Characterizing the role of long non-coding RNAs as epigenetic regulators in disease model

Kankadeb Mishra

Department of Medical Biochemistry and Cell Biology

Institute of Biomedicine

Sahlgrenska Academy at university of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2017

Characterizing the role of long non-coding RNAs as epigenetic regulators in disease model

kankadeb@gmail.com

ISBN: 978-91-629-0368-8 (PDF)

Printed in Gothenburg, Sweden 2017

Ineko AB, Gothenburg

Dedicated to my parents, Manikarnika, Souvik and uncle (Chotokaku)

Characterizing the role of long non-coding RNAs as epigenetic regulators in disease model

Kankadeb Mishra

Department of Medical Biochemistry and Cell Biology

Institute of Biomedicine

Sahlgrenska Academy at University of Gothenburg

ABSTRACT

Long non-coding RNAs (lncRNAs) are a class of biological molecules which are transcribed from DNA but are not translated into any protein. LncRNAs have been identified as critical players in gene regulation. Misregulation of lncRNAs has been considered as one of the underlying causes for cancer pathogenesis and in other human diseases. In the current thesis, I have addressed the epigenetic roles of lncRNAs in regulating gene expression in cell line based and disease model systems.

We investigated the functional role of lncRNAs in the maintenance of active chromatin by sequencing lncRNAs associated with active chromatin enriched with H3K4me2 and WDR5. We identified 209 lncRNAs to be commonly enriched in H3K4me2 and WDR5 pulldown chromatin fractions and we named them as active chromatin associated RNAs (active CARs). Interestingly, 41% of active CARs mapped to divergent transcription units having transcription factor genes as their partner. CARs were found to regulate the expression of partner protein coding genes at the transcriptional level by recruiting WDR5 to maintain the active histone marks H3K4me2/H3K4me3 at these promoters. Depletion of active CARs results in reduced WDR5and H3K4me2/H3K4me3 occupancy at these promoters. However, in absence of WDR5, we found the levels of H3K4me2 to remain unchanged at divergent promoters. Taken together our findings indicates that, conversion of H3K4me2 to H3K4me3 is mediated via active CARs-WDR5 interaction at the active divergent promoter, whereas, the maintenance of H3K4me2 marks appears to be WDR5 independent.

Additionally, we used transcriptome profiling approach, to identify lncRNAs that are differentially expressed between low- and high- risk neuroblastoma tumours. We report *NBAT-1* lncRNA as an independent prognostic biomarker in predicting clinical outcome of neuroblastoma patients. The expression profile analysis showed *NBAT-1* to be lowly expressed in high-risk tumours relative to low-risk tumours. Using cell line and mouse models we characterized *NBAT-1* as a tumour suppressor lncRNA which regulates gene expression by interaction with PRC2 repressive chromatin complex. *NBAT-1* lncRNA promotes differentiation and acts as a tumour suppressor by epigenetic regulation of genes to inhibit cell proliferation and invasion.

Thirdly, we sought to study genomic imprinting in a disease model. Genomic imprinting is an epigenetic regulation of gene expression in a parent of origin-specific manner. Studies in mouse have identified *Kncq1* imprinted domain to be epigenetically regulated by a 91kb long lncRNA *Kcnq1ot1* which is expressed from the paternal chromosome to silence imprinted genes in *cis*. Using BW-syndrome human disease model, we identified a maternal 11p15.5 micro duplication which included the 5′ 20 kb of the non-coding *KCNQ1OT1* gene. Its maternal transmission was associated with ICR2 hypomethylation and familial BWS phenotype. Normally ICR2 is methylated to repress *KCNQ1OT1*, thereby allowing

maternal copies of the imprinted genes including growth inhibitor *CDKN1C* to be expressed. We demonstrated that this duplicated maternal *KCNQ1OT1* RNA also interacts with chromatin through its most 5′ 20 kb sequence to silence *CDKN1C*. This provides a mechanism for biallelic silencing of *CDKN1C* which contributes to the BWS disease phenotype.

In summary, by ChRIP-seq, RNA expression profiling in tumours and human patient-derived cell line based model systems, we have uncovered new roles of lncRNA in epigenetic gene regulation.

Keywords: Long non-coding RNA, Epigenetics, Active Chromatin, Genomic Imprinting, Neuroblastoma, Beckwith-Wiedemann Syndrome.

ISBN: 978-91-629-0368-8 (PDF)

SAMMANFATTNING PÅ SVENSKA

En grupp långa icke-kodande RNA molekyler ofta kallade long non-coding RNAs eller lncRNAs transkriberas från DNA men translateras inte till protein. LncRNAs är av betydelse för genreglering och anses spela en roll i utvecklingen av cancer och andra sjukdomar som drabbar människan. Jag har i denna avhandling undersökt den roll lncRNA spelar för epigenetisk reglering av genexpression cell linjer samt sjukdomsmodeller.

Den funktionella betydelsen av lncRNA för aktivt kroomatin har studerats genom sekvenering av lncRNA associerat med H3K4Me2 och WDR5. Vi fann 209 lncRNA molekyler anrikade i kromatinfraktioner innehållande H3K4Me2 och WDR5. Dessa benämndes aktivt chromatin associerat RNA eller "active CAR". 41 % av kartlagda "active CARs" sammanföll med divergerande transkriptionsenheter där den ena delen utgjordes av gener kodande för transkriptionsfaktorer. CARs visades reglera uttrycket av transkriptionsfaktorgenerna på transkriptionell nivå genom rekrytering av WDR5. Detta resulterade is sin tur att de aktivitetskopplade markörerna WDR5 och H3K4Me2/H3K4me3 bevarades på dessa promotorer. Sänkta nivåer av "active CARs" medförde lägre nivåer av WDR5and H3K4me2/H3K4me3 vid promotorerna. I frånvaro av WDR5 var dock nivåerna av WDR5and H3K4me2 oförändrade. Dessa observationer indikerar att för aktiva divergerande promotorer medieras förändringen av H3K4me2 till H3K4me3 via en inetraktion mellan "active CARs" och WDR5, medan bibehållande av WDR5and H3K4me2 är oberoende av WDR5.

Med utnyttjande av transkriptions analys, "transcriptional profiling", kunde vi vidare identifiera skillnader i lncRNA uttryck i neuroblastom associerade med låg- respektive hög risk. Vi redovisar att lncRNA *NBAT-1* kan utgöra en biomarkör med förmåga att förutsäga den kliniska bilden av neuroblastom där lågt uttryck av *NBAT-1* framförallt ses vid hög sjukdomsrisk. Vi fann såväl i cell linjer som i möss att *NBAT-1* stimulerar celldifferentiering och därmed utgör den en tumörsuppressor med förmåga att hämma cell proliferation och cell invasion.

För det tredje försökte vi studera genomisk imprinting i en sjukdomsmodell. Med genomisk imprinting menas föräldraspecifik epigenetisk reglering av genexpression. I möss har visats att *Kcnq1ot1* regionen utgör en "imprinted" domän. En 91 kb lång lncRNA molekyl, *Kcnq1ot1*, vilken uttrycks från faderns kromosom, kan inaktivera gener i *cis*. Med BW-syndromet som mänsklig sjukdomsmodell kunde vi indentifiera en mikroduplikation på moderns kromosomala locus 11p15.5 vilket innehöll 20kb av den icke-kodande *KCNQ1OT1* genes 5´ del. Nedärvning av denna gen var kopplad till hypometylering av ICR2 and egenskaper karaktäristiska för ärftlig BWS. Under normala omständigheter metyleras ICR2 för att nedreglera *KCNQ1OT1*, vilket i sin tur medför att moderns kopior av "imprinted genes'', inkluderande till växtinhibitorn *CDKN1C*, kommer till uttryck. Vi viasade också att *KCNQ1OT1* RNA uttryckt från moderns duplicerade gen interagerade med kromatin via sin 5´del för att nedreglera *CDKN1C*. Våra resultat visar på en mechanism genom vilken *CDKN1C* bidrar till BWS sjukdom.

Sammanfattningsvis har vi med hjälp av ChRIP-seq på odlade celler samt RNA expressions analys i tumörer och cell linjer från patienter upptäckt nya funktioner för lncRNA vid epigenetisk reglering.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals. Articles are reprinted with permission from publishers.

I. H3K4me2 and WDR5 enriched chromatin interacting long non-coding RNAs maintain transcriptionally competent chromatin at divergent transcriptional units.
Mishra K, Subhash S, Akhade VS, Kanduri M, Mondal T and Kanduri C (manuscript).

II. The Risk-Associated Long Noncoding RNA NBAT-1 Controls Neuroblastoma Progression by Regulating Cell Proliferation and Neuronal Differentiation.
Pandey GK, Mitra S, Subhash S, Hertwig F, Kanduri M, **Mishra K**, Fransson S, Ganeshram A, Mondal T, Bandaru S, Ostensson M, Akyürek LM, Abrahamsson J, Pfeifer S, Larsson E, Shi L, Peng Z, Fischer M, Martinsson T, Hedborg F, Kogner P, Kanduri C. ***Cancer Cell*****,** 2014 Nov 10; 26(5):722-37. doi: 10.1016/j.ccell.2014.09.014.

III. The KCNQ1OT1 imprinting control region and non-coding RNA: new properties derived from the study of Beckwith–Wiedemann syndrome and Silver–Russell syndrome cases. Chiesa N, De Crescenzo A, **Mishra K**, et al. ***Human Molecular Genetics***. 2012; 21(1):10-25. doi:10.1093/hmg/ddr419.

Additional publication not included in the thesis:

MCPH1 maintains long-term epigenetic silencing of ANGPT2 in chronic lymphocytic leukemia. Kopparapu PK, Miranda C, Fogelstrand L, **Mishra K**, Andersson PO, Kanduri C, Kanduri M. ***FEBS J.*** 2015 May; 282(10):1939-52. doi: 10,1111/febs.13245.

ABBREVIATIONS

ncRNA	Non-coding RNA
LincRNA	Long intergenic RNA
LncRNA	Long non-coding RNA
ICR	Imprinting control region
DMR	Differentially methylated regions
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
ES	Embryonic stem cells
iPSCs	Induced Pluripotent stem cells
UIGs	Ubiquitously imprinted genes
PIGs	Placental imprinted gene
NIGs	Non imprinted genes
XCI	X chromosome inactivation
XIC	X inactivation center
DCC	Dosage compensation complex
HOX	Homeobox
PCAT	Prostate cancer associated transcript
NBAT-1	Neuroblastoma associated transcript-1
Ezh2	Enhancer of zeste homologue 2
CAGE	Cape analysis of gene expression
RIP	RNA immuno precipitation
ChRIP	Chromatin RNA immuno precipitation
ChOP	Chromatin oligo-affinity precipitation
CAR	Chromatin associated RNA
SHAPE	Selective 2′-hydroxyl acylation analyzed by primer extension
HMT	Histone methyl transferase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas9	CRISPR associated protein 9
DNMT1	DNA methyl transferase1
RIA-seq	RNA interactome analysis and sequencing
CTCF	CCCTC-Binding factor
YY1	Yin Yang 1
LSD1	Lysine-specific histone demethylase 1A

PROMPTs	Promoter upstream transcripts
eRNAs	Enhancer RNAs
NATs	Natural antisense transcripts
snoRNA	Small nucleolar RNA
ciRNA	Circular RNA
HAT	Histone acetyl transferase
WDR5	WD Repeat Domain 5
MLL1	Mixed-lineage leukemia 1
PCGF	Polycomb group ring fingers

CONTENTS

CHAPTER 1: INTRODUCTION

Evolution of organismal complexity has long raised intriguing questions regarding the mechanism by which higher eukaryotes have gradually developed cellular complexity. Many hypotheses have been generated to explain the cellular diversity as well as the complexity of gene regulation in higher organisms. One of the key initial theories was the Central Dogma, which proposed that DNA is transcribed into RNA, which is then translated into functional molecules called proteins which are the sole regulators of gene regulatory network. Thus, increase in the number of protein coding genes was initially thought to correlate positively with increasing complexity. Several evidences that challenged this theory have started accumulating over the years. However, the most convincing evidences against this theory came from the data of high throughput sequencing studies which proved beyond doubt that the number of protein coding genes do not increase with increasing complexity. A simple nematode like *C. elegans* and more complex organism like human beings both contain approximately 20,000 protein coding genes, such inconsistency between complexity and number of genes is referred to as G-value paradox [1]. Interestingly, while the genomes of higher eukaryotes pervasively transcribe to form RNA, only a small percentage of the transcribed RNAs were found to be translated into proteins [2] which are called as non-coding RNAs (ncRNAs). This non coding portion of the genome have consistently increased with complexity [3]. These observations underline the importance of non-coding portion of the genome in the lights of evolution of organismal complexity.

Thus, how does the same number of protein coding genes regulate both a worm with around 1000 cells and a human with diverse cell types and complex body plan? To counteract this discrepancy for number of genes, multilayered mechanisms of gene regulation must be present in in highly evolved organism order to efficiently control and coordinate the system. Interestingly non-coding RNAs have been found to form a significant part of the mammalian transcriptome that have been implicated in the regulation of gene expression at multiple levels [4, 5]. NcRNAs can be broadly classified into two categories based on their functions *i.e.* those involved in housekeeping function (tRNA, rRNA) and others which are regulatory in nature. Regulatory ncRNA are further classified based on their size as small ncRNA (snoRNA, microRNA, siRNA, snRNA, piRNA etc.) and as long ncRNA (lncRNA). LncRNAs are arbitrarily defined as transcripts of 200 nucleotides (nt) or more. Taken together, all these ncRNAs generate a complex transcriptional output in mammals in addition to the limited number of protein coding genes.

Interestingly, detailed analysis of the mouse transcriptome by FANTOM3 consortium indicated that more than 72% of all mapped transcripts overlap with an antisense transcription [6]. Subsequently, the expression of some long antisense ncRNAs has been shown to be coupled with the silencing of protein-coding genes situated on both sides of the antisense transcripts [7]. Particularly gene clusters with parent of origin-specific mono allelic expression patterns (defined as genomically imprinted locus, where either the maternal or the paternal alleles of these genes are expressed) often contain one or more such antisense lncRNAs as their partners [8] [9]. Some of these lncRNAs are very large, ranging in size from

fifty to several hundred nucleotides. In most cases, the promoters of these lncRNAs, which are associated with an imprinted gene cluster, map to differentially methylated imprinting control regions (ICRs, which are defined as the critical region within an imprinted cluster, whose DNA methylation determines the outcome of expression patterns of all other genes in that cluster). Transcription from these promoters occurs in an antisense direction relative to the protein coding genes, and majority of these lncRNAs are expressed only from the paternal chromosomes due to methylation of their promoters on the maternal chromosomes [10-12]. The expression of long antisense ncRNA on the paternal chromosomes has been shown to be correlated with the repression of protein coding genes in *cis*, spreads over several hundred kilo-bases on either side of the antisense transcription unit, indicating a link between lncRNA expression and silencing of neighboring protein-coding genes [13, 14]. Apart from these long antisense RNAs which have been implicated as critical elements in fine tuning gene expression of imprinted gene clusters, several lncRNAs have been reported to regulate gene expression in different biological processes, by deploying diverse mechanisms. Regulation of gene expression is critical in ensuring both precise spatio-temporal expression of genes and for producing isoforms of the same gene, thereby increasing molecular complexity. Over the last decade, several reports have implicated the role of different lncRNAs and small ncRNAs (mechanism of regulation differs) in the regulation of diverse biological functions such as pluripotency, differentiation, carcinogenesis, body pattern and development, imprinting, immune response, metabolism etc. [12, 15-17].

In addition to classical pre/post -transcriptional and -translational control of gene expression, evidence has also shown the role of epigenetic modifications (like DNA methylation, post-translational modifications in the tails of chromatin associated histone proteins) to act as critical regulators of gene expressions. Regulation of gene expression at transcriptional level is also coordinated and orchestrated by the organization of chromatin into poised, active or inactive states. This is brought about either or both by DNA methylation and histone modifications. Several chromatin modifying complexes work in complexes to bring about the DNA methylation as well as the histone tail modifications. Several lncRNAs have been implicated to have functional association with both DNA and histone modifying complexes [16, 18]. It has recently been shown that several lncRNAs regulate gene expression in critical cellular contexts by organizing the chromatin into active and inactive domains through direct interaction with different chromatin modifying enzymes. LncRNAs have also been reported to associate with chromatin and control transcriptional gene regulation of neighboring genes [19]. Functional association of few lncRNAs with chromatin modifying complexes and/or DNA methylation machinery has been shown to be critical in disease and developmental contexts. This highlights the functional importance of lncRNAs which associate with chromatin to either facilitate or repress the expression of genes. Additionally, lncRNAs together with architectural proteins like CTCF and mediators organizes chromatin into distinct three-dimensional structural domains, which may represent functional units of

the genome. This additional layer of gene regulatory mechanism mediated via lncRNAs might also be interesting to investigate further.

Our research group is focused on understanding the role of lncRNAs in the epigenetic regulation of gene expression in two different contexts. Firstly, using human diseases as a model system we look to address the mechanism by which lncRNAs might contribute to disease pathogenesis. Secondly, by using epigenetic mechanism based on global purification approaches in cell culture model systems we try to identify and characterize regulatory lncRNAs.

In the present thesis, I will emphasize on characterizing the role of lncRNAs as epigenetic regulators using disease model and cell culture based model systems. I will begin by explaining the classical epigenetic regulators of gene expression which includes DNA and histone modifications and their respective modifiers. Next, I will discuss in detail about general characteristic of lncRNAs: annotation, genomics, expression, conservation, localization and functional mechanisms. Eventually, I will discuss some of the lncRNAs that have been identified by different approaches to act as epigenetic regulators of gene expression in important biological processes. Finally, I will discuss specific aims, results and conclusions from my own studies.

CHAPTER 2: EPIGENETIC REGULATORS OF GENE EXPRESSION

Different cell types in multicellular organisms have the same genetic composition, yet they differ from each other both phenotypically and functionally which is defined by cell-type specific gene expression patterns. Such differential cell and tissue specific expression pattern of genes can be partly explained by the concept of epigenetics. Genetics is the study of inheritable changes in the DNA sequences. Epigenetics on the contrary is the study of modifications of DNA and DNA associated proteins without any changes in the DNA sequence. Similar to genetic changes, epigenetic modifications are also inherited during cell division and also from one generation to another. Unlike genetic changes, reversibility of epigenetic changes offers a unique flexibility to the genome in response based regulation of its function. Chromatin is the carrier of epigenetic changes which are stable modifications inherited through mitotic cell divisions, and in some cases have also been found to be inherited from one generation to another (transgenerational) [20, 21] . Chromatin was originally thought to be composed of only DNA and histone proteins, however, emerging evidences have established even RNA to be an important functional component of chromatin as well [22, 23]. Epigenetic modifications or chromatin modifications as we may say, are of two main types: DNA methylation and different post-translational modifications of the histone tail proteins. In conjugation with ncRNAs these two chromatin modifications execute transcriptional regulation of gene expression in a cell-type and tissue specific manner. In the following section I will give an overview of the role of DNA methylation and histone modifications in gene regulation.

