 Regulates GATA3 Transcription in T-Helper 2 Cells.” Frontiers in Immunology 9 (2018): 2512. https://doi.org/10.3389/fimmu.2018.02512.
58. Giles, K. E., H. Gowher, R. Ghirlando, C. Jin, and G. Felsenfeld. “Chromatin Boundaries, Insulators, and Long-Range Interactions in the Nucleus.” Cold Spring Harbor Symposia on Quantitative Biology 75 (2010): 79–85. https://doi.org/10.1101/sqb.2010.75.006.
59. Ginno, Paul A., Paul L. Lott, Holly C. Christensen, Ian Korf, and Frédéric Chédin. “R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters.” Molecular Cell 45, no. 6 (March 30, 2012): 814–25. https://doi.org/10.1016/j.molcel.2012.01.017.
60. Gomez, J. Antonio, Orly L. Wapinski, Yul W. Yang, Jean-François Bureau, Smita Gopinath, Denise M. Monack, Howard Y. Chang, Michel Brahic, and Karla Kirkegaard. “The NeST Long NcRNA Controls Microbial Susceptibility and Epigenetic Activation of the Interferon-γ Locus.” Cell 152, no. 4 (February 14, 2013): 743–54. https://doi.org/10.1016/j.cell.2013.01.015.
61. Gonen, Nitzan, Chris R. Futtner, Sophie Wood, S. Alexandra Garcia-Moreno, Isabella M. Salamone, Shiela C. Samson, Ryohei Sekido, Francis Poulat, Danielle M. Maatouk, and Robin Lovell-Badge. “Sex Reversal Following Deletion of a Single Distal Enhancer of Sox9.” Science (New York, N.Y.) 360, no. 6396 (June 29, 2018): 1469–73. https://doi.org/10.1126/science.aas9408.
62. Gonen, Nitzan, Alexander Quinn, Helen C. O’Neill, Peter Koopman, and Robin Lovell-Badge. “Normal Levels of Sox9 Expression in the Developing Mouse Testis Depend on the TES/TESCO Enhancer, but This Does Not Act Alone.” PLoS Genetics 13, no. 1 (January 2017): e1006520. https://doi.org/10.1371/journal.pgen.1006520.
63. Grandori, Carla, Shaun M. Cowley, Leonard P. James, and Robert N. Eisenman. “The Myc/Max/Mad Network and the Transcriptional Control of Cell Behavior.” Annual Review of Cell and Developmental Biology 16, no. 1 (November 2000): 653–99. https://doi.org/10.1146/annurev.cellbio.16.1.653.
64. Guenzl, Philipp M., and Denise P. Barlow. “Macro LncRNAs: A New Layer of Cis-Regulatory Information in the Mammalian Genome.” RNA Biology 9, no. 6 (June 2012): 731–41. https://doi.org/10.4161/rna.19985.
65. Guh, Chia-Yu, Yu-Hung Hsieh, and Hsueh-Ping Chu. “Functions and Properties of Nuclear LncRNAs-from Systematically Mapping the Interactomes of LncRNAs.” Journal of Biomedical Science 27, no. 1 (March 17, 2020): 44. https://doi.org/10.1186/s12929-020-00640-3.
66. Guillou, Emmanuelle, Arkaitz Ibarra, Vincent Coulon, Juan Casado-Vela, Daniel Rico, Ignacio Casal, Etienne Schwob, Ana Losada, and Juan Méndez. “Cohesin Organizes Chromatin Loops at DNA Replication Factories.” Genes & Development 24, no. 24 (December 15, 2010): 2812–22. https://doi.org/10.1101/gad.608210.
67. Guth, Sabine I. E., Michael R. Bösl, Elisabeth Sock, and Michael Wegner. “Evolutionary Conserved Sequence Elements with Embryonic Enhancer Activity in the

Vicinity of the Mammalian Sox8 Gene." The International Journal of Biochemistry & Cell Biology 42, no. 3 (March 2010): 465–71. https://doi.org/10.1016/j.biocel.2009.07.008.
68. Guttman, Mitchell, Ido Amit, Manuel Garber, Courtney French, Michael F. Lin, David Feldser, Maite Huarte, et al. "Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals." Nature 458, no. 7235 (March 12, 2009): 223–27. https://doi.org/10.1038/nature07672.
69. Hacisuleyman, Ezgi, Chinmay J. Shukla, Catherine L. Weiner, and John L. Rinn. "Function and Evolution of Local Repeats in the Firre Locus." Nature Communications 7 (March 24, 2016): 11021. https://doi.org/10.1038/ncomms11021.
70. Handoko, Lusy, Han Xu, Guoliang Li, Chew Yee Ngan, Elaine Chew, Marie Schnapp, Charlie Wah Heng Lee, et al. "CTCF-Mediated Functional Chromatin Interactome in Pluripotent Cells." Nature Genetics 43, no. 7 (June 19, 2011): 630–38. https://doi.org/10.1038/ng.857.
71. Herold, Martin, Marek Bartkuhn, and Rainer Renkawitz. "CTCF: Insights into Insulator Function during Development." Development (Cambridge, England) 139, no. 6 (March 2012): 1045–57. https://doi.org/10.1242/dev.065268.
72. Hildebrand, Erica M., and Job Dekker. "Mechanisms and Functions of Chromosome Compartmentalization." Trends in Biochemical Sciences 45, no. 5 (May 2020): 385–96. https://doi.org/10.1016/j.tibs.2020.01.002.
73. Hoffmeister, Helen, Andreas Fuchs, Fabian Erdel, Sophia Pinz, Regina Gröbner-Ferreira, Astrid Bruckmann, Rainer Deutzmann, et al. "CHD3 and CHD4 Form Distinct NuRD Complexes with Different yet Overlapping Functionality." Nucleic Acids Research 45, no. 18 (October 13, 2017): 10534–54. https://doi.org/10.1093/nar/gkx711.
74. Hondele, Maria, Ruchika Sachdev, Stephanie Heinrich, Juan Wang, Pascal Vallotton, Beatriz M. A. Fontoura, and Karsten Weis. "DEAD-Box ATPases Are Global Regulators of Phase-Separated Organelles." Nature 573, no. 7772 (September 2019): 144–48. https://doi.org/10.1038/s41586-019-1502-y.
75. Horsfield, Julia A., Cristin G. Print, and Maren Mönnich. "Diverse Developmental Disorders from the One Ring: Distinct Molecular Pathways Underlie the Cohesinopathies." Frontiers in Genetics 3 (2012): 171. https://doi.org/10.3389/fgene.2012.00171.
76. Hu, Jiali, Ke Li, Zhanghuan Li, Chao Gao, Fei Guo, Yingmei Wang, and Fengxia Xue. "Sex-Determining Region Y Box-Containing Genes: Regulators and Biomarkers in Gynecological Cancers." Cancer Biology & Medicine 16, no. 3 (August 2019): 462–74. https://doi.org/10.20892/j.issn.2095-3941.2019.0062.
77. Huang, Di, Hanna M. Petrykowska, Brendan F. Miller, Laura Elnitski, and Ivan Ovcharenko. "Identification of Human Silencers by Correlating Cross-Tissue Epigenetic Profiles and Gene Expression." Genome Research 29, no. 4 (April 2019): 657–67. https://doi.org/10.1101/gr.247007.118.
78. Huang, Jin-Zhou, Min Chen, De Chen, Xing-Cheng Gao, Song Zhu, Hongyang Huang, Min Hu, Huifang Zhu, and Guang-Rong Yan. "A Peptide Encoded by a Putative

