Characterization of Nucleolus-Associated Domains in Mouse Embryonic Stem Cells

Aizhan Bizhanova

University of Massachusetts Medical School
CHARACTERIZATION OF NUCLEOLUS-ASSOCIATED DOMAINS
IN MOUSE EMBRYONIC STEM CELLS

A Dissertation Presented

By

AIZHAN BIZHANOVA

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 5, 2020

INTERDISCIPLINARY GRADUATE PROGRAM
CHARACTERIZATION OF NUCLEOLUS-ASSOCIATED DOMAINS

IN MOUSE EMBRYONIC STEM CELLS

A Dissertation Presented

By

AIZHAN BIZHANOVA

This work was undertaken in the Graduate School of Biomedical Sciences
Interdisciplinary Graduate Program

Under the mentorship of
Paul Kaufman, Ph.D., Thesis Advisor
Thomas Fazzio, Ph.D., Member of Committee
Oliver Rando, M.D., Ph.D., Member of Committee
Anthony Imbalzano, Ph.D., Member of Committee
Andrew Stephens, External Member of Committee
Craig Peterson, Ph.D., Chair of Committee

Mary Ellen Lane, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

May 5, 2020
ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to my thesis advisor Paul Kaufman for all his support and guidance during my graduate studies. I was fortunate to join his lab and benefit from his research expertise and mentorship. Dr. Kaufman has always been open to scientific discussions and troubleshooting and has been understanding and caring as an advisor. Truly, the work presented here would not be possible without his mentorship.

I would like to thank all current and past members of the Kaufman lab: Ana Vertii, Osama Garwain, Tim Matroni, Dan Trombly, Xiaoming Sun, Tim Matheson and Yuichi Ichikawa for creating such a friendly lab environment. I am grateful to Ana Vertii for teaching me FISH assay and helping with F121-9 cell culture. I would like to thank Julie Zhu, Aimin Yan and Jun Yu for their incredible bioinformatics work and contribution to my thesis project. I am grateful to Kurtis McCannell from the Fazzio lab for his help in learning how to culture stem cells, to Meg Byron from the Lawrence lab for her tips on FISH assay, and to Jillian Lindblad from the Bergmann lab for helping me with confocal microscope.

I am very grateful to members of my TRAC and dissertation committee: Craig Peterson, Tom Fazzio, Tony Imbalzano, Oliver Rando and Andrew Stephens for their constructive feedback and advice. I also would like to thank all members of MCCB department and GSBS office for aiding in my graduate school experience.

I would like to thank my parents Abay Bizhanov and Bagym Pshayeva, and my sister Aigerim Bijelic for their love and support. Warm memories of my
mom Bagym propelled me to overcome challenges and strive for growth. Memories of my father Abay will encourage me to strive for a meaningful and happy life. I am grateful to my friends Dana Sapyshева, Zhibek Kadyrsizova, Indira Beisenova, Aneth Canale, Alka Singh, Heleen van ’t Spijker, Dana Abeuova and Maryam Akbari for their friendship and encouragement. Finally, I would like to thank my fiancé David Berthiaume for always being there for me.
ABSTRACT

In eukaryotic interphase cells, heterochromatin mostly localizes either at the nucleolar periphery or at the nuclear lamina. Genome localization studies are crucial due to evidence that spatial organization of the genome affects gene function. Nucleolus-associated domains (NADs) are mainly heterochromatic regions that have been mapped only in a handful of mouse and human somatic cells, and in plants. The extent to which changes in NAD localization occur during cellular differentiation remains unknown.

In this thesis, we characterize a map of genome-wide NADs in F121-9 mouse embryonic stem cells (mESCs). We identified NADs by deep sequencing chromatin associated with biochemically purified nucleoli and using \textit{NADfinder} software to call NAD peaks. F121-9 NADs are mostly comprised of genomic regions with inactive or lowly transcribed genes and overlap extensively with lamina-associated domains (LADs) and regions with late replication timing. Similar to somatic mouse embryonic fibroblasts (MEFs), where NADs have been previously characterized by our laboratory, F121-9 mESCs display abundant “Type I” NADs. This subset of NADs frequently associates with nuclear lamina and nucleolar periphery and resembles constitutive heterochromatin. Compared to MEFs, F121-9 mESCs have fewer “Type II” NADs; this subset of NADs is frequently found at the nucleolar periphery but not at the nuclear lamina. mESC NADs are also less enriched in H3K27me3 modified regions compared to MEF NADs. This suggests that Polycomb complex-mediated facultative
heterochromatin expansion is part of NAD maturation during cellular
differentiation. Comparison of MEF and mESC NADs also revealed enrichment
of developmentally regulated genes in NADs specific to these cell types.
Together, these data indicate that NADs are a developmentally dynamic
component of heterochromatin. Our F121-9 mESC NAD studies identified
distinct features of stem cell NADs and will facilitate future studies of genome
organization changes during mammalian development.
**TABLE OF CONTENTS**

Cover Page i  
Title Page ii  
Reviewer Page iii  
Acknowledgements iv  
Abstract vi  
Table of Contents viii  
List of Tables xi  
List of Figures xii  
List of copyrighted Materials Produced by the Author xiii  
List of Third Party Copyrighted Material xiv  
List of Abbreviations and Nomenclature xv  

**Chapter I: Introduction**  

Chromatin organization 1  
Heterochromatin 9  
Constitutive heterochromatin 10  
Facultative heterochromatin 15  
Spatial heterochromatin localization 17  
Nucleolus-associated Domains (NADs) 17  
*Cis* and *trans*-acting factors and mechanisms of nucleolar association 24  
Functional significance of NADs 32  
Lamina-associated Domains (LADs) 36  
*Cis* and *trans*-acting factors mediating the nuclear lamina-LAD interactions 40  
Overlap between NADs and LADs 43  
Features of heterochromatin in embryonic stem cells (ESCs) 45
# Chapter II: Distinct features of nucleolus-associated domains in mouse embryonic stem cells

- **Abstract** 52
- **Introduction** 53
- **Results** 55
- **Discussion** 90
- **Materials and Methods** 96

# Chapter III: Discussion

- Characteristics of mouse ESC NADs 106
- Anomalies of cell type-specific NADs in stem cells 110
- Allele-specific NAD features in F121-9 hybrid cells 113
- Possible mechanisms of nucleolar association 115
- The role of nucleolar association in developmental regulation 119

# Appendix A: Differentiation of mouse ESCs to neural precursor cells (NPCs)

- **Introduction** 121
- **Results** 122
- **Discussion** 129
- **Materials and Methods** 130

# Appendix B: Differentiation of mouse ESCs to epiblast-like cells (EpiLCs)

- **Introduction** 135
- **Results** 136
- **Discussion** 140
- **Materials and Methods** 140
Bibliography
LIST OF TABLES

Table 2.1: mm10 genomic coordinates, laboratory BAC probe names, systematic BACPAC names, FISH and NADfinder results for DNA-FISH probes.

Table A.1: Pre-differentiation medium recipe.

Table A.2: RT-qPCR primers used to validate NPC differentiation.

Table A.3: Antibodies used for mESC and NPC IF.

Table B.1: EpiLC differentiation medium.

Table A.2: RT-qPCR primers used in EpiLC differentiation assays.
LIST OF FIGURES

**Figure 1.1**: The hierarchy of chromatin organization.

**Figure 1.2**: Schematic diagram of nucleolar tripartite structure.

**Figure 2.1**: Analysis of F121-9 NAD sequencing data and comparison with heterochromatin.

**Figure 2.2**: Isolation and characterization of purified nucleoli in mESC.

**Figure 2.3**: Genomic locations of BACs used for FISH experiments.

**Figure 2.4**: 3D DNA-FISH experiments validate nucleolar association of NADs in F121-9 mESC.

**Fig. 2.5**: Two types of NADs in F121-9 mESC.

**Fig. 2.6**: Conserved and cell type-specific NADs.

**Fig. 2.7**: GO analysis of conserved and cell type-specific NADs.

**Fig. 2.8**: Allele-specific chromosome-nucleolus contacts.

**Figure 2.9**: Allele-specific NAD-seq data for chromosomes 5 and 12.

**Figure A.1**: Morphological changes during mESC to NPC differentiation.

**Figure A.2**: RT-qPCR results validate successful differentiation of mESCs into NPCs.

**Figure A.3**: NPCs show decreased expression of Oct3/4 and increased expression of nestin compared to mESCs.

**Figure B.1**: Morphological changes during mESC to EpiLC differentiation.

**Figure B.2**: Figure B.2. RT-qPCR results suggest successful differentiation of mESCs into EpiLCs.
LIST OF COPYRIGHTED MATERIALS PRODUCED BY THE AUTHOR

Chapter II is reprinted from the journal Chromosoma, license number 4800850990681.

# LIST OF THIRD PARTY COPYRIGHTED MATERIAL

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Publisher</th>
<th>License number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Springer Nature</td>
<td>4696610992533</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS AND NOMENCLATURE

3C: chromosome conformation capture
4DN: 4D Nucleome
5-BrdU: 5-bromodeoxyuridine
Ab: antibody
BAC: bacterial artificial chromosome
BMP: bone morphogenetic protein
CAF-1: chromatin assembly factor-1
CBX2: chromobox 2
CD: chromodomain
CENP-A: centromere protein A
ciLAD: constitutive interLAD
cLAD: constitutive LAD
CSD: chromoshadow domain
CTCF: CCCTC binding factor
CTs: chromosome territories
Dam: adenine methyltransferase
DamID: DNA adenine methyltransferase identification
DFC: dense fibrillar component
EED: embryonic ectoderm development
EpiLC: epiblast-like cell
EtoL: early to late
EZH2: enhancer of zeste 2
FBL: fibrillarin
FC: fibrillar center
FGF: fibroblast growth factor
fLAD: facultative LADs
FRAP: fluorescence recovery after photobleaching
GC: granular component
GO: Gene Ontology
GSK3β: glycogen synthase kinase 3β
HBB: human β-globin
HDAC3: histone deacetylase 3
Hox: homeobox genes
HP1: Heterochromatin protein 1
IDR: intrinsically disordered region
IF: immunofluorescence
IGS: intergenic spacer
iLAD: interLAD
INM: inner nuclear membrane
LADs: lamina-associated domain
LAS: lamina-associated sequence
LBR: lamin B receptor
LIF: leukemia inhibitory factor
LINE: long interspersed nuclear element
IncRNA: long non-coding RNA
LTR: long terminal repeat
MAPK/ERK: mitogen-activated protein kinase/extracellular signal-regulated kinase
MEFs: mouse embryonic fibroblasts
mESCs: mouse embryonic stem cells
miRNA: microRNA
NAD: nucleolus-associated domain
ncRNA: non-coding RNA
NL: nuclear lamina
NLP: nucleoplasmin-like protein
Nopp140: nucleolar phosphoprotein 140
NOR: Nucleolus organizer region
NoRC: nucleolar remodeling complex
NPCs: neural precursor cells
NPM-1: nucleophosmin
nt: nucleotide
PcG: Polycomb group
Ptn: Pleiotrophin
PNH: perinucleolar heterochromatin
PRC2: Polycomb Repressive Complex 2
pRNA: promoter-associated RNA

PTM: post-translational modifications

rDNA: Ribosomal DNA

RNA Pol II: RNA polymerase II

RNAi: RNA interference

RT: room temperature

SPRITE: Split-Pool Recognition of Interactions by Tag Extension

SUZ12: suppressor of zeste 12

TAD: topologically associated domain

TF: transcription factor

tRNA: transfer RNA

TTF1: transcription termination factor 1

UBF: upstream binding factor

Xa: active X

XCI: X chromosome inactivation

Xi: inactive X

Xic: X inactivation center

Xist: X inactive specific transcript
CHAPTER I: INTRODUCTION

Chromatin is a dynamic structure that consists of DNA and proteins; it allows for storage and expression of genetic material in the nucleus. Heterochromatin plays a crucial role in preserving genome stability by maintaining the silent state of a portion of the genome. Heterochromatin mainly localizes at the nuclear and nucleolar peripheries. Nucleolus-associated domains (NADs) are regions frequently found at the nucleolar periphery, and they comprise about one-third of a mammalian genome. Lamina-associated domains (LADs) are regions frequently associated with the nuclear periphery, and they show a similar genome coverage as NADs. While LADs have been extensively studied and mapped, the study of NADs is a relatively new field that awaits further discoveries.

This chapter describes the molecular bases of chromatin structure and mechanisms of heterochromatin establishment and maintenance, and outlines heterochromatin features in embryonic stem cells (ESCs). I will also discuss mechanisms of cis/trans-acting factors mediating the nucleolar and nuclear lamina associations, as well as the functional significance of these associations.

Chromatin organization

Eukaryotic cells store their genetic material in a structure called chromatin, which consists of DNA and proteins. The folding and compaction of DNA in this structure allows 2 meters of DNA to fit in the 10-20 μm volume of the nucleus. At
the same time, chromatin is a dynamic structure that regulates and permits the accessibility of DNA in processes such as replication, transcription and DNA repair.

A nucleosome, which is a basic unit of chromatin, consists of 147 bp of DNA wrapped around an octamer that contains two copies each of four histone proteins: H2A, H2B, H3 and H4 (Kornberg 1974; Oudet et al. 1975; Luger et al. 1997). Each nucleosome has a width of ~11 nm, and arrays of nucleosomes can be seen as “beads on a string” structure in electron micrographs (Olins and Olins 1974, 2003). These 11 nm fibers represent the first level of DNA compaction.

It is important to remember that chromatin is a dynamic structure that responds to external cues to regulate the genome. Histone post-translational modifications (PTM) play a key role in this regulation. Mainly found on the amino-terminal and carboxy-terminal histone tails, PTMs include phosphorylation, methylation, acetylation, ubiquitination, SUMOylation, deimination, etc. (Kouzarides 2007). There are two known mechanisms by which PTMs affect chromatin. The first one is a direct chromatin perturbation that disrupts contacts between histones in adjacent nucleosomes or changes DNA-histone interactions (Kouzarides 2007). For example, addition of acetyl group to lysine on histones neutralizes lysine’s positive charge and destabilizes the electrostatic interactions between histones and DNA and inhibits the compaction of nucleosomes into chromatin fibers (Shogren-Knaak et al. 2006). The second mechanism is based on the ability of histone modifications to recruit histone modifying enzymes, which
include histone readers, writers and erasers with protein domains that recognize specific histone PTMs. These histone modifying enzymes can then initiate a cascade of events that regulate DNA processes such as transcription, replication, repair and recombination (Kouzarides 2007). Histone modifications can also be recognized by domains in nucleosome remodelers, e.g. SWI/SNF remodeler contains a bromodomain that recognizes acetylated histones (Hassan et al. 2002). The remodeler complexes use the energy of ATP-hydrolysis to change histone-DNA interactions in nucleosomes. This can lead to nucleosome assembly or disassembly, histone variant exchange and histone dimer or octamer relocation (Becker and Workman 2013). Hence, nucleosome remodelers facilitate the accessibility or inaccessibility of nucleosomes to proteins, thereby affecting various processes in genome function.

Linker DNA connects nucleosomes and associates with linker histone H1, which was postulated to lead to further condensation of nucleosome arrays into 30 nm fibers, representing the second level of DNA compaction (Thoma et al. 1979). 30 nm fiber structure has been observed under cell-free conditions, or only in specialized cells such as starfish sperm and chicken erythrocytes (Woodcock 1994). However, cryogenic electron microscopy, electron spectroscopic imaging and X-ray scattering studies showed no evidence for 30 nm fibers in most other eukaryotic cells (Eltsov et al. 2008; Nishino et al. 2012; Fussner et al. 2012; Chen et al. 2016). Instead, some studies suggest that chromatin possesses a highly random organization with fractal characteristics
(Lebedev et al. 2005; Dekker 2008; Lieberman-Aiden et al. 2009; Bancaud et al. 2009; Nishino et al. 2012). However, a study that determined chromatin fragmentation patterns in living cells via ionizing radiation showed occurrence of variable well-defined oligonucleosomal structures \textit{in vivo}: heterochromatic regions displaying fragmentation patterns similar to zig-zag helix fibers, and euchromatic regions exhibiting patterns similar to extended fibers and solenoids (Risca et al. 2017). Therefore, existence of 30 nm fibers in cells is still debated, and it is likely that no single 30 nm structure exists \textit{in vivo}. Variations in histone variant composition, post-translational modifications and linker DNA lengths are some of the factors that likely result in mixed findings regarding 30 nm fiber structure (Woodcock and Ghosh 2010). Chromatin fibers are the fundamental basis of larger-scale chromatin structures that will be discussed below.

Recent advances in 3C and microscopy techniques allowed to make advances in our understanding of spatial organization of chromatin (Nir et al. 2018; Abbas et al. 2019). 3C techniques make inferences about physical interactions of regions based on frequency of ligation events that follow nuclear fixation and restriction enzyme digestion (Dekker et al. 2002). Different types and levels of chromatin organization were uncovered with the advent of 3C methods: TADs, loops, chromosome territories and compartments (Fig. 1.1A-C). A high-throughput 3C technique, called Hi-C is widely used in studies of genome-wide chromatin interactions. DNA recognition sequence of a restriction enzyme used in the chromatin digestion step is utilized to design biotin-labeled nucleotides,
which are incorporated at DNA junctions after the ligation step. This allows for selective purification of chimeric junctions and their deep sequencing (Belton et al. 2012).

Topologically associated domain (TAD) is defined as a domain which has more interactions within this domain and significantly fewer interactions with structures outside of a TAD, and is usually a tens of kilobases up to a few megabases long (Dixon et al. 2012; Nora et al. 2012; Rao et al. 2014) (Fig. 1.1C). TADs can be distinguished from compartments based on the shorter range of interactions (<3 Mb) in TADs, and the fact that TADs are seen along the diagonal of Hi-C contact map, whereas compartments exhibit checkerboard pattern not only along the diagonal, but also in interchromosomal and intrachromosomal areas of Hi-C map (Mirny et al. 2019). Compartments and TADs were thought to be organized hierarchically (Dixon et al. 2012), but recent high-resolution Hi-C data showed that these regions can overlap and coexist in the genome of interphase cells (Mirny et al. 2019).

TADs were shown to be relatively conserved in different cell types and species (Dixon et al. 2012; Vietri Rudan et al. 2015). The observation that CTCF and cohesin usually bind TAD boundaries led to the concept of a TAD being constrained by a loop forming at TAD boundaries (Dixon et al. 2012; Nora et al. 2012; Rao et al. 2014). According to the loop extrusion model, CTCF is thought to act as a boundary factor, and cohesin to play a role of loop extrusion factor (Sanborn et al. 2015). Importantly, it has been observed that loop-demarcated
TADs almost always form between CTCF binding sites in a convergent orientation (Rao et al. 2014; Vietri Rudan et al. 2015). Studies showed that depletion of cohesin leads to loss of TADs (Rao et al. 2017; Schwarzer et al. 2017), whereas loss of CTCF results in fusion of TADs (Sanborn et al. 2015; Nora et al. 2017), thereby supporting the loop extrusion model. Conversely, mitotic chromosomes lack TADs and compartments, instead forming an array of nested loops emanating from a central scaffold, where loop extrusion is achieved by condensins (Naumova et al. 2013; Gibcus et al. 2018). Such loop arrays display a “bottle-brush” shape, where chromatin achieves ~1000-fold lengthwise compaction (Mirny et al. 2019).

TADs represent not only organizational units of chromatin, but functional ones too (Serizay and Ahringer 2018). Replication timing domains (Pope et al. 2014) and regulatory domains largely overlap with TADs. Studies show that TADs facilitate enhancer and promoter interactions by restricting their contacts to occur within a TAD (Nora et al. 2012; Symmons et al. 2014; Mifsud et al. 2015). Disruption of TAD boundaries have been shown to affect gene expression by causing improper enhancer-promoter interactions (Narendra et al. 2015; Lupiáñez et al. 2015; Guo et al. 2015). TADs also segregate based on their association with histone marks and transcriptional activity of genes within a TAD, hence TADs can be designated as active and repressed ones (Entrevan et al. 2016).
Another type of chromatin spatial organization called compartments organize chromatin in interphase cells in TAD-independent manner into two main categories: A and B compartments (Lieberman-Aiden et al. 2009; Rao et al. 2014) (Fig. 1.1B). These compartments have higher contact frequency with regions of the same compartment type (A-A, B-B), and lower contact frequency between regions of different types (A-B). The A compartment corresponds to active chromatin, which is characterized by chromatin accessibility, transcriptional activity and H3K36me3 enrichment, whereas the B compartment is associated with inactive chromatin: more densely packed than compartment A, transcriptionally inactive and enriched for heterochromatic mark H3K27me3 (Lieberman-Aiden et al. 2009).

Individual interphase chromosomes occupy distinct and relatively reproducible nuclear volumes called chromosome territories (CTs) (Cremer and Cremer 2010), which represent the highest level of chromatin compaction in interphase cells (Fig. 1.1A). CTs near nuclear periphery and nucleolar periphery tend to consist of inactive chromatin, whereas the active CTs tend to occupy internal positions in the nucleus (Croft et al. 1999).

Chromatin structure and organization influence and are interdependent with gene expression and function: e.g., A/B compartments segregate based on these regions’ transcriptional activity, and TAD boundaries facilitate proper enhancer-promoter interactions and gene activity (Rowley and Corces 2018).
Figure 1.1: The hierarchy of chromatin organization.

A) Interphase chromosomes (shown in different colors) occupy distinct chromosome territories.

B) At a smaller scale, chromatin is organized into an active compartment A, and an inactive compartment B. Compartment A regions are often found near nuclear speckles, whereas compartment B regions frequently associate with nucleolar and nuclear peripheries.

C) At a smaller scale, topologically associated domain (TAD) is comprised of regions with more interactions within this domain and significantly fewer interactions with structures outside of a TAD. CTCF-binding sites are frequently found at TAD boundaries. Adapted by permission from Springer Nature (Zheng and Xie 2019), Copyright (2019).
Heterochromatin

The terms euchromatin and heterochromatin were coined by Emil Heitz in 1928, when he observed that parts of chromosomes were not stained after telophase (euchromatin) and other parts were stained throughout the cell cycle (heterochromatin) (Heitz 1928). DNA dyes have a higher affinity for AT-rich heterochromatic regions and concentrate in these compacted regions, hence heterochromatin is stained more darkly compared to euchromatin (Politz et al. 2013). Heterochromatin is defined as regions of chromatin that are condensed and transcriptionally inactive, whereas euchromatic regions are decondensed and show more active gene expression. Euchromatic regions are generally accessible to DNase I cleavage, while heterochromatic regions are more resistant to DNase I treatment (Weintraub and Groudine 1976). Studies of replication timing based on readout of 5-BrdU incorporation into DNA during S-phase showed that euchromatic regions replicate in early S-phase, and heterochromatin regions replicate in mid to late S-phase (O’Keefe et al. 1992; Wu et al. 2005).

Heterochromatin is divided into two categories: constitutive and facultative heterochromatin. Constitutive heterochromatin is thought to localize in the same regions in every cell type, it marks repeat-rich sites and generally lacks genes or is gene-poor (Saksouk et al. 2015). Obtaining a complete map of the repeat-rich sites through next generation sequencing (NGS) still remains a challenge (Nishibuchi and Déjardin 2017). Facultative heterochromatin is found at different genomic regions and is commonly repressed in specific cell types or during
particular developmental stages (Politz et al. 2013). Heterochromatin in general is characterized by global hypoacetylation, which contributes to its condensation. Constitutive heterochromatin is decorated with histone PTM H3K9me2/3, while facultative heterochromatin is enriched for H3K27me3 marks (Saksouk et al. 2015).

**Constitutive heterochromatin**

Constitutive heterochromatin is mainly found at telomeres, centromeres and adjacent silent regions (subtelomeres and pericentric regions, respectively). These regions are highly condensed, highly repetitive, constitutively repressed and contain repressive H3K9me2/3 and H4K20me2/3 marks, cytosine methylation at CpG dinucleotides and are often bound by heterochromatin protein 1 (HP1) (Politz et al. 2013). Indeed, it is thought that the fundamental mechanism of constitutive heterochromatin organization is H3K9me2/3 mark deposition and HP1 binding these marks (Nishibuchi and Déjardin 2017). H3K9 methylation is carried out by Suppressor of variegation (Su-var) 3-9, Enhancer-of-zeste and Trithorax (SET) domain-containing histone methyltransferases (Nicetto and Zaret 2019). G9a/GLP complex monomethylates and dimethylates H3K9 in euchromatic regions (Tachibana et al. 2001, 2005), whereas SETDB1, SUV39H1 and SUV39H2 catalyze H3K9me2/3 formation in mammals (Rea et al. 2000; Schultz et al. 2002). Studies in mice showed that knockout of both SUV39H1 and SUV39H2 leads to decreased H3K9me3 levels specifically in
heterochromatin (Peters et al. 2001). Loss of SETDB1 leads to diminished levels of H3K9me3 in euchromatic regions (Schultz et al. 2002) and pericentric heterochromatin (Loyola et al. 2009).

