2016-04-08

XIST and CoT-1 Repeat RNAs are Integral Components of a Complex Nuclear Scaffold Required to Maintain SAF-A and Modify Chromosome Architecture: A Dissertation

Heather J. Kolpa
University of Massachusetts Medical School

XIST AND COT-1 REPEAT RNAS ARE INTEGRAL COMPONENTS OF A COMPLEX NUCLEAR SCAFFOLD REQUIRED TO MAINTAIN SAF-A AND MODIFY CHROMOSOME ARCHITECTURE

A Dissertation Presented

By

HEATHER J. KOLPA

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

APRIL 8, 2016

CELL BIOLOGY

XIST AND COT-1 REPEAT RNAS ARE INTEGRAL COMPONENTS OF A COMPLEX NUCLEAR SCAFFOLD REQUIRED TO MAINTAIN SAF-A AND MODIFY CHROMOSOME ARCHITECTURE

A Dissertation Presented
By
HEATHER JILL KOLPA

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

The signature of the Thesis Advisor signifies validation of Dissertation content

_____
Jeanne B. Lawrence, Ph.D., Thesis Advisor

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

_____
Jeffrey Bailey, Ph.D., Member of Committee

_____
Job Dekker, Ph.D., Member of Committee

_____
Lisa Hall, Ph.D., Member of Committee

_____
Erica Larschan, Ph.D., External Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

_____
Jeffrey Nickerson, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

_____
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

April 8, 2016

Acknowledgments

I would like to thank my PI, Jeanne Lawrence, for her support (especially during a difficult fifth year while I recovered from two concussions and during maternity leave). I have benefitted immensely from Jeanne's understanding and good humor. She gave me freedom, but generously shared her unique perspective and sharp eye when I needed advice or inspiration. After difficulty finding a grad school "home," I was glad to have found one in her lab. Thank you for believing in me when I didn't believe in myself.

Thank you to all members of the Lawrence lab, past and present, for creating a fun, supportive, and collaborative working environment. Thank you to members of my thesis research advisory committee: Dr. Jeffrey Nickerson, Dr. Lisa Hall, Dr. Job Dekker, Dr. Jeff Bailey, and Dr. Melissa Moore for helpful questions and advice throughout the course of this work. Thank you to Frank Fackelmayer for providing reagents and long emails to help me sort out confusing results from experiments investigating a highly complicated protein, SAF-A.

Finally, my family and friends have been a wonderful source of love and distraction – allowing me to leave work behind during many fun-filled weekends: hiking the White Mountains, camping trips, miles of trail running, cross country skiing, dog walks, and delicious meals. Special thanks go to my wonderful husband Gabriel for patience, advice, sacrifice, many sandwiches, and critically reading portions of this thesis. Most importantly, thank you to the universe for making me a mother to my beautiful daughter, Maren. She is a source of unending joy and no doubt my most amazing accomplishment, far greater than anything written in this thesis.

Abstract

XIST RNA established the precedent for a noncoding RNA that stably associates with and regulates chromatin, however it remains poorly understood how such RNAs structurally associate with the interphase chromosome territory. I demonstrate that transgenic XIST RNA localizes in *cis* to an autosome as it does to the inactive X chromosome, hence the RNA recognizes a structure common to all chromosomes. I reassess the prevalent thinking in the field that a single protein, Scaffold Attachment Factor-A (SAF-A/hnRNP U), provides a single molecule bridge required to directly tether the RNA to DNA. In an extensive series of experiments in multiple cell types, I examine the effects of SAF-A depletion or different SAF-A mutations on XIST RNA localization, and I force XIST RNA retention at mitosis to examine the effect on SAF-A. I find that SAF-A is not required to localize XIST RNA but is one of multiple proteins involved, some of which frequently become lost or compromised in cancer. I additionally examine SAF-A's potential role localizing repeat-rich CoT-1 RNA, a class of abundant RNAs that we show tightly and stably localize to euchromatic interphase chromosome territories, but release upon disruption of the nuclear scaffold.

Overall, findings suggest that instead of "tethering" chromosomal RNAs to the scaffold, SAF-A is one component of a multi-component matrix/scaffold supporting interphase nuclear architecture. Results indicate that Cot-1 and XIST RNAs form integral components of this scaffold and are required to maintain the chromosomal association of SAF-A, substantially advancing understanding of how chromatin-associated RNAs contribute to nuclear structure.

Table of Contents

List of Tables

List of Figures

List Copyrighted Material produced by the author

Portions of this dissertation have appeared in the following publications:

Chapter II:

Jiang, J., Y. Jing, G.J. Cost, J.C. Chiang, **H.J. Kolpa**, A.M. Cotton, D.M. Carone, B.R. Carone, D.A. Shivak, D.Y. Guschin, J.R. Pearl, E.J. Rebar, M. Byron, P.D. Gregory, C.J. Brown, F.D. Urnov, L.L. Hall, and J.B. Lawrence. (2013) Translating dosage compensation to trisomy 21. *Nature.* **500**(7462):296-300.

Chapter IV:

Hall, L.L., D.M. Carone, A.V. Gomez, **H.J. Kolpa**, M. Byron, N. Mehta, F.O. Fackelmayer, and J.B. Lawrence (2014) Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell*. **156**:907-919.

License number: 3824960881945

List of Abbreviations

α-aman	Alpha-amanitin
ActD	Actinomycin
AURKB	Aurora B Kinase
ATRX	Alpha Thalassemia/Mental Retardation Syndrome X-Linked
ChIP	Chromatin immunoprecipitation
ChIRP-MS	Chromatin Isolation by RNA Purification - Mass Spectrometry
CLIP	Crosslinking immunoprecipitation
Col1A1	Collagen A1
CoT-1	C_0 (DNA concentration at time zero) * t (seconds)
DAPI	4',6-diamidino-2-phenylindole
DMSO	Dimethyl sulfoxide
DRB	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole
DS	Down syndrome
EBV	Epstein-Barr virus
EM	Electron microscopy
ES	Embryonic stem
EtOH	ethanol
FISH	Fluorescence in situ hybridization
FUS	Fused in Sarcoma
G1D	Growth 1 Daughter
GFP	Green fluorescent protein

Hek	Human Embryonic Kidney
hnRNA	heterogeneous nuclear RNA
hnRNP	heterogeneous ribonucleoprotein
iDRiP	Identification of Direct RNA Interacting Proteins
IF	immunofluorescence
iPS	induced pluripotent stem
lncRNA	long noncoding RNA
lincRNA	long intergenic noncoding RNA
MEF	Mouse embryonic fibroblast
ncRNA	noncoding RNA
NM	Nuclear matrix
NuMA	Nuclear mitotic apparatus
ORF	Open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PP1	Protein Phosphatase-1
PRC2	Polycomb repressive complex 2
RAP-MS	RNA Antisense Purification - Mass Spectrometry
RIP	RNA immunoprecipitation
RNAi	RNA interference
RNAP	RNA polymerase
RSX	RNA on the silent X

S/MAR	Scaffold or Matrix attachment region
SAF-A	Scaffold Attachment Factor A
SAFB1	Scaffold Attachment Factor B1
SSC	Saline Sodium Citrate
TAUT	Tautomycin
UV	Ultraviolet
XCI	X chromosome inactivation
Xi	Inactive X chromosome
XIST	X inactive specific transcript

Chapter I : Introduction

The complex Nucleus:

Heterochromatin and Euchromatin

The human genome contains roughly 2 meters (6.5 feet) of linear DNA, all of which fits inside the ~10-20 μM diameter cell nucleus. To make it fit, DNA is packaged into chromatin: wrapped around nucleosomes composed of histone proteins, which are further organized into complex higher-order structure that we have yet to fully understand (Gilbert et al., 2005). It is now widely appreciated that this intricate packaging not only exists to compress DNA into nuclear structure, but the packaging is a key component of genome regulation. Moreover, beginning with XIST RNA, long noncoding RNAs are thought to play a role in genome packaging. Scientists have recently discovered a vastly underestimated transcriptome containing numerous previously unknown noncoding RNAs, and whether some of these bind and regulate chromatin at specific loci is a rapidly evolving area of investigation. Work described in this thesis examines the role of noncoding chromosome-associated RNAs in chromatin structure.

Chromatin is subdivided into two types: transcriptionally active euchromatin and silent heterochromatin, which are spatially separated in a cell type-specific manner, suggestive of an epigenetic blueprint that may function to shape cellular identity (Carone and Lawrence, 2013) (Figure 1.1). Euchromatin, generally localized to the nuclear interior, contains loosely packaged active genes that are accessible to transcription factors. Euchromatin stains lightly by DAPI (4',6-diamidino-2-phenylindole), indicative

of its more “open” and mobile structure. In contrast, inactive genes are packaged into highly condensed heterochromatin, which in human cells, localizes primarily to the nuclear or nucleolar periphery. Heterochromatin is further subdivided into two types: constitutive and facultative. Constitutive heterochromatin contains sequences that are silent in all cells (i.e. centromeric satellites), and in mouse cells, these are visible by DAPI staining as “chromocenters,” round foci scattered throughout the nucleus. Facultative heterochromatin contains sequences that are silenced in a cell-type specific manner (i.e. the inactive X chromosome). These regions are generally silenced during cellular differentiation as a cell specifies its lineage (for review, see (Gilbert et al., 2005)).

Pluripotent cells have been shown to contain almost entirely euchromatin, with highly mobile “open” chromatin and relative lack of nuclear structure (Meshorer et al., 2006; Dang-Nguyen and Torres-Padilla, 2015). In contrast, differentiated cells contain a heterochromatin footprint unique to their specific cell type that is visible by various imaging techniques. In addition to compaction and relocation to the nuclear periphery, heterochromatin formation involves DNA methylation and repressive histone modifications (i.e. H3K27 methylation or H4K20 acetylation), marks that are replicated with DNA during S-phase to maintain the silent state and retain cellular identity in progeny (Hathaway et al., 2012).

Within these euchromatic and heterochromatic regions, entire chromosomes are further spatially separated into discrete territories (Lichter et al., 1988; Cremer et al., 2006) (Figure 1.1). Most striking of these is the inactive X chromosome (Xi) territory in

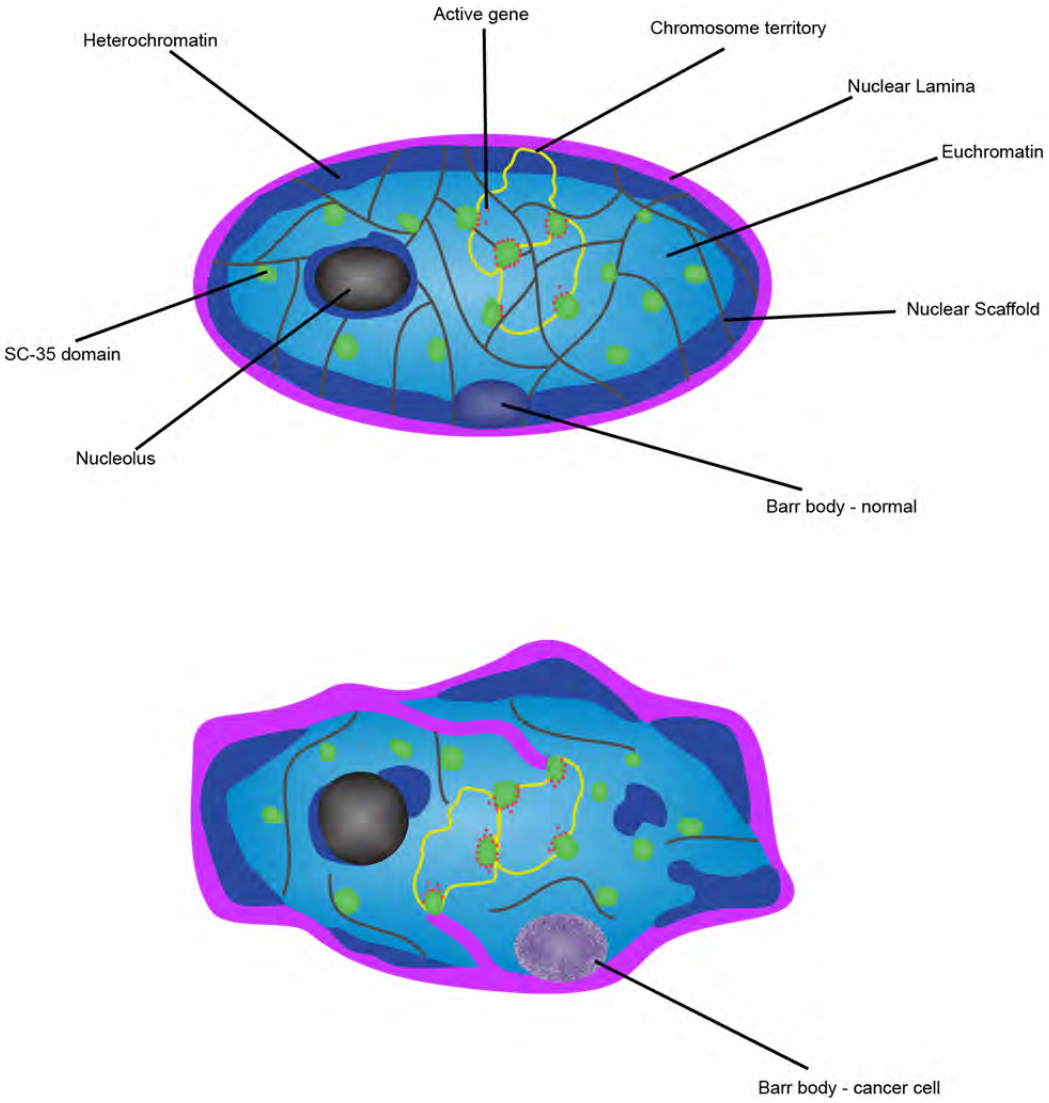

Figure 1.1 Principals of nuclear organization in normal vs. cancer cells.

