UNIVERSITY OF CALIFORNIA,
IRVINE

Cohesin and its Loading Factor NIPBL in Genetic Diseases

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Yen-Yun Chen

Dissertation Committee:
Professor Kyoko Yokomori, Chair
Professor Paolo Sassone-Corsi
Professor Bogi Andersen
Associate Professor Yongsheng Shi

2014

DEDICATION

To

my parents and friends

for their support along this journey.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	vi
ACKNOWLEDGEMENTS	vii
CURRICULUM VITAE	viii
ABSTRACT OF THE DISSERTATION	ix
CHAPTER 1: Introduction	**1**
CHAPTER 2: NIPBL haplinsufficiency and cohesin- mediated gene regulation.	**14**
Abstract	**15**
Introduction	**16**
Materials and methods	**18**
Results	**25**
Discussion	**33**
CHAPTER 3: NIPBL's function in the nucleolus	**50**
Abstract	**51**
Introduction	**52**
Materials and methods	**54**
Results	**60**
Discussion	**66**
CHAPTER 4: D4Z4 Heterochromatin and FSHD	**85**
Abstract	**86**
Introduction	**87**
Materials and methods	**90**
Results	**94**
Discussion	**98**
CHAPTER 5: Conclusions	**110**
REFERENCES	116

LIST OF FIGURES

Page

CHAPTER 1

Fig. 1-1. Schematic diagrams of SMC protein complexes **11**

Fig. 1-2. Regulators of cohesin loading and establishment of sister chromatid cohesion **12**

Fig. 1-3. CTCF-dependent and -independent chromatin loop formation at the β-globin locus. **13**

CHAPTER 2

Fig. 2-1. Global decrease of cohesin binding to chromatin in *Nipbl* heterozygous mutant MEFs. **38**

Fig. 2-2. Nipbl reduction decreases cohesin binding to chromatin. **39**

Fig. 2-3. Correlation of cohesin binding and gene expression changes in mutant MEFs. **40**

Fig. 2-4. Cohesin binding signals at specific gene regions. **41**

Fig. 2-5. Cohesin plays a direct role in adipogenesis gene regulation. **42**

Fig. 2-6. The long distance interaction involving the Cebpβ promoter is decreased in Nipbl +/- MEFs. **43**

Fig. 2-7. *Nipbl* heterozygous mutation and Nipbl siRNA depletion in mES cells generate different expression patterns. **44**

CHAPTER 3

Fig. 3-1. Schematic diagram of the tripartite structure in nucleolus **70**

Fig. 3-2. Nipbl localization in nucleoli in mouse and human cells. (Immunofluorescence) **71**

Fig. 3-3. Nipbl localization in nucleoli in mouse and human cells (Biochemical analyses). **72**

Fig. 3-4. Nipbl nucleolar foci decrease in Nipbl siRNA- treated MEFs and *Nipbl* +/- MEFs. 73

Fig. 3-5. Nipbl binds to the rDNA transcribed region. 74

Fig. 3-6. Nipbl positively affects rRNA transcripts level. 75

Fig. 3-7. Nucleolar localization of Nipbl is dependent on RNA. 76

Fig. 3-8. Nipbl binds to rRNA. 77

Fig. 3-9. RNA-dependent binding of Nipbl to rDNA chromatin. 78

Fig. 3-10. Nipbl's nucleolar localization is affected by H_2O_2, INK128 and AICAR. 79

Fig. 3-11. Nipbl's nucleolar localization is affected by actinomycin D and H_2O_2. 80

Fig. 3-12. Nipbl binding to rDNA chromatin decreases upon nucleolar Stresses. 81

Fig. 3-13. Model of Nipbl's function in the nucleolus. 82

CHAPTER 4

Fig. 4-1. Model of how H3K9me3 / HP1γ / cohesin loss can regulate epigenetic status and expression at distant locations. 102

Fig. 4-2. Inhibition of H3K9me3 results in *DUX4fl* upregulation. 103

Fig. 4-3. Inhibition of H3K9me3 results in the loss of SMCHD1 binding. 104

Fig. 4-4. Schematic models of H3K9me3 heterochromatin at D4Z4. 105

Fig. 4-5. KD3 differentiation and gene expression. 106

CHAPTER 5

Fig. 5-1. Schematic diagram of NIPBL isoforms A and B. 115

LIST OF TABLES

Page

CHAPTER 2

Table 2-1. The list of PCR primers **45**

Table 2-2. Nipbl and Rad21 depletion levels in mutant and siRNA-treated MEFs **46**

Table 2-3. Gene expression changes and cohesin binding status in MEFs **46**

Table 2-4. Ontology analysis of cohesin target genes in MEFs **46**

Table 2-5. Ontology of changed genes in Rad21 siRNA and Nipbl siRNA treated mES cells **48**

Table 2-6. Ontology of changed genes in *Nipbl*+/- mES cells **48**

Table 2-7. 30 developmental related genes down-regulated in the *Nipbl*+/- mES cells **49**

CHAPTER 3

Table 3-1. The list of PCR primers **83**

Table 3-2. Antibodies and their applications in the study **84**

CHAPTER 4

Table 4-1. Myoblasts used for epigenomic analyses, and the percentage of cells that are desmin staining positive. **107**

Table 4-2. The list of PCR primers **107**

Table 4-3. Antibodies and their applications in the study **108**

Table 4-4. Antibodies used in the epigenomic analyses **108**

Table 4-5. FSHD epigenomic analyses **109**

Table 4-6. Undifferentiated and differentiated KD3 epigenomic analyses **109**

ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Kyoko Yokomori, for the opportunity to work on a very interesting thesis project. Her encouragement and advice during my graduate training allowed me to continue forward despite the many experimental obstacles.

I would also like to thank the current and past Yokomori lab members—Daniel Newkirk, Richard Chien, Wanting Chen, Benny Zeng, Xianduo Kong, Alex Ball, Ebony Flowers and John Schmiesing---for all their help, support, encouragement and friendship throughout the years.

CURRICULUM VITAE

Yen-Yun Chen

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE (*if applicable*)	YEAR(s)	FIELD OF STUDY
National Taiwan University	**B.S.**	**2002 – 2006**	**Biology**
University of California, Irvine	**Ph. D.**	**2007 – 2014**	**Biological Sciences**

Research Experience:

September 2005 – June 2006: Jun-Hon Chen Lab, National Taiwan University

September 2006 – September 2007: Carrie Brachmann Lab, UC Irvine, CA

January 2008 – present: Kyoko Yokomori Lab, UC Irvine, CA

Award:

Robert Warner Memorial Award from UCI: Outstanding Achievement in Nucleic Acid Biochemistry (2008)

Publications:

1. Zeng W, de Greef JC, **Chen YY**, Chien R, Kong X, et al. (2009) Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS Genet 5: e1000559.

2. Monserrate JP, **Chen MY**, Brachmann CB (2012) Drosophila larvae lacking the bcl-2 gene, buffy, are sensitive to nutrient stress, maintain increased basal target of rapamycin (Tor) signaling and exhibit characteristics of altered basal energy metabolism. BMC biology 10: 63.

3. Ball AR, Jr., **Chen YY**, Yokomori K (2014) Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies. Biochimica et biophysica acta 1839: 191-202.

4. Zeng W*, **Chen YY***, Newkirk DA, Wu B, Balog J, et al. (2014) Genetic and Epigenetic Characteristics of FSHD-Associated 4q and 10q D4Z4 that are Distinct from Non-4q/10q D4Z4 Homologs. Human mutation. **(Co-first author)**

5. Newkirk DA*, **Chen YY***, Chien R*, Zeng W, Biesinger J, et al. The effect of "Nipped-B-like" haploinsufficiency on genome-wide cohesin binding and target gene expression: modeling Cornelia de Lange Syndrome. (Submitted) **(Co-first author)**

ABSTRACT OF THE DISSERTATION

Cohesin and its Loading Factor NIPBL in Genetic Diseases

by

Yen-Yun Chen

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2014

Professor Kyoko Yokomori, Chair

Cohesin is a protein complex important for chromatin structural organization. NIPBL is required for cohesin loading onto chromatin. Cornelia de Lange syndrome (CdLS) is a developmental disorder caused by heterozygous *NIPBL* (60% of CdLS) or cohesin (~5-6%) mutations. How the mutations lead to developmental abnormalities was not well understood.

We used the the *Nipbl +/-* mouse cells to investigate how *Nipbl* reduction affects cohesin-mediated gene regulation. We found that development-related genes bound by cohesin preferentially changed expression in *Nipbl* mutant cells, suggesting that dysregulation of cohesin target genes directly contributes to the CdLS pathogenesis. We also found that a sub-population of Nipbl localizes in the nucleolus, which is compromised in *Nipbl* heterozygous mutant cells. Nipbl binds to both ribosomal RNA and DNA (rRNA and rDNA) and stimulates pre-rRNA synthesis. Interestingly, binding of Nipbl to rDNA is dependent on RNA, suggesting that the active rRNA synthesis promotes Nipbl recruitment resulting in further stimulation of rRNA transcription. Nucleolar Nipbl dissociates from rDNA and relocalizes to the nucleolar cap

structure in response to stress, suggesting that Nipbl is involved in stress-induced rRNA gene repression. Our results raise the possibility that defective ribosome biogenesis may also contribute to the CdLS phenotypes.

Fascioscapulohumeral muscular dystrophy (FSHD) is associated with locus-restricted dysregulation of cohesin binding at D4Z4 macrosatellite repeat sequences on chromosome 4. The upregulation of the D4Z4-encoded *DUX4* retrogene is associated with FSHD. D4Z4 repeats contraction caused FSHD1 while FSHD2 has no repeat contraction. We previously found that H3K9me3, HP1γ and cohesin form a heterochromatin structure at D4Z4, which is lost in both FSHD1 and FSHD2. How this contributes to the disease, however, was unknown. We found that reducing H3K9me3 at D4Z4 results in *DUX4* upregulation accompanied with diminished binding of SMCHD1. SMCHD1 is an SMC homolog frequently mutated in FSHD2 and was shown to repress *DUX4*. Thus, the loss of H3K9me3 at D4Z4 contributes to dissociation of SMCHD1 and *DUX4* expression.

My thesis research provided insights into the mechanisms of two cohesin-related developmental disorders, CdLS and FSHD, and may lead to development of new therapeutic strategies.

Chapter One

Introduction

Part of this chapter was published in a review article entitled-

“Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies”

Alexander R. Ball Jr., Yen-Yun Chen, Kyoko Yokomori[¶]

[¶]Corresponding author.

The cohesin complex

The structural maintenance of chromsomes (SMC) proteins is a class of protein conserved from yeast to human. They have conserved protein structure which involves a hinge domain that brings the conserved head and tail globular domains with divided ATPase motifs together. They form three major complexes along with non-SMC protein subunits into the cohesin, condensin and the SMC5–SMC6 complex (Fig.1-1). The common feature of SMC complexes is that they physically associate with chromatin and regulate higher-order chromatin structure. Cohesin consists of the SMC family proteins SMC1 (also known as SMC1A) and SMC3 as a heterodimer with the two non-SMC components Rad21 (also called Mcd1 or Scc1) and Scc3 (also called SA or STAG) [1]. SMC1 and SMC3 interact through their central hinge regions, while their respective paired amino- and carboxyl-terminal globular domains are further bridged by the kleisin family component Rad21 (or Scc1) (Fig. 1-1) [2,3].

The primary function of cohesin is to mediate genome-wide sister chromatid cohesion to ensure proper segregation of chromosomes in mitosis [4,5,6]. Cohesin forms a ring structure. Analyses of purified cohesin-circular minichromosome complexes assembled in vivo, in conjunction with various mutational manipulations of cohesin subunits, support the notion that the cohesin ring traps sister chromatids inside to mediate sister chromatid cohesion [7,8,9,10,11]. However, alternative models of DNA trapping and cohesion by cohesin are still being discussed [12], and the exact mechanism is not yet fully resolved. Recently two SA proteins, SA1 and SA2 (STAG1 and STAG2 in mice), are found in higher eukaryotes to form two distinct cohesin complexes in somatic cells: cohesin-SA1 and cohesin-SA2 [13,14]. The two cohesin complexes have distinct and redundant functions in gene regulation and maintaining genome stability [15,16,17].

The sister chromatid cohesion cycle

Human cohesin requires NIPBL (or delangin, yeast Scc2 homolog) and its partner MAU2 (yeast Scc4 homolog) for chromatin loading (Fig. 1-2) [18,19]. Cohesin loading takes place in telophase in higher eukaryotes. ESCO1 and ESCO2 (Eco1p in yeast) are acetyltransferases, and their acetylation of SMC3 is required for antagonizing Wapl and establishment of sister chromatid cohesion [20,21,22,23,24]. ESCO-mediated acetylation of SMC3 is reversed by the deacetylase Hos1 in *S. cerevisiae* and HDAC8 in human cells, which is required for the next cycle of cohesion establishment [20,25,26,27]. Wapl releases cohesin from chromatin by opening the gate between SMC3 and Rad21 [11,28,29]. Both PDS5 and sororin are important for maintenance of cohesion in vertebrates [30,31].

In higher eukaryotes, cohesin is removed from chromosomes in a two-step process during mitosis that results in chromosome separation in anaphase [32]. The first step is removal of the majority of cohesin from chromatin in prophase, which depends on Polo kinase that phosphorylates SA2 [33]. The second step is destruction of the residual cohesin remaining primarily at centromeres by separase-mediated Rad21 cleavage at the end of metaphase, which leads to chromosome segregation in anaphase [34].

Cohesin functions in gene regulation

An expanding body of literature is documenting cohesin as a key regulator of gene expression. Cohesin was shown to mediate chromatin looping at multiple gene loci important for imprinting and differential gene expression during development [35,36,37,38,39,40,41]. These interactions include CCCTC-binding factor (CTCF)-dependent insulator interaction, that blocks enhancer activity and/or inhibits the spreading of heterochromatic domains, as well as distal enhancer–promoter interactions important for gene activation. However, a significant number of cohesin sites appear to be CTCF-free and often overlap with binding sites for cell

type-specific transcription factors [42,43,44]. CTCF-free cohesin binding sites coincide significantly with enhancer elements and genes that exhibit tissue/cell type-specific patterns of expression, and cohesin appears to help stabilize transcription factor binding to these sites [42].

At the β-globin locus, both CTCF-dependent insulator interaction and CTCF-independent enhancer–promoter interactions can be observed [38]. Both types of interaction involve cohesin in mouse and human erythroid lineage cells as detected by chromatin conformation capture (3C) and 3C combined with ChIP (ChIP-loop) (Fig. 1-3) [38]. The distal enhancer in the locus control region (LCR) interacts with the developmental stage-specific globin genes, which correlates with their specific expression [45,46]. Both Nipbl and cohesin binding rapidly increase at chromatin loop anchoring sites upon cellular differentiation [38]. Depletion of either cohesin or Nipbl decreased both the insulator interaction and the LCR enhancer–promoter interaction, while CTCF depletion only affected the insulator interaction [38]. Consistent with this, cohesin depletion, but not CTCF depletion, decreased β-globin gene expression [38].

Recent studies examined cohesin-mediated chromatin interactions genome-wide using high-resolution high-throughput 3C-based techniques, and found that cohesin mediates chromatin interactions at a global scale. An SMC1 ChIA-PET study, in which chromatin interactions involving cohesin were selectively analyzed in developing mouse limb, identified over 2200 interactions at both CTCF-positive and -negative cohesin binding sites [47]. In either the promoter or intergenic/intronic regions, ~65% of chromatin interaction sites coincided with CTCF occupancy. The study revealed that in addition to tissue-specific promoter–enhancer interactions and constitutive chromatin domain demarcations, a subset of promoter–enhancer interactions reflect the poised state in embryonic stem cells (ESCs) and are maintained in multiple tissues even when the genes are not expressed.

Cohesin loading by NIPBL- MAU2 complex

In metazoans, genome-wide cohesin loading occurs at the end of mitosis during telophase, which requires Scc2 and Scc4 (NIPBL and MAU2 in human, respectively) [19,48,49]. Chromatin loading requires ATP hydrolysis [50,51,52]. In Xenopus and human cells, pre-replication complex components, including ORC, Cdc6, Cdt1, and MCM2-7, were shown to be required for loading of Scc2–Scc4 (NIPBL–MAU2) and subsequent cohesin binding to chromatin[48,49,53]. Despite the early discovery of this cohesin loading factor, the exact loading mechanism remains enigmatic. Interestingly, all three SMC complexes (cohesin, condensin, and the SMC5–SMC6 complex) independently require Scc2 in *S. cerevisiae* [54,55]. The relationship between other SMC complexes and NIPBL–MAU2 is unclear in mammalian cells.

Cohesinopathies

Human syndromes caused by cohesin and cohesin-associated factor mutations, resulting in cohesin dysfunction, are called “cohesinopathies” (Fig. 1-2) [56,57]. The two classic examples are Roberts Syndrome (RBS) and Cornelia de Lange Syndrome (CdLS).

Roberts Syndrome, or SC phocomelia (OMIM 268300), is caused by mutations of both alleles of *ESCO2* (Fig. 1-2) [58]. RBS patients have a wide range of clinical phenotypes that include upper and lower limb defects, growth retardation, craniofacial anomalies, and mental retardation with limited similarity to the CdLS phenotype [58,59]. Importantly, RBS chromosomes exhibit premature centromere separation and heterochromatin puffing, indicative of a sister chromatid cohesion defect [60]. Centromeric cohesion defects and cell cycle aberrations are observed in ESCO2 knockout mice and zebrafish [61,62].

CdLS (OMIM 122470, 300590, 610759) is a dominant multisystem developmental disorder characterized by facial dysmorphism, hirsutism, upper limb abnormalities, cognitive retardation, and growth abnormalities [63,64]. Mutations in the NIPBL gene on chromosome

5p13 are linked to more than 55% of CdLS cases (Fig. 1-2) [65,66]. Frameshift or nonsense mutations of NIPBL that result in NIPBL haploinsufficiency often exhibit more severe phenotypes compared to missense mutations [67]. Mutations in the cohesin subunits SMC1 and hSMC3 were also found in a minor subset of clinically milder CdLS cases (~5% and ~1%, respectively) [68,69]. SMC1 or SMC3 mutations are always missense mutations and patients often show mental retardation as the primary symptom, with other abnormalities being fewer and/or milder [69]. More recently, mutations in HDAC8, which regulates cohesin dissociation from chromatin in mitosis, were also found in a subset of CdLS patients (OMIM 300882) [26]. HDAC8 functions to deacetylate SMC3 and therefore facilitates cohesin displacement from chromatin during mitotic progression [26]. Nonsense or missense mutations that cause loss of HDAC8 activity resulted in SMC3 hyperacetylation and chromatin retention of the cohesin complex during mitosis [26]. CdLS patients with HDAC8 mutations display similar phenotypes as the patients with NIPBL mutations [26]. Furthermore, cohesin component Rad21 mutations were found in patients with a CdLS-like phenotype (OMIM 614701) [70]. In contrast to SMC1 and SMC3 mutations, patients with RAD21 mutations exhibit classical CdLS physical phenotypic characteristics (growth retardation,minor skeletal anomalies, and facial features) but have mild or no cognitive impairment [70]. Taken together, mutations of cohesin subunits and the regulators of cohesin loading cause phenotypically related developmental disorders [64,71].

While mutations in these proteins (NIPBL, HDAC8, SMC1A, SMC3, and possibly RAD21) may explain approximately 65% of CdLS patients, the cause of the remaining 35% remains unclear. For example, mutations in Pds5A and Pds5B, additional factors important for proper cohesin function in sister chromatid cohesion, also result in phenotypes in mouse models reminiscent of those observed in CdLS patients. However, no significant association of Pds5A or Pds5B mutations with CdLS has been observed [72,73]. Nevertheless, mutations in additional genes involved in the cohesin pathway are expected to contribute to CdLS pathogenesis.

Mechanisms of CdLS

NIPBL haploinsufficiency is the major cause of CdLS [64,74,75]. *Nipbl* heterozygous mutant (*Nipbl*+/−) mice exhibit wide-ranging defects characteristic of CdLS including small size, craniofacial anomalies, microbrachycephaly, heart defects, hearing abnormalities, low body fat, and delayed bone maturation, confirming that partial reduction of Nipbl is sufficient to cause a CdLS-like phenotype [76]. The mutant mice demonstrated only a 25–30% decrease in Nipbl transcripts, suggesting compensatory upregulation of the intact allele, which apparently is not sufficient to block development of the phenotype. Consistent with this, as little as a 15% decrease in NIPBL expression was shown to cause CdLS with mild phenotypes [77,78]. These observations indicate the extreme sensitivity of mammalian development to NIPBL/Nipbl gene dosage.

There appears to be a functional hierarchy for cohesin in which the most essential function, which is resistant to partial reduction of cohesin, is its role in sister chromatid cohesion and proper segregation of chromosomes [79]. The differential sensitivities of cohesin functions to cohesin depletion were most systematically demonstrated in yeast with different degrees of cohesin protein reduction [80]. Namely, mitotic sister chromatid cohesion is most resistant to partial reduction of cohesin. Similar observations were made in *Drosophila* and in human cells, in which partial depletion of cohesin by siRNA does not lead to any significant sister chromatid cohesion defect [81,82]. Consistent with these findings, CdLS patient cells do not exhibit any obvious sister chromatid cohesion abnormalities [83,84,85,86]. This is in contrast to RBS, in which premature sister chromatid separation serves as a prototypical cellular phenotype for the disorder [60]. Though it is currently unclear how sister chromatid cohesion defects specifically contribute to the pathogenesis of RBS, distinct mechanisms are likely involved in the development of this cohesinopathy as opposed to CdLS.

Because NIPBL mutations in both CdLS patient cells and in mouse models cause little or no chromatid cohesion defect, the developmental abnormalities are likely to be the result of defective cohesin-mediated gene regulation [76,78]. In both patient lymphoblasts and Nipbl-mutant mouse tissues and cells, the partial decrease of Nipbl expression is associated with pervasive, though small, alterations in gene expression. It was proposed that relatively mild expression perturbations collectively contribute to the developmental defective phenotypes. Supporting this model, combined depletion of Nipbl target genes indeed recapitulates the Nipbl depletion phenotype in zebrafish [87]. As noted above, it was shown that Nipbl haploinsufficiency causes both decreased cohesin binding at the *β-globin* locus in embryonic liver as well as decreased long-distance chromatin interactions. In particular, reduced chromatin interactions between the enhancer and adult globin genes appear to contribute to decreased globin gene expression [38]. One can envision that diminished cohesin-mediated long-distance chromatin interactions could affect gene regulation genome-wide, resulting in widespread disruption of normal gene expression in a cell type- and differentiation stage-specific manner.

FSHD as a new cohesinopathy?

FSHD is the third most common heritable muscular dystrophy in the U.S. It is characterized by progressive wasting of facial, shoulder, and upper arm musculature [88,89]. The genetics underlying FSHD are very unusual; the majority of FSHD cases (~95%) are associated with monoallelic deletion of D4Z4 macrosatellite repeat sequences clustered at the subtelomeric region of chromosome 4q (4qter D4Z4) (FSHD1(MIM 158900)) [88,90]. There are between one and ten repeats in the contracted 4qter allele in FSHD1 patient cells, in contrast to 11–150 copies in normal cells. In the more rare form of FSHD (b5% of cases) (FSHD2) there is no D4Z4 repeat contraction, though phenotypically FSHD1 and FSHD2 are largely identical [91].

D4Z4 is a 3.3 kb repeat that contains an open reading frame for the double-homeobox transcription factor DUX4 retrogene [92,93,94]. Artificial overexpression of the full-length DUX4 (DUX4fl) protein caused a myoblast differentiation defect in human myoblasts and mouse C2C12 cells [95,96]. Only those individuals with a 4qA haplotype with specific single-nucleotide polymorphisms in the last D4Z4 repeat (creating a canonical polyadenylation signal for the DUX4 transcript) develop FSHD, strongly suggesting that *DUX4fl* mRNA expression is critical for FSHD pathogenesis [97]. A recent study found that SMCHD1, an epigenetic gene silencer involved in the maintenance of DNA methylation and X inactivation [98,99], binds to D4Z4 and plays a role in *DUX4* gene repression [100]. Importantly, this gene is mutated in many FSHD2 patients (OMIM 158901) as well as in severe cases of FSHD1 in conjunction with D4Z4 contraction [100]. How SMCHD1 is recruited to D4Z4 remains unclear.

D4Z4 chromatin normally harbors the transcriptionally repressive histone modification marks histone H3 lysine 9 trimethylation (H3K9me3) and H3K27me3 [101]. Interestingly, H3K9me3 is significantly diminished at D4Z4 repeat regions in both FSHD1 and FSHD2 patient cells, but not in other muscular dystrophies [101,102]. This change is also found in FSHD patient lymphoblasts, indicating that the loss is not an epiphenomenon of the dystrophic phenotype and suggesting that it occurs early in development before lineage separation [101]. Loss of H3K9me3 at D4Z4 indicates that FSHD is an epigenetic abnormality disease. Cohesin and HP1γ are recruited to D4Z4 in an H3K9me3-dependent manner and are therefore lost in FSHD cells [101]. Interestingly, the two factors require each other for D4Z4 binding, demonstrating the active role of cohesin in heterochromatin organization in human cells [101]. This is analogous to the subtelomeric heterochromatin repeats in *S. pombe* in which cohesin and Swi6 are recruited in a mutually dependent manner and function in gene silencing [103]. Thus, FSHD may be considered to be a cohesinopathy, in which D4Z4 heterochromatin-associated cohesin function is specifically disrupted. It is speculated that the loss of

heterochromatin contributes to the expression of *DUX4fl* in FSHD. However, this has not been explicitly demonstrated.

NIPBL and cohesin are important chromatin organizers and influencing virtually all aspects of genomic functions. Impaired NIPBL/ cohesin pathways thus can have huge impact on developmental processes and therefore causing diseases. The goal of my thesis project is to understand how impaired NIPBL/ cohesin pathways cause genetic diseases. The first part of my thesis was the investigation of how NIPBL haploinsufficiency affects cohesin's gene-regulatory function; second part was about NIPBL's nucleolar functions and its relevance to CdLS pathogensis; the last part was to understand the role of D4Z4 heterochromatin in FSHD.

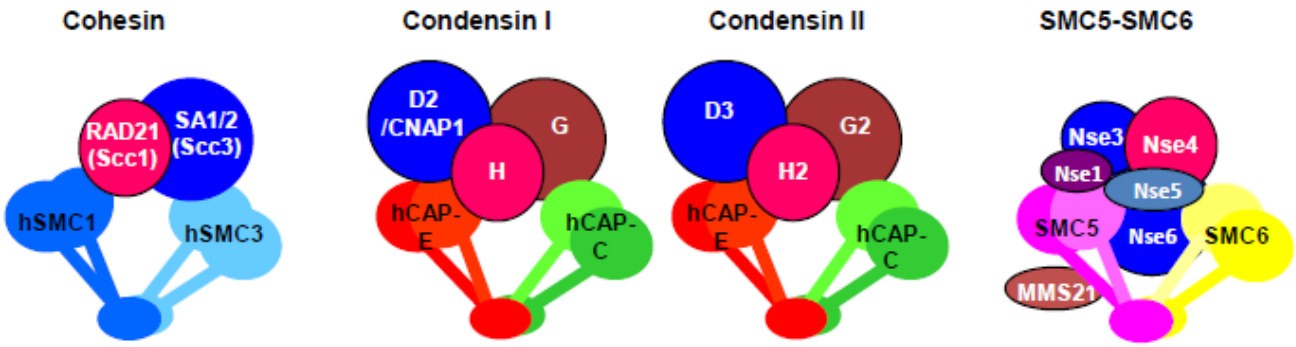


Figure 1-1 Schematic diagrams of SMC protein complexes

The SMC protein complexes are cohesin, condensin I , condesin II and the SMC5/6 complex. SMC protein complexes are formed by the SMC heterodimers (SMC1 with SMC3, CAP-E with CAP-C, and SMC5 with SMC6) and non- SMC protein subunits. All SMC complexes functions in regulating chromatin structure of different aspects.

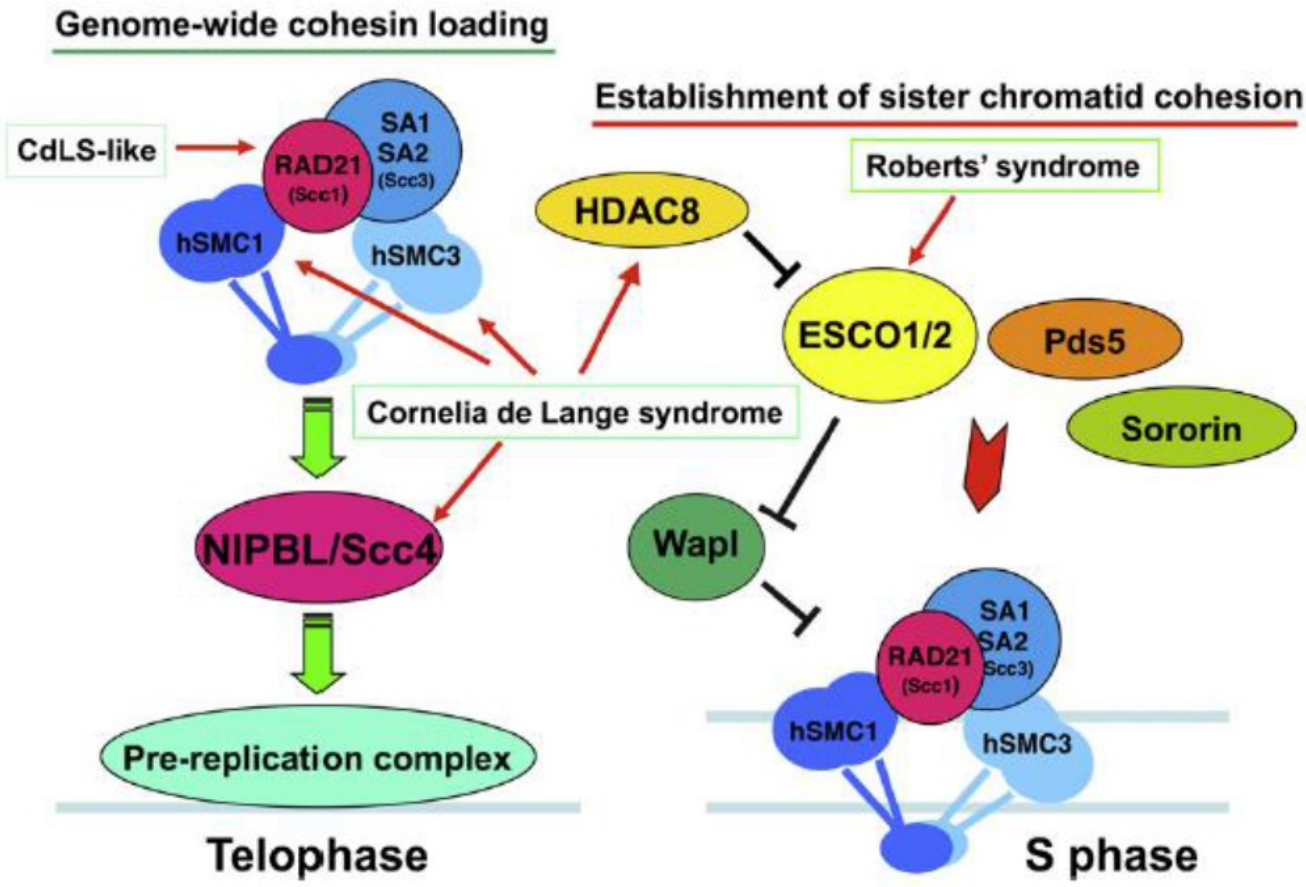


Figure 1-2. Regulators of cohesin loading and establishment of sister chromatid cohesion

The cohesin loading factor, NIPBL–MAU2 (Scc2–Scc4), is required for cohesin loading onto chromatin in telophase in mammalian cells. The initial loading of NIPBL–MAU2 is dependent on the pre-replication machinery. In S phase, the establishment of sister chromatid cohesion requires sororin and Pds5A/B as well as the ESCO1/2 (Eco) acetyltransferases that coordinately antagonize the activity of the cohesin destabilizing factor Wapl. ESCO-mediated acetylation of SMC3 must be reversed by histone deacetylase HDAC8 in order to refresh and recycle cohesin for the subsequent cell cycle. Mutations associated with the cohesinopathies RBS and CdLS are indicated.

(The figure was published in [104].)

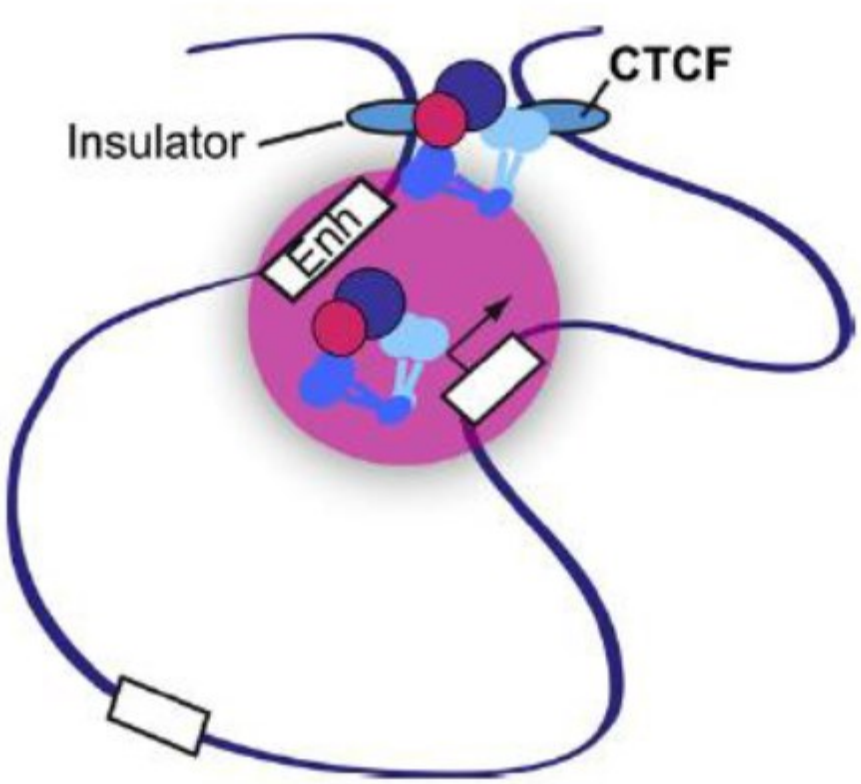


Fig. 1-3. CTCF-dependent and -independent chromatin loop formation at the β-globin locus.

Cohesin binds to and mediates the long-distance interactions of CTCF-bound insulator elements flanking the locus as well as between the distal enhancer (Enh) in the locus control region and the adult globin genes (white boxwith an arrow) [38]. The pink circle represents the presence of various transcription factors involved in globin gene expression, such as EKLF (Klf1), GATA-1, Fog-1, Ldb1, and NF-E2 [38,105,106,107,108]. A white box without an arrow represents the inactive gene, which is not interacting with the enhancer.

(The figure was published in [104].)

Chapter Two

NIPBL haplinsufficiency and cohesin- mediated gene regulation.

The majority of this chapter was included in the submitted paper entitled-

“The effect of “*Nipped-B-like*” haploinsufficiency on genome-wide cohesin binding and target gene expression: modeling Cornelia de Lange Syndrome”

Daniel A. Newkirk,* Yen-Yun Chen,* Richard Chien,* Weihua Zeng, Jacob Biesinger, Ebony Flowers, Aniello M. Infante, Shimako Kawauchi, Rosaysela Santos, Anne L. Calof, Arthur D. Lander, Xiaohui Xie,¶ and Kyoko Yokomori.¶

* These authors contributed equally to the work.

¶ Corresponding authors.