LncRNA HOXB-AS3 Suppresses Colon Cancer Growth." Molecular Cell 68, no. 1 (October 5, 2017): 171-184.e6. https://doi.org/10.1016/j.molcel.2017.09.015.
79. Huang, Mingyan, Huamin Wang, Xiang Hu, and Xuetao Cao. "LncRNA MALAT1 Binds Chromatin Remodeling Subunit BRG1 to Epigenetically Promote Inflammation-Related Hepatocellular Carcinoma Progression." Oncoimmunology 8, no. 1 (2019): e1518628. https://doi.org/10.1080/2162402X.2018.1518628.
80. Isoda, Takeshi, Amanda J. Moore, Zhaoren He, Vivek Chandra, Masatoshi Aida, Matthew Denholtz, Jan Piet van Hamburg, et al. "Non-Coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate." Cell 171, no. 1 (September 21, 2017): 103-119.e18. https://doi.org/10.1016/j.cell.2017.09.001.
81. Jabbari, Kamel, Peter Heger, Ranu Sharma, and Thomas Wiehe. "The Diverging Routes of BORIS and CTCF: An Interactomic and Phylogenomic Analysis." Life (Basel, Switzerland) 8, no. 1 (January 30, 2018): E4. https://doi.org/10.3390/life8010004.
82. Jarroux, Julien, Antonin Morillon, and Marina Pinskaya. "History, Discovery, and Classification of LncRNAs." Advances in Experimental Medicine and Biology 1008 (2017): 1–46. https://doi.org/10.1007/978-981-10-5203-3_1.
83. Kallen, Amanda N., Xiao-Bo Zhou, Jie Xu, Chong Qiao, Jing Ma, Lei Yan, Lingeng Lu, et al. "The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs." Molecular Cell 52, no. 1 (October 10, 2013): 101–12. https://doi.org/10.1016/j.molcel.2013.08.027.
84. Kalwa, Marie, Sonja Hänzelmann, Sabrina Otto, Chao-Chung Kuo, Julia Franzen, Sylvia Joussen, Eduardo Fernandez-Rebollo, et al. "The LncRNA HOTAIR Impacts on Mesenchymal Stem Cells via Triple Helix Formation." Nucleic Acids Research 44, no. 22 (December 15, 2016): 10631–43. https://doi.org/10.1093/nar/gkw802.
85. Kataruka, Shubhangini, Vijay Suresh Akhade, Bhavana Kayyar, and Manchanahalli R. Satyanarayana Rao. "Mrhl Long Noncoding RNA Mediates Meiotic Commitment of Mouse Spermatogonial Cells by Regulating Sox8 Expression." Molecular and Cellular Biology 37, no. 14 (July 15, 2017). https://doi.org/10.1128/MCB.00632-16.
86. Kawaguchi, Tetsuya, Akie Tanigawa, Takao Naganuma, Yasuyuki Ohkawa, Sylvie Souquere, Gerard Pierron, and Tetsuro Hirose. "SWI/SNF Chromatin-Remodeling Complexes Function in Noncoding RNA-Dependent Assembly of Nuclear Bodies." Proceedings of the National Academy of Sciences of the United States of America 112, no. 14 (April 7, 2015): 4304–9. https://doi.org/10.1073/pnas.1423819112.
87. Kentepozidou, Elissavet, Sarah J. Aitken, Christine Feig, Klara Stefflova, Ximena Ibarra-Soria, Duncan T. Odom, Maša Roller, and Paul Flicek. "Clustered CTCF Binding Is an Evolutionary Mechanism to Maintain Topologically Associating Domains." Genome Biology 21, no. 1 (December 2020): 5. https://doi.org/10.1186/s13059-019-1894-x.
88. Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L Salzberg. "TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions." Genome Biology 14, no. 4 (2013): R36. https://doi.org/10.1186/gb-2013-14-4-r36.

89. Kim, Somi, Nam-Kyung Yu, and Bong-Kiun Kaang. “CTCF as a Multifunctional Protein in Genome Regulation and Gene Expression.” Experimental & Molecular Medicine 47, no. 6 (June 2015): e166–e166. https://doi.org/10.1038/emm.2015.33.
90. Kino, Tomoshige, Darrell E. Hurt, Takamasa Ichijo, Nancy Nader, and George P. Chrousos. “Noncoding RNA Gas5 Is a Growth Arrest- and Starvation-Associated Repressor of the Glucocorticoid Receptor.” Science Signaling 3, no. 107 (February 2, 2010): ra8. https://doi.org/10.1126/scisignal.2000568.
91. Kopp, Florian, and Joshua T. Mendell. “Functional Classification and Experimental Dissection of Long Noncoding RNAs.” Cell 172, no. 3 (January 25, 2018): 393–407. https://doi.org/10.1016/j.cell.2018.01.011.
92. Kotake, Y., T. Nakagawa, K. Kitagawa, S. Suzuki, N. Liu, M. Kitagawa, and Y. Xiong. “Long Non-Coding RNA ANRIL Is Required for the PRC2 Recruitment to and Silencing of P15(INK4B) Tumor Suppressor Gene.” Oncogene 30, no. 16 (April 21, 2011): 1956–62. https://doi.org/10.1038/onc.2010.568.
93. Kretz, Markus, Zurab Siprashvili, Ci Chu, Dan E. Webster, Ashley Zehnder, Kun Qu, Carolyn S. Lee, et al. “Control of Somatic Tissue Differentiation by the Long Non-Coding RNA TINCR.” Nature 493, no. 7431 (January 10, 2013): 231–35. https://doi.org/10.1038/nature11661.
94. Krijger, Peter H. L., Geert Geeven, Valerio Bianchi, Catharina R. E. Hilvering, and Wouter de Laat. “4C-Seq from Beginning to End: A Detailed Protocol for Sample Preparation and Data Analysis.” Methods (San Diego, Calif.) 170 (January 1, 2020): 17–32. https://doi.org/10.1016/j.ymeth.2019.07.014.
95. Kung, Johnny T., Barry Kesner, Jee Young An, Janice Y. Ahn, Catherine Cifuentes-Rojas, David Colognori, Yesu Jeon, et al. “Locus-Specific Targeting to the X Chromosome Revealed by the RNA Interactome of CTCF.” Molecular Cell 57, no. 2 (January 22, 2015): 361–75. https://doi.org/10.1016/j.molcel.2014.12.006.
96. Langmead, Ben, and Steven L Salzberg. “Fast Gapped-Read Alignment with Bowtie 2.” Nature Methods 9, no. 4 (April 2012): 357–59. https://doi.org/10.1038/nmeth.1923.
97. Latos, Paulina A., Florian M. Pauler, Martha V. Koerner, H. Başak Şenergin, Quanah J. Hudson, Roman R. Stocsits, Wolfgang Allhoff, et al. “Airn Transcriptional Overlap, but Not Its LncRNA Products, Induces Imprinted Igf2r Silencing.” Science (New York, N.Y.) 338, no. 6113 (December 14, 2012): 1469–72. https://doi.org/10.1126/science.1228110.
98. Lefebvre, Véronique. “Roles and Regulation of SOX Transcription Factors in Skeletogenesis.” Current Topics in Developmental Biology 133 (2019): 171–93. https://doi.org/10.1016/bs.ctdb.2019.01.007.
99. Lefebvre, Véronique, Bogdan Dumitriu, Alfredo Penzo-Méndez, Yu Han, and Bhattaram Pallavi. “Control of Cell Fate and Differentiation by Sry-Related High-Mobility-Group Box (Sox) Transcription Factors.” The International Journal of Biochemistry & Cell Biology 39, no. 12 (2007): 2195–2214. https://doi.org/10.1016/j.biocel.2007.05.019.
100. Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing

Subgroup. “The Sequence Alignment/Map Format and SAMtools.” Bioinformatics (Oxford, England) 25, no. 16 (August 15, 2009): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

101. Li, Yue, Junetha Syed, and Hiroshi Sugiyama. “RNA-DNA Triplex Formation by Long Noncoding RNAs.” Cell Chemical Biology 23, no. 11 (November 17, 2016): 1325–33. https://doi.org/10.1016/j.chembiol.2016.09.011.
102. Liang, Junnan, Jingyuan Wen, Zhao Huang, Xiao-Ping Chen, Bi-Xiang Zhang, and Liang Chu. “Small Nucleolar RNAs: Insight Into Their Function in Cancer.” Frontiers in Oncology 9 (2019): 587. https://doi.org/10.3389/fonc.2019.00587.
103. Liu, Bodu, Lijuan Sun, Qiang Liu, Chang Gong, Yandan Yao, Xiaobin Lv, Ling Lin, et al. “A Cytoplasmic NF-KB Interacting Long Noncoding RNA Blocks IκB Phosphorylation and Suppresses Breast Cancer Metastasis.” Cancer Cell 27, no. 3 (March 9, 2015): 370–81. https://doi.org/10.1016/j.ccell.2015.02.004.
104. Loewer, Sabine, Moran N. Cabili, Mitchell Guttman, Yuin-Han Loh, Kelly Thomas, In Hyun Park, Manuel Garber, et al. “Large Intergenic Non-Coding RNA-RoR Modulates Reprogramming of Human Induced Pluripotent Stem Cells.” Nature Genetics 42, no. 12 (December 2010): 1113–17. https://doi.org/10.1038/ng.710.
105. Logan, Catriona Y., and Roel Nusse. “The Wnt Signaling Pathway in Development and Disease.” Annual Review of Cell and Developmental Biology 20 (2004): 781–810. https://doi.org/10.1146/annurev.cellbio.20.010403.113126.
106. Lubelsky, Yoav, and Igor Ulitsky. “Sequences Enriched in Alu Repeats Drive Nuclear Localization of Long RNAs in Human Cells.” Nature 555, no. 7694 (March 1, 2018): 107–11. https://doi.org/10.1038/nature25757.
107. Lukassen, S., E. Bosch, A. B. Ekici, and A. Winterpacht. “Characterization of Germ Cell Differentiation in the Male Mouse through Single-Cell RNA Sequencing.” Scientific Reports 8, no. 1 (April 25, 2018): 6521. https://doi.org/10.1038/s41598-018-24725-0.
108. Luo, Huacheng, Ganqian Zhu, Tsz Kan Fung, Yi Qiu, Mingjiang Xu, Chi Wai Eric So, and Suming Huang. “Hottip Lncrna Reinforces CTCF Defined Chromatin Boundaries and Drives Wnt Target Gene Expression in AML Leukemogenesis.” Blood 134, no. Supplement_1 (November 13, 2019): 277–277. https://doi.org/10.1182/blood-2019-127323.
109. Luo, Huacheng, Ganqian Zhu, Jianfeng Xu, Qian Lai, Bowen Yan, Ying Guo, Tsz Kan Fung, et al. “HOTTIP LncRNA Promotes Hematopoietic Stem Cell Self-Renewal Leading to AML-like Disease in Mice.” Cancer Cell 36, no. 6 (December 9, 2019): 645-659.e8. https://doi.org/10.1016/j.ccell.2019.10.011.
110. Ma, Lina, Vladimir B. Bajic, and Zhang Zhang. “On the Classification of Long Non-Coding RNAs.” RNA Biology 10, no. 6 (June 2013): 925–33. https://doi.org/10.4161/rna.24604.
111. Makarewich, Catherine A., Amir Z. Munir, Gabriele G. Schiattarella, Svetlana Bezprozvannaya, Olga N. Raguimova, Ellen E. Cho, Alexander H. Vidal, Seth L. Robia, Rhonda Bassel-Duby, and Eric N. Olson. “The DWORF Micropeptide Enhances Contractility and Prevents Heart Failure in a Mouse Model of Dilated Cardiomyopathy.” ELife 7 (October 9, 2018): e38319. https://doi.org/10.7554/eLife.38319.