Above, the Eukaryotic nucleus is well-organized into heterochromatic and euchromatic compartments. The nuclear lamina and internal nuclear matrix/scaffold organizes chromosome territories and other nuclear bodies. Below, the cancer cell nucleus undergoes several changes during transformation including altered shape, compromised nuclear scaffold, enlarged nucleoli, re-positioning of chromosome territories, and disorganized heterochromatin, including aggregation and heterochromatin/Barr body loss.

female cells, termed the Barr body, highly condensed heterochromatin that is visible by DAPI stain (Figure 1.1). More euchromatic chromosome territories can be visualized after hybridization with chromosome-specific plasmid libraries that “paint” the DNA (Lichter et al., 1988). The active and inactive X chromosome territories were found to have similar volumes, but the active X had a much greater surface area and reduced roundness factor (Eils et al., 1996). While chromosome territories contain a mix of both heterochromatin and euchromatin, in general, gene-rich chromosomes tend to localize to more interior euchromatic regions, and gene-poor chromosomes closer to the peripheral heterochromatin (Croft et al., 1999; Kupper et al., 2007; Dechat et al., 2008).

Facultative heterochromatin: X Chromosome Inactivation

X chromosome inactivation (XCI), the near-complete transcriptional silencing of one X chromosome, is a premier example of specialized facultative heterochromatin formation, as a major means of epigenetic regulation throughout the genome (for review, see (Heard, 2005)). Eutherian mammals use XCI to correct disparity in X-linked gene dosage between males (XY) and females (XX). X inactivation initiates in pluripotent female cells when XIST (X inactivation specific transcript) RNA is transcribed from a single, randomly selected X chromosome. Both the human *XIST* and mouse *Xist* genes lack an open reading frame, and instead of coding for protein, produce long noncoding RNAs (lncRNA) that spatially associate with the inactive X-chromosome DNA territory (Brown et al., 1992; Clemson et al., 1996).

The *XIST* gene is located in the 2.3 Mb region called the X inactivation center (XIC) (Chureau et al., 2002), which also contains several noncoding genes that regulate

XIST RNA (for review, see (Galupa and Heard, 2015)). Most of what we know about regulation of the *XIST/Xist* gene itself comes from studies in the mouse, which contains a 750 kb orthologous Xic region (Chureau et al., 2002). While *Xist* is the primary factor required for chromosome silencing, other genes in the Xic function in chromosome counting, choice, and regulation of *Xist* expression. For example, transcription of mouse *Tsix*, antisense to *Xist*, is sufficient to repress Xist RNA accumulation prior to inactivation and on the active X (Stavropoulos et al., 2001). *Tsix* is also implicated in chromosome choice (Lee and Lu, 1999), but the exact mechanism remains to be elucidated (Galupa and Heard, 2015). Like *Xist*, *Tsix* expression is regulated by genes in the Xic, including *DXPas34* and *Xite*, both of which are also transcribed but do not code for protein (Ogawa and Lee, 2003; Cohen et al., 2007; Pollex and Heard, 2012).

While *Tsix* is involved in *Xist* repression, *Jpx*, *Ftx*, and *Xpr* appear to contribute to its activation. However, the precise mechanism, including whether they act in *cis* or in *trans*, remains unclear (Sun et al., 2013; Barakat et al., 2014; Galupa and Heard, 2015). *Xist* is also regulated in trans by pluripotency factors including *Oct4* and *Nanog*, which respectively activate *Tsix* and repress *Xist* (Galupa and Heard, 2015).

Although well characterized in mouse, less is known about genes in the human XIC. For example, the human ortholog *TSIX* lacks the 5’ CpG island required for its function. Therefore, although it is transcribed, human *TSIX* has no repressive effect on XIST RNA expression, and may be the product of an evolutionary breakpoint (Migeon et al., 2001a; Migeon et al., 2002). Furthermore, *JPX* and *FTX* have not been shown to activate human *XIST*, and *FTX* lies outside of the human XIC (Migeon, 2014).

Once transcribed, however, both human and mouse XIST/Xist RNAs recruit a host of epigenetic factors that decorate, structurally reorganize, and silence most genes on the chromosome. Silencing is maintained in all cells of the organism by several layers of repressive chromatin modifications (reviewed by: (Wutz and Gribnau, 2007; Gendrel and Heard, 2014). XIST RNA forms a tightly-bordered nuclear territory coincident with the inactive X chromosome (Xi) DNA territory, providing a readily visualized model for RNA binding to chromatin and determining its epigenetic state.

Nuclear Organization and Structure

Nuclear bodies

Far more than just a bag of chromatin, the nucleus is highly dynamic, facilitating DNA replication, RNA transcription, splicing, and export in an efficient and coordinated manner. To support these activities, the nucleus is further compartmentalized into specialized domains or “bodies” that likely function as localized concentrations of substrates, enzymes and other factors required to facilitate efficient and carefully controlled gene expression (Dundr, 2012). For example, the euchromatic region contains nucleoli, hubs of mRNA metabolism and ribosome assembly (Boisvert et al., 2007), and splicing factor-rich SC-35 domains, which tend to localize near large clusters of active genes (Lamond and Spector, 2003) (Figure 1.1). While the protein components of nuclear bodies can be highly dynamic (Phair and Misteli, 2000), the bodies themselves resist extraction of DNA and chromatin (Nickerson et al., 1997), suggesting they are not free-floating in the nuclear space but rather tightly bound to the nuclear sub-structure

(Shopland and Lawrence, 2000; Nickerson, 2001). This suggests the existence of a scaffold that organizes chromatin and various functional nuclear compartments.

Evidence for a nuclear "matrix" or "scaffold"

Evidence for a non-chromatin scaffold or "matrix" first came to light many years ago when electron microscopy (EM) and 2-D gel analysis revealed that the nucleus contains a large amount of non-chromatin insoluble protein and heterogeneous RNA resistant to extensive biochemical extraction (Herman et al., 1978; Capco et al., 1982). This fibrogranular structure was present in all cells and tissues examined and shown to organize chromatin by attaching to the bases of DNA loops (Matrix/Scaffold attachment region (S/MAR) DNA), while also serving as a platform for DNA replication (Vogelstein et al., 1980; Nickerson, 2001). It was termed the nuclear matrix and proposed to form an architectural scaffold for nuclear domains and chromatin. In this thesis, I will use the terms "matrix" and "scaffold" interchangeably.

The nuclear matrix (NM) has been a somewhat controversial entity (Pederson, 1998; Pederson, 2000; Nickerson, 2001) in part due to the extensive manipulation of the nucleus necessary to visualize it. While protocols vary somewhat, a general approach (nuclear matrix fractionation) involves 1) Triton-X-100 detergent to remove membranes and soluble proteins, 2) 0.25M ammonium sulfate to remove histones, 3) DNAse I digestion to remove 90-95% of DNA, and 4) a final high-salt wash to remove any remaining fragments. This leaves behind ~75% of nuclear RNA and insoluble protein (Herman et al., 1978; Berezney and Jeon, 1995).

The nuclear matrix contains both the largely peripheral nuclear lamina and an associated network of core filaments that stretches throughout the nuclear interior (Nickerson, 2001) (Figure 1.1). The nuclear lamina is comprised of a meshwork of polymerized intermediate filament proteins (Lamins A/C, B1, and B2) that form a nuclear "skeleton," analogous to the cytoskeleton, providing mechanical stability to the whole nucleus (Dechat et al., 2008). In addition, the lamina has been shown to organize chromatin, including controlling the position of entire chromosome territories (Malhas et al., 2007; Meaburn et al., 2007). Essential for proper cell function, more than 300 known mutations in lamin proteins result in several debilitating laminopathies that fall broadly into four categories: striated muscle disease, lipodystrophy syndromes, peripheral neuropathy, and accelerated aging disorders including Hutchinson–Gilford progeria syndrome (Worman and Bonne, 2007).

Distinct from the nuclear lamina, the internal nuclear matrix is comprised of thick irregular fibers revealed by DNAse digestion and ammonium sulfate extraction (He et al., 1990). These fibers contain a wide variety of proteins that vary by cell type (Fey and Penman, 1988). Upon further salt extraction, the majority of these proteins are stripped away leaving behind what one study termed a network of branched 10nm "core filaments" connected to the nuclear lamina (He et al., 1990). These core filaments were shown to contain heterogeneous nuclear RNA (hnRNA), further raising the possibility that RNA may serve as a structural component nuclear architecture (Fey et al., 1986; Nickerson et al., 1989; He et al., 1990).

Some fraction of the RNA component was suggested to be nascent RNAs undergoing transcription, splicing and export associated with various nuclear bodies. For example, in Epstein-Barr Virus (EBV)-infected cells, long EBV RNA “tracks” remain visible after nuclear matrix fractionation. This suggested that these nascent pre-mRNA transcripts attach to some underlying structure, consistent with the concept of a nuclear scaffold that plays an important role in RNA processing and/or nuclear export (Lawrence et al., 1989; Xing and Lawrence, 1991). A larger component of this RNA was identified as high-molecular weight hnRNA. RNAse digestion of this RNA led to release of hnRNA associated proteins (hnRNPs), leaving what appeared to be largely hollow nuclei (Herman et al., 1978; Nickerson et al., 1989), which may suggest that the hnRNA provides structural support to the matrix. However, fueling the controversy, another study suggested incubation at 37°C after RNAse treatment retained the hnRNPs in the nucleus (Martelli et al., 2002), thus challenging the concept that an RNA component is required for nuclear structure. Today, the controversial concept of the nuclear matrix and RNA’s role in nuclear structure remains essentially unresolved (Razin et al., 2014), however emerging technologies will enable future studies to elucidate the crucial proteins and nucleic acids comprising nuclear structure. Studies described here advance this fundamental question of cell biology in new ways, moving the field past the long standing impasse.

Cancer effects on nuclear organization

The nuclear matrix is also implicated in tumor progression. Cancer cells commonly have vastly altered nuclear structure including irregular nuclear shape,

formation or loss of heterochromatin aggregates, and loss of the Barr body (Zink et al., 2004) (Figure 1.1). Some evidence implicates deletion or mutation in lamina proteins, as the nuclear lamina is primarily responsible for nuclear shape and connecting heterochromatin to the nuclear periphery (Zink et al., 2004), however changes to the overall protein composition of the nuclear matrix may also impact nuclear shape (Konety and Getzenberg, 1999).

Furthermore, numerous nuclear matrix proteins have been shown to have altered expression levels in cancer cells (Fey and Penman, 1988; Getzenberg et al., 1991) for review: (Konety and Getzenberg, 1999). In general, tumor grade is inversely proportional to the number of nuclear matrix proteins, consistent with decreased heterochromatin content and increased transcriptional activity (Konety and Getzenberg, 1999; Barboro et al., 2012). However, in de-differentiated prostate cancer cells, despite an increase in the copy number and expression level of NM proteins, fewer proteins bound to S/MAR DNA, resulting in larger DNA loops and overall loss of chromatin structure (Barboro et al., 2012). Preliminary work has been done to identify specific nuclear matrix proteins that may serve as biomarkers for certain cancers, and one such test, NMP22, is currently used to detect bladder cancer (Konety et al., 2000; Kibar et al., 2006). While this is promising, further study is needed to develop similar tests for more cancers.

In addition to altered expression levels of NM proteins, the conserved positioning of chromosome territories (with more gene-poor chromosomes localized toward the nuclear periphery) is often lost in cancer cells (Figure 1.1), supporting that changes in NM composition compromise nuclear organization and structure (Cremer et al., 2003).

These findings further support the hypothesis that nuclear organization is fundamental for controlling the epigenome, and essential for keeping cancer at bay. While the goal of this thesis was not to study cancer per se, unanticipated results encountered in the course of this work have significant implications for fundamental cancer biology, particularly in regard to a compromised relationship of RNA to the nuclear scaffold.