Transcription of repetitive DNA and subsequent binding of RNA interference (RNAi) machinery are known to recruit SUV39 histone lysine methyltransferase homologs in yeast and plants (Allshire and Madhani 2018), but it is unknown whether this mechanism of heterochromatin establishment is conserved in mammals (Janssen et al. 2018). There is evidence that transient expression of pericentric repeats promotes chromocenter formation in mice (Probst et al. 2010). Chromocenters are heterochromatic clusters formed through the coalescence of pericentric repeats (Guenatri et al. 2004), where SUV39 and HP1 play important role in heterochromatin establishment and proper centromere function (Peters et al. 2001; Probst and Almouzni 2011). Recent studies provide additional evidence that RNA-dependent mechanisms may play role in SUV39H1/H2 recruitment in human and mice (Johnson et al. 2017; Shirai et al. 2017; Velazquez Camacho et al. 2017).

Following H3K9me2/3 establishment, HP1 binds these marks (Bannister et al. 2001) through its chromodomain (CD) at the N-terminus and self-dimerizes and interacts with various proteins through its chromoshadow domain (CSD) at the C-terminus, while the linker domain called the hinge binds RNA and DNA (Zeng et al. 2010). HP1 was first identified as a heterochromatin-associated protein in Drosophila (James and Elgin 1986), and was the first heterochromatic
protein determined to be a modifier of position-effect variegation in fruit flies (Eissenberg et al. 1990). Mammalian HP1 variants are HP1α, HP1β and HP1γ. HP1α and HP1β are found in constitutive heterochromatic regions, whereas HP1γ localizes in both euchromatic and heterochromatic regions (Minc et al. 1999, 2000).

HP1 self-dimerizes and facilitates compaction of heterochromatic regions (Canzio et al. 2013). Importantly, HP1 serves as a platform for many other regulatory proteins. HP1β is known to interact with DNA methyltransferases DNMT1 and DNMT3a (Fuks et al. 2003), DNA methylation being known to elicit gene repressive effects on transcription. Notably, HP1 binds SUV39 (Aagaard et al. 1999) and deacetylases (Motamedi et al. 2008; Fischer et al. 2009), which promotes a cycle of deacetylation, methylation, HP1 binding and recruitment of histone deacetylases and methyltransferases. This is one of major mechanisms of heterochromatin maintenance and spreading, well-described in fission yeast (Janssen et al. 2018).

Recent studies showed that HP1α (Larson et al. 2017; Strom et al. 2017) and HP1β (Wang et al. 2019) can undergo liquid-liquid phase separation. Liquid droplet formation is driven by the intrinsically disordered region (IDR) found in the N-terminal and hinge domains of HP1 and requires either N-terminal phosphorylation of HP1, or DNA binding (Larson et al. 2017), or H3K9me3 binding (Wang et al. 2019). The exclusion of transcription factors at the surface
of liquid droplets, rather than chromatin compaction, is postulated to render inaccessibility of heterochromatin domains (Larson et al. 2017).

As mentioned previously, constitutive heterochromatin is mainly found at telomeres and centromeres. Telomeric sequences are located at the end of chromosomes, where various proteins safeguard the ends from being recognized as DNA double-stranded breaks (shelterin complex) and elongate telomere ends that shorten during DNA replication (telomerase, expressed in stem and cancer cells) (Greider and Blackburn 1985; de Lange 2005). Mammalian telomeres harbor an array of tandem repeats of TTAGGG motif that is several kilobases long (Tardat and Déjardin 2018). Tandem repeats consist of units that are adjacent to each other and are organized as direct or inverted repeats (McNulty and Sullivan 2018). Constitutive heterochromatin prevents homologous recombination in telomeric and subtelomeric regions (Nishibuchi and Déjardin 2017). Heterochromatin in general plays important role in structural stability of telomeres, e.g. deletion of Suv39H1/2 mouse methyltransferases leads to abnormal telomere elongation (García-Cao et al. 2004), and deletion of DNA methyltransferase DNMT1 results in telomere elongation and increased recombination events at telomeres (Gonzalo et al. 2006). Telomeric sequences and protein complexes that bind them are largely conserved, which suggests that these domains are not epigenetically regulated (Saksouk et al. 2015).

The centromeres are sites of primary constriction in mitotic chromosomes, first identified by Walter Flemming in 1882. Centromeres are bound by
proteinaceous complex called the kinetochore, which connects centromeres to microtubules for chromosome segregation (Luykx 1965). Centromeres are specified through a distinct histone H3 variant called centromere protein A (CENP-A) (Earnshaw and Rothfield 1985; Palmer et al. 1987). The organization of CENP-A nucleosomes and sequences within centromeric and pericentric regions vary greatly between species and even between chromosomes of the same species (Henikoff et al. 2001), implying that binding of these regions by proteins is not sequence-specific and might be epigenetically regulated (Saksouk et al. 2015). Centromeric chromatin can be generally divided into two regions: CENP-A containing centromere core and heterochromatin-associated pericentric region (Schalch and Steiner 2017). Centromeric and pericentromeric regions are comprised of tandem repeats that are often called satellite DNA. Mouse centromeres and pericentromeres are comprised of 123-bp minor satellite and 234-bp major satellite DNA, respectively (Vissel and Choo 1989; Kipling et al. 1991). Murine pericentric heterochromatin can be easily visualized as DAPI-dense chromocenters, unlike that of primates (Nishibuchi and Déjardin 2017). Centromeric and pericentric regions can be tens of megabases long, and it is estimated that repeats in these regions comprise up to 10% of the genome (Saksouk et al. 2015).

Heterochromatin serves as a guardian of the genome, by keeping repeats in centromeres and telomeres silent, preventing recombination and maintaining stable replication of these repetitive regions (Janssen et al. 2018).
Facultative heterochromatin

Facultative heterochromatin is defined as transcriptionally inactive regions of chromatin that decondense and become transcriptionally active at specific developmental stages, or during nuclear relocalization, or in a parental/heritable context such as monoallelic gene expression (Trojer and Reinberg 2007). Examples of facultative heterochromatic regions are inactive X (Xi), homeobox gene (Hox) clusters, developmental genes and autosomal imprinted loci. These regions are often hypoacetylated and enriched for macroH2A and H4K20me3 marks. Notably, Polycomb group (PcG) proteins play important role in facultative heterochromatin formation, and H3K27me3 PTM is considered a signature mark of facultative heterochromatin (Trojer and Reinberg 2007).

Polycomb Repressive Complex 2 (PRC2) subunit enhancer of zeste 2 (EZH2) catalyzes di/trimethylation of H3K27, and a chromodomain protein in Polycomb Repressive Complex 1 (PRC1) recognizes H3K27me3 (Cao et al. 2002). Along with EZH2, embryonic ectoderm development (EED) and suppressor of zeste 12 (SUZ12) comprise the core complex of PRC2. EED has been shown to bind H3K27me3 and facilitate methyltransferase activity of PRC2 and self-propagation of repressive H3K27me3 mark (Hansen et al. 2008; Margueron et al. 2009). H3K27me3 mark can block the deposition of an activating mark H3K27ac and hinder RNA Polymerase II (RNA Pol II) recruitment (Chopra et al. 2011).
PRC1 subunit RING1B facilitates chromatin compaction (Eskeland et al. 2010) and also monoubiquitinates H2AK119, which is thought to lead to transcriptional repression (Wang et al. 2004). However, the role of H2AK119ub (H2AK118 in Drosophila) in PcG-mediated gene repression is debated, since this PTM is widespread but only partially associated with PRC1 (Lee et al. 2015) and catalytically inactive mutation of Drosophila RING1B homolog did not lead to derepression of PcG target genes (Pengelly et al. 2015). PRC1 has been shown to inhibit the acetyltransferase activity of CREB-binding protein (CBP) towards H3K27, thereby preventing activation of enhancers and promoters (Tie et al. 2016).

Mechanisms of H3K27me3 mark deposition by PRC2 in mammalian cells still remain unclear. One of the existing models suggests that local chromatin environment determines the placement of this mark: presence of constitutive heterochromatin-associated H3K9me3 and HP1, as well as active histone marks H3K36me2/3 and H3K4me3 (except in bivalent domains) and nascent transcripts prevent H3K27me3 deposition (Wiles and Selker 2017). Conversely, presence of PRC1 and H2AK119ub1 was shown to recruit PRC2 (Blackledge et al. 2014; Cooper et al. 2014).

PcG proteins aggregate and form nuclear foci, which are called PcG bodies (Pirrotta and Li 2012). These PcG bodies are clusters of PcG targets that interact through chromatin looping (Entrevan et al. 2016). Studies of long-range PcG chromatin interactions in mouse embryonic stem cells (mESCs) showed that
EED is important for these interactions (Denholtz et al. 2013) and RING1B plays a crucial role in interactions between promoters of PcG targets (Schoenfelder et al. 2015). Hence, PcG proteins affect not only local chromatin environment, but also mediate looping interactions between regulatory elements and shape nuclear organization (Entrevan et al. 2016).

**Spatial heterochromatin localization**

In addition to TADs, the contacts between the genome and nuclear subcompartments represent another layer of genome organization. Heterochromatin is spatially concentrated in two main regions in the nucleus: 1) the perinucleolar region, or nucleolus-associated domains (NADs); and 2) the nuclear periphery, or lamina-associated domains (LADs) (Politz et al. 2016). I will review these regions separately, with a special focus on NADs.

**NADs**

NADs are genomic regions often found associated with nucleolar periphery in interphase cells (Németh et al. 2010; van Koningsbruggen et al. 2010). The nucleolus is the largest nuclear substructure, best known for its ribosome biogenesis function. Ribosomal DNA (rDNA) tandem repeats are found in nucleolus organizer regions (NORs), which cluster and give rise to the nucleolus (Heitz 1931; McClintock 1934). NORs are located on centromeric regions of mouse chromosomes 12, 15, 16, 18 and 19, although the number and
chromosome locations of NORs vary in different mouse strains (Dev et al. 1977; Kurihara et al. 1994). Of note, mouse chromosomes are acrocentric, hence, the mouse centromeres are found near the telomeric end of chromosomes. rDNA repeats consist of units formed by approximately 14 kb of RNA Pol I-transcribed 45S rDNA region and about 30 kb of intergenic spacer (IGS) region (Grummt 2010). 45S pre-rRNA is processed into 18S, 5.8S and 28S rRNA, which, together with 5S RNA that is transcribed elsewhere, constitute the RNA backbone of the ribosome (Grummt 2010). The nucleolus lacks a membrane and is composed of three layers: a large granular component (GC), with one or a few dense fibrillar components (DFC), each of which possesses a fibrillar center (FC) (Pederson 2011) (Fig. 1.2). Nucleolar protein nucleophosmin (NPM-1) is found in the GC, whereas fibrillarin (FBL) is localized in the DFC (Boisvert et al. 2007). Transcription of 45S pre-rRNA occurs in the FC or at the border between FC and DFC. Processing of rRNA happens in the DFC, and the first steps of ribosome assembly occur in the GC (Németh and Längst 2011). The rDNA outside of the nucleolus is usually inactive, whereas the active and poised rDNA genes are found within the nucleolus, close to the DFC and FC (Németh and Längst 2011). Nucleoli disassemble when cells enter mitosis and begin reassembling in early telophase (Boisvert et al. 2007).
Figure 1.2: Schematic diagram of nucleolar tripartite structure.

Nucleoli possess tripartite structure consisting of Fibrillar Center (FC), Dense Fibrillar Component (DFC) and Granular Component (GC). Ribosomal DNA transcription occurs in the FC or at the interface between FC and DFC, and processing of ribosomal RNA takes place in the DFC. Ribosomal subunit assembly occurs in the GC.

Prior to the generation of genome-wide NAD maps, it had been observed that genomic regions besides rDNA repeat-containing NORs associate with nucleolar periphery. For example, centromeres of human chromosomes 1 and 9 (Stahl et al. 1976), transfer RNA (tRNA) gene families in *Saccharomyces cerevisiae* (Thompson et al. 2003), mammalian Xi (Zhang et al. 2007) and 5S rDNA (Fedoriw et al. 2012b) have been known to localize to nucleolar periphery. Repeat elements in telomeres of *S. cerevisiae* were shown via 3C-based method to associate frequently with rDNA repeats and this has been suggested to limit the movement of chromosomes and define the nuclear chromatin architecture.
(O’Sullivan et al. 2006). Similarly, chromatin in human cells was reported to be limited in mobility at nucleolar and nuclear peripheries (Chubb et al. 2002). Perinucleolar heterochromatin (PNH) contains centromeres and pericentric heterochromatin, as well as facultative heterochromatin from NOR and non-NOR bearing chromosomes (McStay and Grummt 2008; Politz et al. 2013).

Two seminal studies identified the first genome-wide map of NADs in human cells (Németh et al. 2010; van Koningsbruggen et al. 2010). The Langst group used formaldehyde crosslinked HeLa cells, whereas the Lamond group utilized non-crosslinked HT1080 fibrosarcoma cells to biochemically purify nucleolus-associated DNA; microarray-based (Langst group) or deep sequencing (Lamond group) approach was used to obtain NAD maps. Both studies reported significant enrichment of gene-poor and transcriptionally silent regions, including satellite repeats, in NADs. However, repeats had to be removed from the NADs, due to the inability to locate these repeats to specific chromosomes during the analysis of NADs (van Koningsbruggen et al. 2010). Additionally, both studies showed that zinc finger genes, olfactory receptors, immunoglobulins and 5S rDNA were highly enriched in NADs. NADs were found to be enriched for repressive histone marks H3K9me3, H3K27me3 and H4K20me3 (Nemeth et al. 2010). The genome coverage of NADs in these studies was less than 5%. The Langst group showed that tRNA genes were highly represented in NADs, and NADs were distinct from LADs (Németh et al. 2010), whereas the Lamond group did not identify tRNA genes in NADs and showed clear overlap between NADs and LADs using whole-
genome sequencing and live cell time lapse fluorescence microscopy (van Koningsbruggen et al. 2010).

A more recent study from the Nemeth group used diploid human primary cell line, specifically, human IMR90 embryonic fibroblast to obtain higher-resolution maps of NADs in young proliferating and old senescent cells (Dillinger et al. 2017). The NADs in proliferating cells cover approximately 38% of human non-repetitive genome, and the authors noted that the lower genome coverage in their previous study (Németh et al. 2010) was due to using higher threshold value. The median length of IMR90 NADs was 361 kb. These NADs were reported to be late replicating regions, characterized by “heterochromatin” and “quiescent/low” chromatin Multivariate Hidden Markov Model (ChromHMM) states, which are distinguished by high levels of DNA methylation, low DNaseI accessibility, low gene density, and low gene expression. Indeed, the authors noted that 74% of NADs are found in B2/B3-type constitutive heterochromatic compartments, and 15% of NADs reside in B1-type facultative heterochromatic compartments as determined by Hi-C experiments. Surprisingly, the senescent IMR90 cells yielded NAD maps highly similar to young IMR90 cells. Most of the changes in NADs between the young and senescent cells had a median size less than 20 kb, hence, they often involved parts of individual NADs. The NADs unique to young or senescent cells were enriched in protein-coding genes. The loss of nucleolar association in either young or senescent cells correlated with higher gene expression; conversely, the gain of NAD status correlated with
decreased gene expression. Importantly, the Nemeth group observed that the satellite repeats at centromeric and pericentric regions had a decreased association with nucleolar periphery in senescent cells. These repetitive genomic regions are not annotated; hence the changes were determined via 3D immuno-FISH assay, but not via microarray experiments which the authors employed to map the NADs. Quantitative immunofluorescence experiments showed that H3K9me3 signal intensity decreased at perinucleolar space in senescent cells, which the authors concluded to be due to senescence-associated distention of satellites. Dillinger et al. speculate that the nucleolus safeguards the maintenance of NADs and 3D genome organization in senescence (Dillinger et al. 2017).

NADs were also sequenced in plant *Arabidopsis thaliana*, where they comprised 4.2% of the genome (Pontvianne et al. 2016). Pseudogenes and tRNA genes were enriched in *A. thaliana* NADs. Subtelomeric regions of all five chromosome pairs, and the entire short arm of an active NOR-bearing chromosome 4 were found to be part of NADs. As in human cells, plant NADs were enriched in heterochromatic marks H3K27me3 and H3K9me2 and depleted from actively transcribed genes.

The first genome-wide map of NADs in mouse cells has recently been published by our group (Vertii et al. 2019). The NAD maps were obtained via deep sequencing of biochemically purified nucleolus-associated DNA (NAD-seq) from formaldehyde crosslinked and non-crosslinked mouse embryonic fibroblasts.
(MEFs). The genome coverages of crosslinked and non-crosslinked NADs were 41% and 30%, respectively, where almost all peaks in the non-crosslinked data set were present in the crosslinked data set. The authors developed a software called NADfinder that uses local background correction to accurately call NAD peaks. NADfinder was more reliable in identifying peaks at chromosome ends distal from the centromere, or distinguishing peaks and valleys on small chromosomes that were otherwise annotated as almost entirely nucleolus-associated by other software packages. MEF NADs exhibited heterochromatic features: they often overlapped with MEF LADs (Peric-Hupkes et al. 2010), late replicating regions in MEFs (Hiratani et al. 2010) and were enriched in heterochromatic marks H3K9me3 and H3K27me3 and were mostly comprised of regions with low gene expression levels. Importantly, Vertii et al. (2019) identified distinct subsets of NADs: Type I NADs-regions that associate with both nucleolar and nuclear peripheries; and Type II NADs-regions that are found to be associated with nucleolar periphery, but not nuclear periphery. Type II NADs are more enriched in facultative heterochromatic mark H3K27me3, rather than in constitutive heterochromatic mark H3K9me3; are often early replicating and have higher gene density and gene expression levels than Type I NADs (Vertii et al. 2019).

Very recently, the Shen group mapped NADs in cJ9 (129 strain) mESCs using non-crosslinked nucleoli isolation protocol (Lu et al. 2020). NADs in their data set comprised only ~7.5% of the mouse genome, which is much smaller
than the genome coverage of NADs in MEFs (Vertii et al. 2019). Intriguingly, Lu et al. (2020) identified 1.6- to 2.3-fold enrichment of long interspersed nuclear element-1 (LINE-1 or L1) repeats in NADs and LADs compared with random genomic regions. Moreover, 59% of genes enriched in L1 in regulatory regions (“L1-enriched regions”) were found in NADs and/or LADs, suggesting that repeats can play a role in nuclear localization (Lu et al. 2020).

**Cis and trans-acting factors and mechanisms of nucleolar association**

There are multiple studies that implicate various cis and trans-acting factors in tethering genomic regions to the perinucleolar space. It is likely that no single mechanism of perinucleolar association exists, since the factors described below act in a context-dependent manner and have other cellular functions besides the nucleolar tethering (e.g. CTCF, CAF-1).

**CTCF** is a DNA-binding protein that binds insulator elements in vertebrates and aids in blocking an enhancer of one gene from activating a promoter of another gene (West et al. 2002). This insulating activity of CTCF is achieved through loop formation, described in the “Chromatin organization” section of this chapter. The Felsenfeld group identified that CTCF co-purifies with NPM-1 in HeLa cells; they expressed transgenes with insulator elements in K562 human leukemia cells, where these insulators preferentially localized to nucleolar periphery (Yusufzai et al. 2004). CTCF and NPM-1 were bound at these
elements, depending on the presence of intact CTCF binding site (Yusufzai et al. 2004).

**Nucleophosmin** (NPM-1) is a nucleolar histone chaperone that plays important role in various cellular processes, such as ribosome biogenesis, chromatin remodeling, DNA damage response, etc. (Box et al. 2016). Its *Drosophila* homolog, nucleoplasmin-like protein (NLP), along with CTCF and Modulo (*Drosophila* homolog of nucleolin, a nucleolar protein) have been demonstrated to mediate centromere clustering around the nucleolar periphery (Padeken et al. 2013). Depletion of either NLP, or Modulo, or CTCF led to de-clustering of centromeres and relocalization of heterochromatin away from the nucleolar periphery. This was accompanied by derepression of centromeric repeats and mitotic defects, such as lagging chromosomes and anaphase bridges (Padeken et al. 2013). Another study implicated NPM-1 in nucleolar chromatin organization through experiments in which NPM-1-depleted human and mouse cells showed altered nucleolar morphology and decreased levels of H3K9me3 and HP1γ foci at perinucleolar space (Holmberg Olausson et al. 2014). The authors of this study suggested that NPM-1 is important in tethering HP1γ to the nucleolus. These studies highlight the challenges of identifying factors responsible for nucleolar association, as depletion of proteins such as nucleolin and NPM-1 lead to changes in nucleolar morphology, heterochromatin organization and mitotic defects, hence it is difficult to discern direct, indirect and specific functions of these factors.
CAF-1 (chromatin assembly factor-1) is a histone chaperone complex that deposits newly synthesized H3/H4 tetramers onto replicating DNA (Kaufman et al. 1995). A study from the Kaufman group determined that depletion of CAF-1 subunit p150 led to relocalization of nucleolar proteins NPM-1, Ki-67, nucleolar phosphoprotein 140 (Nopp140), upstream binding factor (UBF), transcription termination factor 1 (TTF1) and nucleolin in human cells (Smith et al. 2014). Notably, upon p150-depletion, the authors observed decreased nucleolar association of select NAD regions identified by the previous study (Nemeth et al. 2010): D4Z4 repeats on telomere of chromosome 10q, 5S rDNA array, and centromeric satellite DNA (α-satellite) (Smith et al. 2014).

The cell proliferation marker Ki-67 depletion has been shown to result in decreased nucleolar association of CENP-A and reduced nucleolar staining of H3K9me3 and H4K20me3 in human and mouse cells (Sobecki et al. 2016). Ki-67 depletion has also been shown to lead to decreased association of α-satellite DNA with nucleolar periphery in HeLa cells (Matheson and Kaufman 2017). Xi association with nucleolar periphery during S-phase was decreased in Ki-67 depleted cells, which was accompanied by an increase of Xi association with nuclear lamina (Sun et al. 2017). Our group observed a 2-hr delay in Xi-nucleolus association in Ki-67 depleted cells, which was the same delay detected in S-phase entry as a result of Ki-67 depletion (Sun et al. 2017). These studies suggest a link between the cell cycle progression, heterochromatin organization and nucleolar association.
A recent study implicated a ribonucleoprotein complex **MiCEE**, which contains *Mirlet7d* microRNA (miRNA), and components of exosome and PRC2, in nucleolar tethering and silencing of bidirectionally expressed genes (Singh et al. 2018). The authors showed that there is a substantial overlap (49%) between binding sites of the *let-7* miRNA family member *Mirlet7d* and binding sites of H3K27me3 mark and nucleolar marker named UBF. The loss of either *Mirlet7d*, or the exosome complex, or PRC2 complex components (EZH2, SUZ12, and EED), led to significantly reduced staining of H3K27me3 mark at PNH and derepression of *Mirlet7d* targets in mouse and human cells. Notably, the deletion of a *Mirlet7d* target non-coding RNA (ncRNA) resulted in the relocalization of the rest of this locus away from the nucleolar periphery to the nuclear periphery (Singh et al. 2018). This study suggests that RNA, specifically, ncRNAs can initiate nucleolar tethering of specific loci.

Indeed, **PRC2** has been shown to contribute to the NAD localization, as its depletion by chemical inhibitors led to decreased nucleolar association of Type I & Type II NADs (Vertii et al. 2019). However, PRC2 is unlikely to be a nucleolar-specific *trans*-acting factor, as its depletion also led to decreased lamina association of Type I NADs (Vertii et al., 2019), and knockdown of EZH2 resulted in disrupted lamina association of lamina-associated sequences (LASs) (Harr et al. 2015).

Endogenous **5S rDNA** array frequently localizes to the nucleolar periphery in mESCs, and insertions of reporter constructs with 5S rRNA gene into these
cells resulted in frequent nucleolar association and transcriptional silencing of the reporter transgene (Fedoriw et al. 2012b). Surprisingly, Hi-C map of 5S and 45S rDNA array contacts did not reveal direct 5S-45S rDNA interactions, although they shared many common interaction sites (Yu and Lemos 2018).

NAD enrichment for centromeric repeats suggest that they have an important role in the PNH association (Németh et al. 2010; Németh and Längst 2011). For example, centromeric RNA facilitates nucleolar association of centromeric proteins INCENP and CENPC1 in interphase, which are then released into the nucleoplasm for kinetochore assembly in mitotic human cells (Wong et al. 2007).

