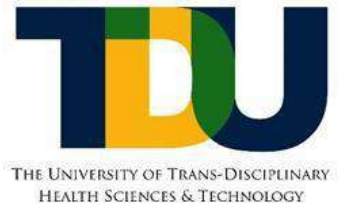
Interplay between enhancer hierarchies and histone demethylase JMJD3

A Thesis submitted to

The University of Transdisciplinary Health Sciences and Technology



For the award of the degree of

Doctor of Philosophy

By

UMER FAROOQ

Under the guidance of

Dr. DIMPLE NOTANI

National Centre for Biological Sciences (NCBS)
Bangalore, 560065

April 2023

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY
Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064

DECLARATION BY THE CANDIDATE

I declare that this thesis entitled "**Interplay between enhancer hierarchies and histone demethylase JMJD3**" submitted for the award of Doctor of Philosophy to THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY, Bengaluru, is my original work, conducted under the supervision of my guide **Dr. Dimple Notani** at National Centre for Biological Sciences (NCBS). I also wish to inform that no part of the research has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged.

I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

Place: Bengaluru

Signature of the Candidate

Date: 30/03/2023

Umer Farooq

Reg. No.: 21117020496

10/11/2017 Provisional Registration

03/12/2018 Registration

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

Private University Established in Karnataka by ACT 35 of 2013

BENGALURU - 560064

CERTIFICATE

This is to certify that the work incorporated in this thesis "**Interplay between enhancer hierarchies and histone demethylase JMJD3**" submitted by **Umer Farooq** was carried out under my supervision. No part of this thesis has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged. I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

Dr. Dimple Notani

Supervisor

NCBS, TIFR, Bangalore - 560065

Date: 31/03/2023

Acknowledgments

There are several people to whom I am grateful for their assistance and encouragement throughout my Ph.D. First of all, I am immensely thankful to my supervisor, Dr. Dimple Notani, for giving me the opportunity to work in her lab. This work would not have been possible without her guidance and unwavering support. She gave me the freedom to explore, think, and design experiments. She has always been highly tolerant of my mistakes throughout these years. I genuinely thank her and will always be grateful to her. I sincerely thank my thesis committee members, Dr. Dasaradhi Palakodeti and Dr. Sabarinathan Radhakrishnan, for their time, insightful suggestions, and discussions. I am also thankful to my teachers, who have guided me since the beginning. I want to thank Dr. Khalid Majid Fazili, Dr. Shaida Andrabi, and Dr. Mahboob ul Hussain in particular.

My heartfelt thanks to all the former and current lab members for their help, suggestions, contributions to my projects, thought-provoking discussions, and entertaining conversations that made my time in the lab pleasant. Bharath Saravanan for the assistance with bioinformatic analysis, suggestions, and discussions. Zubairul Islam for the assistance in various experiments and discussions. Sweety Meel for the help in immunostaining and FACS experiments. Deepanshu Soota and Rajat Mann for their timely assistance, suggestions, and scientific conversations. Arif Hussain for the help with immunostaining and microscopy experiments. Kaivalya Walavalkar for the help with the 4C experiments. Anurag Kumar Singh for the help with patient data analysis. Sudha Swaminathan for reviewing my manuscripts and discussions. I would like to thank Nidharshan, Vinay, Tripti, Rakesh, Ishfaq, Sachin, Urooj, and Shreyas for being wonderful lab mates. I had the pleasure of mentoring many JRFs, rotation students, trainees, etc., in the lab. I am grateful to all of them because I learned more from them than they did from me.

I want to thank NCBS for its services and technical assistance. I want to thank Dr. Awadhesh Pandit and the NGS team for assisting with next-generation sequencing experiments. I thank everyone who works at NCBS, particularly in the administrative, instrumentation, purchase, laboratory support, and stores sections. I am immensely grateful to the laboratory kitchen employees for being prompt and efficient in their jobs,

allowing our experiments to go smoothly. I would also like to thank TDU administration, particularly Mr. Ravi Kumar, for assisting me with university administrative procedures. I want to thank the Council of Scientific & Industrial Research (CSIR), Government of India (GoI), for funding my Ph.D. for five years.

I have had the privilege of knowing many wonderful people, and I will remember them fondly, notably Ashiq Hussain for being a wonderful friend, flatmate, and teacher. I am grateful for the affection and support of my friends back home, especially Jehangir and Suhail.

I am at a loss for words to express how fortunate I am to have my family's unconditional support and affection, particularly my dear parents, Muhammad Farooq and Maryam. I want to pay tribute to them for their everlasting affection, support, and love. Their teachings have molded me into the person I am today. I will never forget the countless sacrifices they made to raise me. I want to thank my grandparents, Late Abdul Qadir Wani, and Noora, for their tremendous support in helping me reach where I am today. I want to express my gratitude to my sisters, Sumiya and Mariya, for their love and support. Sumiya for several reasons, mainly for her support, love, and the blessed gift of Noorain. Mariya for her love and support, and for making our home a fun place. I am grateful to my brother Ubaid, and I praise him for taking on all of my family obligations during this phase and executing them better than I could. I want to thank Erfan for being an invaluable addition to our family. I want to thank Noorain, my niece who just joined us, bringing us so much happiness and joy.

Finally, I want to thank my wife, Salsabeel, for her unconditional love, support, and confidence in me. Her arrival in my life has been a blessing in numerous ways. I cannot describe how grateful I am to her for her patience over the past few years. I would also like to thank her for reviewing my manuscripts.

Dedicated to my family.

Their unending love, endurance, and compassion deserve my eternal appreciation.

Table of Contents

List of Tables

List of Figures

List of Acronyms

TSS	Transcription start site
SV40	Simian Virus 40
E-P	Enhancer-Promoter
eRNAs	Enhancer RNAs
Pol II	RNA Polymerase II
MPRA	Massively parallel reporter assay
sgRNAs	Single guide RNAs
TADs	Topologically associating domains
GRO-seq	Global run-on sequencing
PRO-seq	Precision nuclear run-on sequencing
LLPS	Liquid-liquid phase-separation
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
lncRNAs	Long noncoding RNAs
IFNγ	Interferon gamma
ChIP	Chromatin immunoprecipitation
SEs	Super Enhancers
PCR	Polymerase chain reaction
hESCs	Human Embryonic Stem Cells
siRNA	small interfering RNA
shRNAs	short hairpin RNAs
ATAC-seq	Assay for Transposase-Accessible Chromatin with high-throughput sequencing
GWAS	Genome-wide association studies
SNPs	Single nucleotide polymorphisms
IDRs	Intrinsically disordered regions
INK4a	Inhibitor of CDK4
INK4b	Inhibitor of CDK4
ARF	Alternate open reading frame
CDKN2A	Cyclin-Dependent Kinase Inhibitor 2A
CDKN2B	Cyclin-Dependent Kinase Inhibitor 2B
CDKN2BAS	Cyclin-Dependent Kinase Inhibitor 2B Anti-Sense

PcGs	Polycomb group genes
ANRIL	Antisense non-coding RNA in the INK4 locus
JMJD3	Jumonji domain-containing protein D3
JmjC	Jumonji C
EZH2	Enhancer of zeste homolog 2
Suz12	Suppressor of zeste
CRISPRi	CRISPR-mediated repression
TCGA	The Cancer Genome Atlas

Synopsis

Introduction

Enhancers in transcription regulation

Cell differentiation, development, responses to external stimuli, and other processes rely on precise and accurate gene expression, which is regulated at various stages by several protein factors, RNAs, and DNA regulatory elements (Buccitelli and Selbach 2020; Pope and Medzhitov 2018). The key DNA regulators of transcription are cis-regulatory DNA elements called enhancers. They are short DNA elements ranging from 100 to 1500 bp and induce target gene transcription regardless of their proximity, orientation, or placement with respect to the gene promoter (Panigrahi and O'Malley 2021; Banerji, Rusconi, and Schaffner 1981). Enhancers are cell type-specific, meaning a specific set of enhancers that are active in one cell type may be inactive in another. Certain modifications, such as H3K4me1 and H3K27ac, are found on the histones of active enhancers (Li, Notani, and Rosenfeld 2016). Active enhancers exhibit transcription factor and cofactor binding and are hypersensitive to DNase I treatment due to their core accessibility. Furthermore, because of their ability to recruit PolII, active enhancers are transcribed into non-coding RNAs known as enhancer RNAs (eRNAs) (Li, Notani, and Rosenfeld 2016). These non-coding RNAs have been shown to increase target promoter transcription by enhancing enhancer activity (Arnold, Wells, and Li 2019; Li et al. 2013). Enhancer-mediated gene transcription regulation is thought to occur via looping between the enhancer and the target gene promoter (Panigrahi and O'Malley 2021). Although enhancer-promoter (E-P) looping is a widely accepted model of gene regulation, the underlying mechanism is unclear.

Enhancers can appear in clusters composed of multiple individual enhancers that coexist in close linear proximity. These clustered enhancers are called super-enhancers if the enhancers within the cluster occur in close linear proximity to one another and exhibit exceptionally high enrichment of enhancer-associated marks (Hnisz et al. 2013). These enhancers can phase separate by liquid-liquid phase-separated (LLPS) due to the abundance of transcription factors and cofactors such as Mediator (Sabari et al. 2018; Whyte et al. 2013). Super-enhancers are primarily responsible for driving the expression of genes important for cell identity (Hnisz et al. 2013). These enhancers, for example, have been discovered near pluripotency genes such as Sox2, Nanog, and

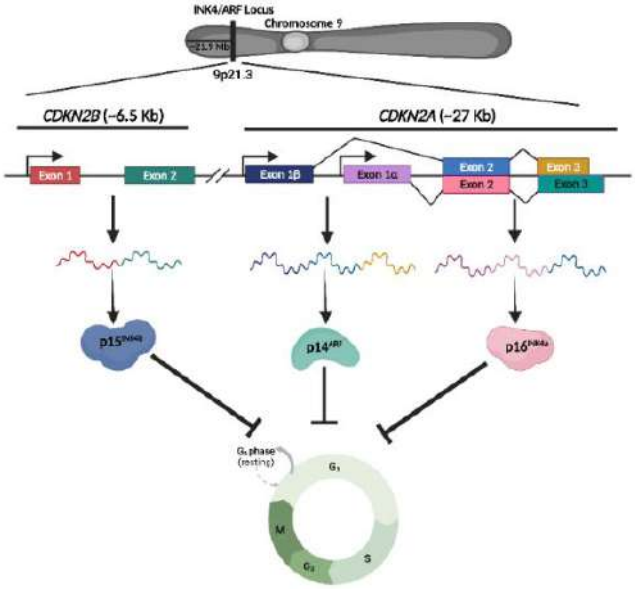
others in mESCs (Li et al. 2014; Blinka et al. 2016). Furthermore, super-enhancers are abundant around oncogenes in cancer cells and have been shown to increase the expression of these genes dramatically. For instance, they have been found in the MYC locus in several cancers (Schuijers et al. 2018). Thus, understanding how super-enhancers work holds enormous therapeutic potential.

The thesis aimed to understand how dense enhancer clusters function. Do all the individual enhancers that form the cluster regulate target gene transcription or only a subset is important? What is the relationship between functional enhancers? Thus, to dissect the individual regulatory enhancer elements from the dense enhancer cluster and to understand the intrinsic properties of these regulatory elements, we chose *INK4/ARF* locus. This locus harbors a very dense enhancer cluster. Moreover, the enhancer cluster is already associated with several aging-associated diseases like coronary artery disease (CAD), type 2 diabetes, and several cancers.

***INK4/ARF* locus:**

The *INK4/ARF* locus contains two crucial cell cycle regulatory genes: *CDKN2A*, which codes for $p14^{ARF}$ and $p16^{INK4a}$, and *CDKN2B*, which codes for $p15^{INK4b}$. Another gene at this locus, *CDKN2BAS*, codes for the lncRNA ANRIL. Although the mRNA sequences of $p14^{ARF}$ and $p16^{INK4a}$ are very similar, the resulting proteins differ due to different reading frames. However, the amino acid sequences of $p16^{INK4a}$ and $p15^{INK4b}$ encoded by different genes are very similar. Thus, gene duplication events are thought to have resulted in the *CDKN2A* and *CDKN2B* genes (López et al. 2017). Despite their differences, all three proteins play cell cycle regulatory roles and cause cell cycle arrest, primarily in the G1 phase. In normal cells, these genes are suppressed, but they become active as the cell ages and enters senescence. Furthermore, this locus is either deleted or silenced by repressive machinery in most cancers. These proteins, however, are overexpressed in a few cancers, including head and neck carcinoma and HPV-expressing cancers like cervical cancer. This is made possible by two viral proteins, E6 and E7. Because E6 induces degradation of p53 and E7 inhibits Rb, these proteins

cannot activate their downstream targets which results in cell cycle inhibition (Tommasino and Crawford 1995).



Schematic representing the *INK4/ARF* locus. *This locus contains two critical cell cycle regulatory coding genes: CDKN2A and CKDN2B. CDKN2A encodes two proteins, p14ARF and p16INK4a, whereas CDKN2B encodes p15INK4b. Though distinct, all three proteins perform cell cycle regulatory functions and arrest the cell cycle at the G1 phase.*

Regulation of *INK4/ARF* locus via distal regulatory elements

A long gene desert region exists upstream of the *INK4/ARF* locus, and SNPs in this region have been linked to various diseases, such as coronary artery disease, type 2 diabetes, and multiple cancers. In mice, deletion of this syntenic CAD interval reduced the expression of these genes significantly (Visel et al. 2010). A seminal study discovered several enhancers present in the gene desert to understand the functional relevance of gene desert in regulating this locus. Furthermore, SNP in one of the enhancers disrupts the STAT1 motif, causing the genes to be downregulated in response to IFNγ signaling (Harismendy et al. 2011). TGFβ is required to activate upstream enhancers and binds to one of the CAD interval enhancers, promoting ARF expression during development. TGFβ also induces three H3K27ac peaks in the gene desert, and deletion of the entire interval reduces INK4b and ARF expression (Zheng et al. 2013; Liu et al. 2019). Furthermore, people with the SNP rs10757278 have significantly lower expression of all three genes in T cells in their peripheral blood. A cis-regulatory element near the ARF promoter, in addition to upstream regulatory elements, inhibits

INK4a gene expression by forming a loop with its promoter (Zhang, Hyle, et al. 2019). All studies link gene desert enhancers to *INK4/ARF* locus transcription regulation. However, how multiple enhancers in this gene desert activate the locus was unknown. My work towards the doctoral thesis has revealed an intricate relationship among functional enhancers in this gene desert.

Regulation of *INK4/ARF* locus via chromatin modifiers

1. **By Polycomb Complexes:**

Polycomb group proteins (PcGs) are methyltransferase enzyme complexes necessary for development, tissue repair, differentiation, aging, and other processes. PcGs form two different multi-subunit complexes: PRC1 and PRC2 (Chittock et al. 2017). PRC1 is comprised of several core subunits, including RING1A/B, PCGF2/4, CBX2/4/6/7/8, and PHC1/2/3. In contrast, the PRC2 complex is comprised of core subunits such as SUZ12, EZH2, RbAp46/48, and EED (Chittock et al. 2017; Kerppola 2009). In addition to core subunits, other auxiliary subunits have been discovered that aid the function of both these complexes (Kerppola 2009). Both PRC1 and PRC2 complexes are required for *INK4/ARF* locus suppression in proliferating cells. PRC2 accomplishes this by binding to and trimethylating the lysine 27 of histone 3 (H3K27me3) at promoters of *INK4/ARF* genes (Bracken et al. 2007). Ectopic expression of PRC1 and PRC2 subunits suppresses this locus and delays aging (Dietrich et al. 2007). On the other hand, downregulation of the PcG subunits results in the activation of this locus and premature senescence.

2. **By JMJD3:**

JMJD3 is a histone demethylase essential for development, infectious disease progression, cancer progression, senescence, etc. (Xiang et al. 2007; Zhang, Liu, et al. 2019). It acts as an antagonist to the PRC2 complex by removing methyl marks from Histone 3 trimethylated at lysine 27 (H3K27me3). JMJD3 is a well-known positive regulator of the *INK4/ARF* locus and during senescence, it activates this locus by removing the repressive methylation (Barradas et al. 2009). JMJD3 is activated by cellular signals such as oncogene overexpression which activates the *INK4/ARF* locus

(Agger et al. 2009). As a result, *INK4/ARF*-mediated cell cycle arrest occurs in MEFs and IMR90 cells. Furthermore, JMJD3 acts as a negative regulator for reprogramming by activating the *INK4/ARF* locus (Zhao et al. 2013). Silencing JMJD3 alone increases reprogramming efficiency by orders of magnitude. Surprisingly, inhibiting JMJD3 with INK4a or ARF improves reprogramming efficiency even further (Zhao et al. 2013).

The study found that only a few enhancers that looped with the promoter were important and they had an intricate mutual relationship. Later, the JMJD3 (a histone demethylase) was functionally tied with the enhancer functions and hierarchy. A brief outline of thesis chapters that include, the description of the *INK4/ARF* locus, its regulation, results of the study, material and methods used, interpretations, references, and publications are mentioned below.

Thesis Organization

Chapter 1 (Introduction) of the thesis discusses the existing literature on DNA regulatory elements called enhancers and their role in transcription regulation. It also introduces other aspects of enhancers, including active enhancer signatures, methods and approaches for identifying regulatory enhancers, and the role of enhancers in higher-order chromatin organization. Furthermore, it also summarizes the effects of SNPs in regulatory enhancers and their impact on target gene transcription and subsequent association with several diseases. This chapter also introduces the *INK4/ARF* locus and its role in cell cycle regulation. Furthermore, it also explains the role of this locus in cancers, reprogramming, and senescence. Moreover, this chapter also focuses on the known cis and trans mechanisms of transcription regulation of this locus.

Chapter 2 (Findings-Part I) explains how individual functional enhancers that regulate the *INK4/ARF* locus were dissected from a dense enhancer cluster. The first section of this chapter discusses the identification of an enhancer cluster upstream of the *INK4/ARF* locus and its relationship to *CDKN2A* and *CDKN2B* gene expression. This enhancer cluster was found to contain 24 individual enhancers in HeLa, only 15 of which are active. The active enhancers also correlate with the expression of *INK4/ARF* genes in this cell line. The second section discusses the dissection of

regulatory enhancers from this dense enhancer cluster. Only 5 of the 15 active enhancers were found to be interacting with the promoters. CRISPR-Cas9 perturbations were used to test the regulatory potential of promoter-interacting enhancers. It was found that promoter-interacting enhancers are required for *INK4/ARF* transcription, and they were found to be interdependent on one another. The third section discusses the impact of perturbation of these enhancers on the cancer phenotype of the cells. The deletion of the promoter-interacting enhancers affects the cancerous properties of the cells, such as the colony-forming potential, cell migration, etc. The final section discusses the mechanism of gene expression downregulation in the absence of enhancers. When promoter-interacting enhancers are disrupted, EZH2, a component of the repressive complex, is loaded onto the promoters of these genes, explaining why transcription is downregulated in the absence of promoters. Furthermore, the genes that are downregulated when the enhancers are deleted corroborate with the disease association of this locus.

Chapter 3 (Findings-Part II) shows that JMJD3 binding to *INK4/ARF* enhancers and other genome-wide regions depends on its N-terminal region containing an intrinsically disordered region (IDR) and the RNA binding domain. The first part discusses the genome-wide binding of JMJD3 on chromatin. It was observed that JMJD3 binds more to enhancer elements than promoters across the genome. Its binding pattern on regulatory elements differs from that of UTX, which binds more to promoters than enhancers. The second section explains how JMJD3's chromatin binding depends on its IDR. When the IDR of JMJD3 is removed, it no longer binds to chromatin. However, removing half of its IDR alters its binding pattern across the genome. The third section discusses JMJD3's chromatin binding dependence on RNA. We found that JMJD3 is an RNA-binding protein and using domain mapping, it was found that its C-terminus, a Zinc finger-like domain, binds to RNA. Furthermore, JMJD3 chromatin targeting is RNA-dependent, and it binds to the *INK4/ARF* locus enhancers in an eRNA-dependent manner. The final section discusses the effects of removing its IDR region and mutating its RNA binding on its catalytic activity. Removing half of JMJD3's IDR does not affect its catalytic activity, but removing the entire IDR reduces its catalytic activity. Moreover, RNA-binding mutants also show reduced catalytic activity.

Chapter 4 (Materials and Methods) describes the materials and methods used in this

study. To address the questions of Chapter 2 (Finding Part-I), we took advantage of techniques and approaches like Chromatin Immunoprecipitation (ChIP), CRISPR-Cas9 assays, Chromosome conformation capture techniques (4C), Gene expression, etc. To address the questions of Chapter 3 (Finding Part II), we utilized approaches like ChIP-seq, RNA Immunoprecipitations (UV-RIP), RNA Pulldown Assays, Site-directed mutagenesis, and Subcellular fractionation.

Chapter 5 (Conclusions and Discussion) summarizes the key findings from chapter 2 and chapter 3. Our results in chapter 2 highlight that only a subset of enhancers from a dense enhancer cluster can regulate the target genes. The enhancers interacting with the promoters are critical for the transcriptional regulation of the *INK4/ARF* locus. The promoter-interacting enhancers are mutually dependent on each other for function. The perturbation of one enhancer in the cluster affects the other enhancers in the same cluster. The deletion of promoter-interacting enhancers affects the cancerous phenotype of the cells by reducing their colony-forming potential and delaying migration confirming, the role of this locus in the cell cycle and cancer. When regulatory enhancers are deleted, EZH2 is loaded onto *INK4/ARF*, causing the genes to be down-regulated. Furthermore, the dysregulated genes after enhancer deletion correlate with the diseases this locus is associated with. Our findings in Chapter 3 show that JMJD3 binds to enhancers more than promoters. The binding of JMJD3 on chromatin is dependent on its N-terminal IDR, the complete removal of which affects its chromatin binding whereas, the removal of half IDR does not affect the chromatin binding but its target specificity is altered. These results suggest that the IDR of JMJD3 is essential for chromatin binding and determining target specificity. Furthermore, JMJD3 chromatin binding is RNA-dependent, and mutations in its RNA binding domain affect chromatin binding.

List of Publication

Farooq, U., Saravanan, B., Islam, Z., Walavalkar, K., Singh, A.K., Jayani, R.S., Meel, S., Swaminathan, S., Notani, D. An interdependent network of functional enhancers regulate transcription and EZH2 loading at INK4a/ARF locus. Cell Reports, 34 (12), 108898. (2021)

Farooq, U., Notani, D. Optimized protocol to create deletion in adherent cell lines using CRISPR/Cas9 system. STAR Protocols, 2(4), 100857. (2021)

Farooq, U.* and Notani, D*, Transcriptional regulation of INK4/ARF locus by cis and trans mechanisms. Front. Cell Dev. Biol. 10:948351. (2022) *Co-Corresponding author

Walavalkar, K., Saravanan, B., Singh, A.K., Jayani, R.S., Nair, A., **Farooq, U.**, Islam, Z., Soota, D., Mann, R., Padubidri, S.V., Freedman, M.L. Sabarinathan. R., Haiman, C., Notani, D. A rare variant of African ancestry activates 8q24 lncRNA hub by modulating a cancer. Nature communications, 11, 3598. (2020)

Saravanan, B., Soota, D., Islam, Z., Majumdar, S., Mann, R., Meel, S., **Farooq, U.**, Walavalkar, K., Gayen, S., Singh, A.K., Hannenhalli, S. and Notani, D. Ligand dependent gene regulation by transient ERα clustered enhancers. PLoS genetics, 16(1), p.e1008516. (2020)

Islam, Z.*, Saravanan, B.*, Walavalkar, K., **Farooq, U.**, Singh, A.K., Thakur, J., Pandit, A., Sabarinathan, R., Henikoff, S., Notani, D. Active enhancers strengthen insulation by RNA-mediated CTCF binding at chromatin domain boundaries. Genome Res. Jan;33(1):1-17 (2023)

Chapter 1: Introduction

Gene regulation at the transcriptional level

Complex multicellular organisms are composed of various cell types derived from the same cell. Despite sharing the same DNA code, these cell types exhibit striking differences in physiology, morphology, function, etc. These differences primarily result from differential gene regulation at various stages, such as transcriptional, post-transcriptional, translational, etc. (Casamassimi and Ciccodicola 2019). Precise and accurate gene expression is critical for cell differentiation, development, homeostasis, reactions to external stimuli, disease, and other processes. Transcription initiation is one of the fundamental stages at which gene expression is regulated. Transcription is a process in which the information stored in DNA is decoded by RNA Polymerases, a multi-subunit enzyme, to synthesize RNA. A diverse set of transcription factors, cofactors, and chromatin-modifying enzymes perform the intricate task of controlling gene transcription. These proteins bind to DNA elements like proximal and distal regulatory elements. Proximal DNA elements, including core promoters and promoter-proximal elements, are mainly present near the transcription start site (TSS) of the genes. These elements can direct transcription independently for some genes, but only to basal levels. On the other hand, the distal regulatory elements are usually present several kilobases to even megabases away from the target genes; these include enhancers, insulators, silencers, etc. Among the distal regulatory elements, enhancers mainly activate gene transcription and significantly affect gene expression (Lee and Young 2013).

Enhancers as transcription regulatory elements

Enhancers are cis-regulatory DNA elements essential for gene regulation at the transcriptional level. Enhancers are short DNA sequences ranging from 100 to 1500 bp in length that induce target gene transcription regardless of their distance, orientation, or placement from the gene promoter (Banerji, Rusconi, and Schaffner 1981). These DNA elements were initially discovered 40 years ago in Simian Virus 40 (SV40) as a 72 bp DNA repeat sequence capable of triggering transcription of the human β-globin gene regardless of its distance and orientation from the promoter (Banerji, Rusconi, and Schaffner 1981). Shortly after this finding, enhancers were discovered in bacteria and

metazoans including humans (Xu and Hoover 2001). Interestingly, enhancers exhibit cell type specificity; thus, a specific set of enhancers active in one cell type may be inactive in another. In addition, though enhancers have many characteristics in common with other regulatory elements, such as promoters, their ability to stimulate gene transcription over long distances distinguishes them from the promoters.

Molecular properties of active enhancers

Active enhancers share some characteristics with other DNA regulatory elements, such as promoters, but some features are unique to enhancers. The histones on the active enhancers are enriched with certain modifications, such as H3K4me1 and H3K27ac (Calo and Wysocka 2013; Creyghton et al. 2010). The presence of H3K4me1 and lack of H3K4me3 distinguishes enhancers from promoters. Thus, the H3K4me1 to H3K4me3 ratio has been used to distinguish enhancers from promoters. Furthermore, the active enhancers have a nucleosome-free core that contains motifs for sequence-specific transcription factors. Due to the core accessibility, active enhancers exhibit transcription factor and cofactor binding and show hypersensitivity to DNase I treatment (Spitz and Furlong 2012; Thurman et al. 2012). Moreover, active enhancers are also transcribed into non-coding RNAs known as enhancer RNAs (eRNAs), due to their ability to recruit Pol II (Lam et al. 2014; Li, Notani, and Rosenfeld 2016a). These non-coding RNAs have been shown to aid enhancer activity, thereby increasing target promoter transcription (Arnold, Wells, and Li 2020).

Identification of enhancers

Several techniques have been developed to identify regulatory enhancers, although these assays currently have drawbacks. Thus, a single method cannot accurately detect enhancers; combining more than one technique is utilized to discover these elements. Some of the strategies used in the field for identifying enhancers are as follows:

1. Enhancer reporter assay: This is the most extensively used method to identify enhancers in the field. The DNA region of interest is cloned in a construct containing a reporter gene (typically luciferase gene) driven by a minimal promoter (Shlyueva,

Stampfel, and Stark 2014). The region of interest is cloned away from the promoter, and the effect on reporter gene expression is determined.

2. STARR-Seq: A massively parallel reporter assay (MPRA) is employed to detect regulatory enhancers (Melnikov et al. 2012). This method involves fragmenting genomic DNA and inserting individual fragments into a vector downstream to a gene driven by a minimal promoter. If the fragment is an enhancer, it will increase the reporter gene expression, including the enhancer including a distinct barcode. The RNA is then subjected to high throughput sequencing, which allows functional enhancers to be identified (Arnold et al. 2013). When compared to standard reporter-based assays, this technique has greater scalability.

3. CRISPR Interference: CRISPR-based assays have recently enabled us to uncover functional enhancers in the genomic context (Pickar-Oliver and Gersbach 2019a). Specific guide RNAs (gRNAs) are designed against the region of interest in CRISPRi. These gRNAs direct catalytically dead mutant of Cas9 protein called dCas9 fused to a regulatory protein such as VP64 and KRAB to the region of interest. If the region functions as an enhancer for a particular gene, its activation or repression would affect the transcription of that gene accordingly.

4. CRISPR Knockout: CRISPR is a highly effective method for deleting any genomic region from cells/tissues. In this technique, two gRNAs flanking the region of interest are designed that recruit Cas9 protein to the target site. Cas9 is an endonuclease that causes a double-strand break at the target site, so two gRNAs flanking a site would result in its deletion (Pickar-Oliver and Gersbach 2019b). However, there are certain limitations to CRISPR. For example, nonspecific targeting of gRNAs might result in genomic lesions, which can cause cell cycle halt or other cellular abnormalities caused by the activation of DNA damage response.

Transcription at enhancers

Pol II recruitment on enhancers induces transcription at these regulatory elements. These non-coding transcripts, mainly <200 nt long, are known as enhancer RNAs (eRNAs) (Li, Notani, and Rosenfeld 2016b). Since enhancers lack a defined

transcription start site, eRNA transcription begins in the nucleosome-free region bound by transcription factors. Most of the enhancers are transcribed into bidirectional eRNAs; thus, the transcription at these enhancers is divergent (Li, Notani, and Rosenfeld 2016b). The majority of the known eRNAs are neither spliced nor polyadenylated (Kim et al. 2010). Moreover, eRNAs have a fast turnover rate than other RNAs and are rapidly degraded by exosomes (Andersson et al. 2014). Consequently, they are difficult to detect using conventional steady-state RNA sequencing. Thus, unique approaches such as GRO-seq or PRO-seq are extensively utilized to detect the eRNAs (Core, Waterfall, and Lis 2008; Kwak et al. 2013).

Enhancer-promoter communication

Enhancer-mediated control of gene transcription is thought to occur by looping between the enhancer and the target gene promoter. Although enhancer-promoter (E-P) looping is a commonly accepted model of gene regulation, the underlying mechanism remains unclear (Novo et al. 2018). Several studies have shown that chromosome structural proteins like cohesin play a vital role in E-P communication (Pherson et al. 2019). Cohesin degradation results in the loss of most E-P interactions, but a few interactions are retained, and new interactions are gained. This highlights that both cohesin-dependent and independent mechanisms might exist that help establish E-P loops (Thiecke et al. 2020). Furthermore, the stability of E-P interaction is also debated. Recently, multiple studies indicated that E-P communication is highly dynamic compared to inactive regions in the genome (Hatzis and Talianidis 2002; Yokoshi and Fukaya 2019). These studies primarily took advantage of live cell imaging, where enhancers and promoters were tagged with distinct fluorophores (Gu et al. 2018). The dynamicity of E-P interactions is consistent with the idea of transcription bursting, which posits that transcription of any single gene happens in occasional surges rather than continuously. The bursting could thus be the result of the dynamic nature of E-P interactions.

Enhancers and Genome Organization

Genomic DNA stores genetic information, which is then expressed as RNA and, finally, as proteins. Human diploid DNA is over 2 meters long when laid end to end; however, the typical human nucleus is only 5μM to 10μM in diameter. Consequently, DNA must be extensively compacted to fit inside the nucleus. However, critical processes such as transcription, DNA replication, and DNA repair that require loose chromatin should also be possible with compacted DNA. This emphasizes the importance of organizing the genome to balance compaction and accessibility for other processes. Thus, DNA is organized in complex higher-order structures in the nucleus that emerge due to multilevel organization (Figure 1.1). The first level of organization is formed when DNA is wrapped around histone proteins forming chromatin fibers. This process efficiently reduces the length of DNA by 7-folds. Next, different DNA regulatory elements, such as enhancers, promoters, insulators, etc., interact with one another to further compact the chromatin. These interactions, which can occur between enhancers and promoters, insulators and insulators, etc., primarily result in the formation of chromatin loops. The formation of loops not only compacts chromatin but also performs gene regulatory functions. Topologically associating domains (TADs) are formed when multiple such loops are contained within a large loop. TAD formation requires structural proteins such as CTCF and cohesin, which do so via a process known as loop extrusion. Furthermore, TADs segregate into gene active (compartment A) and gene inactive (compartment B) compartments, formed due to multiple TADs interacting with each other based on their transcription status. Finally, at the highest level, the genome is organized into chromosomes that occupy specific regions of the nucleus, resulting in the formation of chromosome territories.

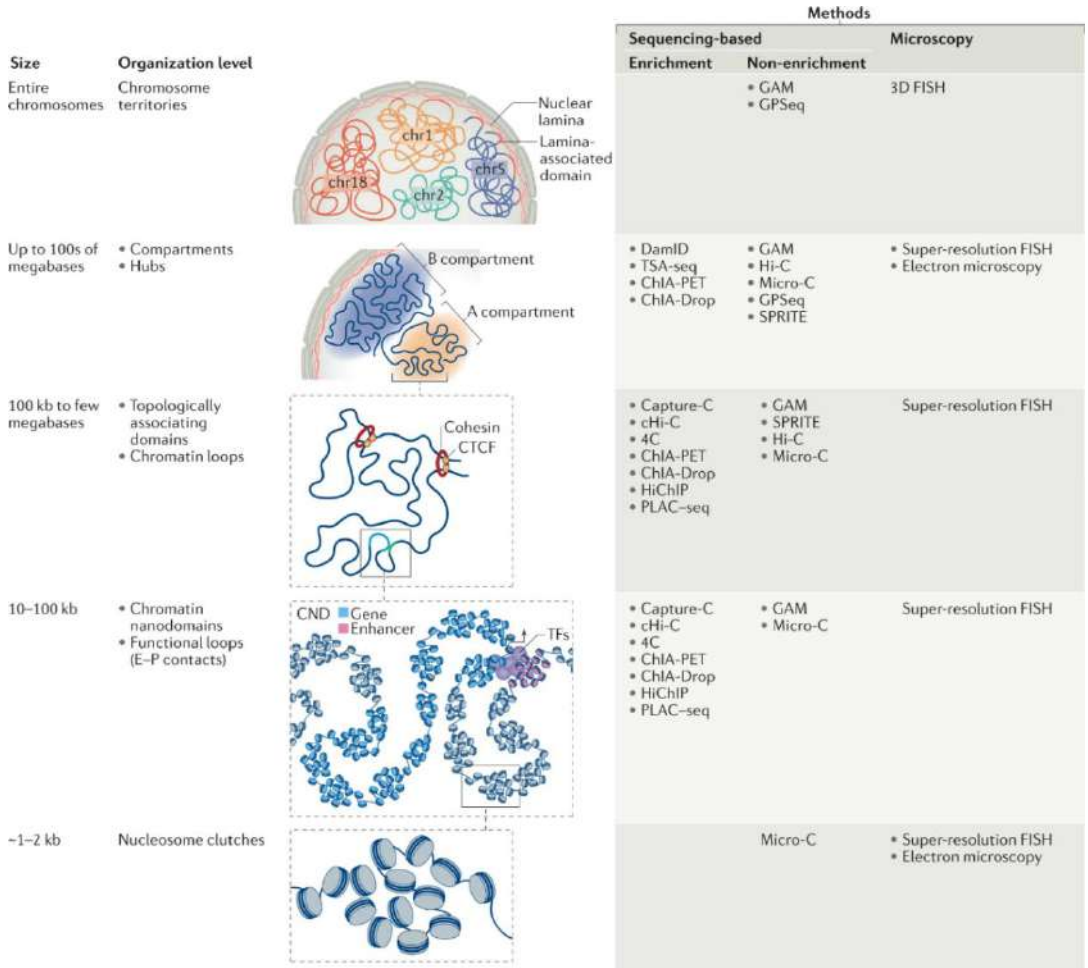


Figure 1.1: Different levels of genome organization. *This schematic shows the different levels of genome organization and the size of the domains. Furthermore, it highlights the sequencing and microscopy-based methods currently utilized to study the levels of genome organization. (Adapted from Jerkovic I, Cavalli G. Nature reviews. Molecular Cell Biology. (2021)).*

Super enhancers

The term 'super enhancer' refers to the cluster of multiple enhancers that occur in close linear proximity to one another and exhibit very high enrichment of enhancer-associated marks (Hnisz et al. 2013; Whyte et al. 2013). Super-enhancers are identified based on the linear distance between individual units of super-enhancers, which should not exceed 12.5 kb (Peng and Zhang 2018; Pott and Lieb 2015). Individual units within the distance limit are stitched together and regarded as being part of the same cluster, which can range in size from 10 kb to 60 kb. These enhancers are not restricted to a

single cell type but can be found in nearly all types of cells. Super-enhancers exhibit cell type specificity; hence, a particular combination of super-enhancers active in one cell type might not be active in a different cell type. Because of the high abundance of transcription factors, these enhancers can generate liquid-liquid phase-separated (LLPS) entities (Sabari et al. 2018; Jia, Chng, and Zhou 2019). Super-enhancers primarily drive the expression of genes critical for cell identity. In mESCs, for example, super-enhancers have been discovered near pluripotency genes such as sox2, oct4, and others (Whyte et al. 2013). Furthermore, super-enhancers are abundant around oncogenes in cancer cells and have been demonstrated to stimulate the expression of these genes to extremely high levels. For example, super-enhancers have been discovered in the MYC locus in several malignancies (Dave et al. 2017). Understanding how super-enhancers work would thus hold enormous therapeutic promise.

Enhancer mutations and disease

Given the role of enhancers in transcriptional control, changes in enhancer activity can lead to changes in gene expression. Single nucleotide polymorphisms (SNPs) have been found in enhancer elements in several genome-wide association studies (GWAS). Some SNPs have been linked to disorders such as Crohn's disease, systemic lupus, multiple sclerosis, and several cancers (Claringbould and Zaugg 2021). SNPs within enhancers influence the activity of the enhancer by modifying the transcription factor binding motifs, which in turn affects the transcription of the target gene. The SNPs could either lead to the gain or loss of the transcription factor binding motif. The effect of SNP would thus be determined by the type of transcription factor (activator or repressor) that acquires or loses the motif. For instance, it has been discovered that the 9p21 locus contains a dense enhancer cluster, and SNP in one of the enhancers mutates the STAT1 binding motif. Upon interferon-gamma signaling, this SNP alters the expression of the *CDKN2A, CDKN2B*, and *CDKN2BAS* genes (Harismendy et al. 2011). Moreover, different intervals in the 8q24 locus have been linked to prostate, colorectal, and other malignancies (Haiman et al. 2007). Several enhancer elements have been shown to interact with one another to regulate MYC gene transcription (Sotelo et al. 2010; Tuupanen et al. 2009). Multiple independent studies discovered an SNP in one of the enhancers linked to colorectal and prostate cancer. The presence of the SNP in the enhancer results in the alteration of the TCF7L2 motif (Tuupanen et al. 2009;

Pomerantz et al. 2009). Another SNP in a separate enhancer impairs FOXA1 binding, resulting in changes in the MYC gene expression (Jia et al. 2009).

INK4/ARF locus

INK4/ARF is a well-known locus that harbors two crucial cell cycle regulatory genes: *CDKN2A*, which codes for $p14^{ARF}$ and $p16^{INK4a}$, and *CDKN2B*, which codes for $p15^{INK4b}$. This locus also contains another gene, *CDKN2BAS*, which codes for a lncRNA ANRIL (Farooq and Notani 2022). The mRNA sequences of $p14^{ARF}$ and $p16^{INK4a}$ are very similar, but the resulting proteins differ due to alternative reading frames. However, $p16^{INK4a}$ and $p15^{INK4b}$, encoded by different genes, have very similar amino acid sequences (López et al. 2017). Thus, the *CDKN2A* and *CDKN2B* genes are thought to have arisen due to gene duplication events. Despite their differences, all three proteins are involved in cell cycle regulatory functions and cause cell cycle arrest, primarily at the G1 phase (Figure 1.2). These genes are repressed in normal cells and expressed when the cell ages and enters senescence (Farooq and Notani 2022). Furthermore, this locus is either deleted or silenced by repressive machinery in most cancers. However, these proteins are highly expressed in a few cancers, including head and neck carcinoma and HPV-expressing cancers, such as cervical cancer. This is made possible by two important viral proteins known as E7 and E6. E6 inhibits Rb, and E7 inhibits p53, and as a result, these proteins are unable to activate their downstream targets that lead to cell cycle inhibition (Kanao et al. 2004; Münger et al. 1992).

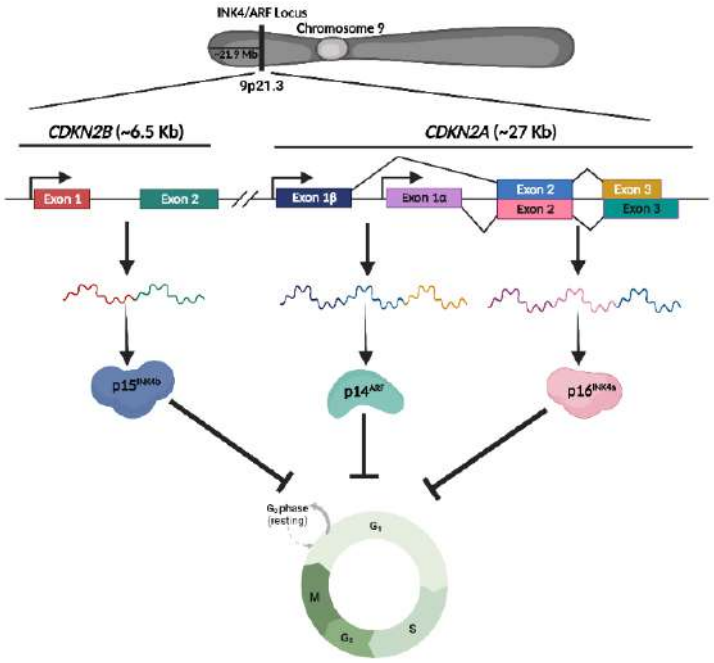


Figure 1.2: Schematic depicting the *INK4/ARF* locus. *INK4/ARF locus harbors two coding genes: CDKN2A and CKDN2B. CDKN2A encodes two proteins, $p14^{ARF,}$ and $p16^{INK4a}$, whereas CDKN2B encodes $p15^{INK4b}$. All three proteins are cell cycle regulatory proteins and arrest the cell cycle at the G1 phase once expressed. (Adapted from Farooq U and Notani D. Frontiers in Cell and Developmental Biology (2022)).*

INK4/ARF locus and cell cycle

Various stress signals, including DNA damage and oncogene overexpression, activate the *INK4/ARF* locus (Romagosa et al. 2011). When these proteins are expressed, they activate their downstream targets that cause cell cycle arrest (Ivanchuk et al. 2001). $p14^{ARF}$, a well-studied tumor suppressor protein, prevents p53 degradation. $p14^{ARF}$ accomplishes this by interacting with MDM2, a ubiquitin ligase, resulting in MDM2 sequestration in the nucleolus (Maggi et al. 2014; Weber et al. 1999). As a result, MDM2 is prevented from interacting and subsequent degradation of p53. Consequently, p53 enters the nucleus and activates genes required for G1 arrest (Figure 1.3A) (Maggi et al. 2014). $p16^{INK4a}$ and $p15^{INK4b}$ control a distinct pathway leading to cell cycle arrest at the G1 phase (Kim and Sharpless 2006). These proteins mediate cell cycle arrest by activating Rb (retinoblastoma) pathway. CDK4/6 interacts with cyclin D to form an active complex (Cobrinik 2005). This complex binds to the Rb protein and phosphorylates it at multiple sites. The phosphorylation prevents Rb from interacting with and inhibiting E2F, a transcription factor required for the G1-S transition (Dimova and Dyson 2005). Once expressed, $p16^{INK4a}$ and $p15^{INK4a}$ bind to CDK4/6 and prevent it from forming a complex with cyclin D. This keeps Rb hypophosphorylated and allows it to interact with E2F, preventing E2F from activating the genes required for the G1 to S phase transition and effectively arresting the cell cycle at the G1 stage (Figure 1.3B) (Hannon and Beach 1994; Russo et al. 1998; Giacinti and Giordano 2006).

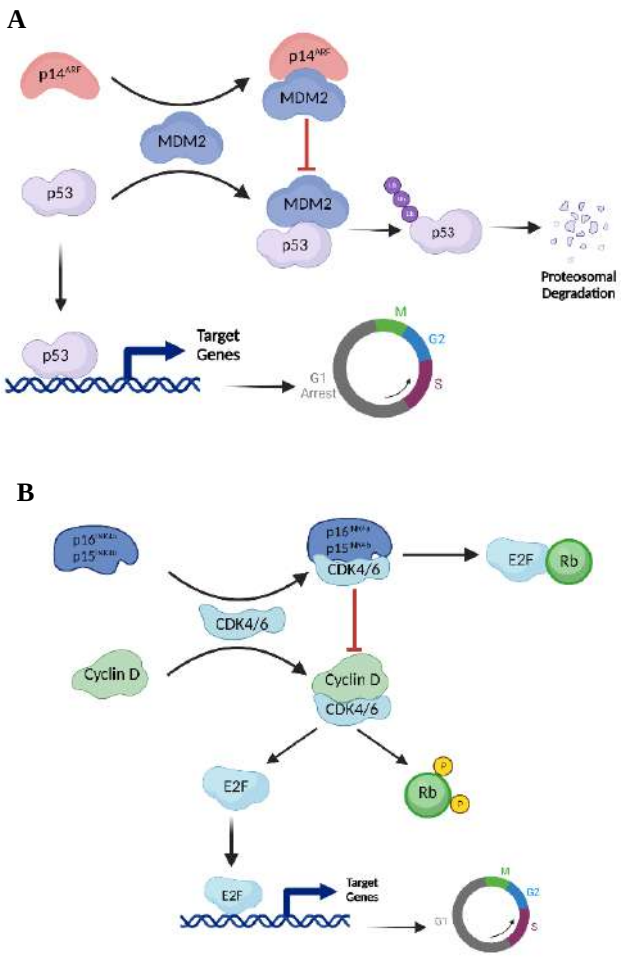


Figure 1.3: *INK4/ARF* locus in action. ***A)*** *$p14^{ARF}$ regulates the p53-MDM2 pathway. Its interaction with MDM2 results in p53 stabilization. P53 moves to the nucleus to activate its target genes, resulting in cell cycle arrest at the G1 phase.* ***B)*** *$p16^{INK4a}$/$p15^{INK4b}$ participate in activating a completely distinct pathway. These proteins prevent the complex formation of Cyclin D-CDK4/6, resulting in the hypophosphorylation of Rb. Hypophosphorylated Rb binds to E2F, which doesn't allow E2F to activate its target genes required for G1 to S phase transition. (Adapted from Farooq U and Notani D. Frontiers in Cell and Developmental Biology (2022)).*

INK4/ARF locus in cellular senescence

Cellular senescence is the permanent arrest of cell division in response to certain intrinsic or extrinsic stimuli. This phenomenon serves as the first line of defense against cancer progression (Prieto and Baker 2019). However, senescence leads to aging-

related disorders by impairing repair and regeneration (McHugh and Gil 2018). $p16^{INK4a}$ is a well-known senescence driver (Krishnamurthy et al. 2004; Rayess, Wang, and Srivatsan 2012). In response to oncogene insult, $p16^{INK4a}$ has been shown to cause senescence (Lin et al. 1998; Zhu et al. 1998). The expression of $p16^{INK4a}$ increases, when fibroblasts and other cell types approach replicative senescence (Mirzayans et al. 2012).

INK4/ARF locus in cancers

Cancer cells divide indefinitely because their cell cycle is no longer regulated. Mutations in cell cycle regulatory genes are one of the common events that allow cancer cells to divide indefinitely (Collins, Jacks, and Pavletich 1997). Because the *INK4/ARF* locus contains cell cycle regulatory genes, it is either deleted or silenced in the majority of cancers (Sherr 2012; Forbes et al. 2006). In other cancers, these genes are highly silenced by DNA methylation or histone modification at promoters. Furthermore, in some cancers point mutations that render $p16^{INK4a}$ inactive have also been reported (Forbes et al. 2006).