Noncoding "chromosome-associated" RNA

Chromosomal RNAs in genome regulation

Noncoding RNAs (ncRNAs) are increasingly recognized as essential players in genome regulation. Diverse types of ncRNAs have been shown to participate in activities such as gene silencing, cell cycle regulation, stem cell pluripotency, immunity, and metastasis, (Rinn et al., 2007; Gupta et al., 2010; Guttman et al., 2011; Gomez et al., 2013; Tripathi et al., 2013), but their mechanisms of action are diverse and incompletely understood. Some ncRNAs are hypothesized to function by binding to chromatin. XIST RNA is the best-characterized ncRNA that functions by binding to chromatin, raising the possibility that other ncRNAs may also. However, we still lack a clear understanding of how XIST RNA localizes, and knowledge of this will not only enhance understanding of how this ncRNA induces the structural transformation and silencing of an entire chromosome, but also will provide insight into how other ncRNAs may interact with chromatin.

Below I will summarize the current understanding of two types of "chromosomal-RNAs" that are the focus of this thesis: XIST RNA, which is well established to bind and

function in heterochromatin (Chapters II and III), and repeat-rich sequences in CoT-1 RNA, which we show bind to euchromatin, and results suggest likely function there (Chapters IV and V).

XIST RNA

As described above, XIST RNA is transcribed from the inactive chromosome and physically “paints” the chromosome territory, bringing about its transcriptional silencing. It was previously demonstrated that two populations of XIST RNA accumulate on the Xi: a smaller concentration of intron-containing pre-mRNA that is continuously transcribed and spliced, and a much larger concentration of mature, functional RNA that is structurally associated or embedded within the chromosome architecture (Clemson et al., 1996). Mature XIST RNA has a half-life of about 5 hours following transcriptional inhibition (Clemson et al., 1998), whereas several pre-mRNAs tested are undetectable after 1-3 hours (Hall et al., 2014). Mature XIST RNA remains unperturbed in cells treated with RNAse H, which specifically targets RNA/DNA hybrids (Clemson et al., 1996). This supports the hypothesis that the majority of XIST RNA is not hybridized to DNA, but it is likely tethered by specific proteins (Clemson et al., 1996). Finally, the XIST RNA territory remains unperturbed after removal of chromatin and DNA (nuclear matrix fractionation) (Clemson et al., 1996). This suggests that the proteins responsible for localizing XIST RNA to the Xi territory are components of the insoluble nuclear scaffold or matrix.

Despite substantial progress, we still lack a complete understanding of the proteins that bind XIST RNA including the chromatin modifiers it recruits to initiate the

silencing cascade and those that localize it to the chromosome. Several studies have catalogued a host of histone modifications that can be visualized on the inactive chromatin by light microscopy (e.g.: (Jeppesen and Turner, 1993; Costanzi and Pehrson, 1998; Boggs et al., 2002; Plath et al., 2003), and several non-histone proteins have been observed enriched on the inactive chromosome territory, and therefore are hypothesized to play a role in inducing these chromatin modifications (Helbig and Fackelmayer, 2003; Blewitt et al., 2008; Baumann and De La Fuente, 2009; Pullirsch et al., 2010). However, the sequence in which these modifications occur remains unresolved. For example, a prior report concluded that Xist RNA first directly binds PRC2 to induce H3K27 tri-methylation, but this is now debated (for review, see (Cerase et al. (2015)).

The development of RNA FISH (Lawrence et al., 1989) and the initial demonstration that XIST RNA uniquely coats its interphase chromosome territory in *cis* (Lawrence et al., 1989; Clemson et al., 1996) became the mainstay of many X inactivation studies, however this approach does not typically provide information on a molecular scale. This approach may miss highly transient interactions, or occasionally an epitope may be masked on the condensed Barr body. Hence, other approaches are needed to determine the direct molecular interactions of the functional XIST RNA *in vivo.* However, as will be discussed in more detail below, extraction-based biochemical methods have also proven insufficient, hindered in part by the large size of Xist/XIST RNA (14- 17kb) (Brockdorff et al., 1992; Brown et al., 1992), its association with the insoluble nuclear structure, and the transient developmental window during which X inactivation is initiated in pluripotent cells (for review, see (Heard and Disteche, 2006)).

Highlighting this difficulty, it was previously thought that BRCA1 tumor suppressor protein directly bound XIST RNA and anchored it to the chromosome (Ganesan et al., 2002). The authors used RNA immunoprecipitation (RIP), then a relatively new technique, to find what appeared to be a direct interaction between XIST RNA and BRCA1. To isolate XIST RNA, they used formaldehyde to crosslink nuclei, immunoprecipitated BRCA1, and adapted the chromatin immunoprecipitation (ChIP) protocol to isolate RNA instead of DNA. They performed RT-PCR and detected XIST RNA in the BRCA1 fraction. In addition, they reported complete overlap of BRCA1 signal with that of the whole XIST RNA territory in many cells, by combined immunofluorescence (IF) and fluorescence in situ hybridization (FISH). However, this was later shown to reflect a general relationship of BRCA1 to replicating satellite heterochromatin rather than specifically Xi or XIST RNA (Pageau and Lawrence, 2006; Pageau et al., 2007a; Pageau et al., 2007b; Xiao et al., 2007). Hence molecular cytological analyses can lead to misinterpretations, such as antibody specificity or bright signals in one fluorescence channel “bleeding through” to create a dim signal in another, and RIP-based approaches can be fraught with technical artifacts.

More recently, three independent groups developed techniques to pull Xist RNA out of the nucleus and characterize associated proteins using mass spectrometry. These groups produced three very different lists of putative Xist interacting proteins. While this provides a starting point for future studies, as discussed in detail below, it also further highlights challenges of studying chromosome-associated RNAs like Xist, particularly

when using extraction-based techniques. The nature of XIST RNA's relationship to its parent chromosome is the subject of Chapters II and III.

The repeat genome and CoT-1 RNA

It has long been recognized that genomes of most organisms contain more than protein-coding sequences. Beginning with the observation that different organisms contain differing amounts of DNA, it was determined that the C-value, the amount of DNA per haploid genome, varies >80,000 fold between eukaryotes (Pagel and Johnstone, 1992). This was followed by the surprising discovery that genomic size does not correlate with organismal complexity. This apparent contradiction, coined the C-value paradox (Thomas, 1971; Gregory and Hebert, 1999), led to various hypotheses explaining the existence of up to 98% of the human genome which does not code for protein. Early attempts to explain this paradox included the proposition that non-coding human DNA is simply "junk" (Ohno, 1972; Pagel and Johnstone, 1992) resulting from genetic drift, chromosome segment duplications, or maintenance of extinct genes. These hypotheses supported the view that evolutionary pressure led to increased genome size over time (Gregory and Hebert, 1999), and DNA was only eliminated when the metabolic cost associated with its replication became detrimental to the organism (Pagel and Johnstone, 1992).

Among the junk are intronic sequences, which on average comprise the vast majority of pre-mRNA sequence. Why introns would be so large, or even exist in such prevalence, is not firmly understood. Most introns and intergenic DNA is replete with repetitive sequences, and more than 50% of the human genome attributed to junk DNA is

comprised of repetitive sequences that include transposon-derived short and long interspersed nuclear elements (SINEs and LINEs), simple sequence repeats (e.g., $(A)_n$, $(AT)_n$ or $(CGG)_n$), and tandemly repeated sequences, such as centromeric satellites and telomeres (Lander et al., 2001). These repeat sequences are part of CoT-1 DNA, the fraction of the genome that anneals most quickly after melting. CoT-1 is calculated by C_0 (the DNA concentration at time zero) * t seconds. Abundant, highly repetitive DNA reanneals faster than unique sequences and therefore has lower CoT values (Britten et al., 1974).

Despite an overall lack of evidence for the junk hypothesis, (Gregory and Hebert, 1999) these views have been widely adopted (Germain et al., 2014) and have had a lasting impact on the direction of biological research. Repetitive sequences in CoT-1 are routinely screened out of biochemical analyses by repeat masker, and only a few studies propose biological importance for interspersed repeats. For example, studies of X-autosome translocations and *Xist* transgene insertions suggest that genes in LINE-rich regions may be silenced better than those in LINE-poor regions. This led to the suggestion that LINE elements potentially function as DNA signals to facilitate the spread of silencing across the X chromosome (Lyon, 1998), and this was supported by the observation that LINE-1 (L1) DNA elements are enriched on the human X chromosome compared to the genomic average (and X escape region) (Bailey et al., 2000), and to any individual chromosome ((McNeil et al., 2006) reviewed in (Hall and Lawrence, 2010)). However, the role of L1 in X inactivation remains speculative, for example another study found, for a very different reason, that XIST RNA preferentially

associates with gene-rich (and LINE-poor) regions during its initial spread across the chromosome (Engreitz et al., 2013).

Recently, the ENCODE project helped to turn the notion of junk DNA on its head by ascribing a putative function to 80% of DNA in the human genome (Consortium, 2012; Germain et al., 2014). These claims are controversial, in large part due to the definition of "function" (Graur et al., 2013). However a distinction can be made between true "junk" (DNA with no biological relevance) and DNA that may not engage in relevant biochemical activities, however still cannot be ignored by biomedical research (Germain et al., 2014).

As a result of their historical label as genomic junk, most investigators assume most interspersed repeats and other CoT-1 sequences are packaged into transcriptionally silent heterochromatin, and thus few studies have considered the possibility that repeat elements may be transcribed into functional RNAs. However, it is possible that expression of repetitive RNAs has been overlooked by numerous biochemical analyses. In addition to widespread use of repeat-masker, RNA extraction protocols designed for cytoplasmic RNAs may fail to pull repetitive RNAs of the cell, as these sequences may be more likely to form complex and relatively insoluble structures. As described above, previous studies have suggested substantial nuclear RNA remains after extensive biochemical extraction removes most DNA and protein (Fey et al., 1986).

To circumvent the limitations of extraction-based and bioinformatic approaches, the potential expression and distribution of repeat RNAs may best be revealed *in situ*. In fact, one such study did find a suggestion that a subset of L1 elements are expressed

specifically on the Xi during initiation of XCI in mouse ESCs (Chow et al., 2010). Deep sequencing has also vastly improved our knowledge of the transcriptome, and there are now several examples of individual transposons impacting the promoter or transcription factor binding sites of nearby genes (reviewed in: (Faulkner and Carninci, 2009)), setting a precedent for functional RNAs transcribed from transposon-derived repeats.

There is also long-standing evidence that SINEs are expressed in response to stress-stimuli such as heat shock or viral infection (Liu et al., 1995; Li et al., 1999). Early studies focused on whether stress can induce transposons to become mobile and jump around the genome, potentially causing damaging mutations (Capy et al., 2000). More recently, SINEs (human Alu, and mouse B1 and B2 RNA) transcribed by RNA Pol III during the stress response were shown to broadly inhibit cellular gene transcription by specifically binding to Pol II (Allen et al., 2004; Espinoza et al., 2004; Mariner et al., 2008). These studies provide evidence that interspersed repeat RNAs can regulate genome-wide transcription and set a precedent for interspersed repeat ncRNAs with a biological function.

This implies that any perturbation impacting the balance of repeat RNA expression could have physiological repercussions for the cell. Scientists routinely stress cells while investigating a variety of hypotheses, including, but not limited to drug treatments, transient protein overexpression, and osmotic stress. These treatments are assumed to elicit a stress-response, however it is usually ignored. In Chapter IV of this thesis, I provide evidence that transient plasmid transfections of GFP-tagged proteins can significantly

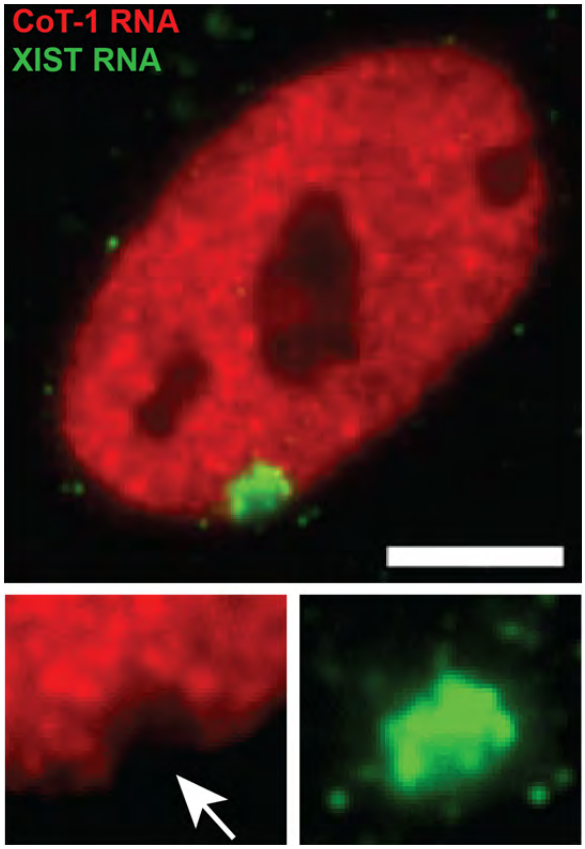

Figure 1.2 Xist and Cot-1 RNAs localize to heterochromatin and euchromatin, respectively.

XIST RNA (green) and CoT-1 RNA (red) localization in the nucleus is mutually exclusive. XIST RNA localizes to the heterochromatic Barr body, in a CoT-1 RNA hole (arrow).

increase transcription of individual repeat families, including both SINEs and LINEs. This has potential consequences for both gene expression and nuclear organization.