A study from the Shen group suggests that L1 repeats, which belong to non-long terminal repeat (non-LTR) retrotransposons, can facilitate localization of L1-enriched genes at nuclear and nucleolar peripheries (Lu et al. 2020). Specifically, depletion of L1 RNA transcripts resulted in relocalization of L1-rich DNA sequences away from these heterochromatic regions. Additionally, both knockdown of L1 RNA transcripts and depletion of nucleolin led to upregulation of L1-enriched genes, suggesting that L1 transcripts interact with nucleolar and nuclear lamina proteins to tether L1-rich DNA sequences to repressive domains (Lu et al. 2020).

The nucleolar association of Xi is facilitated by the activity of long non-coding RNAs (IncRNAs), a type of ncRNA that is longer than 200 nucleotides (nt). During random X chromosome inactivation (XCI) in early female early
embryos, the X inactivation center (Xic) locus produces lncRNA named X inactive specific transcript (Xist), which binds in cis and initiates silencing of the X chromosome, which becomes Xi (Brown et al. 1991; Brockdorff et al. 1991; Clemson et al. 1996). Xist has been shown to target Xi to the perinucleolar space during S-phase in mouse cells (Zhang et al. 2007). In this study, the loss of Xist led to the detachment of Xi from the nucleolar periphery, loss of H3K27me3 mark on Xi, and reactivation of genes on Xi in some of the tested subclones. The authors suggested that nucleolar association of Xi maintains and replicates its repressive chromatin states. Interestingly, Zhang et al. found that the nucleolar periphery is enriched for Snf2h, the catalytic subunit of ACF1-ISWI that is needed for the DNA replication fork progression through heterochromatin (Collins et al. 2002). Hence, it is possible that apart from acting as a silencing compartment, nucleolar periphery maintains the fidelity of heterochromatin replication (Zhang et al. 2007).

Another lncRNA named Firre escapes XCI and is found a long distance (54 Mb) away from the Xic (Yang et al. 2010). Of note, XCI escapers were among the NADs in human IMR90 embryonic fibroblasts (Dillinger et al. 2017). Firre locus is a macrosatellite repeat (repeating unit is several kb in length), that showed enrichment for CTCF and cohesin binding on the Xi (Yang et al. 2015). Additionally, CTCF was shown to bind Firre lncRNA. The authors noted the more frequent nucleolar association of Firre lncRNA in female than male somatic cells, which was mediated by Firre expressed from the Xi chromosome. Knockdown of
either Firre, or Ctcf led to reduced nucleolar association of the Xi in mouse fibroblasts. Also, Firre knockdown, and to a lesser extent, Ctcf knockdown led to reduced H3K27me3 enrichment on the Xi. However, Firre knockdown did not reactivate X-linked genes, which suggests there are additional mechanisms that prevent their reactivation (Yang et al. 2015). Firre depletion did not affect Xist IncRNA levels, or the shape of Xist RNA clouds, suggesting that Firre acts independent of Xist. of Firre and Ctcf knockdown also led to the reduced nucleolar association of another X-linked IncRNA Dxz4 (Yang et al. 2015), but Dxz4 deletion did not result in changes in nucleolar association of Xi, or its H3K27me3 enrichment (Bonora et al. 2018). It is likely that the interactions between CTCF, cohesin, Firre IncRNA and possibly other unknown players result in nucleolar association of the Xi, independent of the role Xist IncRNA plays in tethering Xi to the nucleolus.

LncRNA Kcnq1ot1 has been shown to regulate the ~1 Mb Kcnq1-imprinted domain in mouse chromosome 7 by being paternally expressed and silencing neighboring genes in cis (Fitzpatrick et al. 2002; Thakur et al. 2004; Mancini-DiNardo et al. 2006). Insertion of a silencing domain found at the 5’ end of Kcnq1ot1 transcript into an episomal vector led to the nucleolar association of this vector and silencing of a flanking reporter gene (Mohammad et al. 2008). Kcnq1ot1 domain was more enriched for H3K9me3 and H3K27me3 marks, as well as for G9a and PRC2 occupancy in mouse placenta than in fetal livers cells (Pandey et al. 2008). This correlated with increased nucleolar association of
*Kcnq1* domain observed in placenta, but not in liver cells, which suggested that nucleolar localization is linked to heterochromatin establishment at paternal *Kcnq1* domain (Pandey et al. 2008). However, the study by Magnuson group showed that EED knockout mouse trophoblast stem cells still exhibited frequent nucleolar association of *Kcnq1ot1* lncRNA, regardless of the derepression of *Kcnq1*-imprinted domain genes in these PRC2-deficient cells (Fedoriw et al. 2012a). Hence, the nucleolar localization is not always coupled with transcriptional repression, at least in specific cell types that Fedoriw et al. investigated.

Recent studies have proposed **phase separation** as a mechanism for membraneless nuclear organelle formation and function. Nucleoli have been demonstrated to form as a result of liquid-liquid phase separation (Brangwynne et al. 2011), and NPM-1 and FBL were shown to form immiscible phases within nucleoli *in vitro*, recapitulating the distinct tripartite structure of nucleoli *in vivo* (Feric et al. 2016). FBL and NPM-1 are proteins with intrinsically disordered regions (IDRs) that drive their liquid droplet formation, whereas RNA binding to these low complexity proteins drives them to their respective subcompartments (Feric et al. 2016; Mitrea et al. 2016). Another study used temperature dependence and reversibility parameters to discern two distinct mechanisms of nucleolar protein assembly: IDR-driven phase separation of FBL and Nopp140, and active recruitment of nucleolin homolog Modulo and Nucleostemin 1 via rDNA (Falahati and Wieschaus 2017). However, other *Drosophila* nucleolar
protein RNA Polymerase I Subunit 135 (Rpl135) showed characteristics of both mechanisms: it was presumably initially actively recruited to the nucleolus, after which it was stabilized through a secondary mechanism, perhaps through association with RNA transcripts (Falahati and Wieschaus 2017).

Analogous to the mechanisms of nucleolar protein assembly described above, genomic regions might also be driven to associate with nucleolar periphery through similar mechanisms, such as phase separation, active recruitment (e.g. mediated by CTCF-NPM-1-Firre lncRNA complex, or 5S rDNA array), and a combination of these two mechanisms. It is also conceivable that nucleolar formation processes could facilitate nucleolar association events. For example, aluRNAs transcribed by Pol II from introns of Alu repeats have been shown to mediate nucleolar structure and nucleolar association through their specific interactions with nucleolin and NPM-1, possibly through a phase separation mechanism (Caudron-Herger et al. 2015). Indeed, hexanediol treatment, which is known to disrupt liquid-liquid phase separation, was shown to reduce the nucleolar association levels of Type II NADs, and to a lesser degree, the nucleolar association of Type I NADs (Vertii et al. 2019).

Functional significance of NADs

**Silencing hub and chromatin organization.** Genome-wide maps of NADs (Németh et al. 2010; van Koningsbruggen et al. 2010; Pontvianne et al. 2016; Dillinger et al. 2017; Vertii et al. 2019) and studies of trans-factor mediated
chromatin or vector localizations at the nucleolar periphery (Zhang et al. 2007; Mohammad et al. 2008; Pandey et al. 2008; Fedoriw et al. 2012b; Yang et al. 2015; Singh et al. 2018) suggest that the perinucleolar space is a heterochromatic subcompartment. Comparison of NADs with Hi-C revealed that NADs correspond to B2/B3-type constitutive heterochromatin and B1-type facultative heterochromatin compartments (Dillinger et al. 2017). A recent study from the Guttman group also implicated the nucleolar periphery as a silencing hub (Quinodoz et al. 2018). Quinodoz et al. developed a technique called Split-Pool Recognition of Interactions by Tag Extension (SPRITE) that measures multiple genomic interactions that take place simultaneously in the nucleus. This study determined that both mouse and human cells contain two discrete hubs where the genomic regions interact at higher frequencies. The “active” hub that was enriched for transcriptionally active genes was found to be around nuclear speckles, whereas the “silencing” hub enriched for inactive genes was located around the nucleolus. Interestingly, genomic DNA showed preference either for nucleolar periphery, or nuclear speckles, i.e. regions found in these two hubs were mutually exclusive (Quinodoz et al. 2018).

A recent study from the Dekker group used liquid Hi-C technique to determine the strength of chromatin interactions in a genome-wide manner (Belaghzal et al. 2019). This technique involves extensive chromatin fragmentation via restriction enzyme digestion in situ in the nucleus, followed by fixation and Hi-C analysis to identify the kinetics of chromatin dissociation.
Regions around the nucleoli exhibited one of the most stable interactions based on half-life of chromatin interactions upon digestion, with heterochromatin in general showing the most stable chromatin associations in the genome (Belaghzal et al. 2019). This is in agreement with the hypothesis that stable interactions in heterochromatin drive the A-B compartmentalization (Falk et al. 2019).

The SPRITE and liquid Hi-C data support the notion that the nucleolar periphery serves as heterochromatin docking site, thereby limiting chromatin movement and influencing chromatin organization, as it has been shown in human and yeast cells (Chubb et al. 2002; O'Sullivan et al. 2006).

**Developmental regulation.** Mouse nucleolar-specific NADs (Type II NADs) are enriched for developmentally regulated genes, e.g. Gene Ontology (GO) terms organ morphogenesis and sensory organ development are among the most enriched in genes within Type II NADs (Vertii et al. 2019), suggesting that nucleolar association could promote stage-specific silencing of genes, i.e. harbor facultative heterochromatin. Indeed, Vertii et al. (2019) showed that Type II NADs are enriched for facultative heterochromatin mark H3K27me3.

A study on IncRNA *Firre* demonstrated that CTCF binds *Firre* locus similarly in female and male mESC before the onset of random XCI; but after the occurrence of random XCI, CTCF was preferentially enriched in female differentiated cells, consistent with the finding that CTCF is enriched on the Xi and lost on the active X (Xa) (Yang et al. 2015). This correlated with increased
nucleolar association of *Firre* in female cells, where *Firre*/CTCF promote nucleolar association of Xi (Yang et al. 2015). In this case, XCI occurs first, after which it is found to be frequently associated with nucleolar periphery, hence in this case, silencing precedes nucleolar localization. In this specific case, it can be hypothesized that nucleolar periphery serves to maintain the inactive status of Xi, not to initiate it. However, the mechanisms of recruitment and silencing of developmentally regulated genes at the nucleolar periphery remain to be determined.

**Genome stability.** Studies have shown that depletion of *Su(var)3-9*, an H3K9 methyltransferase in *Drosophila* led to a dispersal of satellite and rDNA repeats and appearance of multiple nucleoli instead of one nucleolus that is usually found in fruit flies (Peng and Karpen 2007). In another study, *Su(var)3-9* mutant *Drosophila* cells showed increased DNA damage in heterochromatin and chromosomal defects, such as loss of heterozygocity and hypo-condensation (Peng and Karpen 2009). Depletion of either NLP (*Drosophila* homolog of NPM-1), or CTCF, or centromeric H3 resulted in reduced centromere clustering near the nucleoli and derepression of repeats found in pericentric heterochromatin (Padeken et al. 2013). Notably, NLP depletion was accompanied by increased DNA damage and mitotic defects, such as lagging chromosomes and anaphase bridges (Padeken et al. 2013), resembling the defects observed in *Su(var)3-9* mutants (Peng and Karpen 2009). A study of NADs in senescent human fibroblasts showed decreased association of centromeric repeats with the
nucleoli, and reduced H3K9me3 staining at the nucleolar periphery, which is likely to be due to senescence-associated distension of satellites (Dillinger et al. 2017). It is conceivable that the centromere and pericentromeric region clustering at the PNH facilitates silencing and prevents recombination of repeats found in these regions. Hence, the loss of nucleolar association could lead to derepression and recombination of the repeats, resulting in genome instability and mitotic defects.

LADs

LADs are genomic regions that frequently associate with the nuclear lamina (NL) (Kind and van Steensel 2010). NL is comprised of a meshwork of proteins associated with the inner nuclear membrane (INM) and attached to chromatin. NL is mainly composed of type V intermediate filaments- the nuclear lamins (Dechat et al. 2008). Mammalian lamins are: lamin B1 and B2 encoded by LMNB1 and LMNB2 genes, respectively, and lamin A and C, derived from one gene LMNA by alternative splicing (Burke and Stewart 2013). LADs were mapped in various species and cell types, such as Caenorhabditis elegans, Drosophila melanogaster, mouse and human cells (Pickersgill et al. 2006; Guelen et al. 2008; Ikegami et al. 2010; Peric-Hupkes et al. 2010; Kind et al. 2015; Borsos et al. 2019). Most of the LAD studies use a technique called DNA adenine methyltransferase identification (DamID), which utilizes bacterial adenine methyltransferase (Dam) that is tethered to a NL protein such as Lamin
B, which results in adenine methylation of genomic regions that interact with the NL protein (Pickersgill et al. 2006). The advent of this method helped to obtain genome-wide maps of LADs, as well as to visualize LADs by microscopy via the use of GFP-tagged protein that recognizes the adenine methylated regions, i.e. m6A-Tracer technology built upon DamID (Kind et al. 2013).

Mammalian LADs are typically 10 kb-10 Mb long, with median size 0.5 Mb, and 40% genome coverage (Guelen et al. 2008; Peric-Hupkes et al. 2010). LADs have low gene density, and genes in these domains are mostly late-replicating and transcriptionally inactive (Guelen et al. 2008; Peric-Hupkes et al. 2010; Pope et al. 2014). LADs contain pericentric heterochromatin and a subset of telomeric regions (Guelen et al. 2008); LADs are enriched for heterochromatic marks H3K9me2/3 and H3K27me3 (Guelen et al. 2008; Wen et al. 2009; Harr et al. 2015; Kind et al. 2015). Studies in mammalian cell types showed that all of them invariably possess a subset of LADs called constitutive LADs (cLADs), whereas some cells have cell type-specific LADs called facultative LADs (fLADs) (Peric-Hupkes et al. 2010; Meuleman et al. 2013). Also, regions that are not found to be lamina-associated in the investigated cell types are called constitutive interLADs (ciLADs) (Peric-Hupkes et al. 2010; Meuleman et al. 2013). cLADs are the most gene-poor subset of LADs, they are enriched for AT-rich DNA segments and LINEs (Meuleman et al. 2013). Since cLAD genomic positions and sizes are strongly conserved in mouse and human, cLADs are
suggested to be a “structural backbone” of chromatin organization and folding in interphase nuclei (Meuleman et al. 2013; van Steensel and Belmont 2017).

NL is considered to be a repressive nuclear compartment (van Steensel and Belmont 2017; Lochs et al. 2019). During the differentiation of mESCs to neural precursor cells (NPCs) and astrocytes, many genes detached from the NL, which was frequently accompanied by gene activation, and vice versa, regions that increased their frequency of NL interactions often showed decreased gene expression (Peric-Hupkes et al. 2010). The NL-tethering of a hygromycin-LacO reported gene in mouse fibroblasts using Lacl-Emerin (an INM protein) resulted in transcriptional repression of hygromycin gene and its decreased H4 acetylation (Reddy et al. 2008). However, a study that used LacO reporter with Lacl-Lamin B1 in human U2OS cells did not show changes in reporter gene expression upon its relocalization to the NL (Kumaran and Spector 2008). These differences could be due to the chromatin context of reporter integration sites or differential sensitivities of promoters to the LAD environment (Lochs et al. 2019).

A recent study that utilized two promoter transplantation strategies showed that LADs are generally repressive but also highly heterogeneous chromatin domains (Leemans et al. 2019). Leemans et al. demonstrated that many promoters become active when they were inserted into non-LAD (interLAD, or iLAD) positions. Additionally, there was a subset of promoters that were less sensitive to the repressive effect of LADs. These escaper promoters were locally detached from the NL despite them being located inside LADs, but they did not show
significant difference in their chromatin mark enrichment compared to other promoters that were repressed in LADs. The authors speculate that the insensitivity of these escaper promoters could be due to their recruitment of transcription factors (TFs) that are more efficient in resisting LAD repression (Leemans et al. 2019).

There are several models of how LADs can promote repression. LADs could inactivate genes by the virtue of repressive enzymes' activities in the NL, e.g. histone deacetylase 3 (HDAC3) activity (van Steensel and Belmont 2017). Emerin, a protein found in the NL, has been shown to bind and catalytically activate HDAC3, which then facilitates gene repression at the nuclear periphery (Demmerle et al. 2012). Another study showed that HDAC3 in a complex with transcriptional repressor cKrox and INM protein Lap2β binds to lamina-associated sequences (LASs) and promotes the localization to the NL and gene repression (Zullo et al. 2012). A different model suggests that genes in LADs are inactive due to being shielded from transcriptionally active nuclear compartments (van Steensel and Belmont 2017). A study from the Reddy group integrated Hi-C and DamID data and demonstrated that LADs correspond to an inactive B-compartment, and that LADs tend to self-associate and form a more compact organization compared to non-LADs (Luperchio et al. 2017). Hence, segregation of LADs from the more active A-compartment is a possible mechanism for limiting TF access to LADs (Buchwalter et al. 2019).
A study from the Torres-Padilla and Kind groups determined that LADs were already established in zygotes during early embryonic development, whereas TADs boundaries were not well-defined at this stage (Borsos et al. 2019). Thus, these results suggest that LAD establishment precedes TAD consolidation, and chromatin attachment at the NL can instruct the formation of chromatin higher-order organization in the early embryonic development (Borsos et al. 2019). Another study proposed a model, where the interactions between lamina and heterochromatin are necessary to obtain the conventional nuclear organization, i.e. when heterochromatin segregates at the nuclear and nucleolar peripheries (Falk et al. 2019). In the absence of lamina, the simulations predicted that heterochromatin would localize in the nuclear interior, forming an inverted nucleus (Falk et al. 2019). Hence, it is likely that NL serves as a scaffold that determines the heterochromatin organization and orientation.

**Cis and trans-acting factors mediating the NL-LAD interactions**

Various studies have investigated the proteins and genomic sequences that are responsible for NL-LAD interactions. Lamins were shown to be important for NL-tethering in *Drosophila* and *C. elegans* (Mattout et al. 2011; Kohwi et al. 2013). However, triple knockout of Lamin B1, Lamin B2 and Lamin A/C in mESC did not result in dramatic changes in LADs according to Emerin-DamID (Amendola and van Steensel 2015). However, another study showed that the triple knockout of lamins led to changes in inter-TAD interactions,
decondensation of cLADs, and detachment of fLADs from the NL (Zheng et al. 2018), which implicates lamins in TAD and LAD organization. Lamin B receptor (LBR) is an integral membrane protein found in the INM, which together with lamin A/C was shown to mediate chromatin attachment to the nuclear periphery (Solovei et al. 2013). Solovei et al. demonstrated that the absence of both LBR and lamin A/C in mouse cells results in inverted nuclei, where the heterochromatin fuses in the nuclear interior. In another study, lamin A/C knockdown in mouse fibroblasts resulted in loss of the NL attachment of LADs, decompaction of LAD chromosome territories and mixing of LAD-iLAD regions, all of which could be discerned at the single-cell level via microscopy (Luperchio et al. 2017). However, these changes were not obvious at the population level, since DamID profile showed no changes from wild-type cells (Luperchio et al. 2017). These studies can be interpreted that lamins, LBR, emerin and possibly other INM proteins act redundantly to promote chromatin attachment to the NL (van Steensel and Belmont 2017).

H3K9me2-enriched domains largely overlap with LADs in mammalian cells (Wen et al. 2009; Kind et al. 2013). Several groups investigated whether H3K9me2 modification mediates the recruitment of chromatin to the NL. Indeed, G9a depletion resulted in decreased frequency of NL-LAD interactions (Kind et al. 2013), or even dissociation of LADs from the NL (Harr et al. 2015). H3K9me3 modification might also aid in recruiting LADs to the NL, as it was shown in C. elegans (Towbin et al. 2012). This study from the Gasser group demonstrated
that H3K9me1/2 promotes the NL attachment of repeat-rich chromosome arms, whereas H3K9me3 reinforces this attachment and establishes silencing at these regions. Another study that explored the epigenetic mechanisms of peripheral localization of human β-globin (HBB) bacterial artificial chromosome (BAC) in mouse fibroblasts showed that Suv39H1/2 knockdown, i.e. H3K9me3 depletion led to decreased association of HBB BAC with the NL (Bian et al. 2013). This study also showed that G9a knockdown, i.e. H3K9me2 depletion results in the lamina dissociation of the LAD region adjacent to the HBB locus. However, only combined G9a and Suv39H1/H2 knockdown led to the detachment of the entire HBB BAC with the adjacent LAD region (Bian et al. 2013).

H3K27me3 marks were shown to be enriched at LAD borders (Guelen et al. 2008). The Reddy group showed that inhibition of H3K27me3 either via RNAi knockdown or chemical inhibition led to relocalization of ectopically integrated LASs and endogenous LADs away from the lamina (Harr et al. 2015).

A previous study from the Reddy group identified LASs as genomic sequences in LADs that conferred NL association when inserted ectopically (Zullo et al. 2012). These LASs were enriched for GAGA motif, which is bound by cKrox in a complex with Lap2β and HDAC3; this complex was shown to promote NL attachment (Zullo et al. 2012). However, cKrox knockdown in mouse fibroblasts with integrated HBB BAC transgene (where GAGA motif clusters were present adjacent to HBB site) did not alter the peripheral localization of HBB (Bian et al. 2013). The Belmont group showed that one of three redundantly
acting peripheral targeting regions they identified in the HBB BAC confers
pericentric heterochromatin targeting to another BAC, hence this DNA sequence
targeting mechanism is likely to be epigenetic (Bian et al. 2013). It is likely that
multiple factors act to tether chromatin to the NL; some of them might act in a
cell-type or context-specific manner (e.g. LASs enriched for GAGA motif,
peripheral targeting cis-element in the HBB BAC).

**Overlap between NADs and LADs**

Genome-wide mapping of NADs in human and mouse cells revealed an
extensive overlap between NADs and LADs (Németh et al. 2010; van
Koningsbruggen et al. 2010; Dillinger et al. 2017; Vertii et al. 2019). The Lamond
group used a photoactivation and time-lapse fluorescent microscopy technique
and demonstrated that a subset of chromatin associated with nucleolar periphery
in mother HeLa cells relocalize to the nuclear periphery in daughter cells (van
Koningsbruggen et al. 2010). Another time-lapse microscopy assay of the m6A-
Tracer revealed that a subset of LADs do not re-associate with the NL, and
instead are found near nucleoli after cell division (Kind et al. 2013). In a follow-up
experiment, the van Steensel group demonstrated that Lamin A m6A-Tracer signal
accumulates at the nucleolar periphery in 35±7% of cells (Kind and van Steensel
2014). The authors proposed that the overlap between LADs and NADs are due
to Lamin A contacting the regions at the nuclear and nucleolar peripheries.
Additionally, it has been shown that Lamin B2 localizes at the nucleolar border in
close proximity to nucleolin, and depletion of Lamin B2 leads to disrupted nuclear and nucleolar morphologies (Sen Gupta and Sengupta 2017). Hence, it is possible that lamins tether regions to the NL and nucleolar periphery, resulting in the overlap between LADs and NADs.

There is also evidence of a redundancy in nuclear localization of mammalian heterochromatin. Ragoszy et al. showed that late-replicating regions are found at the NL, pericentric heterochromatin, and nucleolar periphery in human lymphoblastoid cells (Ragoczy et al. 2014). In general, late-replicating regions on large chromosomes were frequently found at the nuclear periphery, whereas late-replicating regions on small chromosomes showed a tendency to localize in perinucleolar space. When the Groudine group treated cells with actinomycin D, which inhibits RNA Pol I and disrupts nucleoli, the association of heterochromatic regions with the nucleolar periphery drastically decreased, whereas the frequency of association of these regions with the NL increased (Ragoczy et al. 2014). As mentioned previously, our study also showed that Ki-67 depleted cells showed reduced Xi association with the nucleolar periphery, and increased association of Xi with the nuclear periphery (Sun et al. 2017). Hence, when one of the heterochromatic nuclear subcompartments is disrupted, the heterochromatic regions tend to relocalize to other intact subcompartments. This suggests at least a partial redundancy of perinucleolar space, nuclear periphery and pericentric regions as subcompartments where heterochromatic regions can localize and maintain their silent state.
Indeed, about 70% of human NADs overlap with LADs (Dillinger et al. 2017). Similarly, around 66% of MEF NADs overlap with LADs (Vertii et al. 2019). Both nuclear and nucleolar peripheries contain centromeric and pericentric heterochromatin (Solovei et al. 2004), and both NADs and LADs are enriched for repetitive elements such as LINEs and LTR retrotransposons (Dillinger et al. 2017). However, the overlap between LADs and NADs is incomplete. The Nemeth group showed that about one third of human NADs are non-overlapping with LADs (Dillinger et al. 2017). Our group has determined that there is a subset of MEF NADs that is found only at the nucleolar periphery, i.e. Type II NADs (Vertii et al. 2019). The mechanisms and functional significance of nucleolar-specific association remain to be elucidated. One confounding factor is the difference in methodologies: LADs are determined via DamID, which reveals the transient interactions between a NL protein and chromatin that occur within a certain period, whereas NAD-seq captures a snapshot of chromatin associations with the nucleoli at a given time (Németh and Längst 2011).