Abstract

Cornelia de Lange Syndrome (CdLS) is a developmental disorder frequently associated with heterozygous *NIPBL* mutations. NIPBL loads cohesin onto chromatin. Cohesin mediates sister chromatid cohesion and also acts as a regulator of gene expression. In CdLS patient cells and animal models, the presence of multiple gene expression changes with little or no cohesion defect suggests that disruption of gene regulation underlies this disease. However, the effect of *NIPBL* haploinsufficiency on cohesin binding, and how this relates to the clinical presentation of CdLS, has not been fully investigated. We examined genome-wide cohesin binding and its relationship to gene expression using mouse embryonic fibroblasts (MEFs) from *Nipbl* +/- mice that recapitulate the CdLS phenotype. A global decrease in cohesin binding was found, including unique sites and repeat regions. Cohesin-bound genes are mostly downregulated in *Nipbl* +/- MEFs with evidence for reduced promoter-enhancer interaction, suggesting that gene activation is the primary cohesin function sensitive to Nipbl reduction. Over 50% of genes affected in *Nipbl* +/- MEFs are cohesin- bound genes, including those involved in adipogenesis, indicating their direct contributions to the Nipbl haploinsufficiency-induced CdLS phenotype.

Introduction

CdLS (OMIM 122470, 300590, 610759) is a genetic disease estimated to occur in 1 in 10,000 individuals, characterized by facial abnormalites, hirsutism, upper upper limb defects, mental retardation, and growth abnormalities [63,64]. Mutations in the *NIPBL* gene are linked to more than >55% of CdLS cases [65,66]. NIPBL is a conserved protein that loads cohesin onto chromatin [109]. Cohesin is a multiprotein complex that functions in chromosome structural organization and gene regulation [79,104,110]. Mutations in the cohesin subunits SMC1A and hSMC3 were also found in a small subset of CdLS cases (~5% and <1%, respectively) [68,69]. More recently, mutation of HDAC8, which regulates cohesin dissociation from chromatin in mitosis, was found in a subset of CdLS patients (OMIM 300882) [26]. Mutations in the non-SMC cohesin component Rad21 have also been found in patients with a CdLS-like phenotype (OMIM 606462), with much milder cognitive impairment [70]. Thus, mutations of cohesin subunits and regulators of cohesin chromatin association cause related phenotypes, suggesting that impairment of the cohesin pathway is the cause of the disease.

Since *NIPBL* haploinsufficiency is the major cause of CdLS [64], *Nipbl* heterozygous mutant (*Nipbl* +/-) mice were created previously as a CdLS disease model [76]. The *Nipbl* +/- mice exhibited wide-ranging defects characteristic of the disease, including small size, craniofacial anomalies, microbrachycephaly, heart defects, hearing abnormalities, low body fat, and delayed bone maturation [76]. The mutant mice demonstrated only a 25-30% decrease in *Nipbl* transcripts [76]. A similar partial decrease of *NIPBL* was found in CdLS patients, in which only a 15% decrease was sufficient to cause mild CdLS [78]. These results indicate a high sensitivity of mammalian development to *Nipbl* dosage. The partial decrease of *Nipbl* expression in CdLS patients and *Nipbl* +/- mice was not sufficient to cause a significant sister

chromatid cohesion defect [76,83,85,86]. Instead, a distinctive profile of gene expression changes was observed, strongly suggesting that transcriptional dysregulation underlies the disease phenotype [76,78]. Gene expression changes in *Nipbl* +/- mutant mice are pervasive but minor, raising the possibility that small expression perturbations of multiple genes collectively contribute to the disease phenotype [76]. This hypothesis was further tested by combinatorial gene depletion in zebrafish, successfully recapitulating some aspects of the CdLS-like phenotype [87]. However, to what extent Nipbl and cohesin directly regulate affected genes was undetermined.

Using MEFs derived from *Nipbl* +/- mice, we analyzed the effect of *Nipbl* haploinsufficiency on cohesin-mediated gene regulation and identified cohesin target genes that are particularly sensitive to partial reduction of Nipbl. Our results indicate that Nipbl is required for cohesin binding genome-wide. Significant correlation was found between gene expression changes in *Nipbl* +/- MEFs and cohesin binding to the gene regions, suggesting that partial Nipbl reduction significantly affects expression of cohesin-bound genes. Target genes are enriched for developmental genes, including multiple genes that regulate adipogenesis, which is impaired in *Nipbl* +/- mice [76]. The results indicate that Nipbl regulates a significant number of genes through cohesin. Most of the Nipbl/cohesin target genes were downregulated in Nipbl +/- cells, indicating that Nipbl and cohesin are important for activation of these genes. The long-distance interaction of the cohesin-bound promoter and a putative enhancer region is decreased by Nipbl reduction, indicating that reduced cohesin binding by Nipbl haploinsufficiency affects chromatin interactions. Collectively, the results reveal that Nipbl haploinsufficiency globally reduces cohesin binding, and its major transcriptional consequence is downregulation of cohesin target genes.

Materials and Methods

Cells and antibodies

Mouse embryonic fibroblasts (MEFs) derived from E15.5 wild type and Nipbl mutant embryos were used as described previously [76]. In brief, mice heterozygous for Nipbl mutation were generated (*Nipbl* +/-) from gene-trap-inserted ES cells (RRS564 cell line). This mutation resulted in a net 30-50% decrease in Nipbl transcripts in the mice, along with many phenotypes characteristic of human CdLS patients [76]. Wild type and mutant MEF cell lines derived from the siblings were cultured at 37 °C and 5% CO_2 in DMEM (Gibco) supplemented with 10% fetal bovine serum and penicillin-streptomycin (50U/mL). E14TG2a.4 feeder- free mouse embryonic stem (mES) cells and RRS564, the *Nipbl* +/- mES cells which generated from a gene trap mutation in the E14TG2a.4, were cultured in Glasgow MEM (Sigma G5154) supplemented with 2mM L-Glutamine, 1mM sodium pyruvate, 1X non-essential amino acids, 10% fetal bovine serum (Hyclone SH30071.03), 50μM beta-mercaptoethanol, and 6 x 10^5 U LIF (Millipore ESG1107). Both E14TG2a.4 and RRS564 were cultured at 37 °C and 5% CO_2 with passage ratio around 1:4 – 1:6. 0.25% trypsin were used for detachment of cell during passaging, and all vessels and pipettes that come into contact with the mES cells were disposable plastic. Antibodies specific for SMC1 and Rad21 were previously described [111]. Rabbit polyclonal antibody against the NIPBL protein was raised against a bacterially-expressed recombinant polypeptide corresponding to the C-terminal fragment of NIPBL isoform A (NP_597677.2) (amino acids 2429–2804) and antigen affinity-purified. CTCF antibody was from Millipore (07-729) and histone H3 from Abcam (ab1791).

ChIP-sequencing (ChIP-seq) and ChIP-PCR

ChIP was carried out as described previously [38]. Approximately 50 µg DNA was used per IP. Cells were crosslinked 10 mins with 1% formaldehyde, lysed, and sonicated using the Bioruptor from Diagenode to obtain ~200bp fragments using a 30 sec on/off cycle for 1 hr. Samples were diluted and pre-cleared for 1 hr with BSA and Protein A sepharose beads. Pre-cleared extracts were incubated with Rad21, Nipbl, and preimmune antibodies overnight. IP was performed with Protein A beads with subsequent washes. DNA was eluted off beads, reversed crosslinked for 8 hrs, and purified with the Qiagen PCR Purification Kit. Samples were submitted to Ambry Genetics (Aliso Viejo, CA) for library preparation and sequencing using the Illumina protocol and the Illumina Genome Analyzer (GA) system. The total number of reads before alignment were: preimmune IgG, 7,428,656; Rad21 in control WT, 7,200,450; Rad21 in *Nipbl* +/-, 4,668,622; histone H3 in WT, 26,630,000; and histone H3 in Nipbl +/-, 24,952,439. Sequences were aligned to the mouse mm9 reference genome using Bowtie (with parameters–n2, -k20, —best, —strata, —chunkmbs 384) [112]. ChIP-seq data is being submitted to GEO. PCR primers used for manual ChIP confirmation are listed in Table 1. Primers corresponding to repeat sequences (major and minor satellite, rDNA, SINEB1 and B2 repeats) were previously described [113]. For manual ChIP-PCR analysis of selected genomic locations, ChIP signals were normalized with preimmune IgG and input DNA from each cell sample as previously described [38,101,114]. The experiments were repeated at least three times using MEF samples from different litters, which yielded consistent results. PCR reactions were done in duplicates or triplicates.

Peak Finding

Peaks were called using AREM (Aligning ChIP-seq Reads using Expectation Maximization) as previously described [115]. AREM incorporates sequences with one or many

mappings to call peaks as opposed to using only uniquely mapping reads, allowing one to call peaks normally missed due to repetitive sequence. Since many peaks for Rad21 can be found in repetitive sequence [115,116], we used a mixture model to describe the data, assuming K + 1 clusters of sequences (K peaks and background). Maximum likelihood is used to estimate the locations of enrichment, with the read alignment probabilities iteratively updated using EM. Final peaks are called for each window assuming a Poisson distribution, calculating a p-value for each sequence cluster. The false discovery rate for all peaks was determined relative to the pre-immune sample, with EM performed independently for the pre-immune sample as well. Full algorithm details are available, including a systematic comparison to other common peak callers such as SICER and MACS [115]. Overlap between peaks and genomic regions of interest were generated using Perl and Python scripts as well as pybedtools [117,118]. Figures were generated using the R statistical package [119]. Visualization of sequence pileup utilized the UCSC Genome Browser [120,121].

Expression data Analysis

Affymetrix MOE430A 2.0 array data for mouse embryonic fibroblasts (10 data sets for the wild type and nine for *Nipbl* +/- mutant MEFS) were previously published [76]. Expression data were filtered for probe sets with values below 300 and above 20,000, with the remainder used for downstream analysis. Differential expression and associated p-values were determined using Cyber-t, which uses a modified t-test statistic [122]. Probe sets were collapsed into genes by taking the median value across all probe sets representing a particular gene. Raw expression values for each gene are represented as a z-score, which denotes the number of standard deviations that value is away from the mean value across all genes. Gene ontology analysis was performed using PANTHER with a cutoff of $p < 0.05$ [123,124].

Affymetrix GeneChip Mouse Gene 1.0 ST arrays were used for mES gene expression analyses. 3 replicates (3 separate dishes of cells) were obtained for array analyses. Expression data from untreated E14TG2a.4 are compared with RRS564; and data from E14TG2a.4 treated with control siRNA were compared with data from either Rad21 or Nipbl siRNA treated cells. Differential expression and associated p-values were determined using Cyber-t as described above. Probe sets with values below 200 were filtered out, and P value<0.05 and fold change >1.2 (up-regulated) or <-1.2 (down-regulated) were considered as "differentially expressed" genes. DAVID was used for ontology analyses [125,126]. Expression heatmaps were generated using GenePattern [127].

KS test

Genes were sorted by their fold-change and any adjacent ChIP binding sites were identified. We performed a Kolmogorov-Smirnov (KS) test comparing the expression-sorted ChIP binding presence vs. a uniform distribution of binding sites, similar to Gene Set Enrichment Analysis [128]. If ChIP binding significantly correlates with the gene expression fold-change, the KS statistic, d, will also have significant, non-zero magnitude. To better visualize the KS test, we plotted the difference between the presence of cohesin binding at (expression-sorted) genes in Fig. 2-3C. The x axis of this figure is the (fold-change-based) gene rank, and the y axis is the KS statistic d, which behaves like a running enrichment score and is higher (lower) when binding sites co-occur more (less) often than expected if there were no correlation between ChIP binding and expression fold-change. The KS test uses only the d with the highest magnitude, which is indicated in the plots by a vertical red line. To better visualize ChIP binding presence, we further plot an x-mirrored density of peak presence at the top of each plot; the gray "beanplot" [129] at the top of the plots are larger when many of the genes have adjacent ChIP binding sites.

siRNA depletion

Wild type MEFs were transfected using HiPerFect (Qiagen) following the manufacturer's protocol with 10mM siRNA. A mixture of 30µl HiPerFect, 3µl of 20µM siRNA, and 150µl DMEM was incubated for 10 mins and added to 2 x 10^6 cells in 4 ml DMEM. After 6 hrs, 4 ml fresh DMEM with 10% FBS was added. Transfection was repeated the next day. Cells were harvested 48 hrs after the first transfection. SiRNAs against Nipbl (Nipbl-1: 5'-GTGGTCGTTACCGAAACCGAA-3'; Nipbl-2: 5'-AAGGCAGTACTTAGACTTTAA-3') and Rad21 (5'-CTCGAGAATGGTAATTGTATA-3') were made by Qiagen. AllStars Negative Control siRNA was obtained from Qiagen.

RT-q-PCR

Total RNA was extracted using the Qiagen RNeasy Plus kit. First-strand cDNA synthesis was performed with SuperScript II (Invitrogen). Q-PCR was performed using the iCycler iQ Real-time PCR detection system (Bio-Rad) with iQ SYBR Green Supermix (Bio-Rad). Values were generated based on Ct and normalized to *Rnh1* expression. PCR primers specific for major satellite, minor satellite, rDNA, SINE B1 and SINE B2 were previously described [113]. Other unique primers are listed in Table 1. The RT-qPCR analyses of the wild type and mutant cells were done with two biological replicates with consistent results. The gene expression changes after siRNA treatment were evaluated with two to three biological replicates with similar results.

3C analysis

The chromosome conformation capture (3C) protocol was performed as described [38]. Approximately 1 x 10^7 cells were crosslinked with 1% formaldehyde at 37°C for 10 mins. Crosslinking was stopped by adding glycine to a final concentration of 0.125M. Cells were

centrifuged and lysed on ice for 10 minutes. Nuclei were washed with 500µl of 1.2x restriction enzyme buffer and resuspended with another 500µl of 1.2x restriction enzyme buffer with 0.3% SDS and incubated at 37°C for 1 hr. Triton X-100 was added to 2% and incubated for another 1 hr. 800 U of restriction enzyme (HindIII New England Biolabs) was added and incubated overnight at 37°C. The digestion was heat-inactivated the next day with 1.6% SDS at 65°C for 25 minutes. The digested nuclei were added into a 7ml 1x ligation buffer with 1% Triton X-100, followed by 1 hour incubation at 37°C. T4 DNA ligase (2000 U) (New England Biolabs) was added and incubated for 4 hrs at 16°C followed by 30 minutes at room temperature. Proteinase K (300µg) was added and the sample was reverse-crosslinked at 65°C overnight. Qiagen Gel Purification Kits were used to purify DNA. Approximately 250ng of template was used for each PCR reaction. PCR products were run on 2% agarose gels with SYBR safe DNA stain (Invitrogen), visualized on a Fujifilm LAS-4000 imaging system and quantified using Multigauge (Fujifilm).

To calculate interaction frequencies, 3C products were normalized to the constitutive interaction at the excision repair cross-complementing rodent repair deficiency, complementation group 3 (ercc3) locus [130,131], which is unaffected in mutant MEFs. A control template was made to control for primer efficiencies locus-wide as described [132]. PCR fragments spanning the restriction sites examined were gel purified and equimolar amounts were mixed (roughly 15µg total) and digested with 600 U restriction enzyme overnight and subsequently ligated at a high DNA concentration (>300ng/µl). The template was purified with the Qiagen PCR Purification Kit and mixed with an equal amount of digested and ligated genomic DNA. 250ng of the resulting control template was used for each PCR for normalization against PCR primer efficiencies. Two biological replicates with three technical replicates each

were analyzed for both wild type and mutant cells and for control and Nipbl siRNA-treated cells, which yielded consistent results.

Results

Nipbl haploinsufficiency leads to a global reduction of cohesin binding

In order to investigate how *Nipbl* haploinsufficiency leads to CdLS, cohesin binding was examined genome-wide by ChIP-seq analyses using antibody specific for the cohesin subunit Rad21, in wild type and *Nipbl* +/- mutant MEFs derived from E15.5 embryos [76] (Fig. 2-1A). MEFs derived from five wild type and five mutant pups from two litters were combined to obtain sufficient chromatin samples for ChIP-seq analysis. Nipbl +/- mutant MEFs express approximately 30-40% less Nipbl compared to wild type MEFs [76] (Table 2-2). MEFs from this embryonic stage were chosen in order to match with a previous expression microarray study, because they are relatively free of secondary effects caused by Nipbl mutation-induced developmental abnormalities compared to other embryonic tissues [76]. Consistent with this, there is no noticeable difference in growth rate and cell morphology between normal and mutant MEFs [76]. This particular anti-Rad21 antibody was used previously for ChIP analysis and was shown to identify holo-cohesin complex binding sites [38,101,133,134].

Cohesin binding sites were identified using AREM [115], with a significance cut-off based on a p-value less than 1x10-4, resulting in a FDR below 3.0% (Fig. 2-1A). Approximately 35% fewer cohesin binding sites were found in *Nipbl* +/- mutant MEFs compared to the wild type MEFs (Fig. 2-1A). This is not due to variability in sample preparation since no significant difference in the histone H3 ChIP-seq was observed between the wild type and mutant cell samples (R-value=0.96) (Fig. 2-1D). The above results might suggest that a significant number of binding sites are unique to the wild type cells (Fig. 2-1A). When we compared the raw number of reads located within wild type peaks and the corresponding regions in mutant MEFs,

however, we noted a reduced, rather than a complete absence of, cohesin binding in mutant cells (Fig. 2-1B). Those regions in mutant cells corresponding to the “WT only” regions consistently contain one to three tags in a given window, which are below the peak cut-off. However, the signals are significant compared to the negative control of preimmune IgG (Fig. 2-1B). Furthermore, even for those sites that are apparently common between the control and mutant MEFs, the binding signals appear to be weaker in mutant cells (Fig. 2-1B). To validate this observation, we segmented the genome into non-overlapping 100bp bins, and plotted a histogram of the log ratios of read counts between the wild type and mutant samples in each bin, with read counts normalized using reads per kb per million total reads (RPKM) [135]. The plot indicates that the read counts for the mutant bins are generally less than those for the wild type bins, even for the binding sites common to both wild type and mutant cells (Fig. 2-1C). Decreased cohesin binding was further confirmed by manual ChIP-qPCR analysis of individual cohesin binding sites using at least three independent control and mutant MEF samples supporting the reproducibility of the results (see below, Figure 2-2A). Decreased cohesin binding was also observed at additional specific genomic regions in *Nipbl* +/- MEFs [136]. Manual H3 ChIP and CTCF ChIP both showed no difference between wildtype and *Nipbl* +/- cells, again demonstrates that the chromatin preps were of equal quality (Fig. 2-2C, 2-2D). Taken together, the results indicate that cohesin binding is generally decreased at its binding sites found in wild type MEFs, rather than re-distributed, in mutant MEFs.

Repeat sequences are often excluded from ChIP-seq analysis. However, cohesin binding is found at various repeat sequences, including pericentromeric and subtelomeric heterochromatin, and ribosomal DNA regions in the context of heterochromatin in mammalian cells [101,137]. We specifically tested the effect of Nipbl reduction on cohesin binding to repeat sequences by manual ChIP-q-PCR (Fig. 2-2B). Both Nipbl mutation (Fig. 2-2B, top) and Nipbl

depletion using two different siRNAs (Fig. 2-2B, bottom) resulted in decreased cohesin binding at the repeat regions, indicating that Nipbl is also important for cohesin binding to repeat sequences. In contrast, there were no significant differences in the histone H3 ChIP signals between these repeat regions in wild type and mutant MEFs (Fig. 3C). Taken together, the results indicate that Nipbl functions in cohesin loading at both unique sites and repeat regions, confirming the genome-wide decrease of cohesin binding caused by *Nipbl* haploinsufficiency.

Cohesin-bound genes are sensitive to *Nipbl* haploinsufficiency

The distribution of cohesin binding sites in the genomes of both wild type and mutant MEFs were examined. There is a significant enrichment of cohesin binding in promoter regions, and to a lesser extent in the 3' downstream regions, relative to the random genomic distribution generated by sampling from pre-immune ChIP-seq reads. There is no significant redistribution or genomic region-biased loss of cohesin binding sites in Nipbl mutant cells (Fig. 2-3A).

Based on the significant enrichment of cohesin binding in the promoter regions, we next examined the correlation between cohesin binding to the gene regions and the change of gene expression in mutant MEFs using a KS test. This is a nonparametric test for comparing peak binding sites with gene expression changes in the mutant MEFs (Fig. 2-3C). Genes that displayed the greatest expression change in mutant MEFs compared to the wild type MEFs showed a strong correlation with cohesin binding to the gene region, indicating that direct binding to the target genes is the major mechanism by which cohesin mediates gene regulation in a Nipbl dosage-sensitive fashion (Fig. 2-3C, left). Random sampling of a comparable number of simulated peaks in the gene regions yielded no correlation (Fig. 2-3C, right). Interestingly, cohesin binding to the gene region correlates better with decreased gene expression than increased expression in mutant cells, indicating that gene activation, rather than repression, is the major mode of cohesin function at the gene regions (Fig. 2-3C, middle).

Identification of cohesin target genes sensitive to *Nipbl* haploinsufficiency

The results above indicate that cohesin-bound genes sensitive to a partial loss of Nipbl can be considered to be Nipbl/cohesin target genes. Among 218 genes that changed expression significantly in mutant cells compared to the wild type (>1.2-fold change, p-value < 0.05), we found that more than half (115 genes) were bound by cohesin, and thus can be considered Nipbl/cohesin target genes (Table 2-3). This is a conservative estimate of the number of direct target genes since cohesin binding sites beyond the upstream and downstream cut-offs (2.5 kb) were not considered for the analysis. Consistent with the KS test analysis (Fig. 2-3C), ~74% of these cohesin target genes were downregulated in mutant cells, indicating that the positive effect of cohesin on gene expression is particularly sensitive to partial reduction of Nipbl (Table 2-3).

Many of these Nipbl/cohesin-target genes contain cohesin binding sites in more than one region (promoter, gene body and/or downstream), suggesting their collaborative effects (Fig. 2-4B). In particular, the promoter binding of cohesin is often accompanied by its binding to the gene body. However, binding pattern analysis revealed no significant correlation between a particular pattern and/or number of cohesin binding sites and gene activation or repression (Fig. 2-4B). Rad21 ChIP-seq signal intensity profiles of several cohesin target genes (as defined above) reveal decreased cohesin binding in mutant cells at the binding sites originally observed in the wild type cells, supporting the notion that gene expression changes are the direct consequence of the reduced cohesin binding (Fig. 2-2A; Fig. 2-4C, top). There are other genes, however, that did not change expression significantly in mutant MEFs, but nevertheless also have reduced cohesin peaks nearby (Fig. 2-4C, bottom), suggesting that cohesin binding is not the sole determinant of the gene's expression status and that its effect is context-dependent.

Gene ontology analysis revealed that the target genes bound by cohesin at the promoter regions and affected by Nipbl deficiency are most significantly enriched for those involved in development (Table 2-4). The results suggest a direct link between diminished Nipbl/cohesin and the dysregulation of developmental genes, which contributes to the CdLS phenotype.

Nipbl- and cohesin-mediated activation of adipogenesis genes

One of the reported phenotypes of *Nipbl* +/- mice is their substantial reduction of body fat that mirrors what is observed in CdLS patients [76,138]. It was found that *Nipbl* +/- MEFs exhibit dysregulated expression of several genes involved in adipocyte differentiation, and reduced spontaneous adipocyte differentiation in vitro [76,138]. We therefore examined the effect of Nipbl haploinsufficiency on these adipogenesis genes in detail. We found that many of them are bound by cohesin, in some cases at multiple sites, suggesting that cohesin plays a direct role in activation of these genes (Fig. 2-5). Although *Il6* and *Cebpδ* and *Lpar1* were originally not included in the 115 genes due to p-values being higher than 0.05 in the microarray analysis (0.112 and 0.117, respectively), significant expression changes were observed in mutant MEFs compared to the wild type MEFs by manual RT-qPCR. *TNFα* and *PPARγ*, also involved in adipogenesis, do not change their expression in mutant MEFs (p-values 0.622 and 0.759, repspectively) [76]. Importantly, a decrease of gene expression was not only observed in *Nipbl* +/- mutant cells, but also by siRNA depletion of Nipbl, confirming that the effect is specifically caused by Nipbl reduction (Fig. 2-5A). Furthermore, depletion of cohesin itself decreased their expression even more significantly than Nipbl depletion (Fig. 2-5A). The results suggest that multiple genes involved in the adipogenesis pathway are direct cohesin targets that are sensitive to Nipbl haploinsufficiency.

Reduced cohesin binding due to Nipbl reduction leads to a loss of long-distance chromatin interaction

The above results revealed the critical association of cohesin binding to the promoter region and expression of target genes. How does cohesin bound to the promoter affect gene expression? We recently showed that cohesin-mediated long-distance chromatin interaction between distal enhancer and promoter regions was reduced at the β-globin locus, resulting in reduced gene expression, in *Nipbl* mutant mice [38]. Thus, we tested the potential involvement of cohesin binding to the *Cebpβ* gene, one of the target adipogenesis genes described above, in such long-distance chromatin interaction(s) and whether it is affected by Nipbl reduction using chromosome conformation capture (3C) analysis (Fig. 2-6). We tested several flanking sites that are positive for cohesin and RNA polymerase II (pol II) binding as well as H3K4me1 and H3K4me3, the hallmarks for enhancers [139,140,141] (Fig. 2-6A). We observed that the Cebpβ promoter interacts with one such region (Fig. 2-6A and B, the site "c"). Although the site "c" is associated with only a weak Rad21 ChIP-seq signal, SMC1 and SMC3 ChIP-seq signals were found at the same region [142], confirming that this is an authentic cohesin binding site (Fig. 2-6A). The results indicate a selectivity of chromatin interactions among neighboring cohesin binding sites, revealing that not all proximal cohesin binding sites interact with each other. Importantly, this interaction is indeed reduced in both *Nipbl* mutant and Nipbl siRNA-treated MEFs (Fig. 2-6B). The 3C signals at the *Cebpβ* locus were normalized to the constant interaction observed at the Ercc3 locus [130,131], which was not affected by Nipbl reduction. The results indicate that the decrease of long-distance chromatin interaction involving the promoters and distant DNA elements is one of the direct consequences of reduced cohesin binding, which may be one mechanism of gene expression alteration by *Nipbl* haploinsufficiency.

Long- and short- term Nipbl reduction affects gene expression differently in embryonic stem cells

The *Nipbl* +/- mice were generated from RRS564, a gene-trapped mouse embryonic stem (mES) cell line in which the *Nipbl* gene on one allele was inactivated. We examined the expression pattern of the *Nipbl* +/- mES cells comparing with the parental line, E14TG2a.4. Nipbl siRNA depletion and cohesin (Rad21) siRNA depletion were also performed and compared with control siRNA treatment in E14TG2a.4. The *Nipbl* +/- mES cells showed ~30% decrease in *Nipbl* transcript level while the decrease of *Nipbl* by siRNA was ~50% (data not shown). Only 197 genes were found to be differentially expressed in the Nipbl +/- mutant, which is significantly less compared to the transient Nipbl depletion (Fig. 2-7A). Furthermore, there is a minimum overlap of mysregulated genes between the *Nipb* +/-cells and transient Nipbl depleted cells (Fig. 2-7A, 2-7C). These results suggest that long- and short-term Nipbl depletion may affect mES cells differently. Alternatively, the different degrees of Nipbl depletion may affect gene expression differently as observed in Drosophila [143]. Distinct effects of *Nipbl* mutation and transient Nipbl depletion may also be explained by the possible off-target effect of Nipbl siRNA. However, my comparison of gene expression changes between Rad21 and Nipbl siRNA-treated cells revealed significant overlap between the two (Fig. 2-7B, 2-7C), strongly arguing against this concern. Rad21 or Nipbl depletion down-regulate many pluripotency genes. Similar results had been obtained by another group which depleted cohesin in a different mES cell line [39]. However, *Nipbl* +/- mutant samples do not show any significant down-regulation in those genes (Fig. 2-7D). This observation demonstrates that cells can adapt to long term Nipbl reduction.

The top gene categories from ontology analyses are similar for both Rad21 depletion and Nipbl depletion. Up-regulated genes are enriched with cytoskeletal binding proteins and down-

regulated genes are enriched with genes whose functions involve transcription, genes that code for mitochondrial proteins, and tRNA metabolism (Table.2-5). Clusters of developmental-related genes were found in the down-regulated genes in Rad21 and Nipbl depleted samples, but with lower rankings. Clustering for genes that are up-regulated in the *Nipbl* +/- mutant didn't show significant clusters in any categories; however, down-regulated genes show enrichment in developmental categories such as pattern specification processes and axon guidance (Table. 2-6). 30 genes were found, which correspond to 40% of all down-regulated genes in the *Nipbl* +/- mutant, are directly involved in embryonic development (Table. 2-7). The range covers muscle, cranial facial, heart, and nervous system development, which are all known processes defective in CdLS patients. Downregulation of many developmental regulators under long term depletion but not the transient depletion suggest that the cells with long term depletion may model more closely to the actual disease.

Discussion

In this study, we used MEFs derived from *Nipbl* heterozygous mutant mice to analyze the effect of *Nipbl* haploinsufficiency (the primary cause of CdLS) on cohesin binding and its relationship to gene expression. We found a genome-wide decrease in cohesin binding at unique sites and repeat regions, indicating the high sensitivity of cohesin binding to even a partial reduction of the Nipbl protein. Importantly, the expression of genes bound by cohesin, particularly at the promoter regions, is preferentially altered in response to Nipbl reduction. While some genes are activated, the majority of cohesin-bound genes are repressed by decreased cohesin binding, indicating the positive role of cohesin in this context. Our results indicate that more than 50% of genes whose expression is altered significantly in *Nipbl* haploinsufficient cells are cohesin target genes directly influenced by decreased cohesin binding at the individual gene regions. One consequence of reduced cohesin binding at the promoter region is a decrease of a specific long-distance chromatin interaction, raising the possibility that cohesin-dependent higher-order chromatin organization in the nucleus may be globally altered in CdLS patient cells.

How does Nipbl haploinsufficiency affect cohesin target gene expression?

One mechanism of cohesin action in gene regulation is to mediate chromatin loop formation [38,39]. Increased Nipbl and cohesin binding correlates with the induction of the enhancer-promoter interaction and robust gene activation at the β-globin locus [38]. Depletion of cohesin resulted in decreased enhancer-promoter interactions and downregulation of globin genes [38]. Similarly, *Nipbl* haploinsufficiency results in less cohesin binding and decreased promoter-enhancer interactions and β-globin gene expression [38]. In the current study, we also found that the cohesin-bound promoter of one of the target genes, *Cebpβ*, is involved in a long-

distance chromatin interaction with a putative enhancer, which is decreased in *Nipbl* mutant cells, consistent with the decreased gene expression. Thus, *Nipbl* haploinsufficiency affects cohesin target gene expression by decreasing cohesin-mediated chromatin interactions.

It should be noted, however, that not all genes we examined showed significant long-distance chromatin interactions involving cohesin-bound promoters. While this may be because we did not test the correct enhancer regions, it also suggests that cohesin may promote gene activation by a mechanism(s) other than by mediating long-distance promoter interaction. One possibility is gene looping. In *S. cerevisiae*, the promoter and terminator regions of genes interact with each other, which was thought to facilitate transcription re-initiation [144]. Although cohesin is often found at the promoter and terminator regions of genes in MEFs we failed to obtain any evidence for the involvement of these sites in gene looping with our limited analysis. Similarly, we observed that the *Cebpβ* gene promoter interacts with only one out of three neighboring cohesin sites. Taken together, it appears that not all nearby cohesin binding sites interact with each other. What dictates the selective chromatin interactions of specific cohesin binding sites is currently unclear.

Cohesin binding to the gene body regions is found at many of the cohesin target genes. This may represent the cohesin binding at intragenic enhancer elements or may be related to Pol II pausing [104]. While cohesin was shown to facilitate Pol II elongation in *Drosophila* [145,146], cohesin together with CTCF in the intragenic region was found to cause Pol II pausing at the *PUMA* gene in human cells [147], suggesting that cohesin can have both positive and negative effects on transcriptional elongation in a context-dependent manner. Furthermore, not all the cohesin-bound genes changed expression in *Nipbl* +/- MEFs, echoing this notion that the effect of cohesin binding on gene expression is context-dependent. What determines the

effects of cohesin binding at individual binding sites on gene expression requires further investigation.

Nipbl haploinsufficiency vs. cohesin mutation

There are two different cohesin complexes in mammalian somatic cells that differ by one non-SMC subunit (i.e., SA1 (STAG1) or SA2 (STAG2)) [13,14]. A recent report on *SA1* knockout mice revealed some phenotypic similarity to what is seen in mice with *Nipbl* haploinsufficiency [142]. Interestingly, the *SA1* gene is one of the cohesin target genes that is slightly upregulated in *Nipbl* mutant cells [76]. Thus, together with the compensatory increase of *Nipbl* expression from the intact allele, there appears to be a feedback mechanism that attempts to balance the expression of Nipbl and cohesin in response to *Nipbl* mutation. The fact that upregulation was observed with the *SA1*, but not *SA2*, gene may reflect the unique transcriptional role of SA1 [142]. Interestingly, however, only 10% of 215 genes altered in *Nipbl* mutant MEFs are changed significantly in *SA1* KO MEFs [142]. This discrepancy may, as observed in *Drosophila* [82], reflect the different effects of decreased binding versus complete knockout of a cohesin subunit on target gene expression. It could also be a result of the decreased binding of the second cohesin complex, cohesin-SA2.

Cohesin binding was relatively uniformly decreased genome-wide in *Nipbl* haploinsufficient cells with no significant redistribution of cohesin binding sites. Point mutations of different subunits of cohesin cause CdLS and CdLS-like disorders with both overlapping and distinct phenotypes compared to CdLS cases caused by *NIPBL* mutations [68,69,70]. Non-overlapping effects of downregulation of different cohesin subunits have been reported in zebrafish [87,148]. This may reflect an unequal role of each cohesin subunit in gene regulation and it is possible that some of the cohesin target genes may be particularly sensitive to a specific cohesin subunit mutation. For example, similar to the TBP-associating factors (TAFs) in

TFIID [149], cohesin subunits may provide different interaction surfaces for distinct transcription factors, which would dictate their differential recruitment and/or transcriptional activities. Furthermore, recent studies provide evidence for cohesin-independent roles of NIPBL in chromatin compaction and gene regulation [150,151]. Thus, disturbance of cohesin functions as well as impairment of cohesin-independent roles of NIPBL may collectively contribute to CdLS caused by *NIPBL* mutations.

Difference between *Nipbl* mutant and transient Nipbl reduction

The different expression patterns observed in the *Nipbl* mutant and Nipbl transient depleted mES cells may be the results of two factors: the amount of Nipbl reduction and the length of time that Nipbl is reduced. The transient depletion resulted in a 50% reduction of *Nipbl* transcript while in the RRS564 mutant, the reduction was about 30%. It is known that different degree of *Nipbl* reduction can result in different effect on gene expression in *drosophila* [143]. It was shown that CdLS patients with *NIPBL* gene truncation or nonsense mutations that lead to early termination generally have more severe phenotypes than those individuals with missense mutations [67], demonstrating that different doses of Nipbl may generate different effects on gene regulation. Temporally, the cell cultures with Nipbl reduction can undergo adaptation or selection, therefore resulted in a population of cells that are favored for survival under Nipbl reduction. Take the pluripotency genes, for example, Nipbl reduction initially downregulates the pluripotency factors. However, cells with long- term Nipbl reduction seems to have compensatory mechanisms to upregulate pluripotency genes back to normal level, thus the cells with Nipbl reduction are still true embryonic stem cells that are capable of developing into a complete organism. It would be insightful to explore the compensatory mechanisms of the pluripotency genes. It would be useful to create an inducible system for Nipbl depletion in mES cells, which allows us to control the degree of Nipbl reduction and the length of time of Nipbl

reduction, to investigate this matter. We should be mindful that CdLS is also the result of long-term Nipbl reduction, thus investigating how cells adapt to long-term Nipbl reduction would provide us with more understanding of CdLS disease progression.

Conclusion

Our results demonstrate that cohesin binding to chromatin is highly sensitive genome-wide (both at unique and repeat regions) to partial *Nipbl* reduction, resulting in a general decrease in cohesin binding genome-wide. Many genes whose expression is changed by *Nipbl* reduction are actual cohesin target genes. Our results suggest that decreased cohesin binding due to partial reduction of NIPBL at the gene regions directly contributes to disorder-specific gene expression changes and the CdLS phenotype. This work provides important insight into the function of cohesin in gene regulation with direct implications for the mechanism underlying *NIPBL* haploinsufficiency-induced CdLS pathogenesis.