2.1 DNA Methylation

DNA methylation is a stable epigenetic mark that occurs by covalent addition of a methyl (CH3) group at the C5 position of cytosine, mainly in a CpG dinucleotide context, but also in non-CpG regions of human embryonic stem cells [24, 25]. These methyl groups project into the major groove of DNA and inhibit transcription. In eukaryotic genome where more than 50% of genes are associated with CpG islands in their promoter regions 5mC at CpG dinucleotide is one of the major epigenetic silencing mechanism. Generally, low levels or a lack of CpG methylation in the promoter region is correlated with an “on” chromatin configuration that allows interaction of DNA with transcription complexes required for activation of gene expression. By contrast, methylation of CpG islands in gene promoters is correlated with an “off” chromatin configuration that leads to gene silencing. Precise regulation of promoter DNA methylation is thus crucial for maintaining differential gene expression patterns in a tissue and developmental-stage-specific manner.

Mammalian DNA methylation machinery comprises of three families of DNA methyltransferases (DNMTs) of which DNMT3A and DNMT3B are involved in the establishment of methylation patterns (*de novo* methyltransferases) whereas DNMT1 is involved in the faithful propagation of methylation

pattern over cellular divisions (maintenance methyltransferase). DNMT3L, another member of this protein family, lacks methyltransferase activity but is crucial to stimulate the activity of DNMT3A and DNMT3B in the establishment of methylation at most of the imprinted gene loci [26]. Studies have identified roles of additional proteins that are required in addition to DNMT1 in the maintenance of methylation during somatic cell division. UHRF1 protein which specifically binds to hemi methylated CpG nucleotides is required for the maintenance of DNA methylation in embryonic stem cells by regulating the stability of DNMT1 through its association with methylated histone H3 on lysine 9 residue [27].

The CpG dinucleotides are underrepresented in the mammalian genome due to frequent deamination of methylated cytosines to be consequently converted into C > T mutations [28, 29]. In the mammalian genome 50-70% of the CpG dinucleotides are methylated while the remaining hypomethylated CpGs are present as clusters called CpG islands in the promoters of constitutive expressed housekeeping genes [30]. CpG islands are genomic regions with high GC content of varied length, mostly near promoters, that are generally hypomethylated with the exception for CpG islands of imprinted genes and genes undergoing X-chromosome inactivation [31, 32]. As discussed in the imprinting section, other factors are involved in the regulation CpG methylation of these genes. 72% of Refseq annotated gene promoters in human are associated with high CpG [33]. In contrast, at active promoter regions in mammals which are enriched with high H3K4me3, the CpG islands are protected from methylation. The presence of H3K4me3 modification prevents the Dnmt3l from interacting with unmodified histone H3 which prevents the access of DNA methyltransferase to promoters of active genes [34-36]. In addition, Cfp1 protein has been shown to bind non-methylated CpG islands to maintain active chromatin conformation via recruitment of active histone methyl transferases like Set1a [37].

Studies have suggested a functional link between H3K36me3 histone modification catalysed by Setd2 [38] and *de novo* DNA methylation. Unlike promoter CpG methylation, high levels of gene body methylation by Dmnt3a/b have been associated with higher expression of genes in mouse oocytes where the elongating RNA PolII mediates H3K36me3 modification via Setd2 which seem to recruit Dmnt3a/b [39, 40].

2.2 Histone Modifications

Histones are the core protein component of chromatin around which DNA is wrapped to form nucleosomes. The core histone is a hetero-octameric protein consisting of a dimer of four different proteins called as H2A, H2B, H3 and H4. 146bp of nucleosomal DNA is wrapped around this octameric histone. Another type of histone called H1 or linker histone is also a crucial component of nucleosome. The N-terminal tails of each of these four core histones and H1 histone are loosely associated with the octamer and are susceptible to post translational protein modifications by different types of enzymes

which has chromatin binding affinity. The post translational modifications of histone tails include covalent addition of the following groups: Acetyl group (acetylation), Methyl group (methylation), Phosphate group (phosphorylation), Ubiquitin group (Ubiquitination) and Sumoyl group (sumoylation). Most of these modifications occur on lysine (K) and arginine (R) residues at different positions on the histone tail as denoted by a position number. For example, lysine residue at the 9th position in the tail is denoted as K9, while, the one at position 4 is denoted as K4 and likewise. The nomenclature of histone modifications includes the name of histone, residue name, residue position, type of covalent group and number of added groups. For example, addition of one two (2) methyl groups (me) to lysine (K) at position 4 of H3 tail would be called H3K4me2. These modifications of the histone tails define the chromatin state by affecting the accessibility of regulatory complexes to the chromatin thereby regulating transcriptional gene regulation. Certain histone modifications result in a compacted chromatin to repress transcription whereas other modifications result in an open chromatin that facilitates transcription. The ability to predict transcriptional outcome based on histone modification with far-reaching consequences for cell fate decisions and development is also known as histone code hypothesis [41]. Combination of different histone modifications can form a complex regulatory network to control and fine tune spatio-temporal gene expression. Apart from canonical histones, several variants of basic histones also regulate chromatin structure and therefore play important roles in transcriptional regulation, chromatin stability, DNA repair, chromosome segregation, cell cycle progression and apoptosis [42]. In the next section I will focus in greater detail about the role of canonical histone modifications as epigenetic regulators.

2.2.1 Repressive histone modifications:

Methylation of H3 on lysine 9 and lysine 27 and the ubiquitinylation of H2A on lysine 119 are the characteristic repressive histone modifications that are often present at silent gene loci. H3K9me2/3 are associated with the formation of constitutive heterochromatin whereas H3K27me3 and H2AK119Ub1 are associated with the formation of facultative heterochromatin.

H3K27me3 and H2AK119Ub1

The Polycomb Repressive Complex 2 (PRC2) which contains four core subunits, EZH1/2, SUZ12, EED, and RBAP46/8 is responsible for the methylation of lysine 27 [43]. None of the core PRC2 components possess a DNA binding domain. The catalytic subunit is the SET domain-containing protein EZH2 (or the related EZH1) [44], although there are also accessory proteins that can associate with the core PRC2 which includes the JARID2 and AEBP2 subunits [45]. The accessory proteins have been implicated in modulation of PRC2 activity and are hypothesized to play a role in the targeting of PRC2 to chromatin [46, 47]. PRC2 mediated mono-, di-, and tri-methylation of H3K27 may have very

different regulatory roles [48]. Although H3K27me3 has been associated to gene repression, enrichment of H3K27me1 over gene bodies has suggested gene activation role for H3K27me1 [49]. The H3K27me2 modification accounts for 60–80% of all nucleosomes in mESCs, however not much is known about its function or binding proteins [50]. On the other hand, the critical role of H3K27me3 in facultative heterochromatin formation and transcriptional repression during development is well characterized.

The Polycomb repressive complex 1 (PRC1) is an E3 ubiquitin ligase complex that is responsible for monoubiquitylation of H2A on lysine 119. PRC1 complex contains the catalytic RING1A/B subunit, along with one of the six different PCGF proteins [51] that define the class of PRC1 complex (the canonical and the variant form of PRC1). Compared to canonical PRC1, the variant complexes have been associated with higher H2AK119Ub1 activity and recruitment of PRC1 [51, 52]. Both PRC1 and PRC2, and the associated chromatin modifications, H2AK119Ub1 and H3K27me3, have been found to co-localize at many promoters of developmentally regulated genes and at the inactive X chromosome [53-55]. Published evidences have suggested both in favour of and against PRC2 mediated hierarchical recruitment of PRC1 and vice versa to explain the mechanism of co-localization of these two repressive epigenetic modifiers [56-59]. Thus, the establishment of Polycomb repressive domains not only requires EED-binding to H3K27me3 or RYBP-binding to H2AK119Ub1, but also the marks placed by their partner complex so that H3K27me3 can establish or reinforce H2AK119Ub1 modifications, and H2AK119Ub1 can establish or reinforce H3K27me3 deposition. However, once established, the positive feedback mechanisms involving the histone modifications are crucial to maintain the PRC1/2 activity at these target sites [60].

H3K27me3 and H3K9me2/3

Methylation of H3K9 generally associated with constitutive heterochromatin formation and transcriptional silencing, has been implicated to crosstalk with the H3K27me3 modification. The heterodimeric complex of G9a and GLP catalyses H3K9me1 and H3K9me2 modifications [61], SETDB1 and SUV3-9H1/H2 both catalyse H3K9me2 and H3K9me3 modifications [62, 63]. At pericentric heterochromatin, chromodomain-containing protein HP1 binds to H3K9me3 and recruits *de novo* DNA methyltransferases (DNMT3A/B) resulting in DNA methylation which further facilitates MECP2 mediated recruitment of SUV3-9 enzymes, thereby maintaining stable heterochromatic marks during cell division [64-66].

Distribution of H3K9me3 and H3K27me3 modifications are generally mutually exclusive with H3K9me3 associated with transposons while H3K27me3 is enriched over silent gene loci [58, 67, 68]. Gain of H3K27me3 and a loss of H3K9me3 at the pericentric heterochromatin upon SUV3-9H1/H2 depletion, suggests that H3K9me3 normally prevents establishment of H3K27me3 [58, 68]. However, several evidence of H3K27me3 overlapping with H3K9me2 and H3K9me3 modifications at

developmentally repressed genes is suggestive of a cooperative transcriptional silencing mechanism of developmentally labile genes [69-71]. Genomic recruitment of PRC2 is modulated by the direct association with G9a/GLP (H3K9me2) [72] and PRC2 has been shown to be essential for the binding of HP1 to chromatin [73]. On the inactive X-chromosome H3K27me3 and H3K9me2 modifications complementary functions are mediated by both PRC2 and G9a binding to a common protein CDYL [74].

2.2.2 Active histone modifications:

Active histone modifications facilitate transcription by establishing an open chromatin structure. These modifications are H3K27ac, H3K4me1/2/3, H3 and H4 acetylation, H3K79me3, H2BK120u1 and H3K36me3. Positive crosstalk mechanisms between these modifications is essential in the recruitment and maintenance of these modifications at active genes. The histone modifications are catalysed by specific enzymes such as histone methyltransferases (HMTs), histone acetyltransferase (HATs) and histone deacetylases (HDACs).

H3K4me3 and H3K36me3

Actively transcribed genes in general have H3K4me2/H3K4me3 over gene promoters and a high level of H3K36me3 over gene bodies. In mammals the highly conserved COMPASS (complex of proteins associated with Set1) family of methyltransferases catalyses the methylation at Lys4 of histone H3 (H3K4) [75], a mark associated with transcriptionally active chromatin at the promoter and TSS regions.. COMPASS comprises of one of the six related proteins SETD1A, SETD1B, MLL1, MLL2, MLL3, and MLL4 having the SETD1/MLL catalytic subunits and four core subunits WDR5, RBBP5, ASH2L, and DPY30, along with additional complex-specific subunits [60]. Recruitment of Set1 histone methylase in Yeast to targeted 5′ portion of active genes by the interaction of Set1 with the phosphorylated form of Pol II CTD results in a promoter associated higher levels of H3K4me3 [76]. In mammals, higher levels of H3K4me3 levels at active gene promoters is preferentially mediated by CFP1, while H3K4me3 levels on lowly expressed CpG islands containing gene promoters are maintained via MLL2. Developmentally regulated gene promoters in ESCs, are bivalent containing both the repressive H3K27me3 mark as well as active H3K4me3. Evidences has suggested that H3K4me3 and H3/H4 acetylation not only coexist at the promoter and TSS of active genes but that H3K4me3 also promotes downstream H3/H4 acetylation through recruitment of HATs. Histone H3 tails with pre-existing H3K4me3 exhibit dynamic turnover of its lysine acetylation through a combinatorial action of both HAT p300/CBP and HDAC. Since H3K4me3 associates with promoter prior to transcription initiation, H3K4me3-dependent co-targeting of both HATS and HDACs would

facilitate the dynamic turnover of histone acetylation and this cooperation is important in ensuring proper transcriptional regulation [60].

H3K36me3 distribution correlates strongly with transcribed regions of active genes with an increase in distribution towards the 3′ end of gene bodies [77]. Studies in yeast have demonstrated that association of Set2 with the elongating Ser2-phosphorylated CTD of Pol II, results in the predominant distribution over gene bodies of actively transcribing genes [78].

H3K4me2

Although both H3K4me2 and H3K4me3 have concordant enrichment patterns [79, 80], increasing evidences suggest that these two active chromatin specific histone modifications are not functionally equivalent [81-83]. H3K4me2 has been shown to associate specifically with activation, poising for activation or repression of developmentally labile promoters as well as transcriptionally active promoters [80, 84, 85]. For example, H3K4me2 marks have been shown to define transcription factor binding regions [86] and also marks tissue-specific gene regulation [87]. WDR5 plays an important role in assembling distinct histone-modifying complexes to stimulate the H3K4 methyltransferase activity of MLL proteins within the MLL/COMPASS complexes and also in efficient promoter targeting of NSL HAT complexes [88]. WDR5, as a core subunit of the MLL-SET1 (hCOMPASS) methyl transferase, has been shown to specifically interact with H3K4me2 and nucleate the assembly of MLL-SET1 complexes to catalyse the transition of H3K4me2 to H3K4me3 [81, 89]. Additionally, a fraction of certain regions that function as enhancers have correlated H3K4me2 modifications specifically the primed enhancer-like regions are marked with H3K4me1 and H3K4me2 and lack histone acetylation [90]. This H3K4me2 correlation with poised enhancers might also result from chromatin looping of enhancers to the H3K4me2 marked promoters which has already been validated in connection to developmental genes.

H3K4me1 and H3K27ac

Enhancers are demarcated by characteristically high ratio of H3K4me1 to H3K4me3, high H3K27ac and also by specific TF and co-factor binding, such as EP300 [80] producing RNAPII-mediated enhancer transcription of bidirectional, unspliced RNAs [91] named as eRNAs. Based on their length, a fraction of eRNAs can also be classified as lncRNAs. However, eRNAs are rarely spliced (5%) in comparison to 30% of lncRNAs being spliced. Enhancers are known to regulate expression of target genes both in *cis* and *trans* through chromatin looping to gene promoters. Active enhancers are characterized by the presence of both the H3K4me1 (sometimes also H3K4me2/3) and H3K27ac marks. Over 70% of H3K27ac-marked enhancers are active and positively affect transcription in vivo [92]. However, all active enhancers are not always marked with only H3K27ac modification, rather co-occupancy of H3K27ac with BRD4 along transcription of eRNAs can predict active enhancer more

accurately [93]. In general, a combination of H3K4me1/H3K27ac marks has been broadly utilized for epigenomic annotation of active enhancers in several cellular contexts. Both Mll3 and Mll4, two large nuclear proteins of the MLL/COMPASS family containing a C-terminal SET domain have been identified as the major enhancer specific H3K4 mono-methyltransferases. Mll3 and Mll4 binds at enhancers to promote the recruitment of coactivator, p300, which acetylates H3K27 to facilitate enhancer activation. Moreover, H3K4me1 in absence of H3K27ac and, in some cases along with the presence of repressive H3K27me3, has been associated with enhancer states that are repressed or poised/primed for activation [94, 95]. Presence of such bivalent marks over developmentally repressed or poised genes emphasizes the role of epigenetics in fine tuning of gene regulation and for the requirement of an effective cross talk between active and inactive chromatin modifiers during developmental.

In the next section I will introduce and discuss in detail the last, but not the least, player of epigenetic regulator of gene expression i.e. long non-coding RNAs. There are other categories of ncRNAs that has also been extensively documented to act as important epigenetic regulators. However, I will restrict my discussion to the role of lncRNAs as epigenetic regulators.

CHAPTER 3: LONG NON-CODING RNAs (LncRNAs)

LncRNAs are defined as transcripts which are greater than 200 nucleotides in length without any evident protein coding capacity. Just like a stable protein coding mRNAs, lncRNAs are also RNA polymerase II transcribed, capped, polyadenylated and are frequently spliced [96-99].

LncRNA annotation: Methodological advances in high throughput technologies complimented with improved computational approaches have enabled identification and annotation of more numbers of valid lncRNAs. FANTOM project initially sequenced cloned cDNA from different mouse tissues to identify 3,652 non-coding transcripts with confidence [2]. Later RefSeq, GENCODE and Ensembl annotated more lncRNAs based on the analysis from refined EST- and cDNA-sequencing data. Parallel to high throughput cloning based-sequencing approaches, tilling microarray based detection of transcribed regions offered greater sensitivity of detection, although this technology had its own limitation of bias in probe designs and suboptimal genome coverage [96, 97]. Information from genome wide chromatin maps of H3K4me3-H3K36me3 active marks were used to design DNA-based tilling probes that led to the identification of 1,600 long intervening non-coding RNAs (lincRNAs) that mapped to K4-K36 chromatin and were expressed in four mouse cell types. Similar tilling array based study identified 3,300 expressed lincRNAs in a human cell line that mapped to K4-K36 chromatin domain [96, 97]. Next major advancement was achieved with high-throughput sequencing of millions of short RNA fragments (RNA-seq) that yielded billions of strand specific paired end reads of 100-150 nt each of which could be easily reconstructed to identify even very lowly expressed transcripts [100]. High resolution mapping of precise transcription start sites using cap analysis of gene expression (CAGE-seq)[98] combined with genome wide annotation of polyadenylation sites using 3P-seq [99] have further strengthened the identification and annotation of stable, valid lncRNA transcripts. Thus, combinations of such independent evidences for transcription initiation, exon-intron structure and transcription termination enabled reliable identification of mature lncRNAs. At present, different databases based on alternative annotation methods have estimated the total number of annotated lncRNAs to be close to that of protein coding genes. Latest version of GENCODE have annotated 58,288 genes in total, of which only 19,836 code for known protein while the rest are non-coding genes with 15,778 of them being lncRNA genes.

LncRNA genomics: Over the year lncRNA catalogue from several species have been made available in public databases. Comparative studies have found that lncRNAs are generally shorter than coding genes, have fewer but longer exons on an average compared to exons of protein coding genes [101, 102]. Moreover, lncRNAs overlap with more repetitive elements as compared to mRNAs [103]. Chromatin modification patterns, transcriptional and splicing regulations have been found to be similar to mRNAs [100-102]. Recent studies have found that lncRNAs differ from mRNAs in a couple of aspects. First, in general lincRNAs have fewer histone marks and transcription factors bound to their promoters than mRNAs. H3K9me3 generally associated with transcriptional repression, was found to be enriched at promoters of tissue specific active lincRNA loci. Second, lncRNAs are less efficiently

spliced than mRNAs, probably owing to the weaker internal splicing signals and the lower U2AF65 binding in lncRNAs [104].