**Features of heterochromatin in ESCs**

ESCs are pluripotent cells derived from the inner cell mass of blastocysts at the pre-implantation stage of embryos (Evans and Kaufman 1981). ESCs are characterized by their ability to self-renew and to differentiate into any of the three germ layers: ectoderm, mesoderm and endoderm (i.e. pluripotency) (O’Shea 2004). Chromatin in ESCs is thought to be less condensed and contain
sparse, disorganized heterochromatin compared to differentiated cells (Meshorer and Misteli 2006; Gaspar-Maia et al. 2011; Mattout et al. 2015; Atlasi and Stunnenberg 2017).

For example, studies showed that HP1 foci and H3K9me3 staining patterns were diffuse in mESCs, whereas in differentiated cells HP1 and H3K9me3-stained foci number increased and they were well-defined (Meshorer et al. 2006; Aoto et al. 2006). Electron spectroscopic imaging revealed the more compact state of heterochromatin and its localization at the nuclear periphery in lineage-committed cells and NPCs, whereas heterochromatin in mESCs is more dispersed (Hiratani et al. 2010; Ahmed et al. 2010). Fluorescence recovery after photobleaching (FRAP) experiments showed that chromatin binding dynamics of HP1α, linker histone H1 and core histones H2B and H3 in heterochromatic regions was faster in mESCs than in NPCs (Meshorer et al. 2006) or MEFs (Bhattacharya et al. 2009). Higher percentages of HP1, core and linker histones were extractable from mESCs than NPCs during biochemical salt extraction, which combined with FRAP results suggested that stem cells possess a subset of hyperdynamically bound chromatin proteins (Meshorer et al. 2006). Meshorer et al. hypothesized that this hyperdynamic chromatin fraction facilitates the open and plastic state of chromatin in mESCs, allowing the cells to adopt any lineage fate during differentiation, hence contributing to the pluripotency.

In agreement with observations of more sparse and dispersed heterochromatin in stem cells, studies in mouse and human ESC showed that
the abundance of heterochromatic marks H3K9me2/3 and H3K27me3 increased
during differentiation (Lee et al. 2004; Meshorer et al. 2006; Efroni et al. 2008;
Wen et al. 2009; Hawkins et al. 2010). Of note, histone demethylases Jmjd1a
and Jmjd2c demethylate H3K9me2 and H3K9me3, respectively, at promoters of
pluripotency-associated genes in mESCs (Loh et al. 2007). Knockdown of either
demethylase led to differentiation of stem cells, hence demethylation of H3K9 at
stem cell-specific genes can promote pluripotency of ESCs (Loh et al. 2007).
Wen et al. reported only about 4% of the genome of mESCs to contain
H3K9me2-modified regions, whereas the coverage of these H3K9me2 domains
increased to 31% in differentiated mESCs (Wen et al. 2009). Another study’s
findings were contrary: they found H3K9me2-modified regions to comprise over
50% of the mESC genome, which increases by 5% in neurons differentiated from
these mESCs (Lienert et al. 2011). Hence, the abundance of H3K9me2 mark in
mESCs is still debated. Interestingly, knockout of G9a or GLP abolished majority
of H3K9me2 marks in mESCs and had no effect on viability, whereas growth
defects were obvious in differentiated cells, and led to lethality in post-
implantation embryos (Tachibana et al. 2002, 2005). Therefore, it is likely that
repressive H3K9me2 marks are not essential for ESC survival and self-renewal.

The distribution of another repressive histone mark H3K27me3 has been
reported to change in mESCs cultured in feeder-free, serum-free culture medium
containing 2 inhibitors (2i): PD0325901, and CHIR99021 (Marks et al. 2012;
Galonska et al. 2015; McLaughlin et al. 2019). The combination of leukemia
inhibitory factor (LIF) and 2i in medium blocks the mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) and glycogen synthase kinase 3β (GSK3β) pathways and thereby promotes the pluripotency and self-renewal, i.e. the naïve ground state of mESCs (Ying et al. 2008; Silva et al. 2008). Surprisingly, high levels of H3K27me3 found at promoters of repressed genes in serum-grown mESCs are greatly reduced in serum-free 2i medium-grown mESCs, however, these genes do not become derepressed in the absence of H3K27me3 (Marks et al. 2012). Abrogation of PRC2 function through deletion of its critical subunits EED or SUZ12 did not affect the ESC self-renewal, but resulted in defective differentiation (Pasini et al. 2007; Chamberlain et al. 2008). PRC2 is essential in post-implantation embryos, as its deletion is embryonic lethal at the post-implantation stage (Faust et al. 1998; O’Carroll et al. 2001). The levels of H3K27me3 at bivalent promoters, where an activating mark H3K4me3 co-localizes with the repressive mark H3K27me3, was also reduced in serum-free 2i medium, but the bivalent genes did not become derepressed in 2i medium (Marks et al. 2012; Galonska et al. 2015). Hence, H3K27me3 mark does not seem to be critical for ESC pluripotency and self-renewal. In 2i-cultured mESCs, redistribution of H3K27me3 away from Polycomb targets, such as Hox loci, has been linked to DNA hypomethylation and chromatin decompaction of these loci, as well as to loss of long-range Polycomb interactions (McLaughlin et al. 2019). McLaughlin et al. showed that in 2i-grown mESCs, engineered to maintain Polycomb target compaction and long-range interactions, the
transcriptional network is not altered compared to wild type 2i-grown cells, thus
the epigenome does not instruct the transcriptional state of the ESCs.

The changes in late replicating regions, which are often heterochromatic
(Politz et al. 2013) also coincide with expansion and nuclear reorganization of
heterochromatin during cell differentiation (Mattout et al. 2015). Small domains
that replicate at different times in mESCs often became consolidated into one
large replication timing domain after these cells differentiated into NPCs (Hiratani
et al. 2008). Another study from the Gilbert group showed that the majority of
domains with early to late replication timing (EtoL) changes and repositioning
towards nuclear periphery occurs in embryoid body intermediates during mESC
differentiation into epiblasts and NPCs (Hiratani et al. 2010). The authors
hypothesized that EtoL changes are associated with the loss of pluripotency in
mESCs (Hiratani et al. 2010).

One of the results of open chromatin in ESCs seems to be global
pervasive transcription (Mattout et al. 2011). Repetitive regions that are normally
repressed in differentiated cells are expressed in mESCs, and even tissue-
specific genes are expressed at low levels (Efroni et al. 2008). The total mRNA
levels were almost two-fold higher in mESCs than in NPCs (Efroni et al. 2008).
Predictably, the levels of active marks such as H3K4me3, H3K9ac,H3K36me3
were higher in mESCs than in differentiated cells (Meshorer et al. 2006; Efroni et
al. 2008). Inhibition of histone deacetylation resulted in compromised
differentiation of mESCs (Lee et al. 2004) and promoted self-renewal of human
ESCs (Ware et al. 2009), revealing the significance of high histone acetylation levels in pluripotency and conversely, deacetylation as a required step for differentiation.

mESCs have been reported to contain mainly active rRNA, which lack CpG methylation and repressive histone marks H3K9me2, H3K9me3 and H3K27me3 at rDNA promoters (Savić et al. 2014). Promoter-associated RNA (pRNA) is a short, structured RNA encoded within the IGS region between rDNA repeats (Santoro et al. 2010). pRNA binding to the TIP5 protein subunit of the nucleolar remodeling complex (NoRC) activates NoRC function on rDNA repeats, leading to transcriptional silencing of rRNA (Mayer et al. 2006). Normally, mESCs do not express mature pRNA, however, ectopic expression of pRNA in stem cells results in the formation of electron-dense perinucleolar heterochromatin-like structures, as well as increased occurrence of H3K9me2/3 marks at repetitive regions within mESCs, similar to the nuclear heterochromatin observed in NPCs (Savic et al., 2014). Conversely, depletion of TIP5 impairs mESC differentiation (Savic et al., 2014). Therefore, NoRC-mediated rRNA silencing seems to promote not only nucleolar, but also nuclear heterochromatin condensation that is critical for exit from pluripotency.

Chromatin organization units such as compartments and TADs, as well as heterochromatic regions, such as LADs, late replicating regions, H3K9me2/3 and H3K27me3-enriched regions were studied and characterized in mESCs. However, NADs are a relatively new area of chromatin studies, and NADs in
mESCs have been poorly characterized. Our group’s work on MEF NADs
revealed that Type I NADs resemble constitutive heterochromatin, whereas Type
II NADs are similar to facultative heterochromatin (Vertii et al. 2019). Therefore,
studies of NADs in mESCs can unfold the distribution and features of
constitutive and facultative heterochromatin at the nucleolar periphery in
pluripotent cells. Our studies of mESC NADs, which I present in Chapter II,
would lay the groundwork for investigations of NAD changes during cell
differentiation and mammalian development.
CHAPTER II: DISTINCT FEATURE OF NUCLEOLUS-ASSOCIATED DOMAINS IN MOUSE EMBRYONIC STEM CELLS

Contributions

This chapter was published (Bizhanova et al. 2020) as a collaboration between the following authors: Bizhanova Aizhan, Aimin Yan, Jun Yu, Lihua Julie Zhu and Paul D. Kaufman. I performed all the wet lab experiments and analyzed and interpreted data. Aimin Yan, Jun Yu and Lihua Julie Zhu conducted all bioinformatic analyses. Paul Kaufman and I designed experiments, Paul Kaufman directed and analyzed the wet lab experimentation.

Abstract

Heterochromatin in eukaryotic interphase cells frequently localizes to the nucleolar periphery (nucleolus-associated domains (NADs)) and the nuclear lamina (lamina-associated domains (LADs)). Gene expression in somatic cell NADs is generally low, but NADs have not been characterized in mammalian stem cells. Here, we generated the first genome-wide map of NADs in mouse embryonic stem cells (mESCs) via deep sequencing of chromatin associated with biochemically purified nucleoli. As we had observed in mouse embryonic fibroblasts (MEFs), the large Type I subset of NADs overlaps with constitutive LADs and is enriched for features of constitutive heterochromatin, including late replication timing and low gene density and expression levels. Conversely, the
Type II NAD subset overlaps with loci that are not lamina-associated, but in mESCs, Type II NADs are much less abundant than in MEFs. mESC NADs are also much less enriched in H3K27me3 modified regions than are NADs in MEFs. Additionally, comparison of MEF and mESC NADs revealed enrichment of developmentally regulated genes in cell type-specific NADs. Together, these data indicate that NADs are a developmentally dynamic component of heterochromatin. These studies implicate association with the nucleolar periphery as a mechanism for developmentally regulated gene expression and will facilitate future studies of NADs during mESC differentiation.

Introduction

Eukaryotic genomes are broadly subdivided into more accessible, transcriptionally active euchromatin, and less accessible, less active heterochromatin. These functional classifications are accompanied by spatial separation: heterochromatin is mainly found at the nuclear periphery and nucleolar periphery, where they comprise nucleolus-associated domains (NADs) (Németh et al. 2010; van Koningsbruggen et al. 2010) and lamina-associated domains (LADs) (Pickersgill et al. 2006; Guelen et al. 2008; Peric-Hupkes et al. 2010), respectively. Studies in multiple organisms indicate that sequestration of heterochromatin to the nuclear and nucleolar peripheries contributes to gene silencing (Fedoriw et al. 2012b; Zullo et al. 2012; Jakociunas et al. 2013). Therefore, there is great interest in discovering the molecular bases for these
localizations. Notably, some trans-acting factors that specifically affect lamina (Zullo et al. 2012; Harr et al. 2015) or nucleolar (Yusufzai et al. 2004; Zhang et al. 2007; Mohammad et al. 2008; Padeken and Heun 2014; Smith et al. 2014; Singh et al. 2018) associations have been reported, suggesting that distinct mechanisms contribute at the two locations.

Both NADs and LADs are enriched for silent genes and histone modifications characteristic of constitutive heterochromatin, e.g., H3K9me2 and H3K9me3 (Matheson and Kaufman 2017; van Steensel and Belmont 2017). LADs have been mapped and studied in multiple species and cell types (Pickersgill et al. 2006; Guelen et al. 2008; Peric-Hupkes et al. 2010; Kind et al. 2013; Borsos et al. 2019). In contrast, NADs have been characterized in a few human somatic cell lines (Németh et al. 2010; van Koningsbruggen et al. 2010; Dillinger et al. 2017), in the plant Arabidopsis thaliana (Pontvianne et al. 2016), and recently, in mouse embryonic fibroblasts (MEFs) (Vertii et al. 2019). Several experiments indicate that LADs can be redistributed to the nucleolar periphery after passage through mitosis, and vice versa (van Koningsbruggen et al. 2010; Kind et al. 2013). However, the extent of overlap between LADs and NADs is unknown in most organisms and cell types.

Here, we mapped and characterized NADs in mouse embryonic stem cells (mESCs), a tractable system for studying how NADs change during differentiation. As in MEFs (Vertii et al. 2019), we identified a large subset of mESC NADs that overlap with LADs (Type I NADs) and a smaller subset of
NADs that do not overlap LADs (Type II NADs). However, Type II NADs are less prevalent in mESCs than in MEFs. mESC NADs are also notably less enriched in H3K27me3 modifications. Comparisons of MEF and mESC NADs also revealed enrichment of developmentally regulated genes in cell-type-specific NADs. These analyses will facilitate future studies of genome dynamics during stem cell differentiation.

Results

Isolation of nucleoli from crosslinked F121-9 mESCs

We isolated nucleoli from formaldehyde-crosslinked hybrid F121-9 mESC cells using methods previously shown to yield reproducible data using MEF cells (Vertii et al. 2019). In those studies, crosslinked and non-crosslinked MEFs were directly compared and shown to yield highly overlapping results, with crosslinked samples detecting a greater proportion of the genome associated with nucleoli (Vertii et al. 2019). This suggests that crosslinking could assist detection of weak or transient nucleolar interactions. Therefore, we used crosslinking for all nucleoli isolation experiments here (Fig. 2.1A). The purity of isolated nucleoli was confirmed using phase-contrast microscopy (Fig. 2.2A). Immunoblot analysis of nucleolar fractions showed that they were enriched for nucleolar protein fibrillarin relative to beta-actin (Fig. 2.2B). Because our previous experiments showed that H3K27 methylation was important for NAD-nucleolar interactions in MEF cells (Vertii et al. 2019), we measured the ratio of heterochromatin mark H3K27me3 to
total H3 protein levels, and found that it was modestly enriched in nucleolar fractions relative to the total extracts (1.5–1.6-fold enrichment, Fig. 2.2C). Quantitative PCR analysis revealed 9–18-fold enrichment of 45S rDNA sequences in purified nucleolar DNA relative to genomic DNA (Fig. 2.2D). These results indicated the enrichment of nucleoli in our preparations; hence, we proceeded with whole-genome sequencing of nucleolar DNA.

**Bioinformatic analysis of NADs**

We performed two biological replicate preparations of crosslinked F121-9 mESC nucleoli. In each replicate experiment, we extracted nucleolus-associated DNA from nucleoli, along with genomic DNA from whole cells from the same population of cells. We sequenced approximately 50 million paired-end reads from each nucleolar and genomic DNA sample. We note that subsampling analyses of larger MEF datasets previously showed that the number of peaks detected had reached a plateau at this sequencing depth (Vertii et al. 2019). Genomic reads were mostly uniformly distributed across the genome, whereas nucleolar reads contained well-defined peaks and valleys, with peaks overlapping known heterochromatic regions, such as constitutive LADs (cLADs) (Peric-Hupkes et al. 2010) and late replicating regions (Hiratani et al. 2010) (Fig. 2.1B, C). cLADs were previously defined as LADs that are lamina-associated in mESCs and also in neural precursor cells (NPCs) and astrocytes differentiated from these mESCs (Peric-Hupkes et al. 2010). Previous studies of NADs have
identified frequent overlap of NADs with LADs (van Koningsbruggen et al. 2010; Németh et al. 2010; Dillinger et al. 2017; Vertii et al. 2019) and with late-replicating regions (Dillinger et al. 2017; Vertii et al. 2019); thus, we concluded that the nucleolar reads are enriched with bona fide nucleolar heterochromatic regions in F121-9 mESCs.

Calculating the log ratio of nucleolar reads to genomic reads resulted in a raw metric of nucleolar association across the genome (nucleolus/gDNA ratio tracks in Fig. 2.1B, C). As in MEFs, visual inspection of the nucleolus/genomic ratio in mESC revealed a negative slope across most chromosomes, especially noticeable on large chromosomes (Fig. 2.1C). Mouse chromosomes are acrocentric, i.e., the centromere is found at one end of a chromosome, and by convention, these are annotated on the left. Because pericentromeric regions frequently associate with nucleolar periphery (Ragoczy et al. 2014), nucleolar associations on centromeric end of chromosomes are usually more frequent. Peak calling based only on nucleolar/genomic ratios would identify peaks mostly at the centromere-proximal end, with fewer peaks at the chromosome-distal end. For this reason, we used our previously described Bioconductor package named \textit{NADfinder} (Vertii et al. 2019) to call NAD peaks in F121-9 mESCs. This software uses local background correction, which was important for detection of validated NAD peaks distal from centromeres in MEFs (Vertii et al. 2019). \textit{NADfinder} peak calling was performed using the default settings with a 50 kb window size, a testing threshold of \text{log}_2(1.5) for background corrected \text{log}_2(\text{nucleolar}/\text{genomic})
ratio to define the null hypothesis, and adjusted p value < 0.05 (Vertii et al. 2019).
Potential peaks were further filtered to be > 50 kb long and to have log2 ratio > 1.7.

The length of the identified F121-9 NADs ranges up to 8 Mb (Fig. 2.1D), with median length 1.1 Mb, which is slightly larger than median length of MEF NADs, 0.7Mb (Vertii et al. 2019). We noted that NADs in F121-9 cells covered 31% of the non-repetitive genome, a smaller percentage than observed in crosslinked MEF NADs (41%) (Vertii et al. 2019). The 31% fraction of the mESC genome in NADs is also smaller than the fraction of the mouse genome in LADs, either for embryonic stem cells or somatic cells (~ 40%) (Peric-Hupkes et al. 2010), or during early mouse embryogenesis (~ 40–60%) (Borsos et al. 2019) (see “Discussion”).
Figure 2.1: Analysis of F121-9 NAD sequencing data and comparison with heterochromatin.

A) Schematic diagram of nucleoli isolation from crosslinked cells.
B) All of chromosome 19 is shown, which contains strongly nucleoli-associated regions. From the top, tracks shown are constitutive interLADs (ciLADs, cyan) and constitutive LADs (cLADs, red) (Peric-Hupkes et al. 2010); mESC replication timing (Hiratani et al. 2010, early replicating regions in cyan and late replicating regions in red); F121-9 cell NAD peaks (“F121-9 NADs,” called using NADfinder software based on two replicate experiments); nucleolar/gDNA ratios, shown for both replicate experiments; raw read counts from both replicates for nucleoli-associated DNA (“Nucleolus,” brown) and total genomic DNA (“gDNA,” dark blue).
C) As in panel A, with all of chromosome 9 shown.
D) Length distribution of F121-9 NADs compared to those from crosslinked MEF cells (Vertii et al. 2019).
Figure 2.1

A

Cytoplasm
Nucleus

Supernatant 1 ("S1") + Pellet: Semi-purified Nucleoli
Centrifuge

Supernatant 2 ("S2") + Pellet: Purified Nucleoli
Re-sonicate
Centrifuge

Total Extract

1% HCHO Sonicate

B

<table>
<thead>
<tr>
<th>Refseq genes</th>
<th>Chr</th>
<th>Range</th>
<th>Refseq genes</th>
<th>Chr</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21</td>
<td>60</td>
<td></td>
<td>21</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>80</td>
<td></td>
<td>21</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>100</td>
<td></td>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>120</td>
<td></td>
<td>21</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>123</td>
<td></td>
<td>21</td>
<td>123</td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th>Refseq genes</th>
<th>Chr</th>
<th>Range</th>
<th>Refseq genes</th>
<th>Chr</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21</td>
<td>60</td>
<td></td>
<td>21</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>80</td>
<td></td>
<td>21</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>100</td>
<td></td>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>120</td>
<td></td>
<td>21</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>123</td>
<td></td>
<td>21</td>
<td>123</td>
</tr>
</tbody>
</table>

D

length, log10 (bp)

F121-9 NAD MEF XL NAD

Legend:
- Black: constitutive interLAD (ciLAD)
- Blue: constitutive LAD (cLAD)
- Red: mESC Replication Timing
- Green: F121-9 NADs
- Purple: Nucleolus/gDNA ratio (replicate #1)
- Orange: Nucleolus (replicate #1)
- Cyan: gDNA (replicate #1)
- Yellow: Nucleolus/gDNA ratio (replicate #2)
- Brown: Nucleolus (replicate #2)
- Pink: gDNA (replicate #2)
- Pink: gDNA (replicate #2)
Figure 2.2: Isolation and characterization of purified nucleoli in mESC.

A) Phase-contrast microscopy images of F121-9 mESCs grown in colonies (left panel), and nucleoli purified from them (right panel). 20x magnification, scale bar 200 μm. The inset (lower right) shows a 3x magnified image of the purified nucleoli.

B) Immunoblots of fractions generated during nucleoli isolation from two replicate experiments. Fractions are labeled as shown in Fig. 2.1A. Fibrillarin was enriched, and beta-actin depleted, in nucleolar fractions.

C) H3K27me3 and total H3 immunoblots of the same fractions shown in panel B. Ratios of H3K27me3 to H3 signals were normalized to the Total Extract values for each replicate and are presented at the bottom. We observed slightly higher H3K27me3/H3 ratios in the nucleolar fractions relative to the total extracts (~1.5-1.6 fold).

D) RT-qPCR measurement of 45S rDNA enrichment in nucleolar DNA from replicate experiments 1 and 2. Two different primer sets were used. Data are represented as mean enrichment relative to genomic DNA, error bars represent standard deviations for triplicate technical measurements.
Figure 2.2

A. mESC and Purified nucleoli

B. Replicate #1 and #2 for Fibrillarin and Beta-actin

C. Replicate #1 and #2 for H3K27me3 and H3

D. Fold nuclear rDNA enrichment for primer set 1 and 2
3D immuno-FISH confirmation of NAD peaks in F121-9 mESCs

To validate associations of NADs with nucleoli by an orthogonal method, we performed 3D immuno-FISH experiments, scoring association of BAC DNA probes with nucleolar marker proteins fibrillarin or nucleophosmin (Figs. 2.3 and 2.4). We tested the association of a euchromatic negative control probe, pPK871, which lacks nucleolar association in MEFs (Vertii et al. 2019) and did not contain a peak in our F121-9 NAD-seq data. The mean frequency of nucleolar association for this probe was ~ 24% (Fig. 2.4A). Three additional non-NAD BAC probes (pPK825, pPK1000, and pPK1003) displayed similar levels of nucleolar association (Fig. 2.4A). The average association frequency for these non-NAD probes in F121-9 cells was 22%, similar to the 20% frequency observed in MEF cells (Vertii et al. 2019). These observations result from stochastic positioning of loci within the nuclear volume. We note that pPK825 was also not associated with nucleoli in MEFs, whereas pPK1000 and pPK1003 had not been tested in MEFs (Vertii et al. 2019).

We also analyzed BAC probes pPK915 and pPK999 (Fig. 2.3A, B and Fig. 2.4), two centromere-proximal probes which were expected to strongly associate with nucleoli. Indeed, both of these probes displayed more frequent nucleolar association than did the set of non-NAD probes (p < 0.0001). We note that probe pPK915 overlaps a ciLAD region (Fig. 2.3A), which means that it was not observed to associate with lamina in mESCs or MEFs (Peric-Hupkes et al. 2010). Thus, this is an example of preferential nucleolar association, which will be
discussed below. We also note that pPK999 overlaps both a NAD and a LAD in mESCs, but not in MEFs (pPK999, Fig. 2.3B). Furthermore, pPK999 contains the Egfr gene, for which transcript levels are higher in MEFs (FPKM value 51.5) (Delbarre et al. 2017) compared to mESCs (FPKM value 0.2). This is an example of a genomic locus that is nucleolus-associated and transcriptionally repressed in mESCs, and which is no longer associated and becomes more active in MEFs.

In sum, these FISH data demonstrate that the identified NADs include bona fide nucleolar heterochromatic regions in F121-9 mESCs, and that these localizations can vary during cell differentiation.