INK4/ARF locus in reprogramming

Somatic cells can be transformed into pluripotent cells by ectopic expression of a few key transcription factors such as Oct4, Sox2, Nanog, and others (Papp and Plath 2011). The *INK4/ARF* locus has been shown to impede the transition of somatic cells to the pluripotent state (Li et al. 2009). This locus is generally silenced in stem cells and iPSCs but due to highly mitogenic culture conditions during reprogramming, this locus is activated (Sharpless 2005). The activation of this locus drastically reduces reprogramming efficiency. When this locus is silenced during reprogramming, the efficiency increases significantly (Li et al. 2009). It has been demonstrated that silencing INK4a or ARF alone improves reprogramming efficiency, but double knockdown improves efficiency even more (Li et al. 2009). Furthermore, mouse fibroblasts lacking the *INK4/ARF* locus reprogram better than wild-type cells (Li et al. 2009).

Regulation of *INK4/ARF* locus via chromatin modifiers

By Polycomb Complexes

Polycomb group proteins are methyltransferase enzyme complexes essential for development, tissue repair, differentiation, senescence, and other processes (Wang et al. 2015). PcGs form two different multi-subunit complexes: PRC1 and PRC2. PRC1 comprises several core subunits, including RING1A/B, PCGF2/4, CBX2/4/6/7/8, and PHC1/2/3. The PRC2 complex, on the other hand, is made up of core subunits that include Suz12, Ezh2, RbAp46/48, and Eed (Chittock et al. 2017). In addition to core subunits, other auxiliary subunits assist both these complexes (Kerppola 2009). Both complexes are critical in suppressing the *INK4/ARF* locus in proliferating cells. PRC2 accomplishes this by binding to the promoters of these genes and trimethylating them (Bracken et al. 2007). Ectopic expression of PRC1 and PRC2 subunits leads to repression of this locus and delays senescence (Jacobs et al. 1999; Gil et al. 2004; Dietrich et al. 2007). Conversely, downregulation of the PcG subunits results in the activation of this locus and premature senescence (Bracken et al. 2007; Dietrich et al. 2007). PRC2 recruitment on *INK4/ARF* promoters is aided by protein factors discovered in several studies (Shields et al. 2016; Martin et al. 2013). For example, Zfp277, a zinc finger protein, interacts with the PRC1 complex's Bmi1 subunit and aids PRC1 recruitment on *INK4/ARF* promoters (Negishi et al. 2010).

By JMJD3

JMJD3 is a histone demethylase that plays a critical role during development, infectious disease, cancer progression, senescence, etc. It removes methyl marks from Histone 3 trimethylated at lysine 27 (H3K27me3), thus acting as an antagonist to the PRC2 complex (Xiang et al. 2007; Zhang, Liu, et al. 2019). JMJD3 is a well-known positive regulator of the *INK4/ARF* locus and, during senescence, activates this locus by removing repressive methylation from lysine 27 of Histone 3 laid down by the PRC complex (Agger et al. 2009). Cellular signals like oncogene overexpression upregulate JMJD3, activating the *INK4/ARF* locus (Agger et al. 2009). Consequently, MEFs and IMR90 cells undergo *INK4/ARF*-mediated cell cycle arrest. By activating the *INK4/ARF* locus, JMJD3 acts as a negative regulator for reprogramming. JMJD3 silencing alone increases reprogramming efficiency by many folds (Li et al. 2009).

Interestingly, silencing JMJD3 with INK4a or ARF improves reprogramming efficiency even more (Zhao et al. 2013).

By KDM2B

KDM2B is another demethylase that has been demonstrated to regulate the *INK4/ARF* locus. KDM2B is an epigenetic modifier that preferentially demethylates trimethylated lysine 4 (H3K4me3) and dimethylated lysine 36 of histone H3 (H3K36me2) (Frescas et al. 2007). It binds to *INK4/ARF* genes and removes methylation from lysine 4 of Histone 3, resulting in less Pol II binding (Tzatsos et al. 2009). The decrease in Pol II causes the downregulation of *INK4/ARF* genes, which prevents cells from undergoing senescence. Furthermore, its downregulation leads to the upregulation of these genes, resulting in premature senescence. Thus, KDM2B prevents proliferating cells like MEFs from undergoing senescence by keeping *INK4/ARF* locus downregulated (He et al. 2008).

Regulation of *INK4/ARF* locus via transcription factors

Several transcription factors have been shown to regulate *INK4/ARF* locus transcription. Most of these factors have been shown to target these genes' promoters directly (Farooq and Notani 2022; Bouchard et al. 2007; Sreeramaneni et al. 2005). However, factors like TGFβ have been demonstrated to regulate this locus via distal regulatory elements upstream of these genes (Zheng et al. 2013). The transcriptional activators activate this locus during the onset of senescence; on the other hand, the repressive factors repress this locus in proliferating cells under normal conditions. Some of the factors that regulate this locus are listed in the schematic below (Figure 1.4).

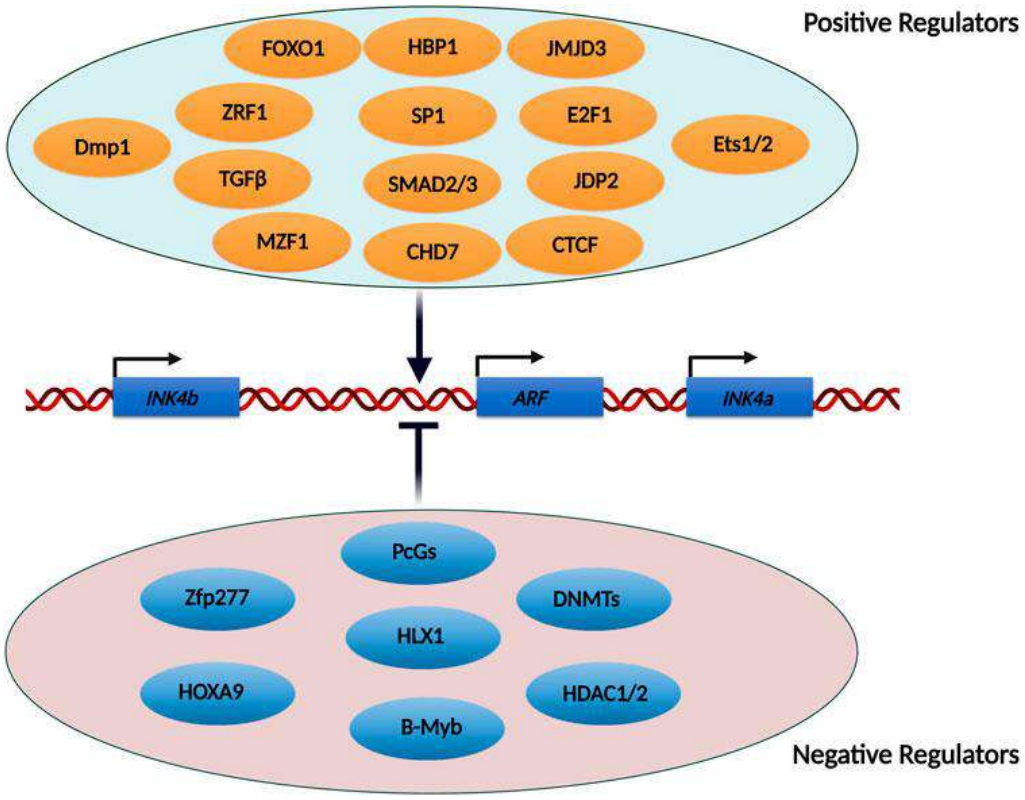


Figure 1.4: Transcription factors and chromatin modifiers involved in *INK4/ARF* locus regulation. *Various transcription factors and chromatin modifiers have been described that regulate this locus. Many factors have been shown to activate this locus, and some have been described to repress it under different scenarios. (Adapted from Farooq U and Notani D. Frontiers in Cell and Developmental Biology (2022)).*

Regulation of *INK4/ARF* locus via lncRNAs

lncRNAs are non-coding RNAs that are <200 nucleotides in length. These RNAs play an essential role in transcriptional regulation (Puvvula 2019). Several lncRNAs have been identified that target this locus both in cis and trans. The majority of those lncRNAs act as repressors of this locus, but some newly discovered lncRNAs have the potential to activate it (Figure 1.5). Repressive lncRNAs primarily function by binding to repressive proteins such as PcG complex subunits and recruiting them to the promoters of these genes (Montes et al. 2015; Sang et al. 2016). ANRIL, a well-studied lncRNA, for example, is transcribed from the same locus (Yap et al. 2010; Kotake et al. 2011). It inhibits *INK4/ARF* gene expression by binding to PcGs and recruiting them

to promoters (Yap et al. 2010; Kotake et al. 2011). Activating lncRNAs, on the other hand, work by removing the repressive complex from the promoters or recruiting transcriptional activators to these promoters (Ma et al. 2020; Lazorthes et al. 2015; Zhou, Sun, and Zhou 2020). Some recently discovered circular forms of ANRIL, for example, have been shown to remove the PRC complex from these promoters, thereby enhancing the activation of this locus (Muniz et al. 2021). Recently, the lncRNA ARHGAP27P1 was discovered to interact with JMJD3, assisting in its recruitment onto the promoters of these genes (Zhang, Xu, et al. 2019).

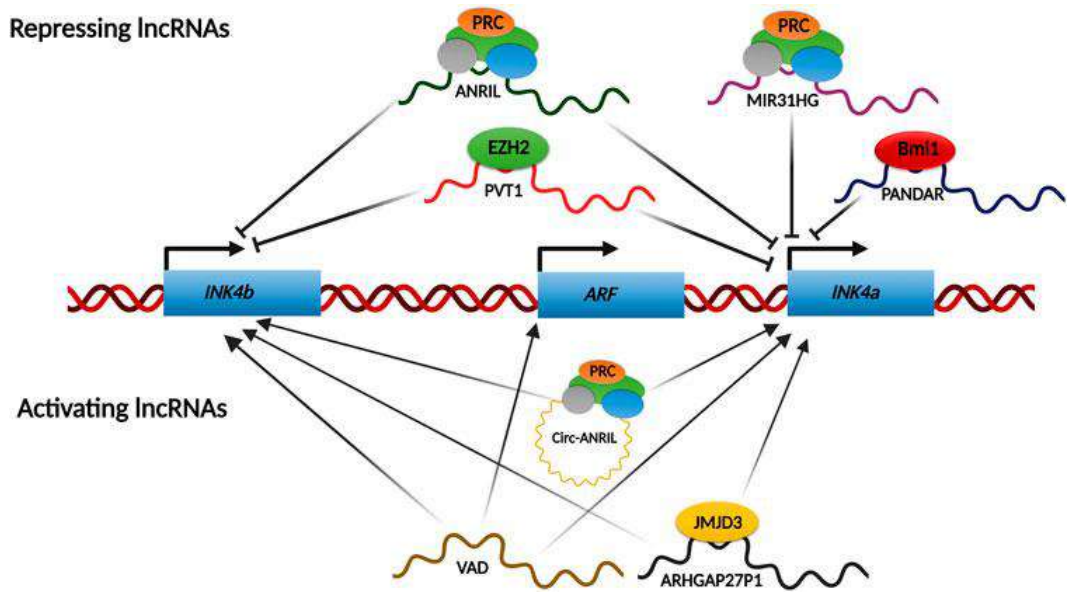


Figure 1.5: lncRNAs regulate *INK4/ARF* locus by cis and trans mechanisms. *Several lncRNAs have been identified that regulate the expression of INK4/ARF genes. Most of the lncRNAs repress this locus in proliferating cells. These repressive lncRNAs mainly function by helping the recruitment of PcGs onto the promoters of these genes. Activating lncRNAs either displace the PcG complexes from the promoter or aid in the recruitment of activators. (Adapted from Farooq U and Notani D. Frontiers in Cell and Developmental Biology (2022)).*

Regulation of *INK4/ARF* locus via distal regulatory elements

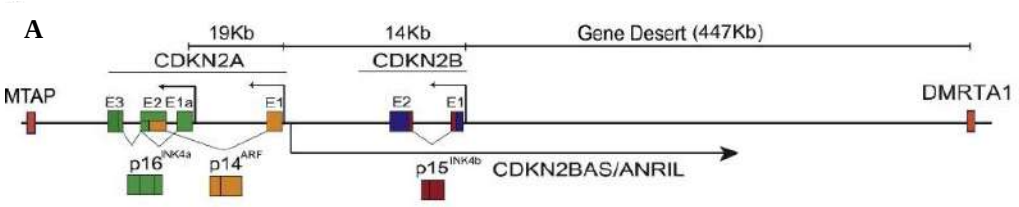
INK4/ARF locus has a long gene desert region upstream of the genes. SNPs in the gene desert have been linked to various diseases, most notably coronary artery disease, and type 2 diabetes (Harismendy et al. 2011). In mice, deleting the CAD interval, a gene

desert region containing SNPs linked to CAD, resulted in a significant decrease in the expression of these genes (Visel et al. 2010). A landmark study discovered several enhancers present in the gene desert to understand the functional relevance of the gene desert in regulating this locus. Moreover, SNP in one of the enhancers disrupts the STAT1 motif resulting in the downregulation of the genes upon IFNγ signaling (Harismendy et al. 2011). Subsequent studies have shown that TGFβ is required to activate upstream enhancers. TGFβ binds to one of the enhancers found in the CAD interval, promoting the expression of ARF during development (Zheng et al. 2013). Furthermore, TGFβ induces three H3K27ac peaks in the gene desert, and deletion of the entire interval results in lower INK4b and ARF expression (Liu et al. 2019). Moreover, individuals with SNP rs10757278 have significantly lower expression of all three genes in T cells in their peripheral blood (Liu et al. 2009). In addition to upstream regulatory elements, a cis-regulatory element near the ARF promoter inhibits INK4a gene expression by forming a loop with its promoter (Zhang, Hyle, et al. 2019). All the studies mentioned above link gene desert enhancers to the regulation of *INK4/ARF* locus transcription.

Chapter 2: Results (Part I)

CDKN2A/B transcription is regulated by an enhancer cluster in the adjacent gene desert

CDKN2A and *CDKN2B* genes encode three critical cell cycle regulators: $p14^{ARF}$; $p15^{INK4b}$ and $p16^{INK4a}$. The *CDKN2A* gene encodes two proteins: $p14^{ARF}$ and $p16^{INK4a}$ (transcribed from two distinct promoters), whereas the *CDKN2B* gene codes for $p15^{INK4b}$. Another gene, *CDKN2BAS,* is transcribed into a lncRNA called ANRIL (Farooq and Notani 2022; Aguilo, Zhou, and Walsh 2011). This gene is transcribed in the opposite direction to protein-coding genes *CDKN2A/2B*, thus the name *CDKN2BAS* (Figure 2.1A). A gene desert located upstream of these genes is one of the most reproducible GWAS hotspots. SNPs in this gene desert are linked to various aging-related diseases such as coronary artery disease, type 2 diabetes, and several cancers (Harismendy et al. 2011). Previous research has shown that the gene desert contains a dense enhancer cluster, but the regulatory potential of this enhancer cluster is not fully understood. Furthermore, the specific enhancers from this cluster that regulate these genes were yet to be uncovered. To explore the relevance of this enhancer cluster in the regulation of these genes, we first looked at the relative expression of these genes in tumors of various origins. Interestingly, tumors of cervical origin displayed the highest expression of both *CDKN2A* and *CDKN2B* genes (Figure 2.1B and 2.1C). To understand why gene expression is high in cervical cancer, we analyzed ATAC-seq data from patients with cervical cancer, which revealed that the gene promoters and some enhancers in the upstream gene desert region have highly open chromatin features, implying a positive relationship between open enhancer chromatin characteristics and gene expression (Figure 2.1D).



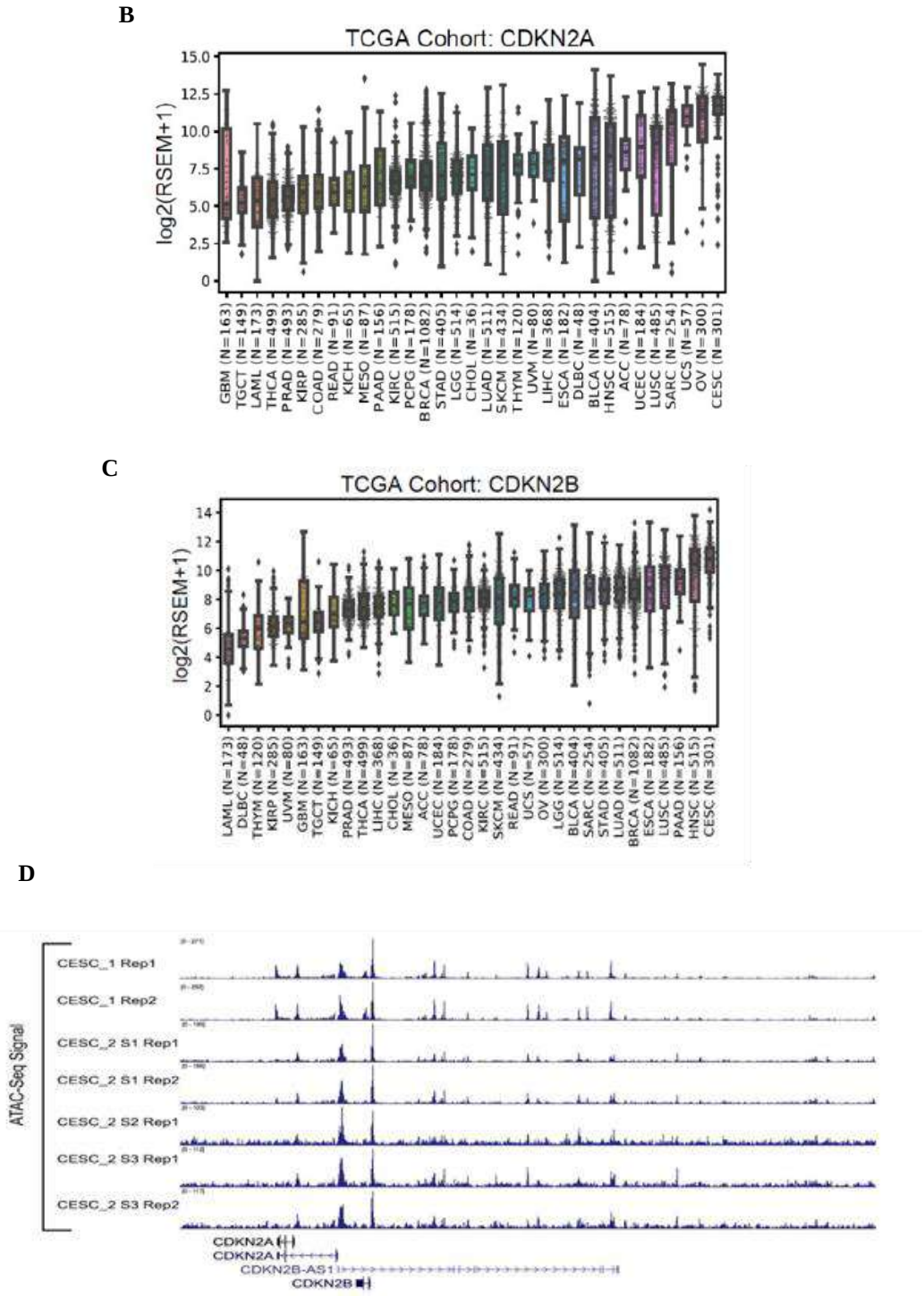


Figure 2.1: The presence of an upstream enhancer cluster correlates with *CDKN2A/B* expression in cervical tumors. ***A)*** *Schematic depicting the INK4/ARF locus. CDKN2A and CDKN2B are the genes that code for* $p16^{INK4a}$, $p14^{ARF}$, *and* $p15^{INK4b}$. ***B)*** *Expression profile of CDKN2A gene in tumors of various origins.* ***C)*** *Expression profile of CDKN2B gene in tumors of different origins. The expression of both CDKN2A and CDKN2B is highest in cervical tumors.* ***D)*** *A screenshot of the IGV*

browser showing ATAC-seq signal on INK4/ARF promoters and the neighboring gene-desert region in cervical cancer tumors from two TCGA patients.

To investigate the potential of the enhancer cluster in the regulation of this locus, we adopted the HeLa cell line as a model for our further study. The HeLa cell line is derived from a cervical carcinoma and expresses very high levels of $p16^{INK4a}$. Several recent studies have established that chromatin is organized into topological associating domains (TADs) that contain the genes and their regulatory elements (Tena and Santos-Pereira 2021). TADs help regulate the interaction of regulatory elements like enhancers with their target promoters (Tena and Santos-Pereira 2021). Using Hi-C data that was already available, we examined the TAD architecture at this locus in HeLa cells (Rao et al. 2014). We found that both *CDKN2A* and *CDKN2B* genes are contained well within one TAD. The *MTAP* gene was observed on the 5` border of the TAD (Figure 2.2A). Since it has already been shown that the gene desert located upstream of these genes contains a very dense enhancer cluster that is also contained well within the same TAD as *CDKN2A/2B* genes. Analysis of H3K27ac and H3K4me1 ChIP-seq datasets revealed the presence of 24 enhancers in the upstream cluster in HeLa cells. 15 of the 24 enhancers were active, displaying both H3K27ac and H3K4me1 marks, while the remaining 6 were only primed with the H3K4me1 mark (Figure 2.2A). The locus is heavily suppressed in stem cells, which is required for their reprogramming ability, and is expressed in some differentiated cell types (Li et al. 2009). To access the relationship of these enhancers with the expression of these genes, we examined RNA-seq datasets in stem cells and certain differentiated cell types spanning multiple lineages. Interestingly, the presence of active enhancers in the upstream cluster strongly correlates with the expression of these genes. These genes are not expressed in stem cells, and the enhancer cluster is devoid of any active enhancer (Figure 2.2B). However, the enhancer cluster is active in other differentiated cells which express these genes. These findings show a strong relationship between the expression of these genes and active enhancers in the upstream cluster in cervical cancer patients and the cell line of cervical tumor origin.

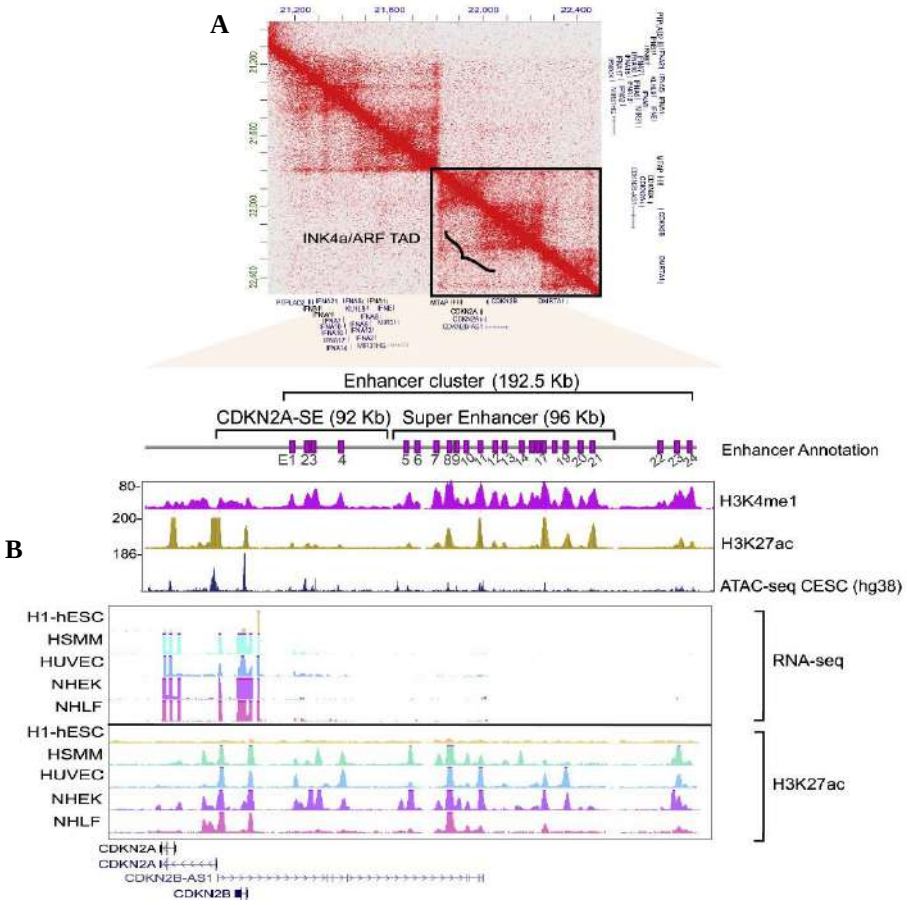


Figure 2.2: The presence of the enhancer cluster is highly correlated with the expression of *INK4/ARF* genes across cell types. ***A)*** *Hi-C heatmap of the TAD structure at the INK4/ARF locus in the HeLa cell line. The INK4/ARF promoters and the enhancer cluster is represented by violet boxes in the zoomed-in region (E1–E24). These are overlaid by H3K4me1 and H3K27ac (hg19) tracks in HeLa and ATAC-seq (hg38) from the cervical tumor.* ***B)*** *UCSC genome browser screenshot of the RNA-seq signal on the locus across different cell lines. H3K27ac ChIP-seq is overlaid with the RNA-seq.*

Only a subset of the enhancers in the upstream cluster interacts with *INK4/ARF* gene promoters

Next, to functionally dissect this enhancer cluster, we compared the number of H3K27ac-marked enhancers in *INK4/ARF* TAD to other TADs in the HeLa cell line. In terms of the number of H3K27ac-marked enhancers per TAD in HeLa cells, the *INK4/ARF* TAD was among the top 30 densest TADs out of 2,740 TADs (Figure 2.3A). Clusters of individual enhancers present within a distance of 12.5 kb and having very high enrichment of Med1 or H3K27ac are termed super-enhancers (Hnisz et al. 2013; Whyte et al. 2013). Thus, to test if the enhancer cluster in the gene desert is a super-enhancer, we used the ROSE algorithm to analyze H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) (Hnisz et al. 2013). Although the enhancer cluster in *INK4/ARF* TAD is one of 719 super-enhancers found in HeLa cells, it is not one of the highest-ranking SEs (Figure 2.3B). After applying a distance cutoff of 12 kb, we discovered that only 3 enhancers out of 24 enhancers were called as SE. However, when the distance cutoff was increased to 20 kb, 17 (E5-E21) of the 24 enhancers were classified as SE (Figure 2.3B). This suggests that the individual enhancers are well spread out in the gene desert. To gain an unbiased understanding of the regulatory potential of the enhancer cluster, we ignored the SE calling and instead focused on all 24 enhancers.

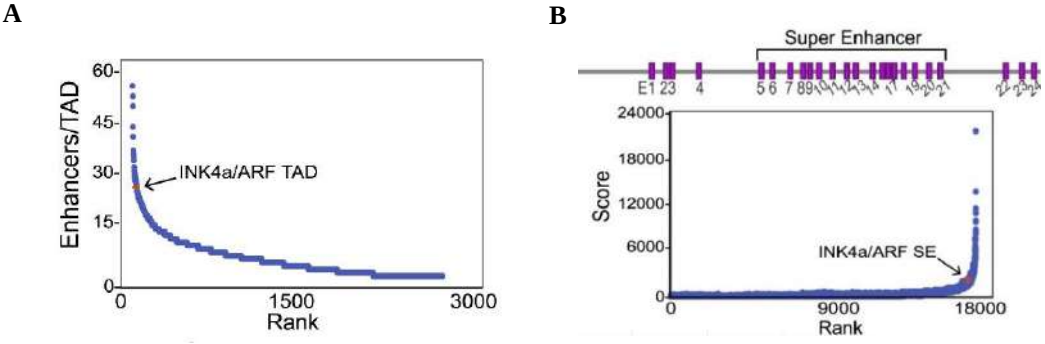
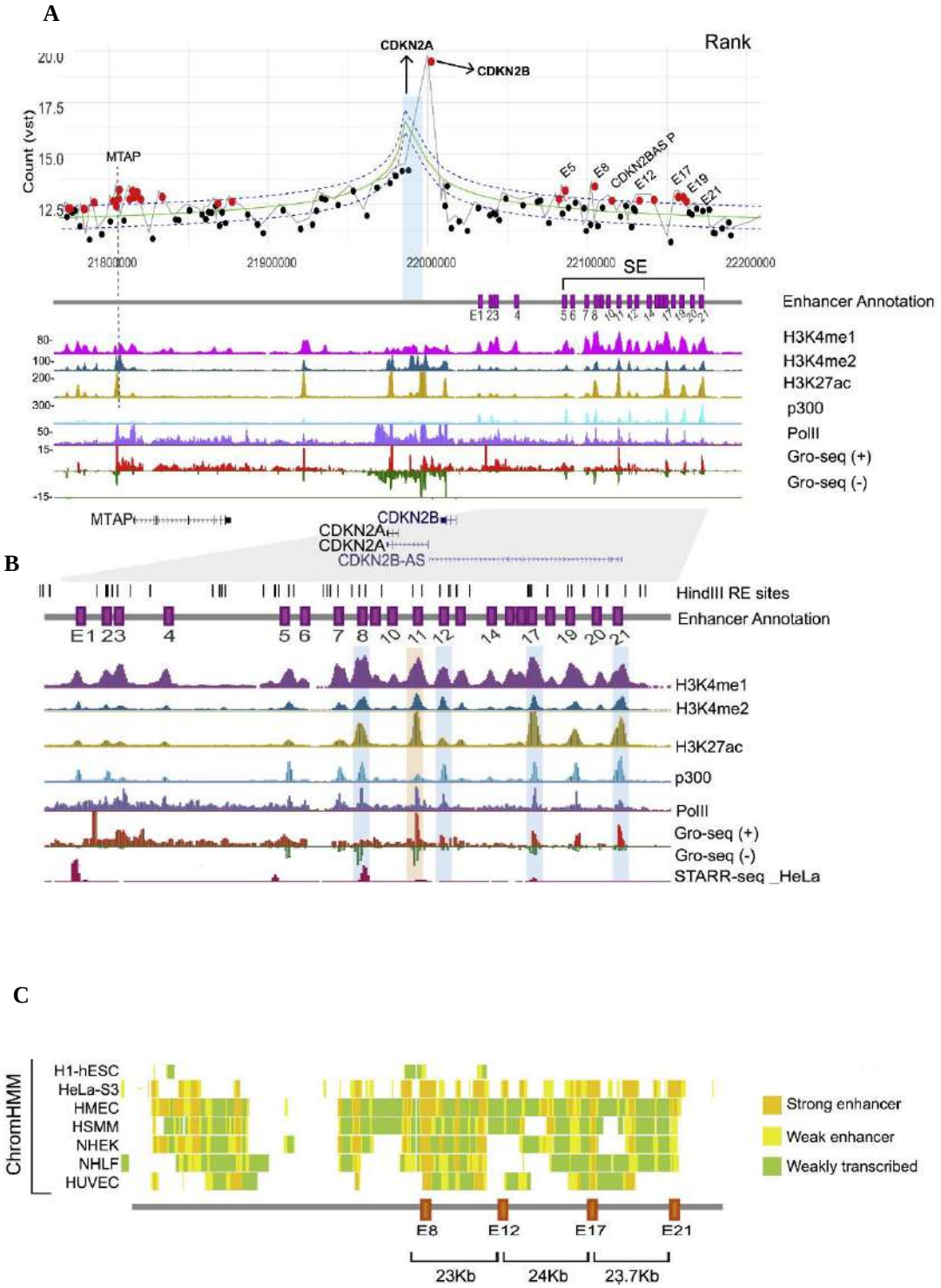


Figure 2.3: *INK4/ARF* TAD harbors a dense enhancer cluster. ***A)*** *TADs ranked according to the number of H3K27ac-marked enhancers.* ***B)*** *ROSE algorithm plot showing that the INK4/ARF enhancer cluster is a SE but is not among the top SEs.*

Typically, functional enhancers regulate target gene transcription by looping with the promoter. Thus, for an enhancer to regulate its target gene, it should be physically close to its promoter (Li et al. 2013). To identify enhancers that interact with promoters, we took advantage of the chromosome conformation capture (4C) technique by taking the *CDKN2A* promoter as a viewpoint. The 4C contact maps revealed that only a subset of enhancers exhibits high-frequency interactions with the promoters (Figure 2.4A). The interactions of the *CDKN2A* promoter were either directed to both ends of the TAD or the enhancers present in the gene desert. Out of 15 active enhancers, only 5 enhancers namely E5, E8, E12, E17, and E19 from the gene desert cluster showed high-frequency interactions with the promoter (Figure 2.4A). Furthermore, the sixth region that showed high-frequency interactions with the *CDKN2A* promoter is the 3` region of the *CDKN2BAS* gene, hence wasn't regarded as an enhancer. Interestingly, the interacting enhancers showed varying levels of H3K27ac, H3K4me1, p300, Pol II, and STARR-seq signal (Figure 2.4B). Pol II-enriched enhancers also showed bidirectional eRNA transcription, but their expression of eRNAs varies among these enhancers (Figure 2.4B).

To test if promoter-interacting enhancers were active in cell types other than HeLa, we used ChromHMM to integrate different datasets across different cell types. We found that E8 and E17 enhancers were consistently active enhancers in other cell types. These cell types included HSMM and HUVEC, which are relevant cell lines to coronary artery disease and atherosclerosis, respectively (Figure 2.4C). Since $p14^{ARF}$, $p15^{INK4b}$ and $p16^{INK4a}$ proteins are key regulators of senescence and stem cell regeneration. This locus was found to be highly upregulated during replicative senescence, and the associated gain of H3K27ac and eRNA expression was strongly induced on E12, E17, E19, and E21 enhancers (Figure 2.4D). All these findings imply that both interacting and non-interacting enhancers might have regulatory potential.


A
Rank
CDKN2A
CDKN2B
Count (vst)
MTAP
E5
E8
CDKN2BAS P
E12
E17
E19
E21
SE
Enhancer Annotation
H3K4me1
H3K4me2
H3K27ac
p300
PolII
Gro-seq (+)
Gro-seq (-)
CDKN2B
CDKN2A
CDKN2A
CDKN2B-AS
B
HindIII RE sites
Enhancer Annotation
H3K4me1
H3K4me2
H3K27ac
p300
PolII
Gro-seq (+)
Gro-seq (-)
STARR-seq _HeLa
C
ChromHMM
H1-hESC
HeLa-S3
HMEC
HSMM
NHEK
NHLF
HUVEC
Strong enhancer
Weak enhancer
Weakly transcribed
E8
E12
E17
E21
23Kb
24Kb
23.7Kb

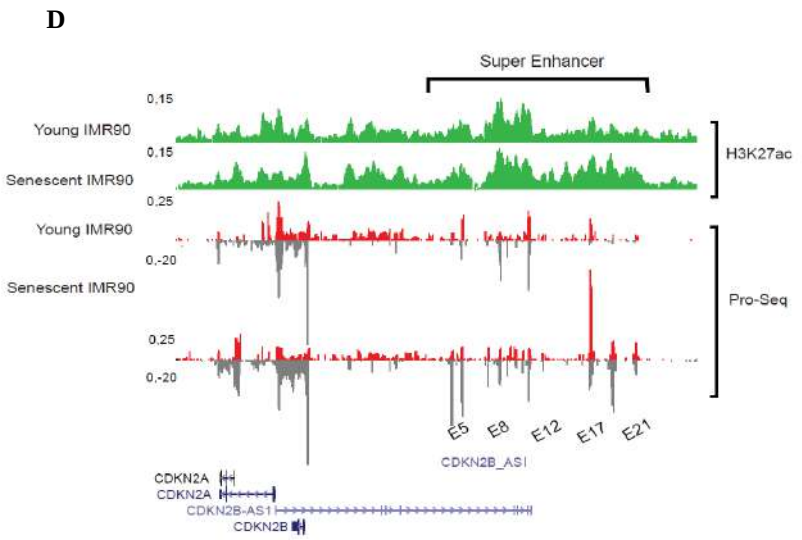


Figure 2.4: Only a subset of enhancers from the dense cluster interacts with the promoters. ***A)*** *4C plot showing the interactions from CDKN2A viewpoint; red dots represent the high-frequency significant interactions. The plot is overlaid with ChIP-seq and GRO-seq tracks from the HeLa cell line.* ***B)*** *The zoomed region from E1 to E21 is shown in the UCSC browser shot. The tracks are overlaid with the HeLa STARR-seq track.* ***C)*** *ChromHMM profiles of the enhancer cluster from E1–E21 across different cell types.* ***D)*** *H3K27ac and PRO-seq snapshots from the UCSC genome browser in young and replicative senescent IMR90 fibroblasts. The enhancers are annotated below the track.*

Promoter interacting enhancers regulate *INK4/ARF* gene transcription

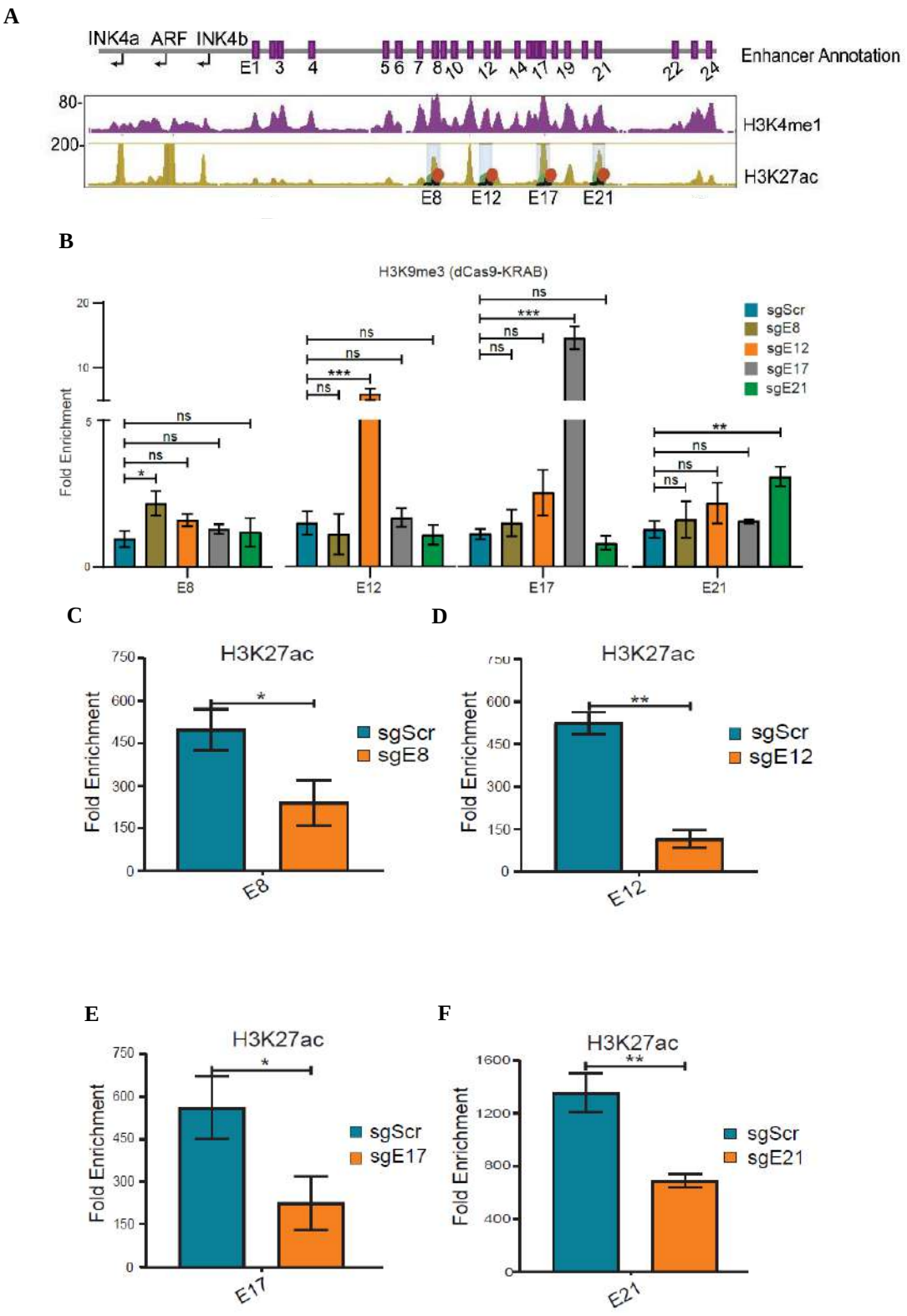
To test the regulatory potential of promoter-interacting enhancers, we chose three out of five, namely E8, E12, and E17 enhancers. These enhancers had varying levels of functional marks like H3K27ac and eRNA transcription. E8 had the highest enrichment of enhancer active marks, E17 had moderate enrichment, and E12 had the lowest enrichment of enhancer active marks. The active enhancer marks of the other two promoter-interacting enhancers E5 and E19 were comparable to E8 and E17, respectively. Furthermore, we chose E21, a promoter non-interacting enhancer with very high enrichment of H3K27ac and eRNA transcription. E21 transcription was also stimulated during the onset of replicative senescence and was found to be an active

enhancer in a variety of cell types. The distribution of functional features of these enhancers is shown in the table below:

Enhancer	H3K27ac	eRNA levels	STARR-Seq Signal	ChromHMM
E8	Medium	High	High	High
E12	Low	Low	Absent	Low
E17	High	Moderate	Moderate	High
E21	High	High	Absent	Moderate

Table 1: Histone marks and eRNA levels of the enhancers examined in the study.

Next, we designed single guide RNAs (sgRNAs) targeting the core of the four selected enhancers. The designed sgRNAs were utilized to block and repress these enhancers with the dCas9KRAB system (Gilbert et al. 2013) (Figure 2.5A). Since KRAB is a repressor that has been shown to associate with SETDB1 on its target sites, the addition of H3K9me3 by SETDB1 thus represses the target region (Schultz et al. 2002). Hence, H3K9me3 gain on targeted enhancers was used to confirm dCas9KRAB targeting. We observed an increase in H3K9me3 on enhancers targeted by the dCas9KRAB system. Interestingly, H3K9me3 gain was restricted to target enhancers and did not spread to other promoter-interacting enhancers (Figure 2.5B). Furthermore, the increase in H3K9me3 was accompanied by a decrease in H3K27ac and eRNA transcription on the targeted enhancers, indicating effective blocking by dCas9KRAB (Figure 2.5C – 2.5J).


A
INK4a ARF INK4b
E1 3 4 5 6 7 8 10 12 14 17 19 21 22 24
Enhancer Annotation
80-
200-
H3K4me1
H3K27ac
E8 E12 E17 E21
B
H3K9me3 (dCas9-KRAB)
Fold Enrichment
sgScr
sgE8
sgE12
sgE17
sgE21
E8 E12 E17 E21
C
H3K27ac
Fold Enrichment
sgScr
sgE8
E8
D
H3K27ac
Fold Enrichment
sgScr
sgE12
E12
E
H3K27ac
Fold Enrichment
sgScr
sgE17
E17
F
H3K27ac
Fold Enrichment
sgScr
sgE21
E21

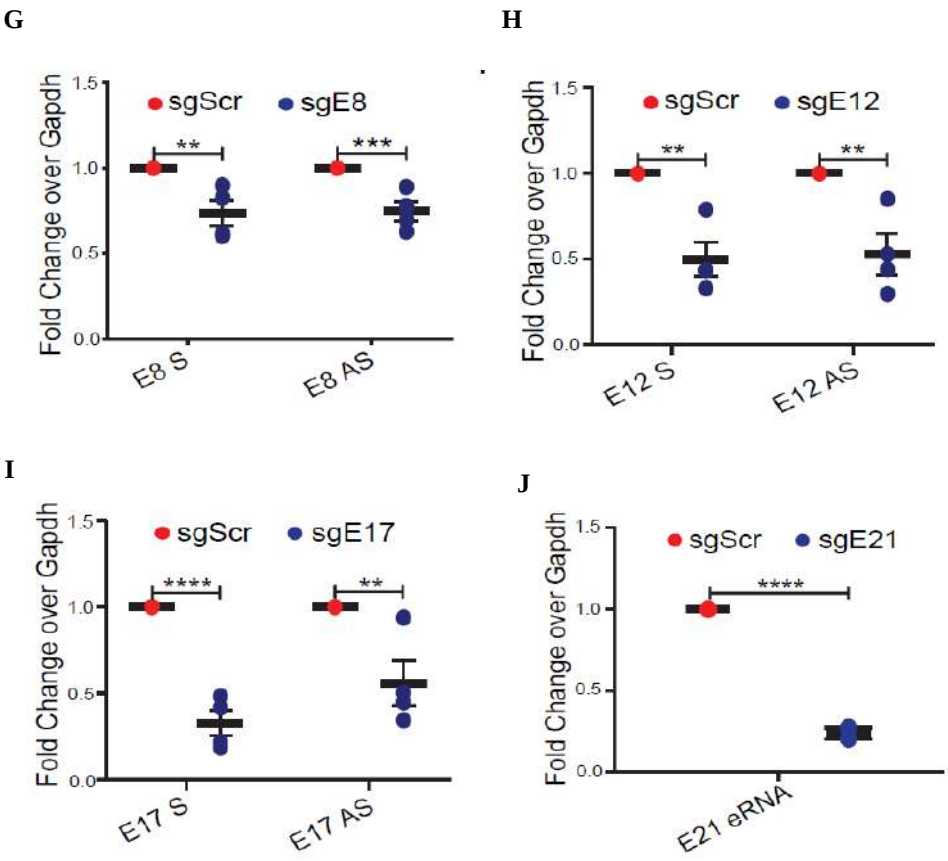
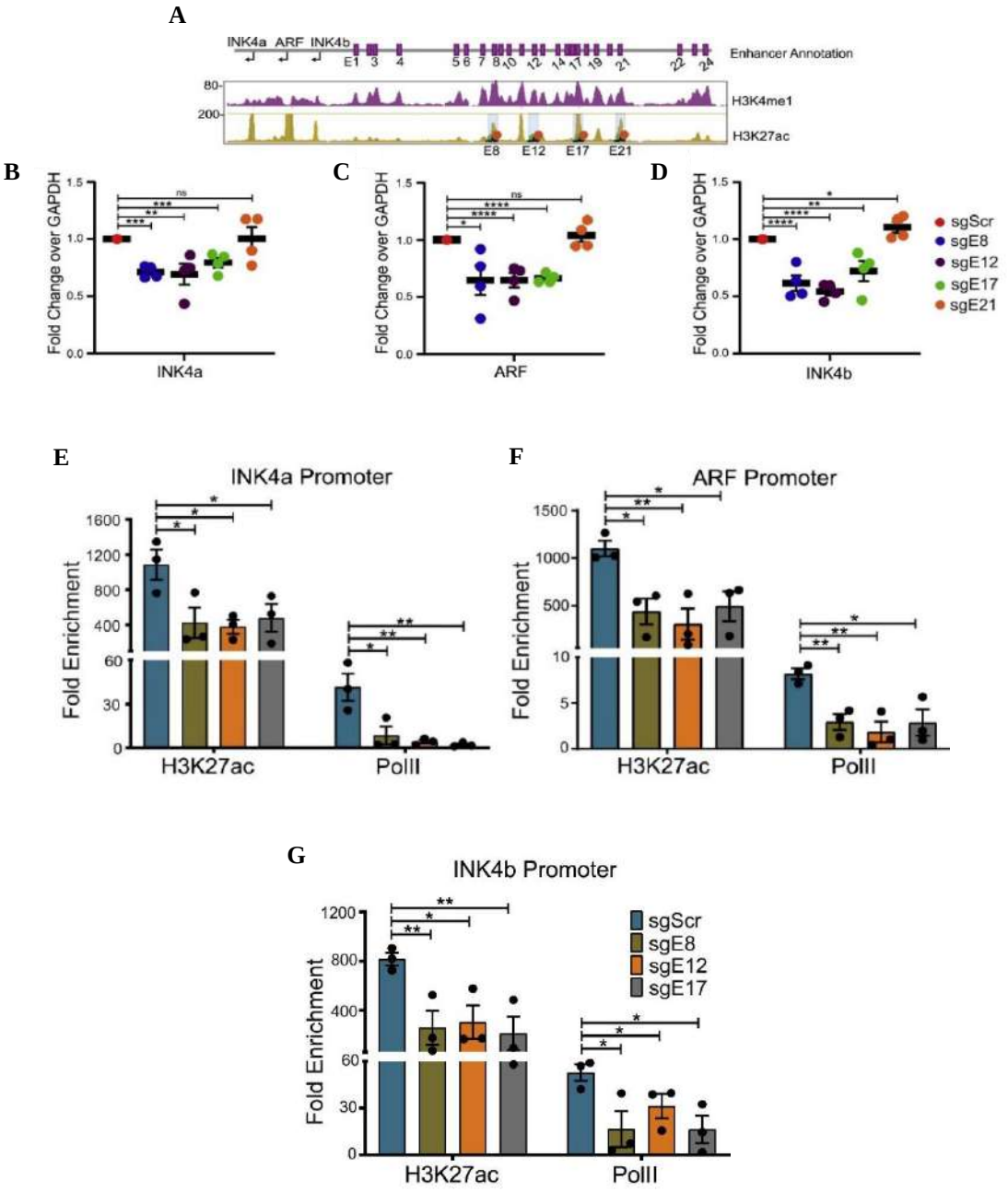


Figure 2.5: sgRNAs effectively repress the target enhancers. ***A)*** *Schematic depicting the enhancers that are targeted with the sgRNAs.* ***B)*** *H3K9me3 levels specifically increase on target enhancers upon dCas9KRAB targeting.* ***C-F)*** *H3K27ac enrichment goes down on the enhancers which are targeted for silencing* ***C)*** *E8 enhancer,* ***D)*** *E12 enhancer,* ***E)*** *E17 enhancer and* ***F)*** *E21 enhancer.* ***G-J)*** *Enhancer silencing mediated via the dCas9-KRAB system results in transcriptional loss of both sense and anti-sense eRNAs* ***G)*** *E8 enhancer,* ***H)*** *E12 enhancer* ***I)*** *E17 enhancer, and* ***J)*** *E21 enhancer.*

After confirming the blocking, we investigated the effect of enhancer blocking on *INK4/ARF* gene transcription. When promoter-interacting enhancers were individually blocked, levels of all three INK4a, ARF, and INK4b transcripts decreased significantly. The non-interacting enhancer E21, on the other hand, did not affect *INK4/ARF* gene transcription (Figures 2.6B – 2.6D). These findings imply that only promoter-interacting enhancers, not non-interacting enhancers, can regulate *INK4/ARF* genes. We further examined the levels of H3K27ac and Pol II at these promoters to see if the

changes in gene expression caused by blocking the promoter-interacting enhancers were accompanied by changes at the promoters of these genes. Expectedly, H3K27ac and Pol II levels were significantly lower when promoter-interacting enhancers were blocked versus non-interacting enhancer E21 (Figures 2.6E – 2.6G and Figures 2.6H – 2.6J). Only the INK4b promoter showed some reduction in the Pol II levels upon blocking E21. These findings are consistent with the downregulation of the target genes.