LncRNA genes are preferentially found to colocalize with a protein coding gene mostly within 10kb [105, 106]. Such a distance distribution resembles to that of adjacent protein coding genes, ruling out the argument that lncRNAs are a by-product of mRNA biogenesis [103]. Interestingly, transcription-factor (TF) genes which are critical developmental regulators are preferably surrounded by lncRNAs within close genomic proximity, and this is consistent across several vertebrate species from zebrafish to mouse and human [102, 103, 107]. Such an organization suggests *cis* regulation of TF genes by lncRNAs or co-expression of these gene pairs to act in coordination. Multiple evidences for either possibility suggest that both mechanisms are exploited by the genome to regulate expression of these TF genes.

LncRNA expression levels: Several studies have shown lncRNAs to have variable tissue expression and have been found to be more cell type specific than protein coding genes [100, 101]. Most of the lncRNAs are expressed in brain and testes. LncRNAs are in general expressed at a much lower level compared to median mRNA levels. It's not clear if inefficient transcription or more efficient degradation accounts for this difference in their steady state levels of expression. In this regard, two independent studies addressed lncRNA stability, using either pulse-chase analysis or using transcription inhibition. Both studies identified several unstable non-coding transcripts, however, both studies concluded that lncRNAs and mRNAs have comparable half-life distribution or stability, an indication which rules out differential degradation efficiency [108, 109].

LncRNA conservation: Compared to protein coding genes, lncRNA sequences are less conserved and might be evolving rapidly. Only 12% of mouse and human lncRNAs have sequence conservation among other species. The selection pressure has been found to be more (i.e. more conserved) over lncRNA exons and promoter regions than introns of lncRNAs and intergenic regions of the genome [97, 100, 101]. Existing approaches for comparing sequence conservation relies on stretches of high sequence conservation at the genomic level rather than at the RNA level. More comprehensive annotation of lncRNAs from other vertebrate species would enable study of sequence conservation at RNA level rather than limited genomic alignment based estimation. Interestingly, many lncRNAs have a conserved genomic organization and exon-intron structure without detectable sequence homology [103]. Perhaps these lncRNAs have conserved sequence dependent functions but the sequences are divergent enough to be detected by existing tools. For example, *Malat1* is a highly expressed, nuclear retained single-exon transcript that was originally considered as a mammalian-specific lncRNA based on sequence conservation by genomic alignment. Recently, *Malat1* orthologues has been found in syntenic genomic position near *Scyl1* in mammals, frogs and fish having the same length, expression pattern and exon-intron structure. However, apart from the 3´terminal region and a short 70 bases

segment homology the mammalian *Malat1* lacks any recognizable sequence conservation with the fish counterpart [103]. Taken together, it would be misleading at this point in predicting the functionality of any lncRNA based on mere sequence conservation.

LncRNA subtypes*:* LncRNAs can be subdivided into several classes based on their positional relationship to protein-coding genes and different mechanisms of processing such transcripts (**Fig 1**). Promoter upstream transcripts (PROMPTs), enhancer RNAs (eRNAs), long intervening/intergenic ncRNAs (lincRNAs) are transcribed either from promoter upstream regions, enhancers or intergenic regions respectively. Natural antisense transcripts (NATs) are transcribed from the opposite strand of protein-coding genes. On the other hand, many other lncRNAs are derived from long primary transcripts with unusual RNA processing pathways, resulting in RNA species with unexpected structures such as small nucleolar RNAs (snoRNA) and circular RNAs (circRNA).

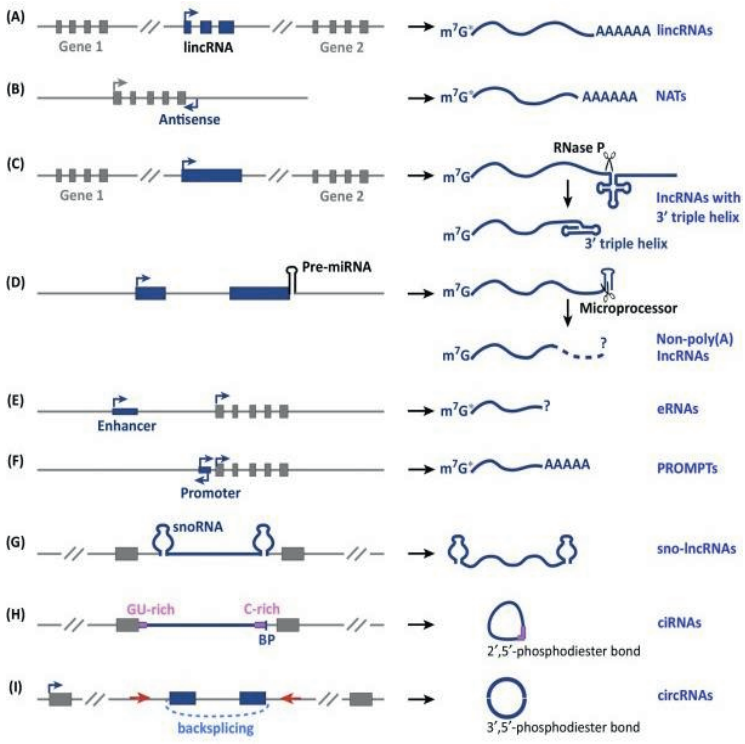

Figure 1. The Diversity of Long Noncoding RNAs (lncRNAs). (A) Large intergenic noncoding RNAs (lincRNAs). **(B)** Natural antisense transcripts (NATs). **(C and D)** lncRNAs with alternatively processed 3′ ends by ribonuclease P (RNase P; C) or microprocessor (D) cleavage. **(E)** Enhancer RNAs (eRNAs). **(F)** Promoter upstream transcripts (PROMPTs). **(G)** Small nucleolar RNA (snoRNA)-ended lncRNAs (sno-lncRNAs). **(H)** Circular intronic RNAs (ciRNAs). ciRNA formation depends on consensus RNA sequences (pink bars) to avoid debranching of the lariat intron. **(I)** Circular RNAs (circRNAs) produced from back splicing of exons. Inverted complementary sequences (red arrows) in introns flanking circularized exons promote circRNA biogenesis. (Adapted from ***Trends in Biochemical Sciences, September 2016, Vol. 41, No. 9,*** **licence number: 4223050570807)**

LncRNA subcellular localization: The function of lncRNAs depends on their subcellular localization. Contrary to the perceived notion of lncRNAs being predominantly nuclear several lncRNAs have been found to have mostly cytoplasmic distribution. Recent large-scale screening investigation using single molecule RNA fluorescence in *situ* hybridization of lncRNAs in human cell lines revealed that lncRNAs exhibited a wide range of subcellular localization patterns, both distinct nuclear localization patterns and localization patterns to both nucleus and cytoplasm [110]. LncRNAs can either (**Fig 2**)

- Accumulate in *cis* to acts in *cis* (**Fig 2A**). Example *Kcnq1ot1, HOTTIP* lncRNA.
- Accumulate in *cis* and act in *trans* (**Fig 2B**). Example *FIRRE* lncRNA.
- Localizes and acts in *trans* (**Fig 2C**). Example *HOTAIR* lncRNA.
- Enriched in nuclear bodies to act in *trans* (**Fig 2D**). Example *MALAT1*, *NEAT1* lncRNA.
- Exported to cytoplasm to act in *trans* (**Fig 2E**). Example *linc-MD1*, *lincRNA-p21*.

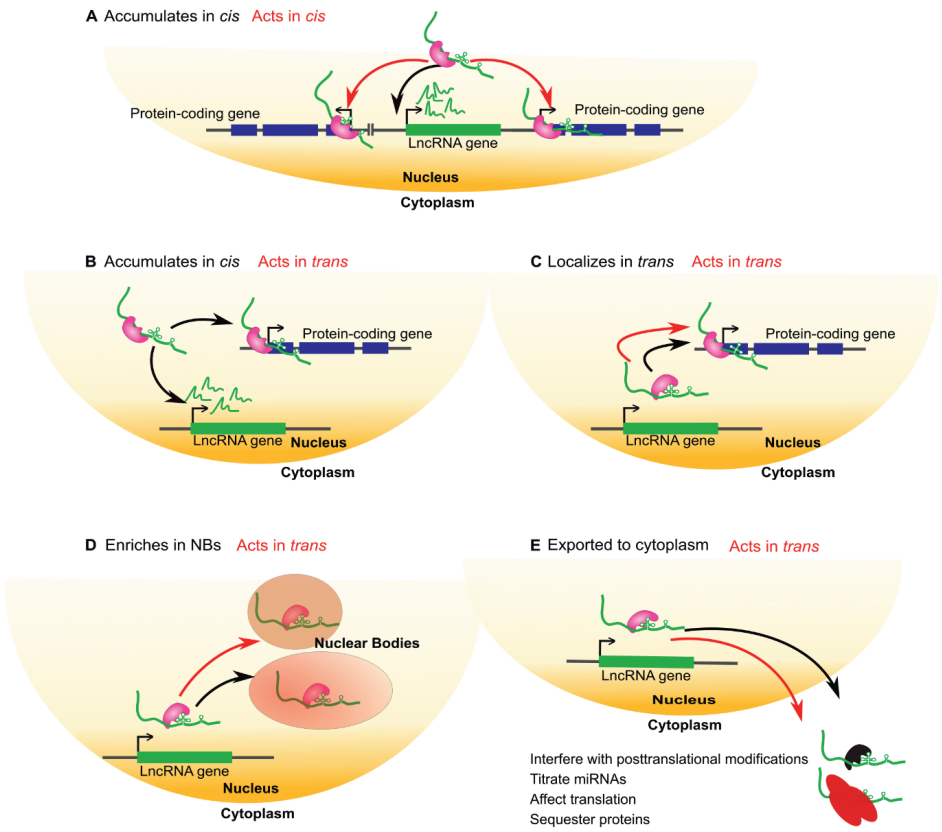

Figure 2. Functions of Long Noncoding RNAs (lncRNAs) associated with their subcellular fates

Mechanism of gene regulation by lncRNA: There is growing evidence for the repertoire of roles performed by lncRNAs in regulating gene expression and their function. Over the last few years, several reports have implicated the role of different lncRNAs in regulation of diverse biological functions. These lncRNAs have been proposed to act primarily via fine tuning the expressions of critical protein coding genes. While in most cases lncRNAs are implicated to play a critical role, the mechanism by which these lncRNAs exert their functions are not always clear and uniform. Rather than any universal mechanism, lncRNAs were in fact found to deploy a wide variety of mechanism in regulating gene expression of protein coding genes, which are as follows:

- The act of lncRNA transcription itself can regulate adjacent gene expression through recruitment of chromatin modifiers. Transcription through regulatory sequences to block its function is termed as transcriptional interference and can inhibit [111, 112] or activate [113] the expression of neighbouring genes.
- The lncRNA can function in *cis* by acting as a scaffold to recruit chromatin modifiers thereby creating a locus specific chromatin structure to activate or repress gene expression. Most well studied examples of this mechanism are lncRNAs like *Xist*, *Kcnq1ot1*, *HOTTIP*.
- LncRNAs can act in *trans* as a scaffold to recruit chromatin modifiers for activation or repression of gene expression. *HOTAIR*, *lincRNA-p21* are known to act in this manner .
- LncRNAs acting as a decoy to bind and sequester transcription factors away from their target chromosomal regions. *TERRA*, *PANDA*, *Gas5*, *MALAT1* lncRNAs are known to act in this manner.
- lncRNAs can act as nucleating domains for accumulation and formation of paraspeckles. *NEAT1*, *MALAT1* are known to act in this way.
- Cytoplasmic lncRNAs can pair with other RNAs and bind to RNA-binding factors to trigger posttranscriptional regulation. *TINCR*, *Linc-MD1* and *LincRNA-p21* are examples of this mechanism.

LncRNA as epigenetic modifier: Several studies have identified lncRNAs to be associated with known chromatin modifying proteins such as CTCF [114], YY1[115], Mediator [116], WDR5 [117-119], LSD1 [120], and PRC2 [121, 122]. With technological advances, genome wide searches for lncRNAs associated with a chromatin modifier by RIP-seq (RNA immunoprecipitation -sequencing) have identified thousands of associated lncRNAs to be associated with PRC2 [15, 122], WDR5 [123, 124] and also with a panel of 24 chromatin regulators and RNA binding proteins [124]. A significant fraction of annotated lncRNAs were found to be associated with each of these chromatin modifiers. While one hypothesis is that lncRNAs act as scaffolds to target these chromatin remodelers to specific genomic targets in order to modify active or inactive chromatin domains. Such lncRNA dependent recruitment is further corroborated by lack of any evidence supporting genomic sequence dependent

recruitment of chromatin remodelers, further supported by the inherent ability of lncRNAs to associate with chromatin and proteins. Apart from the above mentioned histone modifying complexes, several lncRNAs have also been reported to directly interact with DNMTs to play a critical role in regulation of DNA methylation mediated gene repression [18]. LncRNAs directly interact with active and inactive histone modifiers to exert histone modification mediated regulation of gene expression. Regulatory roles of lncRNAs have been implicated in almost all biological processes, where they have been identified as critical regulators of these processes. They contribute to the regulation of biological processes by fine tuning expression of critical protein coding genes. LncRNA interaction with DNA and histone epigenetic modifiers have been implicated in the regulation of important biological processes. In the next section, I will discuss how use of different model systems and alternative experimental strategies have helped elucidating the role of lncRNAs as epigenetic regulators.

3.1 LncRNAs in genomic imprinting

In sexually reproducing organisms, the offspring inherits two copies of the autosomal genes from the parents (paternal copy from father and maternal copy from mother), both of which are biallelically expressed. During early gametogenesis in mammals, some autosomal genes get epigenetically programmed to restrict their expression in a parent of origin-specific manner. Such an epigenetic regulation of parent of origin-specific monoallelic expression of genes is termed as genomic imprinting. Genomic imprinting provides an interesting model system to understand the mechanisms of epigenetic regulation because although both parental alleles have similar DNA sequences, they are differentially regulated for expression. Currently, more than 150 imprinted genes have been reported in human and mouse. Imprinted genes in mammals are associated with the following characteristic features [12]:

- Imprinted genes are often present in clusters to form an imprinted locus of size ranging from a few kilo bases to two to three mega bases, indicating a coordinated regulation of gene expression of the cluster.
- Mono allelic expression of the cluster is regulated by a common *cis*-acting control element called as Locus Control Region (ICR) which is typically 1–3 kb in size.
- Presence of at least one or two lncRNAs in the imprinted cluster as their partners with an inverse expression pattern to their protein-coding counterparts [125, 126].
- Promoters of these imprinted lncRNAs often map to differentially methylated regions (DMRs) which are part of the ICRs.
- DNA methylation and histone modifications coordinate the regulation of temporal monoallelic gene expression with lineage and tissue specificity.

Earlier studies have suggested that the *cis*-dependent local and long-distance gene regulatory mechanisms of genomic imprinting, is characterized by either:

1. ICR functioning as chromatin insulator [127-129] (**Fig 3**).
2. ICRs acting as promoters of lncRNAs [130] (**Fig 4**).

Neither the insulator nor the lncRNA based mechanisms alone can explain the monoallelic parent-of-origin-expression of genes. In this section, we will discuss with specific examples the functional roles of these insulators and lncRNAs in the establishment of parent-of-origin specific gene expression by focusing on antisense, intergenic, and enhancer-derived imprinted lncRNAs.

***H19/Igf2* locus:** This imprinted gene cluster is located on mouse chromosome 7 while its human orthologue is located on chromosome 11. The 2.3 kb long H19 lncRNA, one of the first identified imprinted genes, [131] is expressed from the maternal allele due to silencing by CpG methylation at the promoter of its paternal allele. Insulin growth factor 2 (Igf2) gene which is located 90 kb upstream of H19 gene is expressed from the paternal allele. ICR of this imprinted cluster is located upstream of H19 gene while a tissue specific enhancer common to both these genes is located downstream of H19 gene. This ICR has multiple binding motifs for insulator protein CTCF, the binding of CTCF to the maternal allele blocks the Igf2 promoter from accessing downstream enhancer element, thereby repressing maternal Igf2 expression. This allows the H19 maternal allele to access the enhancer and express. The ICR is methylated on paternal allele, which inhibits the binding of CTCF thereby allowing paternal Igf2 promoter to access the downstream enhancer [129]. Thus, the parent of origin-specific DNA methylation of the cis-acting ICR allele dictates the CTCF dependent chromatin insulator mechanism for regulation of monoallelic expression at this imprinted cluster (**Fig 3**).

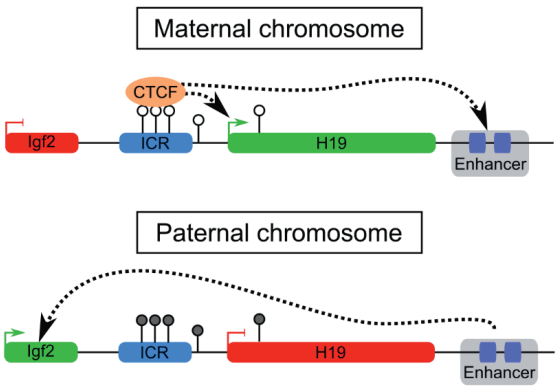

Fig 3. ICR functioning as chromatin insulator. CTCF-dependent chromatin insulation mechanism regulating the parent-of-origin-specific expression of the *H19* lncRNA by controlling the activity of downstream enhancers. Paternal methylation of ICR inhibits CTCF binding to repress *H19*, enabling the enhancer to activate *Igf2* in *cis*.

Additionally, *H19* lncRNA has been shown to be a part of an imprinted gene network (IGN) comprising of 16 co-expressing imprinted genes, including Igf2, Igf2r, and Cdkn1c [132]. Recent investigation demonstrated *H19* to interacts with methyl-CpG-binding protein MBD1, to recruit H3K9 methyltransferase to DMRs of some of IGN members, including Igf2, Slc38a4, and Peg1[132]. Such

H19 lncRNA-dependent recruitment of MBD1-H3K9HMT to DMRs on both the parental alleles fine-tunes the biallelic expression of IGN members, rather than controlling their monoallelic expression. Collectively, maternally expressed *H19* lncRNA can involve in *cis* monoallelic regulation of 90k upstream *Igf2* gene and can also act in *trans* to recruit chromatin modifying complexes to fine tune the repression of biallelically expressed group of genes.