We also examined whether the gaps between called peaks in centromere-proximal regions (e.g., Fig. 2.3C, D and Fig. 2.4) were associated with nucleoli. In our previous analysis of MEF NADs, we showed that such regions, which we termed “NAD splitting regions (NSRs)” displayed much higher gene expression than do the flanking NADs (Vertii et al. 2019). This indicated that the decreased nucleolar sequencing read numbers from NSRs reflect a distinct chromatin state. To test directly whether such regions display reduced nucleolar association, we analyzed two BAC probes (pPK1006 and pPK1007) that display canonical NSR characteristics: they overlap ciLADs and early replicating regions and exhibit high transcriptional activity (Fig. 2.3C, D). FISH assays established that these two probes are more frequently associated with nucleolar periphery compared to non-NAD probes (p < 0.0001); hence, these two BAC regions are NADs (Fig. 2.4A). However, either with or without background correction, NADfinder did not
predict either pPK1006 or pPK1007 to be NADs (Fig. 2.3C, D). We hypothesize that there are both bioinformatic as well as biochemical contributions to these discrepancies in the analysis of the centromere-proximal NADs (see “Discussion”).

We also considered the case of regions at the distal end of chromosomes, far from centromeres. We analyzed two BAC probes, pPK914 and pPK1012, and observed that both of these were significantly more frequently associated with nucleoli than were non-NAD probes (p < 0.0001, Fig. 2.3E, F). These regions were classified as NADs by NADfinder when background correction method was included, whereas in the absence of local background correction, neither were predicted to be NADs (Fig. 2.3E, F). Hence, NADfinder clearly benefits from the use of background correction method in identifying distal NADs.
Figure 2.3: Genomic locations of BACs used for FISH experiments.

For each panel, BAC locations are outlined by a black box and indicated with a red horizontal bar above the top track. From the top, tracks include cLADs (red) and ciLADs (cyan) (Peric-Hupkes et al. 2010), followed by mESC replication timing (Hiratani et al. 2010) and mESC LADs (Peric-Hupkes et al. 2010). The next tracks are raw read distributions from both replicate preparations of nucleolus-associated DNA (“nucleolus”) and total genomic DNA (“gDNA”); F121-9 NAD peaks called by NADfinder using background correction (dark blue) followed F121-9NADpeaks predicted by NADfinder in the absence of background correction (light blue). These are followed by RNA-seq data from the same preparations of F121-9 cells used to generate the NAD data. At the bottom are data from MEF cells for comparison: replication timing (Hiratani et al. 2010), LADs (Peric-Hupkes et al. 2010), NAD peaks from crosslinked cells (Vertii et al. 2019), and RNA-seq (GSM2453368 (ENCODE Project Consortium 2012)).

A) pPK915. This ciLAD-overlapped BAC is a NAD in both F121-9 and MEF cells, encoding solute carrier membrane transport proteins (Scl22a1, 2, 3) and plasminogen (Plg).

B) pPK999. This BAC overlaps a late-replicating LAD that contains the gene encoding epidermal growth factor receptor (Egfr). Note that in MEF cells this region is not identified as a NAD, is early replicating, and displays greater expression of Egfr.

C) pPK1006. This ciLAD-overlapping BAC is in a centromere proximal region that exhibits reduced nucleolar reads compared to neighboring regions (see “Nucleolus” tracks) and was not classified as a peak by NADfinder.

D) pPK1007. This is another centromere-proximal, ciLAD overlapping BAC that has low raw nucleolar read numbers.

E) pPK914. This BAC is within a region distal from the centromere. Its identification by NADfinder required the background correction feature. This region is within a NAD in both F121-9 and MEF cells, and its overlap with a ciLAD region (cyan) indicates a lack of lamina association in these cell types. This NAD contains ion channel genes (Kcnj6, Kcnj15) and Ets family transcription factors (Erg, Ets2).

F) pPK1012. This is another centromere-distal BAC region that is called as a NAD only when local background correction method is used in NADfinder software.
Figure 2.4: 3D DNA-FISH experiments validate nucleolar association of NADs in F121-9 mESC.

A) Graph of percentage of alleles that are nucleolus-associated (individual values and means are graphed for three biological replicates) for the indicated BAC probes (see Table 2.1). The six probes on the left (red points) all displayed significantly more frequent associations that the non-NAD probes (blue points; p < 0.0001 for each red probe).

B) Maximum projection images from 3D immuno-FISH experiments with nuclear DAPI staining in blue, anti-fibrillarin (pPK871, pPK914, and pPK915, lower panels) or anti-nucleophosmin (pPK1006, pPK1007, and pPK1012, upper panels) antibody staining in red, and DNA probes in green. Magnification x63, scale bar 10 μm.
Figure 2.4

A

BAC number

% associated alleles

B

10 μm

pPK1006

pPK1007

pPK1012

pPK871

pPK914

pPK915
Two types of NADs in F121-9 mESCs

In our previous analysis of MEF data, we had defined a “Type I” class of NADs as those overlapping LADs (Vertii et al. 2019). Additionally, a contrasting “Type II” class of NADs was defined which overlaps “constitutive interLADs” (ciLADs), the regions defined as those which were not lamina-associated during multiple steps of cellular differentiation (Peric-Hupkes et al. 2010). In MEFs, Type I NADs are approximately five-fold more abundant, and tend to replicate late; in contrast, the less abundant Type II NADs more frequently overlap with early replicating regions (Vertii et al. 2019). In F121-9 mESC NADs, we also observed abundant Type I NADs that overlap with cLADs (421 Mb of the total 845 Mb NAD population; Fig. 2.5A). However, Type II NADs that overlap with ciLADs comprise only 77 Mb, much less than the 147 Mb observed in similarly crosslinked MEFs (Fig. 2.5A; Vertii et al. 2019). Visual inspection of the distribution of the two classes in a genome browser illustrated the greater size of the Type I subset compared to Type II regions (Fig. 2.5B, E, F). Despite the small size of the F121-9 Type II NAD subset, we note that we have validated nucleolar association of two Type II NAD probes (pPK914, pPK915; Figs. 2.3 and 2.4). These two probes lack significant lamina association in MEFs (Vertii et al. 2019), and both overlap ciLAD regions (Fig. 2.3A, E), indicating that they lack lamina association during multiple steps in the process of differentiation from mES cells to astrocytes (Peric-Hupkes et al. 2010; Meuleman et al. 2013). We conclude that in mES
cells, as in MEFs, a large proportion of NADs overlap LAD regions, but that the amount of ciLAD overlap in mES cells is smaller.

We then analyzed gene density and gene expression characteristics of the different NAD subsets from F121-9 cells. As we had observed in MEFs (Vertii et al. 2019), gene density of Type II NADs was greater than that of NADs as a whole, which in turn have higher gene density compared to Type I NADs (Fig. 2.5C). Using RNA-seq data we obtained from the same preparations of F121-9 cells that were used for nucleolar purification, we analyzed genomic trends in steady-state mRNA levels by plotting the distributions of the FPKM values. As in MEFs (Vertii et al. 2019), F121-9 NADs displayed lower FPKM values than the genome-wide average (p < 0.0001). In addition, FPKM values for the Type I NAD subset were significantly lower than those for NADs as a whole (p < 0.0001) (Fig. 2.5D). Thus, Type I NADs in both MEFs and F121-9 cells display low gene expression levels characteristic of heterochromatin. In contrast, in F121-9 cells Type II NADs displayed mean gene expression levels that are slightly higher than those observed in the whole genome (p < 0.0003) or even in non-NAD regions (p < 0.0233) (Fig. 2.5D). Therefore, in both F121-9 cells and MEFs (Vertii et al. 2019), Type II NADs can become associated with nucleoli without adopting the highly silenced status of Type I NADs.

However, F121-9 NADs displayed a prominent difference from those in MEFs, regarding overlap with H3K27me3 peaks. We note that H3K27me3 is functionally important for heterochromatin localization because Ezh2 inhibitors
that block this modification decrease lamina and nucleolar associations of heterochromatin (Harr et al. 2015; Vertii et al. 2019). In MEFs, we observe frequent overlap of H3K27me3 peaks (Delbarre et al. 2017) with both Type I (117 Mb out of 567 Mb) and Type II NADs (101 Mb out of 147 Mb) (Fig. 2.5 G, H; Vertii et al. 2019). In contrast, in F121-9 cells, we observed that overlap of NADs with H3K27me3-enriched domains (Cruz-Molina et al. 2017) was much smaller than observed in MEFs: only 9 Mb of the 421 Mb of Type I NADs and 22 Mb of 77 Mb of Type II NADs overlap with H3K27me3 domains (Fig. 2.5E, F). These differences likely reflect the lower abundance of repressive histone marks in mESCs compared to differentiated cells; this includes H3K27me3, which becomes more abundant during differentiation ((Martens et al. 2005; Hawkins et al. 2010; Atlasi and Stunnenberg 2017; Zhang et al. 2020); see “Discussion”). Indeed, our analysis of an mESC data set (GSM2416833; Cruz-Molina et al. 2017; see “Materials and methods”) detected 517 Mb of H3K27me3 peak regions in F121-9 cells, and an almost two-fold larger amount (990 Mb) was found in MEFs (GSM1621022; Delbarre et al. 2017)). However, we note that the amount of H3K27me3 peaks in NADs is much more than two-fold greater in MEFs (417 Mb, Fig. 2.5G, H) than in F121-9 cells (66 Mb, Fig. 2.5E, F). Together, these data suggest that H3K27 methylation is a key aspect of NAD chromatin maturation that has not yet occurred fully in mES cells (see “Discussion”).
Figure 2.5: Two types of NADs in F121-9 mESC.

A) Venn diagram illustrating the overlaps among F121-9 NADs, MEF NADs (Vertii et al. 2019), cLAD, and ciLAD regions (Peric-Hupkes et al. 2010). Numbers show the size of the indicated regions in Mb.

B) Chromosomal view of F121-9 NADs overlapping cLADs and ciLADs. The entire chromosome 19 is shown. Euchromatic features (early replication timing, ciLAD) are displayed in cyan, and heterochromatic features (late replication timing, cLAD) are shown in red. From the top, displayed tracks are mESC replication timing (Hiratani et al. 2010), cLAD (Peric-Hupkes et al. 2010), NAD overlap with cLAD (i.e., Type I NADs, magenta), nucleolar genomic ratio and NAD peaks (blue), NAD overlap with ciLAD (i.e., Type II NADs, green), ciLAD (Peric-Hupkes et al. 2010), H3K27me3 domains, and mESC H3K27me3 ChIP-seq data (Cruz-Molina et al. 2017) used for H3K27me3 domain identification (olive green).

C) Gene densities (genes/Mb) of the indicated regions, ranked left to right. “NAD” indicates all F121-9 NADs.

D) A box plot of gene expression levels from F121-9 RNA-seq data, expressed as log10(FPKM+1) for the same indicated genomic regions as in C. The top of the red box indicates the mean value for each population, and the standard deviation is marked by the red error bar.

E) Venn diagram illustrating the overlaps among F121-9 NADs, cLADs (Peric-Hupkes et al. 2010), and mESC H3K27me3 domains (Cruz-Molina et al. 2017). Numbers indicate the size of regions in Mb. The overlaps among all three sets (9 Mb) and between the cLAD and H3K27me3 sets (10 Mb) are left off the diagram because of their small sizes. Diagram was generated using eulerAPE 3.0.

F) As in E, except here the overlap analysis includes ciLADs (Peric-Hupkes et al. 2010) instead of cLADs.

G) As in E, except here Venn diagram illustrates the overlaps among crosslinked MEF NADs (Vertii et al. 2019), cLADs (Peric-Hupkes et al. 2010), and MEF H3K27me3 domains (Delbarre et al. 2017).

H) As in G, except here the overlap analysis includes ciLADs (Peric-Hupkes et al. 2010) instead of cLADs.
Figure 2.5

A

B

C

D

Table

E

F

G

H

Figure 2.5 Representation

- Constitutive LADs (cLAD)
- F121-9 overlap cLADs
- F121-9 NADs
- F121-9 overlap ciLADs
- Constitutive interLADs (ciLAD)
- mESC H3K27me3 domains
- mESC H3K27me3 ChIP
Cell type-specific and conserved NADs

We compared F121-9 stem cell NADs with crosslinked MEF NADs (Vertii et al. 2019), defining overlapped regions on a nucleotide-by-nucleotide basis (e.g., Fig. 2.6A). Close to 80% (660 Mb) of nucleotides in stem cell NADs overlap with nucleotides in MEF NADs (Fig. 2.6A). We designate NADs shared by MEFs and F121-9 stem cells as “conserved NADs.” Analysis of the intersection of conserved NADs with cLAD and ciLAD regions revealed that more than half of conserved NADs overlap cLADs (370 Mb; Fig. 2.5A), which are the most gene-poor subset of LADs and are generally poorly expressed, constitutive heterochromatin (Peric-Hupkes et al. 2010; Meuleman et al. 2013; van Steensel and Belmont 2017). Consistent with these trends, Jaccard similarity coefficient analysis indicated high correlation of conserved NADs with cLADs and late replicating regions (Marchal et al. 2018) (Fig. 2.6B). Furthermore, the conserved NADs display the lowest transcript levels in both cell types (Fig. 2.6C-F), as expected due to the constitutive heterochromatic features of these regions.

We next turned our attention to NADs found only in one of the two analyzed cell types. The Jaccard analysis indicated that these cell type-specific NAD regions (i.e., “MEF-specific NADs” and “F121-9-specific NADs”) are distinct from the conserved NADs, clustering separately from conserved NADs, cLAD, and late replicating regions (Fig. 2.6B). We analyzed steady-state mRNA levels in conserved and cell type-specific NADs by using FPKM values from F121-9 and MEF (Delbarre et al. 2017) RNA-seq data (Fig. 2.6C, D). As we expected,
MEF RNA-seq data revealed lower levels of transcripts from genes within MEF-specific NADs than from F121-9-specific NADs (p value < 0.0001) (Fig. 2.6C), indicating that in MEFs, nucleolar association correlates with transcriptional silencing. In contrast, our RNA-seq data from F121-9 cells showed that transcript levels within both the MEF-specific NADs and the F121-9-specific NADs are statistically indistinguishable (p value = 0.82) (Fig. 2.6D). We observed similar trends in independent sets of MEF and mESC RNA-seq data from the literature (Lowe et al. 2015; Chronis et al. 2017) (Fig. 2.6E, F). These observations were unexpected in that the MEF-specific NADs are not nucleolus-associated in the F121-9 cells, yet are on average less highly expressed than non-NAD genes in these cells. These data suggest that in F121-9 stem cells, gene repression could precede localization to the nucleolar periphery that occurs later during cellular differentiation (see “Discussion”).
Figure 2.6: Conserved and cell type-specific NADs.

A) IGV browser view of entire chromosome 15. Euchromatic features (early replication timing, ciLAD) are displayed in cyan, and heterochromatic features (late replication timing, cLAD) are shown in red. From the top, tracks shown are cLAD, ciLAD (Peric-Hupkes et al. 2010), mESC replication timing (Hiratani et al. 2010), F121-9 nucleolar/genomic ratio and F121-9 NAD peaks (blue), “F121-9-specific NADs,” i.e., NADs found only in F121-9 cells (light blue), “conserved NADs,” or NADs shared between F121-9 and MEFs (magenta), “MEF-specific NADs” (dark green), MEF NAD peaks and MEF nucleolar/genomic ratio (Vertii et al. 2019) in green, and MEF replication timing (Hiratani et al. 2010).

B) Jaccard similarity coefficients were grouped based on similarities among the indicated regions. “F121-9NAD” indicates all NADs identified in F121-9 cells in this study. Conserved NAD” indicates NADs shared between F121-9 and MEF NADs (Vertii et al. 2019), whereas “F121-9-specific NAD” indicates NADs detected in F121-9, but not MEF cells. Conversely, “MEF-specific NAD” indicates NADs found in MEFs, but not in F121-9 cells. “Type I NAD” indicates F121-9 NADs that overlap with cLADs, and “Type II NAD” indicates F121-9 NADs that overlap with ciLADs (Peric-Hupkes et al. 2010). “cLAD” and “ciLAD” regions are from Peric-Hupkes et al. 2010, and F121-9 early replication timing and late replication timing regions are from Marchal et al. 2018. Note that F121-9 NADs, conserved F121-9 NADs, cLADs, and Type I NADs are highly similar. In contrast, Type II NADs are most similar to F121-9-specific NADs.

C) A box plot of gene expression levels from MEF RNA-seq data (GSM1621026; Delbarre et al. 2017), expressed as log10(FPKM+1) for the indicated subsets of NAD, non-NAD and whole genome regions. The statistical significance of pairwise comparisons were all p < 0.0001 (Welch’s t-test).

D) As in C, except our F121-9 RNA-seq data is used for FPKM analysis. The indicated pairwise comparisons were all statistically significant (p < 0.0001), except for that between F121-9 and MEF-specific NADs do not achieve statistical significance (p = 0.82).

E) As in C, except different MEF RNA-seq data (GSE90894; Chronis et al. 2017) was used for FPKM analysis. The changes between cell type-specific NADs achieve statistical significance (p < 0.0001, Welch’s t-test).

F) As in C, except mESC RNA-seq data (GSM1418813; Lowe et al. 2015) is used for FPKM analysis. The changes between F121-9 and MEF-specific NADs do not achieve statistical significance (p = 0.13).
Figure 2.6

A

<table>
<thead>
<tr>
<th>Refseq genes</th>
<th>chr15</th>
<th>qA1</th>
<th>qA2</th>
<th>qB1</th>
<th>qB2</th>
<th>qB3.1</th>
<th>qB3.2</th>
<th>qC</th>
<th>qD1</th>
<th>qD2</th>
<th>qD3</th>
<th>qE1</th>
<th>qE2</th>
<th>qE3</th>
<th>qF1</th>
<th>qF2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mESC Replication Timing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F121-9 NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F121-9-specific NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conserved NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF-specific NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF nucleolus/gDNA ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF Replication Timing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

Color Key

0 1

F121-9 NAD
Conserved NAD
Type I NAD
Late Replication Timing
cLAD
Early Replication Timing
ciLAD
MEF-specific NAD
F121-9-specific NAD
Type II NAD

C

D

E

F

mESC GSM1418813

Figure 2.6

A

<table>
<thead>
<tr>
<th>Refseq genes</th>
<th>chr15</th>
<th>qA1</th>
<th>qA2</th>
<th>qB1</th>
<th>qB2</th>
<th>qB3.1</th>
<th>qB3.2</th>
<th>qC</th>
<th>qD1</th>
<th>qD2</th>
<th>qD3</th>
<th>qE1</th>
<th>qE2</th>
<th>qE3</th>
<th>qF1</th>
<th>qF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>mESC Replication Timing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F121-9 NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F121-9-specific NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conserved NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF-specific NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF NADs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF nucleolus/gDNA ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF Replication Timing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

Color Key

0 1

F121-9 NAD
Conserved NAD
Type I NAD
Late Replication Timing
cLAD
Early Replication Timing
ciLAD
MEF-specific NAD
F121-9-specific NAD
Type II NAD

C

D

E

F

mESC GSM1418813
Gene ontology analysis of conserved and cell type-specific NADs

To further characterize the conserved NADs, we next analyzed enriched GO terms within these. The most significantly enriched Molecular Functions term was “Response to smell detection” (Fig. 2.7A), including olfactory receptor (OR) and vomeronasal receptor genes. These clustered genes are not expressed in either stem cells or fibroblasts and are frequently within NADs in both F121-9 stem cells and MEFs (e.g., the OR genes on chr11, Fig. 2.7B). Among other well-represented gene families in conserved NADs were cytochrome P450 family members: Cyp2a12, Cyp2b10, Cyp2c50 (“heme-interacting genes” in Fig. 2.7A), which are responsible for breaking down toxins, as well as synthesizing steroid hormones, fats and acids, and are most highly expressed in liver (Hannemann et al. 2007). Neurotransmitter receptors were also enriched for conserved NADs, for example, genes that encode for glutamate receptors (Gria2, Grid2, etc.), GABA-A receptors (Gabra5, Gabrb1, etc.) and glycine receptors (Glra1, Glrb, etc.). The common thread among these gene classes is in that they are developmentally regulated, and most strongly induced in lineages not represented by embryonic stem cells or fibroblasts.

We next analyzed the F121-9-specific NADs. Among these, chemotactic cytokines were the GO-derived “Molecular Functions” class with the lowest q-value (Fig. 2.7C). The majority of these chemokines are represented by the CC chemokine ligand family, a cluster of which is shown in Fig. 2.7D. This cluster of Ccl2, Ccl12, and Ccl1 genes has heterochromatic features in the F121-9 cells:
late replication timing, no steady-state mRNA transcripts, presence within both LAD and NAD regions. In contrast, in MEFs, this gene cluster is within neither NAD nor LAD sequences and has euchromatic features, including early replication timing and high gene transcript levels. This is an example of a genomic region in which multiple features are altered, becoming more euchromatic upon differentiation.

We then considered the converse case, the MEF-specific NADs. Among these, the “Biological Processes” GO classifications included genes responsible for differentiation along the anterior-posterior axis (Fig. 2.7E), an example of which is Pcsk6 gene (Fig. 2.7F). This genomic region displays euchromatic features (overlapping a ciLAD region, early replicating timing and high transcript levels: FPKM value 22.2) in mESCs, befitting the need for anterior-posterior axis establishment factors at this early developmental stage. In MEFs, this locus displays altered features, becoming nucleolus-associated, and generating reduced transcript levels (FPKM value 6.6) (Delbarre et al. 2017). In general, both conserved and cell type-specific NADs generally include genes that display reduced expression levels, suggesting that nucleolar localization could contribute to (or be a consequence of) the transcriptional silencing of resident genes. A major question remains as to how functionally distinct classes of NADs (e.g., Type I and Type II NADs) are targeted to nucleoli, and how this has distinct transcriptional consequences in each case (e.g., Fig. 2.5D; see “Discussion”).
**Figure 2.7: GO analysis of conserved and cell type-specific NADs.**

**A)** Molecular Functions subset of GO enrichment analysis of conserved NADs, with \(-\log_{10}(q\text{ values})\) shown.

**B)** Genomic region containing NAD peak (red box) conserved in both MEF and F121-9 cells. This peak contains a cluster of olfactory genes on chromosome 11. ciLAD, mESC, and MEF replication timing tracks are displayed as in Fig. 2.6B. The other tracks shown from the top are mESC LADs (Peric-Hupkes et al. 2010; red), F121-9 nucleolar/genomic ratio, NADs and RNA-seq data (blue), MEF LADs (Peric-Hupkes et al. 2010), MEF nucleolar/genomic ratio, NADs (Vertii et al. 2019) (green), and RNA-seq (GSM2453368 (ENCODE Project Consortium 2012)) (blue).

**C)** Molecular Functions subset of GO enrichment analysis of F121-9-specific NADs.

**D)** As in panel B, showing genomic region corresponding to F121-9-specific NAD (red box), overlapping Ccl family of chemokine ligands.

**E)** Biological Functions subset of GO enrichment analysis of MEF-specific NADs.

**F)** As in B, showing genomic region containing MEF-specific NAD (red box), overlapping the Pcsk6 gene important for differentiation along anterior-posterior axis.
Differentiation along the anterior-posterior axis

SH2/SH3 domain-mediated scaffolding
Interaction with a CCR6 chemokine receptor.
Neurotransmitter signaling
Inhibition of serine-type endopeptidases.
Heme-interacting genes
Response to smell detection

Figure 2.7

A
Molecular Functions: Conserved NADs

Interaction with a CCR6 chemokine receptor.
Neurotransmitter signaling.
Inhibition of serine-type endopeptidases.
Heme-interacting genes.
Response to smell detection.

B

C
Molecular Functions: F121-9-specific NADs

Heparin interacting genes.
SH2/SH3 domain-mediated scaffolding.
Chemokines.