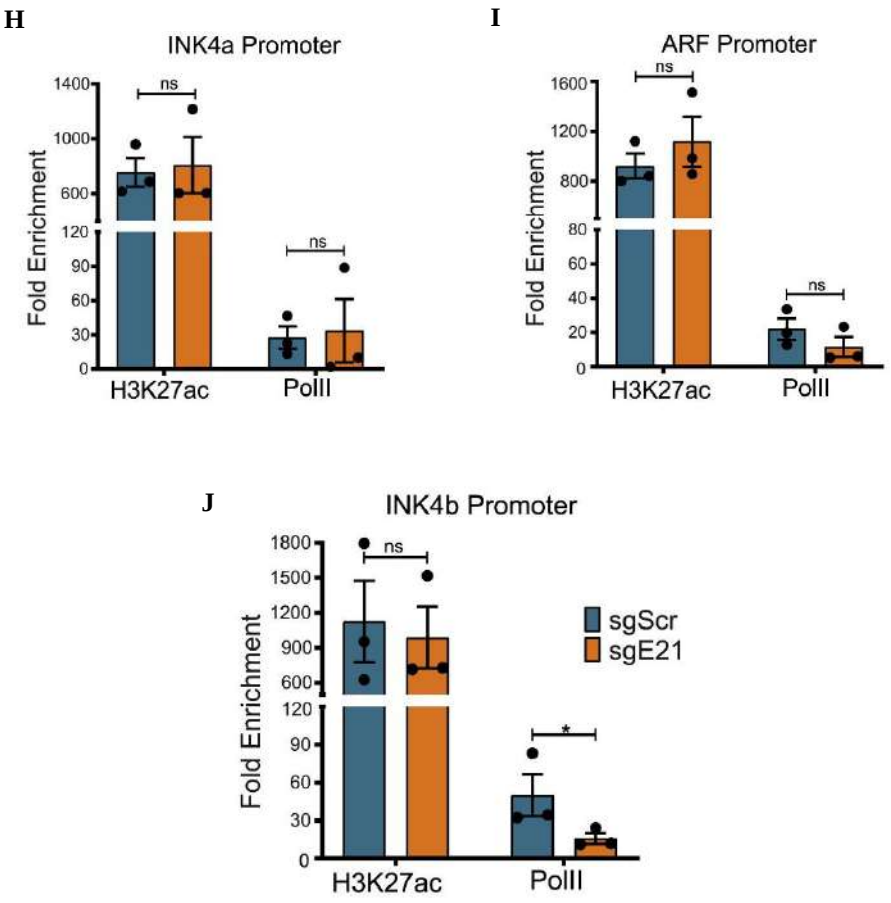
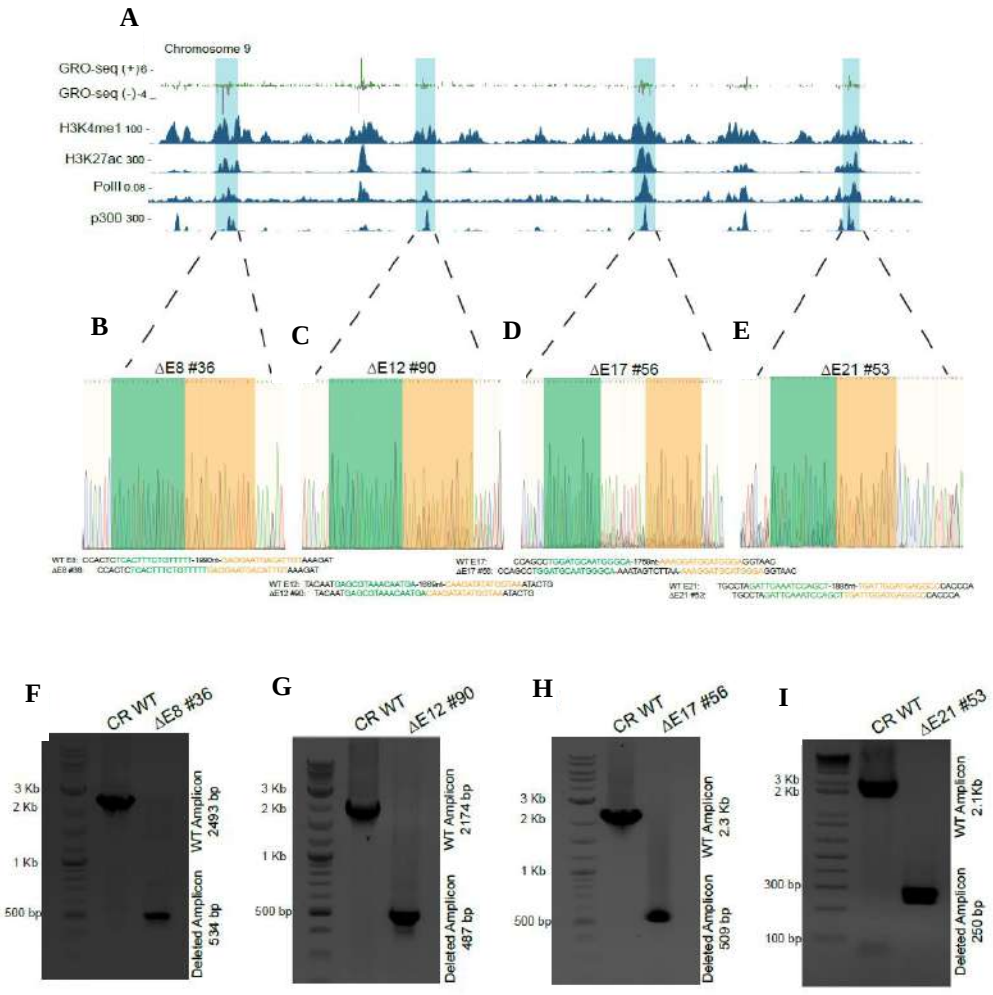


Figure 2.6: Blocking of promoter-interacting enhancers results in the downregulation of *INK4/ARF* genes. ***A)*** *Schematic depicting the enhancers that are targeted with the sgRNAs.* ***B–D)*** *mRNA levels of* ***A)*** *INK4a,* ***B)*** *ARF, and* ***C)*** *INK4b genes upon CRISPRi on E8, E12, E17, and E21 (n = 4). Plots are overlaid with a schematic of the INK4/ARF locus marked with enhancer annotation, H3K4me1, and H3K27ac tracks, and highlighted enhancers were blocked (orange and green solid circles).* ***E–G)*** *H3K27ac and Pol II enrichment on promoters* ***E)*** *INK4a,* ***F)*** *ARF, and* ***G)*** *INK4b upon CRISPRi on E8, E12, and E17 individually or* ***H–J)*** *E21 (n = 3).*

To further investigate the regulatory potential and to understand their intrinsic properties, we created individual knockout lines of these enhancers using the CRISPR/Cas9 system. We generated four distinct lines carrying the deletion of individual enhancers, namely, ΔE8#36, ΔE12#90, ΔE17#56, and ΔE21#53. The knockout lines were confirmed by a PCR-based assay followed by sanger sequencing (Figure 2.7A – 2.7I). Knockout of promoter-interacting enhancers resulted in a strong downregulation of all three genes, whereas knockout of promoter non-interacting enhancer resulted in only a subtle downregulation of INK4b (Figure 2.7J – 2.7L). This suggests that promoter-interacting enhancers are crucial for the transcription of these genes. Notably, the deletion of each interacting enhancer affected promoters that were equal to the sum of all enhancers. Furthermore, heterozygous deletion lines namely ΔE8#58, ΔE12#47, and ΔE17#06 showed similar effects on gene expression (Figure 2.7M-2.7P). Notably, heterozygous ΔE21#61 lines did not affect gene expression, indicating the importance of promoter-interacting enhancers on gene regulation. Furthermore, INK4b was almost completely silenced in the deleted lines, indicating that INK4b is completely dependent on individual interacting enhancers for expression. To test whether activating these three enhancers individually increases the expression of INK4a, we performed CRISPRa in HaCaT cells derived from human keratinocytes (epithelial origin like HeLa cells). CRISPRa on E8, E12, or E17 significantly increased INK4a expression. These findings confirm that the *INK4/ARF* locus is under the control of at least these promoter-interacting enhancers in one of the densest enhancer clusters and that even a single interacting enhancer perturbation results in a similar loss of *INK4/ARF* promoter activity even when the remaining 23 enhancers in the cluster remain intact. Furthermore, we believe that other promoter-interacting enhancers, E5 and E19, could be functional based on these data.


A
Chromosome 9
GRO-seq (+)6 -
GRO-seq (-)-4 -
H3K4me1 100 -
H3K27ac 300 -
PolII 0.08 -
p300 300 -
B
ΔE8 #36
C
ΔE12 #90
D
ΔE17 #56
E
ΔE21 #53
F
CR WT
ΔE8 #36
3 Kb
2 Kb
1 Kb
500 bp
WT Amplicon 2493 bp
Deleted Amplicon 534 bp
G
CR WT
ΔE12 #90
3 Kb
2 Kb
1 Kb
500 bp
WT Amplicon 2174 bp
Deleted Amplicon 487 bp
H
CR WT
ΔE17 #56
3 Kb
2 Kb
1 Kb
500 bp
WT Amplicon 2.3 Kb
Deleted Amplicon 509 bp
I
CR WT
ΔE21 #53
3 Kb
2 Kb
300 bp
100 bp
WT Amplicon 2.1Kb
Deleted Amplicon 250 bp

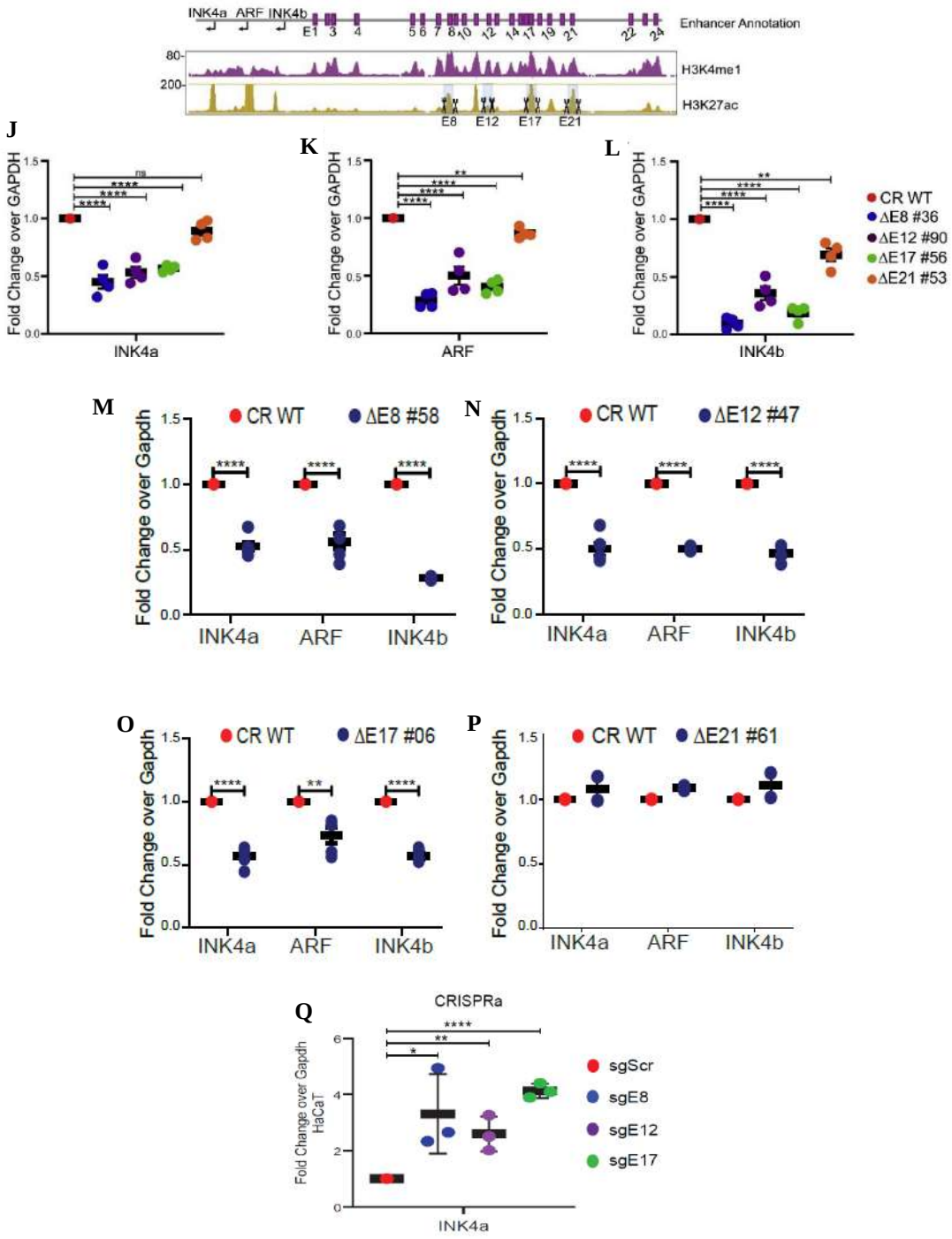


Figure 2.7: Knocking out promoter-interacting enhancers significantly downregulated *INK4/ARF* genes. ***A)*** *ChIP-seq tracks show H3K4me1, H3K27ac, Pol II, p300, and GRO-seq signals at enhancer clusters. Light blue highlights the deleted enhancer regions.* ***B-E)*** *The sanger-seq chromatograms of E8, E12, E17, and E21 enhancer deletion junctions are shown in* ***B)***, ***C)***, ***D)***, *and* ***E)***, *respectively.* ***F-I)*** *Agarose gel images show the confirmation of the deletion of enhancers by a PCR-based assay.* ***J-L)*** *mRNA levels of* ***J)*** *INK4a,* ***K)*** *ARF, and* ***L)*** *INK4b genes after E8, E12, E17, and*

E21 enhancer deletion (n = 4). Plots are overlaid with histone tracks and highlighted enhancers were deleted. ***M-P)*** *Plots depict the levels of expression of the INK4a, ARF, and INK4b genes in the heterozygous deletion clones.* ***Q)*** *Enhancer activation mediated via the dCas9-vpr system shows the gain of transcription of the INK4a gene in the HaCaT cell line.*

An interdependent enhancer network operates within the enhancer cluster

The data above suggest that disrupting a single promoter-interacting enhancer has a significant effect on gene expression, even when the other enhancers in the cluster remain intact. This prompted us to investigate whether the disruption of a single promoter-interacting enhancer affects the activity of other intact enhancers, rendering them inactive. To address this, we first assessed H3K27ac levels on intact enhancers after blocking other individual promoter-interacting enhancers. Surprisingly, when any of the promoter-interacting enhancers is blocked, H3K27ac levels on intact enhancers decrease (Figure 2.8A). Blocking promoter non-interacting enhancers, on the other hand, did not affect intact promoter-interacting enhancers (Figure 2.8B). We observed similar effects on intact enhancers in enhancer knockout lines to enhancer blocking (Figure 2.8C and 2.8D). However, when compared to CRISPRi, the effects of enhancer knockout were far more substantial. For example, blocking or knocking out the E8 enhancer reduces H3K27ac levels in other intact enhancers such as E12 and E17. Similarly, blocking or knocking out the E12 and E17 enhancers results in H3K27ac reduction on other intact enhancers. However, inhibiting or deleting E21, the promoter non-interacting enhancer does not affect the intact promoter interacting enhancers.

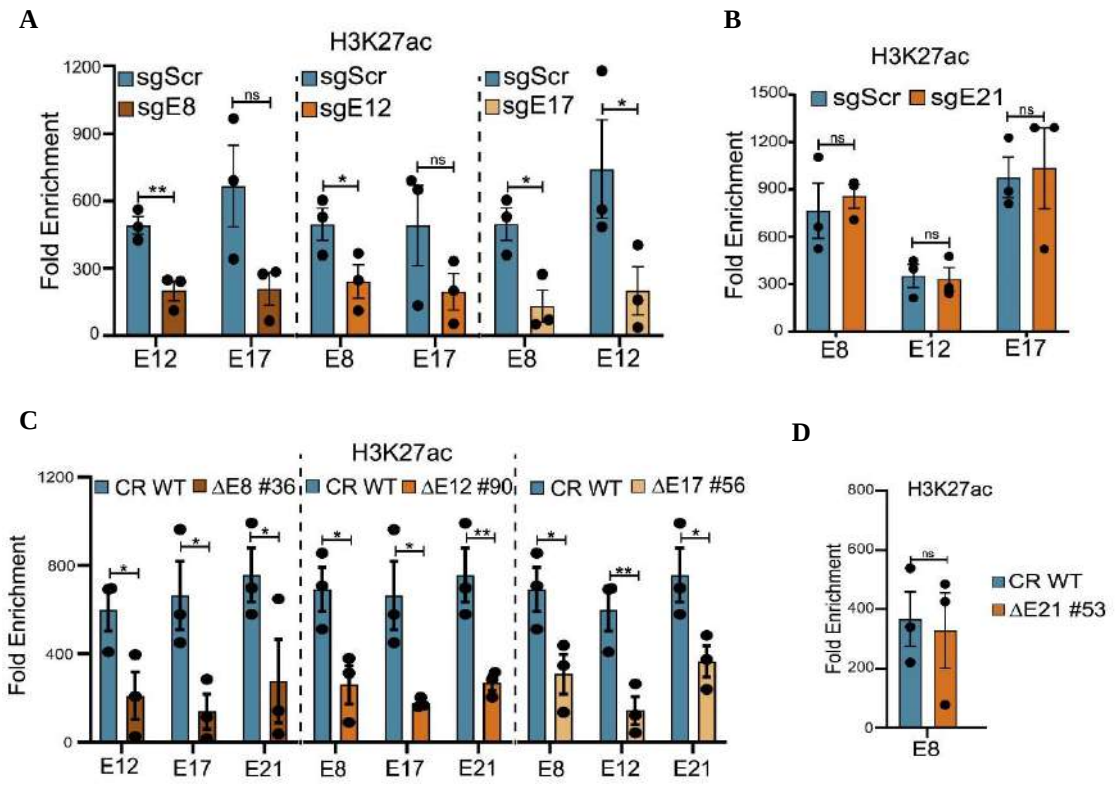


Figure 2.8: Intact enhancers lose H3K27ac upon perturbation of a single regulatory enhancer. ***A)*** *H3K27ac enrichment on intact enhancers in cells with CRISPRi on E8, E12, and E17.* ***B)*** *H3K27ac enrichment on intact enhancers in cells with CRISPRi on E21 (n = 3).* ***C)*** *H3K27ac enrichment on enhancers upon deletion of the individual enhancer (n = 3).* ***D)*** *H3K27ac enrichment on E8 upon E21 deletion (n = 3).*

To further understand the interdependence of the enhancers on each other, we examined the transcription of intact enhancers after perturbating the individual promoter-interacting enhancers. When any of the promoter-interacting enhancers is disrupted, both sense and anti-sense eRNA transcript levels on intact enhancers drop significantly, indicating that promoter-interacting enhancers regulate not only the promoters but also other enhancers in the cluster (Figure 2.9A – 2.9F). Blocking or deleting the E8 enhancer, for example, affects eRNA transcription at E12, E17, and E21 enhancers. Similarly, inhibiting or deleting the E12 and E17 enhancers reduces eRNA levels of other intact enhancers (Figure 2.9A – 2.9F). However, enhancer knockout has a greater impact on eRNA transcription than enhancer blocking. Notably, blocking the E21

enhancer had no effect on the intact promoter-interacting enhancers. However, the deletion of E21 resulted in E8 and E12 eRNA reduction (Figure 2.9D – 2.9F).

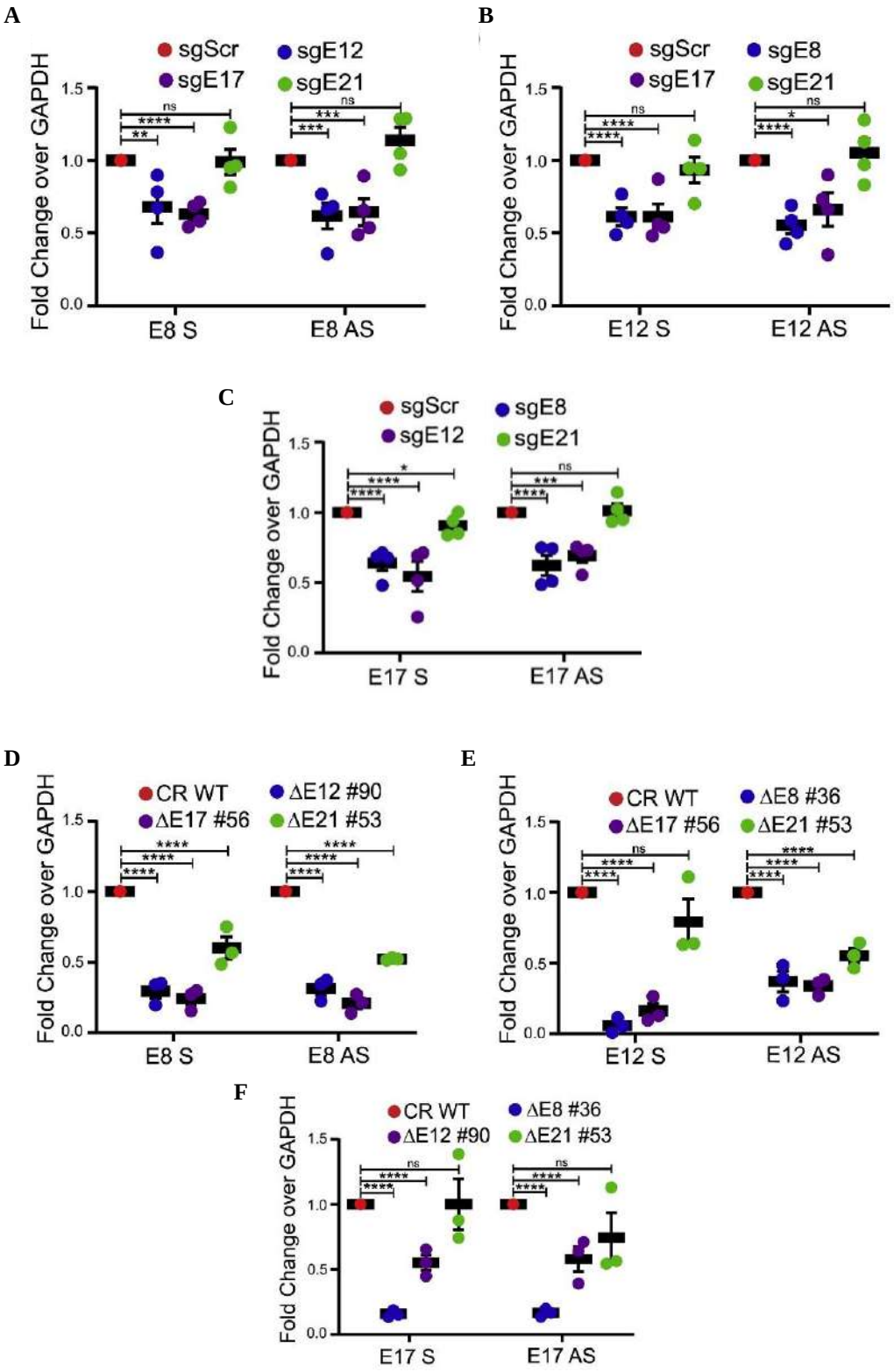


Figure 2.9: eRNA transcription of intact enhancers is affected upon perturbation of a single promoter-interacting enhancer. ***A–C)*** *Sense and antisense eRNA expression from* ***A)*** *E8,* ***B)*** *E12, and* ***C)*** *E17 upon CRISPRi.* ***D–F)*** *Sense and antisense eRNA expression from* ***D)*** *E8,* ***E)*** *E12, and* ***F)*** *E17 enhancers upon enhancer deletions (n = 3).*

The eRNAs were also found to be significantly downregulated in another set of clones that were heterozygous for the deletions (Figure 2.10A – 2.10D). Because perturbing a single promoter-interacting enhancer results in the loss of H3K27ac and eRNA transcription of other intact enhancers, we wondered if these effects were due to enhancer-enhancer interactions. To that end, we performed 4C in WT and E8 enhancer knockout lines taking the E12 enhancer as a viewpoint. We discovered that the E12 enhancer not only interacts with *INK4/ARF* promoters but also with other promoter-interacting enhancers (Figure 2.10E). This suggests that promoters and a few enhancers form a multi-enhancer-promoter network at *INK4/ARF* locus. However, E8 knockout affects the interaction of E12 enhancers with gene promoters and other promoter-interacting enhancers (Figure 2.10E). This implies that the loss of H3K27ac and eRNA transcripts from intact enhancers could be the result of a loss of enhancer-enhancer interactions.

Though we did not measure H3K27ac and eRNA levels on all intact enhancers in the cluster, the fact that none of the intact enhancers could rescue promoter activity in deletion lines strongly suggests that when even a single functional enhancer is deleted, all the intact enhancers lose their activity. These findings imply that functional enhancers within a super-enhancer form a single functional unit with the target promoters, regulating each other in the same way as they regulate their target promoters. Furthermore, even with comparable levels of H3K27ac, non-interacting enhancers have partial effects on transcriptional activity, but they are regulated by interacting enhancers.

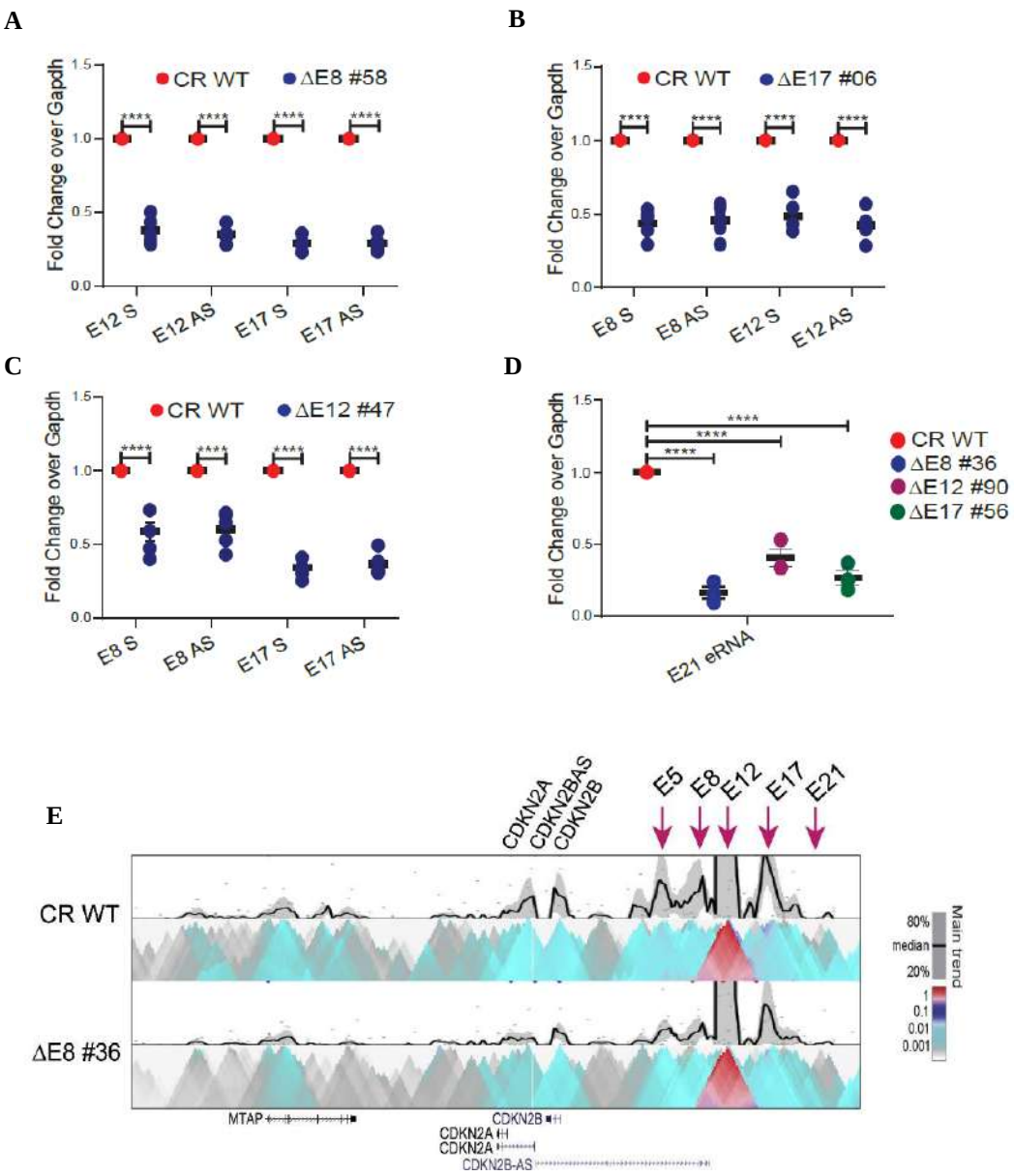
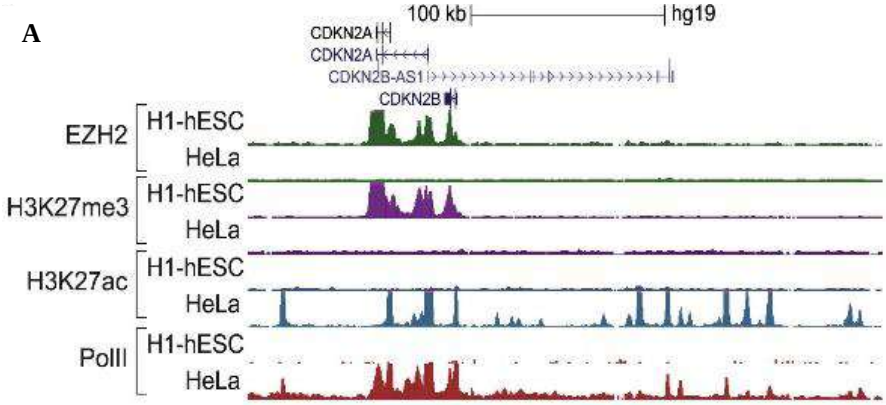


Figure 2.10: A multi-enhancer-promoter network operates at *INK4/ARF* locus. ***A)*** *Expression of E12 and E17 eRNA in cell lines with an E8 heterozygous deletion.* ***B)*** *E8 and E17 eRNA expression in cell lines with an E12 heterozygous deletion.* ***C)*** *E8 and E12 eRNA expression with an E17 heterozygous deletion.* ***D)*** *Expression of E21 eRNA in cell lines with homozygous deletion of promoter-interacting enhancers.* ***E)*** *A 4C heatmap in WT depicting the interactions of E12 with other enhancers and INK4/ARF promoters. When E8 is deleted, these interactions are significantly reduced (n = 2).*

Loss of a single enhancer results in EZH2 enrichment at the *INK4/ARF* promoters

INK4/ARF locus is a well-studied target of the PRC2 complex, which represses these genes in proliferating cells, stem cells, and several cancers (Gamell et al. 2017). To determine whether promoter silencing occurs as a result of PRC2 complex targeting when enhancers are knocked out. We first examined previously published ChIP-seq data sets from HeLa and hESCs. Since these genes are not expressed in hESCs, we found that the promoters of these genes lack active marks like H3K27ac and Pol II while being highly enriched with EZH2 and H3K27me3 (Figure 2.11A). In these cells, the upstream enhancer cluster lacked active marks as well. HeLa cells expressing this locus, on the other hand, do not show EZH2 binding and lack H3K27me3 at the promoters (Figure 2.11A). Moreover, promoters and enhancer clusters in HeLa cells are enriched with very high levels of H3K27ac and Pol II (Figure 2.11A). Thus, to investigate whether PRC2 plays a role in the downregulation of these genes following enhancer knockouts, we assessed the enrichment of EZH2 on these promoters in all three enhancer knockout lines. We found that EZH2 enrichment is high in enhancer knockout lines at these promoters (Figure 2.11B). Furthermore, the loading of EZH2 increases the H3K27me3 at these promoters (Figure 2.11C). However, we found no increase in EZH2 or H3K27me3 on the other intact enhancers after individual enhancer knockouts (Figure 2.11D – 2.11F), implying that enhancers are rendered inactive by a mechanism other than the PRC2 complex.



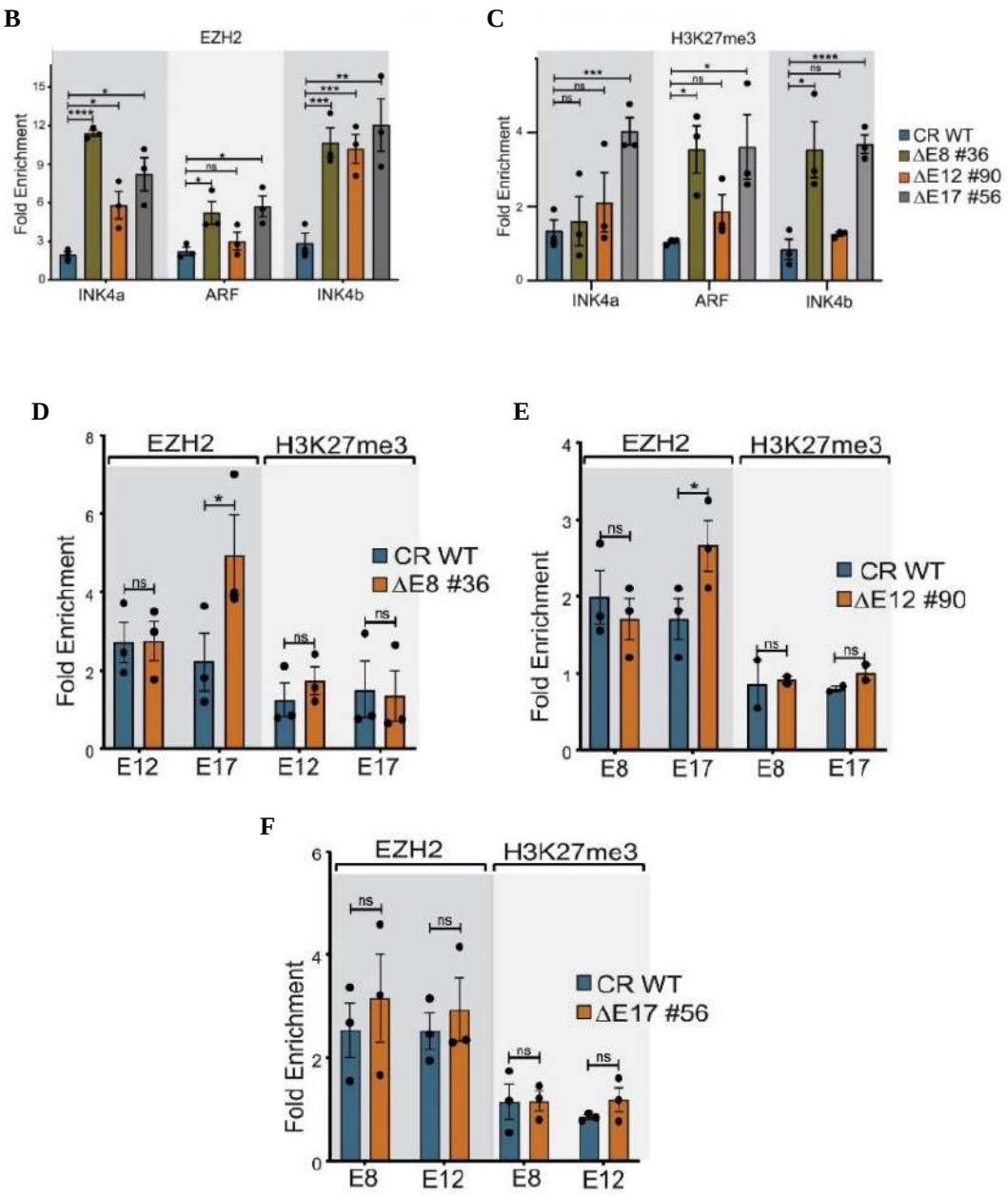
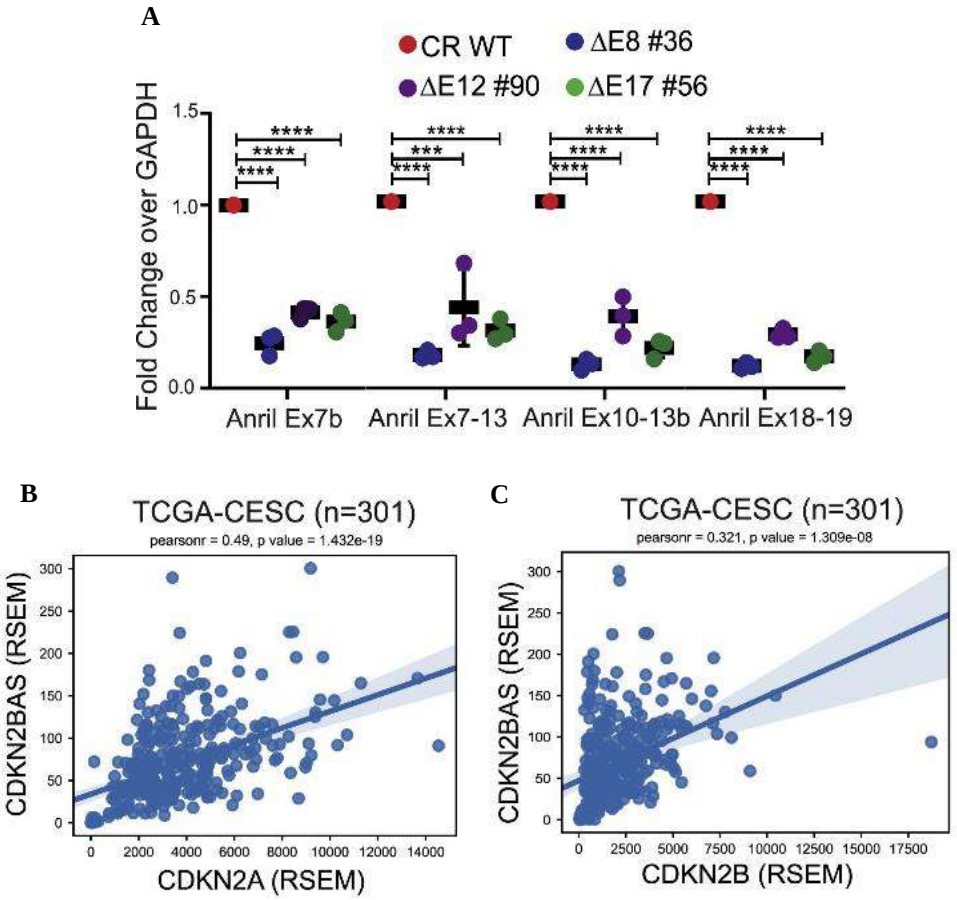


Figure 2.11: *INK4/ARF* genes are silenced by the PRC2 complex in absence of the promoter-interacting enhancers. ***A)*** *UCSC genome browser snapshot at the INK4/ARF locus shows ChIP-seq tracks for EZH2 and H3K27me3, H3K27ac, and Pol II in hESCs and HeLa.* ***B)*** *EZH2 enrichment on INK4a, ARF, and INK4b promoters upon deletion of E8, E12, and E17 enhancers (n = 3).* ***C)*** *H3K27me3 enrichment on INK4a, ARF, and INK4b promoters upon deletion of E8, E12, and E17 enhancers (n = 3).* ***D–F)*** *EZH2 and H3K27me3 enrichment on* ***D)*** *E12 and E17 upon E8 deletion (n = 3),* ***E)*** *E8 and E17 upon E12 deletion (EZH2 n = 3 and H3K27me3 n = 2), and* ***F)*** *E8 and E12 upon E17 deletion (n = 3).*

ANRIL, a lncRNA transcribed from the CDKN2BAS gene in the same locus, has been shown to bind EZH2 and aid in their recruitment on the promoters of *CDKN2A* and *CDKN2B* while being transcribed, resulting in the repression of these genes (Yap et al. 2010). Thus, to understand if EZH2 complexes are recruited on promoters upon enhancer knockout is dependent on ANRIL, we checked the expression of four different isoforms of ANRIL in these enhancer knockout lines. Surprisingly, ANRIL expression goes down upon enhancer knockout suggesting that ANRIL might not be playing any role in recruiting EZH2 onto these promoters in HeLa cells (Figure 2.12A). Furthermore, the ANRIL gene expression is also under the control of these enhancers. We also observed a positive correlation between *CDKN2A*, *CDKN2B*, and ANRIL expression in cervical cancer patients (Figures 2.12B and 2.12C). ANRIL expression was also found to be down in another set of heterozygous knockout lines (Figures 2.12D – 2.12F).



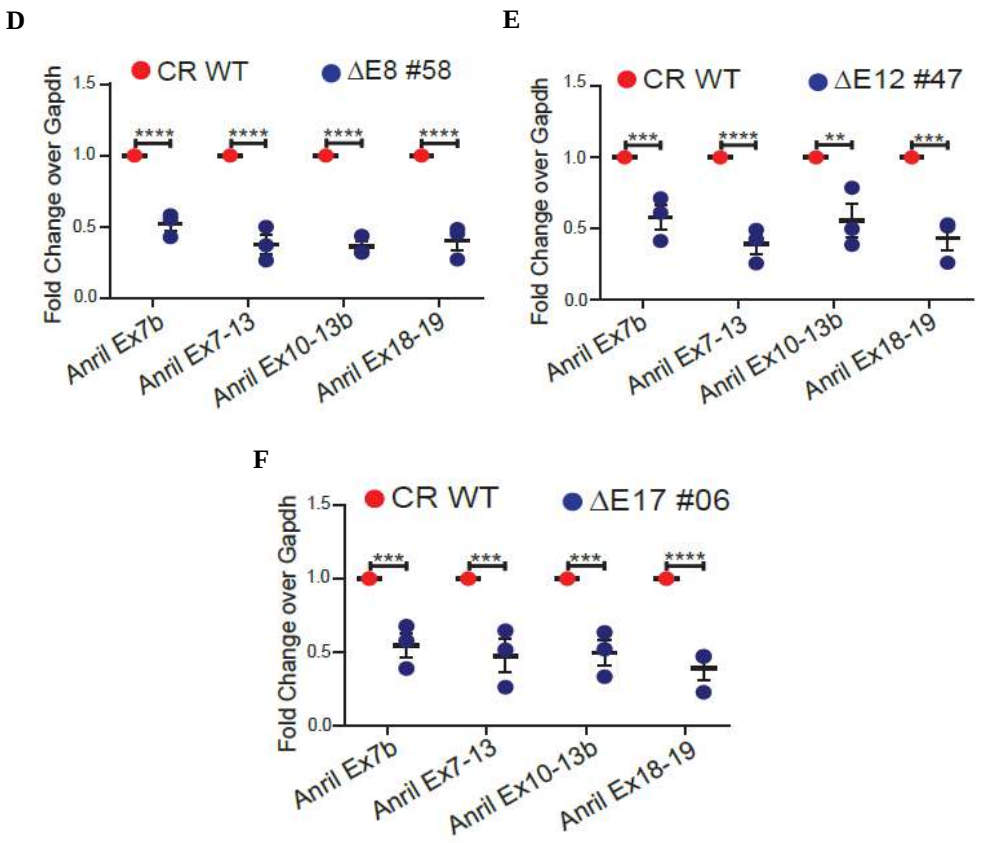
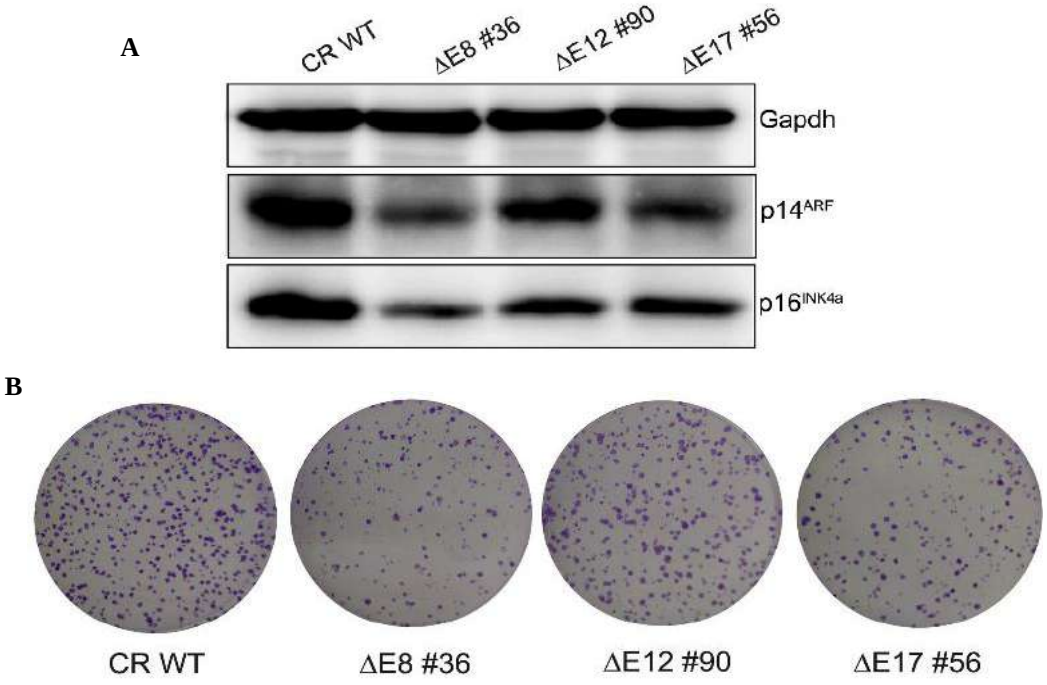


Figure 2.12: EZH2 is recruited on the promoters in an ANRIL-independent manner. ***A)*** *Expression of different isoforms of ANRIL upon deletion of E8, E12, and E17 enhancer (n = 3).* ***B and C)*** *Pearson correlation plots depicting the positive correlation between* ***B)*** *CDKN2A and CDKN2BAS, and* ***C)*** *CDKN2B and CDKN2BAS in cervical cancer tumors from the TCGA cohort; Pearson correlation value and p-value, respectively, are shown at top of each plot.* ***D-F)*** *Expression of different ANRIL isoforms in cells carrying heterozygous deletions of* ***D)*** *E8,* ***E)*** *E12, and* ***F)*** *E17*

Perturbation of promoter interacting enhancers affects the cancerous properties of HeLa cells

High levels of $p16^{INK4a}$ have been demonstrated to be essential for HPV-positive cancer cell proliferation and survival (McLaughlin-Drubin, Park, and Munger 2013; Pauck et al. 2014). The targeting of $p16^{INK4a}$ by small interfering RNA (siRNA) in HPV-positive cancer cells is associated with tumor growth inhibition (McLaughlin-Drubin, Park, and Munger 2013). Because $p16^{INK4a}$ is a target of enhancers, we looked at whether the malignant phenotype, such as cell proliferation and colony-formation potential, is

impacted in HeLa cells when these enhancers are absent. First, we looked at how enhancer knockouts affected $p14^{ARF}$ and $p16^{INK4a}$ protein levels. Both $p16^{INK4a}$ and $p14^{ARF}$ were reduced at the protein level (Figure 2.13A). Not surprisingly, the deletion of the enhancer resulted in a considerable reduction of colonies. The colonies in the enhancer knockout lines were fewer and smaller (Figures 2.13B and 2.13C). Likewise, the rate of proliferation of enhancer knockout lines is also affected (Figure 2.13D). Furthermore, enhancer knockout lines exhibit slowed cell migration (Figure 2.13E). Interestingly, the deletion of these enhancers resulted in increased senescence-associated β-galactosidase (SA-β-gal) staining (Figure 2.13F). These findings suggest that enhancer deletion alters the cancerous properties of HeLa and induces a senescence-like phenotype in an *INK4/ARF*-dependent manner.