Many FISH protocols use the CoT-1 fraction as an unlabeled competitor to block hybridization of genomic probes to repeats. Several years ago our lab developed an *in situ* assay using labeled CoT-1 DNA as a probe under non-denaturing conditions to show that the inactive X chromosome was transcriptionally silent (Hall et al., 2002a). The CoT-1 RNA signal was excluded from the Barr body coated by XIST RNA (Figure 1.2), but not the active X chromosome. This and other findings led to the hypothesis that silencing of repeat-rich hnRNA occurs in parallel to chromosome condensation and formation of the heterochromatic Barr body (Hall and Lawrence, 2010). In addition, hybridization to CoT-1 RNA produced a signal that associates exclusively with euchromatin, and its surprising abundance led us to question whether it represented byproducts of RNA metabolism as was widely assumed. The nature of this abundant repeat RNA and its relationship to interphase chromatin is the focus of Chapters IV and V.

Anchoring chromosomal RNA

Evidence for anchoring by SAF-A/hnRNP U

XIST RNA's association with the insoluble nuclear scaffold has made it difficult to identify the proteins that anchor it to its parent chromosome in *cis*. Insoluble scaffold proteins, particularly those which can bind DNA and/or RNA, are potential chromosomal RNA anchors. Scaffold Attachment Factor-A (SAF-A/hnRNP U) was separately identified as an insoluble nuclear scaffold protein (Romig et al., 1992) and as an hnRNP

protein involved in pre-mRNA processing (Kiledjian and Dreyfuss, 1992). SAF-A is an abundant, ubiquitous protein that binds to all interphase chromosomes, but it is enriched on the Xi (Helbig and Fackelmayer, 2003). SAF-A was shown to bind RNA through its C-terminal RGG box (Kiledjian and Dreyfuss, 1992; Thandapani et al., 2013), which is required for its Xi enrichment (Helbig and Fackelmayer, 2003). Cross-linking immunoprecipitation (CLIP) experiments suggest SAF-A binds a variety of RNAs including Xist (Hasegawa et al., 2010) and many other nuclear noncoding RNAs, and it is thought to play a role in RNA metabolism (Huelga et al., 2012; Xiao et al., 2012). SAF-A was found to bind DNA through its N-terminal SAP (SAF-A/B, Acinus and PIAS) domain, and it has particularly high affinity for AT-rich S/MAR DNA (Romig et al., 1992; Gohring and Fackelmayer, 1997). Hence, since SAF-A can potentially bind both RNA and DNA (Fackelmayer et al., 1994), it alone could provide “the bridge” between XIST RNA and Xi DNA.

A significant finding by Hasegawa et al. (2010) showed that SAF-A knockdown is sufficient to completely release Xist RNA from the interphase chromosome, primarily in mouse Neuro2a neuroblastoma cells. This has led to the perception that SAF-A is the single scaffold factor required to anchor Xist RNA to chromatin (e.g.: (Nakagawa and Prasanth, 2011; Tattermusch and Brockdorff, 2011)). While the study by Hasegawa et al. (2010) implicates SAF-A, how human XIST RNA interacts with its parent chromosome remains unknown, and the potential role of human SAF-A has not been investigated. Several key differences exist in the processes of XCI between the two species (Chow et al., 2005), and the XIST/Xist RNA sequence and XIC region are not well conserved

(Chureau et al., 2002). Human XIST RNA does not localize properly to the human X in a mouse cell, suggesting differences in the protein(s) that localize the RNA (Clemson et al., 1998). In addition, evidence suggests that multiple regions of the long Xist RNA transcript contribute to its localization (Wutz et al., 2002). In light of this, the complexity of the nuclear scaffold, and the functional importance of proper XIST localization, it would be surprising for the RNA's localization to depend on a single anchor protein.

Model: multiple proteins anchor XIST RNA to the nuclear matrix/scaffold

Prior work from our lab provides additional evidence supporting a model for multiple XIST RNA anchors. Given that biochemical techniques had not yielded significant insight into XIST RNA anchor proteins, we pursued an approach to manipulate XIST RNA localization *in situ* based on the fact that XIST RNA releases from the chromosome during mitosis (Clemson et al., 1996; Hall et al., 2009). Hall et al. (2009) tested whether XIST RNA would release in interphase if chromatin proteins were phosphorylated as they are during cell division. They found that AURKB suppression during prophase led to XIST RNA retention during metaphase. However, while activation of AURKB during interphase promoted XIST RNA release, it was not reliable or consistent. Thus, this work supported a model in which XIST RNA has multiple anchor points: perturbation of a single anchor (an AURKB substrate) did not release the RNA during interphase, but retention of that anchor retained XIST RNA on the mitotic Xi. As further explored in Chapter III, I hypothesize that an "AURKB-dependent" anchor, potentially together with multiple components of the nuclear scaffold (such as SAF-A), likely ensure the RNA's strict localization to only its parent chromosome, and

reliable maintenance of silencing during development and beyond. In Chapter V, I investigate the hypothesis that SAF-A is involved in anchoring CoT-1 chromosomal RNA to its parent chromosome in *cis*.

Approaches to identify RNA interacting proteins

The following section was written as part of a review article that is currently accepted pending minor revisions at *The Journal of Cell Biology*. (Kolpa, H.J. and J.B. Lawrence. In search of bona fide Xist/XIST RNA binding proteins: mixed messages from extraction-based approaches).While work in this dissertation relies on *in situ* for the bulk of the experiments, it is important to consider other techniques that may provide an alternative means for investigation, especially when used in parallel as a means to substantiate findings.

Protein-centric methods to investigate direct RNA-protein interactions

It is widely understood that the majority of nuclear ncRNAs function by interacting with proteins. To begin to characterize the mechanisms through which these RNAs function, we need tools to identify their protein binding partners. However, researchers have had only partial success developing biochemical methods to identify direct nuclear RNA-protein interactions, and in particular to discern actual *in vivo* functional interactions with high fidelity (Ule et al., 2005; Keene et al., 2006). Many extraction-based techniques were developed for cytoplasmic RNAs, thus the challenge is especially difficult for structurally embedded nuclear RNAs like XIST. Importantly, Hasegawa et al. (2010) showed that sonication increased extractability of Xist RNA, a

finding replicated by subsequent studies. However, this insolubility remains a significant consideration for anyone looking to pull down Xist RNA.

The first widely-used techniques to identify RNA binding proteins were “protein-centric.” Researchers chose candidate protein(s) thought to bind an RNA-of-interest and immunoprecipitated the protein with its associated RNA(s). These techniques represented a significant technical advance. However as suggested above, they also have several potential pitfalls. Since RNA is inherently sticky, even the most carefully controlled experiments can be compromised by background, creating poor signal to noise ratios. In particular, RIP protocols may or may not involve chemical crosslinking with formaldehyde, and this must be carefully optimized in an assay-dependent manner. With chemical crosslinking, large chemical bridges form not only between proteins immediately contacting the RNA, but also between multiple adjacent proteins (Ule et al., 2005). RNA binding proteins often exist in larger complexes, or, in the case of XIST RNA, in a dense complex structure or inter-connected “scaffold”, and this can lead to erroneous identification of putative interactors that may be quite indirect. However, with no crosslinking, transient interactions are lost (Riley and Steitz, 2013). Importantly, evidence suggests that without crosslinking, proteins and RNAs can re-associate after cell lysis and during processing, further complicating analysis and creating the risk for false positives (Mili and Steitz, 2004; Riley et al., 2012). Despite these limitations, RIP may be more useful when looking at an entire RNA molecule as opposed to small fragments or binding sites (Darnell, 2010), and this is particularly relevant for a long RNA molecule like XIST.

The caveats described above showcase the need for methods with fewer artifacts and increased reproducibility. This led to development of CLIP, which replaced formaldehyde crosslinking with ultraviolet light (UV). CLIP was considered superior because UV does not crosslink proteins to other proteins. In addition, the covalent bond forms only between proteins and nucleic acids within angstroms of each other, which both increases specificity and allows for greater stringency during purification (Ule et al., 2005). UV crosslinking was not adopted earlier because scientists thought it was too inefficient, capturing only 1-5% of all protein-RNA interactions. In addition, the crosslink is irreversible, a potential barrier to RT-PCR identification of the RNA. However these barriers have proven negligible as the technique became more widely adopted and was shown to significantly improve signal:noise. Several studies have successfully used CLIP to identify proteins bound to specific RNAs (reviewed in (Darnell, 2010)).

Protein-centric studies of Xist RNA binding proteins

After BRCA1 was misidentified as an Xist RNA-interacting protein required for the RNA's localization, six years passed before another protein was reported to bind Xist RNA, when Zhao et al. (2008) used RIP to test whether Xist RNA directly bound proteins in PRC2. They did not perform a crosslinking step, hoping to avoid fixation artifacts, which can include reduced lysis efficiency, sequence biases, increased background, and irreversibility (Keene et al., 2006). Zhao et al. (2008) detected Xist RNA by RT-PCR after pull down of Ezh2 and Suz12 in undifferentiated mouse ES cells but not in differentiated MEFs, suggesting PRC2 is primarily bound to Xist RNA during

initiation of XCI. It should be noted that PRC2 also pulled down Xist RNA in a subsequent CLIP study by the same lab (Sarma et al., 2014), however, other labs have attempted to replicate this finding and failed ((Chu et al., 2015b; McHugh et al., 2015) for review: (Cerase et al., 2015)). Hence, the jury remains "out" on the important question of whether a key mechanism of Xist RNA function is direct binding of PRC2 to initiate the silencing cascade. Subsequent CLIP studies have since identified three other proteins that appear to bind Xist RNA: hnRNP U/SAF-A, YY1, and ATRX (Hasegawa et al., 2010; Jeon and Lee, 2011; Sarma et al., 2014), bringing the total number of putative Xist RNA interacting proteins to four.

RIP and CLIP both suffer from the limitation that RNA identification relies on pull down of specific *proteins*. A more impartial investigation of RNA binding proteins would identify proteins after pull down of a specific *RNA*. Below, I will examine three recently developed RNA-centric techniques, used to define what each of three studies concludes is a definitive list of proteins that directly bind to Xist RNA *in vivo*.

ChIRP-MS, RAP-MS and iDRiP: alphabet soup, or reliable identification of specific RNA binding proteins?

The three RNA-centric techniques recently used for Xist RNA differ in substantial ways but share an overall approach: pull down of the RNA using antisense oligonucleotides, followed by mass spectrometry to identify associated proteins. This was a logical but significant extension of RNA antisense purification (RAP) experiments developed to analyze lncRNA association with specific chromatin loci, as done for Xist RNA (Engreitz et al., 2013).

Comprehensive identification of RNA binding proteins by mass spectrometry (ChIRP-MS) (Chu et al., 2011), involves chemically crosslinking cells with formaldehyde or glutaraldehyde (which must be optimized according to the properties of an individual RNA) prior to lysis, sonication and hybridization with biotinylated antisense DNA oligonucleotides, followed by mass spectrometry (Chu et al., 2015a). The authors suggest that ChIRP-MS avoids several pitfalls of previous techniques, including formation of non-specific protein-RNA interactions during purification. ChIRP uses a collection of short (~20-mer) oligonucleotide probes modeled after those used for single-molecule RNA FISH (Chu et al., 2011), which tile across the length of the RNA, excluding repetitive or non-unique sequences. Tiling probes were required to pull down sufficient quantities of RNA, with the advantage that they require no prior knowledge of the RNA sequences accessible to hybridization *in vivo*, and they capture all or most of the transcript, even when fragmented by sonication (Chu et al., 2015a).

Chu et al. (2015b) used different cell conditions/types with either endogenous or transgenic *Xist*, to study proteins associated with either random or imprinted XCI, during both the initiation and maintenance phases. By MS, they generated a list of 62 proteins pulled down with Xist RNA in all four cell types, and an additional 19 enriched only in differentiated cells post-XCI. They obtained similar results from mESCs (with an inducible *Xist* transgene on chromosome 11), epiblast-derived stem cells (random XCI), and trophoblast stem cells (imprinted XCI), supporting the hypothesis that Xist RNA pulls down similar proteins in all of these distinct contexts (Chu et al., 2015b).

The authors examined a handful of candidates using siRNA knockdown and RNA FISH in mouse ES cells, finding that hnRNP K, Spen, and hnRNP U/SAF-A were required to initiate Xi silencing, but hnRNP M was not. Importantly, the authors used northern blotting to test whether hnRNP K, U, and M had an indirect role in Xist biogenesis, and found they did not. hnRNP M may be redundant with some other factor, but it will be important for future studies to further analyze whether it and the remaining 77 proteins directly bind Xist RNA, particularly the mature transcript, in relation to its function.