D

E
Biological Processes: MEF-specific NADs

Granule signaling.
Circulatory system process.
Differentiation along the anterior-posterior axis.
Allele-specific NADs

The F121-9 mES cell line we analyzed was derived from a blastocyst that resulted from a cross between mouse inbred strain 129 and mouse subspecies Mus musculus castaneus (129 x Castaneus) (Rivera-Mulia et al. 2018). The genome of this hybrid cell line has a high single nucleotide polymorphism (SNP) density, with a genome-wide average of 1 SNP per 150 bp (Rivera-Mulia et al. 2018). To determine if we could detect allele-specific differences in our data, we assigned our NAD-seq data to one of the two genomes using SNPsplit_v0.3.2 for allele-specific splitting of alignments (Krueger and Andrews 2016). We obtained approximately 50 million reads in each replicate of the nucleolar DNA samples. We then identified the reads with informative SNPs, yielding approximately 14–16 million allele-specific reads per genome. These allele-specific reads were used for the identification of allele-specific NAD peaks by NADfinder. However, visual inspection of the allele-specific peaks did not reveal correlations with allelic differences in previously published F121 subclone datasets, such as replication timing and ATAC-seq (Rivera-Mulia et al. 2018), Hi-C (Giorgetti et al. 2016) and our F121-9 RNA-seq datasets. However, previous subsampling analysis of larger MEF NAD-seq datasets revealed that the number of peaks detected plateaued at approximately 50 million reads (Vertii et al. 2019). Hence, to detect the maximum number of allele-specific features, we predict needing to obtain ~ 160–180 million reads per sample prior to assigning to specific genomes.
Nevertheless, we did observe chromosome-scale differences in the nucleolar associations in the two genomes. Although the numbers of allele-specific sequencing reads from the genomic samples were comparable for the two alleles throughout the genome, for some chromosomes the numbers of nucleolar sequencing reads differed for the two alleles. For example, more nucleolar reads were detected from the *Castaneus* genome for chromosomes 10 and 11 (Fig. 2.8A, B), whereas more reads from the 129 genome were detected for chromosome 16 (Fig. 2.8C). We note that these differences were more pronounced on the centromere-proximal side of the chromosome. Therefore, one possible explanation for these allele-specific differences could be distinctions in chromosomes bearing nucleolar organizer regions (NORs) between these two mouse genomes. Nucleoli are formed by NORs, which contain arrays of ribosomal DNA (rDNA) repeats (McStay 2016). The identity of NOR-bearing chromosomes vary among mouse subspecies and strains (Britton-Davidian et al. 2012). *Mus musculus castaneus* subspecies harbor NORs on chromosome 11, and less frequently, on chromosome 10, but not on chromosome 16 (Suzuki et al. 1990; Britton-Davidian et al. 2012). Conversely, a laboratory inbred strain 129 has been reported to contain a NOR on chromosome 16 (Kurihara et al. 1994), but there are no reports of NORs on chromosomes 10 and 11 in this strain.

We hypothesized that higher nucleolar signals of a chromosome would correlate with the higher frequency of nucleolar contact. To test our hypothesis, we performed immuno-FISH using whole chromosome paint probes for
chromosomes 5, 10, 12 and nucleophosmin antibody to label nucleoli (Fig. 2.8D-F). We chose to analyze chromosome 10 because the difference between allele-specific nucleolar reads was most noticeable (Fig. 2.8A). We used chromosome 5 as our negative control for NOR-bearing chromosome, due to it being a large chromosome with no literature reports of containing a NOR and exhibiting low nucleolar read intensity in both genomes according to NAD-seq (Fig. 2.9A). Chromosome 12 has been shown to contain a NOR both in the 129 strain (Kurihara et al. 1994) and in castaneus subspecies (Suzuki et al. 1990; Britton-Davidian et al. 2012). Consistent with these reports, both 129 and Castaneus genomes showed robust nucleolar signal for chromosome 12 (Fig. 2.9B). Thus, we utilized chromosome 12 as our positive control for NOR-bearing chromosome. We predicted that chromosome 10 would show intermediate results between chromosomes 5 and 12, due to only Castaneus genome showing robust nucleolar signal. Indeed, the mean frequency of chromosome 10 contacting the nucleoli was 49.7%, whereas for chromosome 5, it was 36.8%, and for chromosome 12, it was 74.3% (Fig. 2.8G). The statistical difference between the association frequencies of chromosomes 5 and 10 ($p = 0.0179$) and chromosomes 10 and 12 ($p = 0.0014$) was statistically significant. We conclude that NAD-seq was able to distinguish different allele-specific frequencies of nucleolar association across broad chromosomal domains.
Figure 2.8: Allele-specific chromosome-nucleolus contacts.

A) IGV browser view of entire chromosome 10. Allele-specific genomic DNA (“gDNA”) reads assigned to the *Castaneus* genome are shown in light blue and genomic DNA reads assigned to the 129 genome are shown in orange. Nucleolus-associated *Castaneus* reads are in dark blue and those from the 129 genome are in red. Note the higher numbers of nucleolus-associated reads, especially in the centromere-proximal region, assigned to the *Castaneus* genome compared to the 129 genome.

B) As in A, except here chromosome 11 is shown. The *Castaneus* genome has more nucleolar read numbers in the centromere-proximal region than the 129 genome.

C) As in A, except here chromosome 16 is shown. In this case, the 129 genome has more nucleolar read numbers in the centromere-proximal region compared to the *Castaneus* genome.

D) Two pairs of individual z-stack slices from two different cells (left and right pairs of images) from 3D immuno-FISH experiments. Nuclear DAPI staining in blue, anti-nucleophosmin antibody staining in red, and whole chromosome paint probe for Chr 5 in green. Magnification x 63, scale bar 10 μm.

E) As in D, except here maximum projection images of two representative nuclei from 3D immuno-FISH experiments with the probe for Chr 10 are shown.

F) As in E, except here the probe for Chr 12 was used.

G) Graph of percentage of Chr 5, Chr 10, and Chr 12 territories contacting nucleoli (individual values and mean of three biological replicates). Chr 10 displays significantly different frequencies of nucleolar contacts than Chr 5 (p=0.0179) or Chr 12 (p = 0.0014).
Figure 2.8

A

B

C

D chr5

E chr10

F chr12

G

80

60

40

20

% chromosomes contacting nucleoli

chromosome

5

10

12
Figure 2.9: Allele-specific NAD-seq data for chromosomes 5 and 12.

A) As in Fig. 2.8A-C, except here the entire chromosome 5 is shown. Note that both Castaneus and 129 genomes exhibit low nucleolar read numbers (“nucleolus” tracks).

B) As in panel A, except here the entire chromosome 12 is displayed. Both Castaneus and 129 genomes show robust nucleolar read numbers.
Figure 2.9

A

B
Discussion

Heterochromatin formation during differentiation

Several types of evidence indicate that compared to differentiated cells, chromatin in mESCs is less condensed, and the ratio of euchromatin to heterochromatin is higher (Gaspar-Maia et al. 2011). For example, fluorescence recovery after photobleaching experiments demonstrated that mESCs display more highly mobile core and linker histones, as well as Heterochromatin Protein 1 (HP1α) than do differentiated cells. These features are thought to contribute to the transcriptional hyperactivity in pluripotent stem cells (Meshorer et al. 2006; Bhattacharya et al. 2009), including many repetitive elements that are silent in somatic cells but are transcribed in mESCs (Efroni et al. 2008). Additionally, microscopy studies showed that electron-dense heterochromatic structures are less condensed and less frequently localize near nuclear lamina in mESCs compared to heterochromatin in differentiated cells (Hiratani et al. 2010; Ahmed et al. 2010; Mattout et al. 2015). Particularly relevant to our studies, more prominent electron-dense perinucleolar heterochromatin-like structures have been observed in differentiated cells, such as NPCs, compared to mESCs (Savić et al. 2014). In concert with changes in the appearance and localization of heterochromatin, the abundance of heterochromatic marks such as H3K27me3 and H3K9me3 increases during differentiation (Lee et al. 2004; Martens et al. 2005; Meshorer et al. 2006; Wen et al. 2009; Hawkins et al. 2010). Some of this regulation is locus-specific, as in a recent report indicating that PRC2 is
prevented from generating repressive H3K27me3 marks at rDNA as part of a mechanism to promote high levels of ribosome biogenesis in pluripotent cells (Zhang et al. 2020). Together, these data are consistent with our observation that NADs in mESCs comprise a smaller fraction of the genome compared to MEFs (31 vs. 41%). Likewise, genome coverage by LADs increases during differentiation. For example, a recent study shows that LADs are first established immediately after fertilization, preceding TAD formation and instructing A/B compartment establishment (Borsos et al. 2019).

The Type II class of NADs is different in stem cells and fibroblasts

Two functionally distinct classes of NADs have recently been reported in mouse embryonic fibroblasts (Vertii et al. 2019). Here, we show that in F121-9 mESCs, Type I NADs that overlap LAD regions are frequently the same as those found in MEFs (Fig. 2.5A), and exhibit similar low gene expression levels as expected for constitutive heterochromatin (Fig. 2.5D). In contrast, the Type II NADs defined by their overlap with ciLAD regions is much smaller in F121-9 than in MEF cells (Fig. 2.5A). We also note that NADs in F121-9 cells display much less overlap with H3K27me3 peaks than do MEF NADs (Fig. 2.5E-H). Together, these data suggest that acquisition of H3K27me3, the hallmark of facultative heterochromatin (Trojer and Reinberg 2007) by NADs is part of the process of cellular differentiation. Indeed, we note that GO analysis of MEF Type II NADs showed enrichment for developmentally regulated GO terms, for example, organ
morphogenesis and sensory organ development (Vertii et al. 2019). Thus, stem cells prevent developmentally important genes from acquiring characteristics of facultative heterochromatin including nucleolar association, whereas these genes can become NADs after they are no longer required during development.

**How are NADs targeted to nucleoli?**

The precise mechanisms for targeting the two distinct classes of NADs to nucleoli remain unclear. Several studies implicate phase separation in the formation of heterochromatin domains (Larson et al. 2017; Strom et al. 2017; Shin et al. 2018) and nuclear bodies, such as nucleoli (Brangwynne et al. 2011; Feric et al. 2016; Mitrea et al. 2016). Our recent data suggest that Type II NADs are more sensitive than Type I NADs to hexanediol treatment (Vertii et al. 2019). Hexanediol perturbs phase separation, likely due to interfering with weak hydrophobic interactions that are important for liquid-like condensate formation (Ribbeck and Görlich 2002). Liquid-liquid demixing reactions frequently involve proteins that have intrinsically disordered regions (IDR) and RNA recognition motifs (Feric et al. 2016), as found for example in nucleolar proteins fibrillarin (FBL) and nucleophosmin (NPM-1). Notably, depletion of Nlp, the Drosophila homolog of NPM-1, led to declustering of centromeres and decreased association of centromeres with nucleolar periphery (Padeken et al. 2013). Therefore, it is possible that Type II NADs are specifically targeted to nucleolar periphery through the interactions between nucleolar proteins with IDRs (e.g.
NPM-1) with RNA species that are yet to be identified. Additionally, Polycomb repressive complex 1 (PRC1) protein chromobox 2 (CBX2) undergoes phase separation and forms liquid-like condensates in mESCs (Tatavosian et al. 2019), and Polycomb proteins are part of the MiCee complex that together with let-7 family miRNAs confers nucleolar association to specific loci (Singh et al. 2018). Therefore, Polycomb group (PcG) proteins are good candidates for nucleolar targeting of Type II NADs via phase separation. This may be especially important during differentiation, when PcG proteins gain special importance (Aloia et al. 2013; Lavarone et al. 2019). However, inhibition of PRC2 enzymatic activity decreases both nucleolar (Singh et al. 2018; Vertii et al. 2019) and lamina heterochromatin localizations (Harr et al. 2015), making it unlikely that PRC2 can target loci to a unique destination. Additionally, nucleolar localization of the Kcnq1 locus can occur in cells lacking functional Polycomb complexes (Fedoriw et al. 2012a), indicating that multiple mechanisms likely exist. Other candidate trans-acting factors that could specifically target genomic regions to the nucleolar periphery are the proteins Ki-67 and the p150 subunit of Chromosome Assembly Factor-1 (CAF-1) (Smith et al. 2014; Matheson and Kaufman 2017), and the Kcnq1ot1 (Mohammad et al. 2008) and Firre (Yang et al. 2015) long non-coding RNAs.
Evaluation of the nucleolar association of “NAD splitting regions” in centromere-proximal regions

We have analyzed the localization of two centromere-proximal probes, pPK1006 and pPK1007, which overlap sequences identified as “NAD-splitting regions (NSRs)” by NADfinder (Fig. 2.3C, D). NSRs are regions between adjacent NADs; previous analysis of these in mouse embryonic fibroblasts indicate that these display much greater levels of gene expression than neighboring NAD sequences (Vertii et al. 2019). However, both of these probes displayed significant nucleolar association in immuno-FISH experiments (Fig. 2.4). We note that both pPK1006 and pPK1007 cover regions with much fewer nucleolar reads than neighboring NADs, and therefore appear as “valleys”, especially in the case of pPK1007 (Fig. 2.3C, D). These observations led us to consider whether background correction could have eliminated positive regions, such as pPK1006 and pPK1007, which are much less frequently associated than neighboring sequences, but are still more frequently associated than the genome-wide average. However, neither pPK1006 nor pPK1007 become positive when background correction is omitted from peak calling by NADfinder (Fig. 2.3C, D). Therefore, we hypothesize that discrepancies could arise from the biochemical purification of nucleoli inherent to our approach. That is, it is possible that centromere-proximal NSR loci represent sonication-sensitive chromatin loops that are frequently lost during purification, but which indeed are often nucleolus-associated in intact cells. Comprehensive testing of this idea would
best come from future studies using an orthogonal method for identifying NADs that does not rely on biochemical purification of sonicated nucleoli.

**Anomalies of MEF-specific NADs in stem cells**

One question of interest is whether nucleolar association leads to, or is a consequence of, transcriptional repression. Notably, previous studies have shown that tethering of loci to the nucleolar periphery via 5S rDNA sequences results in transcriptional silencing (Fedoriw et al. 2012b), so at least in that case a causal relationship has been established. In MEF cells, genes in the MEF-specific NADs display mean expression levels lower than genes in the F121-9-specific NADs \( p < 0.0001 \) (Fig. 2.6C, E). This is the expected situation, in which genes that had been in NADs earlier in development (e.g. in stem cells) become derepressed if that localization is lost. In contrast, in F121-9 cells, genes within MEF-specific NADs showed similar transcript levels as genes within F121-9-specific NADs \( p = 0.82 \), Fig. 2.6D); the same was true in other mES cells analyzed \( p = 0.13 \), Fig. 2.6F). Why aren’t the MEF-specific NADs more transcriptionally active in stem cells, since they haven’t yet acquired nucleolar association? This could be due to other repressive mechanisms acting on regions within MEF-specific NADs, for example, lamina association: 40% of MEFs-specific NADs overlap with cLADs (Fig. 2.5A). Alternatively, additional factors contributing to transcriptional repression may precede (and perhaps contribute to) nucleolar association. Development of
reagents allowing control of perinucleolar associations will be key to exploring the relationship between nucleolar localization and transcriptional repression.

**Materials and Methods**

**F121-9 mESC cell culture and isolation of crosslinked nucleoli**

The F121 mouse embryonic stem cell (mESC) line is a female cell line derived from a cross between a male *Castaneus* and a female 129 mouse (Monkhorst et al. 2008), and F121-9 was subcloned subsequently (Rivera-Mulia et al. 2018). F121-9 cells were obtained from the Gilbert lab (Florida State University) at passage 8. The cells were grown on gelatin-coated plates and cultured in 2i medium. Accutase (EMD Millipore, SF006) was used to detach cells from plates and passage into new dishes. Prior to seeding cells, dishes were coated with 0.1% gelatin (EMD Millipore, SF008) for at least 25 min at room temperature, after which gelatin was aspirated. Dishes were rinsed with DPBS (Gibco, 14190144), which was aspirated, and cells were seeded in these dishes. 2i medium was obtained as described previously (Vertii et al. 2019). Cells were passaged at 3 × 10^4/cm^2 density. 2X HyCryo-STEM cryopreservation medium (GE Healthcare, SR30002.02) was used to freeze cells. For each preparation of nucleoli, cells were grown in eleven 15-cm plates and harvested one or two days after seeding them, with total cell numbers of 3-5 × 10^8 per preparation. One hour prior to nucleoli isolation, old cell culture medium was replaced with fresh
medium. Plates grown in parallel were used for genomic DNA extraction (DNeasy Blood & Tissue kit, Qiagen), and RNA extraction (TRIzol, ThermoFisher Scientific and RNeasy mini kit, Qiagen). Crosslinking followed by isolation of nucleoli was done as described previously (Vertii et al. 2019).

**Quantitative PCR**

DNA was extracted from input whole cells and purified nucleoli using DNeasy Blood & Tissue kit (Qiagen). Quantitative PCR analysis was done as outlined previously (Vertii et al. 2019).

**Antibodies**

The following antibodies were used: fibrillarin (Abcam, ab5821), nucleophosmin (Abcam, ab10530), histone H3K27me3 (Active Motif, 39155), histone H3 (Abcam, ab1791) and actin (Sigma-Aldrich, A1978). Secondary antibody for immunofluorescence was Alexa 594-conjugated donkey anti-rabbit (ThermoFisher, A-21207) and Alexa 594-conjugated goat anti-mouse (ThermoFisher, A-11020). For western blots, horseradish peroxidase (HRP) anti-mouse and anti-rabbit secondary antibodies (Jackson ImmunoResearch) were used.
**Immunoblotting**

Proteins from total cell lysates and purified nucleoli were analyzed as noted previously (Vertii et al. 2019). Proteins from the same fractions were analyzed on 17% gels and transferred to nitrocellulose (replicate #1) or PVDF (replicate #2) membranes in order to detect H3K27me3 and total H3 protein levels. Ratios of H3K27me3 bands to total H3 bands were calculated by estimating adjusted volumes of bands in Bio-Rad Image Lab software.

**DNA isolation, deep sequencing, and read preprocessing and mapping**

Total genomic and nucleolar DNA was purified using DNeasy Blood & Tissue kit (Qiagen). Libraries were generated using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs). The DNA was fragmented to a size of 350 bp, and these fragments were size selected with sample purification beads. 150 bp paired-end sequencing was performed using Illumina reagents. 52.1 and 51.5 million reads were obtained for two replicates of genomic samples, and 49.4 and 52.8 million reads were obtained for two replicates of nucleolar samples. >95% of nucleolar samples, and >96% of genomic samples were mappable. From nucleolus preparation replicate #1, 13.9 million reads were assigned to the *Castaneus* genome and 14.8 million reads were assigned to the *129* genome. From replicate #2, 15.1 million reads were assigned to the *Castaneus* genome and 16.1 million reads were assigned to the *129* genome. The sequencing data files are available at data.4dnucleome.org under accession...
numbers 4DNESXE9K9DB, 4DNESUJZ5FL2. Trimming and alignment of mapped reads to the mouse genome (mm10) was done as previously described (Vertii et al. 2019).

**RNA isolation, deep sequencing, and read preprocessing and mapping**

Total RNA from the two replicate preparations of F121-9 mESCs were extracted using TRIzol (ThermoFisher Scientific) and purified using the RNeasy mini kit (Qiagen). Libraries were constructed using the NEBNext Ultra II RNA Library Prep kit for Illumina (New England Biolabs). The mRNA was fragmented, and double-stranded cDNA library synthesized, and completed through size selection and PCR enrichment. 150 bp paired-end sequencing was achieved using Illumina HiSeq 4000 platform. 22.2 and 26.7 million reads were obtained from the two replicates of mESC RNA. >92% of replicate 1, and >86% of replicate 2 were mappable. These sequencing data files are available at data.4dnucleome.org under accession number 4DNESDHILYLU. The quality of the sequencing reads was evaluated with fastqc (0.11.5) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The paired-end reads were aligned to the mouse genome (ensemble GRCm38) using STAR (version 2.5.3a) with ENCODE standard options as --outFilterMultimapNmax 20, --alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --outFilterMismatchNmax 999, --alignIntronMin 20, --alignIntronMax 1000000, and --alignMatesGapMax 1000000. Additional parameter settings are --outFilterMismatchNoverReadLmax
0.04 and --outSAMattributes NH HI NM MD. To visualize the mapped reads, bigwig files were generated using the bamCoverage function in deepTools2 with the parameter setting --normalizeUsingRPKM.

**DNA-FISH probes**

The bacterial artificial chromosomes (BACs) were obtained from the BACPAC Resource Center of Children’s Hospital Oakland Research Institute (Oakland, CA). DNA was isolated using BAC DNA miniprep Kit (Zymo Research). BAC probes were labeled using BioPrime Labeling Kit (ThermoFisher). Streptavidin, Alexa Fluor 488 conjugate (ThermoFisher, S-32354) was used to stain biotin-labeled BAC probes.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Lab BAC name</th>
<th>BACPAC name</th>
<th>F121-9 FISH</th>
<th>NADfinder AveSig&gt;1.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr17</td>
<td>34,211,624</td>
<td>34,393,919</td>
<td>pPK825</td>
<td>RP23-444J20</td>
<td>negative</td>
<td>negative</td>
</tr>
<tr>
<td>chr12</td>
<td>110,758,289</td>
<td>111,006,192</td>
<td>pPK871</td>
<td>RP23-322M3</td>
<td>negative</td>
<td>negative</td>
</tr>
<tr>
<td>chr16</td>
<td>95,182,165</td>
<td>95,358,805</td>
<td>pPK914</td>
<td>RP24-212E13</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>chr17</td>
<td>12,456,384</td>
<td>12,690,495</td>
<td>pPK915</td>
<td>RP23-3G4</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>chr11</td>
<td>16,745,166</td>
<td>16,937,185</td>
<td>pPK999</td>
<td>RP23-263C13</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>chr7</td>
<td>130,062,90</td>
<td>130,222,254</td>
<td>pPK1000</td>
<td>RP23-227N6</td>
<td>negative</td>
<td>negative</td>
</tr>
</tbody>
</table>

Table 2.1: mm10 genomic coordinates, laboratory BAC probe names, systematic BACPAC names, FISH and *NADfinder* results for DNA-FISH probes.
3D DNA FISH/ immunocytochemistry and microscopy

3D DNA FISH/ immunocytochemistry-labeling was performed as described previously (Vertii et al. 2019), except that DNA FISH-labeling was done after immunocytochemistry, and coverslips were not treated with RNA removal solution. F121-9 mESC were seeded on 0.1% gelatin-coated 22 x 22 mm coverslips (Corning, 2850-22), with total cell number 150-250 x 10³ cells/cover slip, and permeabilized and fixed the next day. Nucleoli were stained with anti-fibrillarin antibodies, except anti-nucleophosmin antibodies were used in the third biological replicates of the pPK999 and pPK1000 and all three biological replicates of pPK1006, pPK1007 and pPK1012 immuno-FISH assays. Cells were fixed with 4% Paraformaldehyde in 1X PBS for 10 min at RT for the second time at the end of immunocytochemistry part of the assay, before proceeding to DNA FISH-labeling in order to preserve nucleolar staining. Images were acquired using Zeiss LSM 700 laser scanning confocal microscope and PMT detector (63x 1.40 Oil DIC M27 Plan-Apochromat objective). DNA-FISH probes were counted.
through z-stacks manually and scored as “associated” if there was no gap between the probe and the nucleolar marker. Each probe was analyzed in at least three biological replicates, with at least 100 alleles scored in each replicate. Z stacks are represented as 2D maximum projections using Fiji software (Schindelin et al. 2012). R, a system for statistical computation and graphics (Ihaka and Gentleman 1996), was used for the analysis of FISH data. Percentage data was first arcsin transformed to homogenize the variance. Levene’s test indicated that the assumption of homogeneity of variances was met. One-way analysis of variance (ANOVA) with Completely Randomized Design was performed using the lsmeans package to test whether there are significant differences among the probes in question. Comparisons are declared statistically significant if Hochberg-adjusted p-value < 0.05 (Hochberg 1988).

**Chromosome paints FISH/immunocytochemistry**

Whole chromosome paint probes for Chr 5 (MetaSystems Probes, D-1405-050-FI), Chr 10 (MetaSystems Probes, D-1410-050-FI) and Chr 12 (MetaSystems Probes, D-1412-050-FI) directly conjugated to a green emitting fluorochrome were purchased from MetaSystems Probes and used according to manufacturer’s directions with some modifications. Briefly, after immunocytochemistry, coverslips were processed for DNA FISH-labeling as described previously (Vertii et al. 2019). 5-7 μl of chromosome paint probes were denatured at 75°C for 2 min, after which the probes were hybridized with
separately denatured coverslips in a humidified chamber at 37°C overnight. The next day coverslips were washed in 0.4X SSC, pH 7.0-7.5 at 72°C for 2 min, then drained and washed in 2X SSC, 0.05% Tween-20, pH 7.0 at RT for 30 s. The coverslips were rinsed in 1X PBS, and stained with DAPI at RT for 2 min. After rinsing the coverslips in 1X PBS and distilled water, they were mounted on slides with ProLong Gold antifade mountant (ThermoFisher Scientific, P36934).

Microscopy and analysis of imaging were done as for the BAC probes, except that chromosomes were counted as contacting nucleoli if chromosome paint probe and nucleolus had at least 3 pixels of overlapping signal. This quantitation criterion was based on chromosome paint immuno-FISH experiments described previously (Strongin et al. 2014).