112.Margueron, Raphaël, and Danny Reinberg. “The Polycomb Complex PRC2 and Its Mark in Life.” Nature 469, no. 7330 (January 20, 2011): 343–49. https://doi.org/10.1038/nature09784.
113.Matera, A. Gregory, Rebecca M. Terns, and Michael P. Terns. “Non-Coding RNAs: Lessons from the Small Nuclear and Small Nucleolar RNAs.” Nature Reviews. Molecular Cell Biology 8, no. 3 (March 2007): 209–20. https://doi.org/10.1038/nrm2124.
114.Mead, Timothy J., Qiuqing Wang, Pallavi Bhattaram, Peter Dy, Solomon Afelik, Jan Jensen, and Véronique Lefebvre. “A Far-Upstream (-70 Kb) Enhancer Mediates Sox9 Auto-Regulation in Somatic Tissues during Development and Adult Regeneration.” Nucleic Acids Research 41, no. 8 (April 2013): 4459–69. https://doi.org/10.1093/nar/gkt140.
115.Mercer, Tim R., Marcel E. Dinger, and John S. Mattick. “Long Non-Coding RNAs: Insights into Functions.” Nature Reviews. Genetics 10, no. 3 (March 2009): 155–59. https://doi.org/10.1038/nrg2521.
116.Mondal, Tanmoy, Santhilal Subhash, Roshan Vaid, Stefan Enroth, Sireesha Uday, Björn Reinius, Sanhita Mitra, et al. “MEG3 Long Noncoding RNA Regulates the TGF-β Pathway Genes through Formation of RNA-DNA Triplex Structures.” Nature Communications 6 (July 24, 2015): 7743. https://doi.org/10.1038/ncomms8743.
117.Mumbach, Maxwell R., Adam J. Rubin, Ryan A. Flynn, Chao Dai, Paul A. Khavari, William J. Greenleaf, and Howard Y. Chang. “HiChIP: Efficient and Sensitive Analysis of Protein-Directed Genome Architecture.” Nature Methods 13, no. 11 (November 2016): 919–22. https://doi.org/10.1038/nmeth.3999.
118.Nakagawa, Shinichi, Tomohiro Yamazaki, and Tetsuro Hirose. “Molecular Dissection of Nuclear Paraspeckles: Towards Understanding the Emerging World of the RNP Milieu.” Open Biology 8, no. 10 (October 24, 2018): 180150. https://doi.org/10.1098/rsob.180150.
119.Naumova, Natalia, Emily M. Smith, Ye Zhan, and Job Dekker. “Analysis of Long-Range Chromatin Interactions Using Chromosome Conformation Capture.” Methods (San Diego, Calif.) 58, no. 3 (November 2012): 192–203. https://doi.org/10.1016/j.ymeth.2012.07.022.
120.Nazar, Ross N. “Ribosomal RNA Processing and Ribosome Biogenesis in Eukaryotes.” IUBMB Life 56, no. 8 (August 2004): 457–65. https://doi.org/10.1080/15216540400010867.
121.Necochea-Campion, Rosalia de, Anahit Ghochikyan, Steven F. Josephs, Shelly Zacharias, Erik Woods, Feridoun Karimi-Busheri, Doru T. Alexandrescu, Chien-Shing Chen, Michael G. Agadjanyan, and Ewa Carrier. “Expression of the Epigenetic Factor BORIS (CTCFL) in the Human Genome.” Journal of Translational Medicine 9 (December 14, 2011): 213. https://doi.org/10.1186/1479-5876-9-213.
122.Nickerson, J. A., G. Krochmalnic, K. M. Wan, and S. Penman. “Chromatin Architecture and Nuclear RNA.” Proceedings of the National Academy of Sciences of the United States of America 86, no. 1 (January 1989): 177–81. https://doi.org/10.1073/pnas.86.1.177.

123.Nigris, F de, C Botti, R Rossiello, E Crimi, V Sica, and C Napoli. “Cooperation between Myc and YY1 Provides Novel Silencing Transcriptional Targets of A3β1-Integrin in Tumour Cells.” Oncogene 26, no. 3 (January 2007): 382–94. https://doi.org/10.1038/sj.onc.1209804.
124.Nishana, Mayilaadumveettil, Caryn Ha, Javier Rodriguez-Hernaez, Ali Ranjbaran, Erica Chio, Elphege P. Nora, Sana B. Badri, et al. “Defining the Relative and Combined Contribution of CTCF and CTCFL to Genomic Regulation.” Genome Biology 21, no. 1 (December 2020): 108. https://doi.org/10.1186/s13059-020-02024-0.
125.Nishant, K. T., H. Ravishankar, and M. R. S. Rao. “Characterization of a Mouse Recombination Hot Spot Locus Encoding a Novel Non-Protein-Coding RNA.” Molecular and Cellular Biology 24, no. 12 (June 2004): 5620–34. https://doi.org/10.1128/MCB.24.12.5620-5634.2004.
126.Nishiyama, Tomoko. “Cohesion and Cohesin-Dependent Chromatin Organization.” Current Opinion in Cell Biology 58 (June 2019): 8–14. https://doi.org/10.1016/j.ceb.2018.11.006.
127.Noh, Ji Heon, Kyoung Mi Kim, Waverly G. McClusky, Kotb Abdelmohsen, and Myriam Gorospe. “Cytoplasmic Functions of Long Noncoding RNAs.” Wiley Interdisciplinary Reviews. RNA 9, no. 3 (May 2018): e1471. https://doi.org/10.1002/wrna.1471.
128.O’Bryan, Moira K., Shuji Takada, Claire L. Kennedy, Greg Scott, Shun-ichi Harada, Manas K. Ray, Qunsheng Dai, et al. “Sox8 Is a Critical Regulator of Adult Sertoli Cell Function and Male Fertility.” Developmental Biology 316, no. 2 (April 15, 2008): 359–70. https://doi.org/10.1016/j.ydbio.2008.01.042.
129.O’Donnell, Kathryn A., Erik A. Wentzel, Karen I. Zeller, Chi V. Dang, and Joshua T. Mendell. “C-Myc-Regulated MicroRNAs Modulate E2F1 Expression.” Nature 435, no. 7043 (June 9, 2005): 839–43. https://doi.org/10.1038/nature03677.
130.Ogbourne, S., and T. M. Antalis. “Transcriptional Control and the Role of Silencers in Transcriptional Regulation in Eukaryotes.” The Biochemical Journal 331 (Pt 1) (April 1, 1998): 1–14. https://doi.org/10.1042/bj3310001.
131.Ohno, Mizuki, Tatsuo Fukagawa, Jeremy S. Lee, and Toshimichi Ikemura. “Triplex-Forming DNAs in the Human Interphase Nucleus Visualized in Situ by Polypurine/Polypyrimidine DNA Probes and Antitriplex Antibodies.” Chromosoma 111, no. 3 (September 2002): 201–13. https://doi.org/10.1007/s00412-002-0198-0.
132.Okazaki, Y., M. Furuno, T. Kasukawa, J. Adachi, H. Bono, S. Kondo, I. Nikaido, et al. “Analysis of the Mouse Transcriptome Based on Functional Annotation of 60,770 Full-Length CDNAs.” Nature 420, no. 6915 (December 5, 2002): 563–73. https://doi.org/10.1038/nature01266.
133.O’Leary, Valerie Bríd, Saak Victor Ovsepian, Laura Garcia Carrascosa, Fabian Andreas Buske, Vanja Radulovic, Maximilian Niyazi, Simone Moertl, Matt Trau, Michael John Atkinson, and Nataša Anastasov. “PARTICLE, a Triplex-Forming Long NcRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation.” Cell Reports 11, no. 3 (April 21, 2015): 474–85. https://doi.org/10.1016/j.celrep.2015.03.043.