Interestingly, *H19* lncRNA has been characterized with additional functional roles apart from above-mentioned control of embryonic growth in mice through epigenetic regulation of imprinted gene or gene clusters. These additional functional roles include, acting as

- Competing endogenous RNA (CeRNA) or as a scaffold in *let-7* microRNA dependent inhibition of myogenic differentiation of multipotent mesenchymal stem cells [133, 134].
- miRNA precursor (exon 1 of the transcript) for miR-675-3p and miR-675-5p, to promote myogenic differentiation [135].
- Tumour suppressor [136, 137] as well as an oncogene [138, 139].

***Kcnq1* locus:** *Kcnq1ot1* is a 91kb long nuclear enriched, RNA Polymerase II (RNAPII) transcribed noncoding transcript that is antisense to the *Kncq1* gene [140] which maps to *Kcnq1*/*Cdkn1c* imprinted chromosomal domain at the distal end of mouse chromosome7 and has orthologous region on human chromosome 11p22. This domain harbours 10-12 imprinted genes, which are maternally expressed whereas *Kcnq1ot1* is paternally expressed since its promoter is methylated on the maternal chromosome [141]. The paternal expression of *Kcnq1ot1* and silencing of paternal alleles of the imprinted genes in this cluster is suggestive of a *cis*-acting regulatory framework. The repression of genes in this cluster and the *Kcnq1ot1* mediated mechanism of repression is tissue and developmental stage dependent [125] and accordingly is classified into three categories of imprinted genes:

- Ubiquitously imprinted genes (UIGs) are imprinted in both embryonic and extra embryonic tissues. Example: *Kcnq1*, *Cdkn1c*, *Slc22a18* and *Phlda2*.
- Placental imprinted genes (PIGs) that are repressed only in placental tissues. Example *Osbpl5*, *Ascl2*, *Tspan32*, *Cd81* and *Tssc4*.
- Non-imprinted genes (NIGs) which escape the RNA mediated silencing. Example: *Cars1* and *Nap1l4*.

Experiments using an episomal system identified an 890 base pairs silencing domain at the 5' end of mouse *Kcnq1ot1* RNA which harbours several conserved repeats, to be crucial for mediating silencing of genes [142]. Deletion, truncation and destabilization of the mouse *Kcnq1ot1* RNA in both transgenic mice and episomal systems have argued for the requirement of RNA *per se* in the allelic transcriptional repression of protein coding genes in *cis* [140, 141, 143]. *Kcnq1ot1* transcript from the paternal chromosome, interacts with DNA (DNMT1) [144-146] and chromatin modifiers (EZH2 and G9a) [140] to recruit them in *cis* and silence both UIGs and PIGs (**Fig 4**).

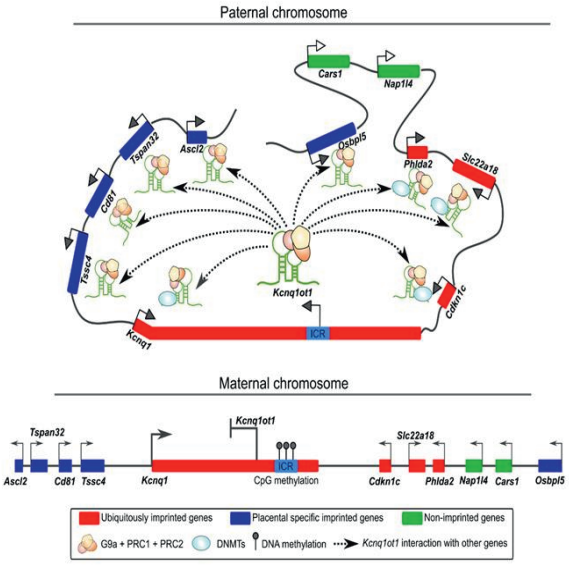

Fig 4. ICRs acting as promoters of lncRNAs. Paternal ICR is unmethylated. Paternally expressed *Kcnq1ot1* lncRNA promotes long-range chromosomal interactions in *cis* through recruitment of DNA modifiers and chromatin to establish higher-order repressive chromatin compartment. Maternal DNA methylation of the ICR silences *Kcnq1ot1* lncRNA expression resulting in the activation of all the target imprinted protein-coding genes. **(**Modified from ***C. Kanduri / Biochimica et Biophysica Acta 1859 (2016) 102–111*****)**

Studies have demonstrated that 890 base pairs silencing domain of *Kcnq1ot1* RNA is essential for both chromatin localization and also for recruitment of DNMT1 for maintaining CpG methylation of somatic DMRs [145] which acquire methylation during post implantation development. This RNA-dependent recruitment of DNMT1 has a functional role only in the maintenance of silencing of ubiquitously imprinted genes [147]. Silencing of PIGs, on the other hand, was shown to be regulated primarily by repressive histone modifications. Knockout studies in mice lacking histone modifiers such as G9a, Eed and Ezh2 have shown to influence silencing of PIGs. *Kcnq1ot1* RNA is required for initiation of silencing of both UIGs and PIGs, and is also required for maintaining the silencing of only UIGs but not for the maintenance of silencing for PIGs [147].

A recent investigation devised an RNA-guided chromosome conformation capture approach (3C; R3C) to unravel the role of *Kcnq1ot1* in long-range gene silencing mediated through establishment of higher order intrachromosomal interactions. Paternally expressed *Kcnq1ot1* RNA was demonstrated to induce *Kcnq1* silencing by promoting higher-order intra chromosomal interactions between the *Kcnq1*

promoter and the *Kcnq1* ICR (also known as KVDMR1) specifically on the paternal chromosome. Continuous presence of *Kcnq1ot1* RNA is required for these higher-order intrachromosomal interactions, as Zinc Finger technology mediated targeted *Kcnq1ot1* promoter methylation or downregulation of *Kcnq1ot1* RNA results in loss of long-range intrachromosomal interactions [148]. Taken together, above observations suggest that chromatin targeting of *Kcnq1ot1* RNA to imprinted promoter regions with DNMT1 or repressive histone modifiers is essential to mediate lineage-specific transcriptional silencing mechanisms to initiate and maintain the silencing of ubiquitously and placental-specific imprinted genes (**Fig 4**).

***Igf2r* locus:** The *Igf2r* imprinted locus present on mouse chromosome 17 is composed of three maternally expressed protein coding genes *Igf2r*, *Slc22a2* and *Slc22a3* and a paternally expressed 108kb long unspliced, polyadenylated non-coding RNA called as *Air* that is transcribed antisense to its host gene *Igf2r*. Like *Kncq1* locus, the ICR on the paternal allele is unmethylated to serve as a promoter for *Air* expression and this correlates with the paternal allele silencing of the three imprinted protein coding genes [149]. *Air* lncRNA has also been shown to interact with and recruit G9a (H3K9 methyltransferase) to the *Slc22a2* and *Slc22a3* gene promoter to mediate epigenetic lineage specific transcriptional repression [150]. The mechanism of *Air* lncRNA mediated allelic silencing of *Igf2r* is dependent both on the act of transcription rather than the transcript per say. This was supported by the observation that act of unmethylated paternal ICR drives *Air* transcription in *cis* through the promoter of *Igf2r* resulting in the eviction of pre-initiation RNAPII complex from promoter region. Whereas maternally methylated ICR restricts the transcription of *Air* lncRNA thereby allowing *Igf2r* expression [112, 150].

***Nespas* locus:** Paternally expressed lncRNA *Nespas*, transcribed antisense to *Nesp* gene, belongs to the *Gnas* imprinted cluster. The ICR of *Gnas* cluster contains two promoters, one each for *Gnasxl* transcript and the *Nespas* lncRNA, and maps to a differentially methylated region (*Nesas-Gnaxl* DMR). ICR is unmethylated on the paternal allele allowing *Nespas* to transcribe through the *Nesp* promoter in the antisense direction, inversely correlating with the transcription of the sense gene *Nesp* [151]. This was found to correlate with increased CpG methylation and reduced H3K4me3 levels at the *Nesp* promoter, while loss of *Nespas* transcription correlated with loss of CpG methylation and increased H3K4me3. Above observations collectively suggest that the act of *Nespas* antisense transcription through the *Nesp* promoter recruit histone demethylases such as KDM1B, resulting in demethylation of *Nesp* promoter H3K4me3 which in turn promotes CpG methylation [152].

3.2 LncRNAs in Dosage compensation

Dosage compensation is the mechanism by which organisms balances the expression of genes from dissimilar sex chromosomes to equalize their expression in both sexes. *Drosophila* and mouse provide two excellent examples where the mechanism of dosage compensation has been investigated in greater details. In both cases the critical role of lncRNAs in dosage compensation has been elucidated providing one of the best understood model for lncRNA mediated epigenetic gene regulation.

***Drosophila* dosage compensation via *roX* RNAs**: *Drosophila* males have single X-chromosome while females have two X chromosomes. The balance of X-linked gene expression between two sexes is achieved in this case by hyperactivation of the single male X-chromosome. Two lncRNAs differing in sequence and length are transcribed from the X-chromosome called as *roX1* and *roX2*. These two lncRNA combine with male specific MSL ribonucleoprotein complex MSL to form the dosage compensation complex (DCC) which localizes exclusively to the transcribed genes on male X-chromosome. This targeting of DCC results in acetylation of H4 lysin16 to establish active chromatin marks to induce transcriptional hyperactivation of genes from the single male X-chromosome [153]. Helicase activity of MLE remodels the folded conformation of the *roX* RNAs thereby unmasking the binding sites for MSL2 to trigger the assembly of DCC required to catalyse X chromosome-wide acetylation of histone H4K16 [154]. A recent study has integrated the *roX* lncRNA/DCC model of X-chromosome targeting mechanism with high-affinity targeting motifs for the MSL complex that are enriched around boundaries of topologically associated domains (TADs) to understand the nucleation and spreading mechanism of *roX* lncRNA/DCC in greater detail [155]. In summary, *roX* lncRNAs provide an interesting example of lncRNA mediated epigenetic regulation in activation of gene expression.

Mammalian dosage compensation: Mammalian males have single X-chromosome while females have two X-chromosomes. The balance of X-linked gene expression between two sexes is achieved in this case by inactivation of one of the female X-chromosomes which is a random process in almost all tissues except extra-embryonic tissues where always the paternal X-chromosome (Xp) gets inactivated. This X-chromosome inactivation (XCI) process is mediated by the orchestrated transcription of several lncRNAs from the X-inactivation centre (XIC) of the future inactive X-chromosome (Xi) [156].
XCI is primarily mediated by a lncRNA called *Xist*, transcribed from the XIC to coat the Xi-chromosome in *cis,* inducing PRC2 mediated H3K27me3 deposition and other repressive chromatin modifications [157]. The association with chromatin and consequent silencing by *Xist* are mediated by independent regions within this RNA. While a chromatin binding domain mediates it chromosomal

coating, the A-repeat sequence in the RNA is transcribed as a smaller *RepA* lncRNA mediates its silencing by interacting with PRC2 complex member EZH2 [158].

Tsix lncRNA transcribed antisense to *Xist* from the active X-chromosome (Xa) represses the expression of *Xist* by DNMT3a mediated methylation at the *Xist* promoter. Loss of *Tsix* results in activation of *Xist* even from the Xa [159, 160]. Unlike *Txis* which antagonizes *Xist* expression, another lncRNA called *jpx/enox* positively regulates *Xist* expression thereby forming two parallel antagonistic RNA switches regulating the *Xist* expression [161]. *Xact* lncRNA that antagonizes *Xist* expression is transcribed from the Xa to coat this chromosome in *cis* [162]. The requirement of *Firre* lncRNA mediated tethering of Xa to the nucleolus also highlights the essential role of lncRNA maintaining higher order chromatin structure [163]. Taken together, dosage compensation in mammals display intricate interplay between several lncRNAs. An orchestrated transcriptional regulation of these lncRNAs epigenetically regulates X-chromosome inactivation process in mouse.

3.3 LncRNAs in Cell-Fate Programming and Reprogramming

One of the most fascinating insight into the regulatory roles of lncRNAs come from studies with cell programming and differentiation based models. LncRNAs have been found to be crucial players involved in the specification, self-renewal, differentiation of stem cells and the differentiated progenies. Following section, I will discuss about the regulatory role of some lncRNAs in cellular differentiation.

***TINCR* lncRNA**: Khavari and colleagues, while examining the pattern of gene expression during epidermal differentiation, discovered two lncRNAs *ANCR* and *TINCR*, which are expressed in epidermal stem cells and their terminally differentiated progeny, respectively. While anti-differentiation noncoding RNA (*ANCR*) provides example of an lncRNA that inhibit differentiation, *TINCR* (terminal differentiation-induced ncRNA) is a 3.7-kilobase lncRNA localized in the cytoplasm that controls human epidermal differentiation by post-transcriptional mechanism. Genome-scale RNA (RIA-seq) interactome identified *TINCR* to interact with a range of mRNAs implicated in epidermal differentiation. A 25-nucleotide 'TINCR box' motif was strongly enriched in the interacting mRNAs and was essential to mediate the *TINCR*–mRNA interaction. A high-throughput screen revealed direct binding of *TINCR* RNA to the staufen1 (STAU1) protein. STAU1 deficient epidermal and TINCR depletion both impaired differentiation. *TINCR*–STAU1 complex mediates stabilization of differentiation mRNAs, such as *KRT80*. *TINCR* lncRNA mediates somatic tissue differentiation via binding to the differentiation specific mRNAs to ensure their expression [164].

LincRNA-p21: Using a OSKM induced somatic cell reprogramming system (ectopic expression of defined factors such as Oct4, Sox2, Klf4 and c-Myc) Bao et al. found that *lincRNA-p21* is induced by p53 that act as a barrier for cell reprogramming without affecting cell apoptosis or cellular senescence. *LincRNAp21* interacts with H3K9 methyl transferase SETDB1 to maintain repressive histone mark H3K9me3 at a subset of pluripotency gene promoters, and also interacts with DNMT1 to methylate the promoters of a different set of pluripotent genes. Mechanistically, both of these interactions were found to be dependent on the RNA binding protein hnRNPk, the knockdown of which attenuated the association between *lincRNA-p21* with DNMT1 or SETDB1 and thus resulted in enhanced reprogramming efficiency. This is an example where a RBP (i.e hnRNPk) serves as a facilitating platform for the functional interaction between a lncRNA and two different inactive epigenetic regulators that is required for efficient somatic cellular reprogramming [165].

***Bvht* lncRNA**: Mouse embryonic stem cell (ESC) based differentiation study, identified a heart-associated lncRNA, Braveheart (*Bvht*) as an important factor in the commitment of cardiac lineage since the depletion of this lncRNA severely reduced the number of beating cardiomyocytes formed during embryoid body differentiation. *Bvht* lncRNA was found to interact with SUZ12, a component of PRC2 and acts in *trans* to activate the heart specific differentiation master driver *Mes*P1 gene. This study provides evidence for the role of lncRNA mediated epigenetic regulation in the establishment of the cardiovascular lineage during mammalian development [166].

Another study experimentally determined the secondary structure of Braveheart (*Bvht*) in mESCs using SHAPE and DMS probing methods to find that this transcript has a 590 nucleotide region with a modular property. CRISPR/Cas9-mediated genomic editing identified that deletion of 11 nt of *Bvht* [5′ asymmetric G-rich internal loop (AGIL)] impairs cardiomyocyte differentiation. Specific interaction between a zinc-finger protein CNBP/ZNF9 and this AGIL loop in the *Bvht* lncRNA is essential to maintain cardiomyocyte differentiation capacity [167]. *Bvht* interaction with CNBP through a well-defined RNA motif to regulate cardiovascular lineage commitment provides supportive mechanistic detail of epigenetic regulation.

***Fendrr* lncRNA**: *Fendrr* lncRNA is expressed in the caudal end of lateral plate mesoderm which eventually gives rise to heart and body wall development. Insertion of a premature polyadenylation (polyA) signal into the mouse *Fendrr* locus promoted a depletion of the full length *Fendrr* lncRNA(KO) that resulted in embryonic lethality at embryonic day 13.5 accompanied by defects of the abdominal wall and heart. *Fendrr* binds to both PRC2 and TrxG/MLL chromatin complexes to exert a dual epigenetic regulation of gene expression In *Fendrr* KO embryos, upregulation in the expression of several lateral plate or cardiac mesoderm differentiation controlling transcription factors coincided with

reduced PRC2 and H3K27me3 occupancy and with or without an increase in H3K4me3 at their promoters, [118].

***Linc-MD1* lncRNA**: *Linc-MD1* was identified to promote muscle differentiation through its ability to sponge out miR-133 and miR-135 thereby preventing their mRNA targets MAML1 and MEF2C which are known transcriptional activators of muscle differentiation. Thus *Linc-MD1* acts as a competing endogenous RNA (ceRNA) for miR-133 and miR-15335. *Linc-MD1* is also a host transcript to a repressor of muscle differentiation miRNA called miR-133b [168]. Additional study has identified that binding of HuR (an RNA Binding Protein, RBP) to *Linc-MD1* prevents the maturation of miR-133b and maintains sponging of miR-133/miR-135 to promote differentiation. Cellular depletion of HuR favours processing of *Linc-MD1* into miR-133b along with lower ceRNA functioning, thereby also promoting the miR-133/miR-135 mediated inhibition of muscle differentiation [169]. This also provides a dual mechanism employed by a single lncRNA in controlling different targets to achieve the same function, in this triggering of muscle differentiation. This also raises the possibility of an additional layer of complex regulatory network that can operate amongst non-coding RNAs in a positive or negative feedback loop to regulate a biological process.

***NoRC-Associated RNA (Promoter-Associated RNA, pRNA)*:** Maturation of 250~300 nucleotide long transcripts from a spacer promoter located 2 kb upstream of the pre-ribosomal RNA transcription start site have been shown to be important in epigenetic silencing ribosomal RNA gene (rDNA) [170]. About 150–300 nt of RNA sequences, named as *pRNA*, that are complementary to the rDNA promoter interacts with TIP5 the large subunit of chromatin remodelling complex NoRC. This interaction of *pRNA* with TIP5 guides NoRC targeting to rDNA promoter to promote H3K9&H4K20 methylation and HP1 recruitment. In *vitro* studies also described the role and mechanism of *pRNAs* in human [171]. *pRNAs* regulate transcription of rRNA genes by forming stable RNA-DNA triplex at the target promoter sequence which is then specifically recognized by DNA methyltransferase DNMT3b, the enzyme that promotes heterochromatin formation in the rDNA promoter by DNA hypermethylation to subsequently silence the rRNA genes [172]. In stem cells, interaction of mature *pRNAs* with the nucleolar transcription terminator factor 1 (TTF1) and TTF1-interacting protein 5 (TIP5) has been found to be essential for the generation of heterochromatic rDNA required for exit from pluripotency during ESC differentiation [173]. This indicates that *pRNA*s forms RNA-DNA triplex structures to mediate epigenetic silencing of rDNA genes, required to promote ESC differentiation.