**NAD identification and annotation**

We used the same workflow for NAD-seq data analysis as described previously (Vertii et al. 2019), except that we removed 20 NAD peaks that are less than 50 kb long (totaling 0.74 MB). Because there are 624 peaks totaling 845 Mb in the F121-9 NAD-seq data, this represents 0.087% of the NAD nucleotides. We used version 1.6.1 of NADfinder for NAD identification in this manuscript. To identify allele-specific NADs, first we performed a single alignment to the N-masked mouse genome mm10 using bowtie2/2.3.2 and then assigned the reads to one of the two genomes based on the SNP found in the
masked positions using SNPsplit_v0.3.2 for allele-specific splitting of alignments (Krueger and Andrews 2016).

**Overlap analyses**

Nucleotide-level overlap analyses of F121-9 NADs with cLADs, ciLADs (Peric-Hupkes et al. 2010), MEF NADs (Vertii et al. 2019), and H3K27me3-enriched domains (GSM2416833; (Cruz-Molina et al. 2017); GSM1621022; (Delbarre et al. 2017)) were performed using GenomicRanges (Lawrence et al. 2013) as described in detail in Vertii et al. 2019. These nucleotide-based overlap analyses in some cases generated small overlapped regions, such that single genes would end up with both Type I and Type II designations, or both MEF-specific and F121-9-specific designations. Because the biology of NADs is centered on large (~1 MB-sized) domains, we removed regions <50 kb in length from overlap analyses of Type I and II NADs and from cell-type-specific NADs to avoid these confounding designations. GO enrichment analyses of conserved and cell type-specific NADs derived from the overlap analysis were performed using ChIPpeakAnno (Zhu et al. 2010; Zhu 2013). mESC H3K27me3-enriched domains were identified based on H3K27me3 ChIP-seq data (GSM2416833; Cruz-Molina et al. 2017) using RSEG (v0.4.9) with 20 iterations for Baum training. MEF H3K27me3-enriched domains were obtained from GSM1621022 (Delbarre et al. 2017). FPKM values based on MEF RNA-seq data were obtained from GSM1621026 (Delbarre et al. 2017) and GSE90894 (Chronis et al. 2017).
FPKM values from mESC RNA-seq data were obtained from GSM1418813 (Lowe et al. 2015). Calculations of the statistical significance of pairwise comparisons were performed using Welch’s t-test in GraphPad Prism.

The NADfinder software is available at:
https://urldefense.proofpoint.com/v2/url?u=https-3A__bioconductor.org_packages_release_bioc_vignettes_NADfinder_inst_doc_NADfinder.html&d=DwIFAw&c=WJBj9sUF1mbpVIAf3biu3CPHX4MeRjY_w4DerPlOhmQ&r=JqQ8_Clm34xp32rT3DzotqsofamUUyNmo3M4_tIEI&m=Lq6n57MH0XVDsayaTs25TVTysYxezReg6cHQXKhVNk&s=BG-jkVe3qQRszk64IZLOGYCqyYe-h9NoghI0rl8l1bM&e=

We calculated Jaccard indexes among NADs, cLAD/ciLAD (Peric-Hupkes et al. 2010), and F121-9 early/late replication timing (GSE95091 (Marchal et al. 2018)). The Jaccard index is the size of the intersect divided by the size of the union of two sets. The higher the Jaccard index, the higher the extent of the overlap. Boxplots and comparisons of gene densities (genes/Mb) and gene expression distributions were performed using R for statistical comparisons, p-values were calculated using Welch’s t-test.
CHAPTER III: DISCUSSION

Heterochromatin in eukaryotic cells mainly localizes at nuclear lamina and nucleolar periphery. Genomic regions frequently found at nucleolar periphery are called NADs, and genome-wide maps of NADs are available only in a handful of human and mouse cells. Spatial genome organization has been linked to gene function and expression, hence studies illuminating genome localization patterns are of essence to understand this link. We have mapped and characterized NADs in hybrid mouse ES cell line F121-9, and we anticipate that this genome-wide map will facilitate future studies of NAD dynamics during cell differentiation.

In this chapter, I will discuss the main findings, conclusions, limitations and future directions of our studies. The main results of our studies are: NADs constitute mostly repressive subcompartment in mESCs, they possess only a small set of nucleolar-specific Type II NADs, and unlike MEFs, Type II NADs in mESCs are not enriched in H3K27me3 marks. Additionally, we defined and characterized cell type-specific NADs, and our allele-specific NAD data allowed us to discern allelic differences on a whole chromosome level.

Characteristics of mouse ESC NADs

We have shown that mESC NADs, similar to previously reported human and mouse cell line NADs (Németh et al. 2010; van Koningsbruggen et al. 2010; Dillinger et al. 2017; Vertii et al. 2019; Lu et al. 2020), are generally repressive
genomic regions. As in MEFs (Vertii et al. 2019), NADs in mESCs frequently overlap with late replicating domains and LADs.

However, there are differences between F121-9 mESC and MEF NADs. NADs comprise 31% of the genome in F121-9 cells, whereas MEF NADs mapped using the same crosslinking method constitute 41% of the mouse genome (Vertii et al. 2019). A smaller genome coverage of NADs in mESCs could be due to less condensed chromatin and higher ratio of euchromatin to heterochromatin in stem cells compared to differentiated cells, as discussed in “Features of heterochromatin in ESCs” section in Chapter I.

Type II NADs, which are found only in the nucleolar periphery and not at the NL, are a smaller subset of total NADs in F121-9 mESCs compared to MEFs. 77 Mb of 845 Mb of NADs, i.e. ~9% of NADs are Type II NADs in mESCs (Fig. 2.5A), whereas 147 Mb of 1,137 Mb of NADs (~13%) are Type II NADs in MEFs (Vertii et al. 2019). Moreover, F121-9 Type II NADs exhibit higher mean gene expression levels compared to the whole genome or even non-NAD genomic regions (Fig. 2.5D), while MEF Type II NADs display lower mean gene expression levels than whole genome or non-NAD regions (Vertii et al. 2019; Fig. 4B). Hence, it seems that Type II NADs comprise a transcriptionally more active subset of NADs in mESCs compared to MEFs. A comparatively small size and transcriptionally active nature of F121-9 mESC Type II NADs can be speculated to be due to smaller PNH layer as seen by microscopy in mESCs (Savić et al. 2014). “Maturation” of perinucleolar heterochromatin in differentiated cells (Savić
et al. 2014) can potentially explain slightly more repressive nature of Type II NADs in MEFs (Vertii et al. 2019) than in F121-9 mESCs. This hypothesis can be tested by mapping NADs in cells differentiated from F121-9 mESCs (e.g. NPCs) and examining whether Type II NADs expand and become less transcriptionally active in these differentiated cells.

Interestingly, F121-9 NADs are less enriched for H3K27me3 marks compared to MEF NADs (Fig. 2.5 E-H), (Vertii et al. 2019). Our analysis of previously published H3K27me3 chromatin immunoprecipitation sequencing (ChIP-seq) dataset derived from mESCs (GSM2416833; Cruz-Molina et al. 2017) revealed 517 Mb of H3K27me3-enriched regions in F121-9 cells. Similar analysis in MEFs (Vertii et al. 2019) revealed an almost two-fold larger amount (990 Mb) of H3K27me3-enriched regions (GSM1621022; Delbarre et al. 2017). However, the enrichment of H3K27me3 in F121-9 NADs is more than two-fold less (66 Mb, Fig. 2.5E, F) compared to H3K27me3 enrichment in MEF NADs (417 Mb, Fig. 2.5G, H). Possible causes are: 1) overall lower levels of H3K27me3 mark in mESC NADs, and 2) decreased enrichment of this PTM in NADs of 2i-grown mESCs compared to mESCs cultured in serum.

The first hypothesis is based on reports of lower levels of H3K27me3 mark in ESCs than in differentiated cells (Hawkins et al. 2010; Atlasi and Stunnenberg 2017). Additionally, recent report by the Shen group showed that in mESCs, PRC2 is inhibited from accessing nucleolar compartments such as DFC and GC, and this is thought to promote high ribosome biogenesis in pluripotent cells
(Zhang et al. 2020). Thus, it is likely that nucleolar periphery would be more enriched for H3K27me3 marks in the course of differentiation.

The second possible reason for reduced H3K27me3 abundance in F121-9 NADs is based on reported differences in this PTM occurrence in 2i cultures (described in Chapter I) vs. serum cultures. We cultured F121-9 mESCs in 2i medium, thus we have used ChIP-seq dataset derived from mESCs grown in 2i (GSM2416833; Cruz-Molina et al. 2017) for our H3K27me3 enrichment analysis. 2i-grown mESCs have been shown to exhibit reduced levels of H3K27me3 at promoters of lowly expressed genes compared to serum cultures (Marks et al. 2012). The total cellular levels of H3K27me3 are similar in 2i and serum mESC cultures, hence the differences at promoters are not due to reduction in total H3K27me3 levels (Marks et al. 2012). It is conceivable, that coupled with pluripotency of F121-9 cells, 2i medium culturing condition can lead to reduction of H3K27me3 abundance in F121-9 NADs. Hypothetically, lowly expressed genes with reduced H3K27me3 occupancy in 2i-grown mESCs could in fact overlap with NADs and account for the decreased levels of H3K27me3 in F121-9 NADs. To study the contribution of mESC culturing condition to the relative enrichment of H3K27me3 in NADs, F121-9 NADs can be mapped in mESCs grown in serum medium.

Immunoblots of fractions generated during nucleolar isolation experiments revealed that nucleolar fractions had a modest enrichment (1.5-1.6-fold) for H3K27me3 protein levels compared to the total cellular extracts (Fig. 2.2C). This
slight enrichment of H3K27me3 protein levels in nucleolar fractions could be due to abundance of H3K27me3 marks at repetitive regions often found in nucleoli, such as rDNA repeats and satellite repeats (Németh et al. 2010; van Koningsbruggen et al. 2010; Dillinger et al. 2017). For example, it has been reported that satellite repeats show increased H3K27me3 occupancy in 2i culture compared to serum culture (Marks et al. 2012; van Mierlo et al. 2019). Since NAD-seq identifies only nonrepetitive nucleolus-associated regions, immunoblots are useful in assessing the proteins associated with repetitive and nonrepetitive genomic regions in nucleoli.

**Anomalies of cell type-specific NADs in stem cells**

The comparison of our F121-9 mESC NADs with previously published MEF NADs (Vertii et al. 2019) allowed us to determine the subsets of conserved NADs, i.e. NADs that are found in both cell types; and cell type-specific NADs: MEF-specific and F121-9-specific NADs. MEF-specific NADs are not associated with nucleolar periphery in F121-9 mESCs; hence, it was expected that in mESCs, genes within these regions would show higher transcriptional activity than genes within F121-9-specific NADs. Surprisingly, in F121-9 mESCs and in another mES cell line (RNA-seq data from GSM1418813 (Lowe et al. 2015)), genes within MEF-specific NADs showed similar mean RNA-seq transcript levels as genes within F121-9 specific NADs (Fig. 2.6D, F). There are two possible scenarios that could explain this unexpected result: 1) Both F121-9-specific and
MEF-specific NADs are repressed in mESCs, or 2) Neither F121-9-specific or MEF-specific NADs are repressed in mESCs.

The first possibility suggests a scenario, where regions destined to become NADs later during differentiation (e.g. in MEFs) are already repressed in mESCs, despite the lack of nucleolar association in current developmental cell stage. Hence, nucleolar association can serve to maintain, but not to initiate heterochromatin state, i.e. it can be a consequence of a region’s already repressed status. One example of repression preceding nucleolar localization is X inactivation, after which Xi becomes frequently associated with nucleolar periphery, at least partly due to CTCF/Firre IncRNA-mediated mechanism (Yang et al. 2015). It is possible that other mechanisms, such as lamina association could account for the relatively low transcriptional activity of MEF-specific NADs in F121-9 mESCs. Indeed, 40% of MEF-specific NADs overlap with cLADs (Fig. 2.5A), i.e. these overlapped regions are found to be lamina-associated in mESCs and MEFs (Peric-Hupkes et al. 2010).

The second possibility is that genes within F121-9-specific NADs are not repressed in mESCs, which results in similar mean transcript levels of cell type-specific NADs in stem cells. It has been reported that mESCs exhibit globally permissive transcriptional hyperactivity, including transcriptional activity of intergenic and intronic regions and silent repeats (Efroni et al. 2008). It is conceivable that genes within F121-9-specific NADs fail to be repressed due to transcriptional hyperactivity in mESCs.
Interestingly, in mESCs, the transcriptional activity of genes within cell type-specific NADs are higher than that of genes within conserved NADs (p<0.0001), but lower than the levels of genes within whole genome (p<0.0001) (Fig. 2.6D, F). Hence, cell-specific NADs in mESCs possess intermediate transcriptional activity. Indeed, the Jaccard analysis demonstrated that conserved NADs cluster with late replicating regions and cLADs, whereas cell-type specific NADs cluster separately (Fig. 2.6B). This suggests that nucleolar associations in conserved and cell type-specific NADs can contribute to different transcriptional outputs, cell type-specific NADs possibly rendering weaker silencing status.

It is possible that a combination of scenarios 1 and 2 described above result in similarity of the mean transcript levels of cell type-specific NADs in mESCs. In other words, in mESCs, some regions within MEF-specific NADs might be transcriptionally repressed despite lack of nucleolar association, and some regions within F121-9-specific NADs might be transcriptionally active due to transcriptional hyperactivity of stem cells. Heterogeneity of gene expression levels within cell type-specific NADs would then lead to their similar mean transcript levels in mESCs. To obtain a clearer picture of expression levels of cell-type specific NADs, elimination of the confounding cell population heterogeneity via the use of single-cell RNA sequencing would be useful.
Allele-specific NAD features in F121-9 hybrid cells

F121-9 mESC line was derived from hybrid F1 blastocysts obtained from a cross between 129/sv strain and *Mus musculus castaneus* subspecies (129 x Castaneus) (Monkhorst et al. 2008; Rivera-Mulia et al. 2018). We observed allelic differences in nucleolar signal on a whole chromosome level for Chr 10, 11 and 16 (Fig. 2.8A-C). Specifically, Castaneus genome shows more robust nucleolar signal on chromosomes 10 and 11 (Fig. 2.8A, B), whereas 129 genome has higher nucleolar read numbers on chromosome 16 (Fig. 2.8C). These differences in nucleolar reads are mostly noticeable on the centromere-proximal side of chromosomes. It is known that nucleoli-forming NORs are found close to centromeres on mouse acrocentric chromosomes (Dev et al. 1977; Suzuki et al. 1990; Kurihara et al. 1994). This led us to hypothesize that the aforementioned allelic nucleolar read differences could be due to variations in NOR-bearing chromosomes between *M. m. castaneus* subspecies and 129/sv strain.

There are five subspecies within house mouse *M. musculus* species: *M. m. domesticus, M. m. castaneus, M. m. musculus, M. m. molossinus* and *M. m. gentilulus* (Britton-Davidian et al. 2012). Based on phylogenetic data obtained from restriction analyses of rDNA (Suzuki et al. 1986) and mitochondrial DNA (Yonekawa et al. 1982), laboratory inbred strains such as 129 were derived mostly from *M. m. domesticus* subspecies. Chr 10 and 11 are frequently found to be NOR-bearing in *M. m. castaneus*, however NORs on these chromosomes are lost in *M. m. domesticus* subspecies and inbred strains derived from it, including
129 (Suzuki et al. 1990; Kurihara et al. 1994; Britton-Davidian et al. 2012). Hence, our allele-specific NAD-seq data for Chr 10 and 11 (Fig. 2.8A, B) corroborate previous findings that *M. musculus castaneus*, unlike 129/sv, frequently contain NORs on these chromosomes. Indeed, we showed via DNA FISH assay that on average, Chr 10 nucleolar contact frequencies are intermediate between Chr 12, which bears NOR in both genomes (Suzuki et al. 1990; Kurihara et al. 1994; Britton-Davidian et al. 2012), and Chr 5, which has no report of bearing NOR in either genomes (Fig. 2.8G). Our allele-specific NAD-seq data also substantiate reported occurrence of NOR on Chr 16 in 129 strain (Kurihara et al. 1994) and low frequency of NOR being detected on this chromosome in *M. m. castaneus* (Suzuki et al. 1990; Britton-Davidian et al. 2012) (Fig. 2.8C).

Intriguingly, there are reports of nucleolar dominance in interspecies hybrids in plants, e.g. a hybrid between *A. thaliana* and *A. arenosa* (Lawrence et al. 2004), and animals, e.g. a hybrid between *Xenopus laevis* and *Xenopus borealis* (Macleod and Bird 1982), where NORs of one genome are repressed by NORs of another genome (reviewed in Preuss and Pikaard 2007). However, the hybrids from crosses between strains and/or subspecies of *M. musculus* completely or partially maintained chromosomal localizations of NORs from both parental genomes (Dev et al. 1977; Suzuki and Sakurai 1992). According to allele-specific NAD-seq data (Fig. 2.8A-C), it is likely that F121-9 hybrid inherited and retained NORs on one homolog each of Chr 10 and 11 from the *Castaneus*
genome, and NOR on one homolog of Chr 16 from the 129 genome. Thus, NAD-seq can be a useful tool in genome-wide studies of nucleolar dominance and NOR evolution in interspecies or inter-subspecies hybrids.

However, we cannot exclude the possibility that the allele-specific differences in nucleolar signal are not due to distinctions in NOR-bearing chromosomes in the two genomes. For instance, there are reports of non-NOR chromosomes contacting nucleoli at nonrandom frequencies (Carvalho et al. 2001; Ragoczy et al. 2014). To validate that Chr 10, 11 and 16 indeed contain NORs, a classical method of NOR identification, in situ hybridization of mitotic chromosomes with rDNA probe needs to be performed.

Possible mechanisms of nucleolar associations

A major question regarding the mechanisms of nucleolar association remains unanswered. It is likely that no single cis or trans-acting factor is responsible for genomic associations with nucleolar periphery. As discussed in Chapter I, various factors have been implicated in nucleolar association. For example, CTCF, CAF-1, Ki-67, NPM-1 proteins, MiCEE complex, 5S rDNA and various IncRNAs were shown to promote nucleolar tethering (Yusufzai et al. 2004; Smith et al. 2014; Matheson and Kaufman 2017; Sun et al. 2017; Padeken et al. 2013; Holmberg Olausson et al. 2014; Singh et al. 2018; Fedoriw et al. 2012b; Zhang et al. 2007; Mohammad et al. 2008; Yang et al. 2015). Recent
evidence suggests phase separation as a mechanism of nucleolar assembly (Brangwynne et al. 2011; Feric et al. 2016; Mitrea et al. 2016), and possibly, of nucleolar association (Caudron-Herger et al. 2015; Vertii et al. 2019).

The existence of Type I and Type II NADs in F121-9 mESCs and MEFs (Vertii et al. 2019) poses an important question: what are the mechanisms of nucleolar tethering of these distinct types of NADs? In MEFs, disruption of liquid-liquid phase separation via hexanediol treatment led to decreased nucleolar association frequencies of Type II NADs, and to a lesser degree, the nucleolar association frequencies of Type I NADs (Vertii et al. 2019). Hence, it is possible that phase separation contributes to Type II NAD formation more than it contributes to Type I NAD assembly. It can be speculated that interactions between nucleolar RNA species and nucleolar proteins with IDR, e.g. NPM-1 or FBL, could lead to phase separation, and subsequently, to nucleolar association of Type II NADs. To investigate the role of phase separation in Type II NADs, genome-wide mapping of NADs after hexanediol treatment can be performed.

Another question regarding these two types of NADs is: how are different features of these NAD types achieved? Type II NADs have higher mean transcript levels and gene density than Type I NADs (Fig. 2.5C, D). Perhaps, association with nucleolar periphery is not enough to cause or maintain robust transcriptional repression, instead, both nucleolar and lamina associations are needed to achieve or maintain silencing. Hence, the stronger repression of Type I NADs can be explained by their frequent association with both nucleolar
periphery and nuclear lamina. One way of discerning this possibility is to disrupt LADs via knockdown of LBR, which was shown to be essential for LAD integrity (Solovei et al. 2013). The prediction would be an increase in transcript levels of genes within Type I NADs due to loss of lamina tethering.

One can envision that strongly repressed Type I NADs possess certain features that facilitate their tethering both to nucleolar periphery and nuclear lamina. One candidate for such feature is H3K9me3 mark, often enriched in constitutive heterochromatin (Saksouk et al. 2015). H3K9me3 mark is abundant in NADs (Németh et al. 2010; van Koningsbruggen et al. 2010; Dillinger et al. 2017; Vertii et al. 2019) and LADs (Kind et al. 2015). H3K9me3 has also been shown to mediate LAD-NL interactions (Towbin et al. 2012; Bian et al. 2013). To test the hypothesis that H3K9me3 promotes associations of Type I NADs with nucleolar and nuclear peripheries, H3K9 methyltransferases SUV39H1/H2 can be knocked down or chemically inhibited and changes in Type I NADs investigated. It is possible that in MEFs, Type II NADs are recruited specifically to nucleolar periphery due to their enrichment for H3K27me3 (Vertii et al. 2019). However, F121-9 Type II NADs are not particularly enriched for H3K27me3 (Fig. 2.5F). Hence, it can be speculated that H3K27me3-mediated nucleolar association of Type II NADs is a MEF-specific phenomenon.

One alternative explanation for distinct properties of Type I and Type II NADs can be hypothetical differences in their proximity to PNH layer. Type I NADs might be frequently found in this heterochromatin layer, which would
correlate with their silent and gene-poor status. On the other hand, Type II NADs could be excluded from the PNH layer. To determine whether this is true, 3D DNA FISH assay can be used: nucleoli can be marked by staining with anti-NPM-1 antibodies and perinucleolar heterochromatin can be visualized on the periphery of nucleoli by staining with anti-H3K9me3 antibodies. The distance of a few selected Type I and Type II NADs, labeled with DNA FISH probes, from the PNH layer can be measured and might be informative about the properties of these NADs.

Future studies of the role of H3K27me3 and H3K9me3 marks, and their effectors PRC2 and HP1, respectively, in nucleolar association are needed to elucidate the mechanisms of nucleolar tethering. Similarly, acute depletion experiments of candidate trans-factors of nucleolar tethering, such as NPM-1 and CAF-1 in conjunction with NAD-seq would help to determine the factors with causative roles in nucleolar association. Emerging new technologies, such as CUT&RUN (Skene and Henikoff 2017), CUT&Tag (Kaya-Okur et al. 2019) Protect-seq (Spracklin and Pradhan 2019) and tyramide signal amplification-sequencing (TSA-seq) (Chen et al. 2018) might be used as orthogonal methods to further advance and refine maps of NADs, and to study mechanisms of NAD formation.
The role of nucleolar association in developmental regulation

The existence of cell type-specific NADs in F121-9 mESCs and MEFs suggests that NADs are likely to be involved in developmental regulation of genomic regions. For example, *Pcsk6* gene, important for anterior-posterior axis establishment in early embryogenesis (Constam and Robertson 2000), is part of a NAD in MEFs, but not in F121-9 cells (Fig. 2.7F). Correspondingly, this gene has higher transcriptional activity in F121-9 mESCs (FPKM value 22.2) than in MEFs (FPKM value 6.6; Delbarre et al. 2017); thus, reinforcing nucleolar localization as a possible repressive mechanism in developmental regulation.

As discussed earlier, Type II NADs constitute a larger subset of NADs in MEFs (Vertii et al. 2019) than in F121-9 mESCs. Type II NADs in MEFs exhibit higher transcriptional activity than Type I NADs and are enriched for facultative heterochromatin mark H3K27me3 (Vertii et al. 2019); hence facultative heterochromatin at the nucleolar periphery is likely to expand during the course of cell differentiation, at least in the case of MEFs vs. mESCs. To study the role of NADs in developmental regulation, F121-9 mESCs can be differentiated into each of the three germ layers and differences in their genome-wide NAD compositions investigated.

Recent study from the Shen group showed that L1-enriched genes, frequently found at nuclear and nucleolar peripheries, are highly enriched in genes with specialized functions, such as genes with olfactory, vomeronasal and pheromone receptor activities and immunoglobulin function (Lu et al. 2020).
Of note, olfactory receptor genes were highly enriched in both F121-9 and MEF NADs (“Response to smell detection” GO term in Fig. 2.7A). The Shen group detected upregulation of immunoglobulin and olfactory receptor genes upon depletion of L1 transcripts (Lu et al. 2020). This study also showed concomitant detachment of NADs and LADs upon knockdown of L1 transcripts (Lu et al. 2020), hence it would be informative to investigate whether the differentiation potential of mESCs would be impaired upon knockdown of L1 transcripts. Olfactory receptor and immunoglobulin gene clusters relocalize from the nuclear periphery and become activated in terminally differentiated cells- olfactory sensory neurons and pro-B cells, respectively (Yoon et al. 2015; Rother et al. 2016). Therefore, it can be speculated that localization of these genes at the lamina and nucleolar peripheries serves to repress them in a developmentally regulated and cell-specific fashion.
APPENDIX A: DIFFERENTIATION OF mESCs TO NPCs

Introduction

The pluripotency of mESCs allows to differentiate them into any of the lineages of three germ layers: ectoderm, mesoderm and endoderm (O'Shea 2004). The differentiation of mESCs into neural lineage is accompanied by increased signaling of fibroblast growth factor (FGF) and Wnt pathways, and inhibition of bone morphogenetic protein (BMP) signaling (Wilson and Edlund 2001; Muñoz-Sanjuán and Brivanlou 2002; Kunath et al. 2007; Stavridis et al. 2007; Turner et al. 2014). The emergence of mESC to NPC monolayer culture differentiation protocols (Ying et al. 2003; Abranches et al. 2009) has simplified and streamlined the derivation of NPCs, which benefits basic and clinical research studies of embryogenesis and neurogenesis.