134.Ong, Chin-Tong, and Victor G. Corces. “CTCF: An Architectural Protein Bridging Genome Topology and Function.” Nature Reviews. Genetics 15, no. 4 (April 2014): 234–46. https://doi.org/10.1038/nrg3663.
135.Ong, Chin-Tong, and Victor G. Corces. “Enhancer Function: New Insights into the Regulation of Tissue-Specific Gene Expression.” Nature Reviews Genetics 12, no. 4 (April 2011): 283–93. https://doi.org/10.1038/nrg2957.
136.Paauw, N. D., A. T. Lely, J. A. Joles, A. Franx, P. G. Nikkels, M. Mokry, and B. B. van Rijn. “H3K27 Acetylation and Gene Expression Analysis Reveals Differences in Placental Chromatin Activity in Fetal Growth Restriction.” Clinical Epigenetics 10 (2018): 85. https://doi.org/10.1186/s13148-018-0508-x.
137.Pal, Debosree, C. V. Neha, Utsa Bhaduri, Zenia Zenia, Sangeeta Dutta, Subbulakshmi Chidambaram, and M. R. S. Rao. “LncRNA Mrhl Orchestrates Differentiation Programs in Mouse Embryonic Stem Cells through Chromatin Mediated Regulation.” Stem Cell Research 53 (May 2021): 102250. https://doi.org/10.1016/j.scr.2021.102250.
138.Patty, Benjamin J., and Sarah J. Hainer. “Non-Coding RNAs and Nucleosome Remodeling Complexes: An Intricate Regulatory Relationship.” Biology 9, no. 8 (August 7, 2020): E213. https://doi.org/10.3390/biology9080213.
139.Pentland, Ieisha, Karen Campos-León, Marius Cotic, Kelli-Jo Davies, C. David Wood, Ian J. Groves, Megan Burley, et al. “Disruption of CTCF-YY1–Dependent Looping of the Human Papillomavirus Genome Activates Differentiation-Induced Viral Oncogene Transcription.” Edited by Bill Sugden. PLOS Biology 16, no. 10 (October 25, 2018): e2005752. https://doi.org/10.1371/journal.pbio.2005752.
140.Petrykowska, Hanna M., Christopher M. Vockley, and Laura Elnitski. “Detection and Characterization of Silencers and Enhancer-Blockers in the Greater CFTR Locus.” Genome Research 18, no. 8 (August 2008): 1238–46. https://doi.org/10.1101/gr.073817.107.
141.Phillips, Jennifer E., and Victor G. Corces. “CTCF: Master Weaver of the Genome.” Cell 137, no. 7 (June 26, 2009): 1194–1211. https://doi.org/10.1016/j.cell.2009.06.001.
142.Porro, Antonio, Sascha Feuerhahn, Patrick Reichenbach, and Joachim Lingner. “Molecular Dissection of Telomeric Repeat-Containing RNA Biogenesis Unveils the Presence of Distinct and Multiple Regulatory Pathways.” Molecular and Cellular Biology 30, no. 20 (October 2010): 4808–17. https://doi.org/10.1128/MCB.00460-10.
143.Portnoi, Marie-France, Marie-Charlotte Dumargne, Sandra Rojo, Selma F Witchel, Andrew J Duncan, Caroline Eozenou, Joelle Bignon-Topalovic, et al. “Mutations Involving the SRY-Related Gene SOX8 Are Associated with a Spectrum of Human Reproductive Anomalies.” Human Molecular Genetics 27, no. 7 (April 1, 2018): 1228–40. https://doi.org/10.1093/hmg/ddy037.
144.Pugacheva, Elena M., Naoki Kubo, Dmitri Loukinov, Md Tajmul, Sungyun Kang, Alexander L. Kovalchuk, Alexander V. Strunnikov, Gabriel E. Zentner, Bing Ren, and Victor V. Lobanenkov. “CTCF Mediates Chromatin Looping via N-Terminal Domain-Dependent Cohesin Retention.” Proceedings of the National Academy of Sciences 117, no. 4 (January 28, 2020): 2020–31. https://doi.org/10.1073/pnas.1911708117.

145.Pugacheva, Elena M., Samuel Rivero-Hinojosa, Celso A. Espinoza, Claudia Fabiola Méndez-Catalá, Sungyun Kang, Teruhiko Suzuki, Natsuki Kosaka-Suzuki, et al. “Comparative Analyses of CTCF and BORIS Occupancies Uncover Two Distinct Classes of CTCF Binding Genomic Regions.” Genome Biology 16 (August 14, 2015): 161. https://doi.org/10.1186/s13059-015-0736-8.
146.Radhakrishnan, I., and D. J. Patel. “Solution Structure of a Pyrimidine.Purine.Pyrimidine DNA Triplex Containing T.AT, C+.GC and G.TA Triples.” Structure (London, England: 1993) 2, no. 1 (January 15, 1994): 17–32. https://doi.org/10.1016/s0969-2126(00)00005-8.
147.Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. “A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.” Cell 159, no. 7 (December 18, 2014): 1665–80. https://doi.org/10.1016/j.cell.2014.11.021.
148.Redon, Sophie, Patrick Reichenbach, and Joachim Lingner. “The Non-Coding RNA TERRA Is a Natural Ligand and Direct Inhibitor of Human Telomerase.” Nucleic Acids Research 38, no. 17 (September 2010): 5797–5806. https://doi.org/10.1093/nar/gkq296.
149.Richardson, Nainoa, Isabelle Gillot, Elodie P. Gregoire, Sameh A. Youssef, Dirk de Rooij, Alain de Bruin, Marie-Cécile De Cian, and Marie-Christine Chaboissier. “Sox8 and Sox9 Act Redundantly for Ovarian-to-Testicular Fate Reprogramming in the Absence of R-Spondin1 in Mouse Sex Reversals.” ELife 9 (May 26, 2020): e53972. https://doi.org/10.7554/eLife.53972.
150.Rinn, John L., Michael Kertesz, Jordon K. Wang, Sharon L. Squazzo, Xiao Xu, Samantha A. Brugmann, L. Henry Goodnough, et al. “Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs.” Cell 129, no. 7 (June 29, 2007): 1311–23. https://doi.org/10.1016/j.cell.2007.05.022.
151.Rion, Nathalie, and Markus A. Rüegg. “LncRNA-Encoded Peptides: More than Translational Noise?” Cell Research 27, no. 5 (May 2017): 604–5. https://doi.org/10.1038/cr.2017.35.
152.Robinson, James T, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov. “Integrative Genomics Viewer.” Nature Biotechnology 29, no. 1 (January 2011): 24–26. https://doi.org/10.1038/nbt.1754.
153.Rothschild, Gerson, Wanwei Zhang, Junghyun Lim, Pankaj Kumar Giri, Brice Laffleur, Yiyun Chen, Mingyan Fang, et al. “Noncoding RNA Transcription Alters Chromosomal Topology to Promote Isotype-Specific Class Switch Recombination.” Science Immunology 5, no. 44 (February 7, 2020): eaay5864. https://doi.org/10.1126/sciimmunol.aay5864.
154.Rowley, M. Jordan, and Victor G. Corces. “Organizational Principles of 3D Genome Architecture.” Nature Reviews. Genetics 19, no. 12 (December 2018): 789–800. https://doi.org/10.1038/s41576-018-0060-8.
155.Ruiz-Orera, Jorge, Xavier Messeguer, Juan Antonio Subirana, and M. Mar Alba. “Long Non-Coding RNAs as a Source of New Peptides.” ELife 3 (September 16, 2014): e03523. https://doi.org/10.7554/eLife.03523.