A

Rad21 ChIP-seq	# Tags	% aligned	# Peaks	p-value	FDR
Control MEF	5,501,724	64%	25,407	< 1X10^{-4}	< 3%
Nipbl (+/-) MEF	4,740,463	75%	16,528	< 1X10^{-4}	< 3%

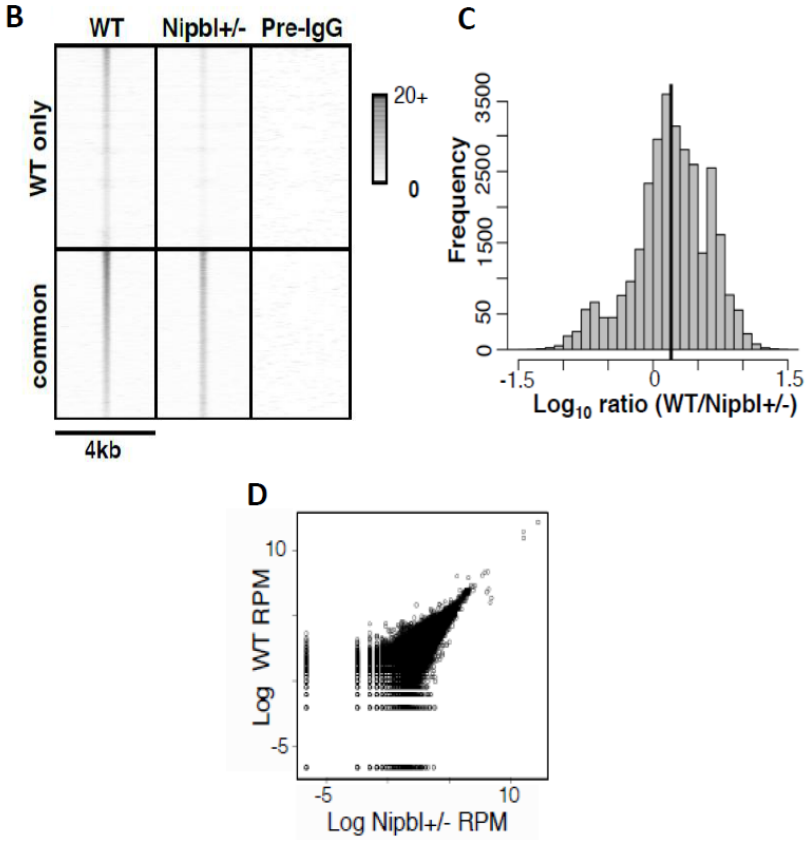


Figure 2-1. Global decrease of cohesin binding to chromatin in *Nipbl* heterozygous mutant MEFs.

(A) Cohesin binding sites identified by ChIP-sequencing using antibody specific for Rad21 in control wild type and *Nipbl* +/- MEFs. The p-value and FDR are shown. **(B)** Heatmap analysis of cohesin binding in wild type (WT) MEFs and corresponding peak signals in *Nipbl* +/- MEFs. The normalized (reads per million) tag densities in a 4 kb window around each peak are plotted, with peaks sorted from the highest number of tags in the wild type to the lowest. Peaks are separated into two categories, those that are found only in wild type ("WT only") and those that overlap between wild type and *Nipbl* +/- ("common"). Preimmune IgG ChIP-seq signals in the corresponding regions are also shown as a control. The color scale indicates the number of tags in a given region. **(C)** Histogram of the ratio between normalized (reads per million total reads) wild type and mutant reads in peaks common to both. Positive values indicate more wild type tags. The black line indicates the mean ratio between wild type and mutant tag counts. **(D)** Scatter plot of histone H3 ChIP-seq tag counts in wild type and mutant MEFs in 500 bp bins across the mouse genome. The values are plotted in log reads per million (RPM).

(Rad21 ChIPseq was performed by Richard Chien; Histone H3 ChIPseq performed by Yen-Yun Chen, and data analyses were done by Daniel Newkirk.)

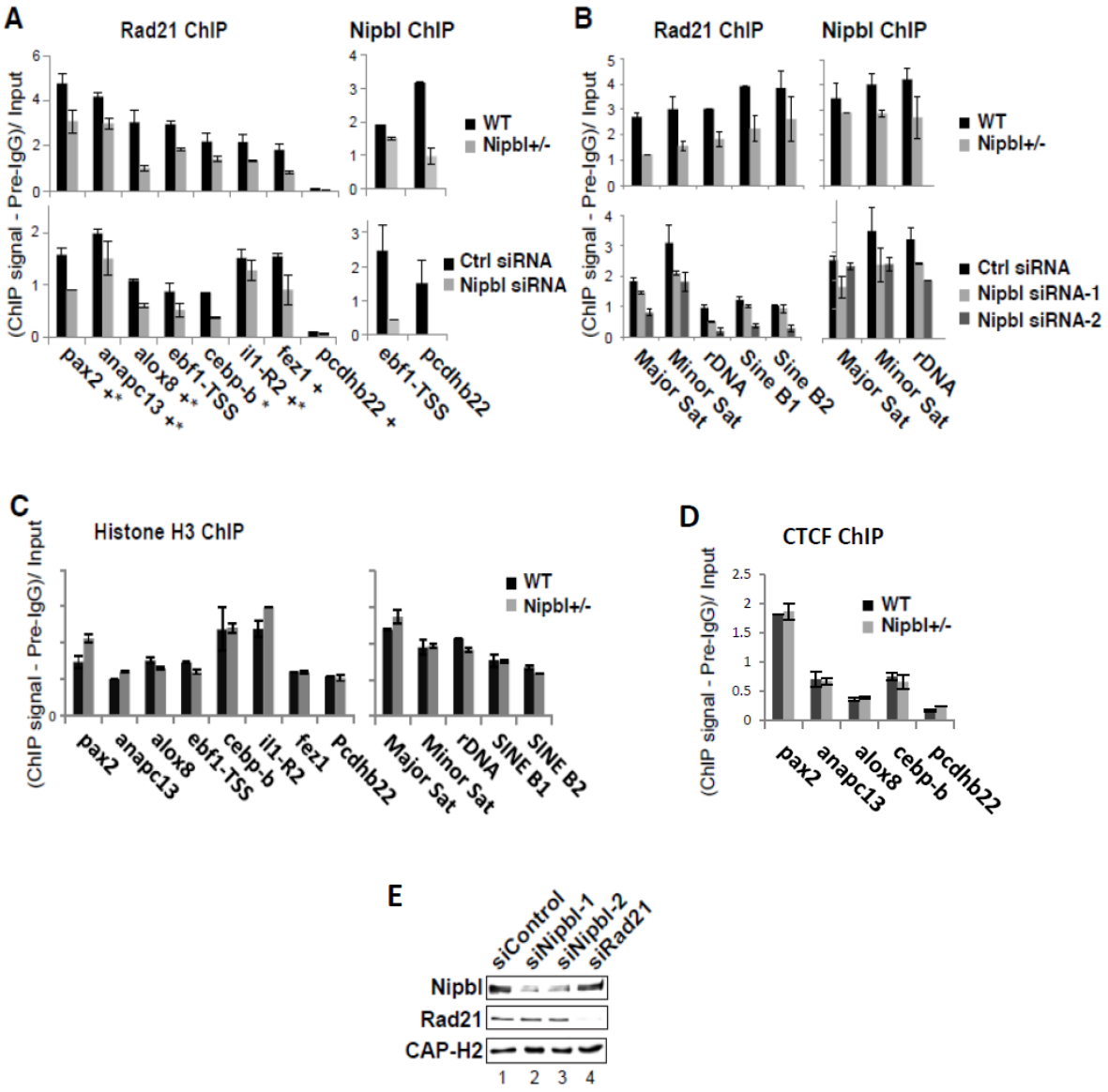


Figure 2-2. Nipbl reduction decreases cohesin binding to chromatin.

(A) Manual ChIP-q-PCR of cohesin binding sites using anti-Rad21 antibody in mutant and wild type MEFs (top panel) and Nipbl or control siRNA-treated MEFs (bottom panel) as indicated. Representative examples of Nipbl ChIP are also shown. "+" indicates CTCF binding and "*" indicates the presence of motif. PCR signals were normalized with preimmune IgG (pre-IgG) and input. **(B)** Similar manual ChIP-q-PCR analysis as in (A) of repeat regions in wild type and *Nipbl* mutant MEFs (top) and control and Nipbl siRNA-treated MEFs (bottom). **(C)** Histone H3 ChIP-q-PCR of both unique gene regions and repeat regions in WT and *Nipbl* +/- MEFs. **(D)** CTCF ChIP-qPCR of some unique gene regions in WT and *Nipbl*+/- MEFs. **(E)** Western blot analysis of siRNA-treated cells is shown using antibodies indicated. Depletion efficiency and specificity of Nipbl siRNA were also examined by RT-q-PCR (Table 2).

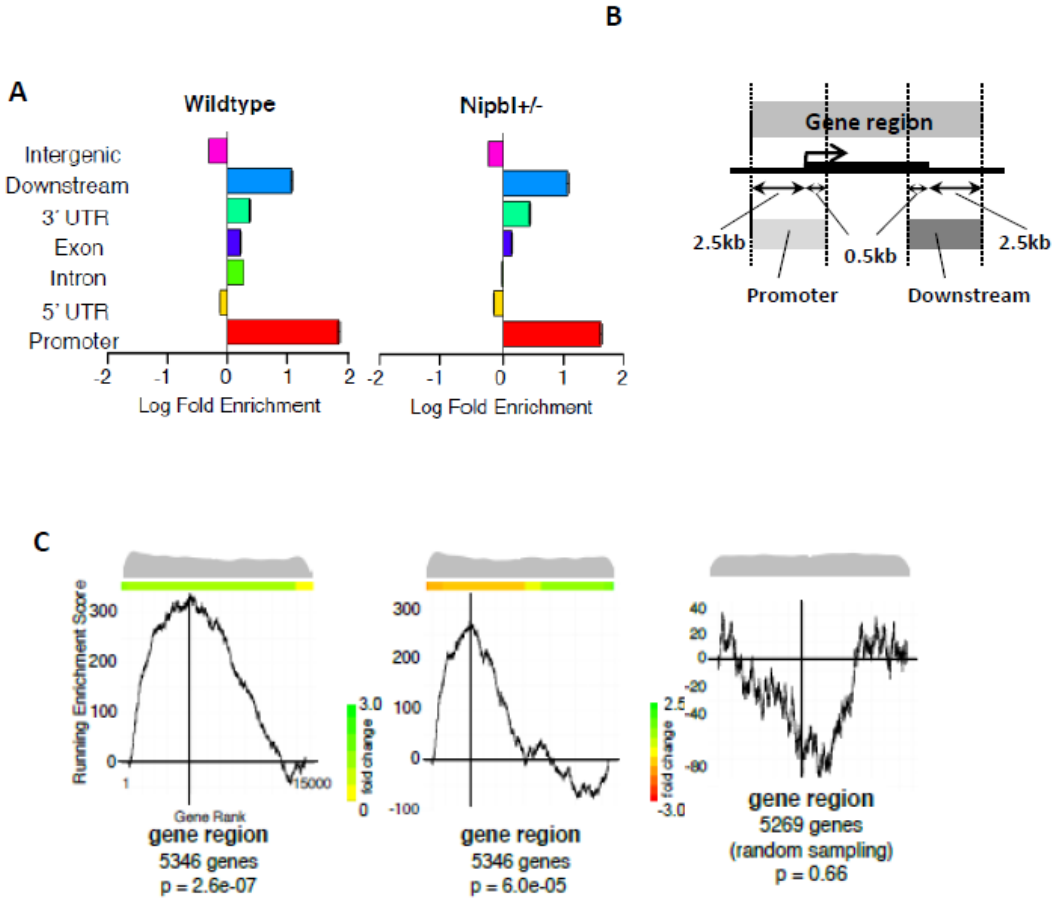


Figure 2-3. Correlation of cohesin binding and gene expression changes in mutant MEFs.

(A) Enrichment of cohesin peaks across genomic regions as compared to randomly sampled genomic sequence. A comparable number of peaks (25,407 and 16,528 peaks in wild type and mutant MEFs, respectively), with the same length as the input set, were randomly chosen 1000 times and the average used as a baseline to determine enrichment in each genomic region category. **(B)** The schematic diagram showing the definition of the gene regions, promoter (2.5kb upstream and 0.5kb downstream of TSS), and downstream (2.5kb downstream and 0.5kb upstream of TTS) regions is shown on the right. **(C)** KS test indicating the degree of cohesin binding to genes changing expression in *Nipbl* +/- MEFs. X-axis represents all 13,587 genes from the microarray data [76] ranked by absolute fold expression changes from biggest on the left to the smallest on the right in the left panel. Fold changes are shown in different colors as indicated on the side. In the middle panel, gene expression changes were ranked from negative to positive with the color scale shown on the side. Both color scales apply to the rest of the Figure. The Y-axis is the running enrichment score for cohesin binding (see METHODS for details). Distribution of cohesin-bound genes among 13,587 genes examined is shown as a beanplot [129] at the top, and the number of cohesin-bound genes and p-values are shown underneath. In the right panel, Lack of correlation between the mutant expression changes and randomly chosen genes are shown as a negative control.

(This part was done by Daniel Newkirk).

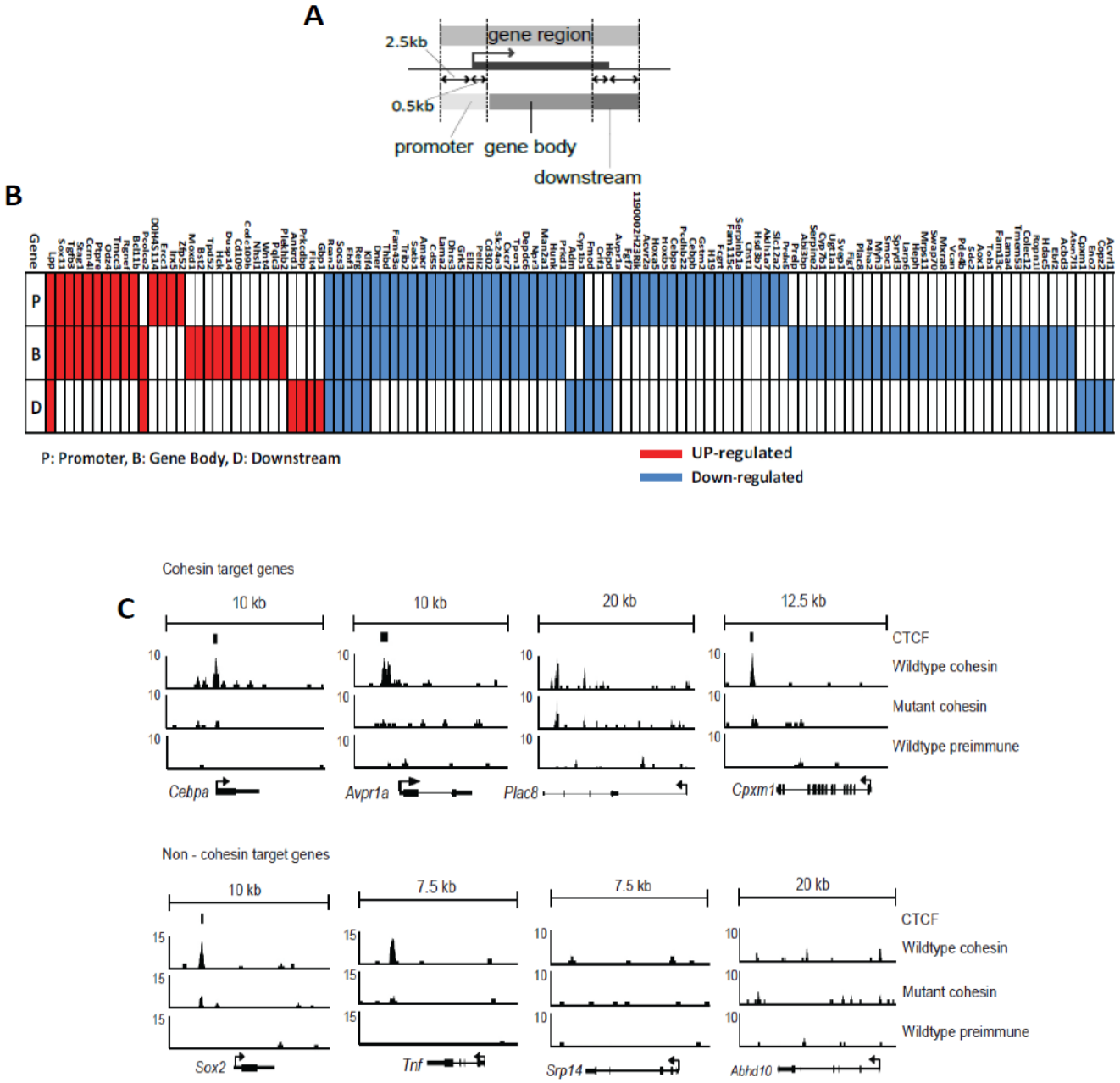


Figure 2-4. Cohesin binding signals at specific gene regions.

(A) Diagram showing the definition of gene regions, promoter, gene body, and downstream. **(B)** Cohesin binding site distribution in cohesin target genes as defined in Table 1. Cohesin binding to the promoter ("P"), gene body ("B"), and downstream region ("D") are indicated for each cohesin target gene in red (upregulated) and blue (downregulated) boxes. **(C)** Signal intensity profiles of Rad21 ChIP-seq at specific gene regions in wild type and Nipbl mutant MEFs. Preimmune IgG ChIP-seq signals are shown as a negative control. Experimentally determined CTCF binding peaks in MEFs [39] are also indicated. Examples of genes that are bound by cohesin and changed expression in *Nipbl* +/- MEFs (top) and those genes that did not change expression (bottom) are shown. No cohesin binding peaks were found at the *Srp14* gene region.

A

Gene	Nipbl $^{+/-}$ Mutant	Nipbl siRNA depletion	Rad21 siRNA depletion	Rad21 binding
Cebpβ	0.67±0.04 *	0.71±0.00 **	0.50±0.13 *	P
Cebpδ	0.70±0.01 *	0.77±0.01 *	0.25±0.07 **	P
Ebf1	0.71±0.01 *	0.72±0.07 *	0.65±0.10 *	P, B, D
IL6	0.63±0.03 *	0.63±0.07 *	0.34±0.03 *	(D)
Avpr1a	0.78±0.02 *	0.49±0.16	0.68±0.06	P
Lpar1	0.71±0.01 *	0.63±0.03 *	0.54±0.04 *	P, B
Adm	0.73±0.03 *	0.71±0.00 **	0.52±0.01 *	P, D

B

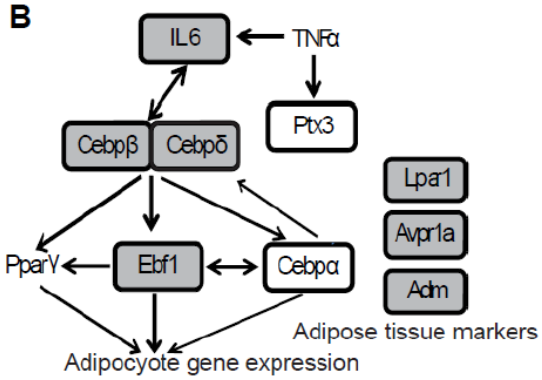


Figure 2-5. Cohesin plays a direct role in adipogenesis gene regulation.

(A) RT-q-PCR analysis of gene expression changes in *Nipbl* +/- mutant MEFs and MEFs treated with siRNA against Nipbl and Rad21 (* P<0.05, ** P<0.01). Cohesin binding status is also shown. P: promoter, B: gene body, and D: downstream as in Figure 5 with the exception of *IL6*. For *IL6*, the cohesin binding site in the downstream region is 3 kb away from TSS. **(B)** A schematic diagram of genes involved in the adipogenesis pathway. Genes that changed expression in *Nipbl* +/- mutant MEFs are circled, and those bound by cohesin and examined in (A) are shown with shaded circles.

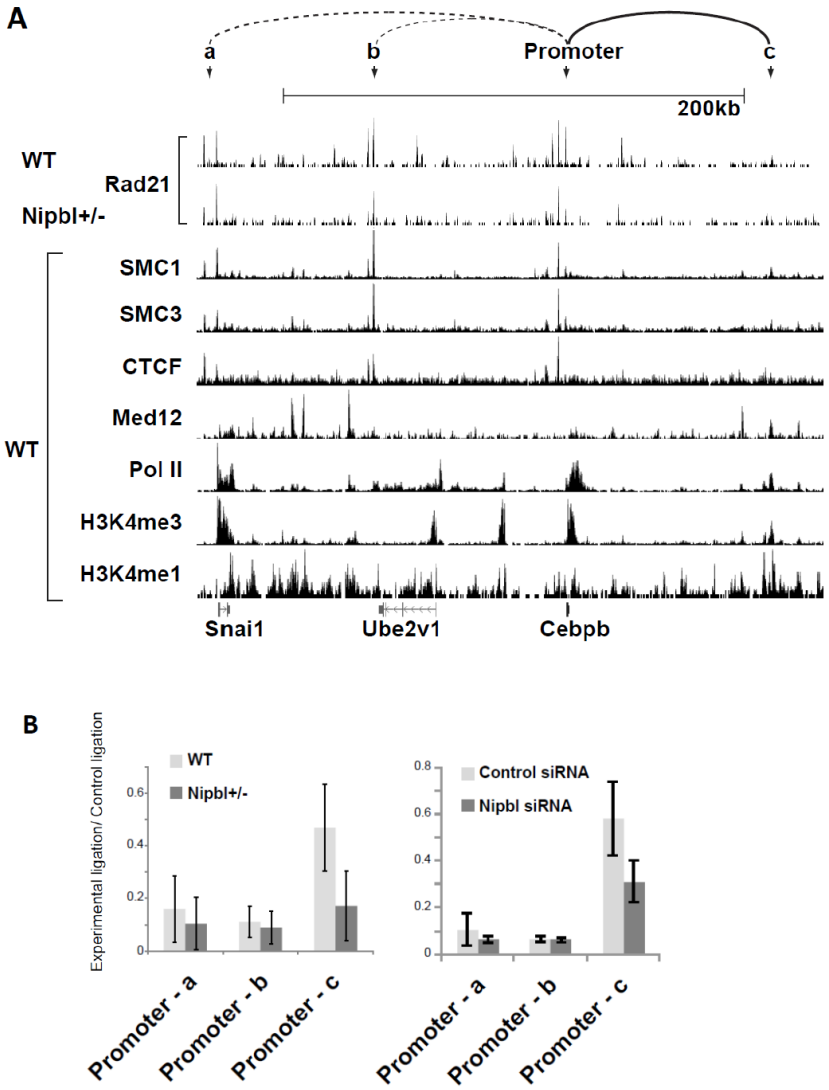


Figure 2-6. The long distance interaction involving the Cebpβ promoter is decreased in Nipbl +/- MEFs.

(A) Comparison of Rad21 binding peaks in wild type (WT) and *Nipbl* +/- mutant MEFs with SMC1 and SMC3, CTCF, and Mediator subunit 12 (Med12) [39] (GSE22562), pol II (GSE22302), H3K4me3 (GSE26657), and H3K4me1 (GSE31039) in WT MEFs in the genomic region surrounding the *Cebpβ* gene. The positions of primers for the 3C analysis (a, b, c and the promoter as the bait) are indicated. These regions were chosen based on the overlapping peaks of cohesin and CTCF, and/or cohesin, pol II and Med12 with H3K4me1/me3. The interaction observed by 3C in (B) is shown in a solid line and other interactions examined but weak are shown in dotted lines at the top. **(B)** The 3C analysis of *Cebpβ* promoter interactions with regions a, b, and c (as indicated in (A)). The chromatin interactions between WT and *Nipbl* mutant MEFs (top panel) and between control and Nipbl siRNA-treated MEFs (bottom) were quantified and normalized as described in METHODS.

(3C in WT and mutant cells were done by Richard Chien; 3C in siRNA treated cells were done by Yen-Yun Chen.)

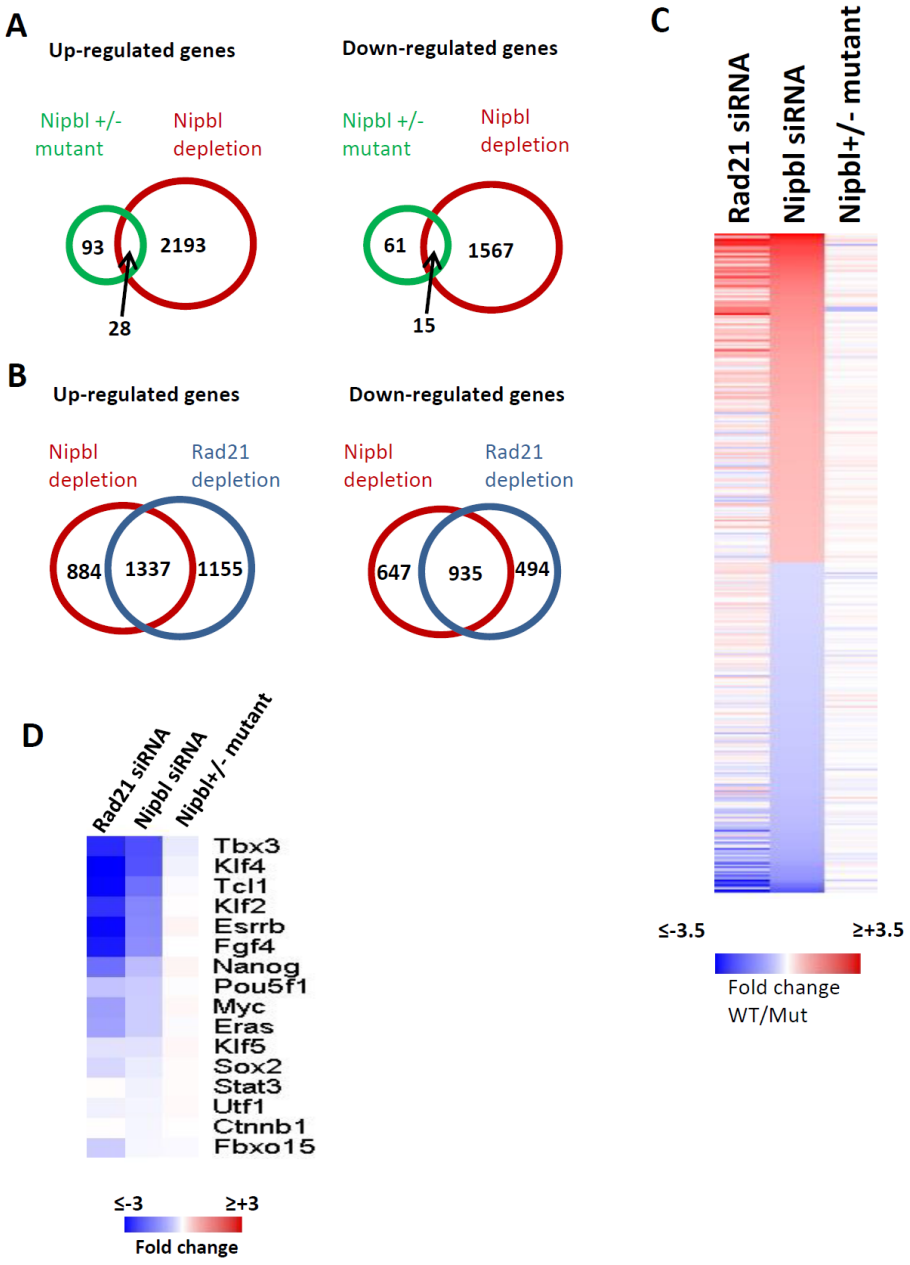


Figure 2-7. *Nipbl* heterozygous mutation and Nipbl siRNA depletion in mES cells generate different expression patterns

(A) Overlap of up- and down- regulated genes in *Nipbl* mutant mES cells and Nipbl siRNA treated mES cells. (Fold change >1.2, P<0.05) **(B)** Overlap of up- and down- regulated genes in Nipbl siRNA treated and Rad21 siRNA treated mES cells. (Fold change >1.2, P<0.05) **(C)** Expression heatmap of the top 200 most up- and down-regulated genes in Nipbl siRNA treated cells, and compare these genes' expression with Rad21 siRNA treated cells and *Nipbl* mutant mES cells. **(D)** Expression heatmap of genes involved in pluripotency in Nipbl siRNA treated cells, Rad21 siRNA treated cells and *Nipbl* mutant mES cells.

Table 2-1. The list of PCR primers

Unique regions ChIP primers	
pax2-F	CTGGCACTGACATCTTGTGG
pax2-R	TGGGACCTGTAGTCCTGACC
anapc13-F	TCCTAAGCCGTCCTGTAGTCC
anapc13-R	GGGTGTCCATCATCTGAGTCC
alox8-F	GTATGAGGTGGGCCTGAGTG
alox8-R	AAGCCCTGCCTAAATGTGTG
ebf1-F	AACTGAGCCTTAGGGGAAGC
ebf1-R	TCAGGGTTCAATCTCCAAGG
cebpb-F	AGAGTTCTGCTTCCCAGGAGT
cebpb-R	GGAAACAGATCGTTCCTCCA
il1R2-F	TGGAGGCAGTGGAAGAATCA
il1R2-R	ATCCTTGGCAGTGAACCAGA
fez1-F	GAGGGTGGGACGTATTTCAGT
fez1-R	CAGCCTTCTTTCCCTCACAA
pcdhb22-F	GCAGTAATGCCAGCAATGG
pcdhb22-R	TCCAGTTGGTTGGGTTTCAT
RT-qPCR primers	
Rnh1-F (Housing keeping gene)	TCCAGTGTGAGCAGCTGAG
Rnh1-R (Housing keeping gene)	TGCAGGCACTGAAGCACCA
Nipbl-F	AGTCCATATGCCCCACAGAG
Nipbl-R	ACCGGCAACAATAGGACTTG
Rad21-F	AGCCAAGAGGAAGAGGAAGC
Rad21-R	AGCCAGGTCCAGAGTCGTAA
Cebpb-F	GCGGGGTTGTTGATGTTT
Cebpb-R	ATGCTCGAAACGGAAAAGG
Cebpd-F	ACAGGTGGGCAGTGGAGTAA
Cebpd-R	GTGGCACTGTCACCCATACA
Ebf1-F	GCGAGAATCTCCTTCAAGACTTC
Ebf1-R	ACCTACTTGCCTTTGTGGGTT
Il6-F	TAGTCCTTCCTACCCCAATTTCC
Il6-R	TTGGTCCTTAGCCACTCCTTC
Avpr1a-F	TGGTGGCCGTGCTGGGTAATAG
Avpr1a-R	GCGGAAGCGGTAGGTGATGTC
Lpar1-F	ATTTCACAGCCCCAGTTCAC
Lpar1-R	CACCAGCTTGCTCACTGTGT
Adm-F	TATCAGAGCATCGCCACAGA
Adm-R	TTAGCGCCCACTTATTCCAC
Cebpb 3C primers	
cebpb-promoter	ACTCCGAATCCTCCATCCTT
cebpb-region-b	CCTGCCCTGTATCAAAGCAT
cebpb-region-a	CTGCCCAAATCAGTGAGGTT
cebpb-region-c	CCTCTGTGAGGTCTGGTCGT
cebpb-promoter-R	GGTGGCTGCGTTAGACAGTA
cebpb-region-a-R	GTTGTATCCCAAGCCAGCTC
cebpb-region-b-R	CTCCCCACTCTGTTCAGGAC
cebpb-region-c-R	TAACAGCAGGGATGGGTTCT

Table 2-2. Nipbl and Rad21 depletion levels in mutant and siRNA-treated MEFs

Gene	*Nipbl*+/- mutant	Nipbl siRNA	Rad21 siRNA
Nipbl	0.68±0.00	0.68±0.00	1.04±0.05
Rad21	0.94±0.02	0.99±0.02	0.26±0.02
CTCF	0.95±0.05	0.96±0.07	0.84±0.07

Table 2-3. Gene expression changes and cohesin binding status in MEFs

		Cohesin binding				
	Total	Gene region	Promoter	Gene body	Downstream	None
Total	218	115	61	83	20	103
Up-regulated	62	30	14	22	6	32
Down-regulated	156	85	47	61	14	71

(Fold change>1.2, *p*-value<0.05)

Table 2-4. Ontology analysis of cohesin target genes in MEFs

Biological processes enriched in cohesin target genes with cohesin binding at either promoters or gene regions. "Gene number" is the number of cohesin target genes that belong to a specific category; "Expected number" is the expected gene numbers that belong to a specific category at random.

Altered gene expression in Nipbl+/- MEFs associated with cohesin binding to the promoters					
Biological process	P value	Enrich-ment	Gene number	Expected number	Genes
development	2.96E-04	2.38	18	7.55	Avpr1a, Dner, Fgf7, Thbd, Hoxa5,Hoxb5, Cebpa, Cebpb, Rcan2, Lama2, Ebf1, Klf4, Hunk, Tgfb3, Irx5, Odz4, Ptpre, Lpp
metabolism	2.90E-03	1.50	33	22	Dner, Acvr2a, Hoxa5, Hoxb5, Trib2, Satb1, Cebpa, Cebpb, Gstm2, Amacr, Cd55, Dhrs3, Grk5, Ell2, Serpinb1a, Cyp1b1, Chst1, Hsd3b7, Aldh1a7, Npr3, Man2a1, Klf4, Hunk, Prkd1,Prdx5, Ercc1, Irx5, Odz4, Sox11, Ptpre, Ccrn4l, Rgnef, Bcl11b
cell communication	2.96E-03	1.82	21	11.53	Dner, Acvr2a, Trib2, Cd55, Grk5, Hunk, Odz4, Ptpre, Rgnef, Avpr1a, Fgf7, Thbd, Fam43a, Rcan2, Socs3, Lama2, Cxcr7, Tpcn1, Rerg, Tgfb3, Lpp
immune system	6.44E-03	2.06	14	6.81	Dner, Cd55, Hunk, Ptpre, Thbd, Lama2, Cxcr7, Cebpa, Cebpb, Gstm2, Klf4, Prdx5, Fcgrt, Cd302

Altered gene expression in Nipbl+/- MEFs associated with cohesin binding to the gene regions					
Biological process	P value	Enrichment	Gene number	Expected number	Genes
immune system	6.60E-06	2.34	30	12.83	Klf4, Dner, Thbd, Cd55, Lama2, Cd302, Cxcr7, Hunk, Cebpa, Cebpb, Gstm2, Fcgrt, Prdx5, Fmod, Crlf1, Prelp, Svep1, Plac8, Heph, Swap70, Mxra8, Sdc2, Colec12, Pcolce2, Flt4, Gbp1, Hck, Dusp14, Cd109, Ptpre
cell adhesion	1.33E-05	3.05	19	6.22	Dner, Cd55, Lama2, Fmod, Prelp, Svep1, Plac8, Heph, Mxra8, Sdc2, Colec12, Pcolce2, Flt4, Hck, Ptpre, Rerg, Vcan, Odz4, Rgnef
cell communication	1.65E-05	1.89	41	21.72	Dner, Cd55, Lama2, Fmod, Prelp, Svep1, Heph, Sdc2, Colec12, Pcolce2, Flt4, Hck, Ptpre, Rerg, Vcan, Odz4, Rgnef, Thbd, Cxcr7, Hunk, Crlf1, Dusp14, Cd109, Rcan2, Socs3, Fam43a, Trib2, Grk5, Tpcn1, Avpr1a, Fgf7, Acvr2a, Figf, Myh3, Tob1, Acvrl1, Moxd1, Tgfb3, Lpp, Wnt4
development	4.81E-05	2.11	30	14.22	Dner, Lama2, Fmod, Prelp, Heph, Sdc2, Colec12, Pcolce2, Flt4, Ebf1, Hck, Ptpre, Vcan, Odz4, Thbd, Hunk, Crlf1, Rcan2, Socs3, Avpr1a, Fgf7, Figf, Myh3, Tgfb3, Lpp, Klf4, Cebpa, Cebpb, Hoxa5, Hoxb5, Irx5
metabolism	1.91E-03	1.38	57	41.44	Dner, Heph, Pcolce2, Flt4, Hck, Ptpre, Odz4, Hunk, Klf4, Cebpa, Cebpb, Hoxa5, Hoxb5, Irx5, Cd55, Svep1, Rgnef, Dusp14, Cd109, Trib2, Grk5, Acvr2a, Acvrl1, Moxd1, Prdx5, Swap70, Satb1, Amacr, Dhrs3, Ell2, Npr3, Man2a1, Prkd1, Cyp1b1, Serpinb1a, Chst1, Hsd3b7, Aldh1a7, H6pd, Serpine2, Cyp7b1, P4ha2, Larp6, Mrps11, Aox1, Hdac5, Cpxm1, Eno2, Sox11, Prkcdbp, Ccrn4l, Ercc1, Pqlc3, Bcl11b

Table 2-5 Ontology of changed genes in Rad21 siRNA and Nipbl siRNA treated mES cells

	Category	Sample	Count	%	P Value	Fold
Up-regulated	cytoskeletal protein binding	Rad21 siRNA	106	4.65	1.21E-15	2.22
		Nipbl siRNA	70	3.78	8.93E-09	2.06
Down-regulated	transcription	Rad21 siRNA	155	11.76	2.08E-05	1.37
		Nipbl siRNA	230	15.52	1.41E-18	1.73
	mitochondrion	Rad21 siRNA	170	12.9	7.97E-28	2.35
		Nipbl siRNA	150	10.12	2.53E-13	1.80
	tRNA metabolic process	Rad21 siRNA	25	1.9	1.74E-07	3.42
		Nipbl siRNA	25	1.69	2.92E-06	2.93

Table 2-6 Ontology of changed genes in *Nipbl*+/- mES cells

	Category	Count	%	P Value	Fold
Down-regulated	pattern specification process	8	11.94	1.31E-05	9.64
	Homeobox	7	10.45	2.38E-05	11.75
	cell motion	8	11.94	6.38E-05	7.53
	axon guidance	5	7.46	1.64E-04	17.52

Table 2-7 30 developmental related genes down-regulated in the *Nipbl*+/- mES cells.