3.4 LncRNAs in pattern formation

Development in multicellular organisms is an orchestrated process during which cells with identical genotype attain different cellular identities based on their three-dimensional positioning in the embryo. This spatio-temporal regulation of cell fate during development is called pattern formation. From Drosophila to mammals, lncRNAs employs a variety of mechanisms to regulate the expression of evolutionarily conserved homeotic gene cassette (*HOX* genes) to accomplish morphologically distinct body plan. In mammals 39 *HOX* genes are organized on different chromosomes in four clusters called *HOXA-HOXD.* The spatial organization of the genes should correspond to their temporal expression pattern for regulating anterior-posterior axis formation during embryogenesis. Differential expression of *HOX* genes is under the regulation of both active and inactive epigenetic modifiers mediated by several lncRNAs.

DNA microarray based tilling all the four human HOX loci at five base pair (bp) resolution and a 2 mega base control regions was used to profile polyadenylated transcripts from fibroblasts representing 11 distinct positional identities [121]. This led to the identification of 231 *HOX* ncRNAs, 64% of which also exhibited *HOX* like distinct spatio-temporal expression patterns. They also identified *HOTAIR*, a 2.2kb long lncRNA in the *HOXC* locus, which was found to exert transcriptional repression in *trans* at the *HOXD* locus via interaction with PRC2. *HOTAIR*–PRC2 interaction enhanced the PRC2 activity at *HOXD* locus and was required for both PRC2 and histone H3K27me3 occupancy at the HOXD locus. Thus, lncRNAs were found to exert epigenetic regulation by their ability to demarcate active and inactive chromatin domains essential for fine tuning the spatio-temporal expression of developmental genes [121]. PRC2 was by then found to be a methyltransferase that trimethylates H3K27 to repress transcription of specific genes [53, 174]. *HOTAIR* lincRNA was found to serve as a scaffold for two distinct histone modification complexes. A 5′ region of HOTAIR transcript binds to PRC2 while a 3′ region of *HOTAIR* binds the LSD1/CoREST/REST complex. The ability to tether two distinct complexes enables RNA-mediated assembly of PRC2 and LSD1, and coordinates targeting of PRC2 and LSD1 to chromatin for coupled histone H3K27 methylation and H3K4 demethylation [120]. This observation supported the *trans* acting scaffolding role of lncRNAs for the efficient assembly and co-targeting of different functionally complimenting protein complexes to target chromatin region.

High throughput chromosome conformation capture (5C) across the *HOXA* locus along with H3K4me3, H3K27me3 and Pol II ChIP-seq data based analysis lead to the identification of *HOTTIP* (HOXA transcript at the distal tip) with bivalent H3K4me3 and H3K27me3 occupancy characteristic of poised regulatory sequences. The 3.7kb long, spliced, polyadenylated *HOTTIP* transcribed antisense to *HOXA13* gene that was found to coordinate activation of several 5′ *HOXA* genes (five genes found to be regulated) in *vivo*. This is regulated via *HOTTIP* directly binding to and recruiting the WDR5/MLL complex to the target promoters through chromosomal looping to drive H3K 4 tri-methylation mediated

gene transcriptional activation [175]. *HOTTIP* provides a rare example of lncRNA involved in the activation of gene expression by interacting only with active chromatin modifiers.

3.5 LncRNAs in Disease

In this section I would like to discuss about few lncRNAs that have been implicated to have a regulatory role in diseases pathogenesis. I would try to focus how use of such disease model helped in better understanding of the mechanism of lncRNA mediated epigenetic gene regulation.

***ANRIL* lncRNA**: A 42kb germline deletion encompassing the *INK4b* locus containing three tumour suppressor genes; *p15*/*CDKN2B*/*INK4b*, *p16*/*CDKN2A*/*INK4a* and *p14*/*ARF* and antisense lncRNA *ANRIL* was first reported from genetic analysis of familial melanoma-neural tumours [176]. In many other tumours these genes have been found to have altered expression with disease associated SNP being mapped to the *ANRIL* gene [177]. *ANRIL* lncRNA transcribed antisense to *INK4* gene and was found to interact with a member of PRC1 complex called chromobox 7 (CBX7) protein to transcriptionally repress the expression of *INK4a*. In prostate cancer patients, elevated levels of both *ANRIL* and CBX7 correlates with reduced expression of *INK4a.* Additional study has also identified *ANRIL* to interact with SUZ12 thereby recruiting PRC2 to silence *INK4a*. This lncRNA was thus shown to regulate epigenetic repression of a tumour suppressor gene by interacting with both PRC1 and PRC2 complexes. However, the detailed mechanism and role of this lncRNA in the disease pathogenesis has not been investigated in detail.

***HOTAIR* lncRNA**: *HOTAIR* expression was found to be upregulated in primary breast tumours and its expression level in primary tumours was used as a predictor of eventual metastasis and death. Overexpression of *HOTAIR* in epithelial cancer cells induced genome-wide re-targeting PRC2 leading to altered H3K27me3 modification, gene expression resulting in increased cancer invasiveness and metastasis. Loss of *HOTAIR* conversely, inhibited cancer invasiveness [178]. However, the PRC2 dependent *trans* acting regulation of *HOTAIR* could not be validated in MDA MB-231breast cancer cell line studies [179]. Using RNA-tethering system to investigated *HOTAIR*-PRC2 interaction in gene silencing, artificial tethering of *HOTAIR* to ectopic target chromatin caused PRC2 independent transcriptional repression. Also, overexpression effect of *HOTAIR* was subtle but PRC2 independent. PRC2 recruitment was a consequence of gene silencing. This study raises the possibility of PRC2 function independent of RNA-mediated chromatin targeting. In addition to breast cancer, *HOTAIR* overexpression has also been associated with poor prognosis in 17 other cancer types [180] where the

oncogenic function of this lncRNA is mediated via PCR2 recruitment to epigenetically repress genes involved in several processes.

***MALAT1* lncRNA**: *MALAT1* was first reported as a prognostic parameter for patient survival and risk of developing metastasis in stage I non-small cell lung cancers [181]. *MALAT1* transcript is a highly expressed nuclear enriched lncRNA transcribed from chromosome 11q13 having high sequence-conservation across species. *MALAT1* localizes to nuclear speckles although its does not regulate the integrity of these speckles, rather it interacts with pre-splicing proteins factors (SR proteins) to regulate their abundance and distribution in the nuclear speckles [182]. In addition to this, *MALAT1* has been shown to promote cell proliferation by modulating the levels of cell cycle regulated oncogenic transcription factor B-MYB [183]. Taken together these observations suggest that *MALAT1* is acting as a molecular sponge to titrate away pre-mRNA splicing factors leading to misregulation of cell cycle progression genes resulting in tumour progression.

Overexpression of *MALAT1* has been reported in various cancers and in most cases expression is associated with higher cell proliferation and metastasis. Several molecular mechanisms have been proposed for the role of *MALAT1* in cancer pathogenies. This is because although it is overexpressed in 17 types of cancer, the target genes in each of these are rather varying depending on cell type specificity. For example, in bladder cancer, *MALAT1* associates with SUZ12 to induce N-cadherin expression, concomitantly reducing E-cadherin expression, such that these cooperative changes results in the malignancy of the tumour cells. In renal carcinoma cells, MALAT1 gene locus is translocated to that of transcription factor EB (*TFEB*) gene, to express a fusion *MALAT1* transcript. The *MALAT1*–TFEB fusion causes the upregulation of *TFEB* transcription, which consequently results in oncogenic function [184]. *MALAT1* acting as an oncogene in various cancers has been implied to function in the control of both alternative splicing and gene expression. Although several reports have described the implication of *MALAT1* in cancers, the mechanism of action in cancer pathogenesis is not yet understood.

***PCAT* lncRNAs**: Use of whole transcriptomic profiling approach with RNA from a cohort of 102 prostate cancer patients identified 121 unannotated prostate cancer associated transcripts (*PCATs*). *PCAT-1* was identified as the most upregulated lncRNA in prostate cancer tissues, and independent cohort studies have also confirmed the overexpression of *PCAT-1* in prostate cancer [185]. Additionally, *PCAT-1* has been implicated as a prognostic biomarker for poor patient survival and colorectal cancer metastasis. A study in prostate cancer, has shown that *PCAT-1* promotes cell proliferation through up-regulation of the cMyc protein (encoded by the *MYC* gene) by stabilization of cMyc post-transcriptionally by interfering with the regulation of *MYC* by miR-34a at the *MYC* 3′ untranslated region (UTR). This explains a mechanism of cMyc stabilization in prostate cancer and

illustrate a post-transcriptional function of *PCAT-1* in the cytoplasmic compartment to antagonize microRNAs [186]. A fraction of *PCAT-1* transcripts localizes to the nucleus where it interacts with PRC2 to repress gene expression, while most *PCAT-1* transcripts that are in the cytoplasm likely operates independently of PRC2 in this cellular compartment. Cytoplasmic fraction of *PCAT-1* transcript has also been shown to repress the expression of tumour suppressor gene *BRCA2*.

Another prostate cancer associated lncRNA *PCAT-114* termed as *SChLAP1* (*S*econd *C*hromosome *L*ocus *A*ssociated with *P*rostate-*1*) is overexpressed only in a subset of prostate cancers. *SChLAP1* levels is a strong predictor of metastasis and prostate cancer specific mortality. Both *in vitro* and *in vivo* gain- and loss-of-function experiments validated that *SChLAP1* contributes to invasiveness and metastasis of cancer cell. *SChLAP1* was found to antagonize the genome-wide occupancy and regulatory functions of the SWI/SNF chromatin-modifying complex. These observations suggest that *SChLAP1* promotes the development of cancer lethality partly by antagonizing tumour-suppressive functions of the SWI/SNF complex [187].

***Pint* lncRNA**: One of the initial experimentally validated connection between p53 pathway and epigenetic regulation of Polycomb that is mediated by a lincRNA came from studies on ubiquitously expressed mouse lincRNA *Pint,* which is precisely regulated by p53 [188]. In mouse cells, *Pint* acts as a positive regulator of cell proliferation and survival by regulating genes in the *TGF-β*, *MAPK* and *p53* pathways. *Pint* human orthologue was also identified (*PINT*) to be regulated by p53, interacts directly with PRC2 to regulate target gene repression via PRC2 mediated H3K27 tri-methylation. Low expression in primary colon tumours and overexpression induced inhibition of proliferating tumours suggest a tumour suppressor role of *PINT*. This provides the evidence of p53 activated lncRNA with PRC2 mediated epigenetic silencing.

LincRNA-p21: In addition to lncRNAs regulating DNA methylation through physical association with DNMTs, some lncRNAs can modulate DNA methylation even through indirect interaction with other RNA binding proteins. In a screen for p53-regulated lncRNAs, Huarte M et al. identified *lincRNA-p21* as a repressor in the p53-dependent transcriptional responses [189]. Depletion of the *lincRNA-p21* affected the expression of p53 target genes that are normally repressed by p53. This transcriptional repression by *lincRNA-p21* was mediated through direct interaction with hnRNP-k. 780 nt in the 5` ends of the transcript was essential for this regulatory interaction that was required for proper genomic localization of hnRNP-k at target repressed genes involved in p53 mediated apoptosis. *LincRNA-p21* was thus a *trans* acting lncRNA with epigenetic regulatory role in repression of p53 target genes via interaction with hnRNP-k.

However, apart from in *trans* transcriptional repression mechanism, *lincRNA-p21* was also found to be posttranscriptional inhibitor of translation. Antagonistic to p53 induction, *lincRNA-p21* is negatively

regulated via let-7/Ago2 destabilization of the RNA in presence of RNA binding protein HuR. However, in absence of HuR, accumulation of *lincRNA-p21* in HeLa (Human cervical carcinoma) cells increases the association with JUNB and CTNNB1 mRNAs. This lncRNA-mRNA interaction selectively decreased the translation of corresponding mRNAs by inhibiting polysomes. Evidences from *lincRNA-p21* studies show that the same lncRNA can adopt different mechanisms to regulate gene expression depending on whether they are localized in the nuclear or cytoplasmic compartment.

***MEG3* lncRNA**: Maternally expressed *Meg3* lncRNA (maps to *Dlk1-Dio3* locus on mouse chromosome 12 and *DLK1-DIO3* locus on human chromosome 14) and other noncoding RNAs of *Dlk1-Dio3* locus have been identified as critical players in the reprogramming of mouse fibroblasts cells into iPS cells with *Meg3* expression acting as a marker for iPS. In several cancers, including gliomas, colorectal cancers, and pancreatic neuroendocrine cancers have aberrant *MEG3* expression [190]. In many of the cancers, *MEG3* is expressed at low levels, while its overexpression in cancer cell lines promotes apoptosis by inhibits cell proliferation. Recently, *TGF-β* pathway genes were identified to be direct targets of *MEG3*, which regulates these genes by directly binding and recruiting EZH2 to the promoter–distal regulatory regions. This target recognition by *MEG3* occurs via triplex formation between GA-rich sequences within *MEG3* and lncRNA GA-rich sequences of target genes [191]. This provides a possible mechanistic explanation for the *trans* action of *MEG3* lncRNA in cancer etiology.

***UBE3A-ATS* lncRNA**: Angelman syndrome (AS) is a severe neurodevelopmental disorder associated with speech impairment and intellectual disability. This is caused by biallelic activation of imprinted *UBE3A-ATS* causing biallelic silencing of *UBE3A*, an E3 ubiquitin ligase, leading to impairment of important neural-related functions such as synaptic development, signal transduction, and plasticity. Several mouse models, with either pre-mature termination of *UBE3A-ATS* or deletion of its promoter have been generated to study the role of *UBE3A-ATS* in AS [12].

LncRNA research is at a very interesting juncture. Diverse functional and mechanistic understanding of some well-studied examples in important and diverse aspects of biology suggest that many more of these might play critical roles. Latest research is focused on characterising the regulatory landscape of the genome with an integrative approach to identify the regulatory nexus between enhancers, promoters, lncRNAs, and lncRNA-binding proteins in the regulation of specific biological processes. To do so, experimental strategies are expanding for interrogating lincRNA structure, targets, localization, function and interaction by developing improved tools for comparative genomics and precise high-

throughput gene editing technologies. Additionally, lncRNAs are providing a new perspective towards fresh therapeutic opportunities. In fact higher tissue specificity of lncRNAs over protein coding genes is making them attractive as biomarkers and therapeutic targets. LncRNAs are also remarkably stable in body fluids and tissues and their levels can be evaluated with the help of various techniques to assess the disease progression and/or recovery with a particular treatment regimen. LncRNAs can be targeted therapeutically by a variety of approaches including RNAi mediated gene silencing, antisense oligonucleotides, plasmid based targeting, through small molecule inhibitors, CRISPR/Cas9 and by gene therapy. Combinatorial approaches targeting both lncRNAs and protein-coding genes are coming up as alternative therapeutic strategies in several cancer studies. Thus integrative molecular and therapeutic approaches are rapidly growing for better understanding of this remarkable class of biomolecule.

AIMS

The main objective of this thesis was to investigate the mechanism of lncRNA mediated epigenetic regulation of gene expression in disease model. In the first paper, in breast cancer cell line, using a genome wide approach we have tried to identify lncRNAs that are functionally associated with active chromatin marks and understand the mechanism by which they might regulate gene expression. In the next paper, using genome wide transcriptomic approach on paediatric tumours as disease model we identified regulatory lncRNA in disease condition. Finally, in the third paper, we looked for the role and mechanism of *KCNQ1OT1* lncRNA in familial cases of Beckwith-Wiedemann and Silver–Russell syndrome.

Paper I:

- Using ChRIP-seq we identified lncRNAs that are associated with H3K4me2 and WDR5 active chromatin domains. We named these as Active CARs (chromatin associated RNAs).
- We characterized the mechanism by which active CARs might mediate epigenetic regulation of gene expression.

Paper II:

- Using transcriptomic approach, we identified lncRNAs that are differentially expressed between low- and high- risk neuroblastoma tumours.
- Identified a lncRNA neuroblastoma associated transcript-1 (*NBAT-1*) as a biomarker significantly predicting clinical outcome of neuroblastoma.
- Characterized the role of *NBAT-1* in neuroblastoma pathogenesis that is mediated via epigenetic silencing of target genes.

Paper III:

- Identified a maternal 11p15.5 micro duplications including the most 5′ 20 kb of the non-coding *KCNQ1OT1* gene. Its maternal transmission was associated with ICR2 hypomethylation and familial BWS phenotype.
- Demonstrated that this duplicated *KCNQ1OT1* RNA also interacts with chromatin through its most 5′ 20 kb sequence to silence *CDKN1C*, providing a mechanism that likely facilitates the *CDKN1C* silencing mediated impact to BWS disease phenotype.

MATERIALS AND METHODS

Chromatin RNA Immuno-Precipitation

Chromatin RNA immunoprecipitation (**ChRIP)** is performed to identify lncRNAs on a global scale that are associated with chromatin. Here, we have used this technique to identify lncRNAs that are associated with chromatin enriched with H3K4me2 and WDR5 marks. For that we have used antibodies anti-H3K4me2, anti-WDR5 to purify chromatin enriched with these marks and anti-IgG as a negative control. In brief, BT549 cells were treated with 5µg/µl of Actinomycin D (ActD) (Sigma) for 40 minutes to inhibit nascent transcription. ActD treated cells were first cross-linked ($20x10^6$ cells were fixed in 20ml volume) with 1% Formaldehyde (Sigma-) at room temperature (RT) for 10 minutes and the crosslinking reaction was quenched with Glycine at 125 µM of final concentration (Sigma-) at RT for 5 minutes. These cells were additionally cross-linked with UV light on ice at (0.15 J/cm^2 at 365 nm) in a CL1000 cross-linker. Cells were then scrapped and washed twice with ice cold 1X PBS. Nuclei was isolated using 1X Nuclei Isolation Buffer (5X Nuclei isolation Buffer: 1.28 M sucrose, 40 mM Tris–HCl, pH 7.5, 20 mM MgCl2, 4% TritonX-100 supplemented with protease inhibitor cocktail and RNasin Ribonuclease inhibitor 50 U/mL), given a cold 1X PBS wash and finally resuspended in Nuclear Lysis buffer (0.1% SDS, 0.5% Triton X-100, 20 mM Tris–HCl pH 7.5, 150 mM NaCl. Supplemented with protease inhibitor cocktail and Ribonuclease inhibitor 50 U/mL). Nuclei was sheared using Bioruptor sonicator for 20-30 cycles in order to solubilize the chromatin. Protein A beads were washed with nuclear lysis buffer and then blocked at 4 °C for 20 minutes in nuclear lysis buffer supplemented with 10 mg/mL BSA. 6 µg of specific antibody was then added to washed and pre-blocked beads in nuclear lysis buffer and kept for 2 h at 4 °C. After this, beads were cleared and incubated with 50-60 µg of soluble chromatin to a final volume of 500 µL on a at 4 °C rocker for overnight. Following day, beads were washed once with low salt wash buffer (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris–HCl, pH 7.5, 150 mM NaCl) and high salt wash buffer (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris–HCl, pH 7.5, 500 mM NaCl) followed by two washes in LiCl Buffer (250 mM LiCl, 1% NP40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris–HCl, pH 7.5) and a single final wash with TE buffer (10 mM Tris–HCl, pH 7.5, 0.1 mM EDTA). Washed beads finally resuspended in Elution buffer (100 mM NaCl, 10 mM Tris–HCl, pH 7, 1 mM EDTA, 0.5% SDS) supplemented with Proteinase K (5 µL of 20 mg/mL), incubated at 55 °C 45 min for isolation of chromatin-bound RNA. Formaldehyde crosslinking was reversed by heating the beads at 95 °C 10 min. chromatin-bound RNA was isolated using standard TRIzol-based extraction, DNase I treated and re-extracted the RNA for the final time.