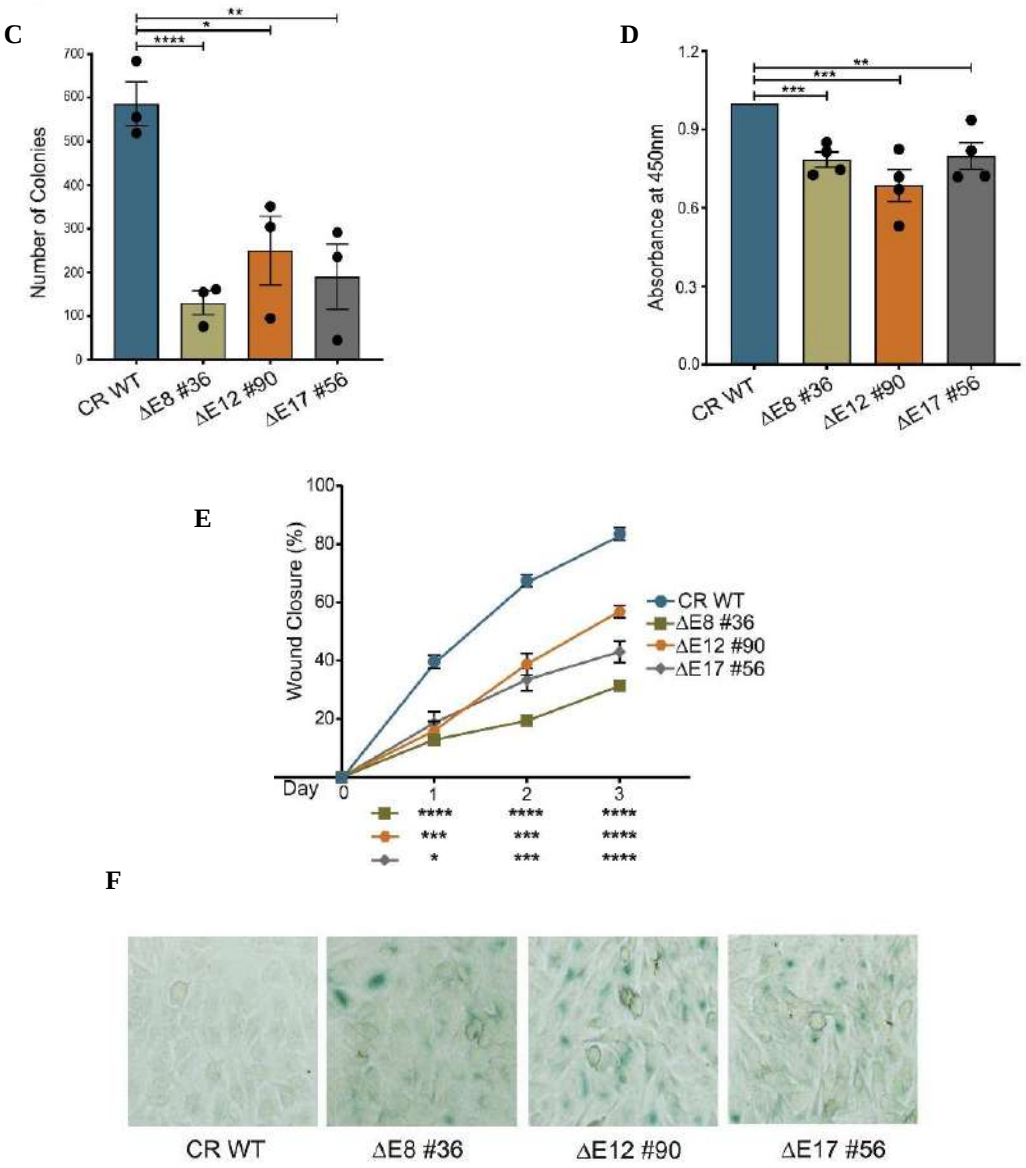


Figure 2.13: Perturbation in the enhancer network affects the cancerous properties of HeLa. ***A)*** *Immunoblots show the protein levels of* $p14^{ARF}$ *and* $p16^{INK4a}$ *upon deletion of E8, E12, and E17. Gapdh was used as a loading control (n = 2).* ***B)*** *Representative images of clonogenicity assay.* ***C)*** *Quantification of colonies using ImageJ upon deletion of E8, E12, and E17 enhancers (n = 3).* ***D)*** *Bromodeoxyuridine (BrdU) cell-proliferation assay in the absence of individual enhancers (n = 4).* ***E)*** *The percentage of wound closure in WT and upon deletion of E8, E12, and E17 (n = 2).* ***F)*** *Representative images of β-galactosidase activity in WT and enhancer deletions (n = 2).*

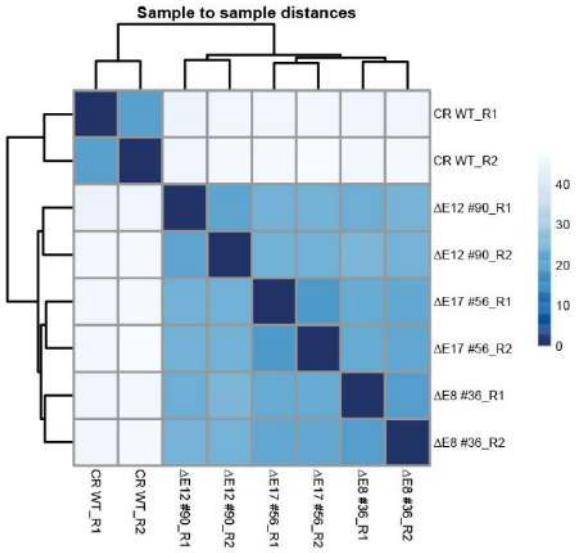
Genes dysregulated upon enhancer deletions corroborate with disease association of 9p21 locus

9p21 locus is the most reproducible GWAS locus linked to type 2 diabetes and coronary artery disease across races and ethnicities (Schunkert et al. 2011; Helgadottir et al. 2008; Nikpay et al. 2015). The *CDKN2BAS* gene and the upstream gene desert region contain the GWAS variations associated with these diseases (Figure 2.1A). However, the causative SNPs in this vast region are mainly unknown. We reasoned that the enhancer network is responsible for the transcriptional regulation of this locus, its loss of function should resemble the transcriptional patterns of the associated disorders of this locus. Even though HeLa is not the relevant cell type, E8 and E17 enhancers were seen to be active in cells of different lineages, including endothelium and smooth muscle cells (Figure 2.4C and Figure 2.4D), suggesting that the regulation by these enhancers is probably similar in these tissues. To achieve this, we first identified the disease association of the entire enhancer cluster (96 kb) extending from E5 to E21 using the GWAS catalogue tool with default parameters (Welter et al. 2014). We found that the SE variations are associated with endometrial cancer, type-2-diabetes, coronary artery disease, lifespan, and other diseases (Figure 2.14A). This implies that the SE alone contains the variants linked to the diseases with which the entire gene desert between the *CDKN2B* promoter and the TAD 3` boundary is associated (Figure 2.1A). We then investigated whether genes that are dysregulated when individual enhancers are deleted are linked to the disorders associated with this enhancer cluster. We performed Total RNA sequencing (RNA-seq) in two replicates in the wild-type (WT), E8, E12, and E17 deletion lines. The MA plots and positional clustering demonstrated a good correlation between the samples (Figures 2.14B – 2.14D).

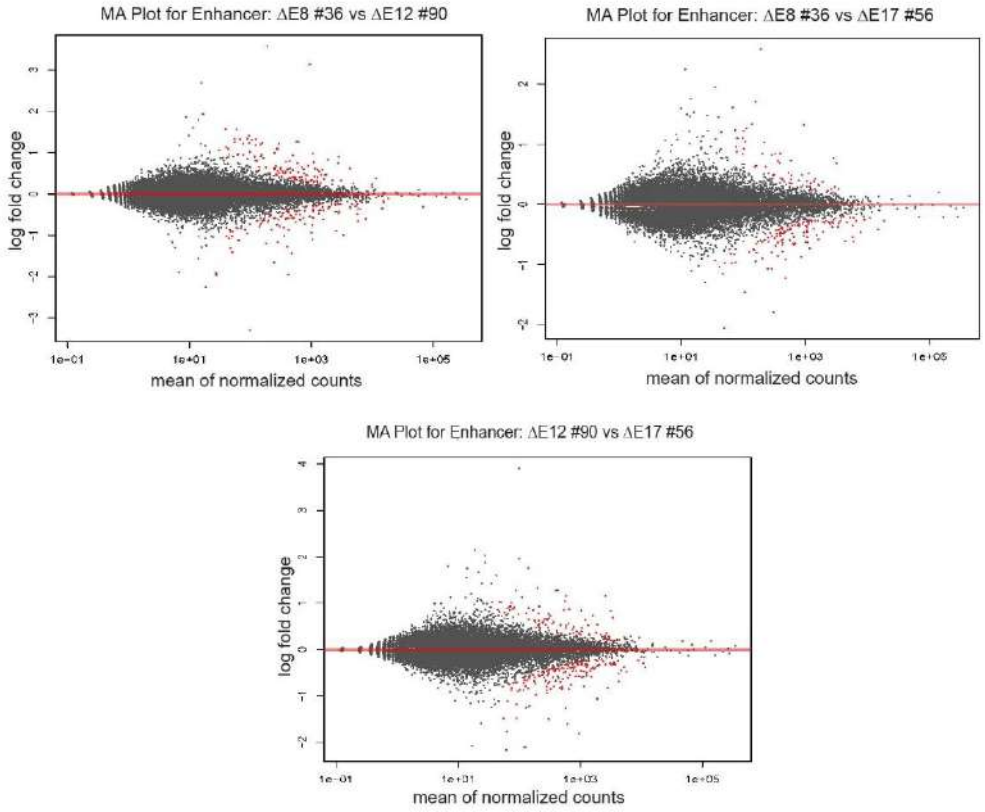
A

Reported trait(s)	Association count with Super Enhancer
Type 2 diabetes	31
Coronary artery disease (myocardial infarction, angina or ischemic heart disease)	21
Parental longevity, Parental lifespan	9
Endometriosis, Deep ovarian and/or rectovaginal disease with dense adhesions	7
Myocardial infarction, Myocardial infarction (early onset)	5
Intracranial aneurysm	5
Fasting blood glucose, Fasting plasma glucose	3
Pulse pressure	3
White blood cell count	3
Glycated hemoglobin levels, Hemoglobin A1c levels	3
Coronary artery calcification	2
Abdominal aortic aneurysm	2

B


Sample to sample distances
CR WT_R1
CR WT_R2
ΔE12 #90_R1
ΔE12 #90_R2
ΔE17 #56_R1
ΔE17 #56_R2
ΔE8 #36_R1
ΔE8 #36_R2
40
30
20
10
0

C


MA Plot for Enhancer: ΔE8 #36 vs ΔE12 #90
log fold change
mean of normalized counts
MA Plot for Enhancer: ΔE8 #36 vs ΔE17 #56
log fold change
mean of normalized counts
MA Plot for Enhancer: ΔE12 #90 vs ΔE17 #56
log fold change
mean of normalized counts

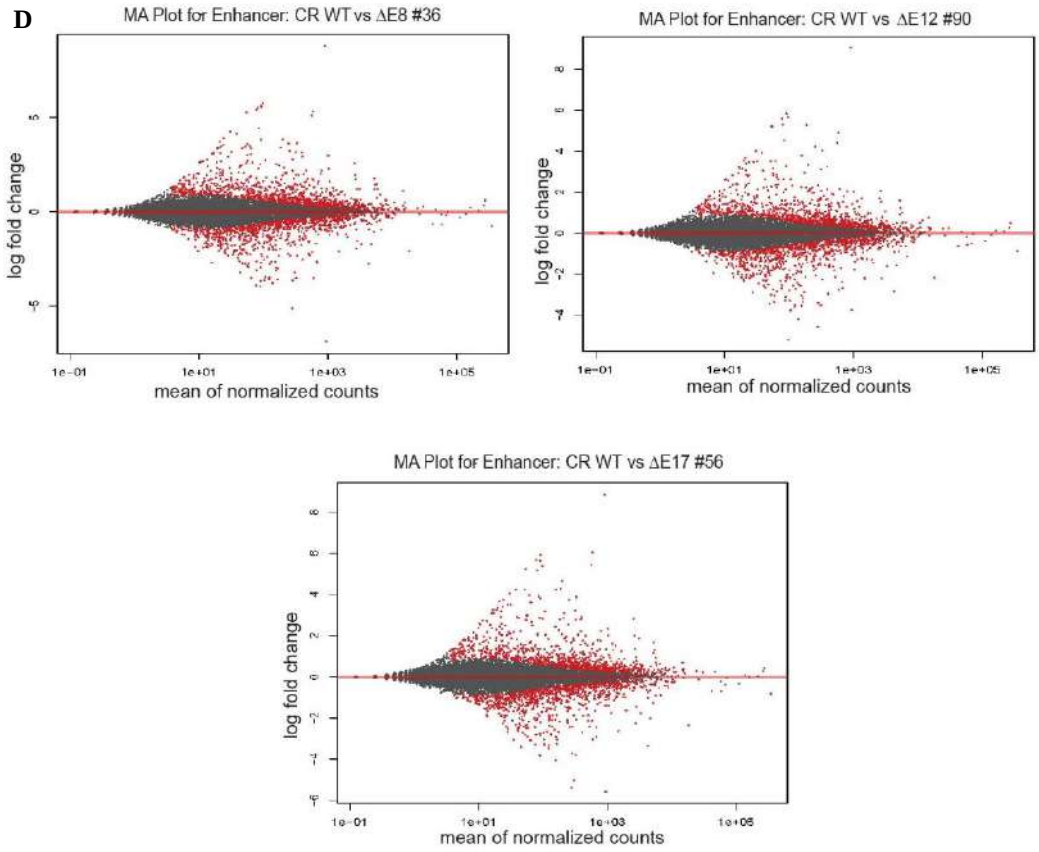


Figure 2.14: Alterations in genome-wide transcription are similar upon any enhancer deletion. ***A)*** *Chart from the GWAS catalog tool shows the disease terms of GWAS variants and their counts present within the enhancer region (E8–E21).* ***B)*** *Positional matrix positions together the replicates of RNA-seq upon an enhancer deletion or WT cells and* ***C)*** *MA plots comparing the RNA-seq data from E8 with E12, E8 with E17, E12 with E17.* ***D)*** *MA plots comparing RNA-seq from WT cells with E8, E12, or E17 deletions.*

CDKN2A intronic and exonic regions, as well as *CDKN2B*, *CDKN2BAS* (ANRIL), and *MTAP*, were all downregulated (Figures 2.15A – 2.15D), consistent with our earlier findings. Another region on chr9 remained unaffected, indicating that these enhancers specifically affect the *INK4/ARF* genes (Figure 2.15E).

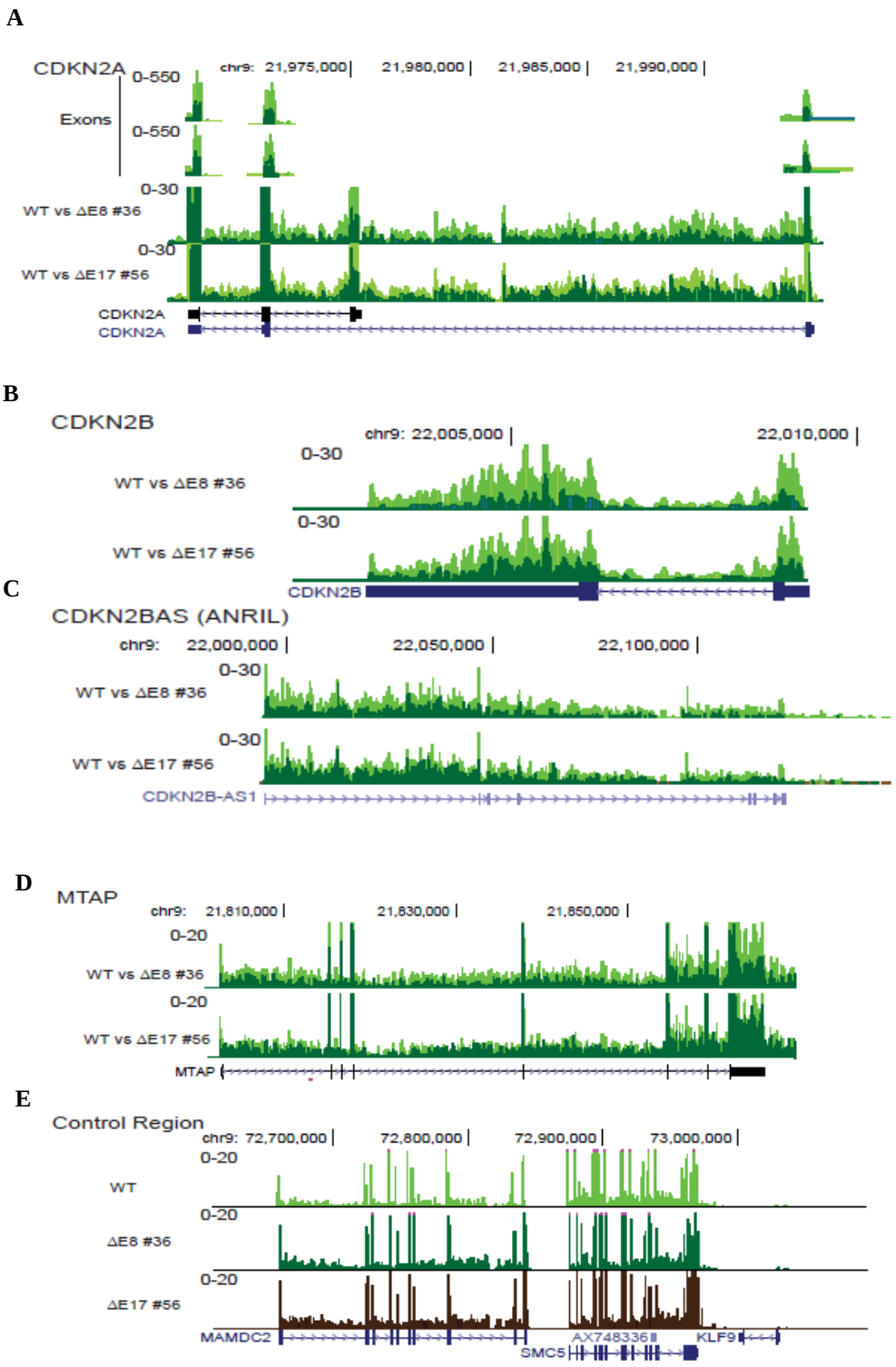
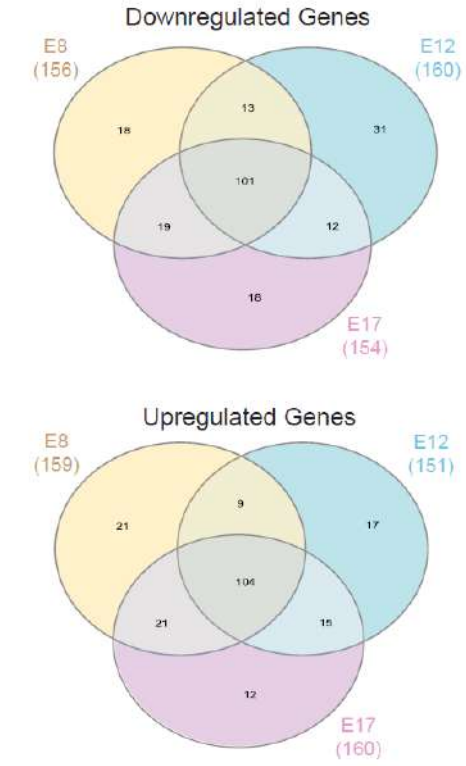
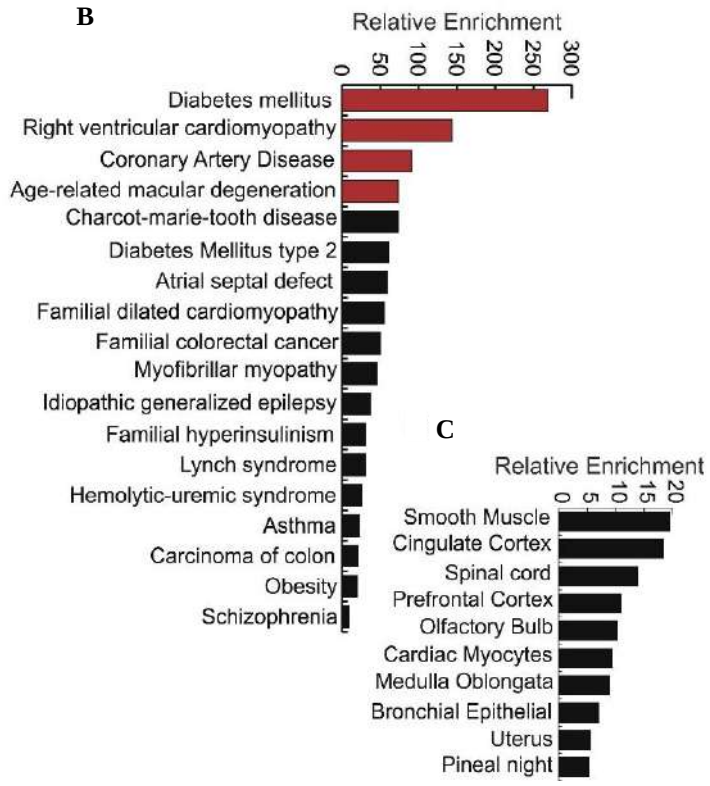


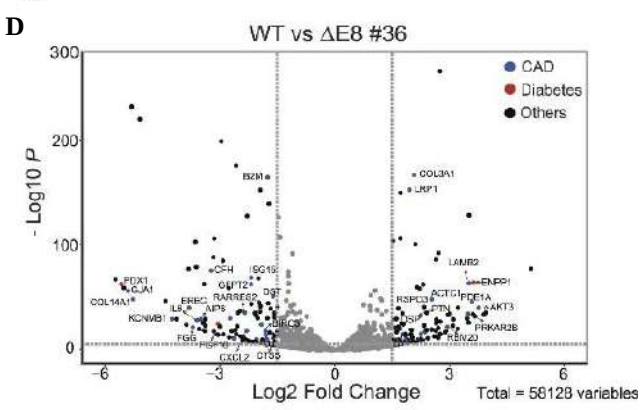
Figure 2.15: Expression of *INK4/ARF* genes is perturbed upon enhancer deletions. ***A-D)*** *UCSC browser shots of RNA-seq showing the signal from exons and entire gene body of* ***A)*** *CDKN2A,* ***B)*** *CDKN2B,* ***C)*** *CDKN2BAS,* ***D)*** *MTAP; expression in E8 and E17 enhancer deletion lines.* ***E)*** *Control genes outside INK4/ARF TAD show no change in expression upon E8 and* E17 deletion.

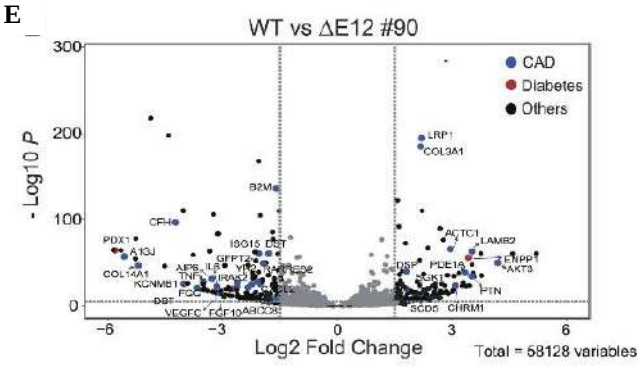
Then, when comparing the WT and E8 deletion, we found about 200 genes that had differential regulation (Figure 2.16A). Notably, we found that these genes were similarly impacted in E12 and E17 (Figure 2.16A), demonstrating that the direct or indirect targets of these enhancers are common. About 100 dysregulated genes were upregulated and 100 were downregulated (Figure 2.16A). The enriched disease terms associated with these genes included age-related macular degeneration, coronary artery disease, diabetes mellitus, cardiomyopathy, and cardiomyopathy (Figure 2.16B). Additionally, the majority of the dysregulated genes were discovered in the cortex and smooth muscle cells (Figure 2.16C). These differentially regulated genes, which are well-known candidates with significant roles, were discovered to be related to CAD and T2D (Figures 2.16D – 2.16F). For instance, all deletion lines showed a substantial downregulation of the important insulin receptor PDX1. The gene COL3A1, which is frequently changed or mutated in aortic and arterial aneurism, was one of the genes with the highest fold upregulation and lowest p-value. Similarly, COL3A1 was elevated to similar levels as LRP1, an essential component in the clearance of apoptotic cells and lipid metabolism. Dysregulated interleukin-6 (IL-6) levels have been linked to endothelial function in *INK4/ARF*-deficient mice (Mosteiro et al. 2018; Yang et al. 2012).

A




B
Relative Enrichment
0
50
100
150
200
250
300
Diabetes mellitus
Right ventricular cardiomyopathy
Coronary Artery Disease
Age-related macular degeneration
Charcot-marie-tooth disease
Diabetes Mellitus type 2
Atrial septal defect
Familial dilated cardiomyopathy
Familial colorectal cancer
Myofibrillar myopathy
Idiopathic generalized epilepsy
Familial hyperinsulinism
Lynch syndrome
Hemolytic-uremic syndrome
Asthma
Carcinoma of colon
Obesity
Schizophrenia
C
Relative Enrichment
0
5
10
15
20
Smooth Muscle
Cingulate Cortex
Spinal cord
Prefrontal Cortex
Olfactory Bulb
Cardiac Myocytes
Medulla Oblongata
Bronchial Epithelial
Uterus
Pineal night


D
WT vs ΔE8 #36
CAD
Diabetes
Others
- Log10 P
Log2 Fold Change
Total = 58128 variables


E
WT vs ΔE12 #90
CAD
Diabetes
Others
- Log10 P
Log2 Fold Change
Total = 58128 variables

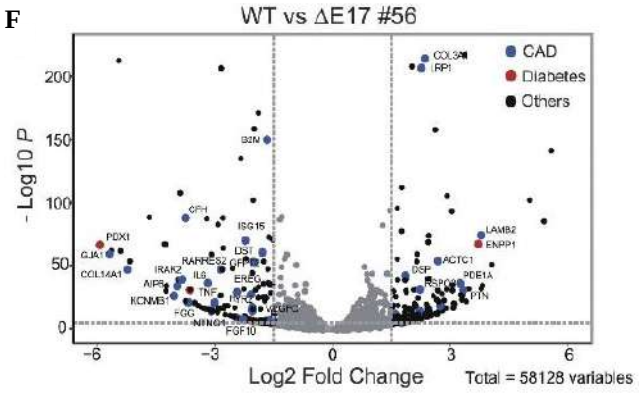


Figure 2.16: Dysregulated genes upon enhancer deletions corroborate with the disease association of the 9p21 locus. A) *Venn diagram showing the number of common, unique, downregulated genes (upper diagram) and upregulated genes (lower diagram) upon enhancer deletions.* ***B)*** *Graph from the Enrichr tool shows the relative enrichment of disease terms based on dysregulated genes upon the E8, E12, and E17 enhancer deletions.* ***C)*** *Graph from the Enrichr tool shows the relative enrichment of dysregulated genes, also expressed in various cell types.* ***D-F)*** *Volcano plots show the log2 fold changes versus –log10p values of genes comparing WT with* ***D)*** *E8,* ***E)*** *E12, and* ***F)*** *E17 deletions.*

Chapter 3: Results (Part II)

JMJD3 regulates *INK4/ARF* locus in HeLa cells

JMJD3 is a histone demethylase that removes trimethyl marks from lysine 27 of Histone 3 (H3K27me3). JMJD3 has been demonstrated to activate the *INK4/ARF* locus during the onset of senescence by removing the H3K27me3 mark from the promoters of the *CDKN2A* and *CDKN2B* genes (Agger et al. 2009). Upon oncogene-induced senescence (OIS), JMJD3 stimulates both INK4a and ARF expression in mouse embryonic fibroblasts (MEFs), but only INK4a expression in human fibroblasts (Agger et al. 2009). Thus, JMJD3 causes $p16^{INK4a}$-mediated cell cycle arrest in fibroblasts. In addition, JMJD3 has been shown to operate as a barrier to somatic cell reprogramming by catalytically activating the *INK4/ARF* locus (Li et al. 2009; Price et al. 2014). JMJD3, for instance, influences Pol II to modulate gene transcription elongation (Chen et al. 2012). As previously described, HeLa cells express *INK4/ARF* transcripts at very high levels. To test if the expression of these genes in HeLa cells is dependent on JMJD3, we knocked down JMJD3 using siRNAs specific to it. JMJD3 knockdown caused the downregulation of all three transcripts, INK4a, ARF, and INK4b in HeLa cells (Figure 3.1A). As a result, JMJD3 appears to be a positive regulator of this locus in HeLa cells as well. Even though JMJD3 is an enzyme that removes H3K27me3 marks from regulatory elements, multiple studies have revealed the novel roles of JMJD3 in a catalytic-independent manner. JMJD3, for example, restricts somatic cell reprogramming by targeting PHF20 for ubiquitination and subsequent degradation in a catalytic-independent manner. JMJD3 accomplishes this by recruiting Trim26, an E2 ubiquitin ligase (Zhao et al. 2013). To test if JMJD3 regulates *INK4/ARF* genes in a catalytic-dependent or independent manner in HeLa cells, we treated cells with 30μM GSK-J4 for 24 hours. GSK-J4 is a cell-permeable drug that specifically inhibits JMJD3 by blocking its catalytic activity (Li et al. 2018). We observed significant downregulation of all three transcripts, INK4a, ARF, and INK4b upon GSK-J4 treatment (Figure 3.1B). According to these results, the catalytic activity of JMJD3 is essential for *INK4/ARF* locus transcription in HeLa cells. PRC2 complex acts as an antagonist to JMJD3 and is known to repress this locus in a variety of cell types such as hESCs and normal proliferating cells (Bracken et al. 2007; Maertens et al. 2009). PRC2 achieves the repression of this locus by adding a trimethyl group to lysine 27 of Histone 3. As mentioned previously, JMJD3 removes the trimethyl mark upon

receiving different cues, thereby activating the locus. To further validate if JMJD3 and PRC2 interplay also takes place in the HeLa cell line, we inhibited the catalytic activity of EZH2 which is the catalytic subunit of PRC2 by treating cells with GSK343 for 24 hours. Upon inhibition of the catalytic activity of EZH2 all three transcripts show modest upregulation (Figure 3.1C). This data suggests that JMJD3 regulates *INK4a/ARF* gene expression in a catalytic-dependent manner and the interplay between JMJD3 and PRC2 is important even in the HeLa cell line. However, JMJD3 outcompetes PRC2 and keeps the locus active.

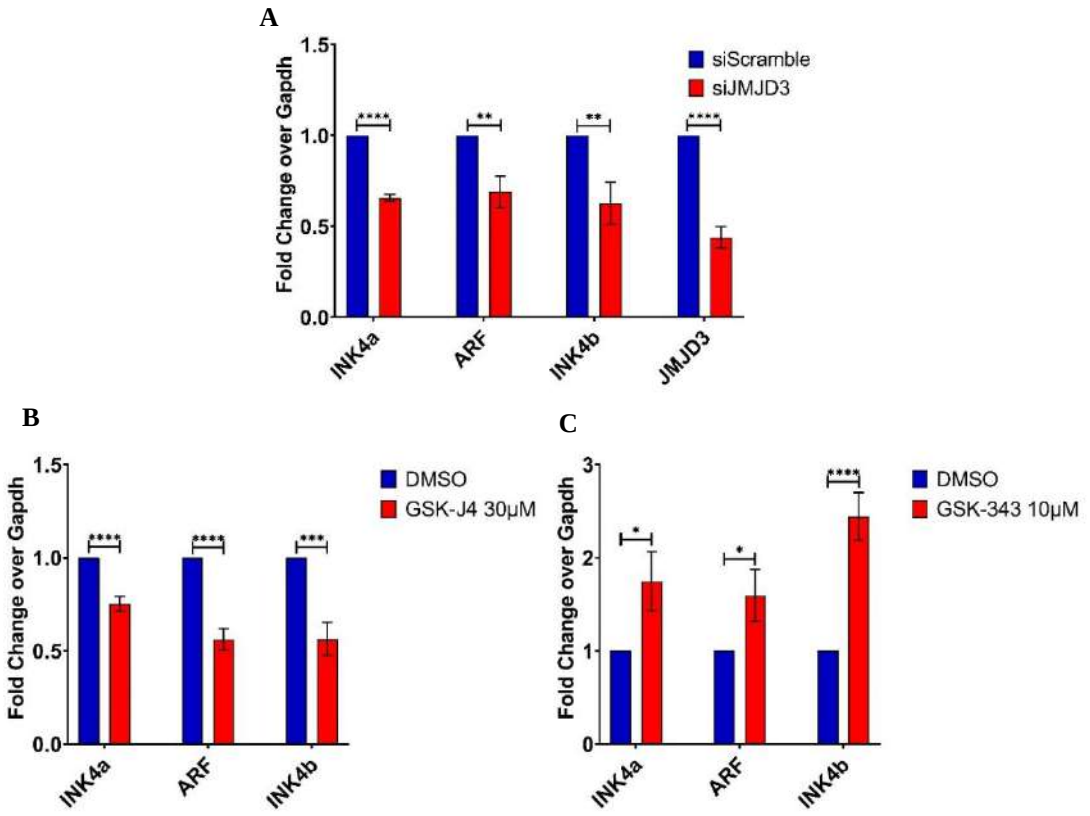
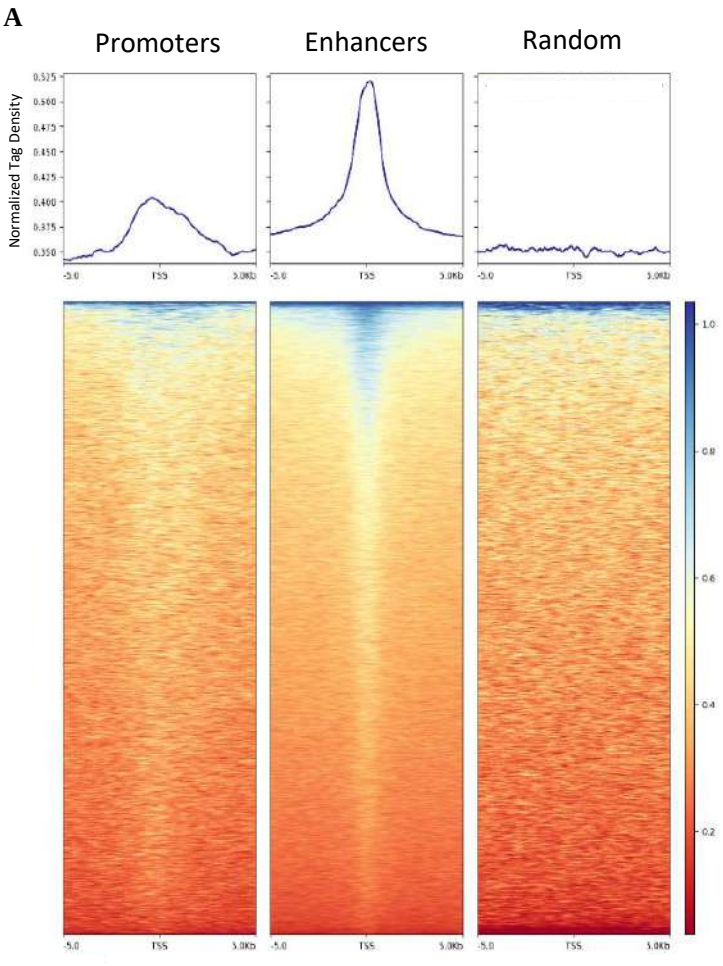


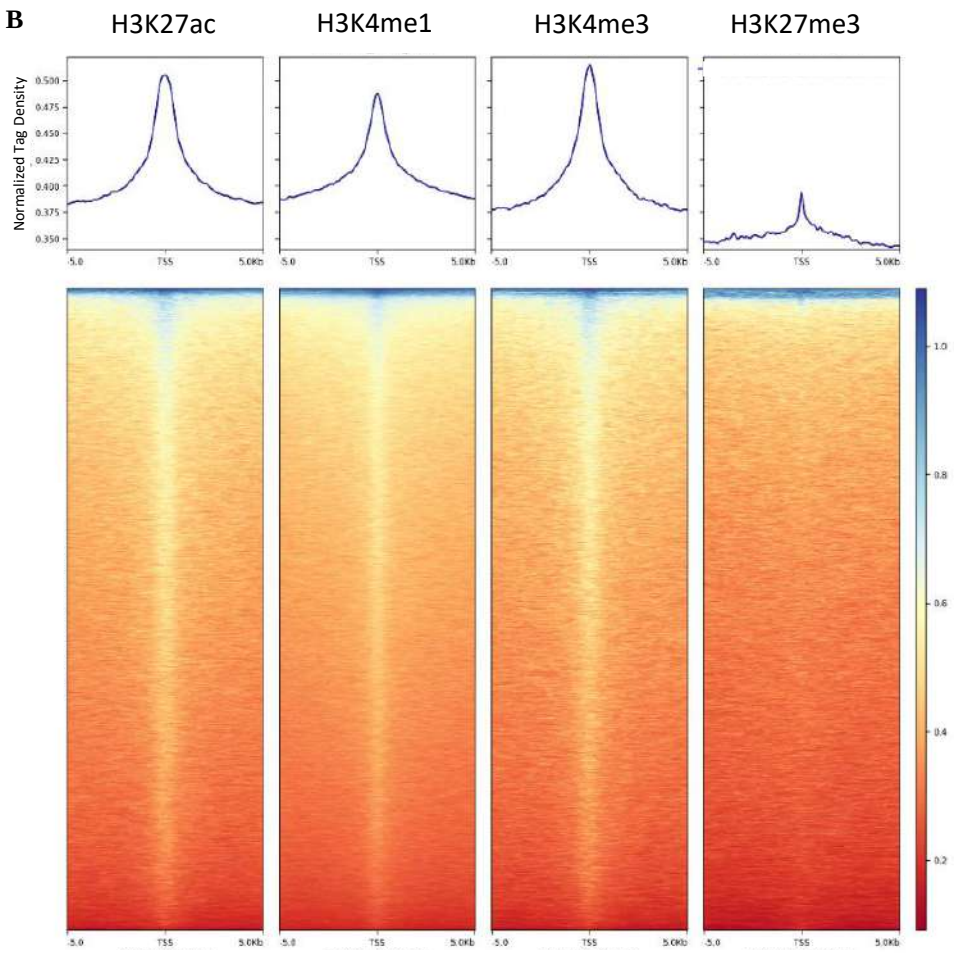
Figure 3.1: JMJD3 regulates *INK4/ARF* locus in a catalytic-dependent manner in the HeLa cell line. ***A)*** *Expression of INK4a, ARF, and INK4b transcripts is downregulated upon JMJD3 knockdown.* ***B)*** *Catalytic inhibition of JMJD3 results in the downregulation of all three INK4/ARF transcripts. Cells were treated with 30uM GSK-J4 for 24 hours.* ***C)*** *Inhibition of the catalytic subunit of PRC2 results in the upregulation of all three INK4/ARF transcripts. Cells were treated with 10uM GSK343 for 24 hours.*

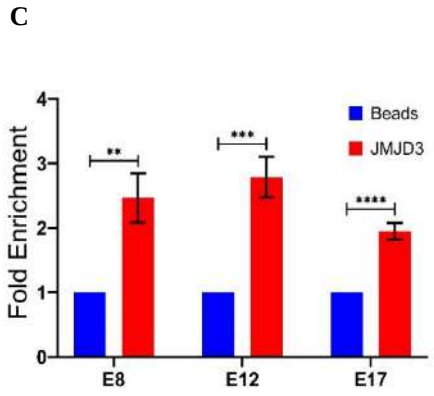
JMJD3 binds to regulatory enhancers of the *INK4/ARF* locus as well as other active enhancers across the genome

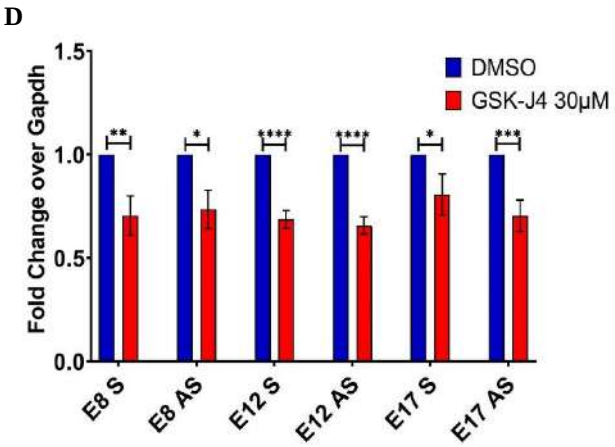
JMJD3 has been demonstrated to bind promoters and enhancers across the genome by cooperating with various transcription factors (Svotelis et al. 2011). For instance, p53 recruits JMJD3 to p53 responsive sites, and its binding to p53 response elements (PREs) increases in response to DNA damage (Williams et al. 2014). Furthermore, JMJD3 is recruited by KLF4 to specific regions where it removes trimethyl marks, thereby activating the target genes (Huang et al. 2020). To identify the regulatory elements which are occupied by JMJD3 in HeLa cells, we performed ChIP-seq for JMJD3. ChIP-seq data showed JMJD3 to be enriched on both enhancers and gene promoters, however, it was more enriched on enhancers than on promoters (Figure 3.2A). The enrichment of JMJD3 on enhancers suggests that it may primarily regulate its target genes via their regulatory enhancers. Further examination of JMJD3-bound sites revealed that JMJD3 binds to enhancers and promoters that are enriched with active histone marks. JMJD3-bound regions, for example, have high levels of H3K27ac, H3K4me1, and H3K4me3. Moreover, JMJD3-bound regions show very less enrichment for H3K27me3 (Figure 3.2B). All of these findings indicate that JMJD3-bound regions, whether promoters or enhancers are active regions. The predominant binding of JMJD3 on enhancers motivated us to investigate if JMJD3 binds to previously identified regulatory enhancers at *INK4/ARF* locus. To test this, we performed ChIP qPCRs for JMJD3 and observed JMJD3 binding to the regulatory enhancers in the locus. JMJD3 was found to bind to all three functional enhancers, E8, E12, and E17 (Figure 3.2C). However, it was more enriched on E8 and E12 than on E17. To validate if JMJD3 is required for the activity of these enhancers, we inhibited the catalytic activity of JMJD3 by treating cells with GSK-J4 for 24 hours. We observed that upon catalytic inhibition of JMJD3, eRNAs transcribed from these regulatory enhancers are down-regulated (Figure 3.2D). This suggests that the transcription of these enhancers is also dependent on the catalytic activity of JMJD3. Interestingly, the eRNA expression correlates with the *INK4/ARF* gene expression when the catalytic activity of JMJD3 is inhibited. Moreover, the inhibition of the catalytic activity of PRC2 by GSK343 for 24 hours resulted in the upregulation of eRNAs (Figure 3.2E). eRNA upregulation by EZH2 inhibition encouraged us to investigate whether JMJD3 knockdown causes PRC2 binding to these enhancers. Towards that, we knocked down

JMJD3 using a pool of siRNAs and subsequently performed ChIP qPCRs for EZH2 and repressive mark, H3K27me3. Surprisingly, the E8 and E12 enhancers did not show a substantial increase in the enrichment of EZH2 and H3K27me3 (Figures 3.2F and 3.2G). These data imply that JMJD3 is essential for enhancer activity, however, its perturbation does not result in an increase in EZH2 and H3K27me3 on these enhancers. The increase of eRNA by PRC2 inhibition could be attributed to the indirect effect of PRC2. PRC2 may be catalytically suppressing several transcription factors essential for the activation of these enhancers. Taken together, these findings indicate that JMJD3 binds to regulatory elements and that its binding to these enhancers is required for their activation.