Gene	Description	Functions in development	Gene	Description	Functions in development
Ankrd6	ankyrin repeat domain 6	Brain, heart	Lix1	limb expression 1 homolog (chicken)	cortex and hindbrain
cdh2	cadherin 2; similar to N-cadherin	neural differentiation, myogenesis	Mbnl1	muscleblind-like 1 (Drosophila)	muscle
cdh4	cadherin 4, R-cadherin	axon guidance	Meis2	Meis homeobox 2	brain and eye
Chrdl1	chordin-like 1	CNS, ossification	mmp9	matrix metallopeptidase 9	skeleton
Cntfr	ciliary neurotrophic factor receptor	CNS	Pbx1	pre B-cell leukemia transcription factor 1	AP patterning, ventricular septum
Ebf1	early B-cell factor 1	Adipogenesis, B lymphopoeisis, CNS	Pknox2	Pbx/knotted 1 homeobox 2	development
Efnb1	ephrin B1	axon guidance, neural crest migration	Pmp22	peripheral myelin protein 22	myelin formation
Efnb2	ephrin B2	axon guidance, lymph vessel	rxrg	retinoid X receptor gamma	development
Emx2	empty spiracles homolog 2 (Drosophila)	cerebral cortex	Slit3	slit homolog 3 (Drosophila)	axon guidence, CNS
Fjx1	four jointed box 1 (Drosophila)	Limb, dendrite extension	Sp8	trans-acting transcription factor 8	cerebral cortex, DV patterning
Flrt2	fibronectin leucine rich transmembrane protein 2	cranialfacial morphogenesis	Tgfb2	transforming growth factor, beta 2	axon guidance, hematopoesis, heart
Fst	Follistatin	Pattern specification	Tnc	tenascin C	Neuron plasticity, olfactory bulb
Gfra1	glial cell line derived neurotrophic factor family receptor alpha 1	CNS	Zeb2	zinc finger E-box binding homeobox 2	CNS, neural crest, myogeneic differentiation
Hoxa1	homeo box A1	AP patterning, CNS	Zic2	zinc finger protein of the cerebellum 2	Neural tube closure
HoxB1	homeo box B1	AP patterning, facial nerve	Zic5	zinc finger protein of the cerebellum 5	Neural tube closure

Chapter Three

NIPBL's function in the nucleolus

Abstract

NIPBL haploinsufficiency is the major cause of Cornelia de Lange Syndrome (CdLS). The canonical function of NIPBL is to load cohesin onto chromatin. Since cohesin subunit mutations also cause CdLS or a CdLS-like disorder, the dogma is that CdLS is caused by dysfunction of the NIPBL-cohesin pathway. It is unclear how NIPBL mediates cohesin loading and whether it has any additional function(s) that may contribute to CdLS pathogenesis. We found that NIPBL binds to RNA and localizes to the nucleolus and interacts with ribosomal DNA (rDNA) repeats in an RNA-dependent manner. Mouse Nipbl depletion inhibits nascent ribosomal RNA transcription, suggesting that *NIPBL* haploinsufficiency affects nucleolar function and ribosome biogenesis. Nucleolar localization of Nipbl is affected by environmental stresses, and stress decrease Nipbl's association with rDNA. These results suggest that Nipbl plays a role in the nucleolar stress response pathways.

Introduction

Cohesin is a multifunctional protein complex essential for higher-order chromatin organization in all eukaryotes. The cohesin complex is composed of four evolutionarily conserved subunits: two SMC (Structural Maintenance of Chromosomes) family proteins designated SMC1 and SMC3, and the non-SMC proteins Rad21 and SA1/2. Cohesin mediates sister chromatid cohesion and functions in DNA damage repair and gene regulation [79,104,152,153]. It is loaded onto the chromatin at the end of telophase in mammalian cells by a loading factor complex composed of NIPBL and MAU2 [19,49,109].

Disruption of the cohesin pathway can result in developmental disorders such as Cornelia de Lange Syndrome (CdLS) (OMIM 122470,300590, 610759, 614701, 300882). CdLS patients demonstrate multisystemic abnormalities including growth and mental retardation, upper limb malformations, and heart and gastrointestinal defects [154,155]. CdLS is estimated to occur in about 1:10,000 births. Approximately 60% of CdLS cases are caused by mutations in the *NIPBL* gene, with a smaller percentage of cases caused by mutations in the *SMC1*, *SMC3* and *Rad21* genes (cohesin subunits) as well as the *HDAC8* gene (a deacetylase that affects the cohesin acetylation cycle) [26,65,66,68,69,70]. Although NIPBL is responsible for loading cohesin, patients with *NIPBL* mutations and cohesin mutations do not always share the same phenotypes. For example, *NIPBL* mutations cause limb abnormalities; however, patients with mutations in *SMC1* or *SMC3* usually have normal limbs and show mild mental retardation [65,66,67,68,69]. Distinct phenotypes between NIPBL and cohesin mutations raise the possibility that NIPBL may function beyond cohesin loading.

A *SMC1* missense mutation found in CdLS was introduced into budding yeast and caused defects in ribosomal RNA production and protein translation, suggesting that cohesin plays a role in ribosome biogenesis [156]. On the other hand, *NIPBL* (called *Scc2* in yeast) missense

mutation found in CdLS caused abnormal nucleolar morphologies in budding yeast [157]. Whether NIPBL has any functional significance regarding ribosome biogenesis in mammalian cells has not been investigated. The nucleolus is a non-membranous sub-nuclear compartment that forms around clusters of ribosomal DNA (rDNA) [158]. The main function of the nucleolus lies in the rapid production of ribosomal subunits. The nucleolus contains a "tripartite structure": the Fibrillar Center (FC), the Dense Fibrillar Component (DFC), and the Granular Component (GC) (Fig. 3-1) [158]. The three structures are functionally distinct. The FC is the site for pre-ribosomal RNA (pre-rRNA) transcription and is enriched in RNA polymerase I (Pol I) machineries. Pre-rRNA is processed in the DFC, which contains many RNA processing factors such as snoRNA and fibrillarin. The GC is the site where pre-ribosome assembly takes place, which contains ribosome assembly factors such as B23 (nucleophosmin). The nucleolus also acts as a sensor for different types of environmental stresses [158,159]. In response to stress, the nucleolus often undergoes dramatic reorganization that resulted in the disruption of the tripartite structure [158]. The ribosomal biogenesis shuts down as a protective mechanism for the cell to conserve energy, and p53 becomes activated in order to repair damages or undergo apoptosis [158,159].

Here we report that a sub-population of NIPBL protein localizes specifically to the FC/DFC regions in the nucleolus and is involved in ribosome biogenesis. NIPBL appears to have strong binding preference at rDNA transcribed regions and positively affects pre-rRNA level. Interestingly, NIPBL localization in the nucleolus is dependent on RNA. Consistent with this, we discovered that NIPBL binding to rDNA is partially RNA-dependent, and NIPBL directly interacts with rRNA. The results reveal a novel RNA-mediated mechanism of NIPBL recruitment to DNA. Nucleolar localization of Nipbl is affected by environmental stresses, and Nipbl's association with rDNA decreases upon stress. Our results demonstrate that NIPBL plays a functional role in mammalian nucleoli.

Materials & Methods

Cells and cell lines

Primary mouse embryonic fibroblasts (pMEFs) derived from E15.5 wild type and Nipbl mutant embryos were described previously [76]. In summary, mice heterozygous for *Nipbl* mutation were generated (*Nipbl* +/-) from a gene-trap-inserted ES cells. This mutation resulted in a net 30-50% decrease in *Nipbl* transcripts in the mice, along with many phenotypes characteristic of human CdLS patients [76]. Wildtype immortalized MEFs were described previously [160]. Wild type and *Nipbl* +/- pMEFs, immortalized wildtype MEFs, U2OS and 293t cells were cultured at 37°C and 5% CO_2 in DMEM (Gibco 31600-034) supplemented with 10% fetal bovine serum and penicillin-streptomycin (50U/mL). The KD3 human immortalized myoblasts were maintained as previously described [161].

Antibodies

Rabbit polyclonal antibodies against NIPBL and MAU2 proteins were raised against bacterially-expressed recombinant polypeptides and antigen affinity-purified. NIPBL-N antibody targets the sequences near the N terminus (a.a. 540 -919) while NIPBL-C antibody targets the C terminus (a.a. 2430 -2804) of human NIPBL variant A (NP_597677.2). MAU2 antibody targets a.a. 125-467 of human MAU2 (NP_056144.3). Rad21, SMC1, SMC3 antibodies and the preimmune IgG control were published previously [111]. The rest of antibodies used in this study and the applications were listed in table 3-2.

Drugs used to induce nucleolar stress

All drugs were added to cell culture media and incubated for different lengths of time in 37°C tissue culture incubator. Actinomycin D (Sigma) were used at a final concentration of 50ng/ml and incubated for 2 hrs. 2mM of AICAR (Abcam ab120358) was added to cell culture and incubated for 12 hours. Cells were treated with 0.5mM H_2O_2 (Ricca chemical, 381916) for 3 hrs. mTOR inhibitor INK128 (Active Biochem, A-1023) was used at 100 nM and treated cells for 24 hrs.

Immunofluorescence staining

MEFs and 293t were grown on coverslips in 24 well plates, fixed with 2% paraformaldehyde in PBS for 10 minutes, extracted with 0.2% triton X-100 in PBS for 4 minutes, and blocked in PBS/ 2.5% BSA/ 5% heat inactivated horse serum (Life Technologies 26050-070)/ 5% goat serum (Life Technologies 16210-064) for 1 hr at room temperature. Primary and secondary antibodies were diluted in 1X PBS/ 1% BSA/ 5% heat inactivated horse serum/ 5% goat serum. Coverslips were incubated in primary antibodies for 1 hr at room temperature followed by three PBS washes. Coverslips were incubated in secondary antibodies for 45 minutes at room temperature followed by three PBS washes. Then coverslips were stain with DAPI, washed by water and mounted by antifade (Life Technologies P-7481). For RNase treated cell staining, MEFs grown on coverslips were first treated with 100ug/ml RNase A (Life Technologies 12091-021) for 10 minutes at room temperature, then followed by the staining protocol mentioned above. Staining in KD3 cells were done using a different protocol. Cells were fixed in 4% paraformaldehyde in PBS for 10 min, then blocked/ extracted in SNBP/ 0.1% Gelatin/ 4% heat inactivated horse serum/ 4% goat serum/ 0.1% triton-X for 30 min at 37°C. SNBP buffer is PBS with 0.02% saponin, 0.05% NaN_3 and 1% BSA. Primary and secondary antibodies were diluted in SNBP/ 0.05% gelatin/ 1% heat inactivated horse serum/ 1% goat

serum. Coverslips were incubated in primary antibodies for 30 min at 37°C followed by three SNBP washes. Then coverslips were incubated in secondary antibodies for 30 min at 37°C followed by three SNBP washes. Then coverslips were stain with DAPI, washed by water and mounted by antifade. Antibodies used for staining and specific dilutions were listed in table 3-2.

Nucleolar fractionation

Nucleolar fractionation in MEFs was done according to the following protocol (http://www.lamondlab.com/pdf/noprotocol.pdf) with some modifications. 5 x 10^7 cells have been used for each nucleolar fractionation. All buffers (S1, S2 and S3) were reduced to half except for buffer A (5 ml). 5ml of cytoplasmic extract and 3ml of nuclear extract were obtained from 5 x 10^7 cells. The pelleted nucleoli were resuspended in 100 µl buffer S2. Same volume of each fraction was loaded onto SDS-PAGE gel for further analysis.

siRNA depletion

MEF cells were transfected using HiPerFect (Qiagen catalog) following the manufacturer's protocol. Media with siRNA and transfection reagents were removed 6 hours post transfection and fresh media were added. Second round of transfection was performed 24 hours after first transfection. Cells were harvested 48 - 72 hours after the first transfection. Nipbl siRNA sequences are as followed: Nipbl-1: 5'-GTGGTCGTTACCGAAACCGAA-3'; Nipbl-2: 5'-AAGGCAGTACTTAGACTTTAA-3'. Rad21 siRNA sequence: 5'-CTCGAGAATGGTAATTGTATA-3'. Control siRNA sequence: 5'-AATTCTCCGAACGTGTCACGT-3'.

FU labeling

5-fluorouridine (Sigma, F5130) was added to cell culture with final concentration of 2mM, followed by 5 min incubation in cell culture incubator. After 5 min, cells were removed of 5-

fluorouridine and kept in growth media for 30 min in cell culture incubator. After 30 min, cells were fixed and followed by immunofluorescence staining.

Chromatin immunoprecipitation

Approximately 6 x 10^6 cells were used per IP. Cells were cross-linked with 1% formaldehyde for 10 min at room temperature. Glycine was added to a final concentration of 0.125 M to stop cross-linking. Cells were washed twice with PBS and collected by scraping. Approximately 2 x 10^7 cells were resuspend in 1ml of Farnham lysis buffer (5 mM PIPES pH 8.0 / 85 mM KCl / 0.5% NP-40/ protease Inhibitors), centrifuged for 5min 2000rpm at 4°C. In terms of RNase treated ChIP, pellets were re-suspended in 1ml RIPA buffer (PBS / 1% NP-40 / 0.5% sodium deoxycholate / 0.1% SDS/ protease Inhibitors). RNase A was added to the lysate with final concentration of 500ug/ml. Both the lysate with and without RNase A were incubated at 37°C for 1 hour. SDS was added to the lysate after RNase treatment to a final concentration of 0.5% to facilitate sonication. For other ChIP experiments in the study, sample pellets were resuspended in SDS buffer (50mM Tris-HCl pH 8.0/ 10mM EdTA/ 1% SDS) after Farnham lysis and subjected to sonication. The lysates were sonicated using bioruptor (Diagenode UCD-200) to fragments averaged around 300-500bp. The extracts were diluted with ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl (pH 8.1), 167 mM NaCl) with protease inhibitors and centrifuged for 10 min at 10,000 g at 4 °C. Extracts were precleared for 1 h with protein A-Sepharose (GE Healthcare) supplemented with 1mg/ml BSA. 10% of the extract for each sample was taken as input DNA. Antibodies was added to the extracts for each IP and incubated overnight at 4°C on a rotator platform. Amount of antibodies used in ChIP is documented in table 3-2. The next day, the antibody-bound complexes were immunoprecipitated with protein A-Sepharose beads for 1 h and subsequently washed with low-salt buffer (0.1% SDS/ 1% Triton X-100/ 2 mM EDTA/ 20 mM Tris-HCl pH 8/ 150 mM NaCl),

high-salt buffer (0.1% SDS/ 1% Triton X-100/ 2 mM EDTA, 20 mM Tris-HCl pH 8/ 500 mM NaCl), lithium salt buffer (0.25 M LiCl/ 1% Nonidet P-40/ 1% deoxycholate/ 1 mM EDTA/ 10mM Tris-HCl pH 8), and TE (10mM Tris-HCl/ 1 mM EDTA pH 8.0). DNA was eluted off the beads with 200ul of elution buffer (1% SDS/ 0.1 M NaHCO3) on a rotator platform at room temperature for 1 hour. The eluates were collected and reverse crosslinked overnight at 65°C along with the input lysates. Quantitative PCR (q-PCR) was performed using the CFX96 real-time PCR detection system (Bio-Rad) with SYBR premix Ex Taq II (Clontech RR820B). ChIP DNA was purified with Qiagen PCR purification kits. ChIP signal was normalized by subtracting the preimmune IgG ChIP signal, and then divided by input DNA signal.

UV cross-linking and immunoprecipitation

UV cross-linking and immunoprecipitation (CLIP) was done according to the previous studies with some modifications [162,163]. The detail protocol of individual nucleoide CLIP sequencing (iCLIPseq) is also published [164]. The modifications are described as the following. After cell lysis, 15ul RQ1 DNase (Promega M610A) and 2ul RnaseOut (Life Technologies 10777-019) were added to the lysate and incubate at 37°C for 3min. RNaseOut were added to the lysate in order to maximize the recovery of intact RNAs. Approximately 5 µg of antibodies were used in each IP. For CLIP followed by RT-qPCR, 2 x 10^6 cells and 50 µg of protein A dynabeads (Life Technologies) were used per IP. After IP overnight and the washes, beads were resuspended in 25 µl RNase free water. 5min incubation at 95°C was done with gentle shaking. Then the 25 µl supernatant were divided into two tubes for reverse transcription using the protocol of SuperScript II reverse transcription (Life Technologies). Reverse transcriptase was added into one tube (RT+) but not the other (RT-). For the comparison of the CLIP signals between the two treatments (eg., control and Nipbl siRNA-treated cells), 20% of the cell samples were taken as input and subjected to total RNA extraction. The same volumes

of eluted total RNA were taken for reverse transcription. The IP signals from different treatments were normalized to the input during data analyses. For CLIP-SDSPAGE, 10^7 cells and 100 µl protein A beads slurry were used per IP. Amount of antibody used per IP is documented in Table 3-2.

RT-qPCR

Total RNA was extracted using the Qiagen RNeasy Plus kit. First-strand cDNA synthesis was performed with SuperScript II reverse transcriptase (Life Technologies). Quantitative PCR (q-PCR) was performed using the CFX96 real-time PCR detection system (Bio-Rad) with SYBR premix Ex Taq II (Clontech RR820B). Standard curves were generated for each primer pair using a serial dilution of cDNA or ChIP Input DNA. Values were generated based on threshold cycles (Ct) with respect to the efficiency of each primer pair. Pre-rRNA expression is normalized to *Rnh1* (house-keeping gene) expression.

Results

A sub-population of Nipbl resides in the nucleolus

By immunofluorescent staining using antibody specific for the C- terminus of human NIPBL protein, we found that a subpopulation of Nipbl specifically localizes to the nucleolus in mouse embryonic fibroblasts (MEFs) (Fig. 3-2A). Nipbl forms multiple small foci within the nucleolus marked by the nucleolar protein B23 (Fig. 3-2A left). B23 specifically resides in the GC region, the most outer layer of the tripartite structure. Nipbl foci appear to be internal to the B23-positive region, suggesting that the Nipbl foci are either in the FC or DFC region (Fig. 3-2A left). Further co-staining analysis with fibrillarin, a marker for the DFC region, revealed that the Nipbl foci are contained within the fibrillarin-positive area (Fig. 3-2A, right, and inset). This suggests that the Nipbl foci are in the FC region, the site for rRNA transcription. Similar Nipbl foci were also observed in mouse fetal brain tissues (Fig 3-2B), indicating that Nipbl foci are present in the nucleolus in both in vitro cell culture and in vivo tissue samples. Furthermore, nucleolar localization of NIPBL was observed in human KD3 immortalized myoblasts (Fig. 3-2C) and osteosarcoma U2OS cells (Fig. 3-2D).

In order to demonstrate the specificity of NIPBL-C antibody, *Nipbl* mutant MEFs and Nipbl siRNA-treated MEFs were used for immunostaining. Significant decrease or loss of the nucleolar Nipbl was observed in both Nipbl mutant and siRNA-depleted MEFs (Fig. 3-4A). Nipbl expression was partially reduced in mutant MEFs compared to the wild type MEFs as determined by RT-qPCR at the level comparable to that in siRNA-treated cells (Fig. 3-4B). While only 10% to 25% of the wildtype and control siRNA-treated cells are without notable Nipbl nucleolar foci, 35-55% of mutant cells or cells treated with siRNAs (siRNA1 or siRNA2) lost Nipbl nucleolar foci (Fig. 3-4C). These results strongly support the notion that the nucleolar foci

detected by NIPBL-C antibody are *bona fide* Nipbl protein. When Nipbl is depleted, it does not affect the B23 signals or the overall nucleolar structure (Fig.3-4D).

To further substantiate the above observation, cytoplasmic, nuclear and nucleolar fractions were isolated from MEFs (Fig. 3-3A) and 293T (Fig. 3-3B) and were subjected to western blot analysis using antibody specific for the N-terminal fragment of NIPBL [165]. The results confirm that a subpopulation of Nipbl/NIPBL is indeed present in the nucleolar fraction consistent with the immunofluorescent results using the NIPBL-C antibody. Taken together, our results demonstrate that a subpopulation of NIPBL/Nipbl localizes and forms foci in the FC region of the nucleolus in human and mouse cells.

Nipbl binds to the transcribed rDNA region and positively regulates pre- rRNA level

Since a subpopulation of Nipbl localizes to the FC region of the nucleolus where the rDNA repeats cluster, we examined whether Nipbl binds to rDNA. We performed ChIP-qPCR analysis using NIPBL-C antibody. We found that Nipbl binds to rDNA, preferentially at the rRNA coding region, rather than the intergenic spacer regions, in both mouse and human cells (MEFs and 293T cells, respectively) (Fig. 3-5B, 3-5D). NIPBL binding to the rDNA transcribed regions was also oberved in a recent NIPBL ChIP-seq analysis in human mammary cells using an antibody directed against the C terminus of *Xenopus* Nipbl [151]. In contrast, ChIP signals of cohesin and CTCF are much weaker at the rRNA coding region compared to the unique cohesin binding sites in the *Ebf1* and *Cebpβ* gene regions (Fig. 3-5B). The cohesin binding sites in the *Ebf1* and *Cebpβ* gene regions contain CTCF motifs (*Cebpβ*: Fig. 2-2A; *Ebf1*: This binding site is -7Kb upstream of *Ebf1*, which contains CTCF motif). Interestingly, Nipbl binds weakly to these two CTCF-positive cohesin binding sites compared to the rRNA coding regions (Fig. 3-5B). Although Nipbl is required for cohesin loading at virtually all cohesin binding sites (Fig. 2-2A, B), Nipbl binds to chromatin independent of cohesin, and Nipbl and cohesin binding signals do not

always show a linear correlation. Thus, the strong binding of Nipbl at the rDNA region with very low cohesin ChIP signal may reflect the cohesin-independent function of Nipbl at the rDNA region.

Since Nipbl localizes to the FC region, the site for rRNA transcription, and binds strongly to rDNA, we assayed whether Nipbl plays a role in regulating rRNA transcription. FU labeling of nascent rRNA transcripts in MEFs were performed in conjunction with Nipbl and B23 immunostaining. Signal intensities of the nucleolar Nipbl in each cell were compared to the FU or B23 signal. The Nipbl signal positively correlates with the FU signal reflecting the rRNA transcritpional activity (R=0.626) (Fig. 3-6B). In contrast, the Nipbl staining intensity does not correlate with that of B23 (R=0.01) (Fig. 3-6B). RT-qPCR analyses of the pre- rRNA transcripts (5'ETS and ITS) showed the significant decrease after Nipbl depletion (Fig. 3-6C). Both FU labeling and RT-qPCR of pre- rRNA showed that Nipbl plays a positive role in regulating pre–rRNA level. Although Myc was shown to stimulate rRNA synthesis [166] and cohesin promotes *Myc* expression [167], there is no significant *Myc* gene expression change in *Nipbl*+/- MEFs [76]. Unlike Eco1 mutation in yeast [156], there was no significant defect in expression of genes involved in ribosome biogenesis by *Nipbl* haploinsufficiency, suggesting that the phenotype is due more to the direct effect on Pol I transcription rather than an indirect effect on upstream genes [76].

Nipbl binds to ribosomal RNA

Since RNA is highly enriched in nucleoli, we tested whether the presence of RNA affects Nipbl localization in the nucleolus. We found that the RNase treatment completely dispersed Nipbl nucleolar foci (Fig. 3-7). The nucleolar protein B23 is not affected by RNase treatment (Fig. 3-7A, 3-7B) while fibrillarin is also dispersed by RNase treatment (Fig. 3-7C, 3-7D). The

results reveal that nucleolar proteins utilize different mechanisms to associate with the nucleolus, and Nipbl is one of the proteins whose nucleolar localization is dependent on RNA.

Since Nipbl depends on RNA to associate with nucleoli, it is possible that Nipbl interacts with RNA. By performing RNA IP after UV crosslinking (CLIP), we detected minimum binding of Rad21 or CTCF with 18S rRNA though it was higher than the preimmune IgG control (Fig. 3-8A). In contrast, robust binding of both 18S and 28S rRNAs was observed with Nipbl CLIP (Fig. 3-8A and B). Nipbl depletion by two different siRNAs resulted in significant decrease of Nipbl binding to both 18S and 28S rRNAs, confirming the specificity of the CLIP signals (Fig. 3-8B). Nipbl depletion level was shown (Fig. 3-4B). Furthermore, Rad21 depletion does not affect Nipbl binding to rRNA, indicating that Nipbl binding to RNA is not mediated by cohesin, consistent with the lack of significant RNA binding by cohesin (Fig. 3-8A and C). Rad21 depletion level was around 70% by RT-qPCR (data not shown). Nipbl CLIP followed by radiolabeling of RNA and SDS-PAGE analysis revealed that the majority of radiolabelled RNAs clustered above 250kDa, corresponding to the size of Nipbl. Although Nipbl antibody used also co-precipitates Mau2, the binding partner of Nipbl, no radiolabeling was observed around 66kDa, corresponding to Mau2 (Fig. 3-8D) [19]. The results indicate that Nipbl, but not Mau2, directly binds to RNA.

Since the association of Nipbl to the nucleoli requires RNA, we tested whether RNA is necessary for Nipbl association with rDNA. Nipbl ChIP in MEFs was performed with and without RNase treatment. RNase decreased Nipbl binding to the rDNA regions, but not to the unique cohesin binding sites (*Ebf1-In*, *Cebpβ* and *Cepbδ*) (Fig. 2-2A, 2-5A, 3-5B, 3-9A). Consistent with our observation that B23 localization to the nucleolus is RNase-insensitive (Fig. 3-7A), B23's association with rDNA is not affected by RNase treatment (Fig. 3-9B). The results indicate that Nipbl binding to the rDNA region is dependent on RNA. This demonstrates for the first time that RNA is involved in determining the binding specificity of Nipbl.

Nipbl responds to nucleolar stresses

One of nucleolus' functions is to sense and respond to environmental stresses. Upon environmental stresses, we found that Nipbl in the nucleolus changes its localization (Fig. 3-10, 11). AICAR is an AMP analog that mimics the energy depleted state in cells, thus creating metabolic stress [168]. INK128 is an mTOR inhibitor, which creates nutrient stress [169]. H_2O_2 induces oxidative stress and low concentration of actinomycin D specifically inhibits Pol I transcription [170,171]. When MEFs were treated by H_2O_2, AICAR and actinomycin D, Nipbl nucleolar foci become clustered to the "nucleolar cap" structure (Fig. 3-10A, 3-10B, and 3-11A) [172]. Actinomycin D is known to relocalize FC/ DFC components to nucleolar caps [172], consistent with our finding that Nipbl is an FC component. Consistently, fibrillarin, a DFC protein, also relocalized to caps in response to actinomycin D treatment; while B23, a protein in GC region, was unaffected (Fig. 3-11A). Mouse upstream binding factor (mUBF) is part of the Pol I initiation complex which binds to rDNA promoter and facilitates rRNA transcription. UBF resides in the FC region of the nulcleoli and relocalizes to the nucleolar cap upon pol I inhibition [172]. Consistent with this, UBF relocalized to the nucleolar cap together with Nipbl upon AICAR treatment (Fig. 3-10B). Interestingly, Nipbl relocalizes and clusters to the center of the nucleolus following the INK128 treatment (Fig. 3-10A bottom panel). In KD3 cells, H_2O_2 treatment almost completely displaced NIPBL from the nucleolar foci (Fig. 3-11B). This effect is different compared to the formation of nucleolar caps in MEFs (Fig. 3-10A, middle panel). The results indicate that NIPBL/Nipbl change its localization in the nucleolus in response to different types of stresses, indicating that NIPBL/Nipbl is part of the stress response in the nucleolus.

It is known that environmental stresses shut down rRNA synthesis, and we showed that Nipbl physically binds to rRNA and associate with rDNA in an RNA-dependent manner. Thus it is possible that when rRNA transcript is decreased upon stress, Nipbl's association with rDNA is also affected. We performed ChIP analyses comparing Nipbl binding to rDNA in MEFs with and

without stress. We found that both H_2O_2 and AICAR treatment resulted in significant decrease of Nipbl binding to rDNA (Fig. 3-12). Our results raised the possibility that decrease of Nipbl binding to rDNA is part of the Pol I silencing pathway central to the nucleolar stress responses in mammalian cells.

Discussion

Our findings demonstrate that Nipbl, besides affecting Pol II genes through cohesin (chapter 2), can also function in the nucleolus to regulate pol I transcription. We found that Nipbl is a positive regulator of rRNA synthesis and is also part of the nucleolar stress response pathway. We found that Nipbl binds to the rDNA region and stimulates rRNA synthesis. In response to stress, Nipbl dissociates from rDNA and relocalized to the nucleolar cap or to the nucleoplasm, contributing to rRNA transcriptional repression. The rRNA-dependent Nipbl binding to rDNA region may represent an important feedback mechanism to ensure rapid upregulation or cessation of rRNA synthesis (Fig.3-13).

The positive feedback model of rRNA transcription regulation by Nipbl

Our findings support a model that rRNA facilitates Nipbl binding to rDNA chromatin, and binding of Nipbl to rDNA further stimulate the production of rRNA, thus forming a positive feedback loop (Fig. 3-13 top panel). Ribosomal RNA transcription is repressed by different types of stress involving distinct stress response pathways. Nipbl dissociates from rDNA in response to the reduced amount of rRNA, resulting in the Nipbl relocalization. Dissociation of Nipbl from rDNA attenuates the rRNA production (Fig. 3-13 bottom panel). We compare the action mode of Nipbl to that of SIRT7, which deacetylates Pol I subunit PAF53 to facilitate Pol I complex binding to rDNA and thus promotes rRNA synthesis [173]. Similar to Nipbl, SIRT7 binds to rRNA and requires rRNA to associate with the nucleolus [173]. SIRT7 also uses this feedback mechanism to reinforce transcriptional repression in response to stress [173]. Whether SIRT7 and Nipbl functionally interact with each other requires further investigation. Interestingly, a recent study succeeded in assembling synthetic nucleoli and discovered that

active rRNA transcription is required for the formation of a functional nucleolus [174]. Without on-going rRNA transcription, nucleolar proteins will not be recruited and no tripartite structure can be formed [174]. The utilization of the feedback loop allows a rapid, switch-like effect on rRNA transcription in response to environmental cues. The similarity between SIRT7 and Nipbl mode of action, and the nucleolar assembly pathway indicates that positive feedback utilizing ribosomal RNA might be a common theme in the nucleolar regulatory pathways.

Functional roles of cohesin and related factors in the nucleolus

Studies have shown that cohesin and associated factors play functional roles in the nucleolus [156,175,176,177]. ESCO2 is an acetyltransferase that can acetylate SMC3, which is important for establishment of cohesion [23]. Roberts syndrome (RBS) is a disease caused by ESCO2 mutations [23]. RBS patient fibroblasts showed impaired ribosome biogenesis and abnormal nucleolar morphology [156]. The phenotypes can be corrected with the introduction of wildtype ESCO2, demonstrating that ESCO2 is important for nucleolar structure and ribosome biogenesis [156]. *SMC1* mutations were found to cause a mild form of CdLS [68,69]. The effects of *ESCO2* and *SMC1* mutations found in RBS and CdLS were assessed by introducing similar mutations in the corresponding yeast homologs, *eco1* and *smc1*, respectively [157]. *eco1-W216G*, a Roberts syndrome mutation, and *smc1-Q843Δ*, a CdLS mutation, both cause defective ribosome biogenesis and nucleolar morphology in *S. cerevisiae*, although defects caused by *smc1-Q843Δ* is not as severe as *eco1-W216G* mutation [156]. Interestingly, however, while *eco1-W216G* affects many Pol II genes that are involved in protein translation [156], ribosome biosynthesis is not a pathway enriched in ontology analyses in *Nipbl* +/- in MEFs (Table 2-4). A recent study showed that rDNA replication origin activity is also disrupted in the *eco1-W216G* mutant [177]. When *fob1*, a gene that codes for rDNA-specific replication fork

blocking protein, is deleted from the *eco1-W216G* mutant, not only DNA replication got restored, rRNA transcription is also recovered [177]. This shows that Eco1 can regulate rRNA transcription through affecting rDNA replication [177]. In contrast, our results suggest that Nipbl is directly acting on Pol I transcription. Thus ESCO2 and Nipbl affect nucleolar functions through distinct molecular pathways.

A mutation found in the human *NIPBL* gene associated with CdLS was introduced into the yeast homolog *scc2* (*scc2-D730V*), which caused abnormal nucleolar morphology [157]. However, this mutation failed to affect rRNA transcription or protein translation in *S. cerevisiae* [156]. On the contrary, we observed that Nipbl reduction lowered RNA transcripts without affecting the overrall nucleolar structure (Fig. 3-5, 3-3D). This might reflect the difference between the point mutation and haploinsufficiency, which cause different severities of the disease phenotypes [67]. Alternatively, how Nipbl functions might be different in budding yeast compared to mammalian cells.

Although Nipbl is the cohesin loading factor, strong Nipbl binding sites do not always correlate with strong cohesin binding sites [151,178]. At the rDNA transcribed region, cohesin subunit Rad21 binds weakly compared to Nipbl, while Nipbl appears to coat the entire transcribed region (Fig. 3-4, 3-8A, 3-10). Previous studies demonstrated cohesin binding to rDNA in both yeast and mammalian cells at the non-transcribed spacer region [101,179]. Furthermore, robust rRNA binding activity appears to be specific to Nipbl but not Rad21 (Fig. 3-4). Taken together, the Nipbl nucleolar pathway may be independent of cohesin. However, whether other cohesin subunits besides Rad21 bind to the coding region of rDNA, and whether they associate with rRNA have not been tested.

The possible contribution of NIPBL nucleolar functions to CdLS pathogenesis

There is a class of diseases called ribosomopathy, which includes syndromes with mutations of genes involved in ribosome biogenesis [166]. Interestingly, the phenotypes of many ribosomopathies overlap with those seen in CdLS [166]. Recently the RBS patient cells, in which ESCO2 is mutated, were found to exhibit impaired ribosomal biogenesis, suggesting that RBS is a ribosomopathy though the underlying mechanism is not well understood [156]. L-leucine had been used for treatment of various ribosomopathies [180,181], and also rescues developemtnal phenotypes in *ESCO2* depleted zebrafish [175]. This further supports the idea that RBS is a ribosomopathy. Our data indicates that Nipbl regulate rRNA synthesis and participate in the nucleolar stress response pathway. Perhaps some of CdLS phenotypes such as smaller body size and skeletal abnormalities may be caused by impaired ribosome biogenesis. A recent study also showed that L-leucine can ameliorate the developmental defects of Nipbl-depleted zebrafish [182]. These evidences suggest that CdLS may also be a ribosomopathy. However, whether the actual protein translation is affected by Nipbl depletion in mammalian cells remains to be tested.