156. Saldaña-Meyer, Ricardo, Javier Rodriguez-Hernaez, Thelma Escobar, Mayilaadumveettil Nishana, Karina Jácome-López, Elphege P. Nora, Benoit G. Bruneau, et al. “RNA Interactions Are Essential for CTCF-Mediated Genome Organization.” Molecular Cell 76, no. 3 (November 7, 2019): 412-422.e5. https://doi.org/10.1016/j.molcel.2019.08.015.
157. Sambrook, Joseph, and David W. Russell. “SDS-Polyacrylamide Gel Electrophoresis of Proteins.” CSH Protocols 2006, no. 4 (September 1, 2006): pdb.prot4540. https://doi.org/10.1101/pdb.prot4540.
158. Schepers, Goslik, Megan Wilson, Dagmar Wilhelm, and Peter Koopman. “SOX8 Is Expressed during Testis Differentiation in Mice and Synergizes with SF1 to Activate the Amh Promoter in Vitro.” The Journal of Biological Chemistry 278, no. 30 (July 25, 2003): 28101–8. https://doi.org/10.1074/jbc.M304067200.
159. Schmitz, Kerstin-Maike, Christine Mayer, Anna Postepska, and Ingrid Grummt. “Interaction of Noncoding RNA with the RDNA Promoter Mediates Recruitment of DNMT3b and Silencing of RRNA Genes.” Genes & Development 24, no. 20 (October 15, 2010): 2264–69. https://doi.org/10.1101/gad.590910.
160. Schneider, Caroline A., Wayne S. Rasband, and Kevin W. Eliceiri. “NIH Image to ImageJ: 25 Years of Image Analysis.” Nature Methods 9, no. 7 (July 2012): 671–75. https://doi.org/10.1038/nmeth.2089.
161. Schock, Elizabeth N., and Carole LaBonne. “Sorting Sox: Diverse Roles for Sox Transcription Factors During Neural Crest and Craniofacial Development.” Frontiers in Physiology 11 (2020): 606889. https://doi.org/10.3389/fphys.2020.606889.
162. Schoenfelder, Stefan, and Peter Fraser. “Long-Range Enhancer-Promoter Contacts in Gene Expression Control.” Nature Reviews. Genetics 20, no. 8 (August 2019): 437–55. https://doi.org/10.1038/s41576-019-0128-0.
163. Scott, Michelle S., and Motoharu Ono. “From SnoRNA to MiRNA: Dual Function Regulatory Non-Coding RNAs.” Biochimie 93, no. 11 (November 2011): 1987–92. https://doi.org/10.1016/j.biochi.2011.05.026.
164. Sherpa, Chringma, Jason W. Rausch, and Stuart Fj Le Grice. “Structural Characterization of Maternally Expressed Gene 3 RNA Reveals Conserved Motifs and Potential Sites of Interaction with Polycomb Repressive Complex 2.” Nucleic Acids Research 46, no. 19 (November 2, 2018): 10432–47. https://doi.org/10.1093/nar/gky722.
165. Shibayama, Youtaro, Stephanie Fanucchi, Loretta Magagula, and Musa M. Mhlanga. “LncRNA and Gene Looping: What’s the Connection?” Transcription 5, no. 3 (2014): e28658. https://doi.org/10.4161/trns.28658.
166. Shukla, Chinmay J., Alexandra L. McCorkindale, Chiara Gerhardinger, Keegan D. Korthauer, Moran N. Cabili, David M. Shechner, Rafael A. Irizarry, Philipp G. Maass, and John L. Rinn. “High-Throughput Identification of RNA Nuclear Enrichment Sequences.” The EMBO Journal 37, no. 6 (March 15, 2018): e98452. https://doi.org/10.15252/embj.201798452.
167. Sirey, Tamara M, Kenny Roberts, Wilfried Haerty, Oscar Bedoya-Reina, Sebastian Rogatti-Granados, Jennifer Y Tan, Nick Li, et al. “The Long Non-Coding RNA Cerox1 Is

a Post Transcriptional Regulator of Mitochondrial Complex I Catalytic Activity." ELife 8 (May 2, 2019): e45051. https://doi.org/10.7554/eLife.45051.
168. Statello, Luisa, Chun-Jie Guo, Ling-Ling Chen, and Maite Huarte. "Gene Regulation by Long Non-Coding RNAs and Its Biological Functions." Nature Reviews. Molecular Cell Biology 22, no. 2 (February 2021): 96–118. https://doi.org/10.1038/s41580-020-00315-9.
169. Struhl, Kevin. "Transcriptional Noise and the Fidelity of Initiation by RNA Polymerase II." Nature Structural & Molecular Biology 14, no. 2 (February 2007): 103–5. https://doi.org/10.1038/nsmb0207-103.
170. Sun, Bryan K., Aimée M. Deaton, and Jeannie T. Lee. "A Transient Heterochromatic State in Xist Preempts X Inactivation Choice without RNA Stabilization." Molecular Cell 21, no. 5 (March 3, 2006): 617–28. https://doi.org/10.1016/j.molcel.2006.01.028.
171. Szabo, Quentin, Frédéric Bantignies, and Giacomo Cavalli. "Principles of Genome Folding into Topologically Associating Domains." Science Advances 5, no. 4 (April 2019): eaaw1668. https://doi.org/10.1126/sciadv.aaw1668.
172. Trapnell, Cole, Brian A Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Marijke J van Baren, Steven L Salzberg, Barbara J Wold, and Lior Pachter. "Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation." Nature Biotechnology 28, no. 5 (May 2010): 511–15. https://doi.org/10.1038/nbt.1621.
173. Tsai, Miao-Chih, Ohad Manor, Yue Wan, Nima Mosammaparast, Jordon K. Wang, Fei Lan, Yang Shi, Eran Segal, and Howard Y. Chang. "Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes." Science (New York, N.Y.) 329, no. 5992 (August 6, 2010): 689–93. https://doi.org/10.1126/science.1192002.
174. Tsai, Pei-Fang, Stefania Dell'Orso, Joseph Rodriguez, Karinna O. Vivanco, Kyung-Dae Ko, Kan Jiang, Aster H. Juan, et al. "A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In Trans." Molecular Cell 71, no. 1 (July 5, 2018): 129-141.e8. https://doi.org/10.1016/j.molcel.2018.06.008.
175. Turnescu, Tanja, Juliane Arter, Simone Reiprich, Ernst R. Tamm, Ari Waisman, and Michael Wegner. "Sox8 and Sox10 Jointly Maintain Myelin Gene Expression in Oligodendrocytes." Glia 66, no. 2 (February 2018): 279–94. https://doi.org/10.1002/glia.23242.
176. Ulitsky, Igor, and David P. Bartel. "LincRNAs: Genomics, Evolution, and Mechanisms." Cell 154, no. 1 (July 3, 2013): 26–46. https://doi.org/10.1016/j.cell.2013.06.020.
177. Vara, Covadonga, Andreu Paytuví-Gallart, Yasmina Cuartero, François Le Dily, Francisca Garcia, Judit Salvà-Castro, Laura Gómez-H, et al. "Three-Dimensional Genomic Structure and Cohesin Occupancy Correlate with Transcriptional Activity during Spermatogenesis." Cell Reports 28, no. 2 (July 2019): 352-367.e9. https://doi.org/10.1016/j.celrep.2019.06.037.
178. Vennin, Constance, Fatima Dahmani, Nathalie Spruyt, and Eric Adriaenssens. "Role of Long Non-Coding RNA in Cells: Example of the <I>H</I>19/<I>IGF</I>2 Locus." Advances in Bioscience and Biotechnology 04, no. 05 (2013): 34–44. https://doi.org/10.4236/abb.2013.45A004.