During mESC to NPC differentiation, various changes in chromatin occur. Heterochromatin consolidates and becomes more organized and pronounced in NPCs (Meshorer and Misteli 2006; Hiratani et al. 2010; Savić et al. 2014). During this transition, numerous LADs have been shown to detach from NL, and vice versa; many of regions with changes in lamina association status also exhibit changes in transcriptional activity (Peric-Hupkes et al. 2010). My goal was to derive NPCs from F121-9 mESCs and obtain genome-wide map of NADs in these NPCs in order to augment our understanding of heterochromatin localization patterns during development.
Results

Changes in morphology during mESC to NPC differentiation

The mESC to NPC differentiation protocol (Rivera-Mulia et al. 2018) I used is based on the mESC monolayer differentiation method developed by the Austin group (Ying et al. 2003) and further modified by the Bekman group (Abranches et al. 2009), where mESCs assume a neural fate upon withdrawal of 2i/LIF in serum-free and defined N2B27 medium. The cells lost their mESC-specific round colony morphology (Fig. A.1, “mESCs”) during pre-differentiation in the absence of LIF and PD0325901, an inhibitor of MAPK (Fig. A.1, “Pre-differentiated mESCs”). In the presence of RHB-A, an N2B27-based neural differentiation medium, cells flattened out, and starting day 8 of NPC differentiation, they formed rosettes that are characteristic of NPCs (Ying et al. 2003; Abranches et al. 2009) (Fig. A.1, “NPCs D1-D10”). Neural rosettes are structures with radial arrangements of neural progenitor cells that are seen in differentiating ESCs (Wilson and Stice 2006).

Figure A.1: Morphological changes during mESC to NPC differentiation.
Phase-contrast microscopy images of F121-9 mESCs (“mESCs”), which were pre-differentiated for 24 hrs (“Pre-differentiated mESCs”) and further differentiated into NPCs for 10 days (“NPCs D1-D10”). 20x magnification, scale bar 200 μm.
Figure A.1

mESCs

Pre-differentiated mESCs

NPCs D1

NPCs D2

NPCs D3

NPCs D5

NPCs D7

NPCs D8

NPCs D9

NPCs D10
Successful differentiation of mESCs into NPCs as validated by RT-qPCR and immunofluorescence assays

RT-qPCR assays using undifferentiated F121-9 mESCs and differentiated NPCs revealed the downregulation of pluripotency factor Oct3/4 and upregulation of neuroepithelial markers nestin and Sox1 in NPCs (Fig. A.2), as has been observed previously (Ying et al. 2003; Abranches et al. 2009). Additionally, neuroectoderm marker Pleiotrophin (Ptn) was markedly upregulated (Fig. A.2), which is consistent with previously reported increase in Ptn-positive cells in NPC intermediates and NPCs (Hiratani et al. 2010). Additionally, RNA-seq analysis of two replicates of NPCs revealed the upregulation of nestin, Sox1, Ptn, and downregulation of Oct3/4 (data not shown).

To verify the increase in neural marker and decrease in pluripotency factor expression during NPC differentiation at a single cell level, immunofluorescence (IF) assay was utilized. Immunostaining of F121-9 mESCs revealed that they were uniformly positive for Oct3/4 expression, and negative for nestin expression (Fig. A.3). Conversely, immunostaining of NPCs showed that majority of cells were Oct3/4-negative and nestin-positive (Fig. A.3).

These IF, RT-qPCR and RNA-seq results suggest that the differentiation of F121-9 mESCs to NPCs was successful.

After obtaining NPCs, I isolated nucleoli and processed them for deep sequencing in order to map NADs in these cells, similar to what we have done in F121-9 mESCs (see Chapter II). However, these experiments failed to yield
distinct NAD peaks (data not shown). It could be due to sensitivity of NPCs to sonication step in our nucleolar isolation method. The optimization of our biochemical nucleolar isolation method, and perhaps use of an orthogonal method is needed to obtain a genome-wide map of NPC NADs.
Figure A.2: RT-qPCR results validate successful differentiation of mESCs into NPCs.

RT-qPCR results for neural markers Sox2, Ptn and nestin, and pluripotency marker Oct3/4 in mESCs, and NPCs derived from them. Data are normalized to undifferentiated mESCs. Error bars represent standard deviations for triplicate technical measurements.
Figure A.3: NPCs show decreased expression of Oct3/4 and increased expression of nestin compared to mESCs.

*Top panels:* Oct3/4 (red) and DAPI (blue) staining in mESCs, and NPCs differentiated from them. *Bottom panels:* nestin (green) and DAPI (blue) staining in mESCs, and NPCs differentiated from them. Nestin staining is weak in NPCs but absent in mESCs. 20x magnification, scale bar 100 μm.
Figure A.3

[Images of DAPI, Oct3/4, and Merged for mESCs and NPCs, showing staining patterns for each cell type and merged images.]
Discussion

My RT-qPCR results ascertained that at the population level, NPCs had a characteristic upregulated expression of neuroectodermal markers, such as Sox1, nestin and Ptn, and downregulated expression of pluripotency marker Oct3/4. At a single cell level, I confirmed the negative staining for Oct3/4 and positive staining for nestin in NPCs, and vice versa in mESCs. Hence, it can be concluded that the differentiation was successful.

There were technical challenges of IF in NPCs: mESCs did not differentiate into NPCs on glass coverslips, hence IF could only be performed in situ on NPCs differentiated on microscopy-grade plastic dishes. The number of NPCs that remained attached to these dishes after IF staining was low, which diminished the efficiency of this method. Additionally, anti-Sox1, anti-Pax6 and anti-Ptn antibodies were not specific. Due to these challenges, I could not confirm the changes in these neural markers at a single cell level during the NPC differentiation. Single-cell RNA-seq could be an alternative and more efficient method of monitoring the outcome of differentiation at a single cell level.

Results and protocols outlined in this chapter would facilitate derivation of NPCs in future studies of heterochromatin in NPCs.
Materials and Methods

Cell culture methods

Undifferentiated F121-9 mES cells were cultured as described in “Materials and Methods” section of Chapter 2. F121-9 mESCs were differentiated into NPCs according to the National Institutes of Health (NIH) 4D Nucleome (4DN) program’s established protocol for F121-9 cells (https://data.4dnucleome.org/protocols/84b5a308-ab0d-46da-bb4b-a17113bce78b/#protocol-info), published by the Gilbert group (Rivera-Mulia et al. 2018), with slight modifications of seeding cells in pre-differentiation medium at higher density (see below). Briefly, mESCs were grown feeder-free at 3.6x10⁴ cells/cm² (2x10⁶ cells/10 cm dish) density in 2i/LIF medium. Cells were seeded in pre-differentiation medium at high density (4.4x10⁵ cells/cm², or 24x10⁶ cells/10 cm dish) onto dishes coated with 0.1% gelatin (EMD Millipore, SF008). This high seeding density was recommended by 4DN consortium’s Cell group member (Takayo Sasaki, personal communication). 24 hours later, pre-differentiated cells were dissociated using Accutase (EMD Millipore, SF006) and plated at 1.6x10⁶ cells/10 cm dish density onto 0.1% gelatin-coated dishes in RHB-A medium (Clontech, Y40001). RHB-A medium was changed every other day. NPCs were analyzed at day 10 of differentiation.
**Table A.1: Pre-differentiation medium recipe.**

<table>
<thead>
<tr>
<th>Solution (stock)</th>
<th>Cat. No.</th>
<th>Final conc.</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum-free ES (SFES) medium</td>
<td>N/A</td>
<td>1x</td>
<td>100ml</td>
</tr>
<tr>
<td>CHIR99021</td>
<td>Stemgent 04-0004-02</td>
<td>3 μM</td>
<td>30 μl</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gibco 25030-081</td>
<td>2mM</td>
<td>1ml</td>
</tr>
<tr>
<td>Monothioglycerol</td>
<td>Sigma M6145-25ML</td>
<td>1.5x10⁻⁴M</td>
<td>1.26 μl</td>
</tr>
</tbody>
</table>

Pre-differentiation, 2i and RHB-A media were kept at 4°C out of light and only the quantity needed for the day was warmed up in 37°C water bath. 2i/LIF and pre-differentiation media expire in 2 weeks, whereas SFES medium expires in 1 month.

**RT-qPCR**

Cell culture dishes with attached F121-9 mESCs and Day 10 NPCs were rinsed with DPBS, then TRIzol Reagent (ThermoFisher Scientific, 15596026) was added directly to dishes (1 ml TRIzol/10 cm dish). The lysates were pipetted up and down, collected in Eppendorf tubes and used for RNA extraction according to TRIzol manufacturer’s instructions, or frozen at -80°C until RNA extraction. Extracted RNA was column purified and DNase I digested using RNeasy mini kit (Qiagen). 1μg of RNA was reverse transcribed using SuperScript II Reverse Transcriptase (ThermoFisher Scientific, 18064014) according to manufacturer’s protocol. Random hexamer primers (Roche, 11034731001) were
used in reverse transcription. cDNA was diluted with RNase-free distilled water 1:30 and analyzed using KAPA SYBR FAST qPCR Master Mix (2X) ABI Prism (Kapa Biosystems, KK4604). The thermocycler program (95°C 3 min, 95°C 3 s, 60°C 20 s, 40 cycles) and the reaction setup were done according to manufacturer’s (Kapa Biosystems) instructions. The table below lists the primers that were used to assess the differentiation of F121-9 mESCs into NPCs; GAPDH primer set was used to normalize RT-qPCR signals.

**Table A.2: RT-qPCR primers used to validate NPC differentiation.**

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nestin</td>
<td>TGGCACACCTCAAGATGTCCCTTA</td>
<td>AAGGAAATGCAGCTTCAGCTTGGG</td>
</tr>
<tr>
<td>Oct3/4</td>
<td>GTGGAGGAAGCCGACAACAATGA</td>
<td>CAAGCTGATTGGCGATGTGAG</td>
</tr>
<tr>
<td>Ptn</td>
<td>GCAACTGGAAGAAGCAGTTTG</td>
<td>TGGAGATGGTGACAGTTTTCTG</td>
</tr>
<tr>
<td>Sox1</td>
<td>AGTGGAAAGGTCTATGTCCGAG</td>
<td>TGTAATCCGGGTGTTCCTTC</td>
</tr>
<tr>
<td>Gapdh</td>
<td>CTGACGTGCCGCTGGAGAAAC</td>
<td>CCCGGCATCGAAGGGTGAAGAG</td>
</tr>
</tbody>
</table>

**mESC and NPC immunofluorescence (IF) protocol**

_In situ NPC IF._ Ibidi 8-well plastic chamber slide (cat # 80826) was coated with gelatin for 30 min at room temperature (RT) and rinsed with DPBS once. Pre-differentiated F121-9 mESCs were plated in 8-well chamber slide at 29x10^3
cells/cm² density. Cells were differentiated on the same slide for 10 days, with RHB-A medium changed every other day. Once differentiated, cells were rinsed twice with PBS, fixed with 4% PFA for 15 min RT, then rinsed 3x with PBS.

*mESC IF.* Coverslips (Corning, 2850-22) were placed into 6-well dishes (Corning, CLS430166) and coated with gelatin for 30 min at RT and rinsed with DPBS once. mESCs were seeded at 15.8x10³ cells/cm², or 150x10³ cells/well density in these 6-well dishes with coverslips. The next day, cells were fixed in the same manner as NPCs.

*NPC and mESC IF.* After fixation, coverslips with mESCs and ibidi slides with NPCs were blocked for 45 minutes with 5% blocking serum in PBST (PBS + 0.2% Triton X-100). Then, samples were incubated with primary antibody in blocking buffer (PBS+1% BSA+0.2% Triton X-100) overnight at 4°C. Cells were washed with PBS 3x5 min, then incubated with secondary antibody in blocking buffer for 1 hr RT. After washing 3x5 min with PBS, cells were stained with DAPI (1:1,000 in PBS) for 1 min RT, and washed 3x5 min with PBS. Non-hardening mounting medium, such as ibidi mounting medium (ibidi, 50001) was used to cover wells in ibidi slide. mESCs on coverslips were mounted on slides with ProLong Gold antifade mountant (ThermoFisher, P36934), after which the slides were cured in the dark at RT for one day. Ibidi dishes were stored at 4°C, whereas the slides were kept at -20°C. The images were taken with Zeiss LSM 700 microscope.
Table A.3: Antibodies used for mESC and NPC IF.

<table>
<thead>
<tr>
<th>Epitope</th>
<th>Primary antibody</th>
<th>Secondary antibody</th>
<th>Blocking serum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct3/4</td>
<td>1:50 dilution, mouse Ab (Santa Cruz, sc-5279)</td>
<td>1:500 dilution, goat anti-mouse Ab, Alexa Fluor 594 (ThermoFisher, A-11020)</td>
<td>Goat serum (Jackson ImmunoResearch, 005-000-121)</td>
</tr>
<tr>
<td>Nestin</td>
<td>1:50 dilution, rat Ab (Santa Cruz, sc-33677)</td>
<td>1:500 dilution, donkey anti-rat Ab, Alexa Fluor 488 (ThermoFisher A-21208)</td>
<td>Donkey serum (Jackson ImmunoResearch, 017-000-121)</td>
</tr>
</tbody>
</table>
APPENDIX B: DIFFERENTIATION OF mESCs TO EpiLCs

Introduction

Naïve pluripotent mouse stem cells are derived from the inner cell mass of blastocysts at embryonic day 3.5 (E3.5) and pre-implantation epiblasts (E4.5); these pluripotent cells can differentiate into all embryonic lineages (Nichols and Smith 2009). Primed pluripotent cells are obtained from post-implantation epiblasts (E5.5-6.5); these cells are able to generate multiple lineages, however they lack naïve pluripotency factors and cannot contribute to blastocyst chimeras (Nichols and Smith 2009; De Los Angeles et al. 2015). The Saitou group pioneered an in vitro method of obtaining epiblast-like cells (EpiLCs), which represent early post-implantation epiblast stage (E5.5-E6.0), from mESCs (Hayashi et al. 2011).

EpiLCs provide an opportunity to study changes in chromatin during the transition from naïve to primed pluripotency. This transition is accompanied by early to late (EtoL) replication timing changes, subnuclear repositioning of these EtoL regions and X chromosome inactivation (Hiratani et al. 2010). Hence, it would be useful to examine whether NADs relocalize during the transition from naïve to primed pluripotency. Additionally, EpiLCs can be further differentiated into primordial germ-like cells (Hayashi et al. 2011; Kurimoto et al. 2015), which emphasizes the utility of EpiLCs in studying heterochromatin changes in multiple
differentiation routes: naïve to primed pluripotency, and during germ cell induction.

Results

Changes in morphology during mESC to EpiLC differentiation

The initial method of obtaining EpiLCs from pluripotent ESCs was developed by the Saitou group (Hayashi et al. 2011), which the Wysocka group then optimized (Buecker et al. 2014). mESCs cultured in serum-free, 2i/LIF conditions undergo differentiation into EpiLCs upon being seeded onto fibronectin-coated dishes, withdrawal of 2i/LIF and addition of FGF and 1% knockout serum replacement (KSR) to medium (Hayashi et al. 2011; Buecker et al. 2014). Accordingly, 2i/LIF-grown F121-9 mESCs (“mESCs”, Fig. B.1) were transferred onto fibronectin-coated dishes, and after the cells had attached to dishes, EpiLC differentiation medium was added (“mESCs 0 hr”, Fig. B.1). F121-9 cells flattened and formed cellular protrusions during the course of 48 hrs of differentiation (“24 hrs”, “48 hrs EpiLCs”, Fig. B.1). This morphology is similar to what had been observed as epithelial structures in previous reports (Hayashi et al. 2011; Buecker et al. 2014).
**Figure A.1: Morphological changes during mESC to EpiLC differentiation.**

Phase-contrast microscopy images of F121-9 mESCs (“mESCs”), to which differentiation medium was added (“mESCs 0 hr”) and cells were differentiated into EpiLCs for 48 hours (“24 hrs”, “EpiLCs 48 hrs”). 20x magnification, scale bar 200 μm.
RT-qPCR results indicate upregulation of epiblast markers in EpiLCs

Previous studies indicate that EpiLCs exhibit silenced naïve pluripotency genes, such as Klf4, Prdm14 and Tbx3, express early post-implantation epiblast markers such as Dnmt3a/b, Oct6 and Fgf5, and retain the expression of key pluripotency factors Oct3/4, Sox2 and Nanog (Hayashi et al. 2011; Buecker et al. 2014). I have compared the expression levels of naïve pluripotency markers Esrrb, Klf4, Prdm14 and Tbx3, post-implantation epiblast markers Fgf5 and Otx2, and pluripotency marker Oct3/4 in EpiLCs and mESCs using RT-qPCR assay (Fig. B.2). According to the protocol from the Wysocka group, EpiLC differentiation is considered successful, if at 48 hours of differentiation there are following changes in transcript levels: less than 2-fold change in Oct3/4 levels; minimum 16-fold induction of Fgf5 and Otx2; minimum 16-fold decrease in Tbx3, 8-fold decrease in Klf4 and Prdm14 (Takayo Sasaki and Joanna Wysocka, personal communication). My RT-qPCR results are within these ranges of transcript expression changes, except that Oct3/4 transcript levels were downregulated 3-fold in EpiLCs relative to mESCs (Fig. B.2). Thus, RT-qPCR assay of additional biological replicates of EpiLC differentiation experiments is needed to validate the efficiency of this differentiation process.
Figure B.2: RT-qPCR results suggest successful differentiation of mESCs into EpiLCs.

RT-qPCR results for naïve pluripotency markers Esrrb, Klf4, Prdm14 and Tbx3, post-implantation epiblast markers Fgf5 and Otx2, and pluripotency marker Oct3/4 in mESCs, and EpiLCs derived from them. Data are normalized to undifferentiated mESCs. Error bars represent standard deviations for triplicate technical measurements. Representative data from one biological replicate experiment are shown.
Discussion

I have differentiated F121-9 mESCs into EpiLCs and observed flattened morphology (Fig. B.1) and upregulation of early post-implantation epiblast markers (Fig. B.2) in EpiLCs. Further NAD-seq experiments failed to yield distinct NAD peaks in EpiLCs (data not shown). Due to these negative NAD-seq results, I did not proceed to further validate the efficiency of EpiLC differentiation. A more complete validation would require performing additional biological replicate differentiation experiments, coupled with RT-qPCR and IF assays.

Materials and Methods

Cell culture methods

F121-9 mESCs were differentiated into EpiLCs using the NIH 4DN protocol from the Wysocka group (https://data.4dnucleome.org/protocols/b7eb526b-5cbe-41b5-bea7-4b343cd1f4f3/). This protocol is based on previously published methods (Hayashi et al. 2011; Buecker et al. 2014). F121-9 mESCs were grown in serum-free 2i/LIF medium at 2x10^6 cells/10 cm dish density. Later passages of mESCs stopped growing in serum-free 2i/LIF medium, this was rescued upon addition of 1% ESC-grade FBS (ThermoFisher, 16141002) to 2i/LIF medium. F121-9 mESCs in 2i/LIF+1% FBS medium grew faster than in serum-free medium, hence the seeding density of cells in this serum-containing medium was lower.
(1x10^6 cells/10 cm dish). However, differentiation protocol steps described below were identical for mESCs grown in serum-free or 2i/LIF+1% FBS medium. 10 cm dishes were coated with 5μg/mL dilution of fibronectin (MilliporeSigma, FC01010MG) in 1X PBS for 30 min RT. F121-9 mESCs were seeded onto these fibronectin-coated dishes at 20x10^3 cells/cm^2 density in serum-free 2i/LIF medium. After 3 hrs, when >90% cells were adherent, 2i/LIF medium was aspirated, and EpiLC differentiation medium added to cells (mESCs “0 hr” time point). Fresh differentiation medium was added at “24 hrs” time point, and at “48 hrs” time point differentiation was terminated.

Table B.1: EpiLC differentiation medium.

<table>
<thead>
<tr>
<th>Solution (stock)</th>
<th>Cat. No.</th>
<th>Final conc.</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFES medium</td>
<td>N/A</td>
<td>1X</td>
<td>50 ml</td>
</tr>
<tr>
<td>Animal-free recombinant human FGF-basic</td>
<td>Peprotech 100-18B-250μg</td>
<td>12 ng/mL</td>
<td>50 μL</td>
</tr>
<tr>
<td>KnockOut serum replacement (KOSR)</td>
<td>ThermoFisher 10828010</td>
<td>1%</td>
<td>500 μL</td>
</tr>
</tbody>
</table>

This differentiation medium was stored at 4°C for 1 week.

**RT-qPCR**

RT-qPCR assay was done as described in Appendix A. GAPDH (for normalization) and Oct3/4 primer sets are listed in Table A.2. The following
primers from the 4DN F121-9 EpiLC differentiation protocol were used to estimate the efficiency of mESC to EpiLC differentiation.

Table A.2: RT-qPCR primers used in EpiLC differentiation assays.

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Esrrb</td>
<td>CAGGCAAGGATGACAGACG</td>
<td>GAGACAGCAGAAGGACTGC</td>
</tr>
<tr>
<td>Fgf5</td>
<td>AAAGTCAATGGCTCCCACGAA</td>
<td>CTTCAGTCTGTACTTCCT</td>
</tr>
<tr>
<td>Klf4</td>
<td>TGGTGCTTTGAGTTGTGG</td>
<td>GCTCCCCGTTTGGTACCTT</td>
</tr>
<tr>
<td>Otx2</td>
<td>CCACTTCGGGTATGGACTTG</td>
<td>GTCCTCTCCCTCGCTGTTC</td>
</tr>
<tr>
<td>Prdm14</td>
<td>ACAGCCAAGCAATTTGCACTAC</td>
<td>TTACCTGGCATTTTCATGTC</td>
</tr>
<tr>
<td>Tbx3</td>
<td>TTATTTCCAGGTCAGGAGATGCC</td>
<td>GGTCGTTTGAACCAAGTCCCTC</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


doi: 10.1016/j.ymeth.2012.05.001


Chromosoma 91:313–321. doi: 10.1007/BF00328227


Heitz E (1928) Das Heterochromatin der Moose. Jahrbücher für Wissenschaftliche Bot


Lochs SJA, Kefalopoulou S, Kind J (2019) Lamina Associated Domains and


Is Required for Early Mouse Development. Mol Cell Biol 21:4330–4336. doi:

replication in mammalian cell nuclei: spatially and temporally defined
replication of chromosome-specific alpha-satellite DNA sequences. J Cell
Biol 116:1095–1110. doi: 10.1083/jcb.116.5.1095


doi: 10.1002/yea.1657

Olins AL, Olins DE (1974) Spheroid Chromatin Units (ngr Bodies). Science (80-)

Mol Cell Biol 4:809–814. doi: 10.1038/nrm1225

Oudet P, Gross-Bellard M, Chambon P (1975) Electron microscopic and
biochemical evidence that chromatin structure is a repeating unit. Cell
4:281–300. doi: 10.1016/0092-8674(75)90149-x

heterochromatin. Curr Opin Cell Biol 28:54–60. doi:
10.1016/j.ceb.2014.03.001

Homolog NLP Mediates Centromere Clustering and Anchoring to the

(CENP-A) copurifies with nucleosome core particles and with histones. J
Cell Biol 104:805–815. doi: 10.1083/jcb.104.4.805

Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing
through Chromatin-Level Regulation. Mol Cell 32:232–246. doi:
10.1016/j.molcel.2008.08.022

Suz12 Is Required for Embryonic Stem Cell Differentiation. Mol Cell Biol

10.1101/cshperspect.a000638

Peng JC, Karpen GH (2009) Heterochromatic Genome Stability Requires
Regulators of Histone H3 K9 Methylation. PLoS Genet 5:e1000435. doi:
10.1371/journal.pgen.1000435

doi: 10.1038/ncb1514


PRC1 spatially constrains the mouse embryonic stem cell genome. Nat Genet 47:1179–1186. doi: 10.1038/ng.3393
inaccessible domains reveals physical properties of chromatin. Nucleic Acids Res. doi: 10.1093/nar/gkz1150