McHugh et al. (2015) used a second technique, RAP and mass spectrometry (RAP-MS), to identify ten proteins that pulled down with Xist RNA in the same male ES cells used by Chu et al. (2015b) which contain an *Xist* transgene on chromosome 11. Similar to ChIRP-MS, RAP-MS uses antisense biotinylated oligos (90-mers) tiled across unique regions of the entire RNA sequence, and then captures and purifies the RNA complexes using streptavidin-coated beads after DNAse treatment and sonication. Using RNAi and a gene silencing assay, McHugh et al. (2015) found three of the ten proteins identified, SHARP, LBR, and SAF-A/hnRNP U are required for gene silencing; however knockdown of the remaining seven candidates (including hnRNP M) had no effect. Thus, further analysis is needed to corroborate whether these proteins indeed specifically and directly bind mature, functional Xist RNA *in vivo*, potentially with a non-essential or functionally redundant role.

RAP-MS differs from ChIRP-MS in at least two key ways: 1) UV rather than chemical crosslinking, and 2) stable isotope labeling of amino acids in culture (SILAC) to

more quantitatively assess proteins pulled down by the assay. Both of these theoretically increase specificity. As described above, UV crosslinking can provide superior specificity to formaldehyde since it captures only “zero-length” interactions. SILAC labeling allows for more quantitative comparison of peptide ratios between experimental and control samples, increasing the likelihood of discerning real from non-specific interactors. The authors also used a stricter cut-off (>3-fold enrichment compared to control) when assessing positive interactors. These differences may help to explain why the number of Xist-interacting proteins found using RAP-MS was much smaller than those generated by the other RNA-centric methods.

Minajigi et al. (2015) developed yet a third variation of the RNA-centric pull-down approach to generate yet another “Xist RNA interactome”, which again differs markedly from the other lists. Identification of direct RNA interacting proteins (iDRiP) uses UV crosslinking and antisense oligo probes to pull down Xist RNA, followed by quantitative mass spectrometry to identify associated proteins. Instead of tiling probes, the authors used nine 25-mers, spaced out across the Xist RNA transcript. They treated isolated complexes with DNAse to eliminate DNA-bound proteins, and emphasize this is an important distinction and advantage of their study (however, McHugh et al. (2015) also treat with DNAse). This approach found more than 200 proteins that were ≥2-fold enriched over background, and more than 80 that were ≥3-fold enriched, all of which the authors concluded bind to Xist RNA. Among these, SAF-A/hnRNP U, PRC2, and ATRX were identified in previous protein-centric CLIP studies, but, importantly, the latter two were not found by the other RNA-centric studies. These authors used fibroblasts

expressing endogenous Xist RNA, which may be more reliable than transgenic Xist, although Chu et al. (2015b) found little difference between the two. Minajigi et al. (2015) used two different quantitative MS techniques (spectral counting and multiplexed quantitative proteomics), however, it is striking to us that these two MS techniques found different lists of interactors, with almost no overlap between them. Furthermore, we note that the top ten candidates on both lists include abundant proteins that do not localize in the nucleus, such as the secreted proteins Collagen Alpha-2 and Fibrinogen Beta Chain. In our view, such observations raise questions about the reliability of the technique to pull down proteins that directly and specifically bind Xist RNA *in vivo*. Testing some of the more logical hits, the authors found they could reactivate genes on Xi by inhibiting multiple proteins at once, in particular DNMT1, Topoisomerase II, SMC1a, SMC3, RAD21, SMARCA4, and AURKB.

Three studies and three different lists of the "Xist RNA interactome"

While similar in concept, the differences between iDRiP-MS, RAP-MS, and chIRP-MS provide different strengths and vulnerabilities. For example, ChIRP-MS uses formaldehyde crosslinking, potentially leading to pulldown of nonspecific yet abundant networks of chromatin-associated factors rather than direct Xist-interacting factors (Chu et al., 2015b). While Chu et al. (2015b) perform their experiments in multiple cell types and XCI conditions, the other studies use only one cell type, including transgenic cells in which Xist RNA expression creates a lethal monosomy of chromosome 11, possibly eliminating factors involved in XCI maintenance. Minajigi et al. (2015) use fibroblasts expressing endogenous Xist, which does not identify factors specifically involved in XCI

initiation. These authors argue that UV crosslinking and DNAse treatment make their method more specific, however they pull down many more proteins than the other groups (Minajigi et al., 2015), and some at the top of their list are clearly non-specific, as noted above. These discrepancies are perhaps best illustrated by the table comparing the top 32 hits from each study (Table 1.1). While the majority of the ten proteins found by RAP-MS were also top hits in the other two studies, overall the lists reveal large variations in the data.

Since tiling or oligo probes would hybridize to either population of Xist RNA, a key question is whether the proteins found by each of these studies associate with Xist RNA as it is being transcribed and spliced, or with the mature RNA that stably localizes with the Xi. It was previously shown that Xist RNA is produced throughout the cell-cycle and that some Xist pre-mRNA accumulates on the Xi along with the larger accumulation of mature transcript (Clemson et al., 1996). Tiling or oligo probes could hybridize to and pull down either population of Xist RNA. Chu et al. (2015b) distinguish 30 of the 81 proteins that were also highly enriched with another ncRNA, either U1, U2, or 7SK, suggesting these interactions are more likely related to Xist biogenesis. It will be important for future studies to determine whether each of the identified proteins bind Xist RNA during transcription and processing. A more definitive way to draw this important distinction may be to also examine the “interactome” of Xist RNA detected with intronic probes, which will pull-down proteins associated with the pre-mRNA.

Table 1. 1 Comparing top hits from three Xist RNA pull down studies.

Minagiji et al.		Chu et al.		McHugh et al.
Abundance (spectral counts)	Abundance (multiplex quantification)	Sheer abundance	Abundance relative to control RNA	Alphabetical order
PLIN1	MINT (SPEN)	HNRNPM	CIZ1	CELF1
Q3UJB0	FIBB	HNRNPU	RNF2	HNRNPC
TPR	COIA2	HNRNPK	TRIM6	HNRNPM
PLIN4	IKIP	MYEF2	RBFOX2	HNRNPU
NB5R3	RGAP1	SPEN	WTAP	LBR
ATRX	RFC1	RBMXL1	PCGF5	MYEF2
MPP10	COCA1	HNRNPAB	RYBP	PTBP1
RFA1	NEP	HNRNPD	TRIM71	RALY
DDX50	NUP88	HNRNPC	L1TD1	RBM15
RFC1	UHRF1	TARDBP	MYEF2	SPEN
HP1B3	WABL	IGF2BP1	SPEN	
TOP2B	ZFR	DDX39B	RALY	
RIF1	BAK	RBM15	YTHDC1	
EPIPL	NU133	HNRNPUL2	RBM15	
PSPC1	Q8BY0	RBM14	TARDBP	
HNRLL	CO1A1	SAFB	THOC4	
RRBP1	NHP2	CIZ1	IGF2BP1	
RL14	HELLS	HNRNPA0	POLDIP3	
SMC1A	HNRPU	YTHDC1	LIN28A	
NOC2L	LRWD1	SRSF7	RBMXL1	
A2AJ72	RCC1	RALY	IGF2BP3	
DNJB6	MBB1A	DDX17	HNRNPC	
KIF4	MYEF2	RNF2	SAFB	
1433T	LRP1	THOC4	SARNP	
SURF6	NXF1	SARNP	CELF1	
KI20A	RL7L	ELAVL1	HNRNPK	
PDS5B	HXA5	HNRPDL	RBM3	
ZN638	SMHD1	SRSF3	ERH	
RAD21	NFIC	SRSF2	RNF20	
SMHD1	P53	TRIM6	SRSF10	
DDX10	CELF2	RBM4	HNRNPM	

	Found in two studies
	Found in all three studies

Techniques that pull down RNA and identify attached proteins theoretically provide a more unbiased identification of proteins that interact with a specific RNA. However, current proteomics technologies can produce large background signals, especially when used for low abundance RNAs or high abundance proteins (Chu et al., 2015a). For example, abundant hnRNP proteins have been found in the majority of RNA purifications performed to date, regardless of whether whole cells or synthetic RNAs with cellular extracts were used, and it is not clear whether these interactions are biological or technical artifacts (McHugh et al., 2014). MS analysis also suffers from technical variation across experiments (Chu et al., 2015a).

Biochemical extraction techniques are a mainstay of many important studies in molecular cell biology. Nonetheless, the largely discordant results of recent studies of the "Xist RNA interactome" illustrate the reality that even seemingly small perturbations to a cellular environment or protocol can produce large discrepancies in data (Riley and Steitz, 2013). Ever since discovery that XIST RNA is embedded in nuclear structure, biochemistry has proven insufficient to reliably characterize the proteins with which it interacts. *In situ* techniques such as RNA FISH and IF as well as functional studies can validate interactions suggested by these and future biochemical methods. As further explored in Chapter III, manipulating Xist RNA localization during the cell cycle such as by forcing its retention on the chromosome during mitosis (Hall et al., 2009) could provide insight into which of these proteins binds Xist RNA *in vivo*. Such experiments would help to transform these from lists of *candidate* to *definitive* Xist RNA-interactors.

Concluding Remarks

Despite some holdouts (Razin et al., 2014), the once highly controversial concept of the nuclear matrix as a “skeleton” providing mechanical structure to the nucleus is no longer considered taboo. At the very least, most no longer think of the nucleus as an unstructured assembly of DNA relying on osmosis or diffusion of enzymes and substrates for replication and RNA metabolism. DNA is inherently structured, and its organization into chromatin allows it to fit inside the nuclear membrane. Sufficient evidence now supports that scaffolding proteins facilitate higher-order chromatin organization and nuclear compartmentalization.

In addition, RNA is increasingly recognized as far more than strictly a messenger transferring information from DNA to protein. The idea that a structural ncRNA comprises part of the nuclear scaffold is not new, and evidence for this was first described in the early days of nuclear matrix discovery (Nickerson et al., 1989). XIST sets a precedent for a scaffolding RNA that paints an interphase chromosome territory to orchestrate its epigenetic transformation into heterochromatin. Scientists are actively investigating the idea that other noncoding RNAs may function in a similar fashion, but they are limited by current extraction-based methods.

This thesis advances our understanding of how RNA may function as part of an overall nuclear scaffold required to maintain nuclear structure. Work in our lab has identified a class of repeat-rich RNA transcribed from sequences previously thought silent junk DNA, and I show evidence that these repetitive RNAs may form part of a scaffold for euchromatin. I encounter significant complexity to the anchoring of XIST

RNA in the Xi chromosome, and provide an important revision to currently widely accepted model of a single anchor protein. This work has important implications for epigenetic and nuclear structural changes that occur in cancer. However, my primary contribution advances knowledge of the basic biology of nuclear structure, which we must understand before we can begin to grasp the disease state.

Chapter II : Silencing of Chr 21 in Down syndrome fibroblasts by insertion of transgenic XIST RNA

Preface

Work presented in this chapter was part of a large effort by members of the Lawrence lab to target an *XIST* transgene to chromosome 21 in Down syndrome cells using zinc finger nucleases (in collaboration with Sangamo BioSciences). This work was primarily done in induced pluripotent stem cells in which we demonstrate comprehensive chromosome silencing, and it resulted in the following publication:

Jiang, J., Y. Jing, G.J. Cost, J.C. Chiang, **H.J. Kolpa**, A.M. Cotton, D.M. Carone, B.R. Carone, D.A. Shivak, D.Y. Guschin, J.R. Pearl, E.J. Rebar, M. Byron, P.D. Gregory, C.J. Brown, F.D. Urnov, L.L. Hall, and J.B. Lawrence (2013) Translating Dosage Compensation to Trisomy 21. Nature. **500**(7462):296-300. doi: 10.1038/nature12394

My contribution to this manuscript involved targeting of the *XIST* transgene to primary Down syndrome fibroblasts. This chapter includes some sections from the cited manuscript (Figures 2.2a, 2.3, and parts of the introduction and discussion), and some unpublished results that I obtained as a result of this effort. This work was supported by NIH grants GM053234, GM085548 and GM096400 RC4 to J.B.L.

Introduction

Evolution has devised a mechanism to compensate for the difference in X-linked gene dosage between sexes by inactivating one X chromosome in all female cells. However, no similar mechanism exists to handle dosage compensation for trisomy of an autosome. In the U.S., about 1 in 300 live births carry a trisomy, and roughly half of these are trisomy for chromosome 21, which causes Down syndrome (DS). DS is the leading genetic cause of cognitive disability and presents many progressive medical issues that affect multiple organ systems. Other chromosome duplication disorders that can result in live birth (trisomy 13 and 18) often are fatal within the first 1-2 years of life. While autosome duplication has extremely deleterious effects, the human sex chromosomes present the same challenge between females (XX) and males (XY) (Lyon, 1961). Our lab set out to test whether the natural mechanism for silencing the X chromosome can be applied to chromosome 21 for dosage compensation in Down syndrome.