Since RNA yield from each ChRIP pull-down was not sufficient for library preparation to perform high-throughput RNA-sequencing using SOLiD (Applied Biosystem) sequencing platform, we had to pool the purified RNA for each antibody from three independent biological pull-down ChRIP

experiments. The RNA sequencing library preparation for SOLiD high throughput RNA sequencing was done following standard library preparation protocol using SOLiD Total RNA-Seq Kit. The RNA sequencing was done at Uppsala Genome Center, Sweden.

Chromatin Immuno Precipitation (ChIP)

ChIP was performed as per manufacture's (Diagenode) recommendations using anti-H3K4me2, anti-H3K4me3, anti-WDR5 and control IgG antibodies. Briefly, $2x10^6$ BT549 cells were seeded in 10cm petri dish and transfected with 500pmol of negative control and siRNAs or ASOs. Cells treated with control and target-specific siRNA were harvested, and equal number of cells were cross-linked with 1% formaldehyde at RT for 5 minutes. Cross-linking was quenched using 125uM of Glycine. Cell lysate from the cross-linked was prepared using the manufacturer's recommendation and the lysate was sonicated for 70 cycles (30sec on, 30 sec off in Diagenode Bioruptor Plus) to obtain fragment size of 100-300 base pairs. 3µg of antibody was used per 7-10µg of sonicated chromatin and binding reaction was incubated overnight. The steps following overnight incubation was performed as per the manufacturer's recommendations. For ChIP-seq in BT549 cell line with H3K4me2, H3K4me3 and WDR5 antibody we used the same kit as mentioned above with suggested scale up as per as manufacturer's suggestion. WDR5 depleted cells and control siRNA treated BT549 cells were also processed in a similar way to prepare ChIP samples using H3K4me2 and H3K4me3 antibodies for high throughput sequencing.

UV-Crosslinked RNA immunoprecipitation (UV-RIP)

UV-crosslinked RIP was carried out to look for direct interaction of protein of interest with lncRNAs. In the thesis, I have used UV-RIP to validate direct interaction of some lncRNAs with WDR5 active chromatin modifier protein. To do this, BT-549 cells treated with Actinomycin-D for 1 hour were cross linked using ultra violet wavelength of light, washed with 1× PBS (3×) and re-suspended in modified RIPA lysis buffer (150 mM NaCl, 50 mM Tris, 0.5% sodium deoxycholate, 0.2% SDS, 1% NP-40) supplemented with RNase inhibitors and protease inhibitors. Cell suspension was sonicated using a Bioruptor Pico Ultra sonicator for 3 × 30 s cycles. 10 µl of DNase was added to sonicated material, incubated at 37 °C for 10 min, and spun down at max speed for 10 min at 4 °C. Protein-A Dyna-beads were washed and pre-incubated with 4µg of WDR5 antibody for minimum of 2 hours at 4 °C. Lysate and beads were incubated at 4 °C for 2 h. Beads were washed 3× using the following wash buffer (1× PBS, 0.1% SDS, 0.5% NP-40) followed by 2× using a high salt wash buffer (5× PBS, 0.1% SDS, 0.5% NP-40) and crosslinks were reversed and proteins were digested with 5 µl proteinase-K at 55 °C for 1 h. RNA was extracted using TRIzol Reagent.

Chromatin Oligo Affinity Precipitation (ChOP)

The ChOP assay was performed to identify the genomic targets of lncRNAs. Here, I have investigated genomic targeting of some active chromatin associated lncRNAs that we identified by ChRIP-seq. For that, BT549 cells (20 x10^6) were cross-linked using 1% formaldehyde for 10 min at room temperature followed by quenching (using 0.125M glycine). For Actinomycin D treated ChOP, BT549 cells were treated with 5µg/µl ActD for 1 hour before cross-linking step. The cross linked pellet was obtained by centrifugation at 1000g at 4^0C for 10 min. Cells were resuspended in 2ml of buffer A (3mM $MgCl_2$, 10mM Tris HCl, pH 7.4, 10mM NaCl, 0.5%v/v NP-40, 0.5mM PMSF and 100 units/ml RNasin) and incubated on ice for 20 min. Nuclei were harvested by centrifugation and resuspended in 1.2 ml of buffer B (50mM Tris HCl, pH 7.4, 10mM EDTA, 0.5% Triton X-100, 0.1%SDS, 0.5mM PMSF and 100 units/ml RNasin) and incubated on ice for 40 min. An equal volume of buffer C (15mM Tris HCl, pH 7.4, 150mM NaCl, 1mM EDTA, 1% Triton X-100, 0.5mM PMSF and 100 units/ml RNasin) was then added and incubated on ice for 15 min. Samples were sonicated using a Bioruptor sonicator (Diagenode) for 45 cycles (30sec on, 30 sec off at High Pulse). Several oligos complimentary to the lncRNAs were pooled with a final concentration of 10µM and then used for the RNA pull down. As a control, a pool of 9 oligos complimentary to LacZ was used. The oligos were added to the chromatin solution along with yeast tRNA (100µg/ml), salmon sperm DNA (100µg/ml), BSA (400µg/ml) and incubated overnight at 4^0C. Samples were then incubated with streptavidin agarose beads for 3h followed by two washes of Low salt buffer (20mM Tris HCl, pH7.9, 150 mM NaCl, 2mM EDTA, 0.1% SDS, 1% TritonX-100, 0.5mM PMSF and 50 units/ml RNasin), two washes of High salt buffer (20mM Tris HCl, pH7.9, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100, 0.5mM PMSF and 50 units/ml RNasin) and one wash of 1X PBS. The beads were then incubated with elution buffer (1%SDS, 100 mM $NaHCO_3$) for 45 min with intermittent mixing at 65^0C followed by Proteinase K (80 µg/ml) digestion and phenol chloroform extraction method for DNA isolation.

RNA Isolation and Library Preparation

Tumours from neuroblastoma patients were cut on dry ice. Total RNA was isolated from the tumours using the Promega Total RNA Isolation Kit (Promega). Isolated RNA was ribo-depleted using the Ribominus Eukaryotic Kit (Invitrogen) according to the manufacturer's instructions. Ribo-depleted RNA was used to prepare a whole transcriptome library (Applied Biosystems) as per the manufacturer's protocol. Quality, size, and concentration of the RNA during library preparation were analysed using Bioanalyzer 2100 (RNA Pico Kit, Agilent). Final amplified cDNA library was analysed using DNA Hypersensitive Kit (Agilent). Amplified cDNA library were finally sentenced for whole transcriptome sequencing.

RESULTS AND CONLUSIONS

Paper I: H3K4me2 and WDR5 enriched chromatin interacting long non-coding RNAs maintain transcriptionally competent chromatin at divergent transcriptional units.

Main observation:

Genome is sub-compartmentalized into active and inactive chromatin regions with distinct epigenetics codes for effective spatio-temporal regulation of gene expression. Active chromatin is characterized by the presence of histone marks such as H3K4me1, H3K4me2, H3K4me3, H3K36me3 and H3K27ac of which H3K4me2 and H3K4me3 are promoter specific while the rest are used along with H3K27ac and H3K27me3 to define different types of enhancers. Chromatin modifier WDR5 specifically interacts with H3K4me2 and facilitates the MLL-SET1 complexes to assemble and convert H3K4me2 to H3K4me3. Recent evidence suggests that lncRNAs can modulate the chromatin structure by either acting as guides or scaffolds for chromatin remodeling complexes.

In the current investigation, using modified ChRIP technique, we have characterized lncRNAs that are associated with active chromatin enriched with histone modification H3K4me2 and the chromatin reader and/or remodeler WDR5. Following are some of the intriguing observations from our current study.

- Sequencing of ChRIP RNAs purified from the H3K4me2 and WDR5 enriched chromatin domains, identified majority of these H3K4me2- and WDR5- associated lncRNAs to be antisense in orientation to any nearby protein coding genes. 209 lncRNAs were commonly enriched in both H3K4me2 and WDR5 pulldown chromatin fractions and we named them as active chromatin associated lncRNAs (active CARs).
- Active CARs were found to have distinct genomic organization as 41% of them mapped to divergent transcription units, within 2kb upstream of some protein-coding partner genes. Functional enrichment analysis revealed these protein coding partners to primarily encode transcription factors.
- Active divergent CARs are targeted to the promoters of partner protein coding genes and can also interact with WDR5. This helps in maintaining transcriptionally competent chromatin at the promoters as depletion of these active divergent CARs resulted in decrease in the levels of mRNA of the partner protein coding genes along with loss of H3K4me2, H3K4me3 and WDR5 occupancy at the promoters.
- Additionally, deletion of WDR5 resulted in the loss of H3K4me3 but not H3K4me2 at these divergent transcription units.

Conclusions:

1. ChRIP-seq identified H3K4me2 and WDR5 enriched CARs with distinct divergent genomic organization.
2. Expression of divergently organized transcription factor genes are epigenetically regulated by their partner active CARs.
3. Divergent active CAR transcripts *per se* and not the act of transcription is responsible in the maintaining the active chromatin architecture at these promoters.
4. Conversion of H3K4me2 to H3K4me3 at divergent transcription units is mediated by active CARs-WDR5 interaction, whereas, the maintenance of H3K4me2 marks at these promoters appears to be WDR5 independent.

Paper II: The Risk-Associated Long Noncoding RNA NBAT-1 Controls Neuroblastoma Progression by Regulating Cell Proliferation and Neuronal Differentiation.

Main observation:

Neuroblastoma (NB) tumors are found in adrenal medulla and sympathetic ganglia and are derived from the improper differentiation of the neural crest cells. NBs account for 10% of all pediatric cancer mortality. NB is heterogeneous disease which can be categorized into two risk groups depending on the several factors like disease identification stage and chromosomal aberration like MYCN amplification. Because of the heterogeneous nature of the disease the treatment regime of the NB patients remains difficult to decide and often the high-risk tumors relapse in spite of multimodal therapies. This suggests need for reliable biomarker for NB risk stratification and better treatment options. In the current manuscript using RNA-seq we have performed transcriptomic profiling of high-risk and low-risk NBs and looked for novel NB specific long non-coding RNA (lncRNA) based biomarker.

Our transcriptomic profiling was able to differentiate low and high-risk NB tumors and we identified several coding mRNAs and lncRNAs which were differentially expressed between these two disease subgroups. We verified that previously identified differentially expressed coding mRNAs were differentially expressed in our NB data set. Our analysis on the differential expression of the lncRNAs between high and low risk tumors revealed a lncRNA, *NBAT*-1 (neuroblastoma associated transcript 1) which showed decrease expression in high-risk NBs. NBAT1 is located in chromosome 6 and NB risk SNP rs6939340 associated with the aggressive disease is located in the intron of the *NBAT-1* lncRNA. Differential expression of the *NBAT-1* was validated in another independent NB cohort comprising 498 tumors. Higher *NBAT-1* expression was a better predictor of NB patient's survival and *NBAT-1* could serve as an independent prognostic marker in predicting event free survival suggesting *NBAT-1* expression can be used in risk assessment for NBs. We observed that the risk genotype (GG) for the SNP rs6939340 and DNA methylation of promoter contributes in the lower expression of *NBAT-1* in high-risk tumors suggesting both genic and epigenetic contribution in NBAT1 regulation.

Using *in vitro* cell culture based assays and mouse xenograft studies we observed *NBAT-1* can acts as tumor suppressor lncRNA in NBs. Consistent with this depletion of *NBAT-1* in NB cells leads to change in the gene expression signature with genes involved in the proliferation, migration, cell invasion. We observed that in NB cells *NBAT-1* interacts with one of the Polycomb complex 2 (PRC2) members EZH2 and *NBAT-1* depletion leads to de-repression of the genes which are silenced by PRC2. The PRC2 and *NBAT-1* regulated common genes participate in the tumor suppressor function of the *NBAT-1*. We also observed *NBAT-1* depletion leads to differential expression of the genes related to neuronal development and differentiation. We observed *NBAT-1* expression was upregulated during retinoic acid

(RA) mediated differentiation of the NB cells and *NBAT-1* depleted cells show perturbation in the neuronal differentiation phenotype. Gene expression profiling during differentiation of the NB cells after *NBAT-1* depletion revealed that wide spread changes in the expression of the genes related to neuronal differentiation and axonogenesis. We observed that a group of genes which were down regulated during neuronal differentiation of the *NBAT-1* depleted cells were bound by neuron restrictive silencing factor REST. Knock down of the REST in the *NBAT-1* depleted cells could rescue the neuronal differentiation phenotype along with reversal in the gene expression changes.

Conclusions:

1. Transcriptomic profile can distinguish low and high-risk NB tumors.
2. *NBAT-1* is tumor suppressor lncRNA and it regulates gene expression by interaction with PRC2.
3. Loss of *NBAT-1* leads to perturbation in the neuronal differentiation of the NB cells.
4. REST regulated genes are down regulated in the *NBAT-1* depleted cells and REST knockdown could rescue the differentiation defect in *NBAT-1* depleted cells.

Paper III: The *KCNQ1OT1* imprinting control region and non-coding RNA: new properties derived from the study of Beckwith–Wiedemann syndrome and Silver–Russell syndrome cases.

Main observation:

Imprinted genes have a monoallelic expression depending on parent of origin-specific manner. This is regulated by epigenetic modifications that are differentially established on the maternal and paternal alleles. The primary targets of these epigenetic modifications are cis-acting master regulatory elements of the imprinted loci, known as imprinting control regions (ICRs). Two independent imprinted domains are located on chromosome 11p15.5 in human, each of which is controlled by a separate ICR. The telomeric and centromeric ICRs (ICR1 and ICR2, respectively). ICR2 is the promoter of an imprinted lncRNA gene *KCNQ1OT1* which is transcribed in an antisense orientation from intron 11 of the protein-coding gene *KCNQ1*. *KCNQ1OT1* transcript bidirectionally silences in *cis* the imprinted genes of the centromeric domain on the paternal chromosome, whereas on the maternal chromosome, ICR2 is methylated, *KCNQ1OT1* is not transcribed and the flanking imprinted genes expressed.

Heterogeneous molecular defects affecting the 11p15.5 imprinted gene cluster are associated with the congenital growth disorders, Beckwith–Wiedemann syndrome (BWS) and Silver–Russell syndrome. The BWS is characterized by overgrowth, macroglossia, abdominal wall defects and predisposition to embryonal tumours in childhood. The SRS is associated with growth restriction, hypotonia and characteristic dysmorphic features. Loss of the maternal-specific ICR2 methylation is the most frequent defect in BWS, resulting in the bi-allelic activation of *KCNQ1OT1* and bi-allelic silencing of the centromeric domain genes, including the cell growth inhibitor *CDKN1C*. Mutations of CDKN1C is also known to account for 5% of the BWS cases. Thus, expression of *CDKN1C* might play as a one of the contributing factors for the BWS disease phenotype.

Here, we report two interesting cases of maternal 11p15.5 micro duplications in *cis* that is associated with opposite disease phenotypes. The first case is a 1.2 Mb long inverted duplication including the entire 11p15.5 imprinted gene cluster and associated with the SRS phenotype. The second is an exceptional case in which a 160kb duplication that included ICR2 and the initial 20 kb of *KCNQ1OT1* co-segregates with the BWS phenotype in a three-generation pedigree.

Using MS-MLPA (methylation specific multiplex ligation dependent probe amplification assay) with 11p15.5 specific probes we identified ICR2 hypomethylation in all the BWs patients which also showed increased copy number (CN) of the *KCNQ1* exons 12–15 and ICR2 (located in *KCNQ1* intron 11) in all the individuals displaying the ICR2 hypomethylation. Next the precise extent of this duplicated regions was analysed by array comparative genomic hybridization using the Affymetrix Genome Wide Human SNP Array 6.0 with DNA from patients. In BWS family the 160kb duplicated region was found to include ICR2 and the first 20kb of *KCNQ1OT1* gene. Bisulphite sequencing strategy was used to

check the allele specific methylation pattern of the normal and duplicated ICRs. In the BWS family upon maternal transmission of the micro duplication, ICR2 is hypomethylated, whereas the truncated *KCNQ1OT1* is expressed and *CDKN1C* is silenced.

Next, we investigated the mechanism of biallelic expression of *CDKN1C*. To this end, we performed chromatin RNA immune purification (ChRIP), using antibodies raised against H3 histone on cross-linked chromatin obtained from lymphoblastoid cell lines derived from one control healthy individual, two familial BWS patients carrying 160kb duplication and a fourth unrelated individual with ICR2 hypomethylation without duplication.

We designed allele specific primer based on the SNP (rs463924) located after 4kb of *KCNQ1OT1* transcription start site, that could distinguish the parental alleles. Using allele specific ChRIP we identified that paternally derived *KCNQ1OT1* (normally expressed) efficiently interacted with chromatin in all four cell lines. However, the enrichment of the maternally derived transcript in the immunoprecipitated chromatin was found to be significantly higher in the BWS cell lines than in the control.

To obtain a further confirmation, the ChRIP RNAs of the normal control, and three patient cell lines were retro-transcribed, PCR-amplified and sequenced with primers around the SNP rs463924. The sequencing electropherograms clearly demonstrated the presence of both parental alleles (A/G) in the ChRIP RNAs of the BWS samples, but only one allele (G) in the control.

Conclusions:

Taken together, we could demonstrate that the human *KCNQ1OT1* interacts with chromatin and that the silencing of *CDKN1C* by this lncRNA in the BWS cells occurs via interaction with chromatin, providing a mechanism that likely facilitates the *CDKN1C* silencing mediated impact to BWS disease phenotype.

ACKNOWLEDGEMENTS

The work presented here has been carried out at Department of Immunology, Genetics and Pathology (IGP), Uppsala University and also at Department of Medical Biochemistry and Cell Biology at Institute of Biomedicine, University of Gothenburg. I would like to express my sincere gratitude to everyone from this these two departments who have directly and indirectly helped me in the completion of my thesis.