B
H3K27ac
H3K4me1
H3K4me3
H3K27me3
Normalized Tag Density
TSS


C
Fold Enrichment
Beads
JMJD3
E8
E12
E17


D
Fold Change over Gapdh
DMSO
GSK-J4 30μM
E8 S
E8 AS
E12 S
E12 AS
E17 S
E17 AS

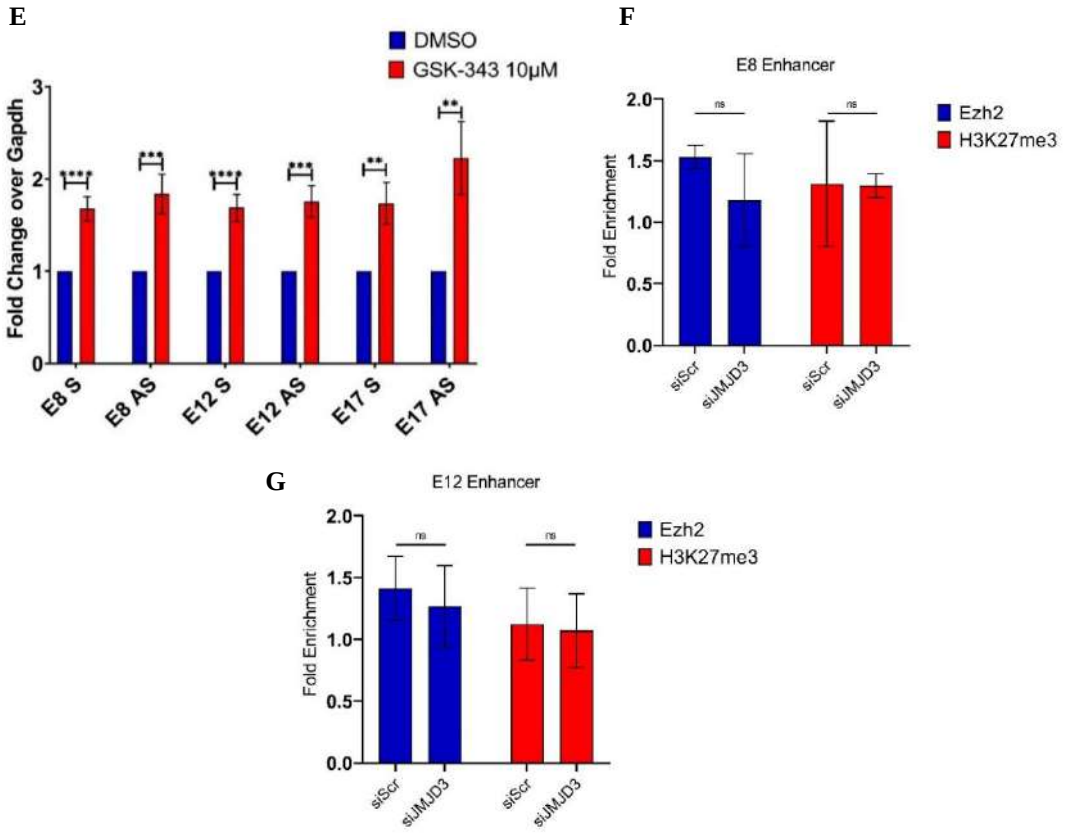


Figure 3.2: JMJD3 binds to active enhancers and promoters genome-wide. ***A)*** *ChIP-seq heatmap showing the binding of JMJD3 on active enhancers and promoters. JMJD3 is more enriched on enhancers than promoters. Random regions (5000) did not show any binding of JMJD3.* ***B)*** *Heatmaps showing JMJD3 bound sites are enriched with active marks like H3K27ac, H3K4me1, and H3K4me3. JMJD3-bound sites show little enrichment of repressive mark like H3K27me3.* ***C)*** *ChIP qPCRs showing JMJD3 enrichment on INK4/ARF regulatory enhancers E8, E12, and E17.* ***D)*** *Catalytic inhibition of JMJD3 results in the downregulation of sense and anti-sense eRNAs of E8, E12, and E17 enhancers. Cells were treated with 30uM GSK-J4 for 24 hours.* ***E)*** *Catalytic inhibition of EZH2 with GSK343 results in the upregulation of sense and anti-sense eRNAs of E8, E12, and E17 enhancers. Cells were treated with 10uM GSK343 for 24 hours* ***F-G)*** *EZH2 and H3K27me3 enrichment on E8 and E12 enhancers upon JMJD3 knockdown. EZH2 and H3K27me3 does not increase on these enhancers upon JMJD3 knockdown.*

INK4/ARF functional enhancers are transcribed into bidirectional eRNAs with regulatory potential

Active enhancers have been demonstrated to attract Pol II, resulting in enhancer transcription. The resulting non-coding transcripts are known as enhancer RNAs (eRNAs). These eRNAs have been implicated to have a variety of regulatory roles, and their dysregulation has been linked to several diseases and malignancies. eRNAs, for example, are critical for enhancer-promoter interactions, chromatin accessibility, and other processes (Li, Notani, and Rosenfeld 2016; Arnold, Wells, and Li 2020). Several studies have shown that eRNAs interact with transcription factors or cofactors and so play a key role in enhancer activity and, as a result, target gene regulation. For example, eRNAs have been demonstrated to interact with and aid the binding of the YY1 transcription factor to its target regions, activating the transcription of their target genes (Sigova et al. 2015). eRNAs help in the release of paused polymerase by interacting and releasing the negative elongation factor, NELF (Gorbovytska et al. 2022; Schaukowitch et al. 2014). As previously described, the functional enhancers of the *INK4/ARF* locus have Pol II enrichment and are transcribed into bidirectional eRNAs. We found that all promoter-interacting enhancers produce bidirectional eRNAs (Figure 3.3A-3.3B). The E8 and E17 enhancers showed more eRNA transcription than the E12 enhancer. The eRNA levels of these enhancers corroborated very well with the Pol II levels bound on these enhancers (Figure 3.3B). To test if eRNAs produced from promoter-interacting enhancers are important for *INK4/ARF* gene transcription, we designed and cloned shRNAs that target both sense and antisense eRNAs transcribed from all three enhancers. To efficiently knock down the eRNAs, three shRNAs were designed for each sense and antisense transcript arising from E8 and E12 enhancers, and two for E17 enhancer. Cell clones carrying shRNAs that target sense and antisense transcripts of individual enhancer were created. When eRNAs produced from the promoter interacting enhancers were knocked down, all three *INK4/ARF* transcripts were significantly downregulated (Figure 3.3C-3.3E). These data suggest that the eRNAs transcribed from the regulatory enhancers may play a role in the transcription of these genes.

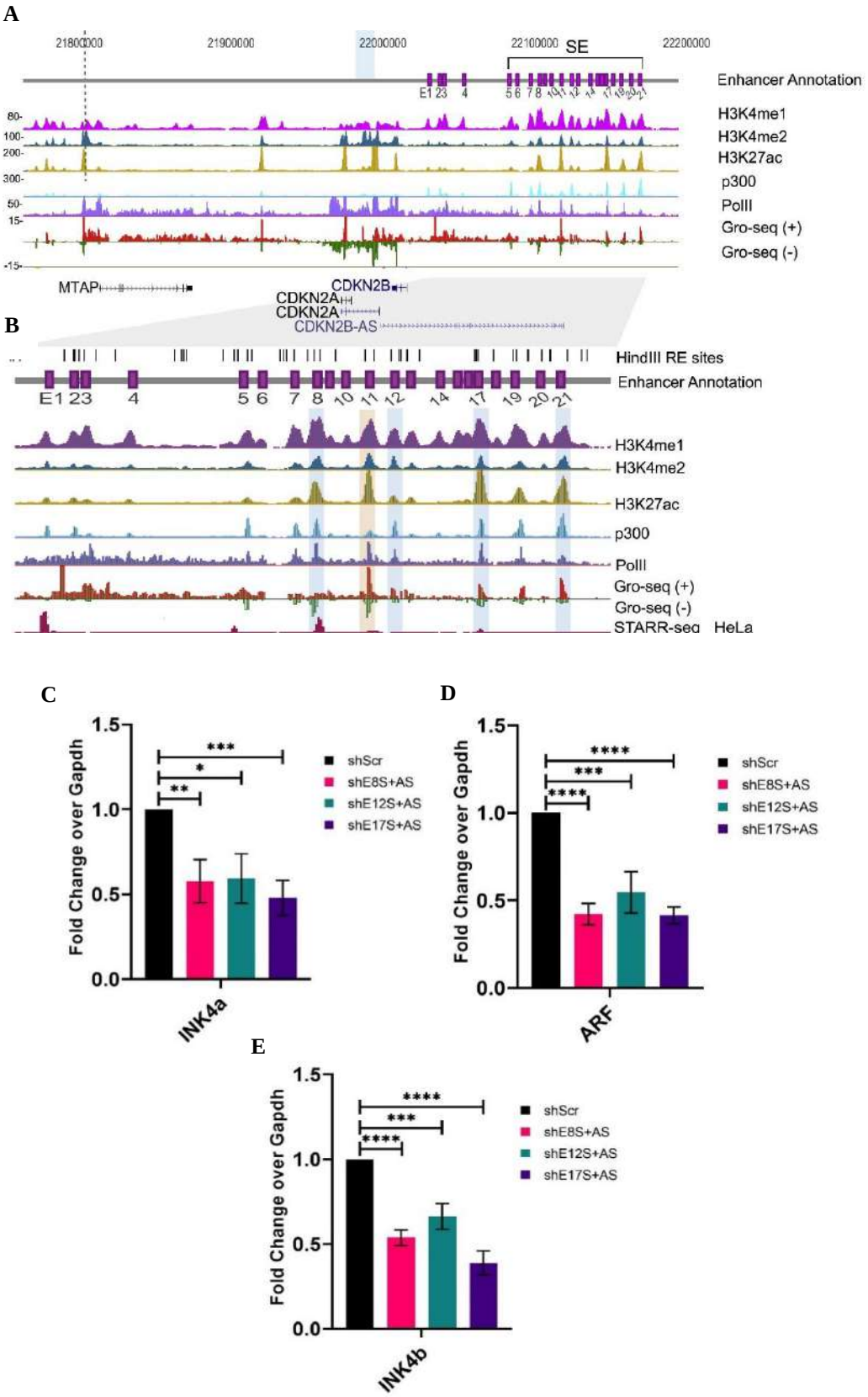


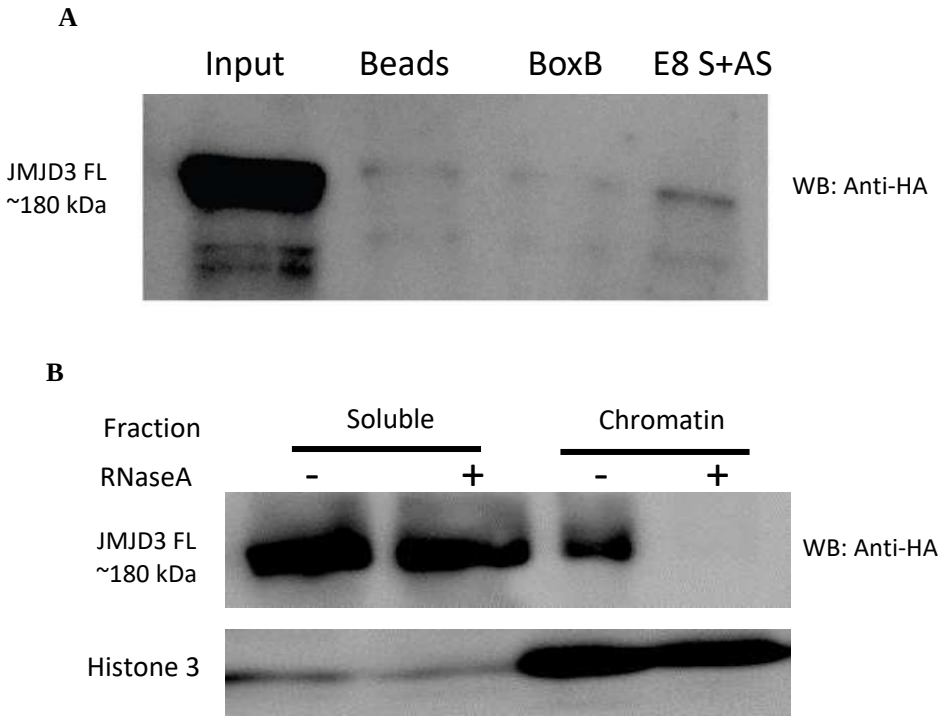
Figure 3.3: Promoter interacting enhancers are transcribed into bidirectional eRNAs with regulatory potential. ***A)*** *UCSC genome browser tracks showing enrichment of different enhancer active marks on INK4/ARF enhancer cluster. The*

ChIP-seq tracks are overlaid with GRO-seq tracks from the HeLa cell line. Gro-seq tracks show that E8, E12, and E17 enhancers are transcribed into bidirectional eRNAs. ***B)*** *The zoomed region from E1 to E21 is shown in the UCSC browser shot. The tracks are overlaid with the HeLa STARR-seq track.* ***C-E)*** *eRNA knockdown of E8, E12 and E17 enhancers result in the downregulation of all three INK4/ARF transcripts* ***C)*** *INK4a;* ***D)*** *ARF and* ***E)*** *INK4b*

JMJD3 is an RNA-interacting histone demethylase that binds *INK4/ARF* enhancers in an eRNA-dependent manner

JMJD3 has been shown to interact with Kaposi's sarcoma-associated herpesvirus (KSHV) PAN RNA and activate the lytic replication cycle by physically associating it with the viral genome (Rossetto and Pari 2012). Another study demonstrated that JMJD3 associates with lncRNA ARHGAP27P1 and this association is required for demethylation and subsequent activation of $p16^{INK4a}$ and $p15^{INK4b}$, thus inhibiting gastric cancer (Zhang et al. 2019). Furthermore, JMJD3 has been also shown to recruit DDX21 to the ENPP2 locus wherein it assists in the prevention of R-loop formation, thereby promoting transcription (Argaud et al. 2019). Though these studies showed the association of JMJD3 with different RNAs but what role this association plays is not known. To understand the role of JMJD3 and RNA association, we tested if JMJD3 interacts with eRNAs transcribed from *INK4/ARF* enhancers. Towards that, we labeled the sense and antisense eRNAs from the E8 enhancer by performing in-vitro transcription using biotin-labeled rNTPs. We incubated the labeled RNA with HeLa protein lysate containing HA-tagged JMJD3 FL. Pulldown of labeled E8 eRNAs showed that JMJD3 is one of its protein binding partners (Figure 3.4A). Labeled BoxB RNA was used as a negative control. This suggests that JMJD3 can interact with E8 eRNA. To understand, if JMJD3's interaction with RNA plays any role in its chromatin binding. We isolated nucleoplasmic and chromatin fractions of JMJD3 in the presence and absence of RNase A. We observed that JMJD3 is almost completely localized inside the nucleus and is present in both the nucleoplasm and chromatin fraction. Interestingly, the removal of RNA by treating the cells with RNase A resulted in the drastic reduction of JMJD3 from chromatin fraction (Figure 3.4B). This suggests that JMJD3 and RNA association assist in its binding to chromatin. Previously, we have

shown that JMJD3 binds to all three regulatory enhancers namely E8, E12, and E17 and JMJD3 binds to eRNAs transcribed from these enhancers. Moreover, fractionation data also revealed that upon RNA removal, a chromatin-bound fraction of JMJD3 is lost. This compelled us to check if JMJD3 binding on the regulatory enhancers of *INK4/ARF* locus is dependent on eRNAs transcribed from these enhancers. Towards this, we knocked down both sense and antisense transcripts from all three enhancers individually and performed ChIP qPCRs of JMJD3. We found that with the knockdown of eRNAs, JMJD3 occupancy on these enhancers is reduced (Figure 3.4C-E) suggesting that JMJD3 binding on these enhancers is dependent on eRNAs. These eRNAs could either be assisting in the recruitment of JMJD3 on these enhancers or they can be aiding in the stabilization of JMJD3 onto these enhancers. Additionally, eRNAs may function in the sequestration of PRC2, allowing JMJD3 to bind to chromatin.



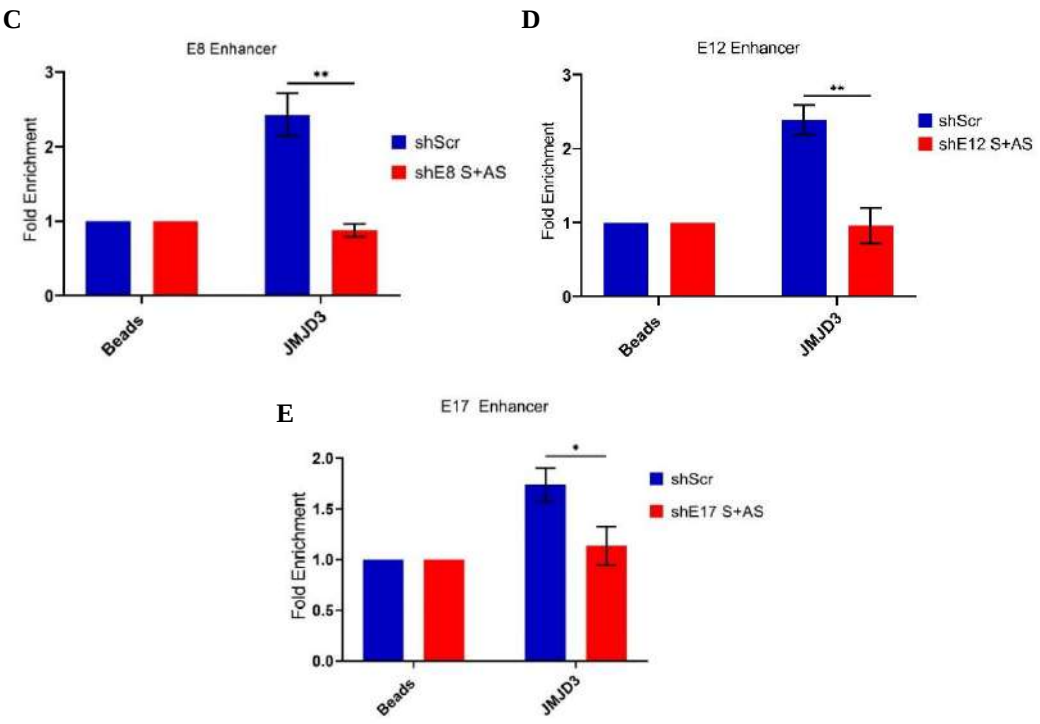
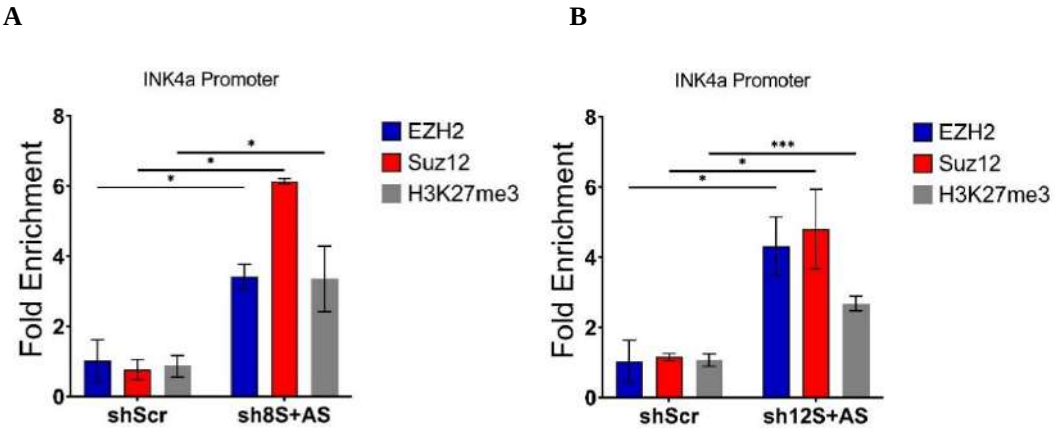


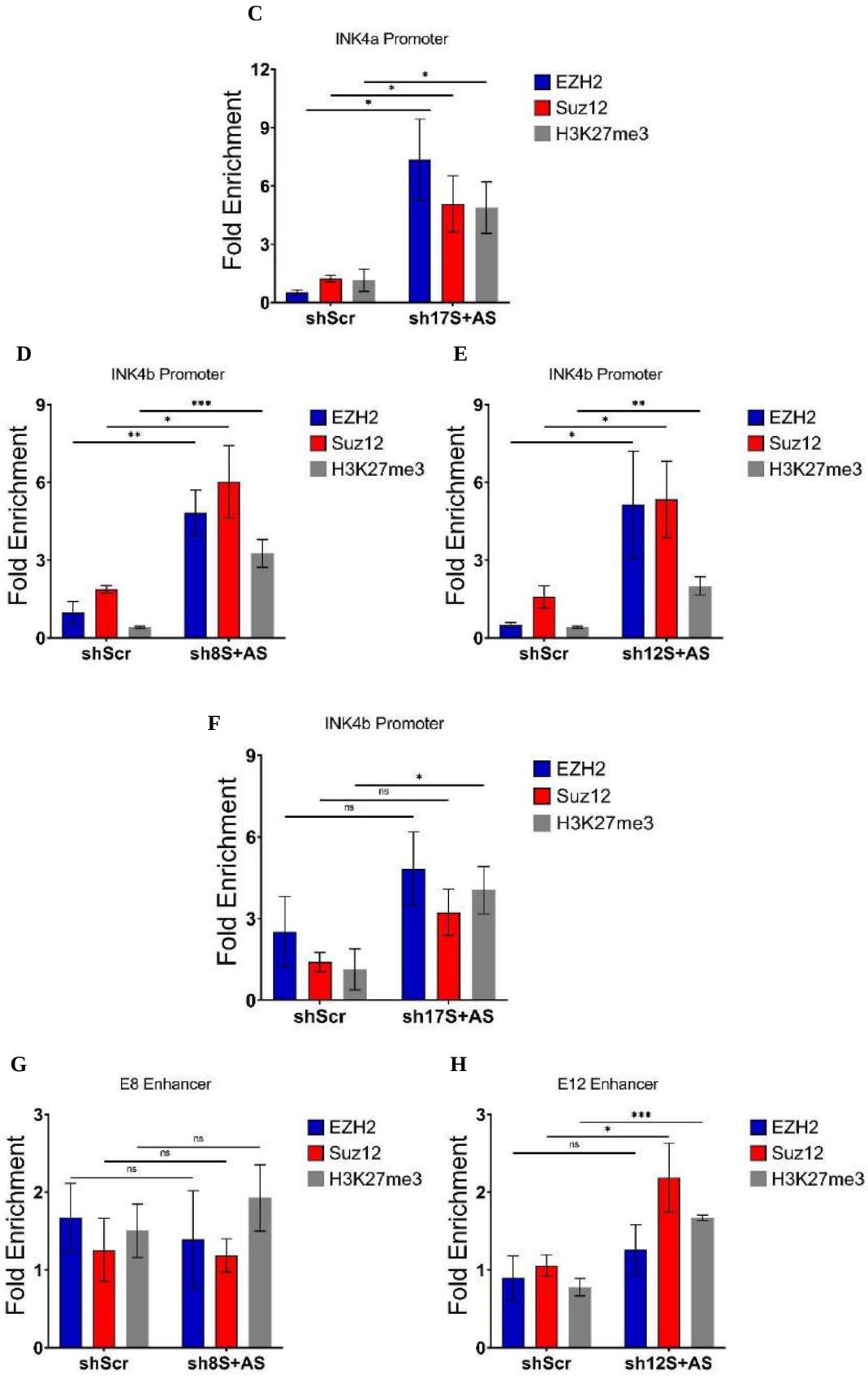
Figure 3.4: JMJD3 is an RNA-binding histone modifier. ***A)*** *RNA pulldown followed by western blot showing JMJD3 interacts specifically with E8 eRNAs and not with BoxB RNA.* ***B)*** *Western blot showing the fractionation of JMJD3 FL in the presence and absence of RNase A. JMJD3 FL is majorly present in nucleoplasmic fraction. JMJD3 FL loses chromatin binding upon RNA degradation.* ***C-E)*** *JMJD3 ChIP qPCRs upon eRNA knockdown with shRNAs. JMJD3 enrichment on* ***C)*** *E8;* ***D)*** *E12; and* ***E)*** *E17 enhancers go down with the knockdown of eRNAs transcribed from these enhancers.*

eRNA loss triggers PRC2 enrichment on *INK4/ARF* genes but not on the enhancers

PRC2 has been demonstrated to silence this locus in several cancers and normal proliferating cells. Several subunits of PRC2 like EED, EZH2, and Suz12 have been shown to interact with RNA (Kretz and Meister 2014). The PRC2 and RNA interaction have been called promiscuous as this complex interacts with almost all types of RNA without specificity, thus without a defining property (Kretz and Meister 2014). The interplay between RNA and PRC2 has been shown to prevent the silencing of actively transcribing genes (Kaneko et al. 2014). The transcribed RNA associates with PRC2 leading to its sequestration, thereby restricting its association with the DNA, thus

preventing silencing. However, in other cases RNAs have been shown to assist PRC2 to bind to its targets, thereby silencing them (Long et al. 2020). Because eRNA knockdown results in the downregulation of *INK4/ARF* genes. We wanted to examine if the PRC2 complex is involved in the downregulation. Towards that, we performed ChIP qPCRs of two subunits of PRC2 complex EZH2 and SUZ12 upon eRNA knockdowns. Interestingly, we found that EZH2 and SUZ12 enrichment was increased on INK4a and INK4b promoters (Figure 3.5A-3.5F). The enrichment of these subunits is also accompanied by increased H3K27me3 (Figure 3.5A-3.5F). This data suggests that the eRNAs transcribed from the functional enhancers prevent PRC2 to target the promoters for their silencing. We also tested the enrichment of EZH2 and SUZ12 on the enhancers upon their eRNA knockdowns. Interestingly, we observed no increase in the enrichment of these subunits on enhancers upon eRNA knockdown (Figure 3.5G-3.5I). This suggests that PRC2 does not target these enhancers for silencing. The data are in line with our previous findings where we observed that enhancer knockout results in PRC2 enrichment on *INK4/ARF* promoters. However, PRC2 enrichment does not increase on the intact enhancers upon individual enhancer knockouts.




C
INK4a Promoter
Fold Enrichment
12
9
6
3
0
shScr
sh17S+AS
EZH2
Suz12
H3K27me3
D
INK4b Promoter
Fold Enrichment
shScr
sh8S+AS
E
INK4b Promoter
Fold Enrichment
shScr
sh12S+AS
F
INK4b Promoter
Fold Enrichment
shScr
sh17S+AS
G
E8 Enhancer
Fold Enrichment
shScr
sh8S+AS
H
E12 Enhancer
Fold Enrichment
shScr
sh12S+AS

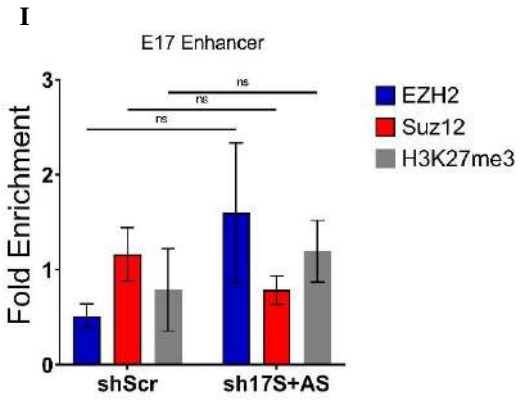
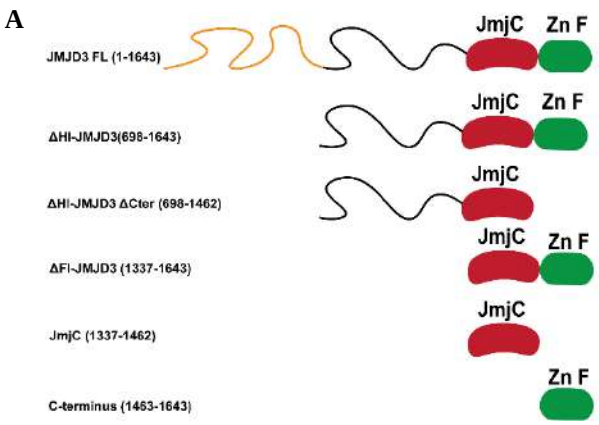


Figure 3.5: eRNAs from functional enhancers prevent PRC2 to target *INK4/ARF* promoters. ***A-C)*** *EZH2, Suz12 and H3K27me3 enrichment on INK4a promoter upon* ***A)*** *E8,* ***B)*** *E12 and* ***C)*** *E17 eRNAs knockdown.* ***D-F)*** *EZH2, Suz12, and H3K27me3 enrichment on INK4b promoter upon* ***D)*** *E8,* ***E)*** *E12, and* ***F)*** *E17 eRNAs knockdown. EZH2, Suz12, and H3K27me3 are enriched on INK4a and INK4b promoters with eRNA knockdowns.* ***G-I)*** *EZH2, Suz12, and H3K27me3 enrichment on* ***G)*** *E8,* ***H)*** *E12, and* ***I)*** *E17 enhancers upon knocking down eRNAs transcribed from these enhancers. eRNA knockdown does not result in PRC2 binding on enhancers.*

The C-terminus region of JMJD3 interacts with RNA

Our data suggest that JMJD3 is an RNA-binding histone demethylase. Moreover, previous studies have also reported an association of JMJD3 with different RNAs (Rossetto and Pari 2012). However, the JMJD3 domain that interacts with RNA is still unknown. Thus, we went on to identify the domain of JMJD3 protein that interacts with RNA. JMJD3 protein has one non-canonical region at its N-terminus (which is an IDR discussed later in the chapter) and one characterized domain, the JmjC domain, at its C-terminus. JmjC is the catalytic domain of JMJD3, and it has been shown to remove trimethyl marks from lysine 27 of histone 3 (Hong et al. 2007). The extreme C-terminal region of JMJD3 after the JmjC domain is a Zn^{2+} binding region. Zinc binding results in the formation of a zinc finger-like domain. The roles of the N-terminal IDR and C-terminal domain in JMJD3 functioning remains unknown. Zinc finger-like domains are known to interact with nucleic acids such as DNA and RNA (Font and Mackay 2010). To better understand the role of IDR (at the N-terminal) and Zinc binding region at the

C-terminal in RNA binding and subsequent chromatin binding of JMJD3, we created several truncation constructs of JMJD3 such as JMJD3 with first half IDR deleted (ΔHI-JMJD3), JMJD3 with full IDR deleted (ΔFI-JMJD3), JMJD3 with C-terminus deleted (ΔCT-JMJD3); JMJD3 with first half IDR and C-terminus deleted (ΔHIΔCT-JMJD3); just JmjC, and C-terminus region separately (Figure 3.6A). We first performed western blotting to examine the expression of all of these truncations. Except for JMJD3 C-terminus deletion (ΔCT-JMJD3), all constructs expressed normally (Figure 3.6B). Moreover, JmjC alone construct expressed at lower levels. We performed in vitro transcription to label E8 sense and antisense eRNAs with biotin. We then carried out RNA pull-down assays with these labeled RNAs. The data indicated that the C-terminal region of JMJD3 interacts with RNA (Figure 3.6C). To further validate the RNA pulldown results, we separated nucleoplasmic and chromatin fractions of the C-terminal domain in the presence and absence of RNase A. The C-terminal domain was found in both the nucleoplasm and the chromatin fraction. Nevertheless, removing RNA using RNase A resulted in the loss of the C-terminal domain from the chromatin fraction (Figure 3.6D). Additionally, fractionation of ΔHIΔCT-JMJD3 without the C-terminus in the presence and absence of RNase A demonstrated that the deletion of the C-terminus results in loss of chromatin binding. Furthermore, the removal of the C-terminus renders JMJD3 resistant to RNase A treatment, leading to no change in the chromatin fraction of this truncation (Figure 3.6E). The results confirm our prior findings and imply that JMJD3's C-terminus is the RNA binding domain and its chromatin binding is dependent on RNA.



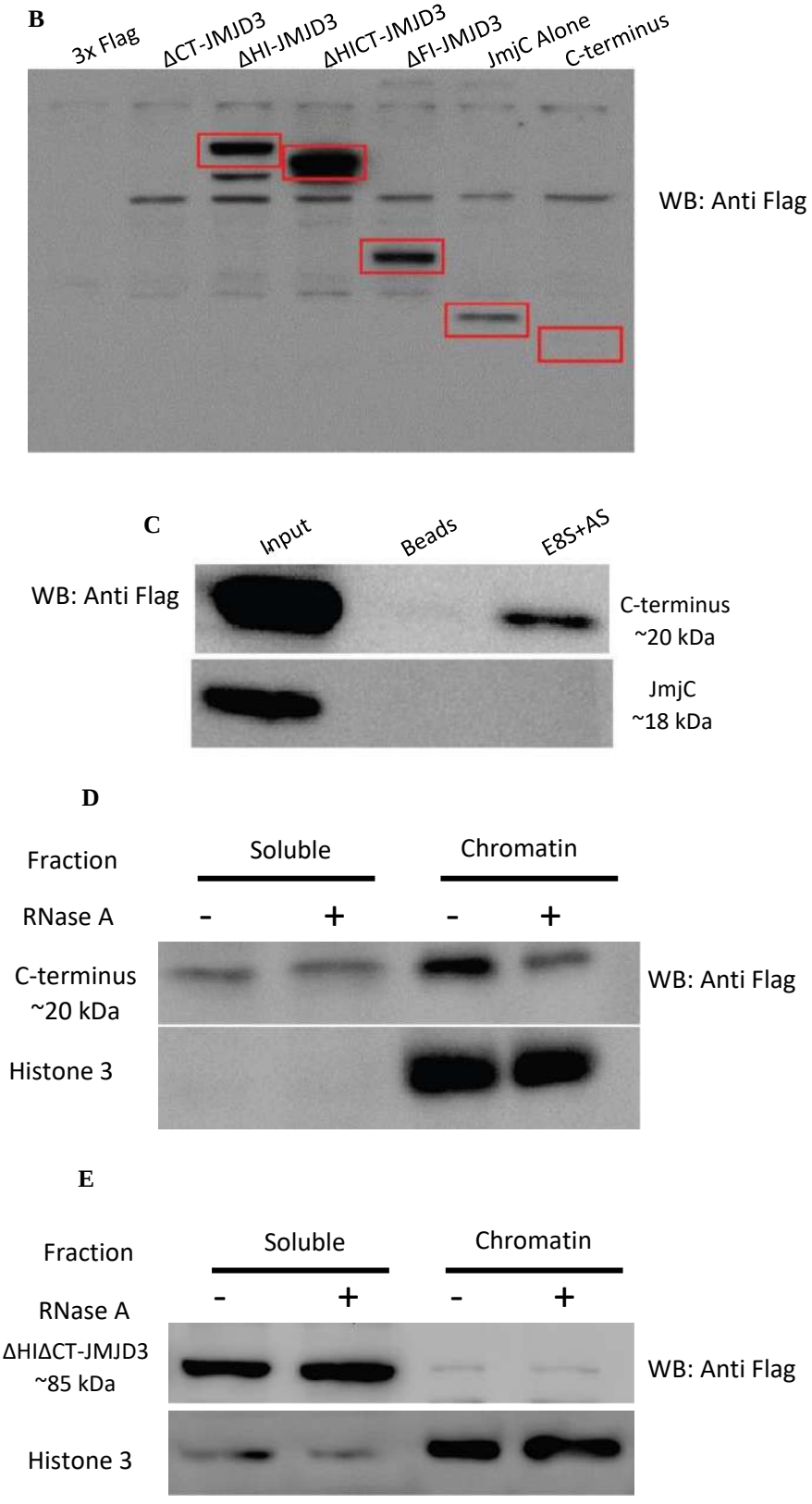


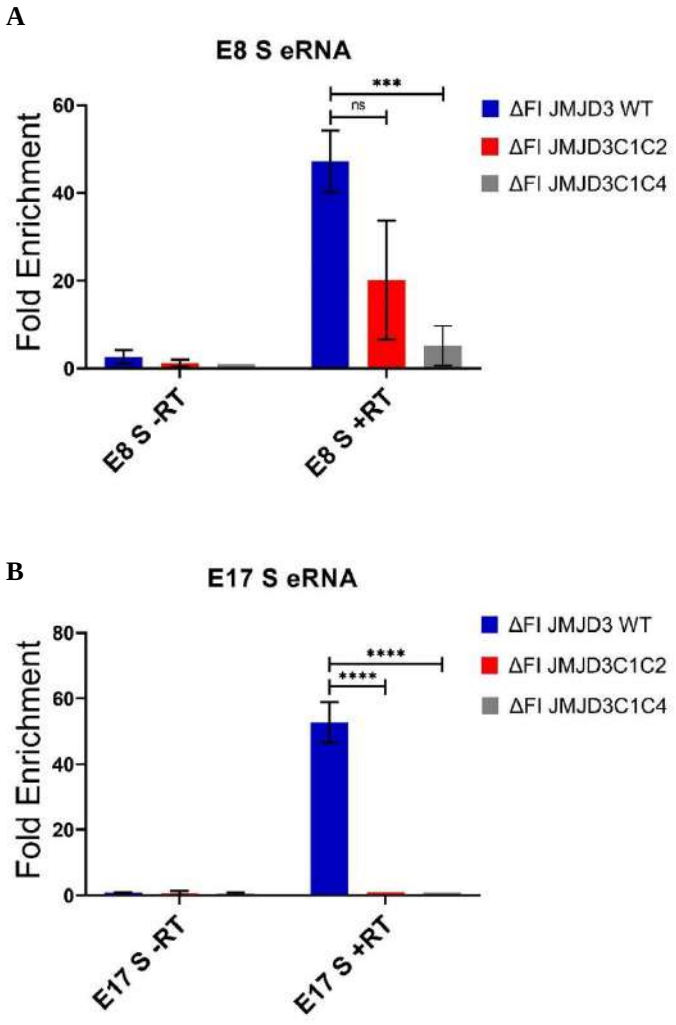
Figure 3.6: C-terminus of JMJD3 interacts with RNA. ***A)*** *Schematic depicting different truncations of JMJD3 that were created to understand chromatin and RNA binding of JMJD3* ***B)*** *Western blot showing that all the variants are expressing properly*

except JMJD3 C-terminus deleted (ΔCT-JMJD3). *Moreover, constructs carrying just the JmjC domain shows less expression compared to others.* ***C)*** *Western blot showing C-terminus of JMJD3 interacts with RNA and not JmjC domain.* ***D)*** *Fractionation of the C-terminus shows that it is mainly present in nucleoplasmic and chromatin fractions. Removal of RNA results in loss of C-terminus from chromatin fraction.* ***E)*** *Fractionation of ΔHIΔCT-JMJD3 shows that it is mainly present in nucleoplasmic fraction. It is also seen in chromatin fraction. Deletion of the C-terminus affects its chromatin binding and makes it resistant to RNase A treatment.*

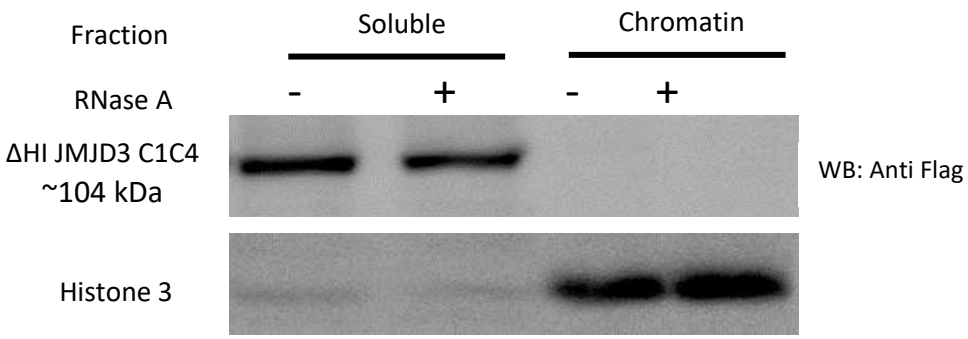
C-terminus perturbations result in loss of RNA binding and subsequent chromatin binding

The C-terminus of JMJD3 contains four cysteines (Cys_{1575}, Cys_{1578}, Cys_{1602}, and Cys_{1605}) that bind to Zn^{2+} ions. Zn^{2+} ion binding folds the C-terminus into a zinc finger-like structure. As previously stated, proteins with zinc finger-like domains have been shown to interact with both DNA and RNA nucleic acids. Thus, to further validate the RNA binding ability of this region, we mutated specific cysteines that bind to Zn^{2+} to alter the structure. We developed two mutants, each with two cysteines altered. Cys_{1575} and cys_{1578} were mutated in one construct, whereas cys_{1575} and cys_{1605} were mutated in the other. These mutations were created in both ΔFI-JMJD3 (ΔFI-JMJD3C1C2; ΔFI-JMJD3C1C4) and ΔHI-JMJD3 (ΔHI-JMJD3C1C2; ΔHI-JMJD3C1C4). Thus, we had four different constructs with these two mutations: ΔHI-JMJD3C1C2; ΔHI-JMJD3C1C4; ΔFI-JMJD3C1C2 and ΔFI-JMJD3C1C4. To investigate the effect of zinc finger disruption on the RNA binding of JMJD3, we performed UV-RIP of ΔFIJMJD3 WT; ΔFI-JMJD3C1C2, and ΔFI-JMJD3C1C4. We observed that the ΔFI-JMJD3 WT binds to eRNAs transcribed from the sense eRNAs of E8 and E17 enhancers but the cysteine mutants fail to interact with these eRNAs (Figure 3.7A-B). These findings demonstrate that JMJD3 can bind to eRNAs without its N-terminus and that cysteine mutations at the C-terminus result in RNA binding loss. We went on to investigate the impact of the loss of JMJD3 and RNA interaction on its chromatin binding. We isolated nucleoplasmic and chromatin-bound fractions of ΔHI-JMJD3C1C4 in the presence and absence of RNase A. Interestingly, we observed that ΔHI-JMJD3C1C4 does not bind to chromatin in both the presence or absence of RNase A (Figure 3.7C). This implies

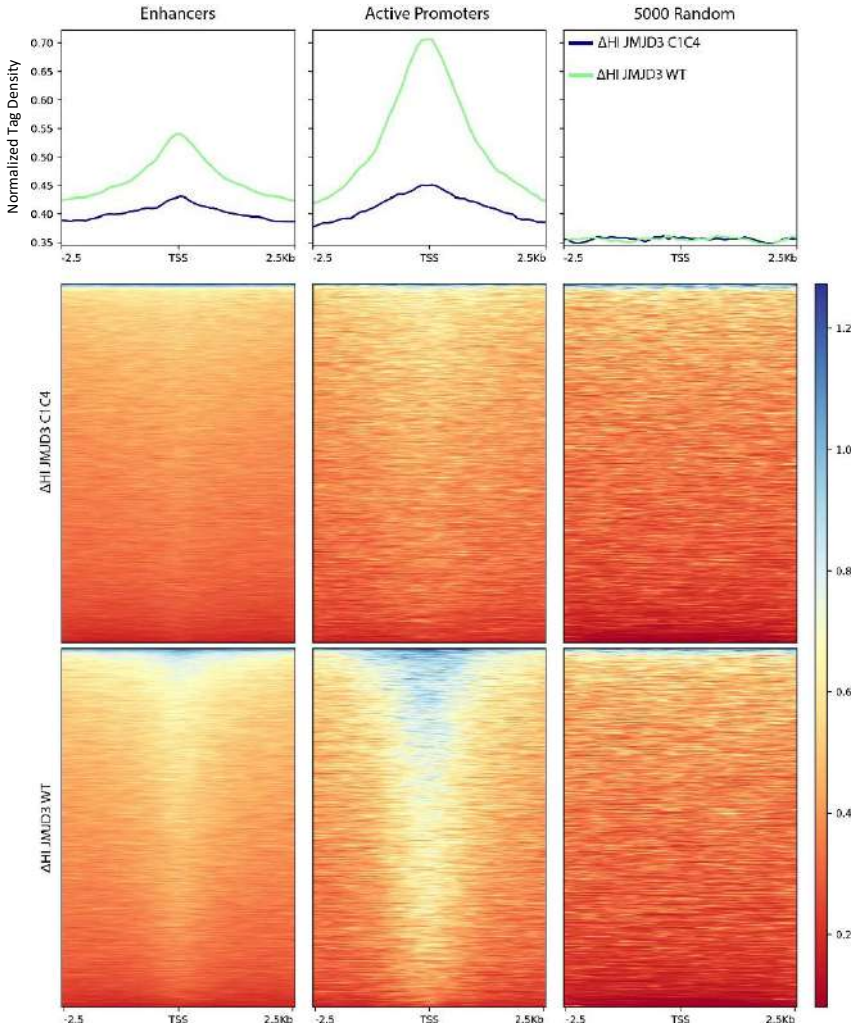
that zinc finger integrity is necessary for RNA binding and that RNA binding may be required for either JMJD3 recruitment or stabilization on chromatin. To validate these findings, we performed ChIP-seq on one of the mutants, ΔHI-JMJD3C1C4. The ChIP-seq data revealed that ΔHI-JMJD3C1C4 loses its chromatin binding compared to ΔHI-JMJD3 WT (Figure 3.7D). Even though ΔHI-JMJD3C1C4 loses its binding profoundly on chromatin but some enrichment on promoters and enhancers is retained. These could be the sites where ΔHI-JMJD3 binds without the assistance of RNA. We also performed ChIP-seq for ΔHIΔCT-JMJD3, lacking the C-terminus completely. Interestingly, this variant also fails to bind chromatin, suggesting that JMJD3 C-terminus is required for it to bind chromatin(Figure 3.7E). These data validate our prior findings and imply that the C-terminal domain is essential for RNA binding, which may be important for JMJD3 chromatin binding at most of the sites.



C


Fraction
Soluble
Chromatin
RNase A
-
+
-
+
ΔHI JMJD3 C1C4
~104 kDa
WB: Anti Flag
Histone 3

D


Enhancers
Active Promoters
5000 Random
ΔHI JMJD3 C1C4
ΔHI JMJD3 WT
Normalized Tag Density
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
-2.5
TSS
2.5Kb
ΔHI JMJD3 C1C4
ΔHI JMJD3 WT
1.2
1.0
0.8
0.6
0.4
0.2

E

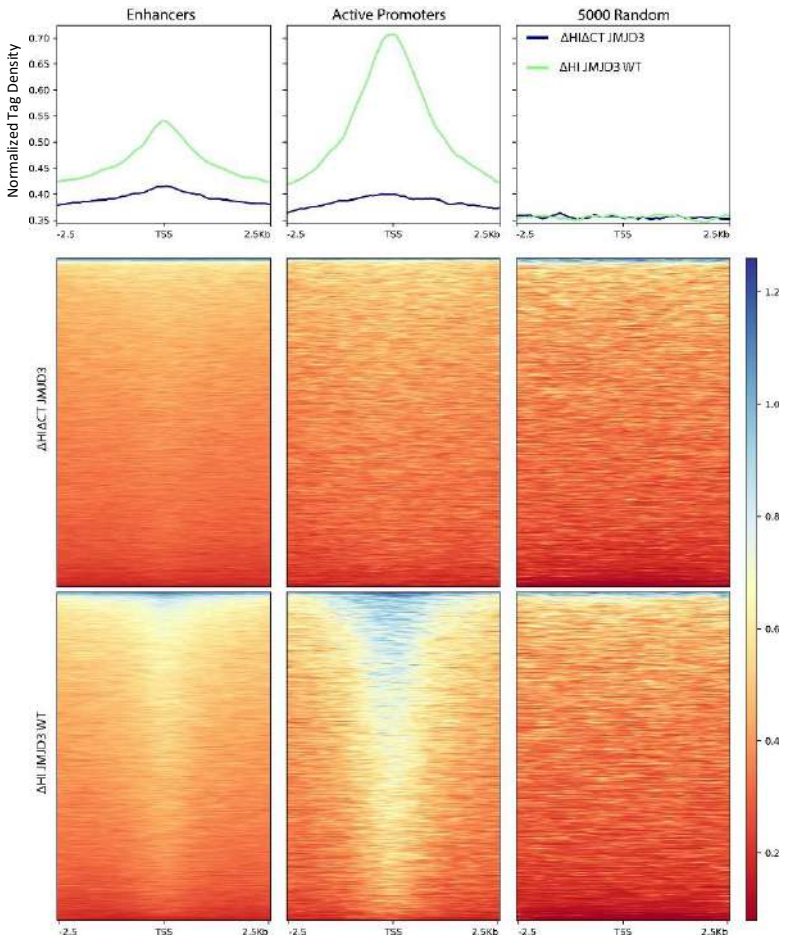


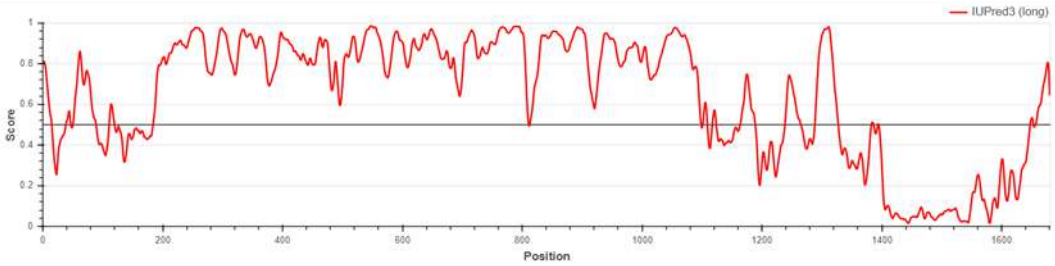
Figure 3.7: C-terminus cysteine mutations result in loss of RNA binding and subsequent chromatin binding. ***A-B)*** *UV-RIP qPCR plots show that ΔFI-JMJD3 interacts with E8 S eRNA and E17 S eRNA. However, ΔFI-JMJD3C1C2 and ΔFI-JMJD3C1C4 have mutations in cysteines that fail to interact with eRNA suggesting the importance of the C-terminus in RNA binding of JMJD3.* **C)** *Fractionation of ΔHI-JMJD3 with cysteines mutated shows loss of chromatin binding. ΔHI-JMJD3C1C4 is present only in the soluble fraction and not in the chromatin fraction.* ***D)*** *ChIP-seq heatmap showing the genome-wide binding of ΔHI-JMJD3 C1C2. ΔHI-JMJD3C1C2 shows some binding on promoters and enhancers, however, ΔHI-JMJD3C1C2 majorly shows a loss of binding compared to ΔHI-JMJD3. The sites that retain binding could be the sites where ΔHI-JMJD3 binds in RNA independent manner.* ***E)*** *ChIP-seq heatmap showing the genome-wide binding of ΔHIΔCT-JMJD3. Removal of the C-terminus results in substantial loss of binding suggesting the importance of RNA in chromatin binding.*

The uncharacterized N-terminal is mostly an intrinsically disordered region (IDR) that dictates the chromatin binding specificity of JMJD3

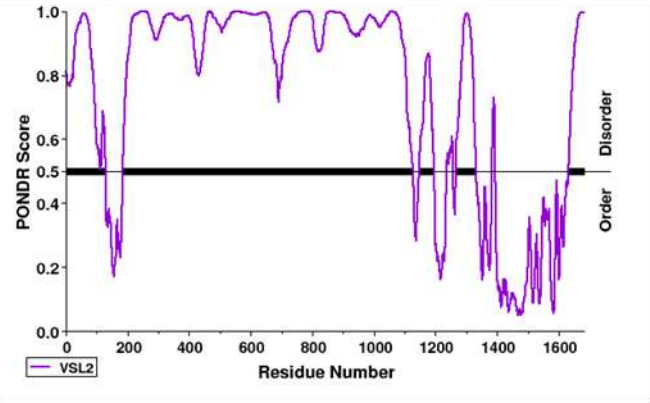
Our data revealed that the uncharacterized C-terminus of JMJD3 is involved in RNA binding. We further went ahead to understand the role of JMJD3's N-terminal in its chromatin binding. When we analyzed the amino acid sequence of JMJD3 with IUPRED3 and PONDR, these tools revealed that the N-terminal of JMJD3 is a highly disordered region (Figure 3.8A and 3.8B) with exception of a stretch containing few amino acids. Moreover, the alphafold-predicted structure also revealed that the N-terminus of JMJD3 is highly disordered (Figure 3.8C). The N-terminal of JMJD3 was of interest to us for a variety of reasons. 1. It is an intrinsically disordered region (IDR). 2. It makes up most of the JMJD3 protein. 3. Apart from carrying DNA binding domains, transcription factors are enriched with IDRs. IDRs have previously been found to influence the specificity of transcription factors. IDR-mediated chromatin binding does not require any other domain and occurs through weak interactions with certain transcription factors (Brodsky et al. 2020). Furthermore, a recent study has demonstrated the importance of this disordered region in the phase separation of JMJD3 (Vicioso-Mantis et al. 2022). To test the role of N-terminal IDR in the chromatin binding of JMJD3, we performed a ChIP-seq of ΔHI-JMJD3 in HeLa cells. Interestingly, the ΔHI-JMJD3 still showed binding to chromatin. As we examined the binding sites in the genome, we discovered that the N-terminal deletion caused the re-distribution of its binding in the genome (Figure 3.8D). Unexpectedly, ΔHI-JMJD3 targets switched from enhancers to promoters when compared to JMJD3 FL binding sites (Figure 3.8D). This suggests that the first half of JMJD3's N-terminal is important for its binding to specific targets, and its deletion leads to the altered binding. Further examination of JMJD3-bound sites revealed that JMJD3 binds to enhancers and promoters that are enriched with active histone marks. JMJD3-bound regions, for example, have high levels of H3K27ac, H3K4me1, and H3K4me3. Moreover, JMJD3-bound regions show very less enrichment for H3K27me3 (Figure 3.8E). Since ΔHI-JMJD3 showed binding on chromatin, we performed fractionation to further validate the ChIP-seq data and check if its binding on chromatin is RNA-dependent like JMJD3 FL. We observed that indeed the ΔHI-JMJD3 is present in both nucleoplasmic and chromatin fraction. The chromatin binding of ΔHI-JMJD3 is dependent on RNA as is for JMJD3 FL. (Figure 3.8F). To better understand the role of JMJD3's N-terminal in

its chromatin binding, we went ahead and performed ChIP-seq for ΔFI-JMJD3, which lacks the complete IDR. Analysis of ChIP-seq data revealed that JMJD3 binds neither enhancers nor promoters when its N-terminal IDR is completely deleted (Figure 3.8G). Thus, the data suggest that the N-terminal IDR of JMJD3 is indispensable for its binding to regulatory elements such as promoters and enhancers. Furthermore, the removal of the half IDR though does not affect the chromatin binding but the target specificity is altered.