It is not known whether XIST RNA has capacity to silence any chromosome, or if its function is limited to the X chromosome. Many human X-chromosome genes escape inactivation, the majority of which lie in a 10 MB X-escape region (Brown et al., 1997; Bailey et al., 2000; Carrel and Willard, 2005). Word frequency analysis has shown that the escape region differs from that of the rest of the chromosome in that it is >10-fold enriched in $[GATA]_n$ repeats (McNeil et al., 2006). As described above, the X chromosome is unique compared to autosomes in that it is enriched in L1 elements, and these have been hypothesized to play a role in Xist RNA spread (for review see (Hall and Lawrence, 2010). This suggests that DNA sequence could potentially influence XIST

RNA localization across its parent chromosome, and this could largely restrict XIST RNA's ability to silence an autosome. While studies of X:autosome translocations (e.g., (White et al., 1998; Hall et al., 2002b)) and randomly integrated *XIST*/*Xist* transgenes (Lee et al., 1996; Wutz and Jaenisch, 2000; Hall et al., 2002a) have indicated that autosomal chromatin has substantial capacity to be silenced, to better understand the potential of an autosome for silencing, this issue needs to be examined under conditions that avoid creation of a deleterious functional monosomy. The strategy pursued here meets that requirement, and provides the first tractable model to study the distinct biology of *human* chromosome inactivation (Heard et al., 1999; Migeon et al., 2001b; Hall et al., 2002a; Chow et al., 2007).

We used zinc finger nucleases (Moehle et al., 2007) to target an inducible *XIST* transgene to the gene-rich core of a trisomic Chr 21 in induced pluripotent stem (iPS) cells and fibroblasts derived from a subject with DS. Given its large size, neither the *XIST* gene nor its cDNA has previously been integrated in a targeted fashion. In the published manuscript, we report (i) an unprecedented efficiency and precision of this addition using by far the largest transgene to date; (ii) the resulting on-demand heterochromatinization of the extra Chr 21 due to acquisition of *de novo* histone modifications and chromatin condensation; (iii) long-range uniform transcriptional repression as gauged by *in situ* analyses, genome-wide expression profiling, and CpG promoter methylation status; (iv) inducible trisomy silencing *in vitro* impacts cell phenotype to improve both cell proliferation and neural rosette formation; and, finally, as described in this chapter, (vi) transgenic XIST RNA has capacity to silence an autosome in differentiated fibroblasts.

This first step toward correction of chromosomal imbalance in DS raises the important prospect of a combined genetic/epigenetic approach to “chromosome therapy.” In addition, our findings establish a unique system to study DS-related cellular pathologies, as well as to investigate the initiation of epigenetic chromosome silencing and its relationship to genomic sequence context. Results described here indicate that XIST RNA has capacity to spread across and initiate silencing of an autosome in a differentiated cell. Recognition of the autosome boundary by XIST RNA indicates the chromosome scaffold may play a role in marking the edge of the chromosome territory.

Materials and Methods

Cell culture: The human female Down syndrome primary fibroblast line (Coriell) (AG13902) was maintained according to supplier's instructions.

ZFN design: ZFNs against the *DYRK1A* locus were designed using an archive of pre-validated zinc finger modules (Urnov et al., 2010; Doyon et al., 2011), and validated for genome editing activity by transfection into K562 cells and Surveyor endonuclease-based measurement of endogenous locus disruption (Miller et al., 2007; Guschin et al., 2010) exactly as described (Doyon et al., 2011). Southern blotting for targeted gene addition was performed exactly as described (Urnov et al., 2005; Moehle et al., 2007) on SphI-digested genomic DNA probed with a fragment corresponding to positions Chr21:38825803+38826056 (hg19).

***XIST* plasmid construction and targeting:** Fourteen-kilobase human *XIST* cDNA, a splicing isoform of full-length *XIST* cDNA, was subcloned into pCMV 2x TetO2 (Invitrogen) containing a hygromycin selectable marker. Two homologous arms (left arm, 690 bp; right arm, 508 bp) of *DYRK1A* gene on chromosome 21 were amplified by PCR from primary Down syndrome fibroblasts (AG13902) (Coriell). When introduced without the Tet-repressor construct, the $TetO_2$ CMV promoter is constitutively active. We transfected two ZFN-containing vectors and the 21-kb *XIST* transgene into primary DS fibroblasts using Stemfect polymer (Stemgent) (10:1 ratio of *XIST* to ZFN, and 13 μg

DNA to 1.3 μl Stemfect per well of 6-well plate) and selected stable clones with hygromycin (75 μg ml^{-1}).

Fixation: Our standard cell fixation protocols have been previously described (Tam et al., 2002). Briefly, cells were grown on glass coverslips and extracted on ice in cytoskeletal buffer, 0.5% Triton X-100, and vanadyl ribonucleoside complex for 3-5 min. Cells were fixed at room temperature for 10 min in 4% paraformaldehyde and stored at 4°C in 70% EtOH.

IF: As previously described (Tam et al., 2002), coverslips were incubated in primary antibody diluted in 1% BSA, 1× PBS for 1 h at 37°C. They were then washed and immunodetected with a 1:500 dilution of fluorescent conjugated anti-mouse or -rabbit antibody in 1× PBS with 1% BSA. Coverslips were counterstained with DAPI and mounted with Vectashield (Vector Laboratories) for imaging. Antibodies were as follows: H3K27me3 (Millipore, 07-449), UbH2A (Cell Signaling, 8240), H4K20me (Abcam, ab9051), macroH2A (Millipore, 07-219).

Fluorescence in situ hybridizations: RNA hybridization was performed under non-denaturing conditions as previously described (Johnson et al., 1991; Tam et al., 2002; Byron et al., 2013). Briefly, hybridizations were done overnight at 37°C in 2× SSC, 1 U/μl of RNasin Plus RNase inhibitor (Promega), and 50% formamide with 2.5 μg/ml DNA probe. Cells were washed with 50% formamide/2× SSC at 37°C for 20 min, 2×

SSC at 37°C for 20 min, 1× SSC at RT for 20 min, and 4× SSC at RT for 1 min. Detection was with antidigoxigenin or fluorescein-conjugated avidin in 1% BSA/4× SSC for 1 h at 37°C. Three 10 min washes were done in 4× SSC, 4× SSC with 0.1% Triton, and 4× SSC at RT in the dark. Coverslips were counterstained with DAPI and mounted with Vectashield (Vector Laboratories). In simultaneous DNA/RNA FISH (interphase targeting assay), cellular DNA was denatured and hybridization performed without eliminating RNA and also treated with 2 U μl^{-1} of RNasin Plus RNase inhibitor (Promega). For immunostaining with RNA FISH, cells were immunostained first with RNasin Plus and fixed in 4% paraformaldehyde before RNA FISH. DNA probes were nick translated with either biotin-11-dUTP or digoxigenin-16-dUTP (Roche).
The *XIST* probe is a cloned 14-kb *XIST* cDNA (the same sequence as *XIST* transgene in Fig. 1b) in pGEM-7Zf(+) (Promega). *DYRK1A* and *APP* gene probes were bacterial artificial chromosomes (BACs) from BACPAC Resources.

Microscopy and image analysis: Digital imaging and analysis was performed using an Axiovert 200 microscope (Carl Zeiss, Inc.) with a 100× NA 1.4 Plan-Apochromat objective and multi-bandpass dichroic and emission filter sets (model 83000; Chroma Technology Corp) set up in a wheel to prevent optical shift. We used AxioVision software (Carl Zeiss, Inc.) and an Orca-ER camera (Hamamatsu Photonics). A narrow bandpass fluorescein filter was inserted to eliminate cross talk between channels. When necessary, we used Photoshop (Adobe) to enhance images for brightness and contrast.

Results

Targeted XIST gene addition in primary DS fibroblasts

To test the translational applicability of chromosome silencing for DS therapy, we questioned whether targeted *XIST* addition may be achievable in primary human cells. Due to infrequent homologous recombination and frequent cell senescence (which precludes sub-cloning from single cells), targeted gene addition in non-immortalized human fibroblasts generally is not feasible. However, we reasoned that ZFN-driven editing may make it possible.

I targeted a single *XIST* transgene with a CMV promoter and selectable marker to the DYRK1A locus on chromosome 21 (Figure 2.1) in fibroblasts from a female DS patient. I obtained numerous colonies that grew into a sparse monolayer. Many cells in the pooled integrant population expressed either transcription foci or localized accumulations of XIST RNA that diminished over several passages, as cell proliferation also diminished. While this precluded in-depth analysis, the majority of XIST RNA signals were accurately targeted to chromosome 21, with a minimum of 74% (46/62) associated with the DYRK1A locus, as seen by both interphase and metaphase FISH analysis (Figure 2.2A-B). This suggests efficient ZFN targeting, expression and localization of the transgene in primary fibroblasts.

XIST RNA expression was robust and localized in a territory over one Chr 21 in the majority of cells, though the transgene territories tended to be smaller than the

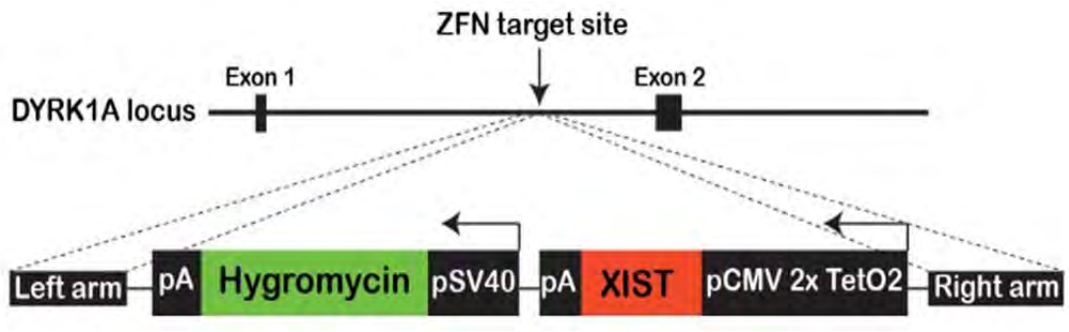

Figure 2.1 The 21 kb *XIST* targeting construct

The 21 kb targeting construct contains two cassettes: a full-length XIST cDNA and a hygromycin selection gene. The *XIST* transgene was driven by a CMV promoter.

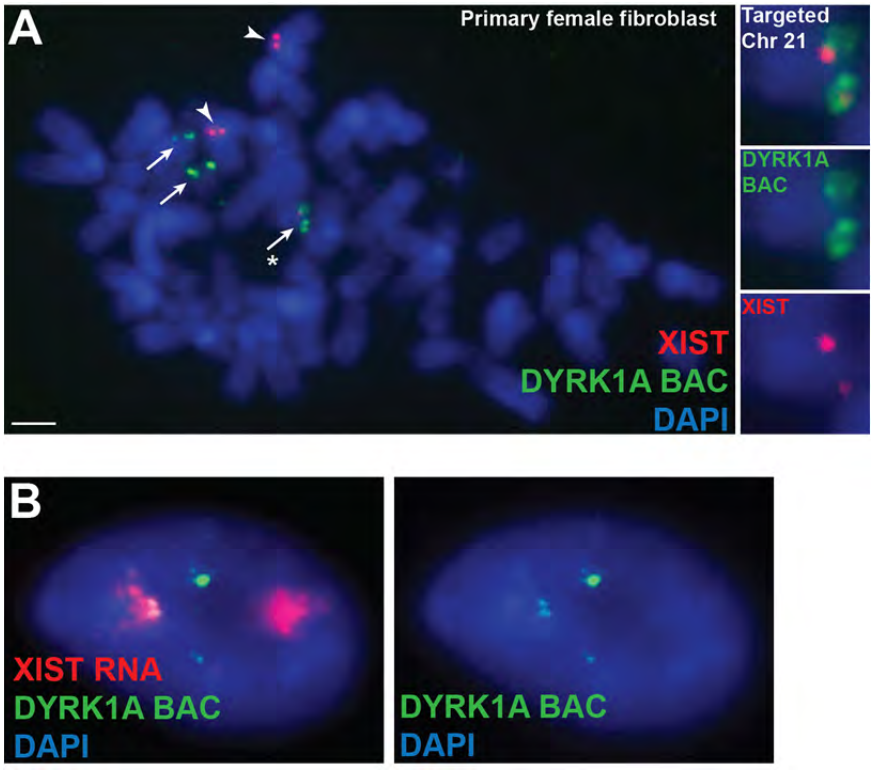

Figure 2.2 The XIST transgene is targeted to Chr 21 and induces an RNA "paint" that localizes to the autosome territory in *cis*.

(A) Metaphase DNA FISH in human DS primary fibroblasts. Long arrows indicate three Chr 21s, and an asterisck indicates the XIST-targeted Chr 21 that is enlarged in the side panel. Arrowheads indicate endogenous XIST genes on the two X chromosomes in the female cell. (B) Transgenic XIST RNA (red) paints a single Chr 21 (DYRK1A DNA, BAC probe) (green) in interphase DS fibroblasts.

endogenous XIST RNA territory at the Xi locus, which may reflect the different sizes of the chromosomes (Chr 21 is the smallest chromosome) (Figure 2.2B). Nevertheless, the transgenic XIST RNA modeled the unique behavior of endogenous XIST RNA which "paints" the inactive X nuclear territory (Clemson et al., 1996).