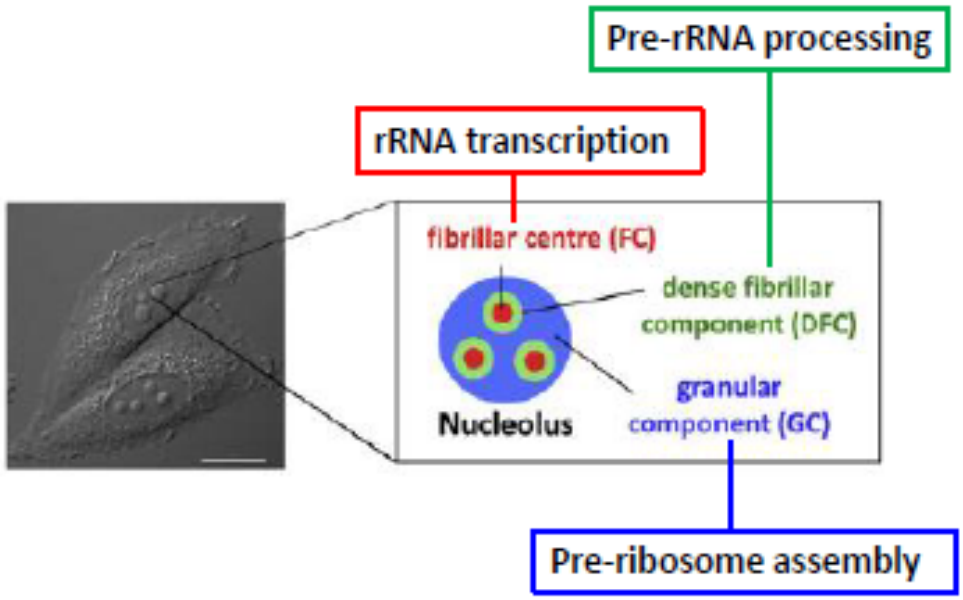


Figure. 3-1 Schematic diagram of the tripartite structure in nucleolus

(Modified from published paper [158].)

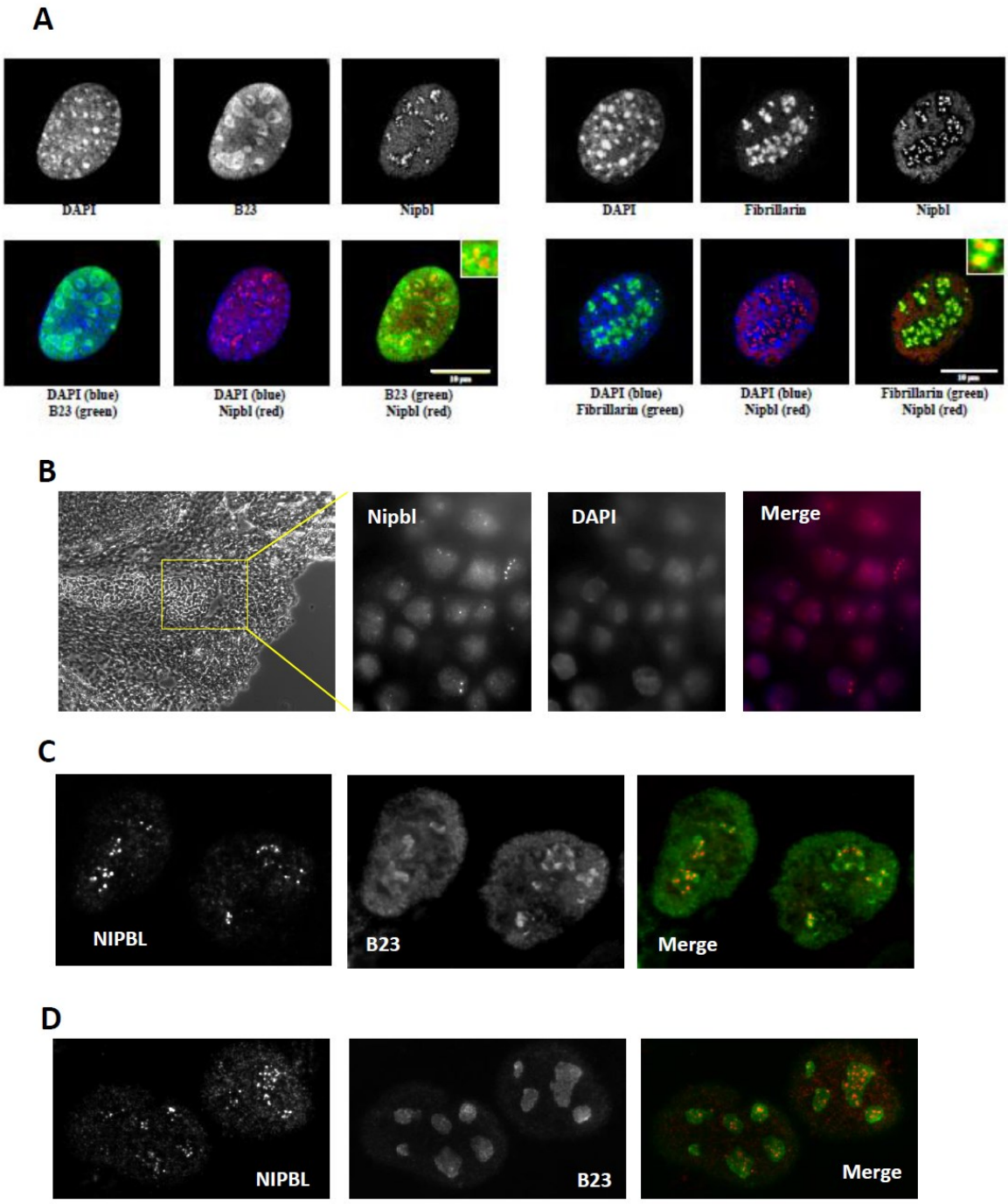


Figure. 3-2 Nipbl localization in nucleoli in mouse and human cells. (Immunofluorescence)

(A) Nipbl and B23 costain in MEFs (Left); Nipbl and fibrillarin costain in MEF (Right). **(B)** Nipbl staining in E15.5 mouse embryonic brain sections. **(C)** NIPBL and B23 costain in KD3 immortalized myoblasts. **(D)** NIPBL and B23 costain in U2OS cells.

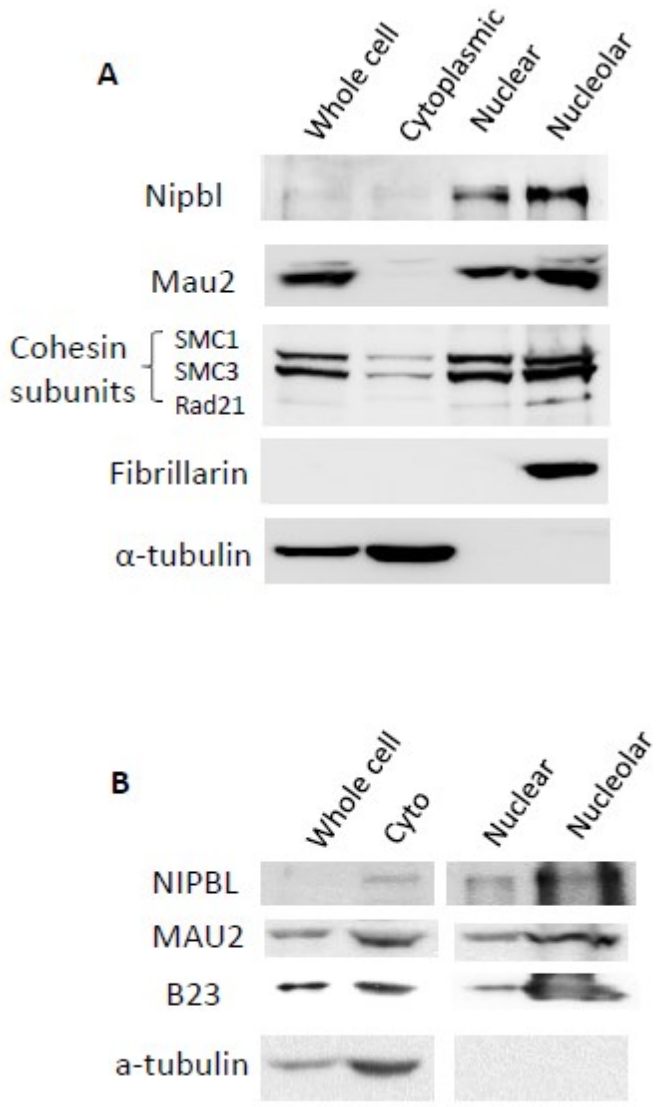
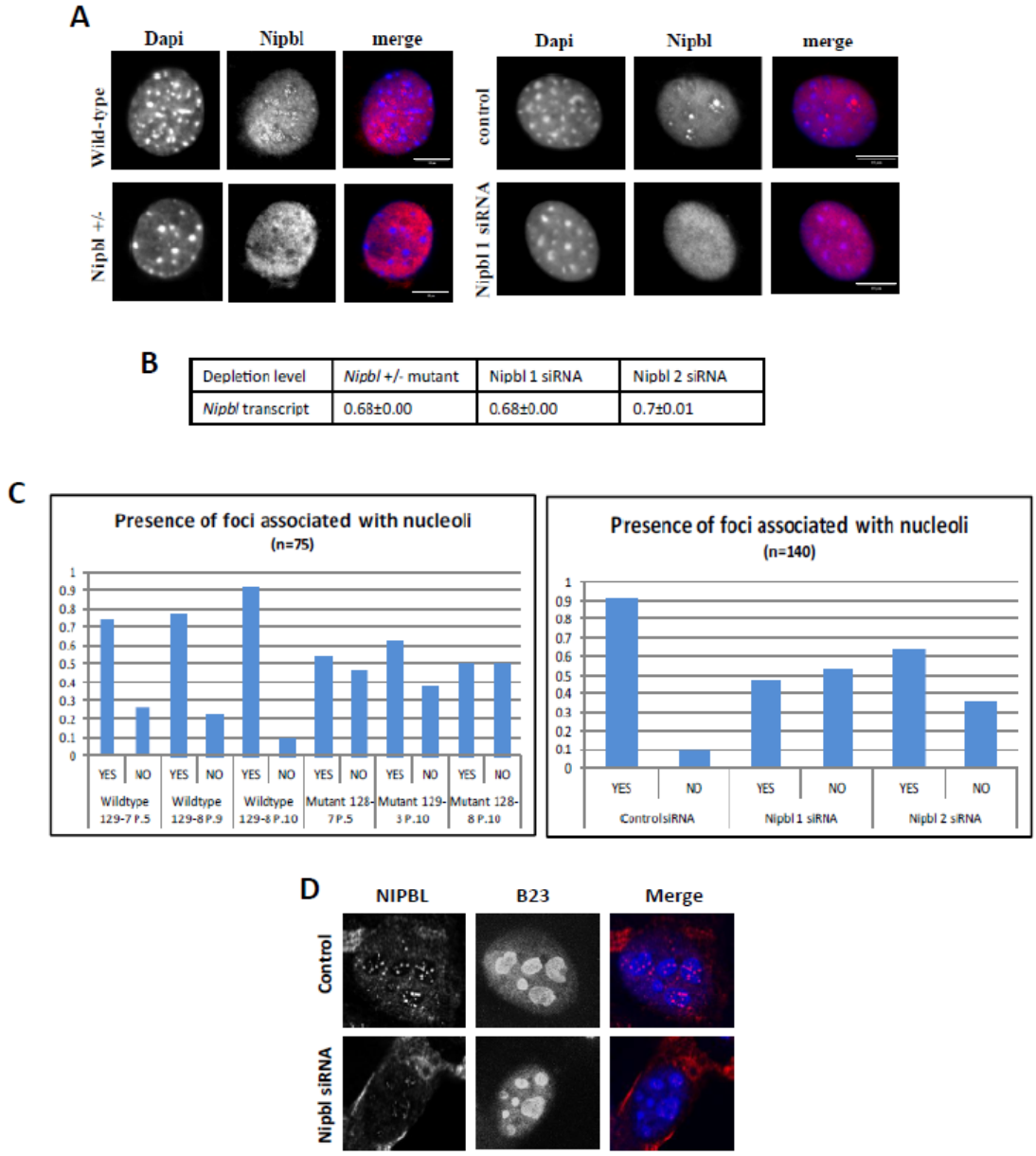


Figure. 3-3 Nipbl localization in nucleoli in mouse and human cells (Biochemical analyses)

(A) Cellular fractionation which separates the cytoplasmic, nuclear and nucleolar fractions in MEFs followed by Western blotting. **(B)** Cellular fractionation in 293T followed by Western blotting.



Depletion level	*Nipbl* +/- mutant	Nipbl 1 siRNA	Nipbl 2 siRNA
Nipbl transcript	0.68±0.00	0.68±0.00	0.7±0.01

Figure. 3-4 Nipbl nucleolar foci decrease in Nipbl siRNA- treated MEFs and *Nipbl* +/- MEFs

(A) Nipbl staining in wildtype and *Nipbl* +/- MEFs (left). Nipbl staining in control and Nipb siRNA treated MEFs (right). **(B)** Depletion level determined by RT q-PCR. Numbers in the table represents "amount of Nipbl when control is set as 1" ± "standard deviation". **(C)** Quantification of proportion of cells that have or don't have nucleolar Nipbl foci. **(D)** Nipbl and B23 co-staining in MEFs treated with control and Nipbl siRNA.

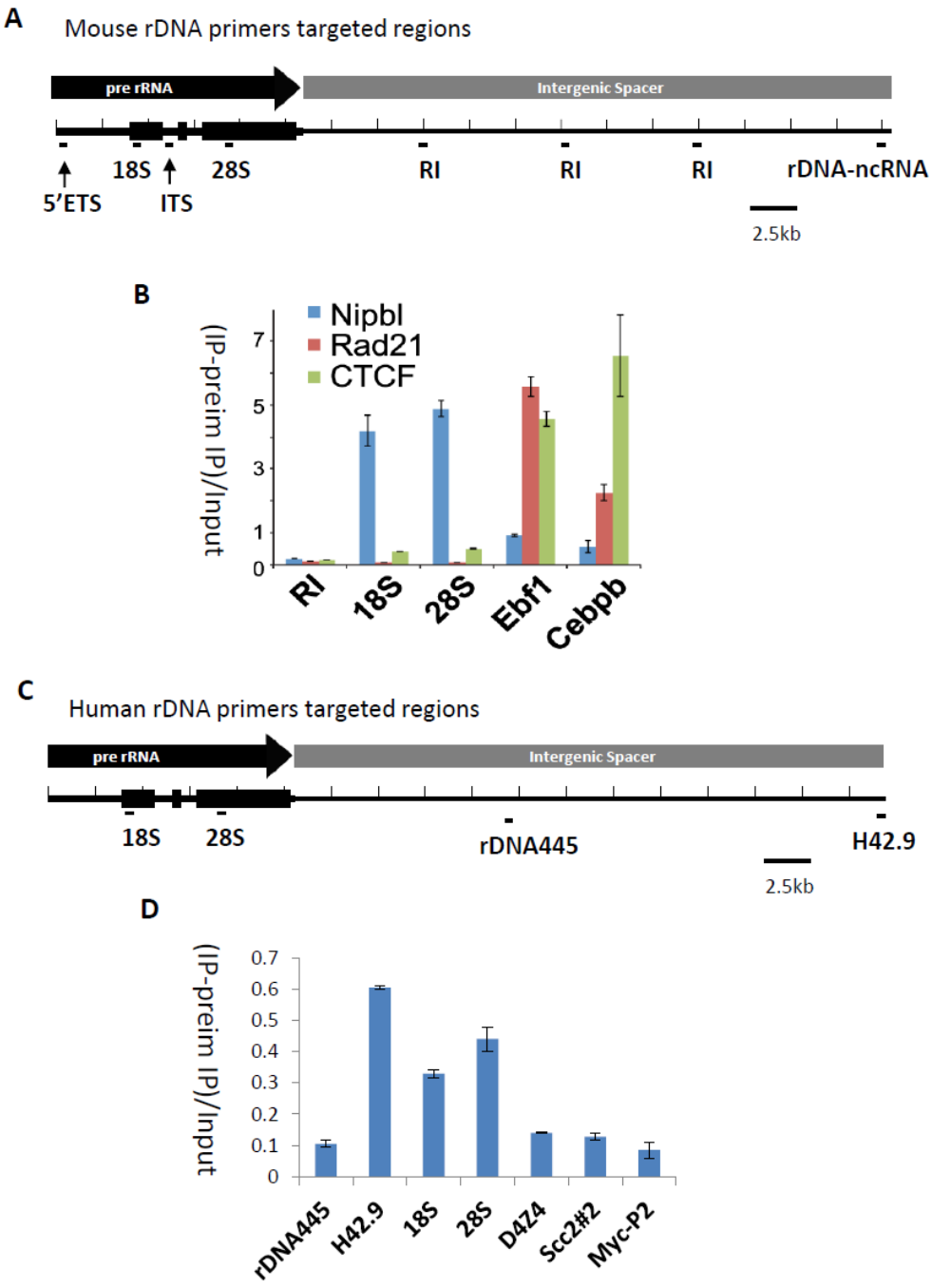


Figure. 3-5 Nipbl binds to the rDNA transcribed region
(A) Schematic diagram showing location of primer sets at the mouse rDNA region. **(B)** Nipbl, Rad21 and CTCF ChIP at rDNA region and two cohesin sites (Ebf1 and Cebpb) in MEFs. **(C)** Schematic diagram showing location of primer sets at the human rDNA region. **(D)** Nipbl ChIP at rDNA region and other known cohesin/ NIPBL sites. D4Z4 is known to be bound by both cohesin and NIPBL [101]. Scc2#2 is a known NIPBL site identified by us previously (data not shown). Myc- P2 is a known cohesin binding site.

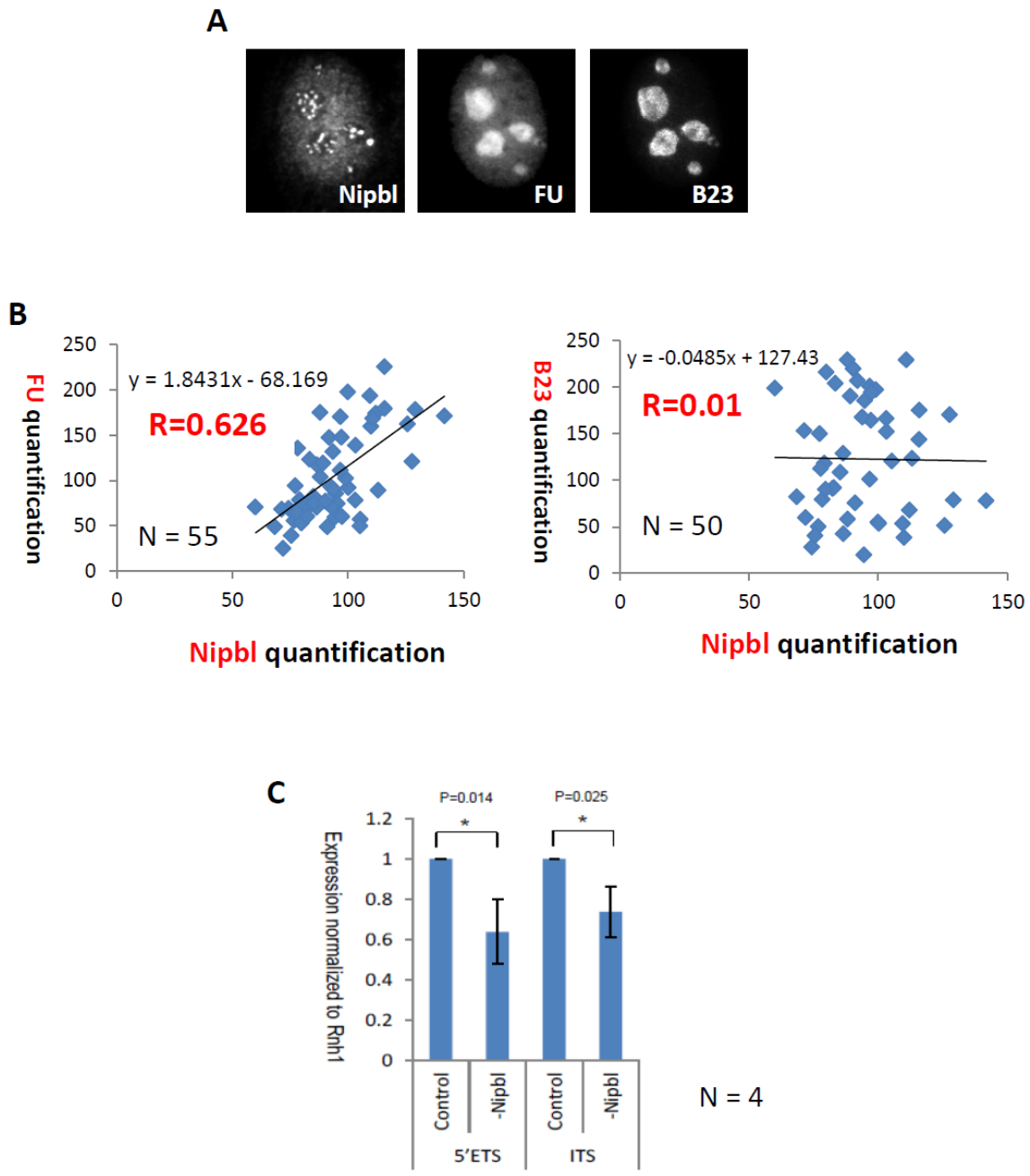


Figure. 3-6 Nipbl positively affects rRNA transcripts level

(A) Example of a MEF cell labeled with FU and triple- stained with BrdU, Nipbl, and B23 antibody. **(B)** Quantification of Nipbl nucleolar staining versus FU labeling signals (left). N= 55. Quantification of Nipbl nucleolar staining versus B23 nucleolar staining (right). N=50. **(C)** RT-qPCR checking the pre- rRNA transcripts (5'ETS and ITS) level after Nipbl depletion in MEFs. The siRNA used here is Nipbl-2 siRNA. Result is averaged from 4 independent experiments.

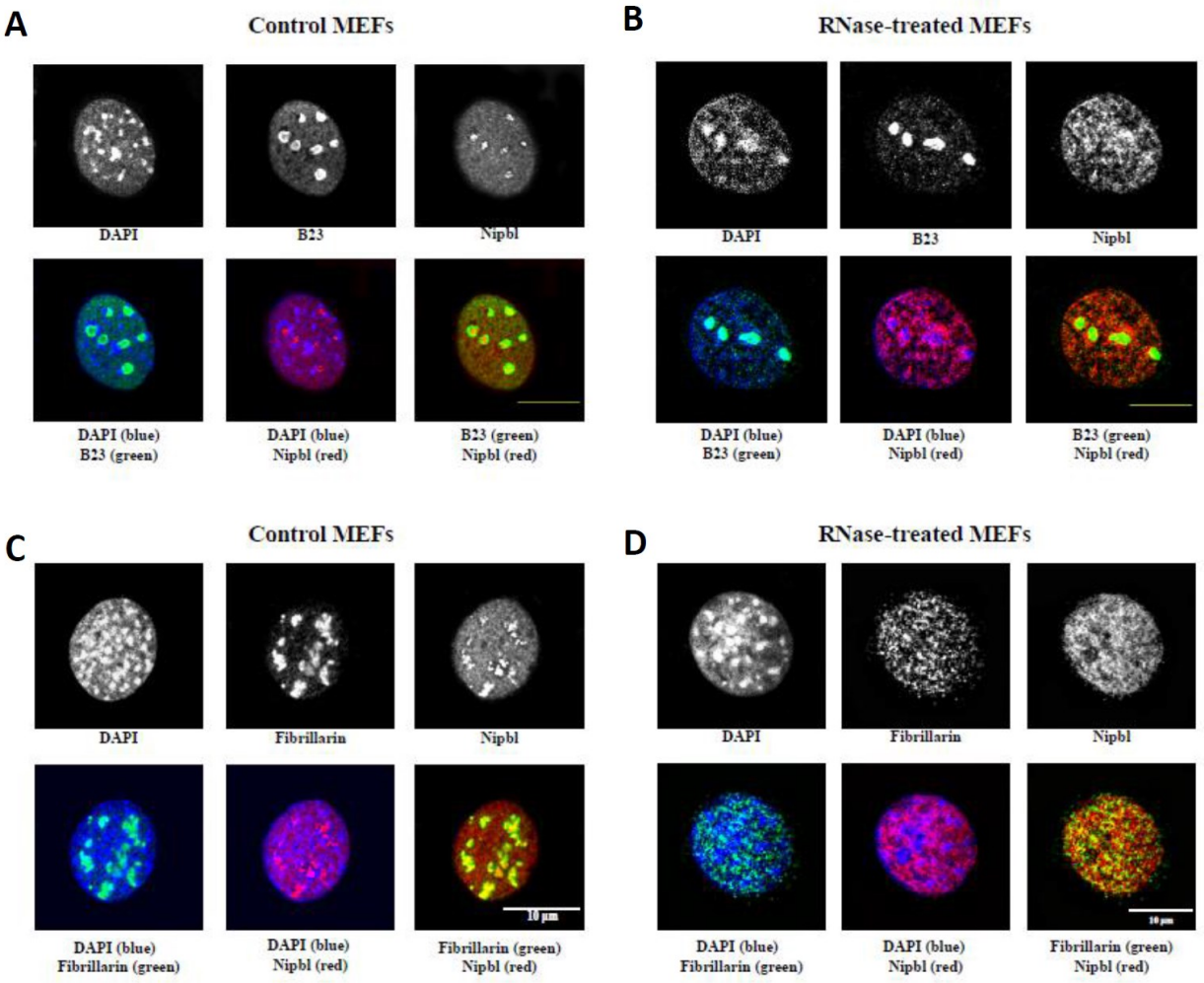


Figure. 3-7 Nucleolar localization of Nipbl is dependent on RNA.

(A) Control (no treatment) MEFs stained with Nipbl and B23 antibodies. **(B)** RNase- treated MEFs stained with Nipbp and B23 antibodies. **(C)** Control (no treatment) MEFs stained with Nipbl and fibrillarin antibodies. **(D)** RNase- treated MEFs stained with Nipbp and fibrillarin antibodies.

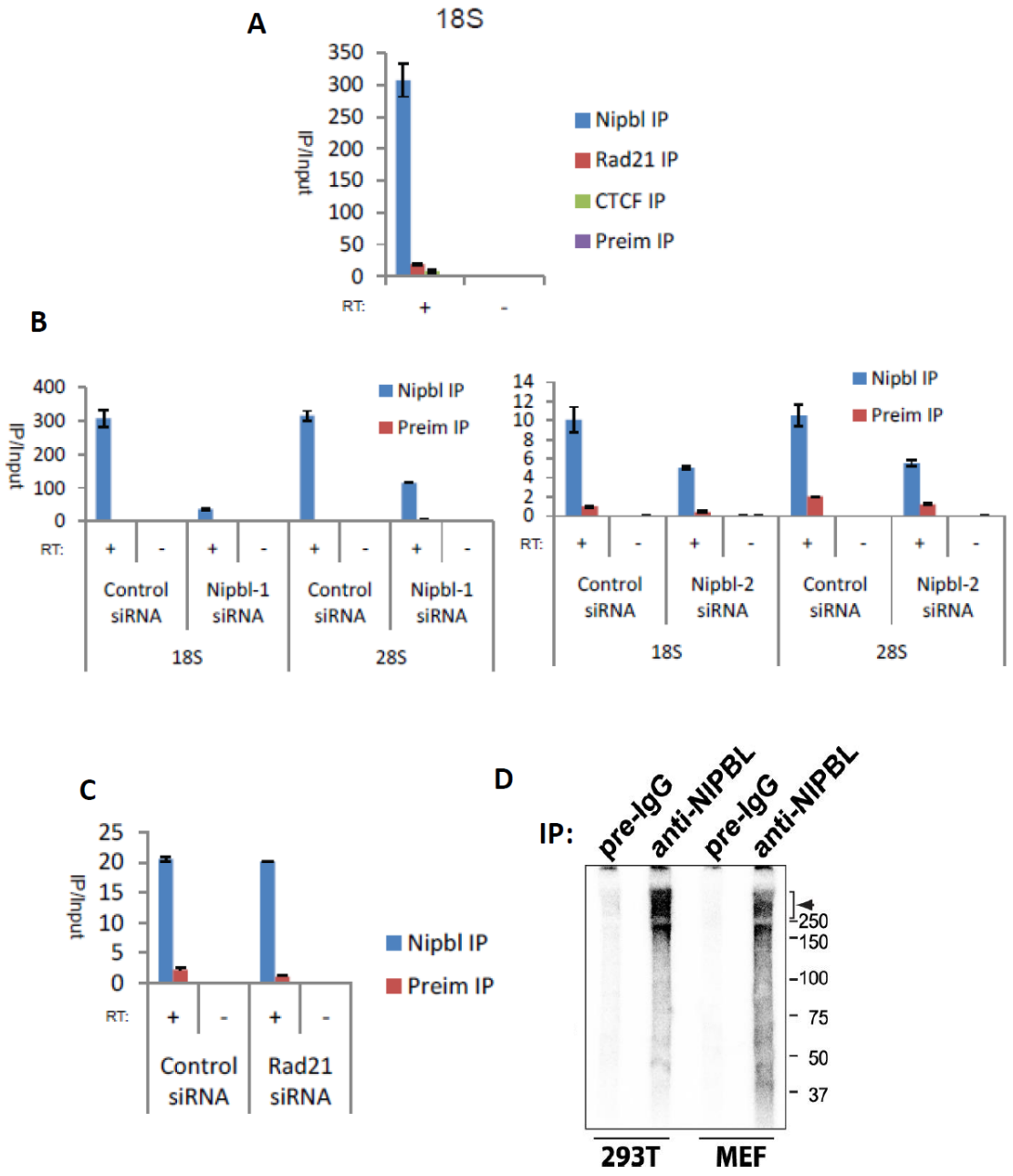


Figure. 3-8 Nipbl binds to rRNA

(A) CLIP analyses using Nipbl, Rad21 and CTCF antibodies for immunoprecipitation. RT-qPCR was done to check 18S rRNA amount in MEFs. **(B)** CLIP signal Comparison between Control and Nipbl-1 siRNA treated cells (Left). CLIP signal Comparison between Control and Nipbl-2 siRNA treated cells (right). **(C)** CLIP signal Comparison between Control and Rad21 siRNA treated MEFs. **(D)** CLIP followed by RNA radiolabeling and western blotting in both MEFs and 293T.

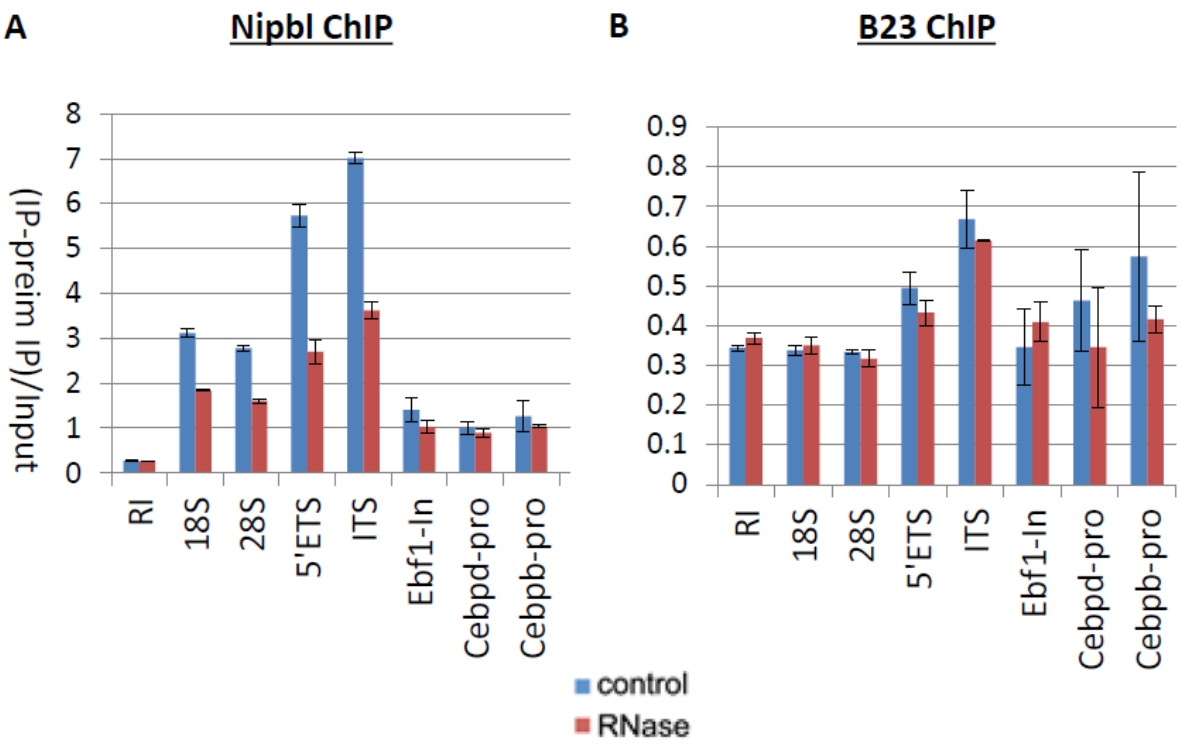


Figure. 3-9 RNA-dependent binding of Nipbl to rDNA chromatin

(A) Rnase treatment followed by Nipbl ChIP in MEFs. **(B)** Rnase treatment followed by B23 ChIP in MEFs.

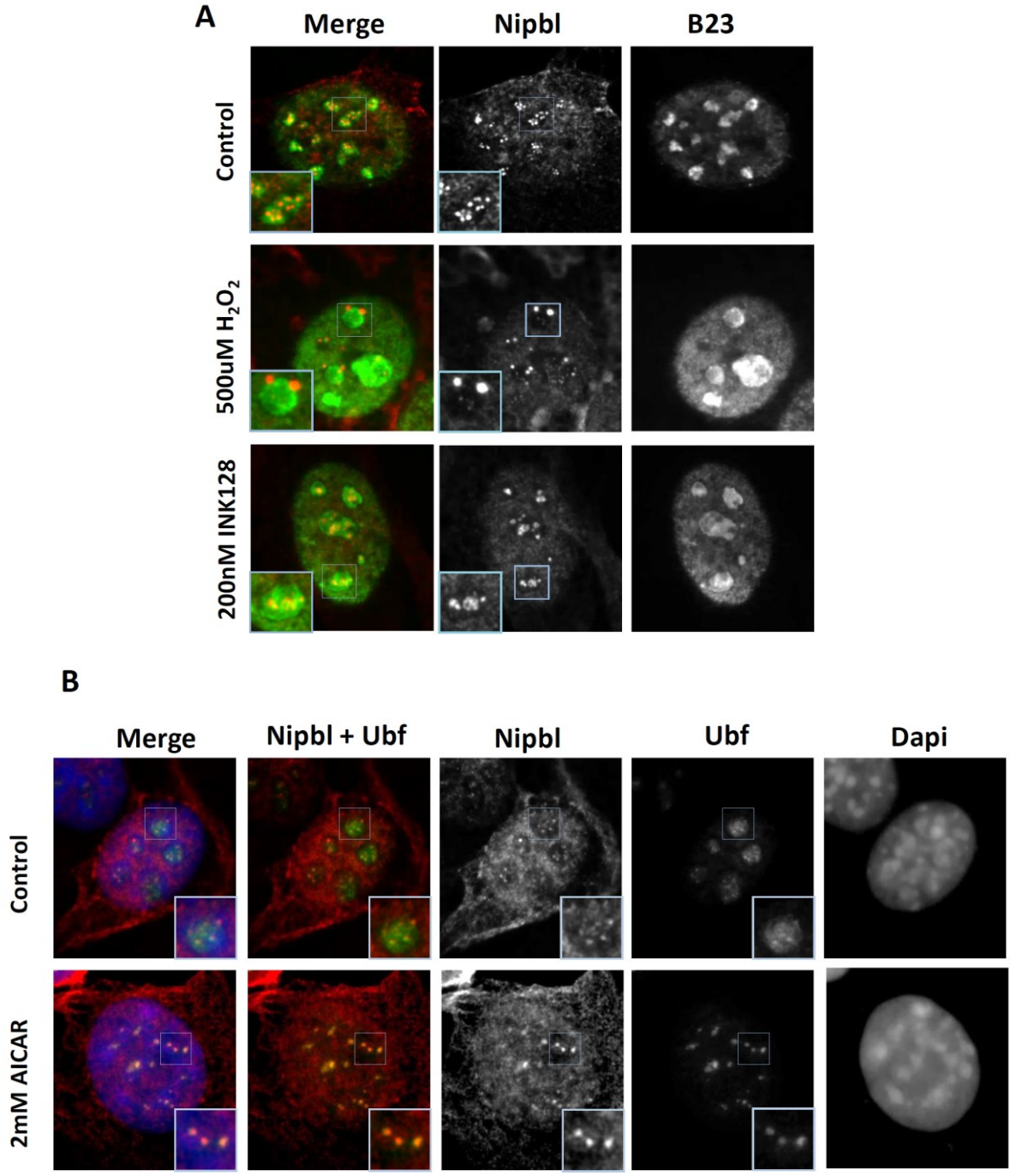


Figure. 3-10 Nipbl's nucleolar localization is affected by H_2O_2, INK128 and AICAR.