179.Verheul, Thijs C. J., Levi van Hijfte, Elena Perenthaler, and Tahsin Stefan Barakat. “The Why of YY1: Mechanisms of Transcriptional Regulation by Yin Yang 1.” Frontiers in Cell and Developmental Biology 8 (September 30, 2020): 592164. https://doi.org/10.3389/fcell.2020.592164.
180.Vernon, Ellen G, and Kevin Gaston. “Myc and YY1 Mediate Activation of the Surf-1 Promoter in Response to Serum Growth Factors.” Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1492, no. 1 (June 2000): 172–79. https://doi.org/10.1016/S0167-4781(00)00116-0.
181.Wang, Kevin C., and Howard Y. Chang. “Molecular Mechanisms of Long Noncoding RNAs.” Molecular Cell 43, no. 6 (September 16, 2011): 904–14. https://doi.org/10.1016/j.molcel.2011.08.018.
182.Wang, Kevin C., Yul W. Yang, Bo Liu, Amartya Sanyal, Ryan Corces-Zimmerman, Yong Chen, Bryan R. Lajoie, et al. “A Long Noncoding RNA Maintains Active Chromatin to Coordinate Homeotic Gene Expression.” Nature 472, no. 7341 (April 7, 2011): 120–24. https://doi.org/10.1038/nature09819.
183.Wang, Pin, Yiquan Xue, Yanmei Han, Li Lin, Cong Wu, Sheng Xu, Zhengping Jiang, Junfang Xu, Qiuyan Liu, and Xuetao Cao. “The STAT3-Binding Long Noncoding RNA Lnc-DC Controls Human Dendritic Cell Differentiation.” Science (New York, N.Y.) 344, no. 6181 (April 18, 2014): 310–13. https://doi.org/10.1126/science.1251456.
184.Wang, Xingwen, Yudong Wang, Li Li, Xuting Xue, Hui Xie, Huaxing Shi, and Ying Hu. “A LncRNA Coordinates with Ezh2 to Inhibit HIF-1α Transcription and Suppress Cancer Cell Adaption to Hypoxia.” Oncogene 39, no. 9 (February 2020): 1860–74. https://doi.org/10.1038/s41388-019-1123-9.
185.Wang, Yan, Yinping Xie, Lili Li, Yuan He, Di Zheng, Pengcheng Yu, Ling Yu, Lixu Tang, Yibin Wang, and Zhihua Wang. “EZH2 RIP-Seq Identifies Tissue-Specific Long Non-Coding RNAs.” Current Gene Therapy 18, no. 5 (2018): 275–85. https://doi.org/10.2174/1566523218666181008125010.
186.Wang, Yanli, Fan Song, Bo Zhang, Lijun Zhang, Jie Xu, Da Kuang, Daofeng Li, et al. “The 3D Genome Browser: A Web-Based Browser for Visualizing 3D Genome Organization and Long-Range Chromatin Interactions.” Genome Biology 19, no. 1 (December 2018): 151. https://doi.org/10.1186/s13059-018-1519-9.
187.Wang, Yue, Zhenyu Xu, Junfeng Jiang, Chen Xu, Jiuhong Kang, Lei Xiao, Minjuan Wu, Jun Xiong, Xiaocan Guo, and Houqi Liu. “Endogenous MiRNA Sponge LincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal.” Developmental Cell 25, no. 1 (April 15, 2013): 69–80. https://doi.org/10.1016/j.devcel.2013.03.002.
188.Wei, Jian-Wei, Kai Huang, Chao Yang, and Chun-Sheng Kang. “Non-Coding RNAs as Regulators in Epigenetics (Review).” Oncology Reports 37, no. 1 (January 2017): 3–9. https://doi.org/10.3892/or.2016.5236.
189.Weick, Eva-Maria, and Eric A. Miska. “PiRNAs: From Biogenesis to Function.” Development (Cambridge, England) 141, no. 18 (September 2014): 3458–71. https://doi.org/10.1242/dev.094037.