XIST RNA induces heterochromatin formation on the targeted chromosome

The Xi in female cells forms a visibly condensed Barr body which carries an epigenetic signature of repressive histone modifications (reviewed in (Wutz, 2011)). Evidence has shown that differentiated mouse fibroblasts lack the capacity to support chromosome silencing (Surralles and Natarajan, 1998; Wutz and Jaenisch, 2000). Surprisingly, my data suggests XIST RNA induced several heterochromatin modifications on Chr 21, and heterochromatic marks were notably enriched with the XIST RNA paint in many cells: H3K27me (39%), H4K20me (38%), UbH2A (66%) and MacroH2A (24%) (Figure 2.3). Further, in a fraction of cells, chromosomal DNA in the XIST-coated territory became visibly condensed into a "Chr 21 Barr body," which appears indistinguishable from that of the endogenous Xi (Figure 2.3, arrows). This supports that XIST RNA function is not limited to the X chromosome, and it can also recruit chromatin modifiers and induce structural changes to transform an autosome from euchromatin to heterochromatin.

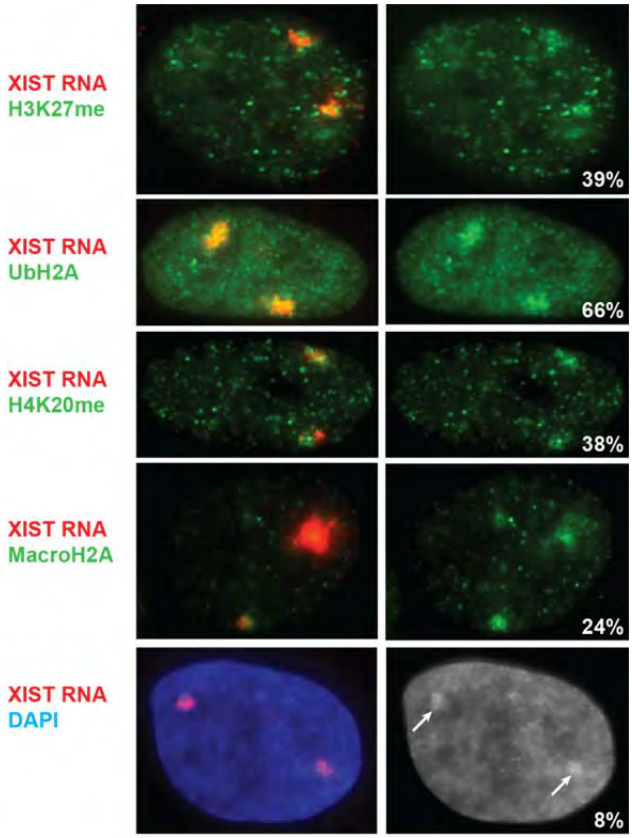

Figure 2.3 XIST RNA induces heterochromatin formation on the targeted chromosome.

Transgenic XIST RNA induces heterochromatin formation on the targeted Chr 21. Heterochromatin markers H3K27me, UbH2A, H4K20me, and MacroH2A become enriched with the XIST RNA paint in a fraction of cells. Chromatin condenses to form a "Chr 21 Barr body" (arrows) in some cells. % in lower right corner indicates fraction of cells with enrichment over both endogenous and transgenic XIST RNA territories.

XIST RNA induces transcriptional silencing on the targeted autosome

Several years ago our lab demonstrated that the inactive X chromosome coated by XIST RNA is transcriptionally silent by hybridizing with a labeled CoT-1 fraction of genomic DNA, which is enriched for repeat sequences and broadly detects heterogeneous nuclear RNA (Hall et al., 2002a). RNA FISH to CoT-1 produces a robust signal on euchromatin that is excluded from the inactive X chromosome territory, suggesting it is transcriptionally inactive. I examined CoT-1 RNA expression in cells containing the XIST RNA-coated Chr 21 and found a CoT-1 RNA hole forms over the XIST RNA territory in 38% of cells (Figure 2.4A). This further suggests that XIST RNA has capacity to induce transcriptional silencing of an autosome.

I next used RNA FISH to examine the presence of transcription foci at each allele of the APP gene, which encodes amyloid beta precursor protein. Mutations in APP (causing accumulation of β–amyloid) lead to early onset familial Alzheimer's disease (Tanzi and Bertram, 2005), and APP over-expression is linked to Alzheimer's disease in DS as well (Webb and Murphy, 2012). RNA FISH with a cDNA probe detects a transcription foci for actively transcribed genes, an established approach to discriminate active versus silenced genes on Xi (Clemson et al., 2006). Cells lacking a robust XIST RNA paint over Chr 21 contained readily visible RNA transcription foci for each APP allele, including one adjacent to the XIST RNA signal (Figure 2.4B). However, 80% of the cells with XIST RNA paint over the targeted Chr 21 contained only two APP transcription foci, which were not adjacent to the XIST RNA signal, indicating XIST RNA had silenced transcription of the targeted chromosome (Figure 2.4B).

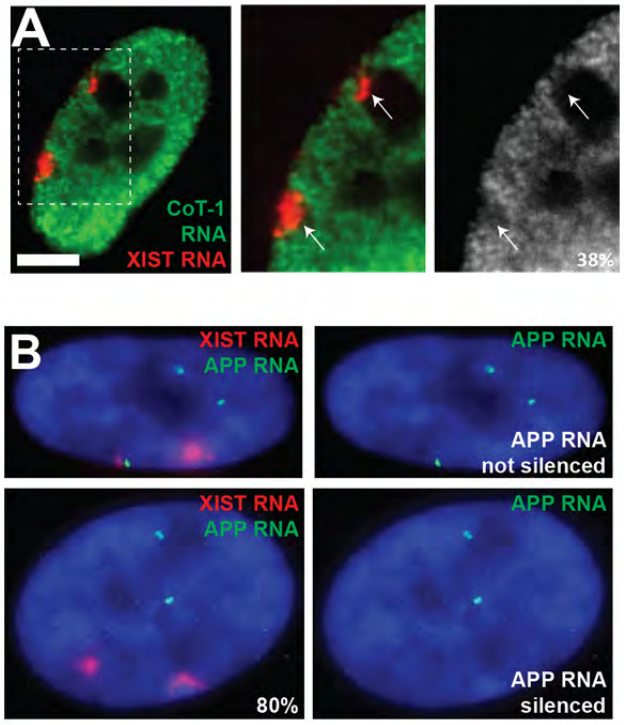

Figure 2.4 Transgenic XIST RNA induces transcriptional silencing of an autosome.

(A) Transgenic XIST RNA induces silencing of heterogeneous RNA (CoT-1 RNA) on the targeted and endogenous loci. XIST RNA (red) localizes in a CoT-1 RNA (green/gray) "hole" in 38% of the targeted loci. (B) Targeted XIST RNA induces silencing of APP mRNA. APP mRNA transcription foci (green) indicate active transcription of the allele. APP transcription is silenced on the the targeted, XIST RNA painted chromosome (lower panels), but not the targeted chromosome with no XIST RNA paint (upper panel).

Discussion

While in-depth characterization is beyond the scope of the present work due to the limited lifespan of primary cells and progressive silencing of the CMV promoter, these data emphasize that targeted *XIST* gene addition to Chr 21 can be attained in primary cells. In addition to providing a new analytical approach to study the cellular impact of trisomy, the targeted integration of XIST cDNA into the DYRK1A locus in primary fibroblasts of a female DS patient establishes feasibility for the unprecedented concept of chromosome therapy. Over the longer-term, this has potential to provide a strategy for DS genetic therapy.

Over the shorter term, this offers a means to study initiation of X chromosome inactivation in human cells, including XIST RNA localization and spreading across its parent chromosome in *cis*. Our *XIST* transgene lacks any counting sequence, thus this strategy translating dosage compensation to an autosomal imbalance equally applies in both male and female cells without interfering with natural X-inactivation. The robust silencing induced by XIST RNA makes it essential that the RNA avoid spreading to and silencing genes on neighboring chromosomes, however it is not known how XIST RNA recognizes the boundaries of the X chromosome. These results support that XIST RNA not only has capacity to bind to an autosome, but that the RNA does not recognize a sequence or scaffold unique to the X chromosome. Instead it may recognize a more universal scaffold that defines the boundaries between the densely-packed chromosomes. In the following chapter, I will further investigate how XIST RNA localizes to the scaffold for the inactive X chromosome.

Chapter III : The complex biology anchoring chromosomal XIST RNA

Preface

Work described in this chapter has been submitted for publication in *Developmental Cell* and is currently in revision.

Kolpa, H.J, F.O Fackelmayer, and J.B. Lawrence. Multiple Anchors localize XIST RNA in normal cells but some cancer lines require SAF-A.

I would like to thank Emiliano Ricci (M. Moore lab) for advice and reagents for making the stable cell lines. This work was supported by NIH grant GM053234 to J.B.L.

Introduction

Work described in this chapter advances understanding of how XIST RNA, a paradigm for RNA that regulates chromatin, localizes to its parent chromosome. Prior work showed mouse Xist RNA localization solely depends on Scaffold Attachment Factor-A (SAF-A/hnRNP U), but human SAF-A's relationship to XIST RNA has not been studied. Here I use RNAi and a series of SAF-A deletion mutants in a variety of cell types to examine *in vivo* the protein's requirement to support XIST RNA localization to the human Xi. Having found some initially discrepant results, I also revisit this question in mouse cells. I show that SAF-A is involved in localizing human XIST RNA to the chromosome territory, and that the DNA binding domain was most critical for this role. However, evidence also suggests that in normal somatic cells (both mouse and human) XIST RNA localization is not singularly dependent on SAF-A. Cell types differ in their requirement for SAF-A, and only the highly transformed or immortal cell types examined here completely require SAF-A. By inhibiting phosphorylation changes that normally release XIST RNA at prophase (Hall et al., 2009), I provide further evidence that XIST RNA has multiple anchors, one (or more) of which is controlled by Aurora B Kinase (AURKB), whereas SAF-A is not. Finally, this novel strategy has utility to identify *in vivo* proteins that have bone fide interactions with XIST RNA, as I show for hnRNP K. Ultimately *in situ* approaches will be required to further dissect the complex biology that underpins RNA's strict localization on chromatin to maintain epigenetic stability. These findings demonstrate unrecognized complexity to the biology of anchoring chromosomal RNA embedded in the nuclear scaffold.

Materials and Methods

Cloning: SAF-A cDNA was cloned into pEGFP-N1 vector (Clontech) and the ΔRGG and G29A mutants made by PCR-based mutagenesis with the primers ΔRGG: Fwd: 5'-TCCGTTAAACTGGTTCTTGCCACTCTTA-3'
Rev: 5'-TCCGGCGGTGGAGGAAGTGGTGGAATC-3' and G29A: Fwd: 5'-ACGCCTTTCTGACAAGGCCCTCAAGGCCGAGCTC-3'
Rev: 5'-GAGCTCGGCCTTGAGGGCCTTGTCAGAAAGGCGT-3'.
Dominant negative C280-GFP was cloned into pEGFP-NLS (Clontech) after PCR amplification with primers Fwd: 5'-AGATCGA ATTCTGTTTAAGAAGCAAATG GCAGAT-3' Rev: 5'-AGATCGGATCCGAATAATATCCTTGGTGATA ATGC-3'.
Flag tags were added by restriction digest to cut GFP from pEGFP-N1and ligating with annealed oligo adaptors: sense: 5'-AGCTTGATTACAAGGATGACGACGATAAGATCTGAGCGGCCGCGGTAC-3'
antisense: 5'-CGCGGCCGCTCAGATCTTATCGTCGTCATCCTTGTAATCA-3'.
For stable lines, Flag-tagged constructs were cloned into pcDNA/FRT/TO (Invitrogen). Mutagenesis to SAF-A siRNA target sites was done by Genscript: GGC → GGA (Glycine), TCG → TCA (Serine), GTT → GTA (Valine), and GTT → GTA (Valine).

Cell culture: Tig-1 (Coriell) and PCS-400-012 (ATCC) are normal female fibroblasts. Female MEFs were a gift from S. Jones [U. of Massachusetts Medical School, Worcester, MA]. Neuro2a cells were a gift from J. Landers [U. of Massachusetts Medical School, Worcester, MA]. Mouse 3x cells were a gift from C. Brown [U. of British Columbia,

Vancouver, British Columbia, Canada]. Hek 293 Flip-in cells were a gift from M. Moore [U. of Massachusetts Medical School, Worcester, MA]. All cell lines were maintained according to supplier's instructions. Poly-D-lysine-coated coverslips were used for Hek 293 cells.

Drug treatment: Hesperadin (EMD Millipore cat #375680) dissolved in DMSO was mixed with cell culture media at 100nM and added to cells for 2 hours. For analysis of mitotics, cells were trypsinized and diluted in cold 1× PBS then cytospun at 5000 RPM onto coverslips as described (Johnson et al., 1991). Tautomycin (EMD Millipore cat #580551) dissolved in DMSO was mixed with cell culture media at 1, 2.5, or 5 µM and added to wells for 4 hours.