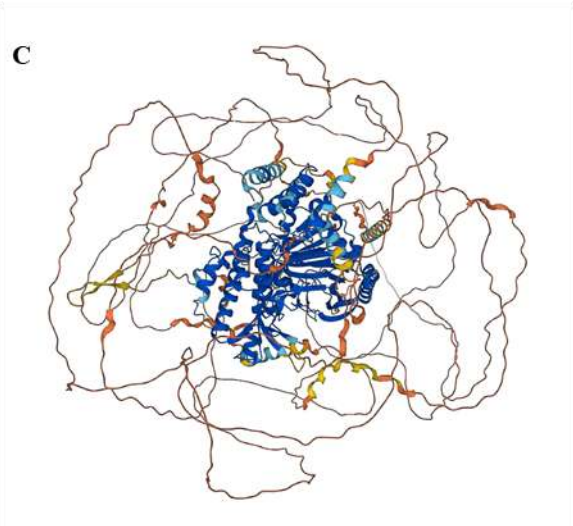
A



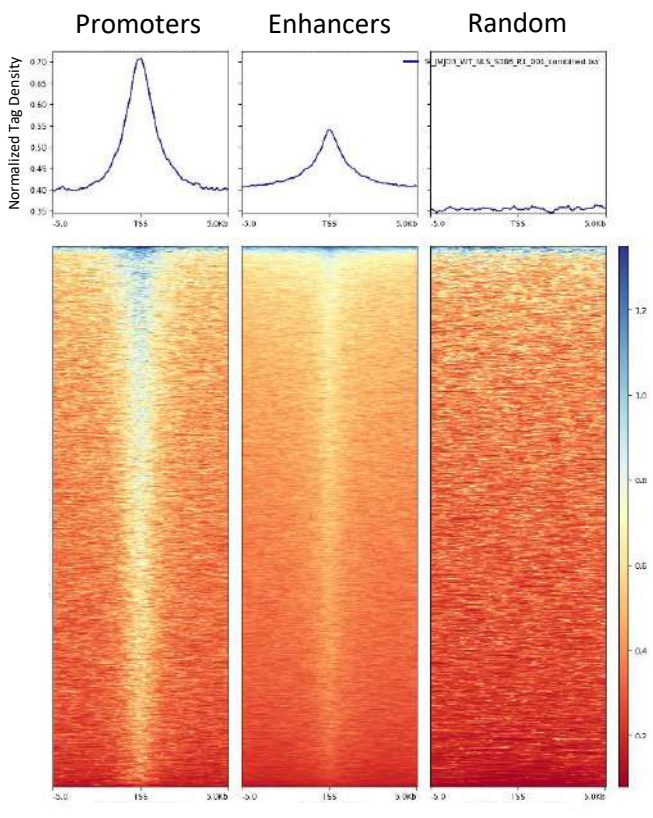
B



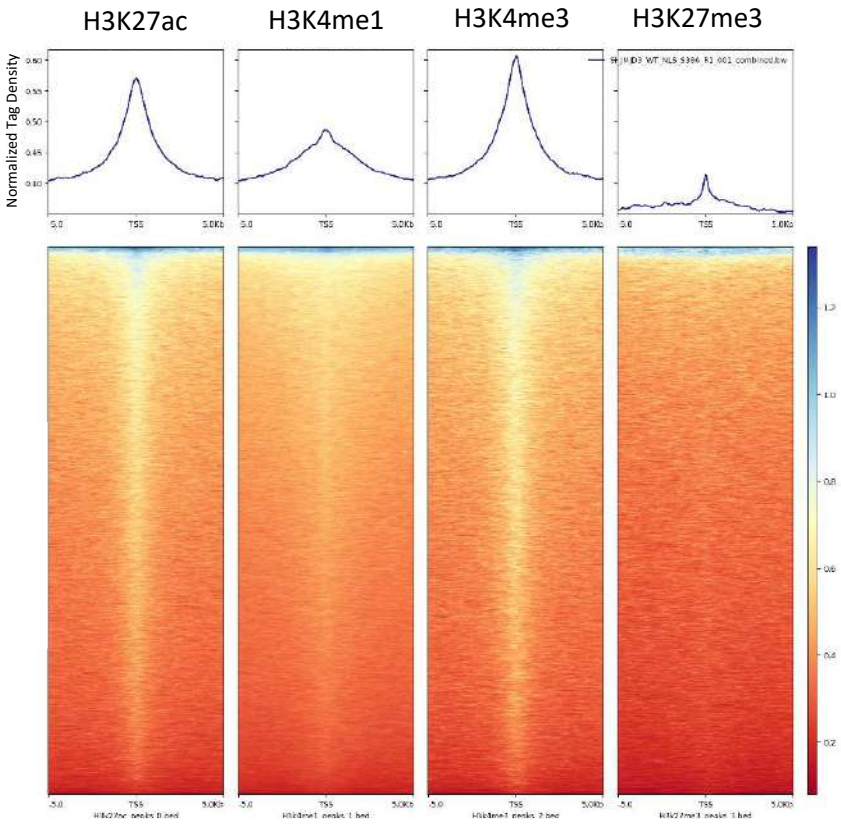
C



D


Promoters
Enhancers
Random
Normalized Tag Density
TSS

E


H3K27ac
H3K4me1
H3K4me3
H3K27me3
Normalized Tag Density
TSS

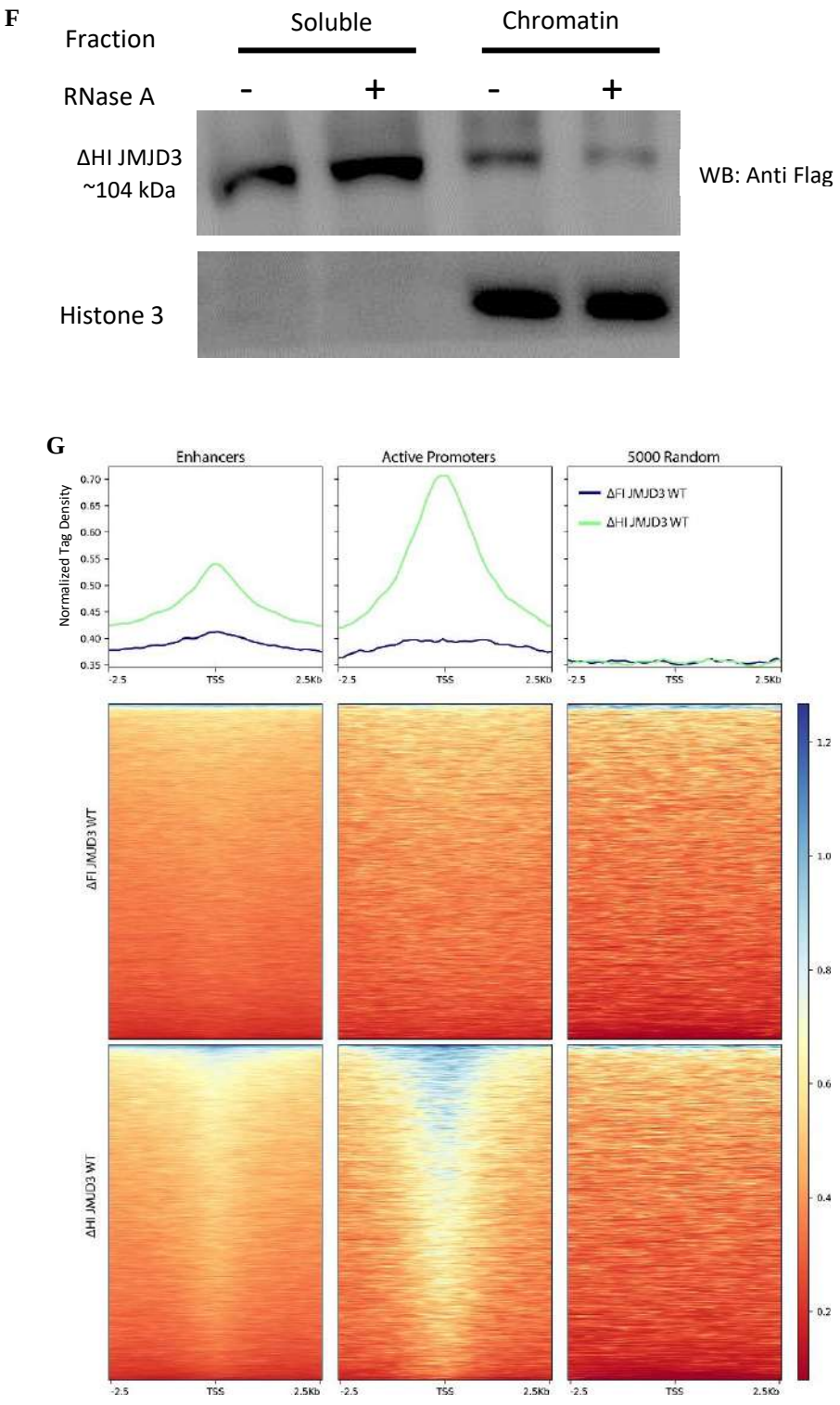


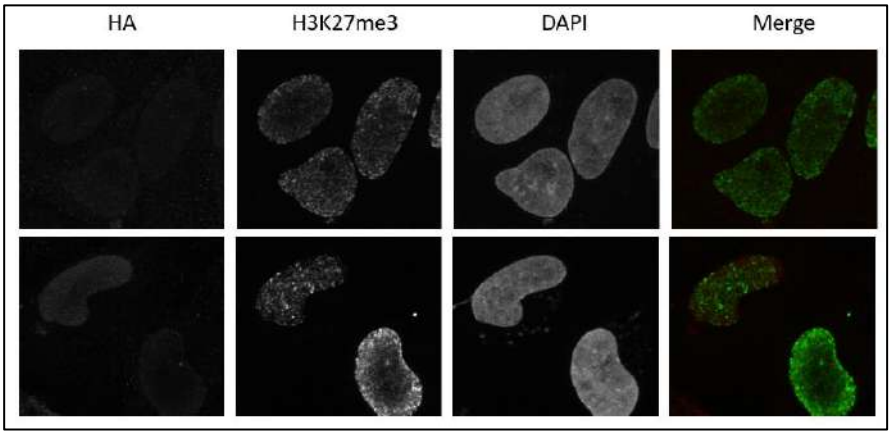
Figure 3.8: Perturbation of IDR affects the chromatin binding of JMJD3. ***A)*** *IUPRED3 plot showing that the JMJD3 N-terminal is highly disordered.* ***B)*** *PONDR plot showing that the JMJD3 N-terminal is highly disordered.* ***C)*** *Alphafold predicted structure of JMJD3 also reveals that the N-terminal of JMJD3 is an IDR.* ***D)*** *Heat map showing the binding of ΔHI-JMJD3 genome-wide. ΔHI-JMJD3 binds to both promoters*

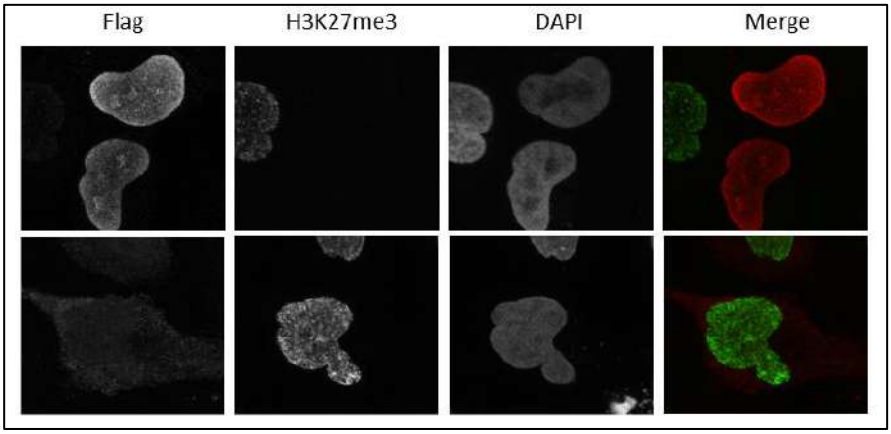
and enhancers. However, it shows more binding on active promoters than enhancers. ***E)*** *Heat maps showing ΔHI-JMJD3 bound sites are enriched with active marks like H3K27ac, H3K4me1, and H3K4me3 while showing little enrichment for a repressive mark like H3K27me3.* ***F)*** *Western blot showing that ΔHI-JMJD3 is present in both soluble and chromatin-bound fractions. ΔHI-JMJD3 loses its chromatin binding upon RNA degradation suggesting that ΔHI-JMJD3 binding on chromatin is dependent like JMJD3 FL.* ***G)*** *Heat map showing the binding of ΔFI-JMJD3 genome-wide. Removal of full IDR from JMJD3 significantly affects its binding on chromatin. ΔFI-JMJD3 shows very little binding on enhancers and promoters.*

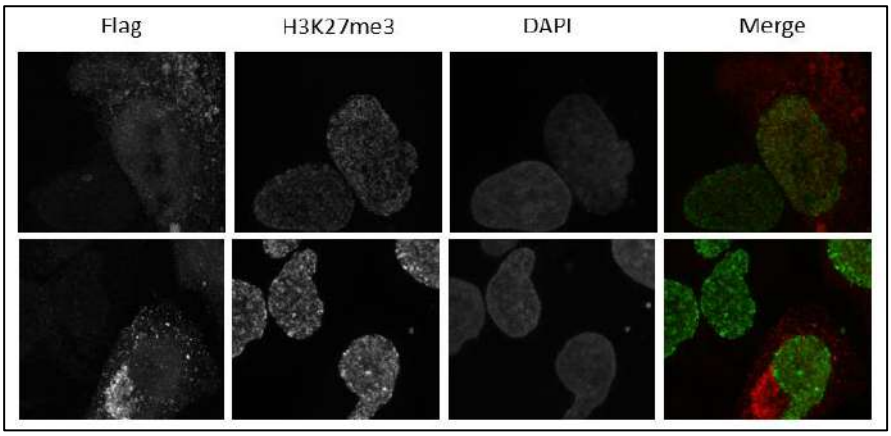
IDR and C-terminal perturbation affects the catalytic activity of JMJD3

Since the well-characterized JmjC domain of JMJD3 removes H3K27me3, we wanted to test if altering the N-terminal IDR and C-terminus of the JMJD3 affects its catalytic activity in vivo. Towards that, we performed coimmunostaining of different truncated versions of JMJD3 and H3K27me3. We observed that H3K27me3 was present more on the nuclear periphery than at the core of the nucleus. We discovered that JMJD3 FL and ΔHI-JMJD3 are catalytically active because we observed a significant decrease in H3K27me3 staining (Figure 3.9A, 3.9B, 3.9C, and 3.9E). Furthermore, we noticed more loss of H3K27me3 with ΔHI-JMJD3 than JMJD3 FL. Conversely, ΔFI-JMJD3 could not reduce the levels of H3K27me3, implying that IDR is essential for JMJD3 to perform its catalytic activity (Figure 3.9E). Similar results were obtained with different cysteine mutant constructs (Figure 3.9D and 3.9F). The loss of catalytic activity could be due to JMJD3 binding on chromatin being disrupted by complete IDR ablation or RNA binding perturbations. JMJD3 must bind to H3K27me3 sites for demethylation to occur. Because these proteins cannot bind chromatin, no methylation changes are detected with these constructs. These findings support our prior findings that the IDR and RNA binding C-terminus are essential for JMJD3 to bind to chromatin and therefore its function as demethylase.

.


A
HA &
H3K27me3
B
JMJD3 FL &
H3K27me3
HA
H3K27me3
DAPI
Merge


C
ΔHI-JMJD3 &
H3K27me3
D
ΔHI-JMJD3C1C4 &
H3K27me3
Flag
H3K27me3
DAPI
Merge


E
ΔFI-JMJD3 &
H3K27me3
F
ΔFI-JMJD3C1C4 &
H3K27me3
Flag
H3K27me3
DAPI
Merge

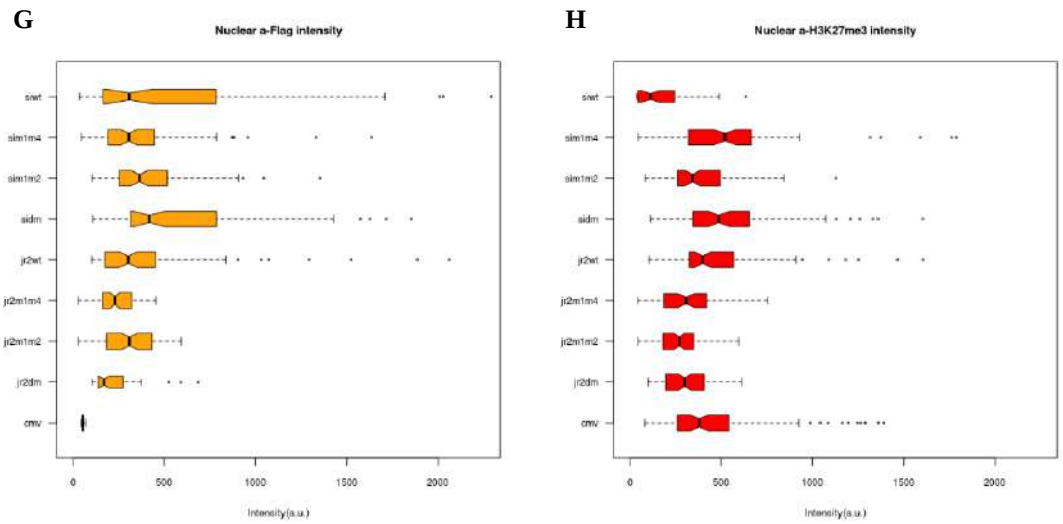
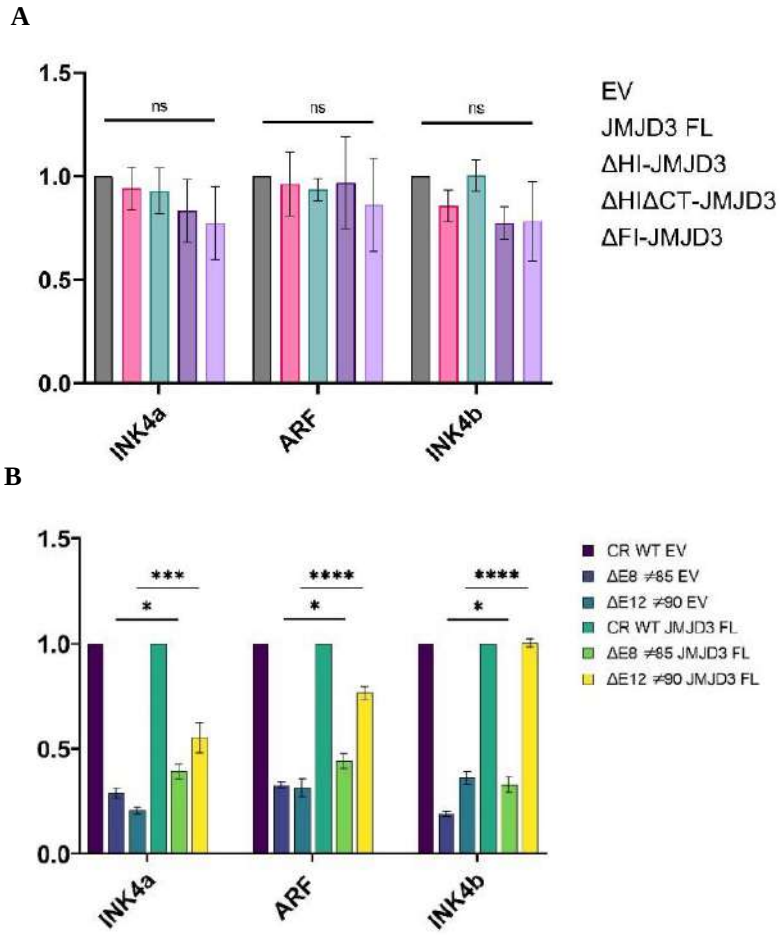


Figure 3.9: IDR and C-terminal perturbation affect the catalytic activity of JMJD3. ***A-F)*** *Representative cells showing co-immunostaining of* ***A)*** *HA Empty Vector and H3K27me3.* ***B)*** *JMJD3 FL and H3K27me3.* ***C)*** *ΔHI-JMJD3 and H3K27me3.* ***D)*** *ΔHI-JMJD3C1C4 and H3K27me3.* ***E)*** *ΔFI-JMJD3 and H3K27me3.* ***E)*** *ΔFI-JMJD3C1C4 and H3K27me3.* ***G)*** *Box plots showing the nuclear intensity of different JMJD3 truncations. The nuclear intensity was calculated from around 70 cells.* ***H)*** *Box plots showing the nuclear intensity of H3K27me3 in cells with expression of different JMJD3 truncations. The nuclear intensity was calculated from around 70 cells.*

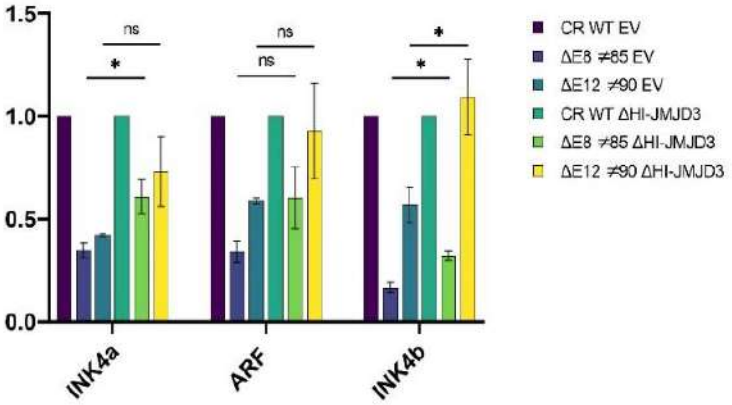
JMJD3 FL and ΔHI-JMJD3 are capable of restoring gene expression in enhancer knockout lines

We know that only JMJD3 FL and ΔHI-JMJD3 can eliminate the trimethyl mark in HeLa cells. Each of these proteins contains IDRs, implying that JMJD3 requires IDR for trimethyl mark removal. Next, we wanted to explore if these truncations of JMJD3 may influence gene transcription. To that end, we transfected HeLa cells with these JMJD3 truncations and tested the gene expression using RT qPCRs. Surprisingly, none of the proteins, including JMJD3 FL, had any effect on gene expression (Figure 3.10A). We reasoned that because this locus is already very active in HeLa cells and does not exhibit H3K27me3 enrichment, therefore, the overexpression of these proteins would not have any effect on gene expression. However, in the HeLa clones where key enhancers are deleted, the EZH2 is loaded on promoters of *INK4/ARF* locus resulting

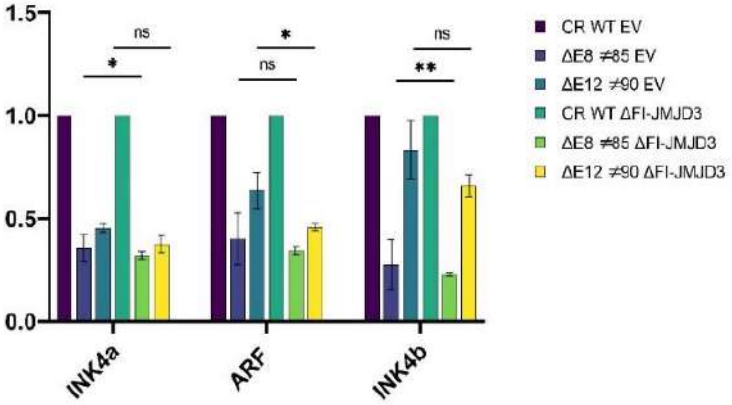
in the very low expression of these genes. Thus, we overexpressed these constructs in E8 and E12 enhancer knockout lines and observed that both, JMJD3 FL and ΔHI-JMJD3 were able to rescue gene expression but not the other constructs (Figure 3.10B-3.10E). This is consistent with the staining and ChIP-seq results, which show that only constructs with chromatin binding capabilities are functional. PRC2 binds to *INK4/ARF* promoters in the absence of E8, E12, and E17 enhancers. We wanted to see if the restoration of gene expression was specific to JMJD3 or if inhibiting the PRC2 complex may also result in restoration. To investigate this, we used GSK343 to inhibit the catalytic activity of PRC2 in enhancer knockout lines. Surprisingly, gene expression was not restored in the enhancer knockout line when PRC2 was inhibited (Figure 3.10F-3.10H). These findings demonstrate that simply inhibiting PRC2 is insufficient to stimulate gene expression. These genes, however, require JMJD3 to be activated.



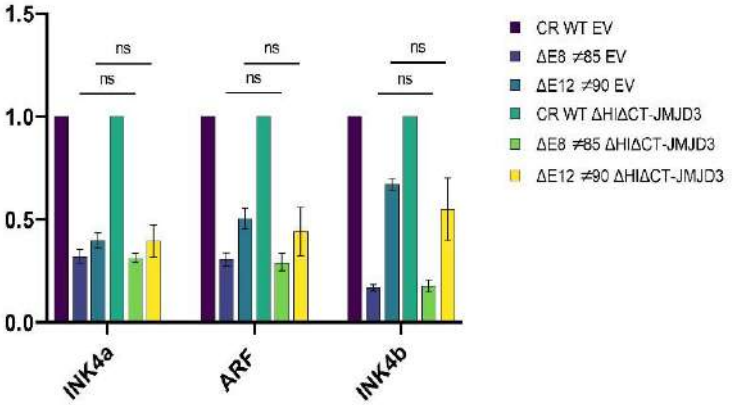
C


1.5
1.0
0.5
0.0
ns
*
ns
ns
*
*
INK4a
ARF
INK4b
CR WT EV
ΔE8 ≠85 EV
ΔE12 ≠90 EV
CR WT ΔHI-JMJD3
ΔE8 ≠85 ΔHI-JMJD3
ΔE12 ≠90 ΔHI-JMJD3

D


1.5
1.0
0.5
0.0
ns
*
*
ns
ns
**
INK4a
ARF
INK4b
CR WT EV
ΔE8 ≠85 EV
ΔE12 ≠90 EV
CR WT ΔFI-JMJD3
ΔE8 ≠85 ΔFI-JMJD3
ΔE12 ≠90 ΔFI-JMJD3

E


1.5
1.0
0.5
0.0
ns
ns
ns
ns
ns
ns
INK4a
ARF
INK4b
CR WT EV
ΔE8 ≠85 EV
ΔE12 ≠90 EV
CR WT ΔHIΔCT-JMJD3
ΔE8 ≠85 ΔHIΔCT-JMJD3
ΔE12 ≠90 ΔHIΔCT-JMJD3

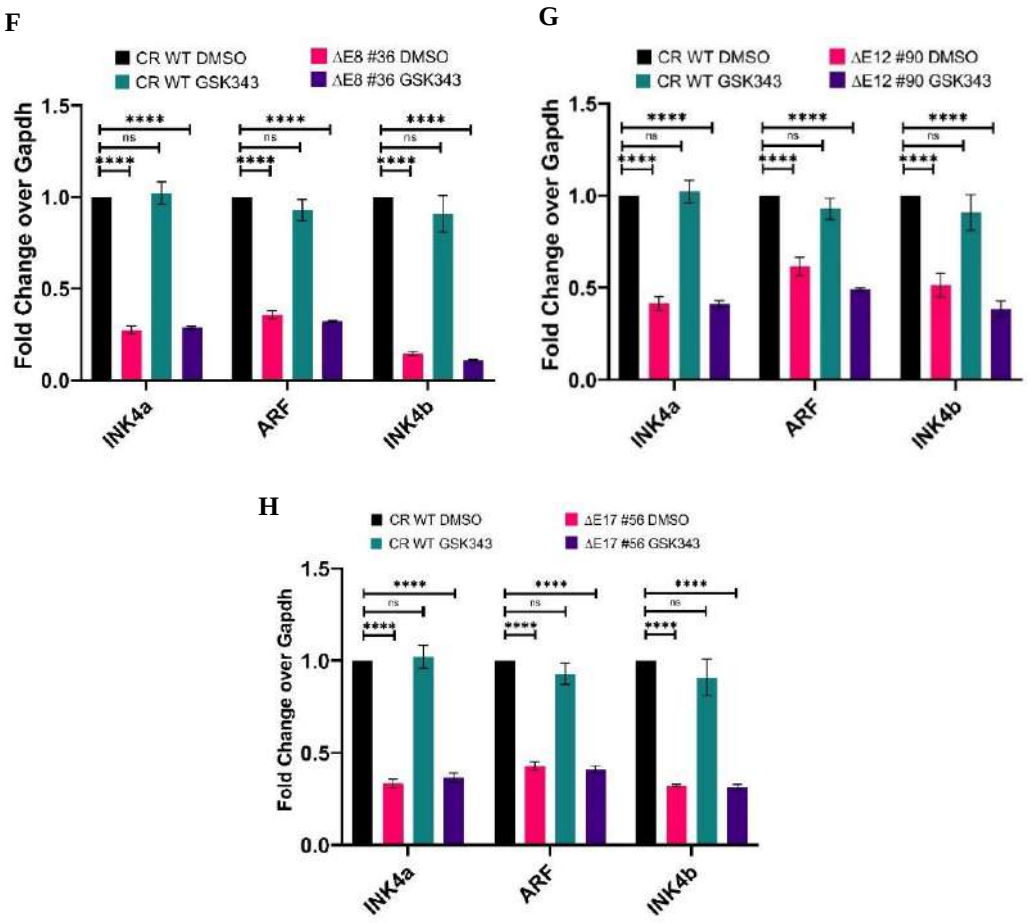


Figure 3.10: JMJD3 FL and ΔHI-JMJD3 are capable of restoring gene expression in enhancer knockout lines. ***A)*** *Gene expression in HeLa with overexpression of different truncations of JMJD3.* ***B-E)*** *Gene expression in E8 and E12 enhancer knockout lines with overexpression of different truncations of JMJD3.* ***B)*** *JMJD3 FL overexpression* ***C)*** *ΔHI-JMJD3 overexpression* ***D)*** *ΔFI-JMJD3 overexpression and* ***E)*** *ΔHIΔCT-JMJD3 overexpression.* ***F-H)*** *Gene expression in* ***F)*** *E8,* ***G)*** *E12, and* ***H)*** *E17 enhancer knockout lines upon EZH2 catalytic inhibition. The catalytic activity of EZH2 was inhibited by treating cells with 10μM GSK343 for 24 hours.*

Chapter 4: Materials and Methods

Cell Culture

HeLa and HEK293FT cells were obtained from the American Type Culture Collection (ATCC). Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptomycin was used to culture both cell lines. The cells were kept at 37°C in a humidified atmosphere with 5% CO2. Both the cell lines were passaged every third day.

Antibodies

Anti-H3K27ac (Abcam, ab4729); Anti Pol II (Santa Cruz, sc-899); Anti-EZH2 (Active Motif, 39875); Anti-p14 (Santa Cruz, sc-53639); Anti-p16 (Cell Signaling, 4824S); Anti-Gapdh (Santa Cruz, sc-32233); Anti JMJD3 (ab38113); Anti Flag (F7425); Anti HA (ab9110); Anti-H3K27me3 (CST #9733); Anti-Flag (CST #8146); Anti HA (CST #2367).

Circular chromosome conformation capture (4C)

4C was performed as per the protocol described in (van de Werken et al. 2012) with minor variations. HeLa cells were fixed with 1.5% fresh formaldehyde for 10 mins at room temperature and quenched with glycine (125 mM) for 5 mins. The cells were washed thrice with ice-cold PBS and scraped, pelleted, and stored at −80ºC. Lysis buffer [Tris-Cl pH 8.0 (10 mM), NaCl (10 mM), NP-40 (0.2%), PIC (1X)] was added to the pellets and homogenized by Dounce homogenizer (20 strokes with pestle A followed by 20 strokes with pestle B). HindIII (400U, NEB) was used for the 3C digestion and T4 DNA ligase was used for ligation along with ligation mix [1% Triton X-100, 1X Ligation buffer [Tris-Cl pH 7.5 (50 mM), $MgCl_2$ (10 mM), DTT (10 mM), BSA (0.0105mg/ml), ATP (0.105 mM)]. The ligated samples were purified by PCI, followed by ethanol precipitation. The pellet was dissolved in 1X TE (pH 8.0) to obtain the 3C library. DpnII (50U, NEB) was used for 4C digestion and the samples were ligated, purified, and precipitated similar to the 3C library to obtain the 4C library. The 4C library was treated with RNase A to remove any trace of contaminating RNA and

purified by the QIAquick PCR purification kit. The library was then subjected to PCR using the oligos designed for the CDKN2A viewpoint (Table S1). The PCR amplicon was purified using the same kit and subjected to next-generation sequencing with Illumina HiSeq2500 using 50bp single-end reads. The number of reads in each replicate is mentioned in the appendices.

Chromatin immunoprecipitation (ChIP)

Cells were crosslinked with 1% formaldehyde (Sigma-F8775) at room temperature for 10 mins with constant shaking. Glycine was added to a final concentration of 125mM to quench the formaldehyde for 5 mins at room temperature. Cells were washed thrice with 1X ice-cold PBS. Cells were scraped in 3 mL of 1X PBS and pelleted down at 3K rpm for 5 mins at 4ºC. Cells were gently resuspended in L1A Buffer [10 mM HEPES/KOH pH 7.9, 85 mM KCl, 1 mM EDTA pH 8.0]. Nuclei were isolated by adding L1B buffer [10 mM HEPES/KOH pH 7.9, 85 mM KCl, 1 mM EDTA pH 8.0, 1% NP40] to the cells, resuspended, and incubated on ice for 10 minutes. Nuclei were obtained by centrifuging at 3.5K rpm for 5 mins. Nuclear lysis buffer (L2) [50 mM Tris-HCl pH 7.4, 1% SDS, 10 mM EDTA pH 8.0] supplemented with 1X PIC was added to the nuclear pellet, resuspended, and incubated on ice for 10 mins. Samples were sonicated using Diagenode biorupter with a setting of 30 Sec ON and 30 Sec OFF for 25 cycles. The cell lysate was cleared by centrifuging samples at 12K rpm for 12 mins. 100 µg of sheared chromatin was taken for each IP. Sheared chromatin was diluted by adding Dilution Buffer (DB) [20 mM Tris-HCl pH 7.4, 100 mM NaCl, 2 mM EDTA pH 8.0, 0.5% Triton X-100] supplemented with 1X PIC in 1:1.5 (1 volume of sheared chromatin and 1.5 volumes of Dilution Buffer) ratio. 5% of diluted chromatin was set aside as Input. 1 µg of antibody was added to immunoprecipitate the DNA and incubated on a rocking platform overnight at 4ºC. Protein G Dynabeads (Invitrogen-140004D) were prepared by blocking in 1% BSA prepared in 1X PBS at 4ºC for 1 hour followed by washing with 1X PBS. Immunoprecipitated DNA was collected by adding 15 µl of BSA-blocked beads to each sample and incubated at 4ºC for 4 hours. Beads were collected, the supernatant was discarded and 600 µl of Wash Buffer I [20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% SDS, 2 mM EDTA pH 8.0, 1% Triton X-100] was added. Washings were carried out at 4ºC on a rocking platform. Washings were

sequentially repeated with Wash Buffer II [20 mM Tris-HCl pH 7.4, 500 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100], Wash Buffer III [10 mM Tris-HCl pH 7.4, 250 mM LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA (pH 8.0)] and 1X TE [10 mM Tris pH 8.0, 1 mM EDTA pH 8.0]. Immunoprecipitated DNA was eluted by adding 200 µl of Elution Buffer [100 mM $NaHCO_3$, 1% SDS] for 45 mins at 37ºC in a thermomixer with an rpm of 1200. Eluate was transferred to fresh tubes and 14 µl of 5M NaCl was added and kept overnight at 65ºC. Immunoprecipitated DNA was purified by phenol:chloroform: isoamyl alcohol (Ambion-AM9732) method, followed by ethanol precipitation. The final dried DNA pellet was dissolved in 100 µl of 1X TE. The ChIP data were plotted as the fold enrichment of percent input. First, percent inputs were calculated for beads and antibody, and then values obtained from beads were divided from antibody values to get the fold enrichment.

RNA Isolation and cDNA synthesis

After discarding the media, 1 mL of Trizol reagent (Invitrogen-15596018) was applied immediately to the adhered cells. Cells were lysed by pipetting and transferred to 1.5 ml nuclease-free Eppendorf tubes. 200 µl of chloroform was added to samples, invert mixed several times and incubated at room temperature for 2 mins. The samples were then centrifuged at 12K rpm for 12 mins at 4ºC. The aqueous phase was carefully collected and transferred to fresh 1.5 ml tubes. 0.7 volumes of isopropanol were added to samples, vortexed, and incubated at room temperature for 10 mins to precipitate the RNA. The samples were centrifuged at 12K rpm for 12 mins at 4ºC, the supernatant was discarded without disturbing the pellet. The pellet obtained was washed twice with 75% ethanol prepared in nuclease-free water. Pellet was allowed to dry for around 20 mins and dissolved in RNase-free water. RNA obtained was treated with ezDNase (Invitrogen-117660) to remove the traces of contaminating DNA. 1 µg of RNA was taken for the cDNA synthesis. cDNA was synthesized with Superscript IV (Invitrogen-18091050) kit.

10 µl of cDNA reaction was set up and the reaction was carried out in two steps. In the first step, the RNA primer mix was prepared by adding the following components to a sterile RNase-free tube.

0.5 µl Random Hexamers;

0.5 µl 10mM dNTPs;

1 µg Isolated RNA;

DEPC-treated water up to 5.5 µl

The components were mixed properly and briefly centrifuged. The RNA primer mix was incubated at 65ºC for 5 mins, followed by 1 min incubation on ice.

In the second step, the RT reaction mix was prepared by adding the following components to a sterile RNase-free tube.

2 µl superscript buffer;

0.5 µl 100mM DTT;

0.5 µl Ribonuclease Inhibitor;

0.5 µl Superscript IV Reverse Transcriptase (200 U/µL)

The contents were mixed properly and briefly centrifuged. The prepared RT mix was added to annealed RNA primer mix. The reaction components were mixed properly and incubated at 50ºC for 10 mins. The cDNA synthesis reaction was stopped by incubating it at 80ºC for 10 mins. The volume of the samples was made up to 100 µls by diluting with nuclease-free water. 4 µl of the diluted sample was taken for RT PCR amplification.

Designing and cloning of gRNAs

CRISPR Cas9-mediated deletions were performed as per the protocol described in (Farooq and Notani 2021). gRNAs were designed with https://zlab.bio/guide-design-resources and http://crispor.tefor.net tools. gRNAs were selected based on the highest score with the least number of off-targets. All the gRNAs were cloned in pgRNA humanized vector (#44248) a gift from Stanley Qi between BstX1 and Xho1 restriction sites.

CRISPR-Cas9 mediated deletion

CRISPR-Cas9-mediated deletions were performed as per the protocol described in (Farooq and Notani 2021). gRNAs were co-transduced with lenti-Cas9 vector (#52962) a gift from Feng Zhang. Cells were selected for pgRNA humanized vector with puromycin (3µg/ml) for 48 hours. The remaining cells were seeded in a 96-well plate such that each well gets a single cell. Wells with single cells were identified under a microscope and allowed to grow till the colonies appeared. The media was changed after every fifth day. The cells were trypsinized and half of the cells were taken for the surveyor assay. The surveyor assay was carried out by PCR-based method.

CRISPRi

For carrying out CRISPRi, gRNAs were designed to target the core of the enhancers as described above. For each enhancer, two gRNAs were designed. Lentiviruses were made carrying dCas9-KRAB (#99372), a gift from Kristen Brennand, and enhancer-specific gRNAs. At the time of transduction, cells were around 75% confluent. Viral soup supplemented with 8 µg/ml of polybrene was added to cells. The infection was stopped after 16 hours of transduction. Cells carrying the vectors were selected with 3 µg/ml of puromycin for 72 hours.

Lentiviral transduction

Lentiviral transductions were carried out as per the protocol described in (Farooq and Notani 2021). Briefly, HEK293FT cells were seeded in culture dishes coated with poly D lysine. Transfection of lentiviral packaging plasmid like pCMV-VSV-G (#8454), a gift from Bob Weinberg's lab along with the plasmid of interest was carried out by Lipofectamine 2000. The media was changed after 6 hours. The viral soup was collected after 48 hours and 72 hours, pooled together, filtered, and finally added to cells along with 8ug/ml of polybrene. The infection was stopped after 16 hours of transduction.

siRNA transfection

siRNAs SMARTpools specifically targeting JMJD3 mRNA (LQ-023013-01-0005) and scrambled siRNA (D-001810-10-05) were purchased from GE Dharmacon. Lipofectamine 2000 (Invitrogen) was used for siRNA transfections as per the manufacturer's recommendations. Two rounds of transfections were done to carry out the knockdown.

shRNA designing and cloning

shRNAs were designed and cloned by following the protocol given at http://www.addgene.org/protocols/plko/shRNAs. Briefly, shRNAs were designed using the https://portals.broadinstitute.org/gpp/public/seq/search tool and cloned in pLKO.1 TRC cloning vector. Three shRNA were designed for each target eRNA to increase the efficiency of RNAi. To minimize the degradation of off-target mRNAs, UCSD's BLAT program was used. All the shRNAs were cloned between EcoRI and AgeI restriction sites. To get the shRNAs sequenced, the clones were digested with the XhoI enzyme that is present in the loop of the shRNAs. Linearization was done to efficiently get the shRNAs sequenced as the hairpin formation was leading to problems in sequencing. The shRNA sequences are provided in appendices.

β-galactosidase staining

Cells were stained for senescence-associated β-Galactosidase activity according to the manufacturer's protocol (Senescence β-Galactosidase Staining from Cell Signaling #9860). Briefly, culture media was discarded and cells were washed once with 1X PBS. 1 mL of 1X Fixative solution was added to each well and allowed to fix for 15 mins at room temperature. Fixed cells were washed twice with 1X PBS. To stain the cells, 1 mL of β-Galactosidase Staining Solution was added to each well. Plates were carefully sealed with parafilm to avoid evaporation and incubated at 37ºC for 48 hours in a dry incubator devoid of CO_2. Images were taken with Olympus IX73. ImageJ was used to quantify the cells stained for senescence-associated β-Galactosidase activity.

Cell proliferation assay

Cell proliferation assay was performed according to the manufacturer's protocol (BrdU Cell Proliferation Assay Kit from cell signaling #6813). Briefly, 2500 cells were seeded in 3 wells (triplicates) of 96 well plate. Cells were incubated for 48 hours. 1X BrdU solution was added to each well and incubated for 24 hours. Growth media was removed and 100 µl of Fixing/Denaturing Solution was added to each well for 30 mins at room temperature. Post 30 mins fixing solution was removed and 100 µl of 1X detection antibody solution was added to each well for 1 hour at room temperature. The detection antibody solution was removed, followed by washing each well 3 times with 1X Wash Buffer. 100 µl of 1X HRP-conjugated secondary antibody solution was added for 30 mins at room temperature. The solution was removed followed by washing 3 times with 1X Wash Buffer. 100 µl of TMB substrate was added for around 20 mins and the reaction was stopped by adding 100 µl stop solution. Absorbance was measured at 450 nm.

Wound healing assay

Cells were grown to confluence in 6 well plates and two scratches per well were created. Cells were washed with DPBS to remove the cellular debris. After removing debris, images were taken and labeled as A0. Cells were allowed to migrate for 72 hours. The extent of migration was recorded every 24 hours and labeled as A1, A2, and A3. Wound closure (2 scratches per replicate) was measured with ImageJ where wound closure (%) = (wounded area after every 24 hours/wounded area at A0) X 100.

Colony formation assay

2000 cells each of CR WT and enhancer knockout lines were seeded in one well of 6 well plates. Cells were allowed to form colonies for 10 days. Culture media was discarded and cells were washed twice with 1X PBS. Colonies were fixed with absolute methanol for 10 mins followed by staining with 0.5% crystal violet (prepared in 25% Methanol) for 20 mins. Cells were washed with 1X PBS 4-5 times. Colonies containing at least 50 cells were counted using ImageJ.

Invitro Transcription Assay

Invitro transcription assay was performed as per the protocol described in (Jayani, Singh, and Notani 2017) with some modifications. Briefly, plasmids containing cloned enhancers were linearized with the appropriate restriction enzyme such that the cut is made right after the transcribing region. Digested plasmids were run on 1% agarose gel and the appropriate digested band was cut and purified using a gel extraction kit from Qiagen. Invitro transcription mix was prepared by adding the following components in a sterile RNase-free tube.