(A) Costaining of Nipbl and B23 in MEFs followed by H_2O_2 and INK128 treatment. **(B)** Costaining of Nipbl and Ubf in MEFs followed by AICAR treatment.

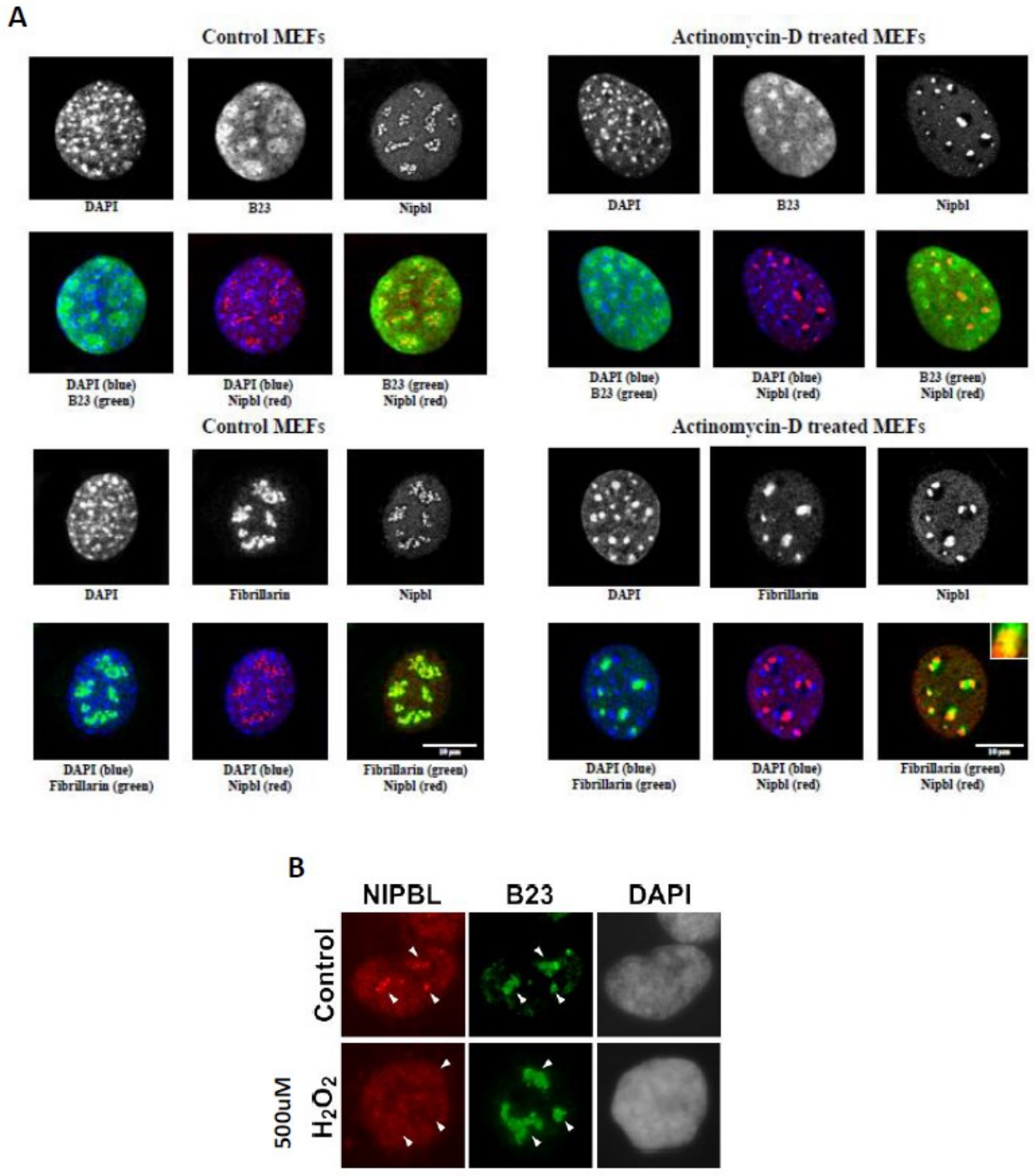


Figure. 3-11 Nipbl's nucleolar localization is affected by actinomycin D and H_2O_2.

(A) Costaining of Nipbl and B23 (top panel) and Nipbl and fibrillarin (bottom panel) after actinomycin D treatment. **(B)** Costaining of NIPBL and B23 in KD3 followed by H_2O_2 treatment.

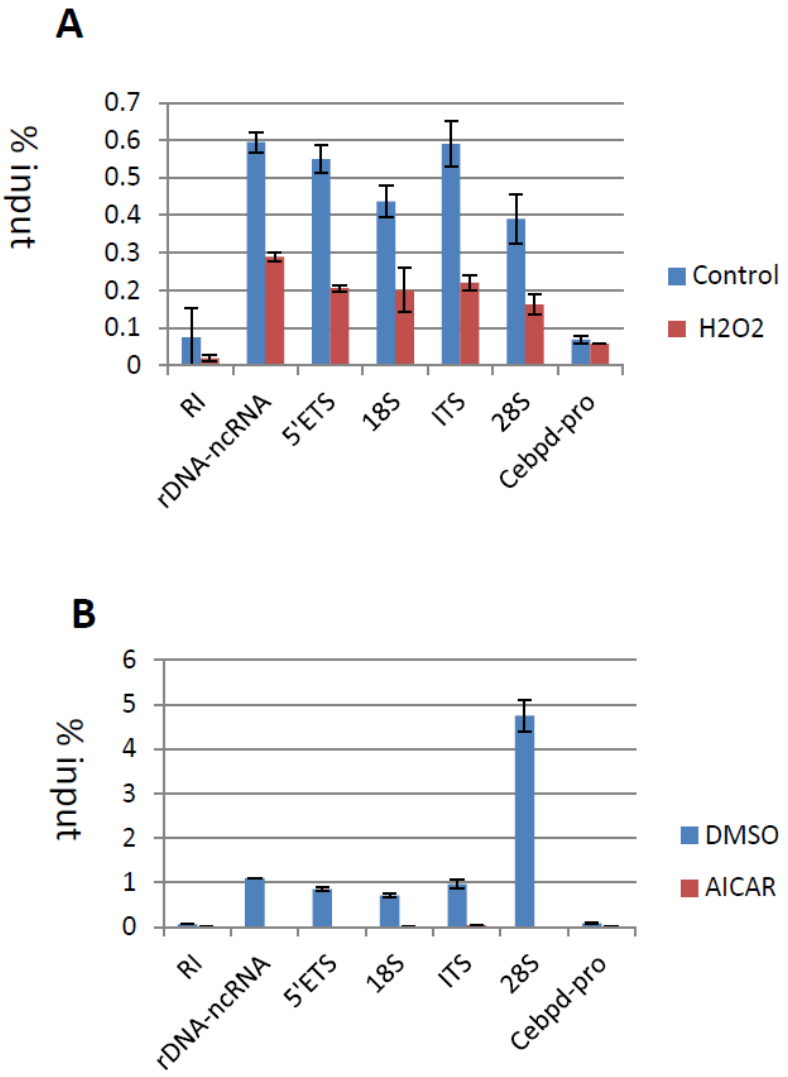


Figure. 3-12 Nipbl binding to rDNA chromatin decreases upon nucleolar stresses

(A) ChIP after H_2O_2 treatment in MEFs. **(B)** ChIP after AICAR treatment in MEFs.

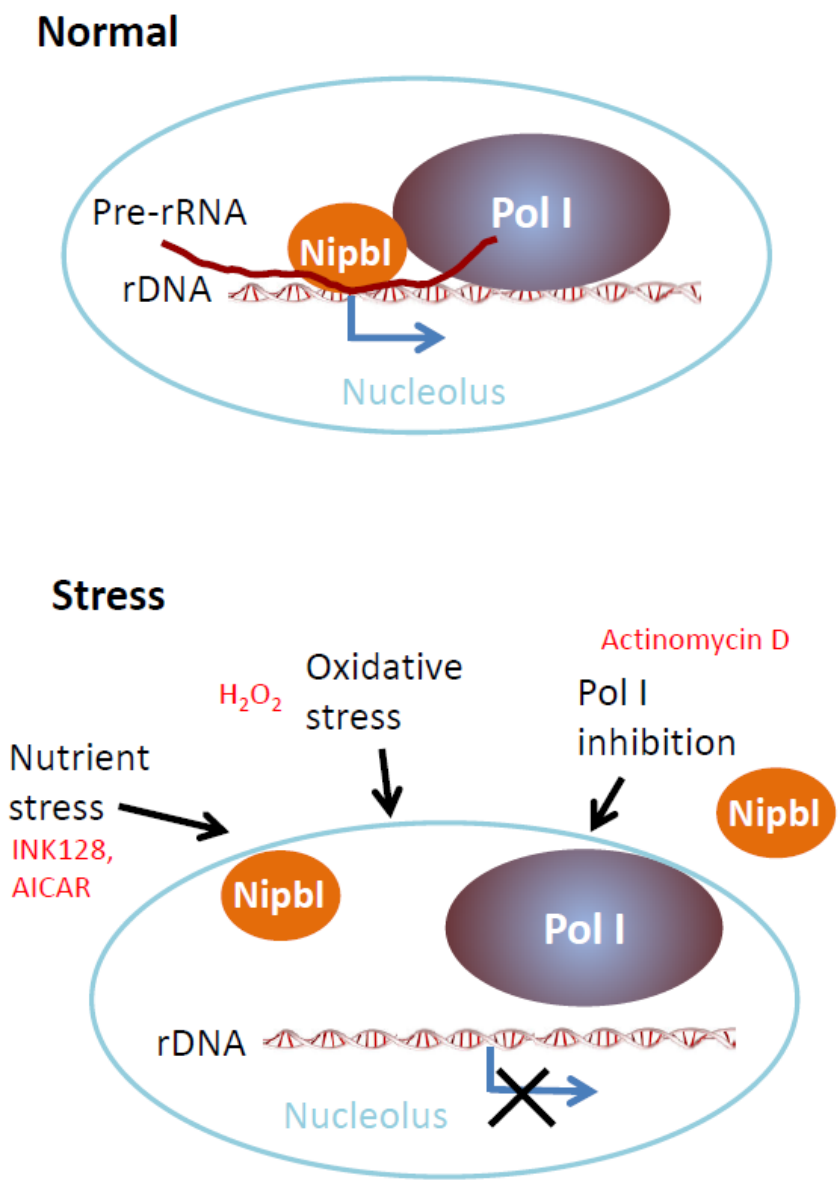


Figure. 3-13 Model of Nipbl's function in the nucleolus

Under normal circumstances, presence of pre-rRNA facilitates Nipbl binding to rDNA. Binding of Nipbl to rDNA further stimulates production of pre-rRNA.

When under adverse environment, stress signals to Pol I machineries to decreased affinity for the rDNA chromatin. rRNA production is in terms decreased. Decreased rRNA amount weakened Nipbl binding to rDNA chromatin, thus further strengthening the inhibition of rRNA transcription.

Table 3-1 The list of PCR primers

	Primer sequence	ChIP-qPCR	RT-qPCR	Species	Source
RI F	CCTGTGAATTCTCTGAACTC	V		Mouse	[113]
RI R	CCTAAACTGCTGACAGGGTG				
rDNA-ncRNA F	TGTTCGGGCGGGACGATG	V		Mouse	[183]
rDNA-ncRNA R	AGGTGTCGCCCGACAATG				
5'ETS F	CCAAGTGTTCATGCCACGTG	V	V	Mouse	[184]
5'ETS R	CGAGCGACTGCCACAAAAA				
18S F	GAGGCCCTGTAATTGGAATGAG	V	V	Mouse/ Human	[185]
18S R	GCAGCAACTTTAATATACGCTATTGG				
ITS F	CCGGCTTGCCCGATTT	V	V	Mouse	[184]
ITS R	GCCAGCAGGAACGAAACG				
28S F	TTGAAAATCCGGGGGAGAG	V	V	Mouse/ Human	[186]
28S R	ACATTGTTCCAACATGCCAG				
rDNA445 F	CCCGTGAGGAGTGATTTCTAA	V		Human	[101]
rDNA445 R	CGTGGGCCGGGCTGGTCTC				
Rnh1 F	TCCAGTGTGAGCAGCTGAG		V	Mouse	Self made
Rnh1 R	TGCAGGCACTGAAGCACCA				
H42.9 (rDNA-TSS) F	CCCGGGGGAGGTATATCTTT	V		Human	[187]
H42.9 (rDNA-TSS) R	CCAACCTCTCCGACGACA				
Ebf1- In F	CACTATGGAATCCGCTCAGG	V		Mouse	Self made
Ebf1- In R	CTCGCGGACAGTTTCTGATT				
Cebpb F	AGAGTTCTGCTTCCCAGGAGT	V		Mouse	Self made
Cebpb R	GGAAACAGATCGTTCCTCCA				
Cebpd F	CTCAGCAAGCCACAATTTCA	V		Mouse	Self made
Cebpd R	GCTCTGGCATTTCTTCTTGG				
D4Z4 F	CCGCGTCCGTCCGTGAAA	V		Human	[101]
D4Z4 R	TCCGTCGCCGTCCTCGTC				
Scc2#2 F	CGGGGTTTCCTACTTAGCAGAGCA	V		Human	Self made
Scc2#2 R	AAAGCCGCACCTATTCCAAACAAG				
Myc-P2 F	AGGGCTTCTCAGAGGCTTG	V		Human	Self made
Myc-P2 R	TGCCTCTCGCTGGAATTACT				

Table 3-2 Antibodies and their applications in the study

Antibody	Vendor	Immunostaining	Western blot	ChIP/ CLIP
Nipbl- C	(In-house)	1:200 - 1:500	1:1000	5ug
Nipbl- N	(In-house)	-	1:1000	-
Mau2	(In-house)	-	1:100	-
B23	GeneTex GTX10530	1:100	1:500	5ug
Fibrillarin	GeneTex GTX24566	1:500	1:2000	-
Ubf	Santa Cruz sc-13125X	1:200	1:2000	-
RPA194	Santa Cruz sc-48385	-	1:200	-
SMC1	(In-house)	-	1:5000	-
SMC3	(In-house)	-	1:5000	-
Rad21	(In-house)	-	1:5000	5ug
CTCF	Millipore 07-729	-	1:2000	10ul
α- tubulin	Sigma T9026	-	1:2000	-
BrdU	GeneTex GTX27384	1:100	-	-
Pre- IgG	(In-house)	-	-	1ul

Chapter Four

D4Z4 Heterochromatin and FSHD

Part of this chapter was published in a paper entitled-

"Genetic and Epigenetic Characteristics of FSHD-Associated 4q and 10q D4Z4 that are Distinct from Non-4q/10q D4Z4 Homologs"

Weihua Zeng,* Yen-Yun Chen,* Daniel A. Newkirk, Beibei Wu, Judit Balog, Xiangduo Kong, Alexander R. Ball Jr., Simona Zanotti, Rabi Tawil, Naohiro Hashimoto, Ali Mortazavi, Silv`ere M. van der Maarel, and Kyoko Yokomori.¶

* These authors contributed equally to the work.

¶ Corresponding author.

Abstract

Facioscapulohumeral dystrophy (FSHD) is one of the most prevalent muscular dystrophies. The majority of FSHD cases are linked to a decreased copy number of D4Z4 macrosatellite repeats on chromosome 4q (FSHD1). Less than 5% of FSHD cases have no repeat contraction (FSHD2), most of which are associated with mutations of *SMCHD1*. FSHD is associated with the transcriptional derepression of *DUX4* encoded within the D4Z4 repeat, and SMCHD1contributes to its regulation. We previously found that the loss of heterochromatin mark (i.e. histone H3 lysine 9 tri-methylation (H3K9me3)) at D4Z4 is a hallmark of both FSHD1 and FSHD2. However, whether this loss contributes to *DUX4* expression was unknown. We found that the suppression of H3K9me3 resulted in displacement of SMCHD1 at D4Z4 and increases *DUX4* expression in myoblasts. The results indicate the significance of the loss of D4Z4 heterochromatin in FSHD pathogenesis, supporting the notion that FSHD is an epigenetic abnormality disease. To investigate whether there are additional heterochromatin changes in FSHD and possibly define the disease-specific epigenome, we attempt to examine potential alterations of H3K9me3 and H3K27me3 as well as HP1γ and cohesin's association with chromatin in primary FSHD myoblasts compared to the control myoblasts in a global scale. We also utilize KD3, an immortalized human myoblasts cell line to interrogate the epigenetic changes associated with myotube differentiation. These experiments will further our understanding of the relationship between gene expression and epigenetic landscapes associated with myogenesis and FSHD development.

Introduction

FSHD is an autosomal dominant muscular dystrophy characterized by progressive wasting of facial, shoulder, and upper arm musculature [88]. The majority of FSHD cases (>95%) are caused by monoallelic partial deletion of D4Z4 repeat sequences at the subtelomeric region of chromosome 4q (4qter D4Z4) (FSHD1 (OMIM 158900)) [88,90]. D4Z4 is a 3.3 kb macrosatellite repeat that contains an open reading frame for the double-homeobox transcription factor DUX4 retrogene (OMIM 606009) [92,93,94]. There are only one to ten D4Z4 repeats in the contracted allele in FSHD1, in contrast to 11~150 copies in the intact allele. In the more rare form of FSHD (<5% of cases), there is no D4Z4 repeat contraction (FSHD2) [91,188]. A recent study found that the *SMCHD1* gene (OMIM 614982) is mutated in >80% of FSHD2 cases (OMIM 158901) [100].

FSHD occurs only in individuals with a 4qA haplotype with specific single-nucleotide polymorphisms in the chromosomal region distal to the last D4Z4 repeat (creating a non-canonical polyadenylation signal for the *DUX4* transcript) [97,189,190,191] with some exceptions [192]. While multiple transcripts encoding different parts of the DUX4 protein have been identified [193], expression of the full-length *DUX4* transcript (*DUX4fl*) is most closely associated with FSHD [94,97]. Overexpression of *DUX4fl* caused differentiation defects in human myoblasts and mouse C2C12 muscle cells, and FSHD-like phenotypes in zebrafish [96,194,195]. Furthermore, though the *DUX4fl* expression can occasionally be observed in unaffected individuals at very low levels (suggesting the presence of additional disease modifier genes), activation of a subset of the *DUX4fl* target genes has been observed in patient cells in multiple studies, supporting the significance of *DUX4fl* in FSHD [93,196,197,198,199].

The chromatin environment plays a significant role in gene regulation in normal development and disease [200]. Epigenetic alteration of D4Z4 chromatin was found to be a common link between FSHD1 and FSHD2 [101,188]. D4Z4 repeats contain transcriptionally repressive heterochromatin harboring DNA hypermethylation and histone H3 lysine 9 trimethylation (H3K9me3) together with H3K27me3 [101,188] . Using chromatin immunoprecipitation (ChIP) analysis, we found a specific loss of H3K9me3 at the D4Z4 repeat sequences in both FSHD1 and FSHD2 patient proliferating cell cultures [101]. Importantly, this change is highly specific for FSHD; no significant change of H3K9me3 was observed in other muscular dystrophies, some of which share similar clinical phenotypes [101]. This change is seen not only in affected muscle cells, but also in patient fibroblasts from skin biopsies and lymphoblasts from blood samples [101]. This indicates that the loss of H3K9me3 is not an epiphenomenon of dystrophic muscle. Although D4Z4 DNA was also shown to be hypomethylated in FSHD, we showed that the H3K9me3 loss is not a downstream consequence of DNA hypomethylation since H3K9me3 is intact in the phenotypically unrelated immunodeficiency-centromeric instability-facial anomalies (ICF) syndrome, in which D4Z4 is severely DNA-hypomethylated [101] due to mutations in the DNA methyltransferase 3B (*DNMT3B*) gene [201,202]. Nevertheless, the loss of DNA methylation and H3K9me3 indicate the perturbation of heterochromatin structure at D4Z4 in FSHD, strongly suggesting that FSHD is an epigenetic abnormality disease associated with the impairment of heterochromatin at D4Z4. We also found that the heterochromatin binding protein HP1γ and the higher-order chromatin organizer cohesin are co-recruited to D4Z4 in a H3K9me3-dependent and cell type-specific manner, and are lost in FSHD as a consequence of the loss of H3K9me3 [101]. FSHD-specific loss of heterochromatin in this region is thought to contribute to the derepression of *DUX4* [93,94,97]. However, this hypothesis has not been explicitly tested. In this study, we report the effect of the inhibition of H3K9me3 on *DUX4fl* expression. We found that decreased H3K9me3

results in the reduced SMCHD1 binding to D4Z4 and derepression of *DUX4fl* expression, demonstrating the significance of the loss of the H3K9me3 heterochromatin at D4Z4 in gene regulation.

Besides *DUX4*, there were evidences of global expression changes associated with FSHD [203,204]. There may be other key genes important for FSHD pathogenesis. HP1γ is known to have the property to spread gene silencing [205] and also has been suggested to promote chromatin interactions [206]. Cohesin is known to mediate long distance chromatin interactions to regulate gene expression [79]. One possibility for the involvement of D4Z4 heterochromatin in gene regulation is that it makes contact with, and represses, distant target genes via long-distance chromatin: chromatin interactions by spreading a silencing effect in normal cells (Fig. 4-1). We hypothesize that in FSHD the loss of H3K9me3, and therefore of HP1γ and cohesin, results in the loss of these chromatin interactions, thereby causing abnormal derepression of these distant target genes that leads to the dystrophic phenotype. In order to test this hypothesis, we began performing epigenomic analyses on primary FSHD and control myoblasts to investigate the possible global distribution changes of H3K9me3, H3K27me3, HP1γ, and cohesin associated with the disease. If D4Z4 affects gene silencing at other distant regions by HP1γ and cohesin, we expect to see similar loss of H3K9me3/ HP1γ/ cohesin in FSHD at these regions. It is also possible that we may be able to identify an epigenetic signature specific to FSHD.

Materials & Methods

Cells and antibodies

KD3 human myoblasts immortalized by the expression of telomerase, cyclin D1, and mutated cyclin-dependent kinase 4 are grown and differentiated as previously described [161]. The primary myoblasts (listed in table 4-1) were kindly given by Drs. Rabi Tawil and Marina Mora. Primary myoblasts were grown in F10 (Life Technologies) supplemented with 20% FBS, 10 ng/ml bFGF (Biopioneer), 20 ng/ml dexamethasone, and 50U/ml penicillin-streptomycin. Passaging ratio for primary myoblasts is between 1:2 ~ 1:4. Freezing media for primary myoblasts consists of 90% serum and 10% DMSO. Rad21 antibody was previous described [111]. H3K9me3 (309M3A) antibody [207] was a kind gift from Dr. Shohei Koide. All antibodies were listed in Tables. 4-3 and 4-4.

Chaetocin, auranofin and H2O2 treatment

KD3 myoblasts that reached approximately 80% confluency were treated with 0.4 mM chaetocin (Sigma C9492) and harvested after 24 hrs. Cells were also treated with the Thioredoxin Reductase (TrxR) inhibitor auranofin (Sigma A6733) (1μM) for 24 hrs or 0.5 mM H2O2 for 3 hrs. The effects of different treatments on D4Z4 H3K9me3 and *DUX4fl* expression were analyzed by ChIP-PCR and RT-nested set PCR as described in Table 4-2.

Lentiviral shRNA transduction

Non-Target shRNA Control (Sigma: SHC002) and the lentiviral shRNA against human SUV39H1 (TRCN0000157251, Sigma) were transfected into 293T cells along with Lentiviral packaging plasmids, using Lipofectamine 2000 (Life Technologies). Supernatants were

collected 36 and 60 hours posttransfection, passed through a 0.45 μm nitrocellulose filter and applied on KD3 cells in the presence of polybrene (1μg/mL). The next day, cells were selected with puromycin for 48 hours (2μg/ml, Sigma). 48 hours after infection, cells were transferred to 10 or 15 cm dishes and maintained 2 days before harvesting for experimental purposes.

DUX4 nested RT-PCR

Total RNA was extracted using the Qiagen RNeasy Plus kit. Two to five μg of RNA was used for double-stranded cDNA synthesis according to manufacturer's protocol (Life Technologies) with the exception of using enzymes purchased from New England Biolabs. cDNA was purified by Qiagen PCR purification kit and eluted in 40 μl EB buffer. The nested PCR was done using the primer sets (182 - 183 and 1A - 184) previously published [94]. The PCR cycling protocol is as follows: 95 °C 2 min, 95 °C 30 sec, 62 °C 30 sec, 72 °C 1 min 40 sec (repeat 34 times), then 72 °C 10 min. PCR enhancer system (Invitrogen 11495-017) was used for the nested PCR. The PCR products were loaded on agarose gel and the observed bands were cut out for sequencing to confirm the DUX4 identity. At least 3 independent experiments were performed and the representative experiment was shown.

ChIP analysis for manual PCR

The ChIP analysis was performed based on the protocol from the Upstate ChIP assay kit with some modifications. Approximately 6 x 10^6 cells were used per IP. Cells were cross-linked with 1% formaldehyde for 10 min at room temperature. Glycine was added to a final concentration of 0.125 M to stop cross-linking. Cells were washed twice with PBS and collected by scraping. Approximately 2 x 10^7 cells were resuspend in 1ml of Farnham lysis buffer (5 mM PIPES pH 8.0 / 85 mM KCl / 0.5% NP-40/ protease Inhibitors), centrifuged for 5min 2000 rpm at 4°C. In terms of RNase treated ChIP, pellets were re-suspended in 1ml RIPA buffer (PBS / 1%

NP-40 / 0.5% sodium deoxycholate / 0.1% SDS/ protease Inhibitors). RNase A was added to the lysate with final concentration of 500ug/ml. Both the lysate with and without RNase A were incubated at 37°C for 1 hour. SDS was added to the lysate after RNase treatment to a final concentration of 0.5% to facilitate sonication. For other ChIP experiments in the study, sample pellets were resuspended in SDS buffer (50mM Tris-HCl pH 8.0/ 10mM EdTA/ 1% SDS) after Farnham lysis and subjected to sonication. The lysates were sonicated using bioruptor (Diagenode UCD-200) to fragments averaged around 300-500bp. The extracts were diluted with ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl (pH 8.1), 167 mM NaCl) with protease inhibitors and centrifuged for 10 min at 10,000 g at 4 °C. Extracts were precleared for 1 h with protein A-Sepharose (GE Healthcare) supplemented with 1mg/ml BSA. 10% of the extract for each sample was taken as input DNA. Antibodies was added to the extracts for each IP and incubated overnight at 4°C on a rotator platform. Amount of antibodies used in ChIP is documented in table 4-3. The next day, the antibody-bound complexes were immunoprecipitated with protein A-Sepharose beads for 1 h and subsequently washed with low-salt buffer (0.1% SDS/ 1% Triton X-100/ 2 mM EDTA/ 20 mM Tris-HCl pH 8/ 150 mM NaCl), high-salt buffer (0.1% SDS/ 1% Triton X-100/ 2 mM EDTA, 20 mM Tris-HCl pH 8/ 500 mM NaCl), lithium salt buffer (0.25 M LiCl/ 1% Nonidet P-40/ 1% deoxycholate/ 1 mM EDTA/ 10mM Tris-HCl pH 8), and TE (10mM Tris-HCl/ 1 mM EDTA pH 8.0). DNA was eluted off the beads with 200ul of elution buffer (1% SDS/ 0.1 M NaHCO3) on a rotator platform at room temperature for 1 hour. The eluates were collected and reverse crosslinked overnight at 65°C along with the input lysates. Quantitative PCR (q-PCR) was performed using the CFX96 real-time PCR detection system (Bio-Rad) with SYBR premix Ex Taq II (Clontech RR820B). ChIP DNA was purified with Qiagen PCR purification kits. ChIP signal was normalized by subtracting the preimmune IgG ChIP signal, then divided by input DNA signal. Alternatively,

H3K9me3 ChIP was normalized with pan-histone H3 antibody ChIP (Abcam ab1791) as recently described [208]. Primers were listed in table 4-2.

Immunostaining in KD3 cells

Cells were fixed in 4% paraformaldehyde in PBS for 10 min, then blocked/ extracted in SNBP/ 0.1% Gelatin/ 4% heat inactivated horse serum/ 4% goat serum/ 0.1% triton-X for 30 min at 37°C. SNBP buffer is PBS with 0.02% saponin, 0.05% NaN3 and 1% BSA. Primary and secondary antibodies were diluted in SNBP/ 0.05% gelatin/ 1% heat inactivated horse serum/ 1% goat serum. Coverslips were incubated in primary antibodies for 30 min at 37°C followed by three SNBP washes. Then coverslips were incubated in secondary antibodies for 30 min at 37°C followed by three SNBP washes. Then coverslips were stained with DAPI, washed by water and mounted by Antifade (Life Technologies). Antibodies used for staining and specific dilutions were listed in Table 4-3.

Sample preparations for epigenomic analyses

ChIP-sequencing (ChIPseq) was done using the protocol from Dr. Richard Myers available online (http://myers.hudsonalpha.org/documents/Myers%20Lab%20ChIP-seq%20Protocol%20v042211.pdf). Antibodies used for the epigenomic analyses were listed in Table 4-4. RNA-sequencing (RNAseq) was done according to the published paper [135]. Recombinant H3K9me3 antibody (309M3A) [207] was a kind gift from Dr. Shohei Koide. 309M3A antibody needs to be linked to streptavidin beads before used for ChIP. The linkage protocol was described in Dr. Koide's paper [207]. NanoString assays were done according to the published paper [209]. DNase-seq will be done according to the published protocol [210].

Results

Inhibition of H3K9me3 results in *DUX4fl* expression

We previously determined that SUV39H1 histone methyltransferase (HMTase) is responsible for H3K9me3 at D4Z4 in HeLa cells [101]. Based on these results, we treated immortalized human KD3 myoblasts with chaetocin, a SUV39 HMTase inhibitor (Fig. 4-2). We confirmed that H3K9me3 is present at D4Z4 in this cell line, and found that the chaetocin treatment indeed decreased H3K9me3 at D4Z4 (Fig. 4-2A, left). Consistent with this, chaetocin treatment resulted in transcriptional derepression of *DUX4fl* (Fig. 4-2A, right). The identity of the DUX4fl-specific PCR product was confirmed by sequencing. Chaetocin was also shown to inhibit TrxR [211]. Upregulation of *HMOX1* is a marker for TrxR inhibition [212,213]. We found that *HMOX1* is indeed upregulated in cells treated with chaetocin similar to cells treated with auranofin, a TrxR inhibitor, indicating that chaetocin under our condition also inhibits TrxR (Fig. 4-2D). Unlike chaetocin, however, treating cells with auranofin failed to affect H3K9me3 at D4Z4 or *DUX4fl* expression, indicating that the observed chaetocin effect on H3K9me3 and *DUX4fl* expression at D4Z4 is not due to TrxR inhibition (Fig. 4-2B). Since TrxR inhibition leads to oxidative damage induction [211], and oxidative stress has been shown to be associated with FSHD [214], we also treated cells with H_2O_2. This failed to exhibit any effect on *DUX4fl* expression, indicating that the chaetocin treatment's effect on *DUX4* expression is not the result of oxidative stress (Fig. 4-2B, right). Chaetocin was also shown to affect G9a HMTase [215]. Under our treatment condition, we found that H3K9me3 at the *c-Myc* region, which was shown to be mediated by G9a in HeLa cells [216], was also suppressed, suggesting that G9a is also inhibited (Fig. 4-2E). Thus, although G9a depletion had no effect on H3K9me3 at D4Z4 in HeLa cells[101], we cannot exclude the possibility that G9a may contribute to D4Z4 heterochromatin

organization in the context of myoblasts. Nevertheless, we observed H3K9me3 reduction at D4Z4 in SUV39H1 shRNA-transduced cells, strongly supporting the significant role of SUV39H1 in H3K9me3 at D4Z4 in myoblasts (Fig. 4-2C left). Importantly, SUV39H1 depletion led to *DUX4fl* induction (Fig. 4-2C, right). Taken together, the results indicate that the inhibition of H3K9me3 at D4Z4 under these experimental conditions contributes to the de-repression of *DUX4*.

Inhibition of H3K9me3 results in the loss of SMCHD1 binding

SMCHD1 was found to bind to D4Z4 and depletion of SMCHD1 results in transcriptional de-repression of *DUX4fl* [100]. Since SMCHD1 associates with the H3K9me3 domains of the inactive X chromosomes to modulate chromatin compaction [217], we tested whether H3K9me3 also dictates SMCHD1 association at D4Z4. Both chaetocin treatment and SUV39H1 shRNA depletion resulted in the significant loss of SMCHD1 binding to D4Z4 (Fig. 4-3). Neither treatment affected the SMCHD1 protein level (Fig. 4-3 right). Thus, the results indicate that one important downstream effector of H3K9me3 is SMCHD1, and suggest that the loss of H3K9me3 at D4Z4 in FSHD contributes to decreased SMCHD1 binding to D4Z4 leading to *DUX4fl* expression.

FSHD Epigenomic analyses

Previously our lab found that the loss of H3K9me3/ HP1γ/ cohesin to be an epigenetic feature associated with FSHD at the D4Z4 region [101]. Whether this signature loss of heterochromatic feature affects the gene expression or epigenome in a global scale is unknown. In order to characterize the epigenomic landscapes associated with FSHD, we began to examine the H3K9me3 and H3k27me3, as well as HP1γ and cohesin binding genome-wide in primary FSHD and control myoblasts by conducting ChIP sequencing using specific antibodies.

We are using two normal myoblasts, three FSHD1 and three FSHD2 myoblasts, and the cells were listed in table 4-1. These cells have been tested and show at least 70% of cells positive of Desmin, a marker of the myo-lineage [218]. To associate the epigenetic changes with expression patterns, we also performed RNA pol II (RNAPII) ChIP-seq and RNA sequencing (RNA-seq) in the same cell samples. Since microRNAs (miRNAs) are known to play a role in muscle development [219], we examine the miRNA expression patterns in the primary myoblasts by NanoString assay. We are in the data collection phase of the project, as shown in table 4-5. The boxes shaded in grey represent the samples that we completed data collection and are ready to begin the data analysis. The boxes that are left in white represent those samples that are still in the process of data collection. This project is in collaboration with Dr. Ali Mortazavi here at UCI. We hope by these experiments we will be able to find distinct epigenomic patterns associated with FSHD, and identify the distant genomic loci possibly controlled by the D4Z4 heterochromatin.