190.Weick, Eva-Maria, Eric A Miska. “PiRNAs: From Biogenesis to Function.” Development (Cambridge, England) 141, no. 18 (September 2014): 3458–71. https://doi.org/10.1242/dev.094037.
191.Weintraub, Abraham S., Charles H. Li, Alicia V. Zamudio, Alla A. Sigova, Nancy M. Hannett, Daniel S. Day, Brian J. Abraham, et al. “YY1 Is a Structural Regulator of Enhancer-Promoter Loops.” Cell 171, no. 7 (December 14, 2017): 1573-1588.e28. https://doi.org/10.1016/j.cell.2017.11.008.
192.Werner, Torsten, Alexander Hammer, Mandy Wahlbuhl, Michael R. Bösl, and Michael Wegner. “Multiple Conserved Regulatory Elements with Overlapping Functions Determine Sox10 Expression in Mouse Embryogenesis.” Nucleic Acids Research 35, no. 19 (2007): 6526–38. https://doi.org/10.1093/nar/gkm727.
193.Wongtrakoongate, Patompon, Gregory Riddick, Suthat Fucharoen, and Gary Felsenfeld. “Association of the Long Non-Coding RNA Steroid Receptor RNA Activator (SRA) with TrxG and PRC2 Complexes.” PLoS Genetics 11, no. 10 (October 2015): e1005615. https://doi.org/10.1371/journal.pgen.1005615.
194.Wutz, Anton, Theodore P. Rasmussen, and Rudolf Jaenisch. “Chromosomal Silencing and Localization Are Mediated by Different Domains of Xist RNA.” Nature Genetics 30, no. 2 (February 2002): 167–74. https://doi.org/10.1038/ng820.
195.Xiang, Jian-Feng, Qing-Fei Yin, Tian Chen, Yang Zhang, Xiao-Ou Zhang, Zheng Wu, Shaofeng Zhang, et al. “Human Colorectal Cancer-Specific CCAT1-L LncRNA Regulates Long-Range Chromatin Interactions at the MYC Locus.” Cell Research 24, no. 5 (May 2014): 513–31. https://doi.org/10.1038/cr.2014.35.
196.Xiao, Tiaojiang, Xin Li, and Gary Felsenfeld. “The Myc-Associated Zinc Finger Protein (MAZ) Works Together with CTCF to Control Cohesin Positioning and Genome Organization.” Proceedings of the National Academy of Sciences of the United States of America 118, no. 7 (February 16, 2021): e2023127118. https://doi.org/10.1073/pnas.2023127118.
197.Yamazaki, Tomohiro, Sylvie Souquere, Takeshi Chujo, Simon Kobelke, Yee Seng Chong, Archa H. Fox, Charles S. Bond, Shinichi Nakagawa, Gerard Pierron, and Tetsuro Hirose. “Functional Domains of NEAT1 Architectural LncRNA Induce Paraspeckle Assembly through Phase Separation.” Molecular Cell 70, no. 6 (June 21, 2018): 1038-1053.e7. https://doi.org/10.1016/j.molcel.2018.05.019.
198.Yang, Fan, Xinxian Deng, Wenxiu Ma, Joel B Berletch, Natalia Rabaia, Gengze Wei, James M Moore, et al. “The LncRNA Firre Anchors the Inactive X Chromosome to the Nucleolus by Binding CTCF and Maintains H3K27me3 Methylation.” Genome Biology 16, no. 1 (December 2015): 52. https://doi.org/10.1186/s13059-015-0618-0.
199.Yang, Fan, Huafeng Zhang, Yide Mei, and Mian Wu. “Reciprocal Regulation of HIF-1α and LincRNA-P21 Modulates the Warburg Effect.” Molecular Cell 53, no. 1 (January 9, 2014): 88–100. https://doi.org/10.1016/j.molcel.2013.11.004.
200.Yang, Yiqi, and Gang Li. “Post-Translational Modifications of PRC2: Signals Directing Its Activity.” Epigenetics & Chromatin 13, no. 1 (December 2020): 47. https://doi.org/10.1186/s13072-020-00369-1.

201.Yao, Hongjie, Kevin Brick, Yvonne Evrard, Tiaojiang Xiao, R. Daniel Camerini-Otero, and Gary Felsenfeld. “Mediation of CTCF Transcriptional Insulation by DEAD-Box RNA-Binding Protein P68 and Steroid Receptor RNA Activator SRA.” Genes & Development 24, no. 22 (November 15, 2010): 2543–55. https://doi.org/10.1101/gad.1967810.
202.Yap, Kyoko L., Side Li, Ana M. Muñoz-Cabello, Selina Raguz, Lei Zeng, Shiraz Mujtaba, Jesús Gil, Martin J. Walsh, and Ming-Ming Zhou. “Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a.” Molecular Cell 38, no. 5 (June 11, 2010): 662–74. https://doi.org/10.1016/j.molcel.2010.03.021.
203.Ye, Buqing, Benyu Liu, Liuliu Yang, Xiaoxiao Zhu, Dongdong Zhang, Wei Wu, Pingping Zhu, et al. “LncKdm2b Controls Self-Renewal of Embryonic Stem Cells via Activating Expression of Transcription Factor Zbtb3.” The EMBO Journal 37, no. 8 (April 13, 2018): e97174. https://doi.org/10.15252/embj.201797174.
204.Yoon, Je-Hyun, Kotb Abdelmohsen, and Myriam Gorospe. “Posttranscriptional Gene Regulation by Long Noncoding RNA.” Journal of Molecular Biology 425, no. 19 (October 9, 2013): 3723–30. https://doi.org/10.1016/j.jmb.2012.11.024.
205.Yoon, Je-Hyun, Kotb Abdelmohsen, Subramanya Srikantan, Xiaoling Yang, Jennifer L. Martindale, Supriyo De, Maite Huarte, Ming Zhan, Kevin G. Becker, and Myriam Gorospe. “LincRNA-P21 Suppresses Target MRNA Translation.” Molecular Cell 47, no. 4 (August 24, 2012): 648–55. https://doi.org/10.1016/j.molcel.2012.06.027.
206.Zhang, Kai, Nan Li, Richard I. Ainsworth, and Wei Wang. “Systematic Identification of Protein Combinations Mediating Chromatin Looping.” Nature Communications 7 (July 27, 2016): 12249. https://doi.org/10.1038/ncomms12249.
207.Zhang, Kun, Zhe-Min Shi, Ya-Nan Chang, Zhi-Mei Hu, Hai-Xia Qi, and Wei Hong. “The Ways of Action of Long Non-Coding RNAs in Cytoplasm and Nucleus.” Gene 547, no. 1 (August 15, 2014): 1–9. https://doi.org/10.1016/j.gene.2014.06.043.
208.Zhao, Jing, Bryan K. Sun, Jennifer A. Erwin, Ji-Joon Song, and Jeannie T. Lee. “Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome.” Science (New York, N.Y.) 322, no. 5902 (October 31, 2008): 750–56. https://doi.org/10.1126/science.1163045.
209.Zhao, Zhongliang, Marcel Andre Dammert, Ingrid Grummt, and Holger Bierhoff. “LncRNA-Induced Nucleosome Repositioning Reinforces Transcriptional Repression of RRNA Genes upon Hypotonic Stress.” Cell Reports 14, no. 8 (March 1, 2016): 1876–82. https://doi.org/10.1016/j.celrep.2016.01.073.
210.Zuckerman, Binyamin, Maya Ron, Martin Mikl, Eran Segal, and Igor Ulitsky. “Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex.” Molecular Cell 79, no. 2 (July 16, 2020): 251-267.e6. https://doi.org/10.1016/j.molcel.2020.05.013.

List of Publications

1) **Bhavana Kayyar**, Anjhana C.R, Utsa Bhaduri, Satyanarayana MR Rao, Chromatin organisation mediated gene regulatory function of mrhl lncRNA at mouse Sox8 locus. **Manuscript under revision** (2021)
2) Shubhangini Kataruka, Vijay S Akhade, **Bhavana Kayyar**, Satyanarayana MR Rao, Mrhl Long Noncoding RNA Mediates Meiotic Commitment of Mouse Spermatogonial Cells by Regulating Sox8 Expression; **Molecular and Cellular Biology**, (2017)
3) **Bhavana Kayyar**, Sangeeta Dutta, Anjhana C.R, Satyanarayana MR Rao, LncRNA as epigenetic regulators in development and disease. **Review Article, Under preparation** (2021)
4) Utsa Bhaduri, **Bhavana Kayyar**, Satyanarayana MR Rao, RNA-DNA Triplexes: From Predictions to Experiments to Machine Learning. **Review Article, Under preparation** (2021)