1 µg of linearized template plasmid

2 µl Biotin RNA labeling mix (10X)

4 µl Transcription buffer (5X)

2 µl DTT (100 mM)

2 µl appropriate RNA polymerase (20U) (T7 polymerase in this case)

x µl RNase-free water to make the final volume of 20 µl

The components were mixed by gently tapping the tube, centrifuged briefly, and incubated at 37ºC for 2 hours. 1 µl of ezDNase was added to digest the template plasmid DNA. The volume of the samples was brought to 100 µl by adding RNAse-free water. 1/10 volume of 3M sodium acetate (pH 5.2) and 2.5 volumes of absolute ethanol were added. The samples were mixed by vortexing and briefly centrifuged and incubated at -20ºC for 3 hours. The precipitated RNA was pelleted down by centrifuging at 12K rpm for 12 mins at 4ºC. The supernatant was removed and the pellet was washed with 75% ethanol. Pellet was air dried for 10 - 12 mins and dissolved in RNase-free water.

RNA Pulldown Assay

Invitro transcription assay was performed as per the protocol described in (Jayani, Singh, and Notani 2017) with some modifications. HeLa cells were grown to a confluency of 90 - 95%. 8 x 10^6 cells were harvested by discarding the media and washing the cells thrice with 1X PBS. The cells were scraped and pelleted at 3000 rpm for 5 mins at 4ºC. The pelleted cells were resuspended in 2 mL PBS and to this 2 mL

of nuclear isolation buffer (NIB) [40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 1.28M sucrose, 4% Triton X-100, 1mM PMSF, protease inhibitors and 20U/ml SUPERase inhibitor (Thermo Fisher Scientific, catalog AM2694)] was added. The contents were mixed gently and 6 mL of distilled water was added to it, gently mixed, and incubated on ice for 20 mins with gentle mixing intermittently. Nuclei were pelleted by centrifuging the samples at 4K rpm for 5 mins at 4ºC. The nuclei pellet was resuspended in 1 mL RNA immunoprecipitation (RIP) buffer: [25 mM Tri-HCl pH 7.4, 150 mM KCl, 0.5 mM DTT, 0.5% NP40, IX PIC and 20U/ml SUPERase inhibitor] and incubated on ice for 5 mins. The samples were transferred to sonication tubes and sonicated for 10 cycles with the setting of 30S ON 30S OFF in a Diagenode biorupter. The lysate was cleared by centrifuging the samples at 12K rpm for 12 mins at 4ºC. The nuclear lysate was equally divided into two tubes. RNA was folded in 50 µl of RNA structure buffer (RSB) [10 mM Tris-HCl pH 7.0, 100 mM KCl, 10 mM $MgCl_2$, 1X PIC, and 20U/ml SUPERase inhibitor] for 5 mins at room temperature. 1 µg of folded RNA of interest and control RNA was added to the tubes containing protein lysate. The RNA protein mix was incubated at 4ºC for 2 hours on a rocker. Dynabeads MyOne Streptavidin beads were prepared by washing the beads thrice with RIP buffer. 15 µl of Dynabeads MyOne Streptavidin beads were added to each tube and incubated at 4ºC for 1 hour on a rocker. Beads were washed thrice with RIP buffer for 5 mins. 2X SDS loading buffer was added to the beads and boiled at 98ºC for 15 mins. The denatured samples were used for western blotting for further steps.

Subcellular fractionation

Subcellular fractionation was performed as per the protocol described in (Caudron-Herger et al. 2019) with minor variations. Briefly, 6 x 10^6 cells were harvested and pelleted down at 3K rpm for 5 mins at 4ºC. The pelleted cells were gently resuspended in 200 µl of ice-cold Hypotonic Lysis Buffer (HLB) buffer [10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.3% (v/v) NP 40, 10% (v/v) glycerol], supplemented with 1X protease inhibitor cocktail. The samples were incubated on ice for 10 mins and vortexed briefly. The samples were then centrifuged at 4000 rpm for 5 mins at 4ºC to obtain the nuclei. The supernatant (cytoplasmic fraction) was transferred to a fresh tube and the NaCl concentration was adjusted to 140 mM by adding the remaining amount

of NaCl. The pellet containing nuclei was washed twice with 200 µl of HLB and centrifuged at 4K rpm for 5 mins at 4ºC. After washing, the nuclei were equally divided into two tubes and resuspended in 150 µL ice-cold Modified Wuarin-Schiebler (MWS) buffer [10 mM Tris-HCl (pH 7.5), 300 mM NaCl, 4 mM EDTA, 1 M urea, 1% (v/v) NP40, 1% (v/v) glycerol], supplemented with 1X protease inhibitor complex. The samples were incubated on ice for 15 mins with or without RNase A (30 µg/ml), vortexed, and centrifuged at 6K rpm for 5 mins at 4ºC. The supernatant (nucleoplasmic fraction) was transferred to a fresh tube. The chromatin pellet was washed once with MWS buffer and centrifuged at 6K rpm for 5 mins at 4ºC. 150 µl of Nuclei Lysis Buffer (NLB) [20 mM Tris-HCl (pH 7.5), 150 mM KCl, 3 mM $MgCl_2$, 0.3% (v/v) NP40, 10% (v/v) glycerol], supplemented with 1X PIC was added to the chromatin pellet and incubated on ice for 10 mins. The pellet was then sonicated in a Diagenode biorupter for 15 cycles with the settings of 30S ON and 30S OFF. All the samples were denatured by adding 1X SDS loading dye and incubating on a heat block at 98ºC for 15 mins.

Ultraviolet-RNA Immunoprecipitation (UV-RIP)

Ultraviolet-RNA Immunoprecipitation (UV-RIP) was performed as per the protocol described in (Rahnamoun et al. 2018) with minor variations. 2 x 10 cm dishes containing around 12 x 10^6 cells were taken for the experiment. Media was discarded and cells were washed thrice with 1X PBS. 5 ml of 1X PBS was added to the dishes and crosslinked by UV irradiation (150 mJ per cm^2 at 254 nm) using a Stratalinker. The cells were then scraped in 3 ml of 1X PBS and pelleted down at 3K rpm. The cells were lysed by adding 1 ml of RNA Immunoprecipitation (RIP) buffer [25 mM HEPES-KOH pH 7.5, 150 mM KCl, 0.5% NP40, 1.5 mM $MgCl_2$, 10% glycerol, 1 mM EDTA, 40U/ml RNase inhibitor (Invitrogen #18091050)], supplemented with 1X protease inhibitor cocktail, resuspending by pipetting and incubated on ice for 30 min. Lysates were cleared by centrifuging the samples at 12K rpm for 12 mins. Cleared cell lysates were divided equally into two tubes and 10% of it was kept as input. 1 µg of target antibody was added to the IP sample along with 15 µl of Protein G beads Dynabeads (Invitrogen) and in the beads sample, only 15 µl of Protein G beads Dynabeads were added. The samples were incubated overnight on a rocker. Beads were subsequently washed three times with ice-cold RIP lysis buffer for 5 mins. RNA samples were eluted using Trizol

reagent as described above. cDNA was prepared from eluted RNA samples as described above. All the stock solutions and buffers were prepared in RNase-free water.

Cloning of JMJD3 variants

pCMV-HA-JMJD3 plasmid was obtained from addgene (Plasmid #24167). Other variants of JMJD3 used in the study were cloned in the p3xFLAG-CMV-10 vector (E7658). pCMV-HA-JMJD3 plasmid was used as a template to PCR amplify all the variants. The region of interest was amplified with KOD Hot Start DNA Polymerase. The oligos used for cloning the variants are mentioned in the appendices. Different restriction sites were used for different variants.

For cloning the variants with nuclear localization signal (NLS), the Simian Virus 40 (SV40) NLS sequence (CCCAAGAAGAAGAGGAAAGTC) was added to the forward oligos.

Site-directed mutagenesis

Oligos were designed manually for creating point mutations in the RNA binding of JMJD3. Oligos were designed in such a way that the mutation was centered in the middle of it. Cysteines that are involved in the formation of the Zinc finger-like domain were mutated to Alanine. PCR amplification was carried out with KOD Hot Start DNA Polymerase by following the manufacturer's protocol given below.

Component	Volume	Final Concentration
10x Buffer	5 µL	1x
25 mM MgSO4	3 µL	1.5 mM
dNTPs (2 mM each)	5 µL	0.2 mM
PCR Grade Water	X µL	
Oligos F+R (10 µM)	3 µL	0.3 µM
KOD Polymerase (1.0 U/µL)	1.0 µL	0.02 U/µL
Template DNA	X µL	50 ng
Total Volume	**50 µL**	

2 x 50 µl reactions were set up to amplify the target regions.

Following thermocycling conditions were used to amplify the target regions:

Polymerase Activation	**Denaturation**	**Annealing**	**Extension**	**Final Extension**
95°C for 2 mins	95°C for 20 s	62°C for 10 s	70°C for 30 s	70°C for 10 mins
	40 Cycles			

The pCMV-HA-JMJD3 plasmid was used as a template for creating mutations. Two mutants with different sets of mutations were created. One mutant had the first and the second cysteine mutated to alanine and the other clone had the first and the fourth cysteine mutated.

Immunostaining

Cells grown on coverslips (after 24hrs of transfection) were gently washed twice with 1X PBS. 4% PFA was added and incubated at room temperature for 15 mins, followed by three washes with 1X PBS. To carry out permeabilization, cells were treated with 0.2% Triton-X in 1% FBS (blocking agent) for 15 minutes, followed by three 1X PBS washes. Then, the coverslips were incubated overnight at 4ºC, with pair of primary antibodies as required (HA-CST/Flag-CST with H3K27me3-CST) prepared in 1% FBS. After washing with 1X PBST (PBS with tween-20) thrice, the coverslips were incubated with secondary antibodies (Alexa Flour mouse 488 and Alexa Flour rabbit 647) for an hour at room temperature (protected from light). Finally, the samples were incubated with Hoechst 33342 to stain the nuclei, following three washes with 1X PBST. The coverslips were then mounted onto slides with 90% glycerol and sealed with nail paint. Imaging was done using Olympus FV3000 with 405, 488, and 631 nm lasers; taking z stacks for all.

Max-intensity mages are shown as representative images.

Imaris 9.4 was used for image analysis. In total ~80 cells were analyzed for every sample, with 2 replicates each. The Surface was created for .tiff images using 405 channels to get signal for H3K27me3 and Flag/HA-jmjd3 variants from nuclei specifically. The mean intensity across stacks: in the surface created was then measured and plotted as boxplots. The significance was tested with respect to empty vector samples using https://astatsa.com/WilcoxonTest/.

Hi-C analysis

Hi-C forward and reverse end reads were trimmed and aligned separately. The raw reads were mapped to hg19 assembly using bowtie 2. The HOMER program makeTagDirectory was first used to create tag directories with tbp 1. Data was further processed by HOMER in order to remove small fragment and self-ligations using makeTagDirectory with the following options: -removePEbg -removeSpikes 10000 5. Next, findTADsAndLoops.pl was used to obtain overlapping TADs, produced at a 20kb resolution with 40kb windows. H3K27ac peaks intersecting with the identified TADs were counted and ranked to obtain the enhancers per TAD slope. The HiC datasets were analyzed using the Juicer pipeline for visualization. The .hic file generated from the juicer pipeline was then visualized using Juicebox. The contact maps were generated using Balanced normalization (Knight-Ruiz balancing algorithm).

Super enhancer calling

Super-enhancers were called using the ROSE (Rank Ordering of Super-enhancers) algorithm (https://bitbucket.org/young_computation/) using the aligned ChIP-seq reads as input with parameters -s 12500.

4C-seq analysis

The sequenced reads were aligned to hg19 assembly using default Bowtie2 options. The output BAM file was used as the input for the FourCSeq pipeline. The first and second restriction site sequence was provided along with the primer sequence in the metadata. The viewpoint information is also provided. The reference genome is then

in-silico digested to obtain the reference fragments. Reads mapping exactly to the fragment ends are then counted. This data is then plotted after smoothening and calculating Z-scores to detect interactions. Interacting regions are defined with the following thresholds: a fragment must have z-scores larger than 3 and an adjusted p value of 0.01. Beside FourCseq pipeline, 4Cseqpipe (http://compgenomics.weizmann.ac.il/tanay/?page_id=367) was also used to process the sequenced data in. 4C-seq images were generated using truncated mean at a 10kb resolution.

RNA-Seq analysis

The raw reads were mapped to hg19 assembly using hisat2 in a strand specific manner. The output BAM file was sorted using Samtools. This was then provided as an input for htseq-count, of the HTSeq pipeline, to count reads in the exonic features. The raw counts from the different datasets were then used as in input in DESeq2 to obtain differentially expressed genes using default thresholds. The volcano plots were plotted using the EnhancedVolcano tool on R.

Gene expression in tumors

Gene expression data from cancer patients belonging to the TCGA cohort were obtained from TCGA Genomic Data Commons (GDC) (https://portal.gdc.cancer.gov/).

ATAC-seq data in cervical tumors

ATAC-seq data of Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) tumors from TCGA cohort were obtained from TCGA study (Corces et al., 2018).

Quantification and statistical analysis

Statistical significance is determined by unpaired t-test ($*p < 0.05$; $**p < 0.01$; $***p < 0.005$; $****p < 0.001$; ns $p > 0.05$). The error bars denote SEM. The statistical details of experiments can be found in the figure legends. n represents the number of biological replicates.

Chapter 5: Discussion and conclusions

Chapter 2

We used the Circular Chromosome Conformation Capture (4C) approach, followed by a series of enhancer blockings and deletions, to identify the regulatory enhancers within the dense cluster that regulate *INK4/ARF* transcription. We uncover a subset of enhancers that regulate gene transcription in a non-redundant manner as an interdependent network. The loss of even a single promoter-interacting enhancer in the network leads to gene inactivation due to EZH2 loading on promoters. Furthermore, genome-wide gene expression profiles upon enhancer deletions show effects that are relevant to the diseases with which the entire 9p21 locus is associated.

CDKN2A/INK4a promoter interacts with only a subset of enhancers

INK4/ARF locus harbors 24 enhancers in the HeLa cell line. These enhancers are present in the gene desert region or overlap with the *CDKN2BAS* gene. Analysis of different epigenomic data sets in HeLa revealed that only 15 of the 24 enhancers were active. Three enhancers overlapped with the 3` TAD boundary, out of the remaining 21 enhancers only 5 enhancers interacted with the promoter as revealed the by 4C-seq experiment. Since the enhancers that interacted with the promoters showed varying levels of enhancer-associated active marks such as H3K27ac alone could not predict the enhancer that would loop with the promoters. For instance, E8 and E17 displayed exceedingly high H3K27ac levels and showed high-frequency interaction with *CDKN2A/INK4a* promoter. E21, on the other hand, has H3K27ac levels similar to E8 and E17 but does not interact with the promoter. The interaction with the promoters is the best predictor for the regulatory enhancers because they could not be functionally predicted from the H3K27ac or eRNA levels.

This enhancer cluster is very dense, and such dense clusters can be found throughout the genome. Since just a few enhancers in this cluster have regulatory potential, the question of what other non-interacting enhancers are doing becomes essential. Because perturbing one of these non-interacting promoters had a mild effect on gene expression, suggesting that non-interacting promoters may be performing a different and lower ranking function. Non-interacting enhancers, for example, may keep the surrounding

chromatin open by attracting various transcription factors and co-factors. Thus, non-interacting enhancers may indirectly help regulatory enhancers by maintaining transcription factor and cofactor density. Furthermore, they may be assisting in the prevention of H3K27me3 spread via PRC2. Enhancer clusters can phase-separate because they recruit transcription factors and cofactors in large quantities, which is required for phase separation. Each constituent enhancer contributes to the valency needed for phase separation. Regulatory enhancers may be essential because they may contribute more to valency than non-interacting enhancers. Thus, the deletion of a regulatory enhancer may lead to the complete disruption of the phase-separated entity. As a result, target genes and the other enhancers show drastic effects on target gene transcription. Furthermore, interacting enhancers and promoters may be in closer proximity in the phase-separated condensate than non-interacting enhancers and are thus captured in 4C.

An interdependent network operates at this locus

We also observed that disruption of promoter-interacting enhancers severely impacts gene transcription and also, the activity of other enhancers in the cluster. H3K27ac and eRNA transcription from the other intact enhancers are lost when a single promoter interacting enhancer is perturbed. This implies that the promoter-interacting enhancers have some sort of reciprocal dependence on one another for transcriptional activity. We also observed that promoter-interacting enhancers interact with promoters and other promoter-interacting enhancers and that the deletion of a single enhancer affects these enhancer-enhancer interactions. Functional enhancers within a super-enhancer can establish an equal-weighted network in which target genes cannot express in the absence of even a single enhancer. In this cluster, the enhancers are not redundant, contrary to previous genome-wide results (Moorthy et al. 2017).

The loss of function mutation in any enhancer within equal-weight networks would result in the dysregulation of all enhancers in the network, resulting in promoter dysregulation. Perhaps the *INK4/ARF* locus being the most reproducible GWAS study is due to such enhancer dependencies, where a change in any enhancer results in a total collapse of the transcriptional output of the locus.

H3K27ac loss reflects enhancer function but not the extent

Through CRISPRi and enhancer knockouts of specific promoter-interacting enhancers, we saw the loss of H3K27ac on intact enhancers. Surprisingly, the extent to which H3K27ac was lost on intact enhancers upon CRISPRi or enhancer knockout was the same. However, compared to CRISPRi lines, enhancer knockout lines exhibited a far more severe loss of eRNA transcription. Additionally, enhancer knockout lines exhibited drastic downregulation of gene transcription than CRISPRi lines do. This implies that while H3K27ac can provide information about the activity of regulatory elements, it cannot provide information about the magnitude of the activity. In this instance, eRNA transcription was a more accurate predictor of the extent of enhancer activity.

In line with this, it has been demonstrated that the enhancer functions are unaffected by the histone H3 mutation at position K27 (Zhang et al. 2020). The CBP/p300, which catalyzes acetylation at H3K27, however, plays a crucial role in enhancer activity in other circumstances (Raisner et al. 2018). Furthermore, high levels of H3K27ac may be the result of eRNA present at these enhancers, implying its existence as a result of enhancer activity (Pnueli et al. 2015).

Enhancer disruption causes EZH2 loading on promoters in an ANRIL-independent manner

The *INK4/ARF* locus is a well-studied target of the PRC2 complex. Several studies have demonstrated that PRC2 binds to the promoters of these genes in dividing cells and certain malignancies. HeLa cells show the absence of the PRC2 complex, as *INK4/ARF* genes are highly active in these cells. We observed that enhancer knockouts cause EZH2 to load onto the *CDKN2A* and *CDKN2B* gene promoters and as a result, exhibit an increase in H3K27me3 levels. This implies that one of the ways these enhancers control the genes is by preventing PRC2 from targeting the gene promoters. Several studies have previously revealed that ANRIL, which is produced from the *CDKN2BAS* gene, interacts with PRC2 complex components (Kotake et al. 2011). It has been demonstrated that it recruits PRC2 to the promoters of *CDKN2A* and *CDKN2B* genes (Kotake et al. 2011; Yap et al. 2010). Given that we suspected that the EZH2 increase could be attributable to ANRIL overexpression in the enhancer knockout lines. To our

surprise, however, the ANRIL expression similarly decreased in the enhancer knockout lines, indicating that these enhancers may potentially be involved in the regulation of *CDKN2BAS* gene transcription as well. The promoters of the *CDKN2A* and *CDKN2BAS* genes are close linearly, and some studies have even labeled them as bidirectional promoters. As a result, these enhancers may have the ability to influence both promoters. The decreased expression of ANRIL in the enhancer knockout lines showed that the PRC2 complex is recruited to *INK4/ARF* promoters in an ANRIL-independent manner. This was evident in the patient RNA sequencing data as well, where ANRIL was found to have a positive correlation with the expression of *CDKN2A* and *CDKN2B* genes.

It has been demonstrated that RNA sequesters or inhibits PRC2 from binding to the neighboring regions, keeping the genes in the vicinity active. Thus, loss of EZH2 sequestration due to the downregulation of gene transcripts and eRNAs is one of the putative mechanisms of EZH2 binding onto these promoters upon enhancer knockouts. Additionally, H3K27ac loss may result in PRC2 binding and H3K27me3 increase. Since we know that promoters and enhancers are involved in genomic interactions, the deletion of an enhancer results in the loss of these interactions. The active promoter and enhancer interactions may prevent PRC2 from targeting the promoters. As a result, the loss of enhancer-promoter interaction may expose promoters, making them vulnerable to PRC2 binding. Furthermore, some circular ANRIL isoforms have recently been found; these isoforms have been demonstrated to keep the *INK4/ARF* locus active by not allowing PRC2 to target promoters (Muniz et al. 2021). Knockdown of these isoforms reduced the expression of *INK4/ARF* genes in a PRC2-dependent manner. Thus, it is also conceivable that such circular isoforms of ANRIL are downregulated in the enhancer knockout lines, resulting in PRC2 accumulation on the promoters.

Genes dysregulated upon enhancer deletions corroborate with disease association of the *INK4/ARF* locus

The *INK4/ARF* locus is one of the most reproducible GWAS hotspots. SNPs in the locus are linked to a variety of aging-related disorders such as type-2-diabetes, coronary artery disease, atherosclerosis, and several cancers. These SNPs have either been found in the gene desert region harboring the enhancers cluster, some of which overlap with

the *CDKN2BAS* gene. RNA-seq data from enhancer knockdown lines revealed that roughly 200 genes were dysregulated. Around 100 genes were upregulated, whereas the remaining were downregulated. Most of the dysregulated genes were common across the enhancer knockout lines, pointing to the potential indirect role of the *INK4/ARF* genes. Despite the fact that the HeLa cell line is not the appropriate model, the dysregulated genes upon enhancer knockouts corroborated with the disease association of this locus. For instance, dysregulated genes in enhancer knockout lines are involved in aging-associated diseases like coronary artery disease, diabetes, etc. This shows that the *INK4/ARF* genes regulate several of these pathways in HeLa cells and the functions are conserved across cell-types.

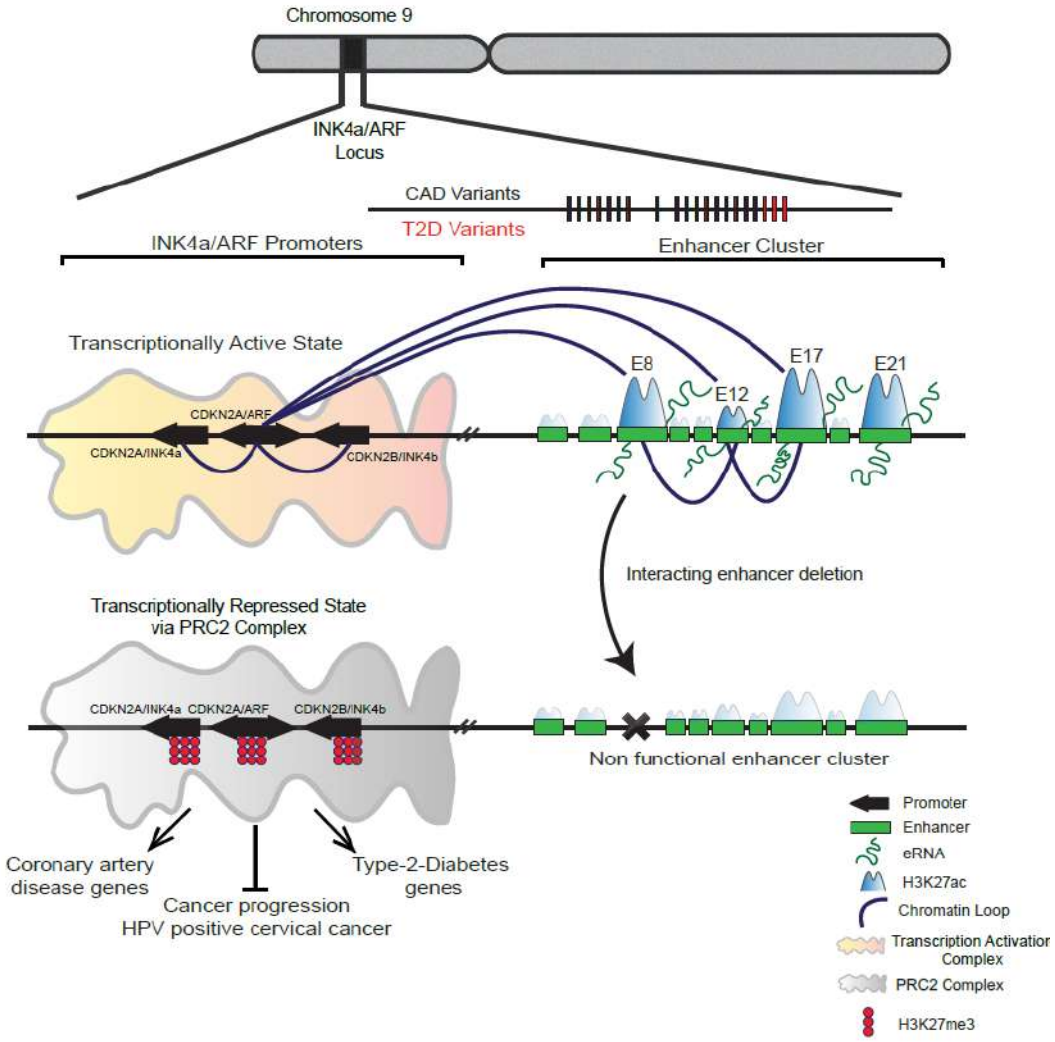


Figure 4.1: Proposed model of enhancer-mediated transcriptional regulation of *INK4/ARF* locus.

INK4/ARF locus contains two genes CDKN2A and CDKN2B that code for three critical cell cycle regulators. This locus is the most reproducible GWAS hotspot and harbors 24 enhancers in HeLa cells. Only a subset of enhancers from the dense multi-enhancer cluster regulates this locus. The regulatory enhancers interact with each other to form a single functional unit that is entirely dependent on each enhancer for target gene regulation. When a single functional enhancer is deleted, the entire enhancer cluster becomes non-functional. The active enhancers prevent EZH2 from binding to the INK4/ARF promoters and downregulating these genes. The genes that are dysregulated upon enhancer perturbations corroborate with the disease association of this locus.

Chapter 3

We created several truncations of the JMJD3 protein to better understand how this chromatin modifier recognizes its target binding sites. We also generated several mutants of JMJD3 from these truncations. We used high throughput approaches to study its chromatin binding specificity. Our data highlights that JMJD3 binding with chromatin is dependent on its N-terminal IDR and C-terminal zinc finger-like domain. When the entire IDR is removed, JMJD3 loses its ability to bind to regulatory components such as enhancers and promoters. Interestingly, JMJD3 binding is switched from enhancers to promoters when a part of the IDR is deleted. Additionally, JMJD3 binding to chromatin is lost when the C-terminal domain is disrupted. Our findings imply that the IDR and C-terminus of JMJD3 is necessary for chromatin binding, but its IDR also dictates binding specificity.

JMJD3 regulates the *INK4/ARF* locus in a catalytically dependent manner

JMJD3 has been shown to activate the *INK4/ARF* locus during the onset of senescence. Several cellular cues such as oncogene overexpression upregulate JMJD3 in MEFs, thereby resulting in INK4a-mediated cell cycle arrest. JMJD3 achieves this in a catalytically dependent manner by eliminating H3K27me3 from the promoters of these genes. *INK4/ARF* locus is highly active in HeLa cells, and we observed that JMJD3 regulates it in HeLa as well. Knockdown of JMJD3 by siRNAs results in the downregulation of all three *INK4/ARF* transcripts, namely INK4a, ARF, and INK4b. JMJD3 has been reported to regulate several genes in a catalytically independent manner and we observed similar effects when JMJD3 catalytic activity was inhibited for 24 hours by treating cells with GSK-J4. The data suggest that the catalytic activity of JMJD3 is essential for the activation of the *INK4/ARF* locus in HeLa cells. Furthermore, inhibiting the catalytic activity of PRC2 leads to modest overexpression of all three genes. This shows that the *INK4/ARF* locus is tightly controlled by JMJD3 and PRC2 in distinct cell types. In HeLa cells, however, JMJD3 outcompetes PRC2 and keeps the locus active.

JMJD3 binds to regulatory enhancers of the *INK4/ARF* locus as well as other active enhancers found across the genome

JMJD3 has previously been demonstrated to bind regulatory elements such as enhancers (Williams et al. 2014). Similarly, JMJD3 is recruited to p53 response elements by p53, and it modulates target gene expression via enhancers. In our study, JMJD3 ChIP-seq revealed that it binds to both enhancers and promoters. However, it is comparatively more enriched on enhancers than on promoters. JMJD3 bound enhancers appeared to be active, as attested by a high enrichment of active enhancer marks such as H3K27ac and H3K4me1. JMJD3 was also found to bind to regulatory enhancers of the *INK4/ARF* locus (as detailed in chapter 2); specifically, E8, E12, and E17. JMJD3 binding to these enhancers is critical for enhancer activity since JMJD3 inhibition results in eRNA downregulation (Figure 3.1A and 3.1B). However, PRC2 or H3K27me3 were not enriched on enhancers after JMJD3 knockout. This implies that JMJD3 is essential for enhancer function, although enhancers are not inhibited by PRC2 in the absence of JMJD3. Enhancers may be affected by the lack of activators caused by JMJD3 knockdown or by another repressor suppressing the enhancer function.

JMJD3 binds *INK4/ARF* enhancers in an eRNA-dependent manner

JMJD3 has been shown to interact with some viral RNAs and lncRNAs. We conducted several experiments to see if it binds to eRNAs. JMJD3 was found to be an RNA-binding histone modifier. RNA has been demonstrated to interact with a number of proteins and RNA-protein interactions have been found to serve a variety of regulatory activities, including; protein sequestration, chromatin binding assistance, stabilization of protein binding on chromatin, protein degradation etc. *INK4/ARF* regulatory enhancers are transcribed into bidirectional eRNAs. Knockdown of eRNAs transcribed from these enhancers resulted in downregulation of all three *INK4/ARF* transcripts. The downregulation occurs because JMJD3 enrichment on these enhancers was lost when eRNAs were knocked down, indicating the relevance of eRNAs in JMJD3 chromatin binding. Interestingly, loss of eRNAs resulted in PRC2 binding on the *INK4/ARF* promoters. We also observed a modest increase in H3K27me3 on these promoters. However, the loss of eRNAs didn't trigger PRC2 binding on enhancers. This is consistent with the data obtained after the JMJD3 knockdown and implies that eRNAs

play a dual role, assisting in JMJD3 recruitment on enhancers while also blocking PRC2 loading on promoters.

This demonstrates that eRNAs derived from these enhancers have a role in enhancer activity. JMJD3 binding to these enhancers is affected by the loss of eRNAs. When JMJD3 binding is lost, PRC2 binds to promoters, causing downregulation of all three transcripts. eRNAs could accomplish this by assisting JMJD3 recruitment to these enhancers while also preventing PRC2 from targeting the promoters. Alternatively, active engagement of promoters with enhancers may prevent the PRC2 complex from targeting promoters. The loss of eRNAs may result in the loss of looping between enhancers and promoters, rendering promoters more susceptible to PRC2 binding.

JMJD3's N-terminus is a disordered region that determines its chromatin binding specificity

JMJD3 is a large protein with only one characterized domain, JmjC. This JMJD3 domain has been shown to catalyze the removal of trimethyl mark from lysine 27 of Histone 3. Apart from this domain, JMJD3 has a region that binds to zinc ions and folds into a zinc finger-like domain. The function of this region is mainly unknown. Nonetheless, these JMJD3 domains are relatively small and account for just a very small percentage of JMJD3 protein. The large N-terminal region is mainly unexplored. We found that this N-terminal region is highly disordered. This disordered region extends from the N-terminus to the JmjC domain indicating that JMJD3 is mostly a disordered protein. One recent study found that this IDR plays a role in JMJD3 phase separation (Vicioso-Mantis et al. 2022) however, its role in the chromatin binding or specificity of jmjd3 binidng is not explored.

By creating truncation of IDR, we observed that the IDR plays a significant role in defining JMJD3 binding specificity on chromatin. JMJD3 binds to both promoters and enhancers, as previously stated, but preferentially to enhancers. JMJD3 binding on chromatin is altered when a portion of the IDR is removed. Surprisingly, JMJD3 with a deleted portion of the IDR prefers to bind to promoters over enhancers. JMJD3's IDR may specifically interact with transcription factors and co-factors and its interactions with protein factors may be determined by IDR and posttranslational modifications of its IDR. By deleting a portion of the IDR, binding sites for certain protein factors may

be eliminated, and the remaining portion may still interact with certain proteins but not all, resulting in a loss of specificity.

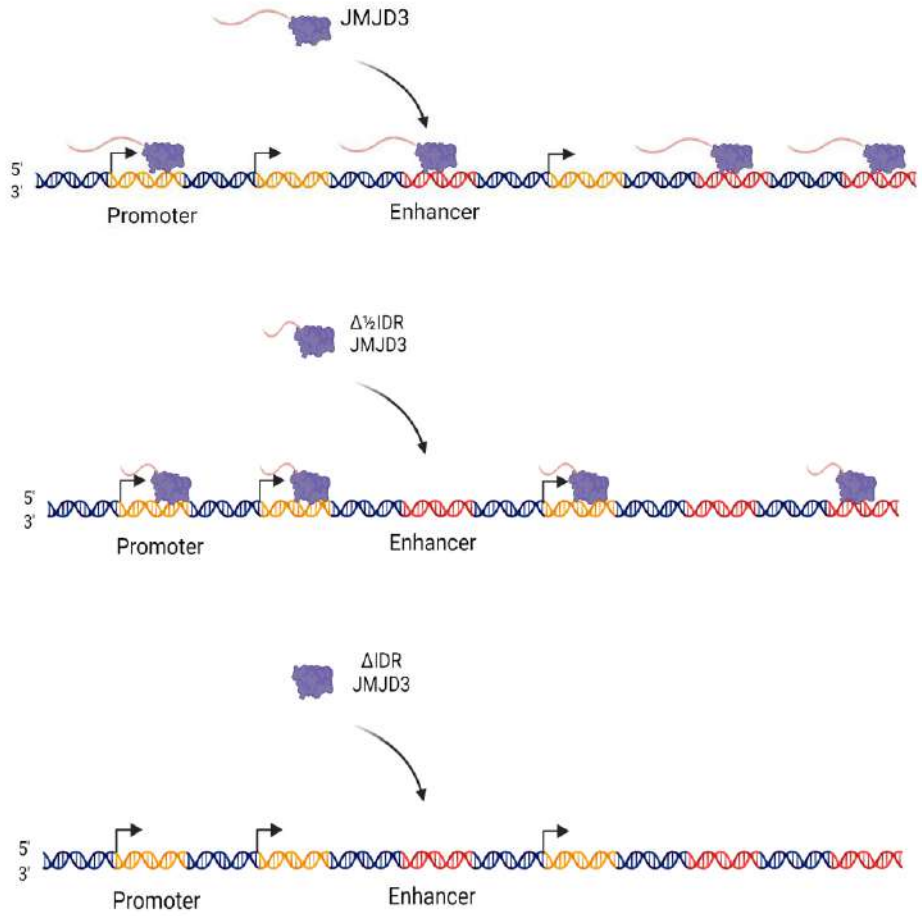
When the entire IDR is removed from JMJD3, it is unable to bind both promoters and enhancers. This implies that JMJD3 binding to chromatin requires its IDR, and the first portion of the IDR regulates binding specificity. This once again suggests that IDR is interacting with transcription factors or cofactors. Complete IDR removal would result in a lack of interaction with these factors and, as a result, chromatin binding of JMJD3.

Full IDR deletion affects the chromatin binding of JMJD3 and its catalytic activity

We co-immunostained JMJD3 and its truncations using antibody against H3K27me3 to test how IDR deletion affected its catalytic activity. As reported, we observed that JMJD3 removes the trimethyl mark from the cells. Its distribution was most noticeable around the nuclear periphery. Similarly, H3K27me3 was more enriched closer to the nuclear periphery. JMJD3-expressing cells lost global H3K27me3 and primarily at the nuclear periphery. JMJD3 with a portion of the IDR deleted (ΔHI-JMJD3) was catalytically active and eliminated H3K27me3. However, ΔHI-JMJD3 showed a much greater decrease of H3K27me3 than JMJD3. This could be explained by the binding preference of the ΔHI-JMJD3. Since H3K27me3 is mostly found on promoters and ΔHI-JMJD3 prefers to bind to promoters. As a result of its binding to H3K27me3-marked promoters, it causes greater loss, as compared to JMJD3.

Truncated JMJD3 with full IDR deletion (ΔFI-JMJD3) was unable to remove H3K27me3 since H3K27me3 levels in ΔFI-JMJD3 transfected cells did not decrease. JMJD3 may require the IDR for catalytic activity, so removing it would render it catalytically inactive. This could also be due to ΔFI-JMJD3 failure to bind to chromatin. Since trimethylation occurs mostly on histones linked to DNA, ΔFI-JMJD3 loses its ability to bind chromatin and so cannot erase the methyl mark. In vitro experiments need to be performed to investigate the involvement of IDR in the catalytic activity of JMJD3.

A


JMJD3
5'
3'
Promoter
Enhancer
Δ½IDR
JMJD3
5'
3'
Promoter
Enhancer
ΔIDR
JMJD3
5'
3'
Promoter
Enhancer

B

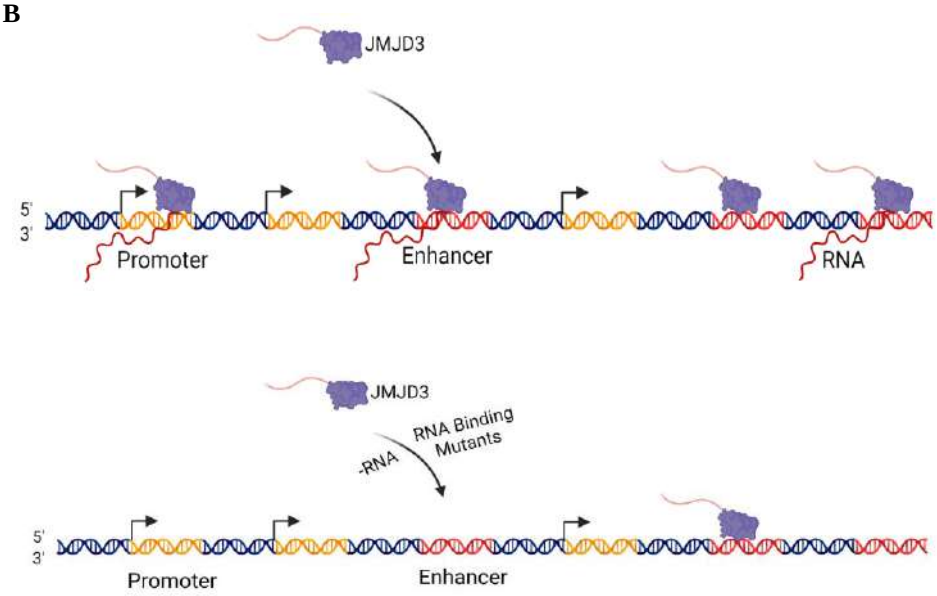

JMJD3
5'
3'
Promoter
Enhancer
RNA
JMJD3
-RNA
RNA Binding
Mutants
5'
3'
Promoter
Enhancer

Figure 4.2: The IDR and C-terminus of JMJD3 are necessary for chromatin binding, but the IDR also dictates specificity. ***A)*** *In addition to the canonical catalytic domain, JmJC, JMJD3 protein has a highly disordered region at its N terminus and a zinc binding region at its C terminus. JMJD3 requires both its IDR and C-terminus to bind to chromatin. JMJD3 FL binds to both promoters and enhancers, but it is more abundant on enhancers. When a portion of JMJD3's IDR is removed, it still retains chromatin binding, but its target binding sites are altered from enhancers to promoters. However, JMJD3 loses its chromatin binding when IDR is completely removed.* ***B)*** *JMJD3 binding to chromatin is also RNA dependent. Its C-terminus interacts with RNA. JMJD3 chromatin binding is lost due to deletion of its RNA binding region or point mutations in its RNA binding region.*

References

1. Agger, K., P. A. Cloos, L. Rudkjaer, K. Williams, G. Andersen, J. Christensen, and K. Helin. 2009. 'The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence', *Genes Dev*, 23: 1171-6.

2. Aguilo, F., M. M. Zhou, and M. J. Walsh. 2011. 'Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression', *Cancer Res*, 71: 5365-9.

3. Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A. R. R. Forrest, P. Carninci, M. Rehli, and A. Sandelin. 2014. 'An atlas of active enhancers across human cell types and tissues', *Nature*, 507: 455-61.

4. Argaud, D., M. C. Boulanger, A. Chignon, G. Mkannez, and P. Mathieu. 2019. 'Enhancer-mediated enrichment of interacting JMJD3-DDX21 to ENPP2 locus prevents R-loop formation and promotes transcription', *Nucleic Acids Res*, 47: 8424-38.

5. Arnold, C. D., D. Gerlach, C. Stelzer, M. Boryń Ł, M. Rath, and A. Stark. 2013. 'Genome-wide quantitative enhancer activity maps identified by STARR-seq', *Science*, 339: 1074-7.

6. Arnold, Preston R., Andrew D. Wells, and Xian C. Li. 2020. 'Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate', *Frontiers in Cell and Developmental Biology*, 7.

7. Banerji, J., S. Rusconi, and W. Schaffner. 1981. 'Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences', *Cell*, 27: 299-308.

8. Bouchard, C., S. Lee, V. Paulus-Hock, C. Loddenkemper, M. Eilers, and C. A. Schmitt. 2007. 'FoxO transcription factors suppress Myc-driven lymphomagenesis via direct activation of Arf', *Genes Dev*, 21: 2775-87.

9. Bracken, A. P., D. Kleine-Kohlbrecher, N. Dietrich, D. Pasini, G. Gargiulo, C. Beekman, K. Theilgaard-Mönch, S. Minucci, B. T. Porse, J. C. Marine, K. H. Hansen, and K. Helin. 2007. 'The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells', *Genes Dev*, 21: 525-30.

10. Brodsky, S., T. Jana, K. Mittelman, M. Chapal, D. K. Kumar, M. Carmi, and N. Barkai. 2020. 'Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity', *Mol Cell*, 79: 459-71.e4.

11. Calo, E., and J. Wysocka. 2013. 'Modification of enhancer chromatin: what, how, and why?', *Mol Cell*, 49: 825-37.

12. Casamassimi, A., and A. Ciccodicola. 2019. 'Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation', *Int J Mol Sci*, 20.

13. Caudron-Herger, M., S. F. Rusin, M. E. Adamo, J. Seiler, V. K. Schmid, E. Barreau, A. N. Kettenbach, and S. Diederichs. 2019. 'R-DeeP: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation', *Mol Cell*, 75: 184-99.e10.

14. Chen, S., J. Ma, F. Wu, L. J. Xiong, H. Ma, W. Xu, R. Lv, X. Li, J. Villen, S. P. Gygi, X. S. Liu, and Y. Shi. 2012. 'The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation', *Genes Dev*, 26: 1364-75.

15. Chittock, E. C., S. Latwiel, T. C. Miller, and C. W. Müller. 2017. 'Molecular architecture of polycomb repressive complexes', *Biochem Soc Trans*, 45: 193-205.

16. Claringbould, Annique, and Judith B. Zaugg. 2021. 'Enhancers in disease: molecular basis and emerging treatment strategies', *Trends in molecular medicine*, 27: 1060-73.

17. Cobrinik, D. 2005. 'Pocket proteins and cell cycle control', *Oncogene*, 24: 2796-809.

18. Collins, Kathleen, Tyler Jacks, and Nikola P. Pavletich. 1997. 'The cell cycle and cancer', *Proceedings of the National Academy of Sciences*, 94: 2776-78.

19. Core, L. J., J. J. Waterfall, and J. T. Lis. 2008. 'Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters', *Science*, 322: 1845-8.

20. Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, Michael A. Lodato, Garrett M. Frampton, Phillip A. Sharp, Laurie A. Boyer, Richard A. Young, and Rudolf Jaenisch. 2010. 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proceedings of the National Academy of Sciences*, 107: 21931-36.

21. Dave, K., I. Sur, J. Yan, J. Zhang, E. Kaasinen, F. Zhong, L. Blaas, X. Li, S. Kharazi, C. Gustafsson, A. De Paepe, R. Månsson, and J. Taipale. 2017. 'Mice deficient of Myc super-enhancer region reveal differential control mechanism between normal and pathological growth', *Elife*, 6.

22. Dietrich, N., A. P. Bracken, E. Trinh, C. K. Schjerling, H. Koseki, J. Rappsilber, K. Helin, and K. H. Hansen. 2007. 'Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus', *Embo j*, 26: 1637-48.

23. Dimova, D. K., and N. J. Dyson. 2005. 'The E2F transcriptional network: old acquaintances with new faces', *Oncogene*, 24: 2810-26.

24. Farooq, Umer, and Dimple Notani. 2021. 'Optimized protocol to create deletion in

adherent cell lines using CRISPR/Cas9 system', *STAR Protocols*, 2: 100857.

25. Farooq, Umer, and Dimple Notani. 2022. 'Transcriptional regulation of INK4/ARF locus by cis and trans mechanisms', *Frontiers in Cell and Developmental Biology*, 10.

26. Font, J., and J. P. Mackay. 2010. 'Beyond DNA: zinc finger domains as RNA-binding modules', *Methods Mol Biol*, 649: 479-91.

27. Forbes, S., J. Clements, E. Dawson, S. Bamford, T. Webb, A. Dogan, A. Flanagan, J. Teague, R. Wooster, P. A. Futreal, and M. R. Stratton. 2006. 'COSMIC 2005', *Br J Cancer*, 94: 318-22.

28. Frescas, D., D. Guardavaccaro, F. Bassermann, R. Koyama-Nasu, and M. Pagano. 2007. 'JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes', *Nature*, 450: 309-13.

29. Gamell, C., D. Ginsberg, S. Haupt, and Y. Haupt. 2017. 'New insights on the regulation of INK4/ARF locus expression', *Oncotarget*, 8: 106147-48.

30. Giacinti, C., and A. Giordano. 2006. 'RB and cell cycle progression', *Oncogene*, 25: 5220-7.

31. Gil, J., D. Bernard, D. Martínez, and D. Beach. 2004. 'Polycomb CBX7 has a unifying role in cellular lifespan', *Nat Cell Biol*, 6: 67-72.

32. Gilbert, Luke A, Matthew H Larson, Leonardo Morsut, Zairan Liu, Gloria A Brar, Sandra E Torres, Noam Stern-Ginossar, Onn Brandman, Evan H Whitehead, Jennifer A Doudna, Wendell A Lim, Jonathan S Weissman, and Lei S Qi. 2013. 'CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes', *Cell*, 154: 442-51.

33. Gorbovytska, Vladyslava, Seung-Kyoon Kim, Filiz Kuybu, Michael Götze, Dahun Um, Keunsoo Kang, Andreas Pittroff, Theresia Brennecke, Lisa-Marie Schneider,

Alexander Leitner, Tae-Kyung Kim, and Claus- D. Kuhn. 2022. 'Enhancer RNAs stimulate Pol II pause release by harnessing multivalent interactions to NELF', *Nature Communications*, 13: 2429.

34. Gu, B., T. Swigut, A. Spencley, M. R. Bauer, M. Chung, T. Meyer, and J. Wysocka. 2018. 'Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements', *Science*, 359: 1050-55.

35. Haiman, C. A., N. Patterson, M. L. Freedman, S. R. Myers, M. C. Pike, A. Waliszewska, J. Neubauer, A. Tandon, C. Schirmer, G. J. McDonald, S. C. Greenway, D. O. Stram, L. Le Marchand, L. N. Kolonel, M. Frasco, D. Wong, L. C. Pooler, K. Ardlie, I. Oakley-Girvan, A. S. Whittemore, K. A. Cooney, E. M. John, S. A. Ingles, D. Altshuler, B. E. Henderson, and D. Reich. 2007. 'Multiple regions within 8q24 independently affect risk for prostate cancer', *Nat Genet*, 39: 638-44.