KD3 as a model to study human myogenesis

Primary myoblasts are difficult to maintain, have limited dividing potential and are prone to alterations of characteristics when grown in culture. Thus it is highly desired to have a model cell line to study human myogenesis in detail. A protocol for human myoblasts immortalization has been created by Shiomi. *et al* [161]. The KD3 cell line has been created with the introduction of mutated Cdk4, cyclinD1 and telomerase [161]. The KD3 line has normal ploidy, high proliferation activity and can be easily induced to differentiate [161]. We aim to establish the KD3 line as a model for human myogenesis. We have verified that the muscle differentiation markers were induced after 3 days of differentiation (Fig. 4-5A). The cells were also fused together and became multinucleated (Fig.4-5A), demonstrating that the cells formed myotubes successfully. RNA-seq analyses demonstrated upregulation of transcripts associated

with muscle differentiation, such as *MYBPH, MYOG,* and *MYH3* (Fig. 4-5B). Genes that are involved in cell cycle progression such as *CENPF* and *BUB1* were downregulated upon differentiation (Fig. 4-5B). *ID1*, which had been shown as a negative regulator of myogensis [220], was indeed downregulated in myotubes (Fig. 4-5B). RNAPII binding to *MYOG* is increased upon differentiation and *ID1* decreased, corresponding to the gene expression changes (Fig. 4-5C). These evidences demonstrated that KD3 differentiation faithfully recaptures the human myogenesis process, and may serve as a good model to study human myogenesis. We are in the process of investigating the epigenetic changes in these cells upon differentiation. We are performing ChIPseq analyses for H3K9me3, H3K27me3, HP1γ, cohesin, CTCF and RNAPII for undifferentiated and differentiated KD3 (Table 4-5). We prepared two sets of samples (cells prepared at different times) in terms of undifferentiated and differentiated cells. RNA-seq and NanoString experiments for miRNA expression were done, and DNase-seq for capturing the open chromatin is in progress. The samples that had been finished with sequencing were colored with grey; the samples that are still in process were left white (Table 4-5). KD3 data will be compared to those of primary normal and FSHD myoblasts in collaboration with Dr. Ali Mortazavi here at UCI. All NanoString and DNase-Seq procedures were/ will be performed by Dr. Weihua Zeng in the Mortazavi lab.

Discussion

In the current study, we demonstrated the significance of the H3K9me3 heterochromatin in *DUX4* gene regulation and SMCHD1 assembly at D4Z4, demonstrating for the first time the biological relevance of the loss of H3K9me3 in FSHD pathogenesis. The epigenomics study for FSHD and KD3 differentiation is undergoing, and the experimental plans will be discussed.

H3K9me3 affects *DUX4fl* expression

Although the loss of H3K9me3 at D4Z4 observed in both FSHD1 and FSHD2 patients was postulated to contribute to disease-specific gene alterations [101], this has not been experimentally proven. We previously demonstrated that SUV39H1 plays a critical role in mediating H3K9me3 at D4Z4 [101]. Using immortalized myoblasts, we showed that *DUX4fl* expression is indeed increased when H3K9me3 is inhibited either by a chemical inhibitor or by shRNA against SUV39H1. This provides important evidence that H3K9me3 plays a role in *DUX4fl* repression, supporting our hypothesis that disruption of D4Z4 H3K9me3 heterochromatin affects gene expression in FSHD. While these immortalized cells retain differentiation capability in vivo and in vitro with expression of appropriate marker genes [161], further analysis is necessary to determine to what extent loss of H3K9me3 affects expression of *DUX4* and other genes in the context of primary patient muscle cells and tissues. Furthermore, other H3K9 HMTases, such as G9a, may contribute to H3K9me3 regulation at D4Z4 in a myogenic context, though SUV39H1 depletion alone was shown to have a significant effect on *DUX4fl* expression in our study.

H3K9me3 affects SMCHD1 binding at D4Z4

SMCHD1, an epigenetic gene silencer mutated in >80% of FSHD2 patients and severe cases of FSHD1, binds to D4Z4 and its depletion results in expression of *DUX4fl* [100,221]. How SMCHD1 is recruited to D4Z4 chromatin, however, was unclear. We found that suppression of H3K9me3 affects SMCHD1 association with D4Z4, indicating that SMCHD1 is a component of D4Z4 H3K9me3 heterochromatin. This raises the possibility that SMCHD1 binding to D4Z4 is already impacted by the loss of H3K9me3 in FSHD even in those FSHD1 and FSHD2 cases in which the SMCHD1 gene is intact, and this may be further worsened in the cases with the haploinsufficiency mutations of SMCHD1. It will be important to test whether SMCHD1 binding to D4Z4 is indeed reduced in primary patient myoblasts. Because SMCHD1 was found to mediate DNA methylation [98,99,222], it is possible that SMCHD1 contributes to DNA hypermethylation observed at D4Z4, which is lost in FSHD [188]. Thus, we propose that H3K9me3 contributes to the recruitment of SMCHD1, which in turn mediates DNA methylation at D4Z4 (Fig. 4-4). Although SMCHD1 associates with the inactive X chromosome through interaction with *XIST* RNA in the chromatin domains enriched for H3K27me3 (though H3K27me3 itself is not required), it also associates with the H3K9me3 domains via the HP1 binding protein HBiX1 [217]. Thus, it would be interesting to further examine the relationship between SMCHD1 and HP1γ/cohesin that assemble at D4Z4 in an H3K9me3-dependent manner [101]. Nevertheless, our results suggest that such diminished binding of downstream effectors of H3K9me3 may be a key event leading to pathogenic alteration of *DUX4* gene expression (Fig. 4-4) and possibly of additional target genes in FSHD.

Hyper-variability in the characteristics of primary myoblasts

While growing and expanding the primary myoblasts for high- throughput sequencing, we noticed a great deal of variability among the samples. Each sample has different dividing

speed and morphological characteristics. Some samples have very low percentage (<30%) of cells that are desmin- positive, indicating that the samples were contaminated with non-myogenic cell populations, or the samples changed characteristics before we tested them. The cells also change properties throughout time. Some samples started with a large population of desmin-positive cells but the population diminished after a few passages; while other samples retained the large desmin-positive population even after 10 rounds of passaging. Proliferation rate and cell morphology may or may not change throughout passages. The percentage of desmin-positive cells, the proliferation speed and the morphological changes do not seem to correlate with each other. We strived to do the experiments using the samples that had high proliferation rate, strong desmin staining and homogenous morphology, but there were always unpredictability even the samples were cultured the same way. In order to get the data of high quality from these cells, it is best to monitor the properties of these cells regularly (eg. check the expression of markers genes) and obtain enough cells for experiment with the lowest passage number as possible.

Epigenomic analyses of primary FSHD and normal myoblasts and immortalized KD3 myoblasts

With the genomic data in hand, we are in the process of analyzing whether there is systematic difference between control, FSHD1 and FSHD2 in terms of both epigenetic marks and gene expression patterns. Regions with the concomitant loss of H3K9me3, HP1γ, and cohesin will be identified and tested for possible chromatin interactions with D4Z4 region using both chromatin conformation capture and fluorescent *in situ* hybridization. SUV39H1, cohesin and HP1γ will be depleted in normal myoblasts (or KD3) to assay the effect on chromatin interactions as well as the effect on gene expression of the interacting loci. The KD3 ChIPseq and DNase-seq data are being compared to RNA-seq and NanoString data to determine

whether there are epigenetic changes associated with gene expression change when myoblasts undergo differentiation. By correlating epigenetic and gene expression changes, we may be able to determine the nuclear reprogramming critical for muscle cell differentiation and development of muscular dystrophy. Another goal is to generate immortalized myoblasts for FHSD using the protocol that generated the KD3 cells [161]. If we have immortalized myoblasts for both normal and FSHD, it would be easy to induce differentiation to compare the normal and FSHD cells differentiation process. By this method we may be able to find important molecular nuances that we would miss with primary cells that are hyper-variable.

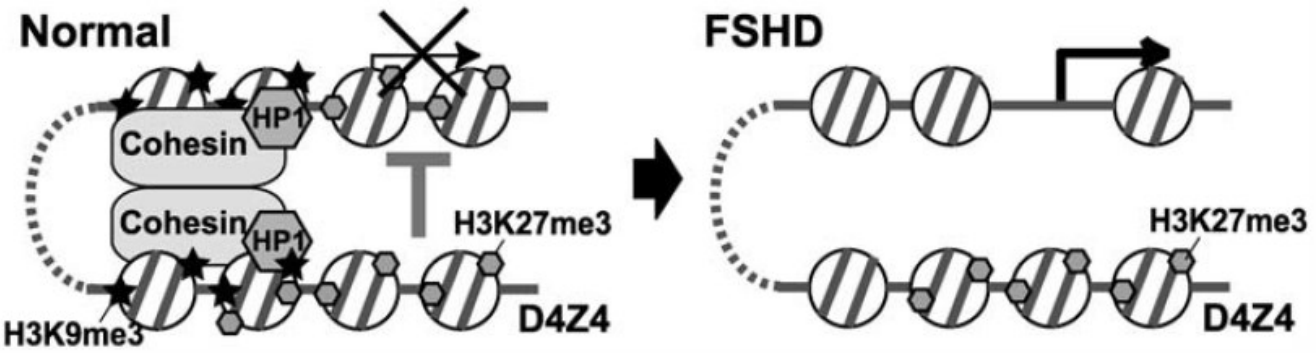


Figure 4-1. Model of how H3K9me3/ HP1γ/ cohesin loss can regulate epigenetic status and expression at distant locations

Since we observed a specific loss of H3K9me3/ HP1γ/ cohesin in FSHD [101], this is a possible model for the spreading of the epigenetic change at D4Z4 to other genomic regions in FSHD. HP1γ and cohesin may contribute to the physical interactions of the heterochromatic D4Z4 region with other genomic regions leading to the spreading of the silencing effect to putative target genes in normal cells. In FSHD, the loss of H3K9me3 (but not H3K27me3), HP1γ, and cohesin from D4Z4 results in loss of chromatin interaction and derepression of these genes leading to muscular dystrophy.

(The model had been published in [101])

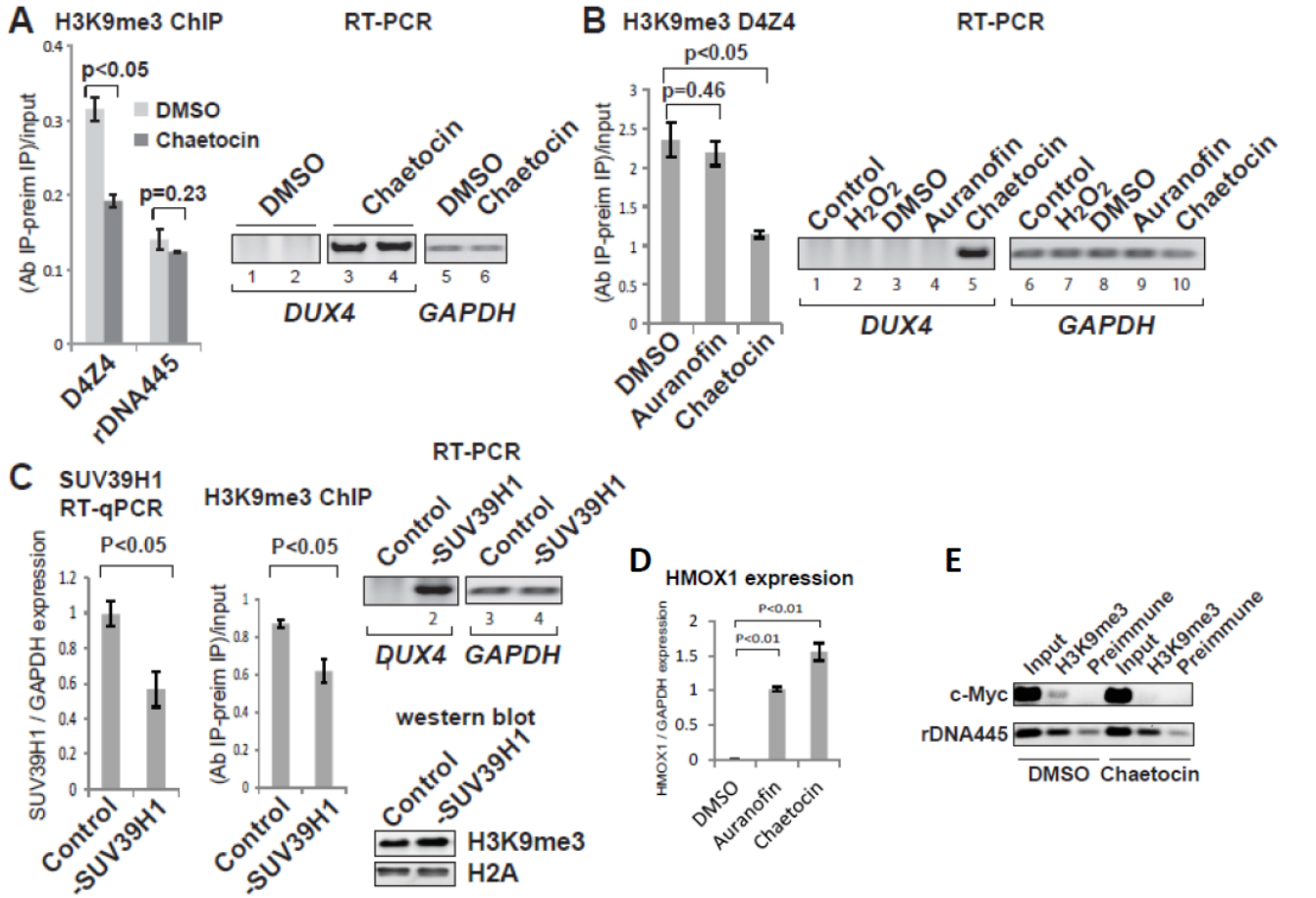


Figure 4-2. Inhibition of H3K9me3 results in *DUX4fl* upregulation.

(A) The effect of chaetocin on H3K9me3 at D4Z4 and DUX4 expression in KD3 myoblasts. Left: H3K9me3 ChIP-PCR analysis of D4Z4 and rDNA regions in DMSO and chaetocin-treated KD3 cells. Right: RT-PCR analysis of DUX4fl expression. Nested set RT-PCR analysis of DUX4fl expression as described (Snider et al., 2010) in KD3 human myoblasts treated with DMSO only or chaetocin as indicated at the top. The PCR products were sequenced to confirm their identity. GAPDH RT-PCR serves as a control. **(B)** Comparison of the effects of chaetocin and auranofin treatments on D4Z4 H3K9me3 and DUX4 expression. Left: H3K9me3 ChIP-PCR analysis of the D4Z4 region in auranofin- and chaetocin-treated cells. Right: DUX4fl and GAPDH RT-PCR analyses as in (A). Untreated control, H2O2, DMSO-treated, auranofin, or cheatocin-treated cells werecompared. **(C)** The effect of SUV39H1 depletion on D4Z4 H3K9me3 and DUX4 expression. Lentiviral shRNA against SUV39H1 was used for depletion. Left: RT-qPCR analysis of SUV39H1 depletion in control and SUV39H1 shRNA-treated cells. Middle: ChIP-PCR analysis of H3K9me3 at D4Z4 in control and SUV39H1 shRNA-treated cells. Right: Top, DUX4fl and GAPDH RT-PCR as in (A); Bottom, western blot analysis of control and SUV39H1-depleted cells with antibodies specific for H3K9me3 and H2A. **(D)** HMOX1 is induced in response to either chaetocin or auranofin, suggesting that TrxR is also inhibited by chaetocin treatment. **(E)** H3K9me3 at the *c-Myc* region, which is primarily mediated by G9a in HeLa cells, is inhibited in chaetocin-treated KD3 myoblast cells, suggesting the possible cross-inhibition of G9a in addition to SUV39H1.

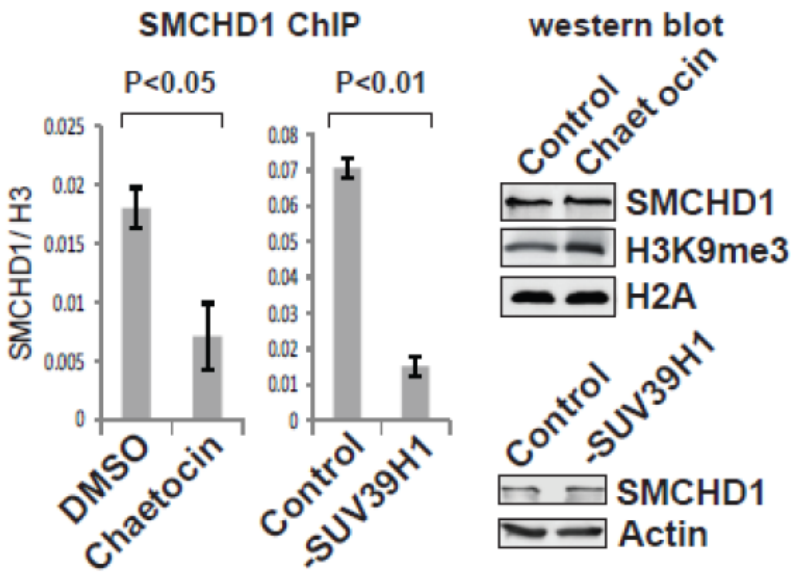


Figure 4-3. Inhibition of H3K9me3 results in the loss of SMCHD1 binding.

Left: ChIP-PCR analysis of SMCHD1 binding at D4Z4 in chaetocin and SUV39H1 shRNA-treated cells. Y-axis: (SMCHD1 ChIP – preimmune IgG)/input, which was further normalized by histone H3 ChIP. Right: western blot analysis of cheatocin- (top) or SUV39H1 shRNA- (bottom) treated cells compared to DMSO (control) or control shRNA-treated cells, respectively, using antibodies indicated. Histone H2A (top) and actin (bottom) served as a loading control.

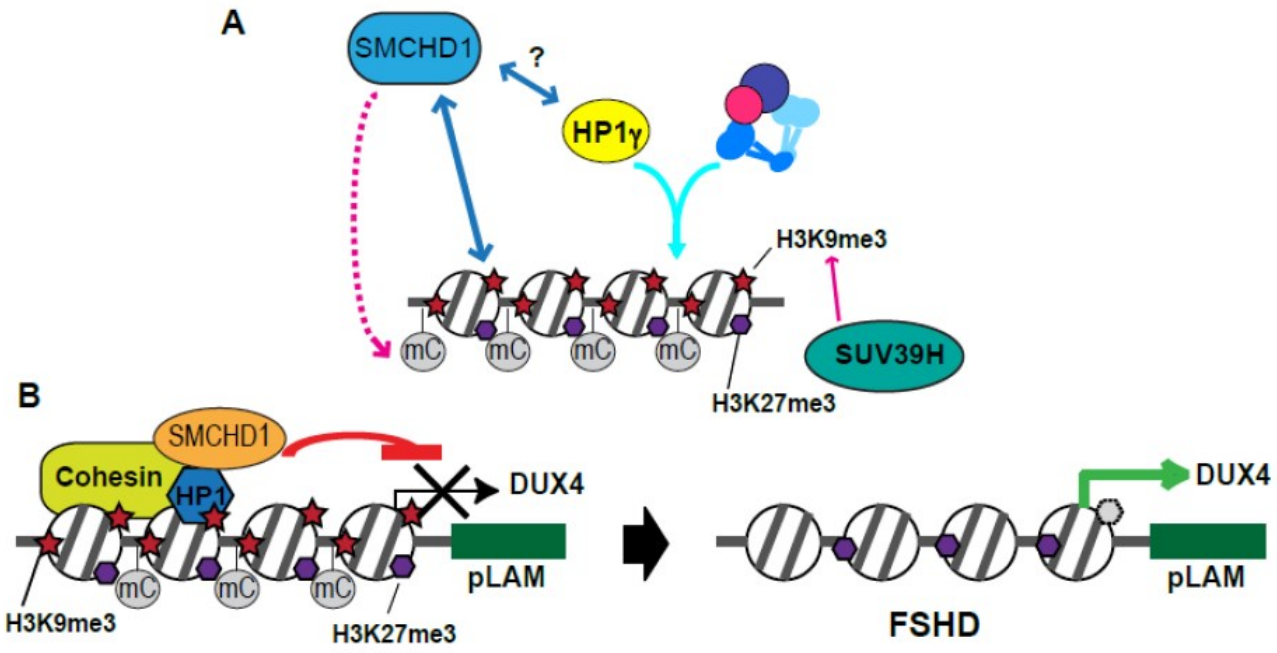


Figure 4-4. Schematic models of H3K9me3 heterochromatin at D4Z4.

(A) The heterochromatin domain of D4Z4 is marked by DNA hypermethylation, H3K9me3 and K27me3. HP1γ and cohesin are recruited to this region in a mutually dependent manner, which requires H3K9me3 mediated primarily by SUV39H. SMCHD1 is also recruited to this domain in an H3K9me3-dependnent manner, and may contribute to DNA methylation. Possible interaction between SMCHD1 and HP1γ is shown with a question mark. **(B)** The H3K9me3 heterochromatin assembled at D4Z4, which is specifically lost in FSHD, contributing to DUX4fl expression.

(The model shown here has been published in [114])

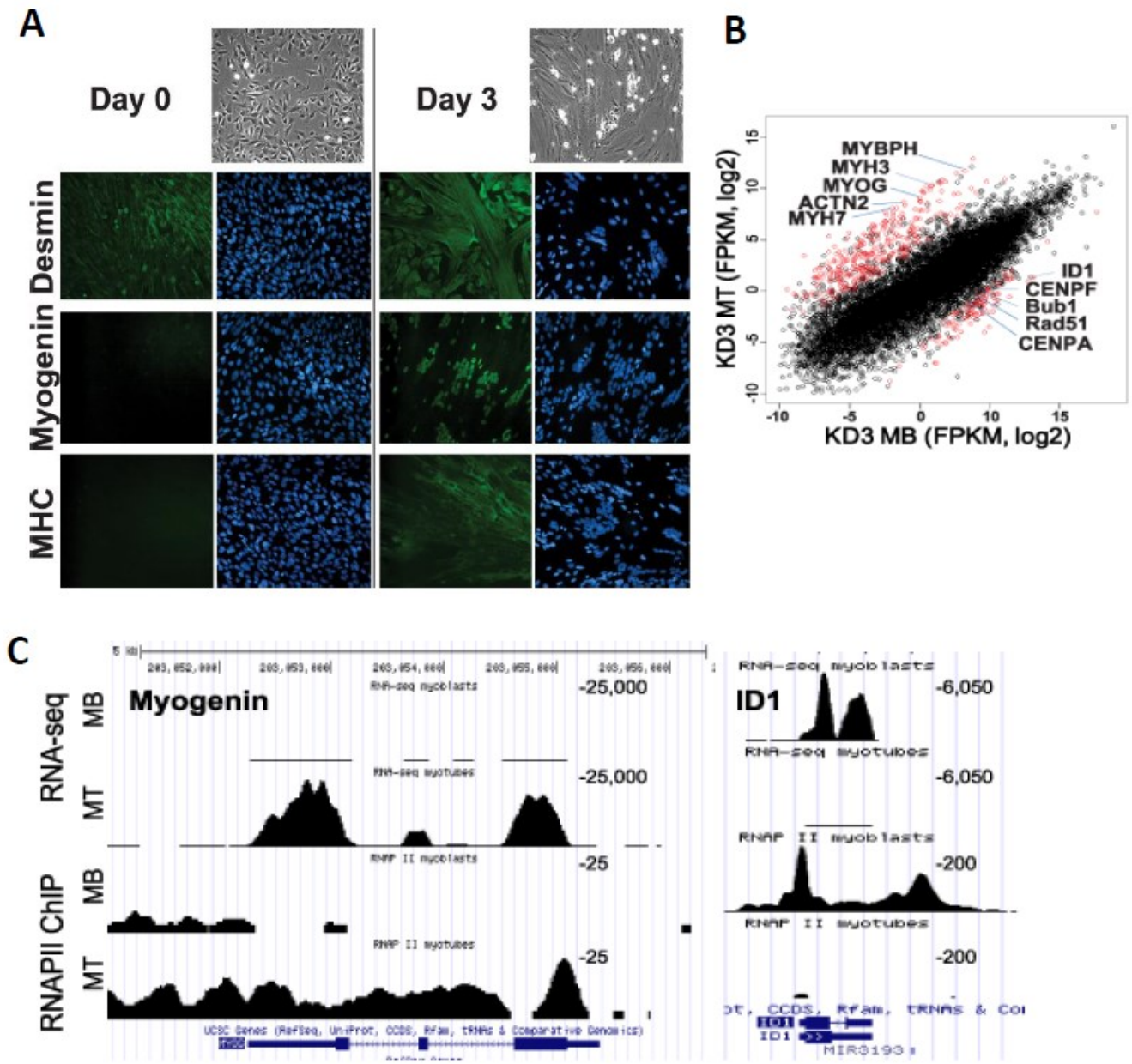


Figure 4-5. KD3 differentiation and gene expression

(A) Immunostaining in undifferentiated KD3 (myoblasts) (Day 0) and after induction of differentiation for 3 days (myotubes) (Day 3). **(B)** Comparison of RNA-seq data from MB (myoblasts) and MT (myotubes). Differentially expressed genes were highlighted in red. **(C)** UCSC genome browser examples of RNA-seq and RNAPII (RNA polymerase II) at two loci, *MYOG* and *ID1*.

(The KD3 stainings were done by Dr. Alexander Ball Jr.)

Table 4-1 Myoblasts used for epigenomic analyses, and the percentage of cells that are desmin staining positive.

Cells	Cell type	Desmin %	Comments
CM135	Normal	95%	
01-189-0022469	Normal	95%	
FM191	FSHD1	80%	
10868	FSHD1	70%	
9155	FSHD1	95%	Female, age 55, FSHD 29kb
01-230-0012334	FSHD2	95%	Male, age 59
01-200-0012062	FSHD2	95%	
01-163-001 NMD1881	FSHD2	95%	SMCHD1 mutation

Table 4-2. The list of PCR primers

Primers for RT-PCR		Reference
DUX4-182	CACTCCCCTGCGGCCTGCTGCTGGATGA	
DUX4-183	CCAGGAGATGTAACTCTAATCCAGGTTTGC	[94]
DUX4-1A	GAGCTCCTGGCGAGCCCGGAGTTTCTG	
DUX4-184	GTAACTCTAATCCAGGTTTGCCTAGACAGC	[94]
GAPDH F	TCGACAGTCAGCCGCATCTTCTT	Self made
GAPFH R	GCGCCCAATACGACCAAATCC	
SUV39H1 F	GTCGTTAGCCGTGGGGAAAG	Self made
SUV39H1 R	GATACCGAGGGCAGGGCAGG	
HMOX1-F	AACTTTCAGAAGGGCCAGGT	Self made
HMOX1-R	TGTTGCGCTCAATCTCCTC	
Primers for ChIP		**Reference**
rDNA445 F	CCCGTGAGGAGTGATTTCTAA	
rDNA445 R	CGTGGGCCGGGCTGGTCTC	[101]
D4Z4 F	CCGCGTCCGTCCGTGAAA	
D4Z4 R	TCCGTCGCCGTCCTCGTC	[101]
C-Myc F	GAAGGTATCCAATCCAGATAGCTGTGC	
C-Myc R	GAGCGTGGGATGTTAGTGTAGATAGGG	[216]

Table 4-3. Antibodies and their applications in the study

Antibody	Vendor	Immunostaining	Western blot	ChIP
Desmin	Dako M0760	1:100	-	-
Myogenin	DSHB F5D	1:20	-	-
Myosin heavy chain	DSHB MF20	1:20		
SMCHD1	Abcam ab31865	-	1:1000	10 µl
H3k9me3	Millipore 17-625	-	1:1000	5 µl
Histone H3	Abcam ab1791	-	-	3.5 µl
Histone H2A	Abcam ab18255	-	1:1000	-
Actin	Sigma A4700	-	1:1000	-

Table 4-4. Antibodies used in the epigenomic analyses

Antibody	Vendor	ChIP
Rad21	(In house)	5 µg
CTCF	Millipore 07-729	10 µl
H3K9me3 (309M3A)	(Given by Dr. Shohei Koide, Univ. Chicago)	0.5 µg
H3K27me3	Millipore 17-622	5 µl
RNAPII	Santa Cruz SC-47701X	2.5 µl
HP1γ	Millipore 17-646	5 µl

Table 4-5. FSHD epigenomic analyses

	Cells	RNA seq	Nano string	WCE	Rad21	H3k9me3	H3k27me3	Pol II	HP1g
Normal	CM135	180M		7M/4M	19M/16M /105K	85M/57M/54K	51M/39M/ 47K		23M/17 M/25K
	2469	205M		10M/3M	27M/19M /34K	73M/49M/63K 73M/50M/57K	100M/80M/ 44K		
FSHD1	9155	78M		7M/5M	15M/11M /29K	67M/48M/57K	37M/29M/?		
	10868	65M		15M/10M		58M/42M/52K			
	FM191	78M		10M/6M		55M/34M/50K	29M/22M/ 45K		
FSHD2	2334	63M		10M/4M	18M/12M /56K	69M/49M/50K	35M/28M/ 35K	20M/16 M/53K	
	2062	58M		15M/9M	21M/10M /38K	73M/45M/55K	37M/28M/ 55K		
	NMD1881	70M		15M/11M	34M/27M /153K	39M/28M/?		17M/13 M/47K	

(Total Reads/Mapped Reads/Peaks)

Table 4-6. Undifferentiated and differentiated KD3 epigenomic analyses

	Cells	RNAseq	Dnase seq	Nano string	WCE	Rad21	CTCF	H3k9me3	H3k27me3	Pol II	HP1g
Rep1	Undiff	74M/62M			8M/5M	12M/3M/27K	12M/4M/40K	42M/21M/50K	28M/17M/32K	9M/5M/28K	
	Diff	102M/84M			16M/11M	14M/8M/28K	14M/9M/28K	58M/32M/51K	26M/16M/37K	11M/7M/9K	
Rep2	Undiff	78M/51M			11M/4M			53M/15M/27K		13M/4M/19K	
	Diff	113M/75M			12M/5M			51M/23M/30K			

Chapter Five

Conclusions

The goal of my thesis work lies in understanding how impaired NIPBL and cohesin functions leads to human diseases. I studied two diseases models- the Cornelia de Lange syndrome (CdLS), and the Facioscapulohumeral muscular dystrophy (FSHD). I obtained informative results which augments our current knowledge of the relationship between NIPBL and the cohesin complex. I established the functional significance of D4Z4 heterochromatin in FSHD. I also discovered a novel function of NIPBL, which provided new directions for future research endeavors and therapeutic opportunities.

The effect of *Nipbl* haploinsufficiency on cohesin function in a CdLS mouse model

The first aim of my thesis was to elucidate how *Nipbl* haploinsufficiency affects cohesin's gene regulatory role. We found that genome-wide cohesin binding is lessened by the reduction of Nipbl, including both unique gene regions and repeat sequences. Multiple genes in the adipogenesis pathway are direct targets of cohesin, and their expression is affected by *Nipbl* haploinsufficiency. Since the CdLS mice are observed to have less body fat, we providing a link between *Nipbl* haploinsufficiency and the actual phenotype. We demonstrated that one of the adipogenesis genes, *Cebpβ*, has reduced chromatin interaction between promoter and a putative enhancer in the *Nipbl* mutant. This shows that a partial reduction of NIPBL can impact chromatin interaction. Thus it is likely that there is a global decrease of chromatin interactions that causes abnormal gene expression in CdLS. Our data provided a plausible mechanism of how reduced NIPBL caused the actual disease phenotype.

A novel role of NIPBL in the nucleolus and its relevance to CdLS

The second part of my thesis involved characterizing NIPBL's function in nucleolus. I found that NIPBL binds to both rRNA and rDNA and positively affects rRNA transcription. This is the first time that NIPBL has been shown to physically bind RNA. It is known that cohesin can

bind to eRNA (enhancer RNA) of the enhancers adjacent to estrogen response genes [223]. Ligand-activated estrogen receptor (ER) upregulates transcription of eRNAs, which act in cis to promote upregulation of nearby ER target genes. The eRNAs bind to cohesin and increase cohesin recruitment to the enhancer regions in response to the ER ligand estradiol, and stimulate the enhancer–promoter interactions in MCF7 breast cancer cells [223]. Nipped-B (NIPBL homolog in *drosophila*) has been shown to promote enhancer-promoter interaction as well [143]. The rDNA locus has a defined enhancer upstream of the rDNA coding region [224]. Thus one can imagine a possible scenario at the rDNA locus, which rRNA facilitates NIPBL binding to rDNA and possibly bridges the interaction between rDNA locus and enhancer to further stimulate rRNA production. To test this, we can deplete NIPBL or treat cells with RNase and perform 3C experiments. The fact that NIPBL interacts with rRNA makes us wonder whether NIPBL can interact with other RNA species as well. Therefore we propose to perform iCLIP-seq (individual nucleotide resolution CLIP followed by high throughput sequencing) [162,163] to identify other possible RNA species that may bind to NIPBL. Furthermore, we found several putative RNA binding domains in NIPBL (Fig. 5-1). Using various deletion mutants, it is possible to substantiate the RNA binding activity of NIPBL by mapping and characterizing the RNA binding domain(s) of NIPBL. We also demonstrated that NIPBL participates in the nucleolar stress response pathway. It would be informative to identify the NIPBL interacting proteins with and without stress to identify the pathways that NIPBL is involved with, and the possible functions NIPBL plays in response to stress. NIPBL's novel role in the nucleolus implies that some of the CdLS phenotypes might be the result of impaired ribosome biogenesis, suggesting that CdLS may be a type of ribosomopahty. The findings also suggest that ribosome biogenesis can be targeted for CdLS disease treatment.

Cohesin/NIPBL at heterochromatin and FSHD as a possible cohesinopathy

In contrast to the global decrease of cohesin binding in CdLS, we found that cohesin binding is specifically impaired at the D4Z4 heterochromatin in FSHD. Furthermore, we previously obtained evidence that cohesin plays an active role in D4Z4 heterochromatin organization [101]. Thus, the third aim of my thesis was to understand the function of D4Z4 heterochromatin in FSHD to explore possible heterochromatin-associated function of cohesin. I found the D4Z4 heterochromatin is important in repressing *DUX4*, whose expression causes the muscular dystrophic phenotypes. Although I failed to demonstrate the role of cohesin in *DUX4* regulation, I found that another SMC homolog SMCHD1 is recruited to D4Z4 in an H3K9me3-dependent manner to regulate *DUX4* expression. A recent study indicated that SMCHD1 is recruited to H3K9me3 domains through interaction with HBiX1, an HP1 binding protein, and contributes to the compaction of the inactive X chromosome [217]. This raises the intriguing possibility that in normal cells SMCHD1 is recruited to D4Z4 by the H3K9me3/HP1γ/cohesin heterochromatin, the loss of which in FSHD results in decreased binding of SMCHD1 and subsequent derepression of the DUX4 gene. To test this hypothesis, we should deplete HP1γ and cohesin and determine how this affects SMCHD1 recruitment to D4Z4. SMCHD1 binding to D4Z4 is decreased in FSHD2 since it is mutated in FSHD2 [100]. However our data showed that H3K9me3 recruits SMCHD1, and that H3K9me3 is lost in both FSHD1 and FSHD2, which would suggest that SMCHD1 binding to D4Z4 in FSHD1 cells is also impaired. Thus we plan to perform SMCHD1 ChIP in FSHD1 and normal myoblasts at D4Z4 to see if indeed this is the case. To clarify the relationship of the protein factors present at D4Z4 regions would lead to a fuller understanding of how cohesin functions at heterochromatin regions.

Taken together, my thesis research provided mechanisms of how defective cohesin/ NIPBL pathways can lead to abnormal developmental phenotypes in genetic diseases. Cohesin/

NIPBL's functional significance in different contexts such as gene activation, gene silencing and at distinct nuclear compartments have been explored. This study contributes to a deeper understanding of cohesin and NIPBL functions in organizing chromatin structure.

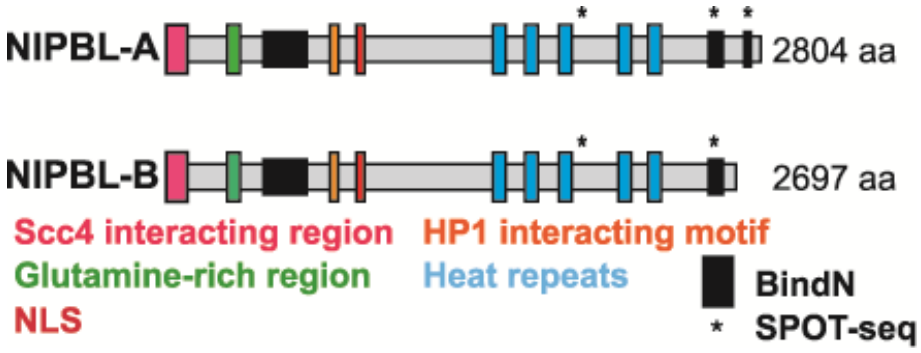


Figure. 5-1 Schematic diagram of NIPBL isoforms A and B
Putative RNA binding domains identified by BindN [225] and SPOT-seq [226] are also indicated. (The figure is modified from [227]).