Firstly, I take this opportunity to thank my supervisor **Prof Chandrasekhar Kanduri** for giving me an opportunity to work with him. It had been a good learning experience in working with you. Thanks for your constant support, critical guidance, motivation and for the freedom in planning of experiments to develop some of the interesting projects of your lab. Your enthusiasm for science and scientific discussion have also motivated me. I applaud your patience for giving me time and space to improve myself to do better science.

Thanks to my co-supervisor, **Prof Claes Gustafsson** for his support and advice on anything that I have approached him with. It has always been nice speaking to you, especially about your wonderful summer travel experiences. Though you have tried your best to motivate me for learning to sail, I have not yet managed to do so, but would learn it one day.

Thanks to **Prof Per Elias**, for being a source of motivation with your sense of humour and helpfulness. It was wonderful to discuss with you on various subjects that ranged from laboratory science to good museums. Your suggestions and guidance on professional aspects have been of great help. Special thanks to you and Claes for making us feel at home when we moved in to Gothenburg from Uppsala.

Thanks to **Prof Sven Enerbäck**, as head of the department has always been supportive and encouraging.

Thanks to **Ann Uv**, for always being so helpful whenever I have approached you. The corridor celebrations for gloggs and paper submission from your group were fun and interactive.

Everyone in the Administration has always been very helpful, accommodative and patient with my queries and problems. I want to thank each one of you for helping me with all the administrative issues for all these years. Special mention goes for **Carina Petersson** and **Carina Ejdeholm** as I have mostly interacted with you both.

Thanks to **Prof Levent Akyürek** and **Erik Larsson Lekholm** for creating a positive, interactive and cordial environment in our working place.

I would like to thank **Tanmoy** and **Gaurav** for their help during my initial period in the lab, teaching me different lab techniques, providing valuable tips and suggestions. It had always been a learning experience for me to discuss science and scientific issues with you both, especially in my initial days.

Tanmoy, you have always been the diplomatically restrained pleasant personality. You have been a guiding force with your scientific temperament and valuable suggestions. I will fondly remember sharing office with you and chatting about life in and outside university. It has been relaxing to discuss with you on a wide range of topics from politics to science.

Gaurav, your sense of humour and knowledge about a lot of things made it fun to work with you. We have had wonderful time while sharing apartment in Uppsala. It was nice to work with you on few projects. Thanks for your invaluable suggestions and comments made specifically during our animal house visits. I formed along with my lab and flatmates, Gaurav and **Prasoon**, a tiny fraternity in Uppsala, the activities of which ranged from political debates to catching movies, with financial suggestion from Prasoon and his never-ending enthusiasm to host a Friday night movie show.

Matthieu, you have been the first person in Sweden in whom I found a friend to confide to. Our regular lunches helped me to keep going when times were tough. Thanks for always understanding me, motivating me and guiding me. Our discussions on scientific non-scientific topics were always interesting. You have always been helpful with work in the lab also.

Sanhita, thanks for all the wonderful parties and food at your place. It was a good experience to work with you on one of the common projects. Yours's take it easy attitude was a good influence in the lab.

Santhilal, you have been constantly supportive with all your bioinformatics help in my projects. Our late evening discussions to shape the chromatin project were very productive. Great help from you with thesis figures and formatting. Your generosity as a host of a bachelors' den on Fridays with wonderful biriyani will stay as fond memories.

A former member of the den, **Prasanna** (Anna), helped with my experiments, chatted late into evenings, provided endless technical suggestions, and never forgot to pull me to his beverage breaks. And, thanks for being the only person in the lab to always support me.

Gendy, your enthusiasm has been a positive influence in our lab. Your humour and ever smiling attitude never let lab life take a nosedive. Special thanks for all the good Egyptian food and sweets that you have supplied us. Also, big thanks to **Walla** for hosting party at your place. Those gathering have always been fun.

Sara, the youngest member, your timely help with the laptop enabled me to complete my thesis on time. You took the initiative to start the fika and other lab outings fund which was nice for all of us to gather and spend quality time. You and Gendy brought a sense of inclusiveness amongst all the members in the lab. I enjoyed your company, specially our interesting discussions in the tram on our way back from lab.

Subazini, thanks for your ever willingness to help with anything and anytime. Subhazini's calm and reassuring presence helped me weather many rough situations.

Vijay, thanks for your help with my experiments in the chromatin projects at crucial time. Vijay's support and urge to not let the world pass by while lab rats are cooped up has been inspiring for us to initiate lab trips and holiday trips. Such trips have been full of fun and memorable.

Luisa, thanks for being nice to me and being helpful. You deserve special thanks for allowing Mochi to stay with me. She is a sweetheart.

Caroline has been a steady partner in laughs and a very dear friend. The invaluable gifts that you got for me from Brazil deserves a mention. As also the Brazilian carrot cakes, parties at your house helped me a lot during my stay here.

Lily, you have influenced everyone in our lab in the most positive way with your presence. Your enthusiasm about everything in life is infectious. You, Sara and Vijay really pulled us out of lab to explore life. Your repeated attempts to teach me driving and Swedish has been inspiring though not successful. Special thanks for the wonderful cakes.

Som, some of our heated yet interesting political/apolitical discussions have been fun. It was nice to have you around.

Abhiarchana, **Roshan** you both made my initial stay in Gothenburg easier with your friendly nature and occasional cooking together events. Roshan thanks with your experimental helps in the lab.

Meena thank you for hosting all the wonderful parties that were so much fun with good food and interesting games. You have always been very spontaneous while helping me with work and reagents

so many times. Whenever, in need of enzymes I used to run to your lab, and you have been very warm in your support.

Nalinee, thank you for hosting wonderful parties at your place, with good food and lots of fun. It was really nice of you to celebrate my birthday at your place.

Thanks to all the fellow present and past researchers from our floor, **Babak**, **Ari**, **Swaraj**, **Sashi**, members of **Ann Uv's group** who have helped in a creating cordial interactive environment. Thanks guys for all the grilling, after work, Christmas dinner parties, floor-fika and activities outside lab.

Thanks to **Ganesh** for your timely support and help during thesis writing. **Pradeep** and **Hamdy** thanks for always being helpful.

Thanks to all the present and past members working on the third floor (**Claes**, **Mia** and **Per's** group) for always being very helpful and welcoming. Our coffee time interactions have been nice and humorous most of the times. **Emily** (third floor) thanks for always willing to help. **Isabella** (third floor) with our common interest in Brazilian soccer, tea time chats with you have always been nice. **Gunilla Petersson** thanks for always being very helpful, especially for your proactive role on departmental glögg and strawberry days.

All my teachers have played a major role in my life since childhood and I acknowledge each one of them. I have been fortunate enough to be able to get in touch with some wonderful personalities as my teachers, who have always supported, encouraged and influenced me. My teachers at **Banaras Hindu University**, particularly the quartet of **Subhas Chandra Lakhotia**, **Rajiva Raman**, **Mercy Raman**, and **Jagat Roy**, have been influential in shaping my research career. The passion and commitment with which you all taught us are unparalleled.

All my teachers the **Garden City College**, Bengaluru, especially everyone from **Department of Genetics**, deserve special mention for nurturing my interest in biology. Each one of you have helped me in not only with studies but also in guiding me as an individual.

Thanks to all my teachers from my school, **Vivekananda Mission School**, Kolkata for helping me to nurture my interest in science.

During this time, I managed to make some friends outside lab during our occasional get togethers where Sanhita fondly introduced me. All the friends from **Ong Bong Chong** have made my stay quite eventful. Thanks to **Namrata**, **Prithu**, **Kiran-da** and **Susmita** di for organizing Saraswati Pujo, Bijoya get together, Christmas and Halloween parties.

Arghya has been a good friend. **Mimi**, thanks for your encouragement and support during final thesis days. Your enthusiasm for travel is inspiring.

Niranjan has been a good friend. Thank you Niru for being there.

I have been fortunate to interact and spend time with some lovely kids here in Gothenburg. **Neha**, **Sanju**, **Sanjith**, **Nushi**, **Noor**, **Hiya** and **Giggi** have always brought great joy to me. **Neha** special thanks for all the lovely cakes that you have been baking for us. These sweet adorable kids and their warm smiles always helped rejuvenating myself time and again. Thanks for making us realize that life is beautiful, and the world is innocent.

My stay in Uppsala was short but memorable. Thanks to **Kali-bhai**, **Sulena** (for being my first host in Sweden), **Snehangshu-da**, **Soumi-di**, **Naren-da** and **Sujata** for all the love and affections. **Umash ji** and **Noopur** have been wonderful hosts and those get together at your place with song and dance are fond memories.

Special thanks to my BSc batchmates, everyone from **Genex-bhavan**, who have stood by me in tough times and for all the unconditional love and affection. It was wonderful to have been able to spend time with bunch of unique characters. Thanks for continuing that bond of friendship and accepting me the way I am without complaining.

Rachel Mam, thanks for your continuous support and encouragement and without your guidance it would have been difficult for me to come abroad for my PhD. **Annapurna**, thanks for all your support and encouragement at a time when I hardly had anyone by my side. My colleagues at the Centre for Cellular & Molecular Biology, especially **Ramrup-da**, **Boudi**, **Somu**, **Surjo**, **Samit**, **Pankaj**, and **Susmita**, and at Indian Statistical Institute, notably **Debuda**, **Sujit-da**, **Mousumi-di**, and **Shalini-di**, are some of the nicest persons whom I am fortunate to have befriended.

Thanks to **Iru-di** for her everyday what's-app messages and effort to help me in not missing any festivals and occasions. It has been a blessing to know and interact with **Partha-kaku**, **Kakima** and **Chintu**. The love affection guidance and support from you all have kept me strong in times of crisis. During the last few years of thesis, whenever in need I have depended on you Kaku. Thanks for always being a support for me. In Chintu I found a loving, caring and ever supportive younger brother.

Chotokaku, without you and your wonderful childhood bedtime stories I could have never learned to love biology. You have always supported me in every possible way to see to it that I can complete my higher education. You have inspired me into research and have taught me with your life what exactly is scientific enthusiasm. You have always motivated me for a PhD, and now that I am completing it, can't thank you enough for all the love affection and blessings.

Manikarnika, my best friend and confidant. I feel blessed to have you in my life, my best gain from Sweden. You have always stood by me, with me and for me. This has been a journey for you too. You never allowed me to feel low or lonely. Nobody have understood me better than you. Thanks for all your love and affection and for your unconditional support, encouragement. I could not have completed these many years in Sweden without you and **Souvik**. Together, we have managed to build a family, my strongest support system. Together we have spent some of the fondest moments of our lives. With his sense of calmness and humour, Souvik have always helped to ease my tension. From late night cooking adventures, memorable trip experiences, to birthday celebrations we have done it all together. Thanks for being my support system and family away from home.

Finally, I would like thank my parents, **Maa** and **Baba**, who made this journey possible. Their unconditional encouragement, motivation and tremendous confidence in me really helped. **Baba**, I learnt from you the meaning of being passionate about academics, of being honest to one's profession, and of being positive about life. **Maa**, I learnt from you to be happy in life without complaints and have patience. Both of you have always taken the trouble to make sure I could continue with my studies and aspirations in life. I am extremely fortunate to have the least demanding parents. Finally, when I am finishing my thesis a big thank you to you both for being so wonderful.

REFERENCES

1. Taft, R.J., M. Pheasant, and J.S. Mattick, *The relationship between non-protein-coding DNA and eukaryotic complexity.* BioEssays, 2007. **29**(3): p. 288-299.
2. Carninci, P., et al., *The transcriptional landscape of the mammalian genome.* Science, 2005. **309**(5740): p. 1559-63.
3. Mattick, J.S., *The central role of RNA in human development and cognition.* FEBS Letters, 2011. **585**(11): p. 1600-1616.
4. Bernstein, E. and C.D. Allis, *RNA meets chromatin.* Genes Dev, 2005. **19**(14): p. 1635-55.
5. Mattick, J.S. and I.V. Makunin, *Non-coding RNA.* Hum Mol Genet, 2006. **15 Spec No 1**: p. R17-29.
6. Katayama, S., et al., *Antisense Transcription in the Mammalian Transcriptome.* Science, 2005. **309**(5740): p. 1564.
7. Wight, M. and A. Werner, *The functions of natural antisense transcripts.* Essays Biochem, 2013. **54**: p. 91-101.
8. Umlauf, D., P. Fraser, and T. Nagano, *The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme.* Biol Chem, 2008. **389**(4): p. 323-31.
9. Raefski, A.S. and M.J. O'Neill, *Identification of a cluster of X-linked imprinted genes in mice.* Nat Genet, 2005. **37**(6): p. 620-4.
10. Bartolomei, M.S. and A.C. Ferguson-Smith, *Mammalian Genomic Imprinting.* Cold Spring Harbor Perspectives in Biology, 2011. **3**(7).
11. Barlow, D.P. and M.S. Bartolomei, *Genomic Imprinting in Mammals.* Cold Spring Harbor Perspectives in Biology, 2014. **6**(2).
12. Kanduri, C., *Long noncoding RNAs: Lessons from genomic imprinting.* Biochim Biophys Acta, 2016. **1859**(1): p. 102-11.
13. Rougeulle, C. and E. Heard, *Antisense RNA in imprinting: spreading silence through Air.* Trends Genet, 2002. **18**(9): p. 434-7.
14. Lewis, A. and W. Reik, *How imprinting centres work.* Cytogenet Genome Res, 2006. **113**(1-4): p. 81-9.
15. Guttman, M., et al., *lincRNAs act in the circuitry controlling pluripotency and differentiation.* Nature, 2011. **477**(7364): p. 295-300.
16. *Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154:26–46*.
17. Schmitt, A.M. and H.Y. Chang, *Long Noncoding RNAs in Cancer Pathways.* Cancer Cell. **29**(4): p. 452-463.
18. Zhao, Y., H. Sun, and H. Wang, *Long noncoding RNAs in DNA methylation: new players stepping into the old game.* Cell & Bioscience, 2016. **6**: p. 45.
19. Mondal, T., et al., *Characterization of the RNA content of chromatin.* Genome Res, 2010. **20**(7): p. 899-907.
20. Miska, E.A. and A.C. Ferguson-Smith, *Transgenerational inheritance: Models and mechanisms of non–DNA sequence–based inheritance.* Science, 2016. **354**(6308): p. 59-63.
21. Moazed, D., *Mechanisms for the Inheritance of Chromatin States.* Cell. **146**(4): p. 510-518.
22. Rodríguez-Campos, A. and F. Azorín, *RNA Is an Integral Component of Chromatin that Contributes to Its Structural Organization.* PLOS ONE, 2007. **2**(11): p. e1182.
23. Rinn, J.L., *lncRNAs: Linking RNA to Chromatin.* Cold Spring Harbor Perspectives in Biology, 2014. **6**(8): p. a018614.

24. Laurent, L., et al., *Dynamic changes in the human methylome during differentiation.* Genome Research, 2010. **20**(3): p. 320-331.
25. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-322.
26. Bourc, et al., *Dnmt3L and the Establishment of Maternal Genomic Imprints.* Science, 2001. **294**(5551): p. 2536.
27. Rothbart, S.B., et al., *Association of UHRF1 with H3K9 methylation directs the maintenance of DNA methylation.* Nature structural & molecular biology, 2012. **19**(11): p. 1155-1160.
28. Pfeifer, G.P., *Mutagenesis at methylated CpG sequences.* Curr Top Microbiol Immunol, 2006. **301**: p. 259-81.
29. Metivier, R., et al., *Cyclical DNA methylation of a transcriptionally active promoter.* Nature, 2008. **452**(7183): p. 45-50.
30. Gruenbaum, Y., et al., *Methylation of CpG sequences in eukaryotic DNA.* FEBS Letters, 1981. **124**(1): p. 67-71.
31. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes & Development, 2011. **25**(10): p. 1010-1022.
32. Elango, N. and S.V. Yi, *Functional Relevance of CpG Island Length for Regulation of Gene Expression.* Genetics, 2011. **187**(4): p. 1077-1083.
33. Saxonov, S., P. Berg, and D.L. Brutlag, *A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.* Proceedings of the National Academy of Sciences, 2006. **103**(5): p. 1412-1417.
34. Brinkman, A.B., et al., *DNA methylation immediately adjacent to active histone marking does not silence transcription.* Nucleic Acids Research, 2007. **35**(3): p. 801-811.
35. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.* Nat Genet, 2007. **39**(4): p. 457-66.
36. Ooi, S.K., et al., *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.* Nature, 2007. **448**(7154): p. 714-7.
37. Thomson, J.P., et al., *CpG islands influence chromatin structure via the CpG-binding protein Cfp1.* Nature, 2010. **464**(7291): p. 1082-1086.
38. Edmunds, J.W., L.C. Mahadevan, and A.L. Clayton, *Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation.* The EMBO Journal, 2008. **27**(2): p. 406-420.
39. Kobayashi, H., et al., *Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks.* PLOS Genetics, 2012. **8**(1): p. e1002440.
40. Dhayalan, A., et al., *The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation.* J Biol Chem, 2010. **285**(34): p. 26114-20.
41. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
42. Zink, L.-M. and S.B. Hake, *Histone variants: nuclear function and disease.* Current Opinion in Genetics & Development, 2016. **37**(Supplement C): p. 82-89.
43. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* Genes & Development, 2002. **16**(22): p. 2893-2905.
44. Cao, R. and Y. Zhang, *SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex.* Mol Cell, 2004. **15**(1): p. 57-67.
45. Alekseyenko, A.A., et al., *Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification.* Proc Natl Acad Sci U S A, 2014. **111**(7): p. 2488-93.
46. Pasini, D., et al., *JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells.* Nature, 2010. **464**(7286): p. 306-10.