36. Hannon, G. J., and D. Beach. 1994. 'p15INK4B is a potential effector of TGF-beta-induced cell cycle arrest', *Nature*, 371: 257-61.

37. Harismendy, Olivier, Dimple Notani, Xiaoyuan Song, Nazli G. Rahim, Bogdan Tanasa, Nathaniel Heintzman, Bing Ren, Xiang-Dong Fu, Eric J. Topol, Michael G. Rosenfeld, and Kelly A. Frazer. 2011. '9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response', *Nature*, 470: 264-68.

38. Hatzis, P., and I. Talianidis. 2002. 'Dynamics of enhancer-promoter communication during differentiation-induced gene activation', *Mol Cell*, 10: 1467-77.

39. He, J., E. M. Kallin, Y. Tsukada, and Y. Zhang. 2008. 'The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b)', *Nat Struct Mol Biol*, 15: 1169-75.

40. Helgadottir, Anna, Gudmar Thorleifsson, Kristinn P. Magnusson, Solveig Grétarsdottir, Valgerdur Steinthorsdottir, Andrei Manolescu, Gregory T. Jones,

Gabriel J. E. Rinkel, Jan D. Blankensteijn, Antti Ronkainen, Juha E. Jääskeläinen, Yoshiki Kyo, Guy M. Lenk, Natzi Sakalihasan, Konstantinos Kostulas, Anders Gottsäter, Andrea Flex, Hreinn Stefansson, Torben Hansen, Gitte Andersen, Shantel Weinsheimer, Knut Borch-Johnsen, Torben Jorgensen, Svati H. Shah, Arshed A. Quyyumi, Christopher B. Granger, Muredach P. Reilly, Harland Austin, Allan I. Levey, Viola Vaccarino, Ebba Palsdottir, G. Bragi Walters, Thorbjorg Jonsdottir, Steinunn Snorradottir, Dana Magnusdottir, Gudmundur Gudmundsson, Robert E. Ferrell, Sigurlaug Sveinbjornsdottir, Juha Hernesniemi, Mika Niemelä, Raymond Limet, Karl Andersen, Gunnar Sigurdsson, Rafn Benediktsson, Eric L. G. Verhoeven, Joep A. W. Teijink, Diederick E. Grobbee, Daniel J. Rader, David A. Collier, Oluf Pedersen, Roberto Pola, Jan Hillert, Bengt Lindblad, Einar M. Valdimarsson, Hulda B. Magnadottir, Cisca Wijmenga, Gerard Tromp, Annette F. Baas, Ynte M. Ruigrok, Andre M. van Rij, Helena Kuivaniemi, Janet T. Powell, Stefan E. Matthiasson, Jeffrey R. Gulcher, Gudmundur Thorgeirsson, Augustine Kong, Unnur Thorsteinsdottir, and Kari Stefansson. 2008. 'The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm', *Nature Genetics*, 40: 217-24.

41. Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, and R. A. Young. 2013. 'Super-enhancers in the control of cell identity and disease', *Cell*, 155: 934-47.

42. Hong, S., Y. W. Cho, L. R. Yu, H. Yu, T. D. Veenstra, and K. Ge. 2007. 'Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases', *Proc Natl Acad Sci U S A*, 104: 18439-44.

43. Huang, Yinghua, Hui Zhang, Lulu Wang, Chuanqing Tang, Xiaogan Qin, Xinyu Wu, Meifang Pan, Yujia Tang, Zhongzhou Yang, Isaac A. Babarinde, Runxia Lin, Guanyu Ji, Yiwei Lai, Xueting Xu, Jianbin Su, Xue Wen, Takashi Satoh, Tanveer Ahmed, Vikas Malik, Carl Ward, Giacomo Volpe, Lin Guo, Jinlong Chen, Li Sun, Yingying Li, Xiaofen Huang, Xichen Bao, Fei Gao, Baohua Liu, Hui Zheng, Ralf Jauch, Liangxue Lai, Guangjin Pan, Jiekai Chen, Giuseppe Testa, Shizuo Akira, Jifan Hu, Duanqing Pei, Andrew P. Hutchins, Miguel A. Esteban, and Baoming Qin. 2020. 'JMJD3 acts in tandem with KLF4 to facilitate reprogramming to

pluripotency', *Nature Communications*, 11: 5061.

44. Ivanchuk, Stacey M., Soma Mondal, Peter B. Dirks, and James T. Rutka. 2001. 'The INK4A/ARF Locus: Role in Cell Cycle Control and Apoptosis and Implications for Glioma Growth', *Journal of Neuro-Oncology*, 51: 219-29.

45. Jacobs, J. J., K. Kieboom, S. Marino, R. A. DePinho, and M. van Lohuizen. 1999. 'The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus', *Nature*, 397: 164-8.

46. Jayani, R. S., A. Singh, and D. Notani. 2017. 'Isolation of Nuclear RNA-Associated Protein Complexes', *Methods Mol Biol*, 1543: 187-93.

47. Jia, L., G. Landan, M. Pomerantz, R. Jaschek, P. Herman, D. Reich, C. Yan, O. Khalid, P. Kantoff, W. Oh, J. R. Manak, B. P. Berman, B. E. Henderson, B. Frenkel, C. A. Haiman, M. Freedman, A. Tanay, and G. A. Coetzee. 2009. 'Functional enhancers at the gene-poor 8q24 cancer-linked locus', *PLoS Genet*, 5: e1000597.

48. Jia, Yunlu, Wee-Joo Chng, and Jianbiao Zhou. 2019. 'Super-enhancers: critical roles and therapeutic targets in hematologic malignancies', *Journal of Hematology & Oncology*, 12: 77.

49. Kanao, H., T. Enomoto, Y. Ueda, M. Fujita, R. Nakashima, Y. Ueno, T. Miyatake, T. Yoshizaki, G. S. Buzard, T. Kimura, K. Yoshino, and Y. Murata. 2004. 'Correlation between p14(ARF)/p16(INK4A) expression and HPV infection in uterine cervical cancer', *Cancer Lett*, 213: 31-7.

50. Kaneko, S., J. Son, R. Bonasio, S. S. Shen, and D. Reinberg. 2014. 'Nascent RNA interaction keeps PRC2 activity poised and in check', *Genes Dev*, 28: 1983-8.

51. Kerppola, T. K. 2009. 'Polycomb group complexes--many combinations, many functions', *Trends Cell Biol*, 19: 692-704.

52. Kim, T. K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin,

M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, and M. E. Greenberg. 2010. 'Widespread transcription at neuronal activity-regulated enhancers', *Nature*, 465: 182-7.

53. Kim, W. Y., and N. E. Sharpless. 2006. 'The regulation of INK4/ARF in cancer and aging', *Cell*, 127: 265-75.

54. Kotake, Y., T. Nakagawa, K. Kitagawa, S. Suzuki, N. Liu, M. Kitagawa, and Y. Xiong. 2011. 'Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene', *Oncogene*, 30: 1956-62.

55. Kretz, M., and G. Meister. 2014. 'RNA binding of PRC2: promiscuous or well ordered?', *Mol Cell*, 55: 157-8.

56. Krishnamurthy, J., C. Torrice, M. R. Ramsey, G. I. Kovalev, K. Al-Regaiey, L. Su, and N. E. Sharpless. 2004. 'Ink4a/Arf expression is a biomarker of aging', *J Clin Invest*, 114: 1299-307.

57. Kwak, H., N. J. Fuda, L. J. Core, and J. T. Lis. 2013. 'Precise maps of RNA polymerase reveal how promoters direct initiation and pausing', *Science*, 339: 950-3.

58. Lam, M. T., W. Li, M. G. Rosenfeld, and C. K. Glass. 2014. 'Enhancer RNAs and regulated transcriptional programs', *Trends Biochem Sci*, 39: 170-82.

59. Lazorthes, S., C. Vallot, S. Briois, M. Aguirrebengoa, J. Y. Thuret, G. St Laurent, C. Rougeulle, P. Kapranov, C. Mann, D. Trouche, and E. Nicolas. 2015. 'A vlincRNA participates in senescence maintenance by relieving H2AZ-mediated repression at the INK4 locus', *Nat Commun*, 6: 5971.

60. Lee, T. I., and R. A. Young. 2013. 'Transcriptional regulation and its misregulation in disease', *Cell*, 152: 1237-51.

61. Li, H., M. Collado, A. Villasante, K. Strati, S. Ortega, M. Cañamero, M. A. Blasco,

and M. Serrano. 2009. 'The Ink4/Arf locus is a barrier for iPS cell reprogramming', *Nature*, 460: 1136-9.

62. Li, Wenbo, Dimple Notani, Qi Ma, Bogdan Tanasa, Esperanza Nunez, Aaron Yun Chen, Daria Merkurjev, Jie Zhang, Kenneth Ohgi, Xiaoyuan Song, Soohwan Oh, Hong-Sook Kim, Christopher K. Glass, and Michael G. Rosenfeld. 2013. 'Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation', *Nature*, 498: 516-20.

63. Li, W., D. Notani, and M. G. Rosenfeld. 2016. 'Enhancers as non-coding RNA transcription units: recent insights and future perspectives', *Nat Rev Genet*, 17: 207-23.

64. Li, Y., M. Zhang, M. Sheng, P. Zhang, Z. Chen, W. Xing, J. Bai, T. Cheng, F. C. Yang, and Y. Zhou. 2018. 'Therapeutic potential of GSK-J4, a histone demethylase KDM6B/JMJD3 inhibitor, for acute myeloid leukemia', *J Cancer Res Clin Oncol*, 144: 1065-77.

65. Lin, A. W., M. Barradas, J. C. Stone, L. van Aelst, M. Serrano, and S. W. Lowe. 1998. 'Premature senescence involving p53 and p16 is activated in response to constitutive MEK/MAPK mitogenic signaling', *Genes Dev*, 12: 3008-19.

66. Liu, Y., H. K. Sanoff, H. Cho, C. E. Burd, C. Torrice, K. L. Mohlke, J. G. Ibrahim, N. E. Thomas, and N. E. Sharpless. 2009. 'INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis', *PLoS One*, 4: e5027.

67. Liu, Y. T., L. Xu, L. Bennett, J. C. Hooks, J. Liu, Q. Zhou, P. Liem, Y. Zheng, and S. X. Skapek. 2019. 'Identification of De Novo Enhancers Activated by TGFβ to Drive Expression of CDKN2A and B in HeLa Cells', *Mol Cancer Res*, 17: 1854-66.

68. Long, Yicheng, Taeyoung Hwang, Anne R. Gooding, Karen J. Goodrich, John L. Rinn, and Thomas R. Cech. 2020. 'RNA is essential for PRC2 chromatin occupancy

and function in human pluripotent stem cells', *Nature Genetics*, 52: 931-38.

69. López, F., T. Sampedro, J. L. Llorente, M. Hermsen, and C. Álvarez-Marcos. 2017. 'Alterations of p14 (ARF) , p15 (INK4b) , and p16 (INK4a) Genes in Primary Laryngeal Squamous Cell Carcinoma', *Pathol Oncol Res*, 23: 63-71.

70. Ma, W., J. Qiao, J. Zhou, L. Gu, and D. Deng. 2020. 'Characterization of novel LncRNA P14AS as a protector of ANRIL through AUF1 binding in human cells', *Mol Cancer*, 19: 42.

71. Maertens, Goedele N., Selma El Messaoudi-Aubert, Tomas Racek, Julie K. Stock, James Nicholls, Marc Rodriguez-Niedenführ, Jesus Gil, and Gordon Peters. 2009. 'Several Distinct Polycomb Complexes Regulate and Co-Localize on the INK4a Tumor Suppressor Locus', *PLOS ONE*, 4: e6380.

72. Maggi, Leonard B., Crystal L. Winkeler, Alexander P. Miceli, Anthony J. Apicelli, Suzanne N. Brady, Michael J. Kuchenreuther, and Jason D. Weber. 2014. 'ARF tumor suppression in the nucleolus', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842: 831-39.

73. Martin, N., N. Popov, F. Aguilo, A. O'Loghlen, S. Raguz, A. P. Snijders, G. Dharmalingam, S. Li, E. Thymiakou, T. Carroll, B. B. Zeisig, C. W. So, G. Peters, V. Episkopou, M. J. Walsh, and J. Gil. 2013. 'Interplay between Homeobox proteins and Polycomb repressive complexes in p16INK⁴a regulation', *Embo j*, 32: 982-95.

74. McHugh, D., and J. Gil. 2018. 'Senescence and aging: Causes, consequences, and therapeutic avenues', *J Cell Biol*, 217: 65-77.

75. McLaughlin-Drubin, M. E., D. Park, and K. Munger. 2013. 'Tumor suppressor p16INK4A is necessary for survival of cervical carcinoma cell lines', *Proc Natl Acad Sci U S A*, 110: 16175-80.

76. Melnikov, Alexandre, Anand Murugan, Xiaolan Zhang, Tiberiu Tesileanu, Li Wang, Peter Rogov, Soheil Feizi, Andreas Gnirke, Curtis G. Callan, Justin B.

Kinney, Manolis Kellis, Eric S. Lander, and Tarjei S. Mikkelsen. 2012. 'Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay', *Nature Biotechnology*, 30: 271-77.

77. Mirzayans, R., B. Andrais, G. Hansen, and D. Murray. 2012. 'Role of p16(INK4A) in Replicative Senescence and DNA Damage-Induced Premature Senescence in p53-Deficient Human Cells', *Biochem Res Int*, 2012: 951574.

78. Montes, M., M. M. Nielsen, G. Maglieri, A. Jacobsen, J. Højfeldt, S. Agrawal-Singh, K. Hansen, K. Helin, H. J. G. van de Werken, J. S. Pedersen, and A. H. Lund. 2015. 'The lncRNA MIR31HG regulates p16(INK4A) expression to modulate senescence', *Nat Commun*, 6: 6967.

79. Moorthy, S. D., S. Davidson, V. M. Shchuka, G. Singh, N. Malek-Gilani, L. Langroudi, A. Martchenko, V. So, N. N. Macpherson, and J. A. Mitchell. 2017. 'Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes', *Genome Res*, 27: 246-58.

80. Mosteiro, L., C. Pantoja, A. de Martino, and M. Serrano. 2018. 'Senescence promotes in vivo reprogramming through p16(INK)(4a) and IL-6', *Aging Cell*, 17.

81. Münger, K., M. Scheffner, J. M. Huibregtse, and P. M. Howley. 1992. 'Interactions of HPV E6 and E7 oncoproteins with tumour suppressor gene products', *Cancer Surv*, 12: 197-217.

82. Muniz, L., S. Lazorthes, M. Delmas, J. Ouvrard, M. Aguirrebengoa, D. Trouche, and E. Nicolas. 2021. 'Circular ANRIL isoforms switch from repressors to activators of p15/CDKN2B expression during RAF1 oncogene-induced senescence', *RNA Biol*, 18: 404-20.

83. Negishi, M., A. Saraya, S. Mochizuki, K. Helin, H. Koseki, and A. Iwama. 2010. 'A novel zinc finger protein Zfp277 mediates transcriptional repression of the Ink4a/arf locus through polycomb repressive complex 1', *PLoS One*, 5: e12373.

84. Nikpay, M., A. Goel, H. H. Won, L. M. Hall, C. Willenborg, S. Kanoni, D. Saleheen, T. Kyriakou, C. P. Nelson, J. C. Hopewell, T. R. Webb, L. Zeng, A. Dehghan, M. Alver, S. M. Armasu, K. Auro, A. Bjonnes, D. I. Chasman, S. Chen, I. Ford, N. Franceschini, C. Gieger, C. Grace, S. Gustafsson, J. Huang, S. J. Hwang, Y. K. Kim, M. E. Kleber, K. W. Lau, X. Lu, Y. Lu, L. P. Lyytikäinen, E. Mihailov, A. C. Morrison, N. Pervjakova, L. Qu, L. M. Rose, E. Salfati, R. Saxena, M. Scholz, A. V. Smith, E. Tikkanen, A. Uitterlinden, X. Yang, W. Zhang, W. Zhao, M. de Andrade, P. S. de Vries, N. R. van Zuydam, S. S. Anand, L. Bertram, F. Beutner, G. Dedoussis, P. Frossard, D. Gauguier, A. H. Goodall, O. Gottesman, M. Haber, B. G. Han, J. Huang, S. Jalilzadeh, T. Kessler, I. R. König, L. Lannfelt, W. Lieb, L. Lind, C. M. Lindgren, M. L. Lokki, P. K. Magnusson, N. H. Mallick, N. Mehra, T. Meitinger, F. U. Memon, A. P. Morris, M. S. Nieminen, N. L. Pedersen, A. Peters, L. S. Rallidis, A. Rasheed, M. Samuel, S. H. Shah, J. Sinisalo, K. E. Stirrups, S. Trompet, L. Wang, K. S. Zaman, D. Ardissino, E. Boerwinkle, I. B. Borecki, E. P. Bottinger, J. E. Buring, J. C. Chambers, R. Collins, L. A. Cupples, J. Danesh, I. Demuth, R. Elosua, S. E. Epstein, T. Esko, M. F. Feitosa, O. H. Franco, M. G. Franzosi, C. B. Granger, D. Gu, V. Gudnason, A. S. Hall, A. Hamsten, T. B. Harris, S. L. Hazen, C. Hengstenberg, A. Hofman, E. Ingelsson, C. Iribarren, J. W. Jukema, P. J. Karhunen, B. J. Kim, J. S. Kooner, I. J. Kullo, T. Lehtimäki, R. J. F. Loos, O. Melander, A. Metspalu, W. März, C. N. Palmer, M. Perola, T. Quertermous, D. J. Rader, P. M. Ridker, S. Ripatti, R. Roberts, V. Salomaa, D. K. Sanghera, S. M. Schwartz, U. Seedorf, A. F. Stewart, D. J. Stott, J. Thiery, P. A. Zalloua, C. J. O'Donnell, M. P. Reilly, T. L. Assimes, J. R. Thompson, J. Erdmann, R. Clarke, H. Watkins, S. Kathiresan, R. McPherson, P. Deloukas, H. Schunkert, N. J. Samani, and M. Farrall. 2015. 'A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease', *Nat Genet*, 47: 1121-30.

85. Novo, C. L., B. M. Javierre, J. Cairns, A. Segonds-Pichon, S. W. Wingett, P. Freire-Pritchett, M. Furlan-Magaril, S. Schoenfelder, P. Fraser, and P. J. Rugg-Gunn. 2018. 'Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition', *Cell Rep*, 22: 2615-27.

86. Papp, B., and K. Plath. 2011. 'Reprogramming to pluripotency: stepwise resetting

of the epigenetic landscape', *Cell Res*, 21: 486-501.

87. Pauck, A., B. Lener, M. Hoell, A. Kaiser, A. M. Kaufmann, W. Zwerschke, and P. Jansen-Dürr. 2014. 'Depletion of the cdk inhibitor p16INK4a differentially affects proliferation of established cervical carcinoma cells', *J Virol*, 88: 5256-62.

88. Peng, Yanling, and Yubo Zhang. 2018. 'Enhancer and super-enhancer: Positive regulators in gene transcription', *Animal Models and Experimental Medicine*, 1: 169-79.

89. Pherson, M., Z. Misulovin, M. Gause, and D. Dorsett. 2019. 'Cohesin occupancy and composition at enhancers and promoters are linked to DNA replication origin proximity in Drosophila', *Genome Res*, 29: 602-12.

90. Pickar-Oliver, A., and C. A. Gersbach. 2019. 'The next generation of CRISPR-Cas technologies and applications', *Nat Rev Mol Cell Biol*, 20: 490-507.

91. Pickar-Oliver, Adrian, and Charles A. Gersbach. 2019. 'The next generation of CRISPR–Cas technologies and applications', *Nature Reviews Molecular Cell Biology*, 20: 490-507.

92. Pnueli, L., S. Rudnizky, Y. Yosefzon, and P. Melamed. 2015. 'RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α-subunit gene', *Proc Natl Acad Sci U S A*, 112: 4369-74.

93. Pomerantz, M. M., N. Ahmadiyeh, L. Jia, P. Herman, M. P. Verzi, H. Doddapaneni, C. A. Beckwith, J. A. Chan, A. Hills, M. Davis, K. Yao, S. M. Kehoe, H. J. Lenz, C. A. Haiman, C. Yan, B. E. Henderson, B. Frenkel, J. Barretina, A. Bass, J. Tabernero, J. Baselga, M. M. Regan, J. R. Manak, R. Shivdasani, G. A. Coetzee, and M. L. Freedman. 2009. 'The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer', *Nat Genet*, 41: 882-4.

94. Pott, Sebastian, and Jason D. Lieb. 2015. 'What are super-enhancers?', *Nature Genetics*, 47: 8-12.

95. Price, J. D., K. Y. Park, J. Chen, R. D. Salinas, M. J. Cho, A. R. Kriegstein, and D. A. Lim. 2014. 'The Ink4a/Arf locus is a barrier to direct neuronal transdifferentiation', *J Neurosci*, 34: 12560-7.

96. Prieto, L. I., and D. J. Baker. 2019. 'Cellular Senescence and the Immune System in Cancer', *Gerontology*, 65: 505-12.

97. Puvvula, P. K. 2019. 'LncRNAs Regulatory Networks in Cellular Senescence', *Int J Mol Sci*, 20.

98. Rahnamoun, H., J. Lee, Z. Sun, H. Lu, K. M. Ramsey, E. A. Komives, and S. M. Lauberth. 2018. 'RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation', *Nat Struct Mol Biol*, 25: 687-97.

99. Raisner, R., S. Kharbanda, L. Jin, E. Jeng, E. Chan, M. Merchant, P. M. Haverty, R. Bainer, T. Cheung, D. Arnott, E. M. Flynn, F. A. Romero, S. Magnuson, and K. E. Gascoigne. 2018. 'Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation', *Cell Rep*, 24: 1722-29.

100. Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. 2014. 'A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping', *Cell*, 159: 1665-80.

101. Rayess, H., M. B. Wang, and E. S. Srivatsan. 2012. 'Cellular senescence and tumor suppressor gene p16', *Int J Cancer*, 130: 1715-25.

102. Romagosa, C., S. Simonetti, L. López-Vicente, A. Mazo, M. E. Lleonart, J. Castellvi, and S. Ramon y Cajal. 2011. 'p16Ink4a overexpression in cancer: a tumor suppressor gene associated with senescence and high-grade tumors', *Oncogene*, 30: 2087-97.

103. Rossetto, C. C., and G. Pari. 2012. 'KSHV PAN RNA associates with

demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome', *PLoS Pathog*, 8: e1002680.

104. Russo, A. A., L. Tong, J. O. Lee, P. D. Jeffrey, and N. P. Pavletich. 1998. 'Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a', *Nature*, 395: 237-43.

105. Sabari, B. R., A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, Cisse, II, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, and R. A. Young. 2018. 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, 361.

106. Sang, Y., J. Tang, S. Li, L. Li, X. Tang, C. Cheng, Y. Luo, X. Qian, L. M. Deng, L. Liu, and X. B. Lv. 2016. 'LncRNA PANDAR regulates the G1/S transition of breast cancer cells by suppressing p16(INK4A) expression', *Sci Rep*, 6: 22366.

107. Schaukowitch, K., J. Y. Joo, X. Liu, J. K. Watts, C. Martinez, and T. K. Kim. 2014. 'Enhancer RNA facilitates NELF release from immediate early genes', *Mol Cell*, 56: 29-42.

108. Schultz, D. C., K. Ayyanathan, D. Negorev, G. G. Maul, and F. J. Rauscher, 3rd. 2002. '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*, 16: 919-32.

109. Schunkert, H., I. R. König, S. Kathiresan, M. P. Reilly, T. L. Assimes, H. Holm, M. Preuss, A. F. Stewart, M. Barbalic, C. Gieger, D. Absher, Z. Aherrahrou, H. Allayee, D. Altshuler, S. S. Anand, K. Andersen, J. L. Anderson, D. Ardissino, S. G. Ball, A. J. Balmforth, T. A. Barnes, D. M. Becker, L. C. Becker, K. Berger, J. C. Bis, S. M. Boekholdt, E. Boerwinkle, P. S. Braund, M. J. Brown, M. S. Burnett, I. Buysschaert, J. F. Carlquist, L. Chen, S. Cichon, V. Codd, R. W. Davies, G. Dedoussis, A. Dehghan, S. Demissie, J. M. Devaney, P. Diemert, R. Do, A.

Doering, S. Eifert, N. E. Mokhtari, S. G. Ellis, R. Elosua, J. C. Engert, S. E. Epstein, U. de Faire, M. Fischer, A. R. Folsom, J. Freyer, B. Gigante, D. Girelli, S. Gretarsdottir, V. Gudnason, J. R. Gulcher, E. Halperin, N. Hammond, S. L. Hazen, A. Hofman, B. D. Horne, T. Illig, C. Iribarren, G. T. Jones, J. W. Jukema, M. A. Kaiser, L. M. Kaplan, J. J. Kastelein, K. T. Khaw, J. W. Knowles, G. Kolovou, A. Kong, R. Laaksonen, D. Lambrechts, K. Leander, G. Lettre, M. Li, W. Lieb, C. Loley, A. J. Lotery, P. M. Mannucci, S. Maouche, N. Martinelli, P. P. McKeown, C. Meisinger, T. Meitinger, O. Melander, P. A. Merlini, V. Mooser, T. Morgan, T. W. Mühleisen, J. B. Muhlestein, T. Münzel, K. Musunuru, J. Nahrstaedt, C. P. Nelson, M. M. Nöthen, O. Olivieri, R. S. Patel, C. C. Patterson, A. Peters, F. Peyvandi, L. Qu, A. A. Quyyumi, D. J. Rader, L. S. Rallidis, C. Rice, F. R. Rosendaal, D. Rubin, V. Salomaa, M. L. Sampietro, M. S. Sandhu, E. Schadt, A. Schäfer, A. Schillert, S. Schreiber, J. Schrezenmeir, S. M. Schwartz, D. S. Siscovick, M. Sivananthan, S. Sivapalaratnam, A. Smith, T. B. Smith, J. D. Snoep, N. Soranzo, J. A. Spertus, K. Stark, K. Stirrups, M. Stoll, W. H. Tang, S. Tennstedt, G. Thorgeirsson, G. Thorleifsson, M. Tomaszewski, A. G. Uitterlinden, A. M. van Rij, B. F. Voight, N. J. Wareham, G. A. Wells, H. E. Wichmann, P. S. Wild, C. Willenborg, J. C. Witteman, B. J. Wright, S. Ye, T. Zeller, A. Ziegler, F. Cambien, A. H. Goodall, L. A. Cupples, T. Quertermous, W. März, C. Hengstenberg, S. Blankenberg, W. H. Ouwehand, A. S. Hall, P. Deloukas, J. R. Thompson, K. Stefansson, R. Roberts, U. Thorsteinsdottir, C. J. O'Donnell, R. McPherson, J. Erdmann, and N. J. Samani. 2011. 'Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease', *Nat Genet*, 43: 333-8.

110. Sharpless, N. E. 2005. 'INK4a/ARF: a multifunctional tumor suppressor locus', *Mutat Res*, 576: 22-38.

111. Sherr, C. J. 2012. 'Ink4-Arf locus in cancer and aging', *Wiley Interdiscip Rev Dev Biol*, 1: 731-41.

112. Shields, B. J., J. T. Jackson, D. Metcalf, W. Shi, Q. Huang, A. L. Garnham, S. P. Glaser, D. Beck, J. E. Pimanda, C. W. Bogue, G. K. Smyth, W. S. Alexander, and M. P. McCormack. 2016. 'Acute myeloid leukemia requires Hhex to enable PRC2-mediated epigenetic repression of Cdkn2a', *Genes Dev*, 30: 78-91.

113. Shlyueva, Daria, Gerald Stampfel, and Alexander Stark. 2014. 'Transcriptional enhancers: from properties to genome-wide predictions', *Nature Reviews Genetics*, 15: 272-86.

114. Sigova, A. A., B. J. Abraham, X. Ji, B. Molinie, N. M. Hannett, Y. E. Guo, M. Jangi, C. C. Giallourakis, P. A. Sharp, and R. A. Young. 2015. 'Transcription factor trapping by RNA in gene regulatory elements', *Science*, 350: 978-81.

115. Sotelo, J., D. Esposito, M. A. Duhagon, K. Banfield, J. Mehalko, H. Liao, R. M. Stephens, T. J. Harris, D. J. Munroe, and X. Wu. 2010. 'Long-range enhancers on 8q24 regulate c-Myc', *Proc Natl Acad Sci U S A*, 107: 3001-5.

116. Spitz, F., and E. E. Furlong. 2012. 'Transcription factors: from enhancer binding to developmental control', *Nat Rev Genet*, 13: 613-26.

117. Sreeramaneni, R., A. Chaudhry, M. McMahon, C. J. Sherr, and K. Inoue. 2005. 'Ras-Raf-Arf signaling critically depends on the Dmp1 transcription factor', *Mol Cell Biol*, 25: 220-32.

118. Svotelis, A., S. Bianco, J. Madore, G. Huppé, A. Nordell-Markovits, A. M. Mes-Masson, and N. Gévry. 2011. 'H3K27 demethylation by JMJD3 at a poised enhancer of anti-apoptotic gene BCL2 determines ERα ligand dependency', *Embo j*, 30: 3947-61.

119. Tena, Juan J., and José M. Santos-Pereira. 2021. 'Topologically Associating Domains and Regulatory Landscapes in Development, Evolution and Disease', *Frontiers in Cell and Developmental Biology*, 9.

120. Thiecke, M. J., G. Wutz, M. Muhar, W. Tang, S. Bevan, V. Malysheva, R. Stocsits, T. Neumann, J. Zuber, P. Fraser, S. Schoenfelder, J. M. Peters, and M. Spivakov. 2020. 'Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers', *Cell Rep*, 32: 107929.

121. Thurman, R. E., E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B. K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos. 2012. 'The accessible chromatin landscape of the human genome', *Nature*, 489: 75-82.

122. Tuupanen, S., M. Turunen, R. Lehtonen, O. Hallikas, S. Vanharanta, T. Kivioja, M. Björklund, G. Wei, J. Yan, I. Niittymäki, J. P. Mecklin, H. Järvinen, A. Ristimäki, M. Di-Bernardo, P. East, L. Carvajal-Carmona, R. S. Houlston, I. Tomlinson, K. Palin, E. Ukkonen, A. Karhu, J. Taipale, and L. A. Aaltonen. 2009. 'The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling', *Nat Genet*, 41: 885-90.

123. Tzatsos, A., R. Pfau, S. C. Kampranis, and P. N. Tsichlis. 2009. 'Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus', *Proc Natl Acad Sci U S A*, 106: 2641-6.

124. van de Werken, H. J., G. Landan, S. J. Holwerda, M. Hoichman, P. Klous, R. Chachik, E. Splinter, C. Valdes-Quezada, Y. Oz, B. A. Bouwman, M. J. Verstegen, E. de Wit, A. Tanay, and W. de Laat. 2012. 'Robust 4C-seq data analysis to screen for regulatory DNA interactions', *Nat Methods*, 9: 969-72.

125. Vicioso-Mantis, Marta, Raquel Fueyo, Claudia Navarro, Sara Cruz-Molina, Wilfred F. J. van Ijcken, Elena Rebollo, Álvaro Rada-Iglesias, and Marian A. Martínez-Balbás. 2022. 'JMJD3 intrinsically disordered region links the 3D-genome structure to TGFβ-dependent transcription activation', *Nature Communications*, 13: 3263.

126. Visel, Axel, Yiwen Zhu, Dalit May, Veena Afzal, Elaine Gong, Catia Attanasio, Matthew J. Blow, Jonathan C. Cohen, Edward M. Rubin, and Len A. Pennacchio. 2010. 'Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice', *Nature*, 464: 409-12.

127. Wang, W., J. J. Qin, S. Voruganti, S. Nag, J. Zhou, and R. Zhang. 2015. 'Polycomb Group (PcG) Proteins and Human Cancers: Multifaceted Functions and Therapeutic Implications', *Med Res Rev*, 35: 1220-67.

128. Weber, J. D., L. J. Taylor, M. F. Roussel, C. J. Sherr, and D. Bar-Sagi. 1999. 'Nucleolar Arf sequesters Mdm2 and activates p53', *Nat Cell Biol*, 1: 20-6.

129. Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, and H. Parkinson. 2014. 'The NHGRI GWAS Catalog, a curated resource of SNP-trait associations', *Nucleic Acids Res*, 42: D1001-6.

130. Whyte, W. A., D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young. 2013. 'Master transcription factors and mediator establish super-enhancers at key cell identity genes', *Cell*, 153: 307-19.

131. Williams, K., J. Christensen, J. Rappsilber, A. L. Nielsen, J. V. Johansen, and K. Helin. 2014. 'The histone lysine demethylase JMJD3/KDM6B is recruited to p53 bound promoters and enhancer elements in a p53 dependent manner', *PLoS One*, 9: e96545.

132. Xiang, Y., Z. Zhu, G. Han, H. Lin, L. Xu, and C. D. Chen. 2007. 'JMJD3 is a histone H3K27 demethylase', *Cell Res*, 17: 850-7.

133. Xu, Hao, and Timothy R. Hoover. 2001. 'Transcriptional regulation at a distance in bacteria', *Current Opinion in Microbiology*, 4: 138-44.

134. Yang, J., S. Kantrow, J. Sai, O. E. Hawkins, M. Boothby, G. D. Ayers, E. D. Young, E. G. Demicco, A. J. Lazar, D. Lev, and A. Richmond. 2012. 'INK4a/ARF

[corrected] inactivation with activation of the NF-κB/IL-6 pathway is sufficient to drive the development and growth of angiosarcoma', *Cancer Res*, 72: 4682-95.

135. Yap, K. L., S. Li, A. M. Muñoz-Cabello, S. Raguz, L. Zeng, S. Mujtaba, J. Gil, M. J. Walsh, and M. M. Zhou. 2010. 'Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a', *Mol Cell*, 38: 662-74.

136. Yokoshi, M., and T. Fukaya. 2019. 'Dynamics of transcriptional enhancers and chromosome topology in gene regulation', *Dev Growth Differ*, 61: 343-52.

137. Zhang, G., Y. Xu, C. Zou, Y. Tang, J. Lu, Z. Gong, G. Ma, W. Zhang, and P. Jiang. 2019. 'Long noncoding RNA ARHGAP27P1 inhibits gastric cancer cell proliferation and cell cycle progression through epigenetically regulating p15 and p16', *Aging (Albany NY)*, 11: 9090-110.

138. Zhang, Tiantian, Zhuqiang Zhang, Qiang Dong, Jun Xiong, and Bing Zhu. 2020. 'Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells', *Genome Biology*, 21: 45.

139. Zhang, X., L. Liu, X. Yuan, Y. Wei, and X. Wei. 2019. 'JMJD3 in the regulation of human diseases', *Protein Cell*, 10: 864-82.

140. Zhang, Y., J. Hyle, S. Wright, Y. Shao, X. Zhao, H. Zhang, and C. Li. 2019. 'A cis-element within the ARF locus mediates repression of p16(INK4A) expression via long-range chromatin interactions', *Proc Natl Acad Sci U S A*, 116: 26644-52.

141. Zhao, W., Q. Li, S. Ayers, Y. Gu, Z. Shi, Q. Zhu, Y. Chen, H. Y. Wang, and R. F. Wang. 2013. 'Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination', *Cell*, 152: 1037-50.

142. Zheng, Y., C. Devitt, J. Liu, J. Mei, and S. X. Skapek. 2013. 'A distant, cis-acting enhancer drives induction of Arf by Tgfβ in the developing eye', *Dev Biol*, 380: 49-57.

143. Zhou, Y. G., F. Sun, and Y. F. Zhou. 2020. 'Low expression of lncRNA TUBA4B promotes proliferation and inhibits apoptosis of colorectal cancer cells via regulating P15 and P16 expressions', *Eur Rev Med Pharmacol Sci*, 24: 3023-29.

144. Zhu, J., D. Woods, M. McMahon, and J. M. Bishop. 1998. 'Senescence of human fibroblasts induced by oncogenic Raf', *Genes Dev*, 12: 2997-3007.

List of oligos used

Oligos used for qRT-PCRs:

Oligo Name	Forward Oligo (5' – 3')	Reverse Oligo (5' – 3')
p14	AACATGGTGCGCAGGTTCT	CACCAGCGTGTCCAGGAAG
p15	CGTTAAGTTTACGGCCAACG	CCATCATCATGACCTGGATCG
p16	ATGGAGCCTTCGGCTGACT	CACCAGCGTGTCCAGGAAG
Gapdh	CGCTCTCTGCTCCTCCTGTT	CCATGGTGTCTGAGCGATGT
Anril Ex7b	AGAATTCTTGATTCTTTGCTTTCC	TCCCTAGTTTTGAGGACTAAGCTACT
Anril Ex7-13	GAACTCCCGACCTCGTGATTCGC	CTTCGTAGGAAATTCCTAGCTCCGTAATC
Anril Ex10-13b	CTGTGGCCACCTTGGAGA	TGGCTTCCATAGCACCAACT
Anril Ex18-19	AATGAGGCTGAGAGCATGGGAGATAC	GAGATATAGGTTCCAGTCCTGGTTCTG
E8 S eRNA	CAGCCAACCCCCTGTATTGT	TGCTGGCTGAGTTGCAATAAC
E8 AS eRNA	CAATACAGGGGGTTGGCTGT	TGCTGCCCAATCAGAAGATG
E12 S eRNA	ATGTCAGGGCCAGAAGTCGT	AAGGTCACAGCCCTGAAGGA
E12 AS eRNA	CACCCTGACTTGTCCCACAG	CAGCAAACCACAATCCCACA
E17 S eRNA	GGGTTTACATCCCCAAAGCA	TGAACAGGGAGCAGGAGTGA
E17 AS eRNA	TGGGGTGACTCAGACCTTCA	GACGAGGAGGCGTTGAAAAC
E21 eRNA	TCACTGTGAGCAGGAAACGT	CACAGACACTTAGGCACACACAT

Oligos used for ChIP qPCRs:

Oligo Name	Forward Oligo (5' – 3')	Reverse Oligo (5' – 3')
p14TSS	ACCCAGGATATTCGGGACTCACTGAC	CGTCTCTAGCCCAGGCTAGGAGG
p15TSS	CCGTCGTCCTTCTGCGGCTTG	AGTGAGGACTCCGCGACGCGT
p16TSS	TCGCCAGGAGGAGGTCTGTGATTAC	CAGGTGGGTAGAGGGTCTGCAGC
E8	GCTGAGGCAAGGGGACATACCAAACAC	GCTCACAAGCTACAGATATGCTGGCTGAG
E12	GGAACTAGAGGTAGTCCTGGCTACTTGGG	CACCTCACCCTATCTTGAAGGCAGGCCACACT
E17	GAATGGCAATTGCGGCAACCATG	GTATCGTCTCCTTCCTCCACAATCC
E21	TGGGTCCTATATAAACCTTCTTC	CACAGACACTTAGGCACACACAT

Oligos used for Surveyor Assay:

Oligo Name	Forward Oligo (5' – 3')	Reverse Oligo (5' – 3')
E8	TACTGTTGGAAGGATCCGTTAGC	CCCAAAACCATGTAGAGAGCATC
E12	GCTATTAGGATGGCCAATGATC	CTAGCGCAATACCACAGTGAACAT
E17	TGACACTGCCAATCAGTTGTAGG	AGGGGACTAAAGAAGACTCCACA
E21	CTGCCACGATATTTAGCAATC	GCTAGATGTTGCTGTGATGCT

Oligos used for 4C:

CDKN2A Viewpoint	TGGGAGGAGCTAGGGCAAGCTT
E12 Viewpoint	AAAGAGGTGAACTAAGCTT

List of sgRNAs

sgRNAs used for CRISPRi:

sgRNA Name	sgRNA Sequence
E8 gRNA 1	AGTGTTGCCCTGCTAAGATC
E8 gRNA 2	CACATATCCCAACTATGACT
E12 gRNA 1	CGTGGAGTCTAGCCATGTCA
E12 gRNA 2	GTGAGGTGTTTTATGACCAC
E17 gRNA 1	TGGAACTTATTCTAGGGCGT
E17 gRNA 2	GCCCTCACTGCTACAACTGC
E21 gRNA 1	ATACATCAACAGAAAGAAGA
E21 gRNA 2	GGACCTCAACTCACACATGC

sgRNAs used for enhancer KOs:

sgRNA Name	sgRNA Sequence
E8 KO gRNA 1	AACTGATCGTTTCAAAGCCG
E8 KO gRNA 2	ATGGCATTGCCATATCGTGG
E12 KO gRNA 1	CGTAAACAATGACAACGGAA
E12 KO gRNA 2	ATCTTGCTTACCTCTGCGAG

E17 KO gRNA 1	GATGTGGGTTAGCGTTTCAG
E17 KO gRNA 2	TAGTAACAAGGCATCTCATG
E21 KO gRNA 1	CCCAAATCCAAGAGTAGAGC
E21 KO gRNA 2	TGTTACAGCCTCCCACTGAT

4C reads and experiments

Sample Name	Total Reads	Aligned Reads	Alignment Percentage
E9_vp_E8_ED_rep1	8894272	6159675	69.25440328
E9_vp_E8_ED_rep2	5321831	4228668	79.458893
E9_vp_E8_WT_rep1	7534258	5463650	72.5174264
E9_vp_E8_WT_rep2	7395234	4645476	62.81716035
Promoter_HeLa_rep1	6796753	6176852	90.87945376
Promoter_HeLa_rep2	4587697	4156286	90.59634932

List of shRNAs

shRNA Name	shRNA Sequence
F shRNA1 E8 S	CCGGGTGATGCCACACCTCAGAAATCTCGAGATTTCTG AGGTGTGGCATCACTTTTTG
R shRNA1 E8 S	AATTCAAAAAGTGATGCCACACCTCAGAAATCTCGAG ATTTCTGAGGTGTGGCATCAC
F shRNA2 E86 S	CCGGTGCGTTCCAGTGACGGTTATTCTCGAGAATAACC GTCACTGGAACGCATTTTTG
R shRNA2 E8 S	AATTCAAAAATGCGTTCCAGTGACGGTTATTCTCGAGA ATAACCGTCACTGGAACGCA
F shRNA3 E8 S	CCGGACCTAGATGGCATGCTATAAACTCGAGTTTATAG CATGCCATCTAGGTTTTTTG
R shRNA3 E8 S	AATTCAAAAAACCTAGATGGCATGCTATAAACTCGAGT TTATAGCATGCCATCTAGGT
F shRNA1 E8 AS	CCGGTGCATTGCACTAAGCAAATAACTCGAGTTATTTG CTTAGTGCAATGCATTTTTG
R shRNA1 E8 AS	AATTCAAAAATGCATTGCACTAAGCAAATAACTCGAGT TATTTGCTTAGTGCAATGCA
F shRNA2 E8 AS	CCGGGGCTGTCTACTCAAATAAATTCTCGAGAATTTAT TTGAGTAGACAGCCTTTTTG
R shRNA2 E8 AS	AATTCAAAAAGGCTGTCTACTCAAATAAATTCTCGAGA ATTTATTTGAGTAGACAGCC
F shRNA3 E8 AS	CCGGCTAAACTCTTCCAGCTATATACTCGAGTATATAG CTGGAAGAGTTTAGTTTTTG
R shRNA3 E8 AS	AATTCAAAAACTAAACTCTTCCAGCTATATACTCGAGT ATATAGCTGGAAGAGTTTAG
F shRNA1 E12 S	CCGGCACAGGCTTTATGAGTTATAGCTCGAGCTATAAC TCATAAAGCCTGTGTTTTTG

R shRNA1 E12 S	AATTCAAAAACACAGGCTTTATGAGTTATAGCTCGAGC TATAACTCATAAAGCCTGTG
F shRNA2 E12 S	CCGGAGAATGACCTTCATGCTATTTCTCGAGAAATAGC ATGAAGGTCATTCTTTTTTG
R shRNA2 E12 S	AATTCAAAAAAGAATGACCTTCATGCTATTTCTCGAGA AATAGCATGAAGGTCATTCT
F shRNA3 E12 S	CCGGTACAGCTATGCTCACAAATATCTCGAGATATTTG TGAGCATAGCTGTATTTTTG
R shRNA3 E12 S	AATTCAAAAATACAGCTATGCTCACAAATATCTCGAGA TATTTGTGAGCATAGCTGTA
F shRNA1 E12 AS	CCGGAGGTTAAGTGATGAGAAATTACTCGAGTAATTTC TCATCACTTAACCTTTTTTG
R shRNA1 E12 AS	AATTCAAAAAAGGTTAAGTGATGAGAAATTACTCGAG TAATTTCTCATCACTTAACCT
F shRNA2 E12 AS	CCGGCCAGTGACACTCCACAAATTGCTCGAGCAATTTG TGGAGTGTCACTGGTTTTTG
R shRNA2 E12 AS	AATTCAAAAACCAGTGACACTCCACAAATTGCTCGAGC AATTTGTGGAGTGTCACTGG
F shRNA3 E12 AS	CCGGATATTTATTGCCTGCAATTTGCTCGAGCAAATTG CAGGCAATAAATATTTTTTG
R shRNA3 E12 AS	AATTCAAAAAATATTTATTGCCTGCAATTTGCTCGAGC AAATTGCAGGCAATAAATAT
F shRNA1 E17 S	CCGGTTAGAGGTTGCTGGAACTTATCTCGAGATAAGTT CCAGCAACCTCTAATTTTTG
R shRNA1 E17 S	AATTCAAAAATTAGAGGTTGCTGGAACTTATCTCGAGA TAAGTTCCAGCAACCTCTAA
F shRNA2 E17 S	CCGGCCAAAGCAAACCTTAACTAATCTCGAGATTAGTT AAGGTTTGCTTTGGTTTTTG
R shRNA2 E17 S	AATTCAAAAACCAAAGCAAACCTTAACTAATCTCGAG ATTAGTTAAGGTTTGCTTTGG
F shRNA1 E17 AS	CCGGTGTGACTTAACACACAAATATCTCGAGATATTTG TGTGTTAAGTCACATTTTTG
R shRNA1 E17 AS	AATTCAAAAATGTGACTTAACACACAAATATCTCGAGA TATTTGTGTGTTAAGTCACA
F shRNA2 E17 AS	CCGGCTTCCTCCACAATCCATATTACTCGAGTAATATG GATTGTGGAGGAAGTTTTTG
R shRNA2 E17 AS	AATTCAAAAACTTCCTCCACAATCCATATTACTCGAGT AATATGGATTGTGGAGGAAG

Cloning Oligos for eRNAs

Oligo Name	Sequence
E8 S eRNA F	AGACTGGCGGCCGCCAGAAGATGTGATGCCACACCTC
E8 S eRNA R	AGACTGCTCGAGGCTGCCCTGTCAGAAGTTCTGC
E8 AS eRNA F	AGACTGGCGGCCGCGCCTCCACGATATGGCAATG
E8 AS eRNA R	AGACTGGCGGCCGCATCTGTTCCTTGCCTCTTCCAC
E12 S eRNA F	AGACTGGAATTCCCAGAAGTCGTCTCAGCTAAG
E12 S eRNA R	AGACTGGCGGCCGCAGGAGAATGGCGTGAACTC
E12 AS eRNA F	AGACTGGCGGCCGCCACCTCACCCTATCTTGAAGGCA
E12 AS eRNA R	AGACTGTCTAGATGGAACTGAGGGAGGTCTCGA

E17 S eRNA F	AGACTGGAATTCGACGAGGAGGCGTTGAAA
E17 S eRNA R	AGACTGTCTAGAGTTACCTCCCATGCATCCT
E17 AS eRNA F	AGACTGGCGGCCGCTGTACCACACAGATACACCTACGCC
E17 AS eRNA R	AGACTGTCTAGAGAGAGCAATGTCCAACCAATGAGG

JMJD3 Variant cloning Oligos

Oligo Name	Sequence
JMJD3 Jmjc F	ATCTAAAGCTTCGGTGGAAGCCCCAGCTG
JMJD3 Jmjc R	ATCTATCTAGACCACTCGTATCGTTCCAGGG
JMJD3 CT F	ATCTAAAGCTTAATGAGGTGAAGAACGTCAAATCC
JMJD3 CT R	ATCTATCTAGATCGCGACGTGCTGGCTGG
ΔHI JMJD3 F	AGTCTAAGCTTAAGATGCTGGACGAATCCATTCGC
ΔHI JMJD3 R	AGTCTTCTAGATCATCGCGACGTGCTGGCTG