References

1. Hirano T (2006) At the heart of the chromosome: SMC proteins in action. Nat Rev Mol Cell Biol 7: 311-322.

2. Nasmyth K, Haering CH (2005) The structure and function of SMC and kleisin complexes. Annu Rev Biochem 74: 595-648.

3. Losada A, Hirano T (2005) Dynamic molecular linkers of the genome: the first decade of SMC proteins. Genes Dev 19: 1269-1287.

4. Guacci V, Koshland D, Strunnikov A (1997) A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*. Cell 91: 47-57.

5. Losada A, Hirano M, Hirano T (1998) Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion. Genes Dev 12: 1986-1997.

6. Michaelis C, Ciosk R, Nasmyth K (1997) Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids. Cell 91: 35-45.

7. Gruber S, Haering CH, Nasmyth K (2003) Chromosomal cohesin forms a ring. Cell 112: 765-777.

8. Ivanov D, Nasmyth K (2005) A topological interaction between cohesin rings and a circular minichromosome. Cell 122: 849-860.

9. Ivanov D, Nasmyth K (2007) A physical assay for sister chromatid cohesion in vitro. Mol Cell 27: 300-310.

10. Haering CH, Farcas AM, Arumugam P, Metson J, Nasmyth K (2008) The cohesin ring concatenates sister DNA molecules. Nature 454: 297-301.

11. Chan KL, Roig MB, Hu B, Beckouët F, Metson J, et al. (2012) Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. Cell 150: 961-974.

12. Nasmyth K (2011) Cohesin: a catenase with separate entry and exit gates? Nat Cell Biol 13: 1170-1177.

13. Losada A, Yokochi T, Kobayashi R, Hirano T (2000) Identification and characterization of SA/Scc3p subunits in the *Xenopus* and human cohesin complexes. J Cell Biol 150: 405-416.

14. Sumara I, Vorlaufer E, Gieffers C, Peters BH, Peters J-M (2000) Characterization of

vertebrate cohesin complexes and their regulation in prophase. J Cell Biol 151: 749-761.

15. Remeseiro S, Cuadrado A, Carretero M, Martínez P, Drosopoulos WC, et al. (2012) Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. EMBO J 31: 2076-2089.

16. Remeseiro S, Cuadrado A, Gómez-López G, Pisano DG, Losada A (2012) A unique role of cohesin-SA1 in gene regulation and development. EMBO J 31: 2090-2102.

17. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, et al. (2011) Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science 333: 1039-1043.

18. Seitan VC, Banks P, Laval S, Majid NA, Dorsett D, et al. (2006) Metazoan Scc4 Homologs Link Sister Chromatid Cohesion to Cell and Axon Migration Guidance. PLoS Biol 4: Epub.

19. Watrin E, Schleiffer A, Tanaka K, Eisenhaber F, Nasmyth K, et al. (2006) Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. Current biology : CB 16: 863-874.

20. Borges V, Lehane C, Lopez-Serra L, Flynn H, Skehel M, et al. (2010) Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. Mol Cell 39: 677-688.

21. Heidinger-Pauli JM, Unal E, Koshland D (2009) Distinct targets of the Eco1 acetyltransferase modulate cohesion in S phase and in response to DNA damage. Mol Cell 34: 311-321.

22. Ünal E, Heidinger-Pauli JM, Kim W, Guacci V, Onn I, et al. (2008) A molecular determinant for the establishment of sister chromatid cohesion. Science 321: 566-569.

23. Zhang J, Shi X, Li Y, Kim BJ, Jia J, et al. (2008) Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Molecular cell 31: 143-151.

24. Higashi TL, Ikeda M, Tanaka H, Nakagawa T, Bando M, et al. (2012) The prereplication complex recruits XEco2 to chromatin to promote cohesin acetylation in Xenopus egg extracts. Curr Biol 22: 977-988.

25. Beckouët F, Hu B, Roig MB, Sutani T, Komata M, et al. (2010) An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. Mol Cell 39: 689-699.

26. Deardorff MA, Bando M, Nakato R, Watrin E, Itoh T, et al. (2012) HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature 489: 313-317.

27. Xiong B, Lu S, Gerton JL (2010) Hos1 is a lysine deacetylase for the Smc3 subunit of cohesin. Curr Biol 20: 1660-1665.

28. Buheitel J, Stemmann O (2013) Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3–Scc1 gate. EMBO J 32: 666-676.

29. Eichinger CS, Kurze A, Oliveira RA, Nasmyth K (2013) Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chro- mosomes during interphase and mitosis. EMBO J 32: 656-665.

30. Losada A, Yokochi T, Hirano T (2005) Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and Xenopus egg extracts. Journal of cell science 118: 2133-2141.

31. Schmitz J, Watrin E, Lenart P, Mechtler K, Peters JM (2007) Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. Current biology : CB 17: 630-636.

32. Waizenegger IC, Hauf S, Meinke A, Peters J-M (2000) Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell 103: 399-410.

33. Hauf S, Roitinger E, Koch B, Dittrich CM, Mechtler K, et al. (2005) Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. PLoS biology 3: e69.

34. Hauf S, Waizenegger IC, Peters JM (2001) Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. Science 293: 1320-1323.

35. Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, et al. (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature 460: 410-413.

36. Hou C, Dale R, Dean A (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. Proceedings of the National Academy of Sciences of the United States of America 107: 3651-3656.

37. Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, et al. (2009) Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet 5: e1000739.

38. Chien R, Zeng W, Kawauchi S, Bender MA, Santos R, et al. (2011) Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. The Journal of biological chemistry 286: 17870-17878.

39. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, et al. (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature.

40. Mishiro T, Ishihara K, Hino S, Tsutsumi S, Aburatani H, et al. (2009) Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. EMBO J 28: 1234-1245.

41. Guo Y, Monahan K, Wu H, Gertz J, Varley KE, et al. (2012) CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. Proc Natl Acad Sci 109: 21081-21086.

42. Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD, et al. (2012) Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. Genome Res 22: 2163-2175.

43. Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, et al. (2010) A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res 20: 578-588.

44. Prickett AR, Barkas N, McCole RB, Hughes S, Amante SM, et al. (2013) Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. Genome Res 23: 1624-1635.

45. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, et al. (2003) The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet 35: 190-194.

46. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P (2002) Long-range chromatin regulatory interactions in vivo. Nat Genet 32: 623-626.

47. Demare LE, Leng J, Cotney J, Reilly SK, Yin J, et al. (2013) The genomic landscape of cohesin-associated chromatin interactions. Genome Res [Epub ahead of print].

48. Gillespie PJ, Hirano T (2004) Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. Curr Biol 14: 1598-1603.

49. Takahashi TS, Yiu P, Chou MF, Gygi S, Walter JC (2004) Recruitment of Xenopus Scc2 and cohesin to chromatin requires the pre-replication complex. Nature cell biology 6: 991-996.

50. Arumugam P, Gruber S, Tanaka K, Haering CH, Mechtler K, et al. (2003) ATP hydrolysis is required for cohesin's association with chromosomes. Curr Biol 13: 1941-1953.

51. Weitzer S, Lehane C, Uhlmann F (2003) A model for ATP hydrolysis-dependent binding of cohesin to DNA. Curr Biol 13: 1930-1940.

52. Lengronne A, McIntyre J, Katou Y, Kanoh Y, Hopfner KP, et al. (2006) Establishment of sister chromatid cohesion at the S. cerevisiae replication fork. Mol Cell 23: 787-799.

53. Bermudez VP, Farina A, Higashi TL, Du F, Tappin I, et al. (2012) In vitro loading of human cohesin on DNA by the human Scc2-Scc4 loader complex. Proceedings of the National Academy of Sciences of the United States of America 109: 9366-9371.

54. Lindroos HB, Strom L, Itoh T, Katou Y, Shirahige K, et al. (2006) Chromosomal association of the Smc5/6 complex reveals that it functions in differently regulated pathways. Mol Cell 22: 755-767.

55. D'Ambrosio C, Schmidt CK, Katou Y, Kelly G, Itoh T, et al. (2008) Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. Genes Dev 22: 2215-2227.

56. Bose T, Gerton JL (2010) Cohesinopathies, gene expression, and chromatin organization. The Journal of cell biology 189: 201-210.

57. Liu J, Krantz ID (2008) Cohesin and human disease. Annu Rev Genomics Hum Genet 9: 303-320.

58. Vega H, Waisfisz Q, Gordillo M, Sakai N, Yanagihara I, et al. (2005) Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet 37: 468-470.

59. Dorsett D (2007) Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. Chromosoma 116: 1-13.

60. Tomkins D, Hunter A, Roberts M (1979) Cytogenetic findings in Roberts-SC phocomelia syndrome(s). Am J Med Genet 4: 17-26.

61. Mönnich M, Kuriger Z, Print CG, Horsfield JA (2011) A zebrafish model of Roberts syndrome reveals that Esco2 depletion interferes with development by disrupting the cell cycle. PLos ONE 6: e20051.

62. Whelan G, Kreidl E, Wutz G, Egner A, Peters JM, et al. (2012) Cohesin acetyltransferase Esco2 is a cell viability factor and is required for cohesion in pericentric heterochromatin. EMBO J 31: 71-82.

63. DeScipio C, Kaur M, Yaeger D, Innis JW, Spinner NB, et al. (2005) Chromosome rearrangements in cornelia de Lange syndrome (CdLS): report of a der(3)t(3;12)(p25.3;p13.3) in two half sibs with features of CdLS and review of reported CdLS cases with chromosome rearrangements. American journal of medical genetics Part A 137A: 276-282.

64. Liu J, Krantz ID (2009) Cornelia de Lange syndrome, cohesin, and beyond. Clin Genet 76: 303-314.

65. Krantz ID, McCallum J, DeScipio C, Kaur M, Gillis LA, et al. (2004) Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet 36: 631-635.

66. Tonkin ET, Wang TJ, Lisgo S, Bamshad MJ, Strachan T (2004) NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nature genetics 36: 636-641.

67. Gillis LA, McCallum J, Kaur M, DeScipio C, Yaeger D, et al. (2004) NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am J Hum Genet 75: 610-623.

68. Musio A, Selicorni A, Focarelli ML, Gervasini C, Milani D, et al. (2006) X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet 38: 528-530.

69. Deardorff MA, Kaur M, Yaeger D, Rampuria A, Korolev S, et al. (2007) Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. Am J Hum Genet 80: 485-494.

70. Deardorff MA, Wilde JJ, Albrecht M, Dickinson E, Tennstedt S, et al. (2012) RAD21 mutations cause a human cohesinopathy. American journal of human genetics 90: 1014-1027.

71. Castronovo P, Delahaye-Duriez A, Gervasini C, Azzollini J, Minier F, et al. (2010) Somatic mosaicism in Cornelia de Lange syndrome: a further contributor to the wide clinical expressivity? Clin Genet 78: 560-564.

72. Zhang B, Jain S, Song H, Fu M, Heuckeroth RO, et al. (2007) Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. Development 134: 3191-3201.

73. Zhang B, Chang J, Fu M, Huang J, Kashyap R, et al. (2009) Dosage effects of cohesin regulatory factor PDS5 on mammalian development: implications for cohesinopathies. PLos ONE 4: e5232.

74. Dorsett D, Krantz ID (2009) On the molecular etiology of Cornelia de Lange syndrome. Ann N Y Acad Sci 1151: 22-37.

75. Selicorni A, Russo S, Gervasini C, Castronovo P, Milani D, et al. (2007) Clinical score of 62 Italian patients with Cornelia de Lange syndrome and correlations with the presence and type of NIPBL mutation. Clin Genet 72: 98-108.

76. Kawauchi S, Calof AL, Santos R, Lopez-Burks ME, Young CM, et al. (2009) Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. PLoS Genet 5: e1000650.

77. Borck G, Zarhrate M, Cluzeau C, Bal E, Bonnefont JP, et al. (2006) Father-to-daughter transmission of Cornelia de Lange syndrome caused by a mutation in the 5' untranslated region of the NIPBL Gene. Hum Mutat 27: 731-735.

78. Liu J, Zhang Z, Bando M, Itoh T, Deardorff MA, et al. (2009) Transcriptional dysregulation in NIPBL and cohesin mutant human cells. PLoS Biol 7: e1000119.

79. Chien R, Zeng W, Ball AR, Yokomori K (2011) Cohesin: a critical chromatin organizer in mammalian gene regulation. Biochemistry and cell biology = Biochimie et biologie cellulaire 89: 445-458.

80. Heidinger-Pauli JM, Mert O, Davenport C, Guacci V, Koshland D (2010) Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. Curr Biol 20: 957-963.

81. Kong X, Ball ARJ, Sonoda E, Feng J, Takeda S, et al. (2009) Cohesin associates with spindle poles in a mitosis-specific manner and functions in spindle assembly in vertebrate cells. Mol Biol Cell 20: 1289-1301.

82. Schaaf CA, Misulovin Z, Sahota G, Siddiqui AM, Schwartz YB, et al. (2009) Regulation of the Drosophila Enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. PLoS One 4: e6202.

83. Kaur M, DeScipio C, McCallum J, Yaeger D, Devoto M, et al. (2005) Precocious sister chromatid separation (PSCS) in Cornelia de Lange syndrome. American journal of medical genetics Part A 138: 27-31.

84. Revenkova E, Focarelli ML, Susani L, Paulis M, Bassi MT, et al. (2009) Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA. Hum Mol Genet 18: 418-427.

85. Castronovo P, Gervasini C, Cereda A, Masciadri M, Milani D, et al. (2009) Premature chromatid separation is not a useful diagnostic marker for Cornelia de Lange syndrome. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology 17: 763-771.

86. Vrouwe MG, Elghalbzouri-Maghrani E, Meijers M, Schouten P, Godthelp BC, et al. (2007) Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. Human molecular genetics 16: 1478-1487.

87. Muto A, Calof AL, Lander AD, Schilling TF (2011) Multifactorial origins of heart and gut defects in nipbl-deficient zebrafish, a model of Cornelia de Lange Syndrome. PLoS biology 9: e1001181.

88. van der Maarel SM, Frants RR (2005) The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. Am J Hum Genet 76: 375-386.

89. Pandya S, King WM, Tawil R (2008) Facioscapulohumeral dystrophy. Phys Ther 88: 105-113.

90. van der Maarel SM, Tawil R, Tapscott SJ (2011) Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends Mol Med 17: 252-258.

91. de Greef JC, Lemmers RJ, Camaño P, Day JW, Sacconi S, et al. (2010) Clinical features of facioscapulohumeral muscular dystrophy 2. Neurol 75: 1548-1554.

92. Gabriëls J, Beckers MC, Ding H, De Vriese A, Plaisance S, et al. (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene 236: 25-32.

93. Geng LN, Yao Z, Snider L, Fong AP, Cech JN, et al. (2012) DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. Dev Cell 22: 38-51.

94. Snider L, Geng LN, Lemmers RJ, Kyba M, Ware CB, et al. (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics 6: e1001181.

95. Bosnakovski D, Xu Z, Gang EJ, Galindo CL, Liu M, et al. (2008) An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. EMBO J 27: 2766-2779.

96. Vanderplanck C, Ansseau E, Charron S, Stricwant N, Tassin A, et al. (2011) The FSHD atrophic myotube phenotype is caused by DUX4 expression. PLos ONE 6: e26820.

97. Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, et al. (2010) A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329: 1650-1653.

98. Ashe A, Morgan DK, Whitelaw NC, Bruxner TJ, Vickaryous NK, et al. (2008) A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. Genome biology 9: R182.

99. Blewitt ME, Gendrel AV, Pang Z, Sparrow DB, Whitelaw N, et al. (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nature genetics 40: 663-669.

100. Lemmers RJLF, Tawil R, Petek LM, Balog J, Block GJ, et al. (2012) Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet Epub.

101. Zeng W, de Greef JC, Chen YY, Chien R, Kong X, et al. (2009) Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated

with facioscapulohumeral dystrophy (FSHD). PLoS Genet 5: e1000559.

102. Balog J, Thijssen PE, de Greef JC, Shah B, van Engelen BG, et al. (2012) Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigeneti 7: 579-584.

103. Dheur S, Saupe SJ, Genier S, Vazquez S, Javerzat JP (2011) Role for cohesin in the formation of a heterochromatic domain at fission yeast subtelomeres. Mol Cell Biol 31: 1088-1097.

104. Ball AR, Jr., Chen YY, Yokomori K (2014) Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies. Biochimica et biophysica acta 1839: 191-202.

105. Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, et al. (2004) The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev 18: 2485-2490.

106. Song SH, Hou C, Dean A (2007) A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Mol Cell 28: 810-822.

107. Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, et al. (2005) Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell 17: 453-462.

108. Brand M, Ranish JA, Kummer NT, Hamilton J, Igarashi K, et al. (2004) Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. Nat Struc Mol Biol 11: 73-80.

109. Ciosk R, Shirayama M, Shevchenko A, Tanaka T, Toth A, et al. (2000) Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Molecular cell 5: 243-254.

110. Xiong B, Gerton JL (2010) Regulators of the cohesin network. Annu Rev Biochem 79: 131-153.

111. Gregson HC, Schmiesing JA, Kim JS, Kobayashi T, Zhou S, et al. (2001) A potential role for human cohesin in mitotic spindle aster assembly. The Journal of biological chemistry 276: 47575-47582.

112. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10: R25.

113. Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, et al. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. The

EMBO journal 24: 800-812.

114. Zeng W, Chen YY, Newkirk DA, Wu B, Balog J, et al. (2014) Genetic and Epigenetic Characteristics of FSHD-Associated 4q and 10q D4Z4 that are Distinct from Non-4q/10q D4Z4 Homologs. Human mutation.

115. Newkirk D, Biesinger J, Chon A, Yokomori K, Xie X (2011) AREM: aligning short reads from ChIP-sequencing by expectation maximization. Journal of computational biology : a journal of computational molecular cell biology 18: 1495-1505.

116. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, et al. (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148: 335-348.

117. Dale RK, Pedersen BS, Quinlan AR (2011) Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics 27: 3423-3424.

118. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.

119. Dean CB, Nielsen JD (2007) Generalized linear mixed models: a review and some extensions. Lifetime data analysis 13: 497-512.

120. Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, et al. (2010) The UCSC Genome Browser database: update 2010. Nucleic acids research 38: D613-619.

121. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome research 12: 996-1006.

122. Long AD, Mangalam HJ, Chan BY, Tolleri L, Hatfield GW, et al. (2001) Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12. The Journal of biological chemistry 276: 19937-19944.

123. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, et al. (2003) PANTHER: a library of protein families and subfamilies indexed by function. Genome research 13: 2129-2141.

124. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, et al. (2005) The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic acids research 33: D284-288.

125. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57.

126. Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths

toward the comprehensive functional analysis of large gene lists. Nucleic acids research 37: 1-13.

127. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, et al. (2006) GenePattern 2.0. Nature genetics 38: 500-501.

128. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545-15550.

129. Kampstra P (2008) Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. J Statistical Software.

130. Kooren J, Palstra RJ, Klous P, Splinter E, von Lindern M, et al. (2007) Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. The Journal of biological chemistry 282: 16544-16552.

131. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, et al. (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes & development 20: 2349-2354.

132. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular cell 10: 1453-1465.

133. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, et al. (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132: 422-433.

134. Hakimi MA, Bochar DA, Schmiesing JA, Dong Y, Barak OG, et al. (2002) A chromatin remodelling complex that loads cohesin onto human chromosomes. Nature 418: 994-998.

135. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods 5: 621-628.

136. Remeseiro S, Cuadrado A, Kawauchi S, Calof AL, Lander AD, et al. (2013) Reduction of Nipbl impairs cohesin loading locally and affects transcription but not cohesion-dependent functions in a mouse model of Cornelia de Lange Syndrome. Biochimica et biophysica acta 1832: 2097-2102.

137. Shimura M, Toyoda Y, Iijima K, Kinomoto M, Tokunaga K, et al. (2011) Epigenetic displacement of HP1 from heterochromatin by HIV-1 Vpr causes premature sister chromatid separation. The Journal of cell biology 194: 721-735.

138. Kline AD, Barr M, Jackson LG (1993) Growth manifestations in the Brachmann-de Lange

syndrome. American journal of medical genetics 47: 1042-1049.

139. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459: 108-112.

140. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics 39: 311-318.

141. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, et al. (2011) H3K4 tri-methylation provides an epigenetic signature of active enhancers. The EMBO journal 30: 4198-4210.

142. Remeseiro S, Cuadrado A, Gomez-Lopez G, Pisano DG, Losada A A unique role of cohesin-SA1 in gene regulation and development. EMBO J.

143. Rollins RA, Morcillo P, Dorsett D (1999) Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics 152: 577-593.

144. Ansari A, Hampsey M (2005) A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. Genes & development 19: 2969-2978.

145. Fay A, Misulovin Z, Li J, Schaaf CA, Gause M, et al. (2011) Cohesin selectively binds and regulates genes with paused RNA polymerase. Curr Biol 21: 1624-1634.

146. Schaaf CA, Kwak H, Koenig A, Misulovin Z, Gohara DW, et al. (2013) Genome-wide control of RNA polymerase II activity by cohesin. PLoS genetics 9: e1003382.

147. Gomes NP, Espinosa JM (2010) Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding. Genes & development 24: 1022-1034.

148. Horsfield JA, Print CG, Monnich M (2012) Diverse developmental disorders from the one ring: distinct molecular pathways underlie the cohesinopathies. Frontiers in genetics 3: 171.

149. Papai G, Weil PA, Schultz P (2011) New insights into the function of transcription factor TFIID from recent structural studies. Current opinion in genetics & development 21: 219-224.

150. Nolen LD, Boyle S, Ansari M, Pritchard E, Bickmore WA (2013) Regional chromatin decompaction in Cornelia de Lange syndrome associated with NIPBL disruption can be uncoupled from cohesin and CTCF. Human molecular genetics 22: 4180-4193.

151. Zuin J, Franke V, van Ijcken WF, van der Sloot A, Krantz ID, et al. (2014) A cohesin-independent role for NIPBL at promoters provides insights in CdLS. PLoS genetics 10: e1004153.

152. Nasmyth K, Haering CH (2009) Cohesin: its roles and mechanisms. Annual review of genetics 43: 525-558.

153. Seitan VC, Merkenschlager M (2012) Cohesin and chromatin organisation. Current opinion in genetics & development 22: 93-100.

154. Jackson L, Kline AD, Barr MA, Koch S (1993) de Lange syndrome: a clinical review of 310 individuals. American journal of medical genetics 47: 940-946.

155. Ireland M, Donnai D, Burn J (1993) Brachmann-de Lange syndrome. Delineation of the clinical phenotype. American journal of medical genetics 47: 959-964.

156. Bose T, Lee KK, Lu S, Xu B, Harris B, et al. (2012) Cohesin proteins promote ribosomal RNA production and protein translation in yeast and human cells. PLoS genetics 8: e1002749.

157. Gard S, Light W, Xiong B, Bose T, McNairn AJ, et al. (2009) Cohesinopathy mutations disrupt the subnuclear organization of chromatin. J Cell Biol 187: 455-462.

158. Boulon S, Westman BJ, Hutten S, Boisvert FM, Lamond AI (2010) The nucleolus under stress. Molecular cell 40: 216-227.

159. Grummt I (2013) The nucleolus-guardian of cellular homeostasis and genome integrity. Chromosoma 122: 487-497.

160. Wang ZQ, Auer B, Stingl L, Berghammer H, Haidacher D, et al. (1995) Mice lacking ADPRT and poly(ADP-ribosyl)ation develop normally but are susceptible to skin disease. Genes & development 9: 509-520.

161. Shiomi K, Kiyono T, Okamura K, Uezumi M, Goto Y, et al. (2011) CDK4 and cyclin D1 allow human myogenic cells to recapture growth property without compromising differentiation potential. Gene therapy 18: 857-866.

162. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, et al. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nature structural & molecular biology 17: 909-915.

163. Yao C, Biesinger J, Wan J, Weng L, Xing Y, et al. (2012) Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. Proceedings of the National Academy of Sciences of the United States of America 109: 18773-18778.

164. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, et al. (2011) iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. Journal of visualized experiments : JoVE.

165. Kong X, Ball AR, Jr., Pham HX, Zeng W, Chen HY, et al. (2014) Distinct functions of human cohesin-SA1 and cohesin-SA2 in double-strand break repair. Molecular and cellular biology 34: 685-698.

166. Hannan KM, Sanij E, Rothblum LI, Hannan RD, Pearson RB (2012) Dysregulation of RNA polymerase I transcription during disease. Biochimica et biophysica acta.

167. Rhodes JM, Bentley FK, Print CG, Dorsett D, Misulovin Z, et al. (2010) Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. Developmental biology 344: 637-649.

168. Corton JM, Gillespie JG, Hawley SA, Hardie DG (1995) 5-aminoimidazole-4-carboxamide ribonucleoside. A specific method for activating AMP-activated protein kinase in intact cells? European journal of biochemistry / FEBS 229: 558-565.

169. Katayoun Jessen SW, Linda Kessler, Xin Guo, Jeff Kucharski, Jocelyn Staunton, Lucy Lan, Marikka Elia, Josh Stewart, Jolene Brown, Liansheng Li, Katrina Chan, Michael Martin, Pingda Ren, Christian Rommel and Yi Liu (2009) INK128 is a potent and selective TORC1/2 inhibitor with broad oral antitumor activity. Molecular Cancer Therapeutics 8.

170. Finkel T, Holbrook NJ (2000) Oxidants, oxidative stress and the biology of ageing. Nature 408: 239-247.

171. Bensaude O (2011) Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? Transcription 2: 103-108.

172. Shav-Tal Y, Blechman J, Darzacq X, Montagna C, Dye BT, et al. (2005) Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. Molecular biology of the cell 16: 2395-2413.

173. Chen S, Seiler J, Santiago-Reichelt M, Felbel K, Grummt I, et al. (2013) Repression of RNA polymerase I upon stress is caused by inhibition of RNA-dependent deacetylation of PAF53 by SIRT7. Molecular cell 52: 303-313.

174. Grob A, Colleran C, McStay B (2014) Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. Genes & development 28: 220-230.

175. Xu B, Lee KK, Zhang L, Gerton JL (2013) Stimulation of mTORC1 with L-leucine rescues defects associated with Roberts syndrome. PLoS genetics 9: e1003857.

176. Harris B, Bose T, Lee KK, Wang F, Lu S, et al. (2014) Cohesion promotes nucleolar structure and function. Molecular biology of the cell 25: 337-346.

177. Lu S, Lee KK, Harris B, Xiong B, Bose T, et al. (2014) The cohesin acetyltransferase Eco1 coordinates rDNA replication and transcription. EMBO reports 15: 609-617.

178. Lengronne A, Katou Y, Mori S, Yokobayashi S, Kelly GP, et al. (2004) Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature 430: 573-578.

179. Laloraya S, Guacci V, Koshland D (2000) Chromosomal addresses of the cohesin component Mcd1p. The Journal of cell biology 151: 1047-1056.

180. Jaako P, Debnath S, Olsson K, Bryder D, Flygare J, et al. (2012) Dietary L-leucine improves the anemia in a mouse model for Diamond-Blackfan anemia. Blood 120: 2225-2228.

181. Payne EM, Virgilio M, Narla A, Sun H, Levine M, et al. (2012) L-Leucine improves the anemia and developmental defects associated with Diamond-Blackfan anemia and del(5q) MDS by activating the mTOR pathway. Blood 120: 2214-2224.

182. Baoshan Xu NSaJG (2014) L-Leucine partially ameliorates developmental deficits in zebrafish models for Cornelia de Lange syndrome. The FASEB Journal 28.

183. van de Nobelen S, Rosa-Garrido M, Leers J, Heath H, Soochit W, et al. CTCF regulates the local epigenetic state of ribosomal DNA repeats. Epigenetics Chromatin 3: 19.

184. Poortinga G, Wall M, Sanij E, Siwicki K, Ellul J, et al. (2011) c-MYC coordinately regulates ribosomal gene chromatin remodeling and Pol I availability during granulocyte differentiation. Nucleic acids research 39: 3267-3281.

185. Bouma GJ, Hart GT, Washburn LL, Recknagel AK, Eicher EM (2004) Using real time RT-PCR analysis to determine multiple gene expression patterns during XX and XY mouse fetal gonad development. Gene expression patterns : GEP 5: 141-149.

186. Hammoud L, Burger DE, Lu X, Feng Q (2009) Tissue inhibitor of metalloproteinase-3 inhibits neonatal mouse cardiomyocyte proliferation via EGFR/JNK/SP-1 signaling. American journal of physiology Cell physiology 296: C735-745.

187. Cong R, Das S, Ugrinova I, Kumar S, Mongelard F, et al. (2012) Interaction of nucleolin with ribosomal RNA genes and its role in RNA polymerase I transcription. Nucleic acids research 40: 9441-9454.

188. van Overveld PG, Lemmers RJ, Sandkuijl LA, Enthoven L, Winokur ST, et al. (2003) Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular

dystrophy. Nat Genet 35: 315-317.

189. Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, et al. (2002) Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nature genetics 32: 235-236.

190. Lemmers RJ, Wohlgemuth M, Frants RR, Padberg GW, Morava E, et al. (2004) Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. American journal of human genetics 75: 1124-1130.

191. Lemmers RJ, Wohlgemuth M, van der Gaag KJ, van der Vliet PJ, van Teijlingen CM, et al. (2007) Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. American journal of human genetics 81: 884-894.

192. Scionti I, Greco F, Ricci G, Govi M, Arashiro P, et al. (2012) Large-scale population analysis challenges the current criteria for the molecular diagnosis of fascioscapulohumeral muscular dystrophy. American journal of human genetics 90: 628-635.

193. Snider L, Asawachaicharn A, Tyler AE, Geng LN, Petek LM, et al. (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. Human molecular genetics 18: 2414-2430.

194. Bosnakovski D, Lamb S, Simsek T, Xu Z, Belayew A, et al. (2008) DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. Experimental neurology 214: 87-96.

195. Mitsuhashi H, Mitsuhashi S, Lynn-Jones T, Kawahara G, Kunkel LM (2013) Expression of DUX4 in zebrafish development recapitulates facioscapulohumeral muscular dystrophy. Human molecular genetics 22: 568-577.

196. Jones TI, Chen JC, Rahimov F, Homma S, Arashiro P, et al. (2012) Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. Human molecular genetics 21: 4419-4430.

197. Rahimov F, King OD, Leung DG, Bibat GM, Emerson CP, Jr., et al. (2012) Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. Proceedings of the National Academy of Sciences of the United States of America 109: 16234-16239.

198. Broucqsault N, Morere J, Gaillard MC, Dumonceaux J, Torrents J, et al. (2013) Dysregulation of 4q35- and muscle-specific genes in fetuses with a short D4Z4 array linked to facio-scapulo-humeral dystrophy. Human molecular genetics 22: 4206-4214.

199. Ferreboeuf M, Mariot V, Bessieres B, Vasiljevic A, Attie-Bitach T, et al. (2013) DUX4 and

DUX4 downstream target genes are expressed in fetal FSHD muscles. Human molecular genetics 23: 171-181.

200. Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. Nature 447: 433-440.

201. Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, et al. (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proceedings of the National Academy of Sciences of the United States of America 96: 14412-14417.

202. Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, et al. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402: 187-191.

203. Winokur ST, Chen YW, Masny PS, Martin JH, Ehmsen JT, et al. (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. Human molecular genetics 12: 2895-2907.

204. Cheli S, Francois S, Bodega B, Ferrari F, Tenedini E, et al. (2011) Expression profiling of FSHD-1 and FSHD-2 cells during myogenic differentiation evidences common and distinctive gene dysregulation patterns. PLoS One 6: e20966.

205. Zeng W, Ball AR, Jr., Yokomori K HP1: heterochromatin binding proteins working the genome. Epigenetics 5: 287-292.

206. Li Y, Danzer JR, Alvarez P, Belmont AS, Wallrath LL (2003) Effects of tethering HP1 to euchromatic regions of the Drosophila genome. Development 130: 1817-1824.

207. Hattori T, Taft JM, Swist KM, Luo H, Witt H, et al. (2013) Recombinant antibodies to histone post-translational modifications. Nature methods 10: 992-995.

208. Thijssen PE, Tobi EW, Balog J, Schouten SG, Kremer D, et al. (2013) Chromatin remodeling of human subtelomeres and TERRA promoters upon cellular senescence: commonalities and differences between chromosomes. Epigenetics : official journal of the DNA Methylation Society 8: 512-521.

209. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, et al. (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature biotechnology 26: 317-325.

210. John S, Sabo PJ, Canfield TK, Lee K, Vong S, et al. (2013) Genome-scale mapping of DNase I hypersensitivity. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] Chapter 27: Unit 21 27.

211. Tibodeau JD, Benson LM, Isham CR, Owen WG, Bible KC (2009) The anticancer agent

chaetocin is a competitive substrate and inhibitor of thioredoxin reductase. Antioxidants & redox signaling 11: 1097-1106.

212. Mostert V, Hill KE, Burk RF (2003) Loss of activity of the selenoenzyme thioredoxin reductase causes induction of hepatic heme oxygenase-1. FEBS letters 541: 85-88.

213. Trigona WL, Mullarky IK, Cao Y, Sordillo LM (2006) Thioredoxin reductase regulates the induction of haem oxygenase-1 expression in aortic endothelial cells. The Biochemical journal 394: 207-216.

214. Turki A, Hayot M, Carnac G, Pillard F, Passerieux E, et al. (2012) Functional muscle impairment in facioscapulohumeral muscular dystrophy is correlated with oxidative stress and mitochondrial dysfunction. Free radical biology & medicine 53: 1068-1079.

215. Iwasa E, Hamashima Y, Fujishiro S, Higuchi E, Ito A, et al. (2010) Total synthesis of (+)-chaetocin and its analogues: their histone methyltransferase G9a inhibitory activity. Journal of the American Chemical Society 132: 4078-4079.

216. Duan Z, Zarebski A, Montoya-Durango D, Grimes HL, Horwitz M (2005) Gfi1 coordinates epigenetic repression of p21Cip/WAF1 by recruitment of histone lysine methyltransferase G9a and histone deacetylase 1. Molecular and cellular biology 25: 10338-10351.

217. Nozawa RS, Nagao K, Igami KT, Shibata S, Shirai N, et al. (2013) Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nature structural & molecular biology 20: 566-573.

218. Debus E, Weber K, Osborn M (1983) Monoclonal antibodies to desmin, the muscle-specific intermediate filament protein. The EMBO journal 2: 2305-2312.

219. Williams AH, Liu N, van Rooij E, Olson EN (2009) MicroRNA control of muscle development and disease. Current opinion in cell biology 21: 461-469.

220. Jen Y, Weintraub H, Benezra R (1992) Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. Genes & development 6: 1466-1479.

221. Sacconi S, Lemmers RJ, Balog J, van der Vliet PJ, Lahaut P, et al. (2013) The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics 93: 744-751.

222. Gendrel AV, Apedaile A, Coker H, Termanis A, Zvetkova I, et al. (2012) Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Developmental cell 23: 265-279.

223. Li W, Notani D, Ma Q, Tanasa B, Nunez E, et al. (2013) Functional roles of enhancer RNAs

for oestrogen-dependent transcriptional activation. Nature: doi: 10.1038/nature12210.

224. Pikaard CS, Pape LK, Henderson SL, Ryan K, Paalman MH, et al. (1990) Enhancers for RNA polymerase I in mouse ribosomal DNA. Molecular and cellular biology 10: 4816-4825.

225. Wang L, Brown SJ (2006) BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic acids research 34: W243-248.

226. Yang Y, Zhao H, Wang J, Zhou Y (2014) SPOT-Seq-RNA: predicting protein-RNA complex structure and RNA-binding function by fold recognition and binding affinity prediction. Methods in molecular biology 1137: 119-130.

227. Oka Y, Suzuki K, Yamauchi M, Mitsutake N, Yamashita S (2011) Recruitment of the cohesin loading factor NIPBL to DNA double-strand breaks depends on MDC1, RNF168 and HP1gamma in human cells. Biochemical and biophysical research communications 411: 762-767.