47. Peng, J.C., et al., *Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells.* Cell, 2009. **139**(7): p. 1290-302.
48. Ferrari, K.J., et al., *Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity.* Mol Cell, 2014. **53**(1): p. 49-62.
49. Ferrari, Karin J., et al., *Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity.* Molecular Cell, 2014. **53**(1): p. 49-62.
50. Jung, H.R., et al., *Quantitative Mass Spectrometry of Histones H3.2 and H3.3 in Suz12-deficient Mouse Embryonic Stem Cells Reveals Distinct, Dynamic Post-translational Modifications at Lys-27 and Lys-36.* Molecular & Cellular Proteomics : MCP, 2010. **9**(5): p. 838-850.
51. Gao, Z., et al., *PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes.* Molecular Cell, 2012. **45**(3): p. 344-356.
52. Morey, L., et al., *RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells.* Cell Reports, 2013. **3**(1): p. 60-69.
53. Ku, M., et al., *Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.* PLoS Genet, 2008. **4**(10): p. e1000242.
54. Plath, K., et al., *Role of histone H3 lysine 27 methylation in X inactivation.* Science, 2003. **300**(5616): p. 131-5.
55. Mak, W., et al., *Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells.* Curr Biol, 2002. **12**(12): p. 1016-20.
56. Wang, L., et al., *Hierarchical recruitment of polycomb group silencing complexes.* Mol Cell, 2004. **14**(5): p. 637-46.
57. Ren, X., C. Vincenz, and T.K. Kerppola, *Changes in the Distributions and Dynamics of Polycomb Repressive Complexes during Embryonic Stem Cell Differentiation.* Molecular and Cellular Biology, 2008. **28**(9): p. 2884-2895.
58. Cooper, S., et al., *Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment.* Cell Rep, 2014. **7**(5): p. 1456-1470.
59. Blackledge, N.P., et al., *Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation.* Cell, 2014. **157**(6): p. 1445-59.
60. Zhang, T., S. Cooper, and N. Brockdorff, *The interplay of histone modifications – writers that read.* EMBO reports, 2015. **16**(11): p. 1467-1481.
61. Shinkai, Y. and M. Tachibana, *H3K9 methyltransferase G9a and the related molecule GLP.* Genes Dev, 2011. **25**(8): p. 781-8.
62. Schultz, D.C., et al., *SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins.* Genes Dev, 2002. **16**(8): p. 919-32.
63. Chin, H.G., et al., *Catalytic properties and kinetic mechanism of human recombinant Lys-9 histone H3 methyltransferase SUV39H1: participation of the chromodomain in enzymatic catalysis.* Biochemistry, 2006. **45**(10): p. 3272-84.
64. Lehnertz, B., et al., *Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin.* Curr Biol, 2003. **13**(14): p. 1192-200.
65. Fuks, F., et al., *The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase.* Nucleic Acids Res, 2003. **31**(9): p. 2305-12.
66. Fuks, F., et al., *The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation.* J Biol Chem, 2003. **278**(6): p. 4035-40.
67. Pauler, F.M., et al., *H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome.* Genome Res, 2009. **19**(2): p. 221-33.
68. Peters, A.H., et al., *Partitioning and plasticity of repressive histone methylation states in mammalian chromatin.* Mol Cell, 2003. **12**(6): p. 1577-89.
69. Bilodeau, S., et al., *SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state.* Genes Dev, 2009. **23**(21): p. 2484-9.

70. Alder, O., et al., *Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment.* Development, 2010. **137**(15): p. 2483-92.
71. Hawkins, R.D., et al., *Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.* Cell Stem Cell, 2010. **6**(5): p. 479-91.
72. Mozzetta, C., et al., *The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing.* Molecular Cell, 2014. **53**(2): p. 277-289.
73. Boros, J., et al., *Polycomb Repressive Complex 2 and H3K27me3 Cooperate with H3K9 Methylation To Maintain Heterochromatin Protein 1α at Chromatin.* Molecular and Cellular Biology, 2014. **34**(19): p. 3662-3674.
74. Escamilla-Del-Arenal, M., et al., *Cdyl, a New Partner of the Inactive X Chromosome and Potential Reader of H3K27me3 and H3K9me2.* Molecular and Cellular Biology, 2013. **33**(24): p. 5005-5020.
75. Shilatifard, A., *The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis.* Annual review of biochemistry, 2012. **81**: p. 65-95.
76. Ng, H.H., et al., *Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity.* Molecular Cell, 2003. **11**(3): p. 709-719.
77. Kolasinska-Zwierz, P., et al., *Differential chromatin marking of introns and expressed exons by H3K36me3.* Nature genetics, 2009. **41**(3): p. 376-381.
78. Xiao, T., et al., *Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast.* Genes & Development, 2003. **17**(5): p. 654-663.
79. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**.
80. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.
81. Wysocka, J., et al., *WDR5 Associates with Histone H3 Methylated at K4 and Is Essential for H3 K4 Methylation and Vertebrate Development.* Cell, 2005. **121**(6): p. 859-872.
82. Bernstein, B.E., et al., *A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells.* Cell, 2006. **125**(2): p. 315-326.
83. Santos-Rosa, H., et al., *Active genes are tri-methylated at K4 of histone H3.* Nature, 2002. **419**(6905): p. 407-411.
84. Orford, K., et al., *Differential H3K4 methylation identifies developmentally poised hematopoietic genes.* Dev Cell, 2008. **14**(5): p. 798-809.
85. Zhang, J.A., et al., *Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity.* Cell, 2012. **149**(2): p. 467-482.
86. Wang, Y., X. Li, and H. Hu, *H3K4me2 reliably defines transcription factor binding regions in different cells.* Genomics, 2014. **103**(2): p. 222-228.
87. Pekowska, A., et al., *A unique H3K4me2 profile marks tissue-specific gene regulation.* Genome research, 2010. **20**(11): p. 1493-502.
88. Dias, J., et al., *Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex.* Genes & Development, 2014. **28**(9): p. 929-942.
89. Trievel, R.C. and A. Shilatifard, *WDR5, a complexed protein.* Nat Struct Mol Biol, 2009. **16**(7): p. 678-80.
90. Calo, E. and J. Wysocka, *Modification of enhancer chromatin: what, how and why?* Molecular cell, 2013. **49**(5): p. 10.1016/j.molcel.2013.01.038.
91. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-461.
92. Nord, A.S., et al., *Rapid and Pervasive Changes in Genome-Wide Enhancer Usage During Mammalian Development.* Cell, 2013. **155**(7): p. 1521-1531.

93. Lin, C.Y., et al., *Active medulloblastoma enhancers reveal subgroup-specific cellular origins.* Nature, 2016. **530**(7588): p. 57-62.
94. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-283.
95. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proceedings of the National Academy of Sciences, 2010. **107**(50): p. 21931-21936.
96. Khalil, A.M., et al., *Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.* Proc Natl Acad Sci U S A., 2009. **106**.
97. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.* Nature., 2009. **458**.
98. Kodzius, R., et al., *CAGE: cap analysis of gene expression.* Nat Meth, 2006. **3**(3): p. 211-222.
99. Ulitsky, I., et al., *Extensive alternative polyadenylation during zebrafish development.* Genome Research, 2012. **22**(10): p. 2054-2066.
100. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.* Genes Dev., 2011. **25**.
101. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.* Genome Res., 2012. **22**.
102. Pauli, A., et al., *Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis.* Genome Res., 2012. **22**.
103. Ulitsky, I., et al., *Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution.* Cell, 2011. **147**(7): p. 1537-50.
104. Mele, M., et al., *Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs.* Genome Res, 2017. **27**(1): p. 27-37.
105. Sigova, A.A., et al., *Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells.* Proc Natl Acad Sci U S A., 2013. **110**.
106. Luo, S., et al., *Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells.* Cell Stem Cell, 2016. **18**(5): p. 637-652.
107. Mercer, T.R., et al., *Specific expression of long noncoding RNAs in the mouse brain.* Proc Natl Acad Sci U S A., 2008. **105**.
108. Clark, M.B., et al., *Genome-wide analysis of long noncoding RNA stability.* Genome Research, 2012. **22**(5): p. 885-898.
109. Tani, H., et al., *Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals.* Genome Research, 2012. **22**(5): p. 947-956.
110. Cabili, M.N., et al., *Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution.* Genome Biology, 2015. **16**(1): p. 20.
111. van Werven, F.J., et al., *Transcription of two long non-coding RNAs mediates mating type control of gametogenesis in budding yeast.* Cell, 2012. **150**(6): p. 1170-1181.
112. Latos, P.A., et al., *Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing.* Science, 2012. **338**(6113): p. 1469-1472.
113. Schmitt, S., M. Prestel, and R. Paro, *Intergenic transcription through a Polycomb group response element counteracts silencing.* Genes & Development, 2005. **19**(6): p. 697-708.
114. Yao, H., et al., *Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA.* Genes & Development, 2010. **24**(22): p. 2543-2555.
115. Jeon, Y. and J.T. Lee, *YY1 tethers Xist RNA to the inactive X nucleation center.* Cell, 2011. **146**(1): p. 119-133.
116. Lai, F., et al., *Activating RNAs associate with Mediator to enhance chromatin architecture and transcription.* Nature, 2013. **494**(7438): p. 497-501.
117. Gomez, J.A., et al., *NeST, a long noncoding RNA, controls microbial susceptibility and epigenetic activation of the Ifng locus.* Cell, 2013. **152**(4): p. 743-754.

118. Grote, P., et al., *The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse.* Developmental cell, 2013. **24**(2): p. 206-214.
119. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature., 2011. **472**.
120. Tsai, M.-C., et al., *Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes.* Science (New York, N.Y.), 2010. **329**(5992): p. 689-693.
121. Rinn, J.L., et al., *Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Non-Coding RNAs.* Cell, 2007. **129**(7): p. 1311-1323.
122. Zhao, J., et al., *Genome-wide identification of Polycomb-associated RNAs by RIP-seq.* Molecular cell, 2010. **40**(6): p. 939-953.
123. Yang, Y.W., et al., *Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency.* eLife, 2014. **3**: p. e02046.
124. G Hendrickson, D., et al., *Widespread RNA binding by chromatin-associated proteins.* Genome Biology, 2016. **17**: p. 28.
125. Mohammad, F., T. Mondal, and C. Kanduri, *Epigenetics of imprinted long non-coding RNAs.* Epigenetics, 2009. **4**(5): p. 277-286.
126. Lee, Jeannie T. and Marisa S. Bartolomei, *X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease.* Cell. **152**(6): p. 1308-1323.
127. Kanduri, C., et al., *Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive.* Current Biology. **10**(14): p. 853-856.
128. Kanduri, C., et al., *A Differentially Methylated Imprinting Control Region within the Kcnq1 Locus Harbors a Methylation-sensitive Chromatin Insulator.* Journal of Biological Chemistry, 2002. **277**(20): p. 18106-18110.
129. Kanduri, C., et al., *The 5′ flank of mouse H19 in an unusual chromatin conformation unidirectionally blocks enhancer–promoter communication.* Current Biology, 2000. **10**(8): p. 449-457.
130. Sleutels, F., R. Zwart, and D.P. Barlow, *The non-coding Air RNA is required for silencing autosomal imprinted genes.* Nature, 2002. **415**(6873): p. 810-813.
131. Bartolomei, M.S., S. Zemel, and S.M. Tilghman, *Parental imprinting of the mouse H19 gene.* Nature, 1991. **351**(6322): p. 153-155.
132. Monnier, P., et al., *H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1.* Proceedings of the National Academy of Sciences, 2013. **110**(51): p. 20693-20698.
133. Kallen, A.N., et al., *The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs.* Molecular cell, 2013. **52**(1): p. 10.1016/j.molcel.2013.08.027.
134. Giovarelli, M., et al., *H19 long noncoding RNA controls the mRNA decay promoting function of KSRP.* Proceedings of the National Academy of Sciences, 2014. **111**(47): p. E5023-E5028.
135. Dey, B.K., K. Pfeifer, and A. Dutta, *The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration.* Genes & Development, 2014. **28**(5): p. 491-501.
136. Yoshimizu, T., et al., *The H19 locus acts in vivo as a tumor suppressor.* Proceedings of the National Academy of Sciences, 2008. **105**(34): p. 12417-12422.
137. Hao, Y., et al., *Tumour-suppressor activity of H19 RNA.* Nature, 1993. **365**(6448): p. 764-767.
138. Moulton, T., et al., *Epigenetic lesions at the H19 locus in Wilms' tumour patients.* Nat Genet, 1994. **7**(3): p. 440-447.
139. Adriaenssens, E., et al., *H19 Overexpression in Breast Adenocarcinoma Stromal Cells Is Associated with Tumor Values and Steroid Receptor Status but Independent of p53 and Ki-67 Expression.* The American Journal of Pathology, 1998. **153**(5): p. 1597-1607.
140. Pandey, R.R., et al., *Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.* Mol Cell, 2008. **32**(2): p. 232-46.

141. Mancini-DiNardo, D., et al., *Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes.* Genes & Development, 2006. **20**(10): p. 1268-1282.
142. Mohammad, F., et al., *Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region.* Mol Cell Biol, 2008. **28**(11): p. 3713-28.
143. Shin, J.-Y., G.V. Fitzpatrick, and M.J. Higgins, *Two distinct mechanisms of silencing by the KvDMR1 imprinting control region.* The EMBO Journal, 2008. **27**(1): p. 168-178.
144. Terranova, R., et al., *Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos.* Dev Cell., 2008. **15**.
145. Mohammad, F., et al., *Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1.* Development, 2010. **137**(15): p. 2493-2499.
146. Redrup, L., et al., *The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing.* Development (Cambridge, England), 2009. **136**(4): p. 525-530.
147. Mohammad, F., et al., *Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing.* Development, 2012. **139**(15): p. 2792-2803.
148. Zhang, H., et al., *Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the Kcnq1 locus.* The Journal of Cell Biology, 2014. **204**(1): p. 61-75.
149. Koerner, M.V., et al., *The function of non-coding RNAs in genomic imprinting.* Development (Cambridge, England), 2009. **136**(11): p. 1771-1783.
150. Nagano, T., et al., *The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin.* Science., 2008. **322**.
151. Williamson, C.M., et al., *Identification of an imprinting control region affecting the expression of all transcripts in the Gnas cluster.* Nat Genet, 2006. **38**(3): p. 350-355.
152. Williamson, C.M., et al., *Uncoupling Antisense-Mediated Silencing and DNA Methylation in the Imprinted Gnas Cluster.* PLOS Genetics, 2011. **7**(3): p. e1001347.
153. Conrad, T. and A. Akhtar, *Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription.* Nat Rev Genet, 2012. **13**(2): p. 123-134.
154. Wutz, A., *Noncoding roX RNA Remodeling Triggers Fly Dosage Compensation Complex Assembly.* Molecular Cell. **51**(2): p. 131-132.
155. Ramirez, F., et al., *High-affinity sites form an interaction network to facilitate spreading of the MSL complex across the X chromosome in Drosophila.* Molecular cell, 2015. **60**(1): p. 146-162.
156. Furlan, G. and C. Rougeulle, *Function and evolution of the long noncoding RNA circuitry orchestrating X-chromosome inactivation in mammals.* Wiley Interdisciplinary Reviews: RNA, 2016. **7**(5): p. 702-722.
157. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.* Science, 2008. **322**(5902): p. 750-6.
158. Wutz, A., T.P. Rasmussen, and R. Jaenisch, *Chromosomal silencing and localization are mediated by different domains of Xist RNA.* Nat Genet, 2002. **30**(2): p. 167-174.
159. Sado, T., Y. Hoki, and H. Sasaki, *Tsix Silences Xist through Modification of Chromatin Structure.* Developmental Cell. **9**(1): p. 159-165.
160. Navarro, P., et al., *Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program.* Genes & Development, 2006. **20**(20): p. 2787-2792.
161. Tian, D., S. Sun, and J.T. Lee, *The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation.* Cell., 2010. **143**.
162. Vallot, C., et al., *XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells.* Nat Genet, 2013. **45**(3): p. 239-241.
163. Yang, F., et al., *The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation.* Genome Biology, 2015. **16**(1): p. 52.
164. Kretz, M., et al., *Control of somatic tissue differentiation by the long non-coding RNA TINCR.* Nature., 2013. **493**.

165. Bao, X., et al., *The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters.* Cell Research, 2015. **25**(1): p. 80-92.
166. Klattenhoff, Carla A., et al., *Braveheart, a Long Noncoding RNA Required for Cardiovascular Lineage Commitment.* Cell. **152**(3): p. 570-583.
167. Xue, Z., et al., *A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage.* Molecular Cell. **64**(1): p. 37-50.
168. Cesana, M., et al., *A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA.* Cell, 2011. **147**(2): p. 358-369.
169. Legnini, I., et al., *A Feedforward Regulatory Loop between HuR and the Long Noncoding RNA linc-MD1 Controls Early Phases of Myogenesis.* Molecular Cell, 2014. **53**(3): p. 506-514.
170. Mayer, C., et al., *Intergenic Transcripts Regulate the Epigenetic State of rRNA Genes.* Molecular Cell, 2006. **22**(3): p. 351-361.
171. Mayer, C., M. Neubert, and I. Grummt, *The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus.* EMBO Reports, 2008. **9**(8): p. 774-780.
172. Schmitz, K.-M., et al., *Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.* Genes & Development, 2010. **24**(20): p. 2264-2269.
173. Savic, N., et al., *lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs.* Cell Stem Cell, 2014. **15**(6): p. 720-34.
174. Bracken, A.P., et al., *Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions.* Genes Dev, 2006. **20**(9): p. 1123-36.
175. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-124.
176. Pasmant, E., et al., *Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF.* Cancer Res, 2007. **67**(8): p. 3963-9.
177. Pasmant, E., et al., *ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS.* The FASEB Journal, 2011. **25**(2): p. 444-448.
178. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.* Nature, 2010. **464**(7291): p. 1071-1076.
179. Portoso, M., et al., *PRC2 is dispensable for HOTAIR-mediated transcriptional repression.* The EMBO Journal, 2017. **36**(8): p. 981-994.
180. Hajjari, M. and A. Salavaty, *HOTAIR: an oncogenic long non-coding RNA in different cancers.* Cancer Biology & Medicine, 2015. **12**(1): p. 1-9.
181. Ji, P., et al., *MALAT-1, a novel noncoding RNA, and thymosin [beta]4 predict metastasis and survival in early-stage non-small cell lung cancer.* Oncogene, 0000. **22**(39): p. 8031-8041.
182. Tripathi, V., et al., *The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation.* Mol Cell., 2010. **39**.
183. Tripathi, V., et al., *Long Noncoding RNA MALAT1 Controls Cell Cycle Progression by Regulating the Expression of Oncogenic Transcription Factor B-MYB.* PLOS Genetics, 2013. **9**(3): p. e1003368.
184. Yoshimoto, R., et al., *MALAT1 long non-coding RNA in cancer.* Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2016. **1859**(1): p. 192-199.
185. Prensner, J.R., et al., *Transcriptome Sequencing Identifies PCAT-1, a Novel lincRNA Implicated in Prostate Cancer Progression.* Nature biotechnology, 2011. **29**(8): p. 742-749.
186. Prensner, J.R., et al., *The Long Non-Coding RNA PCAT-1 Promotes Prostate Cancer Cell Proliferation through cMyc.* Neoplasia, 2014. **16**(11): p. 900-908.
187. Prensner, J.R., et al., *The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex.* Nature genetics, 2013. **45**(11): p. 1392-1398.

188. Marin-Bejar, O., et al., *Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2.* Genome Biol., 2013. **14**.
189. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.* Cell, 2010. **142**(3): p. 409-19.
190. Zhou, Y., X. Zhang, and A. Klibanski, *MEG3 noncoding RNA: a tumor suppressor.* J Mol Endocrinol., 2012. **48**.
191. Mondal, T., et al., *MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures.* Nature Communications, 2015. **6**.