Fixation: Our standard cell fixation protocols were described in Chapter II.

siRNA and Plasmid transfections: For RNAi, cells (70-80% confluent) were transfected with SMARTpool siRNAs and DharmaFECT 1 transfection reagent (GE Dharmacon) according to manufacturer's instructions, and fixed after 72 hours. siGLO siRNA (GE Dharmacon) was used as a control which had no impact on RNA localization. For plasmid transfections, 2-4 ug of DNA was mixed with Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions, and added to 80-90% confluent cells in a 6-well plate. Cells were fixed and assayed 6, 24, 48 or 72 hours after transfection.

Creation of stable cell lines: Hek 293 Flip-in cells were transfected with pOG44 and SAF-A Flag constructs in the expression vector pcDNA5/FRT/TO. Cells were selected in Geneticin (Gibco).

IF: IF was done as previously described in Chapter II. Prior to some SAF-A antibody stains, I treated cells with epitope retrieval by incubating coverslips in 10mM citric acid and .05% tween 20 at 100°C for 20-40 minutes. This allowed us to visualize the enriched layer of endogenous SAF-A on the Barr body, previously seen only after a matrix prep or with the GFP-tagged protein. Antibodies used were anti–Flag M2 (Sigma), anti-SAF-A (Abcam ab20666 and ab10297), anti-FUS (Bethyl Labs), anti-hnRNP C (Abcam ab97541), anti-SafB1 (Abcam ab8060), anti-NuMA (Abcam ab36999), anti-Lamin B1 (Abcam ab16048 and Santa Cruz sc-6217), anti-hnRNP K (Abcam ab18195), anti-hnRNP M (Sigma HPA024344), and anti–phosphorylated histone H3S10 (Cell Signaling 9706).

Fluorescence in situ hybridizations: RNA hybridization was performed under non-denaturing conditions as previously described in Chapter II. Probes were a cloned 14 kb human *XIST* and 6.3 kb mouse *Xist* cDNA in pGEM-7Zf(+) (Promega), XIST introns 1 and 2 (gift from H. Willard [Duke University Medical Center, Durham, NC] and C. Brown [U. of British Columbia, Vancouver, British Columbia, Canada]), and cloned

COL1A1 collagen intron 6 made by PCR with the primers: Fwd: 5'-

GTTCTGCTTGTGTCTGGGTTTC-3'

Rev: 5'-AGTAAGTTCAACCTTCCCCCTCC-3'

Microscopy and image analysis: This was done as described in Chapter II. At least 100 cells in triplicate were scored in each experiment.

Results

SAF-A is not required for XIST RNA localization in normal human fibroblasts

To test whether SAF-A is required for XIST RNA localization in human cells, I used siRNA (sMARTpool Dharmacon/Thermo) to knock down SAF-A in two normal primary human cell lines and analyzed XIST RNA localization in hundreds of single nuclei. To rigorously account for gradations in XIST RNA "phenotype", I distinguished four levels of mislocalization, ranging from fully localized to fully mislocalized (Figure 3.1A). I discriminate between RNA released from the chromosome (partially or fully mislocalized) versus RNA that remains "loosely" localized to a potentially decondensed territory. In primary Tig-1 (lung) fibroblasts, XIST RNA was almost never fully mislocalized after SAF-A knockdown, with 95% of cells showing one of the three more localized phenotypes (Figure 3.1B-C). XIST RNA was even more resistant to mislocalization after SAF-A knockdown in a primary human kidney cell line (PCS-400-012, ATCC). In this cell line, XIST remained fully or loosely localized in 88% of cells (Figure 3.1B, D-E).

The XIST RNA phenotype did not reflect the extent of SAF-A depletion by RNAi, since only cells which lost SAF-A staining were scored, and SAF-A levels averaged 84% (kidney) and 90% (Tig-1) reduced as measured by fluorescence intensity of single nuclei (see Materials and Methods and Figure 3.2A). Nor was the lack of requirement for SAF-A due to an arrest of the cell-cycle. Cells continued to proliferate and the majority of siRNA-treated cells lacking SAF-A had gone through mitosis in the

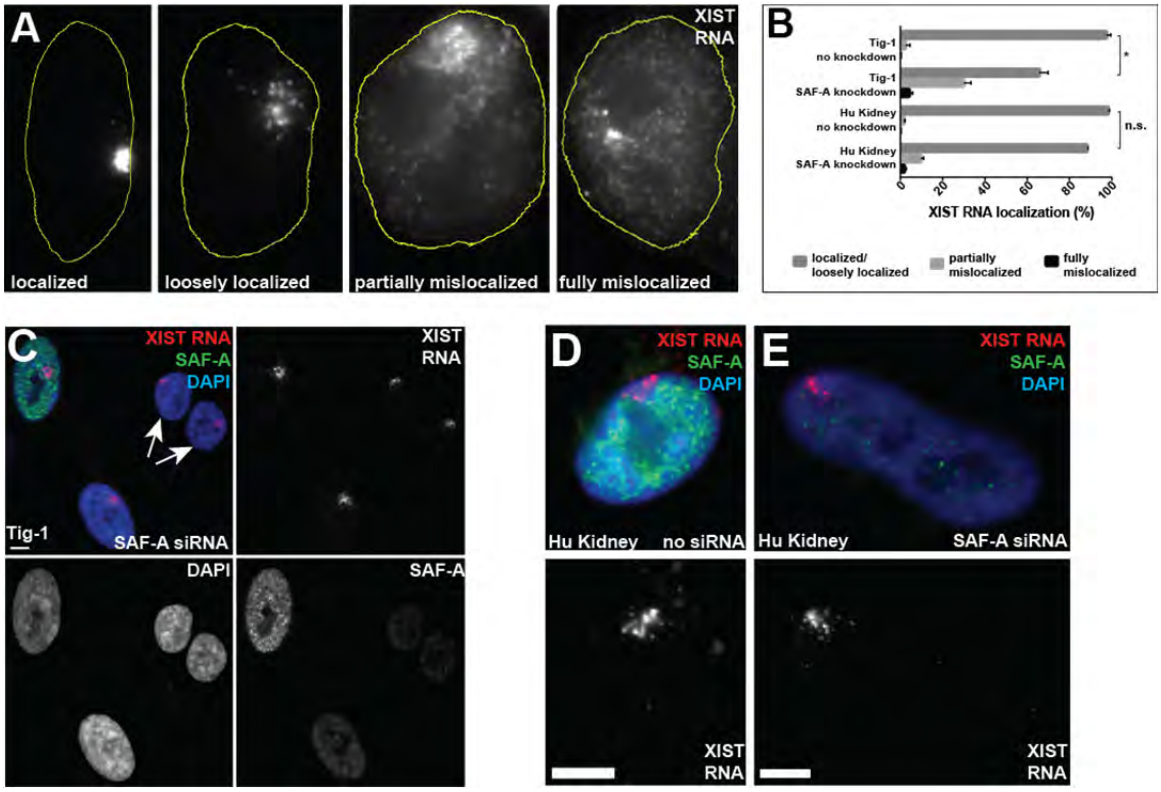

Figure 3.1 SAF-A is not required for XIST RNA localization in some cell types.

(A) Representative images of XIST RNA phenotypes after SAF-A knockdown. (B) Quantification of XIST RNA phenotypes after SAF-A knockdown in Tig-1 and Primary Human Kidney cells. Error bars: standard error of the mean. (C) XIST RNA localization in Tig-1 cells after SAF-A siRNA (72h). XIST RNA is still localized even in SAF-A-depleted post-mitotic G1 daughter cells (arrows). (D-E) XIST RNA localization in control (D) and SAF-A siRNA-treated Primary Human Kidney cell (E). Scale bars, 5 µm.

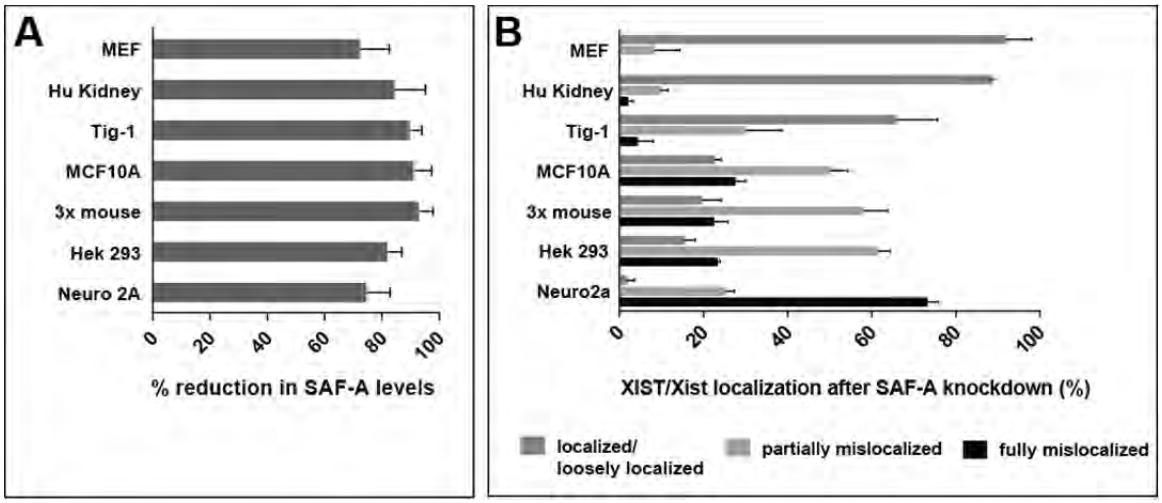

Figure 3.2 XIST RNA phenotype after SAF-A siRNA does not reflect average reduction in SAF-A levels.

(A) Quantification of SAF-A levels in multiple individual nuclei (at least 50 cells per experiment) treated with SAF-A siRNA or no siRNA. Measured fluorescence intensity with Image J. Data represent mean percentage +/- standard deviation of knockdown, compared to control. (B) Quantification of XIST/Xist RNA phenotypes after SAF-A knockdown across seven cell types.

72 hours prior to analysis. These human cells divide ~every 24 hours, and I could readily identify post-mitotic G1 daughter (G1d) cells lacking SAF-A staining which nonetheless localized newly synthesized XIST RNA (Figure 3.1A). Hence, results indicate that SAF-A is not required for XIST RNA localization in these normal human somatic cells. They do not rule out that SAF-A may be involved or sufficient in some cells, as further studied below.

SAF-A is required for XIST RNA localization in more transformed or immortal cell types

Since our results in human cells differed from published results showing SAF-A knockdown alone in mouse Neuro2a neuroblastoma cells was sufficient to release Xist RNA, I repeated this analysis in Neuro2a cells. Xist RNA became fully mislocalized (dispersed RNA with no evident territory) in ~80% of cells, similar to the published results (Hasegawa et al., 2010) (Figure 3.2A-B, 3.3A-B). In the prior study, Xist RNA did not appear to us to be comparably released in the presented images of mouse embryonic fibroblasts (MEFs), though the authors did not describe a difference (Hasegawa et al., 2010; Hasegawa and Nakagawa, 2011). Thus, I repeated SAF-A knockdown in primary MEFs and found that the majority maintained localized or loosely localized Xist (3.2A-B, 3.3C-D). Results confirm a requirement for SAF-A in localizing Xist RNA in Neuro2a cancer cells, however I saw little to no Xist RNA release in primary fibroblasts, indicating the effect may be cell-type specific.

I next tested three transformed or immortalized cell lines: human Hek 293 and MCF10A mammary epithelial cells, and mouse 3X. In each of these lines, I found that

SAF-A RNAi caused partial release of XIST/Xist RNA in the majority of cells and full release in ~25% (3.2A-B, 3.3E). These results show clear differences in the requirement for SAF-A to localize XIST/Xist RNA in different cell types. They further show that there is some heterogeneity between cells within these cultured cell lines. Importantly, only Neuro2a cells showed a complete and consistent release of Xist RNA from the chromosome. These findings raise the important possibility that immortalized or transformed cells depend more on SAF-A to localize XIST RNA than primary cells. I noticed that the XIST RNA territories in many untransfected Hek 293 cells (Figure 3.3E, arrowhead) resemble the loosely localized territories observed in Tig-1 cells after SAF-A knockdown, which may reflect compromised chromatin architectural proteins resulting in less condensed DNA territories in transformed cells, likely related to why the RNA is more easily dispersed by SAF-A knockdown.

Since pluripotent nuclei have more “mobile” chromatin (Meshorer et al., 2006) and cancer cells recapitulate some properties of early stem cells (Carone and Lawrence, 2013), I examined XIST RNA after SAF-A knockdown during inducible initiation of XCI in human iPS cells harboring an *XIST* transgene on one of three chromosome 21s (Jiang et al., 2013). I found that 72 hours after induction, XIST RNA failed to accumulate and form a territory in cells lacking SAF-A staining, instead forming only transcription foci (Figure 3.3F-G). These data support that SAF-A is required for localized accumulation of XIST RNA during initiation of human X inactivation in pluripotent cells. I also find that iPS cells containing SAF-A more loosely localize or partially mislocalize XIST RNA compared to differentiated cells (Figure 3.3F,

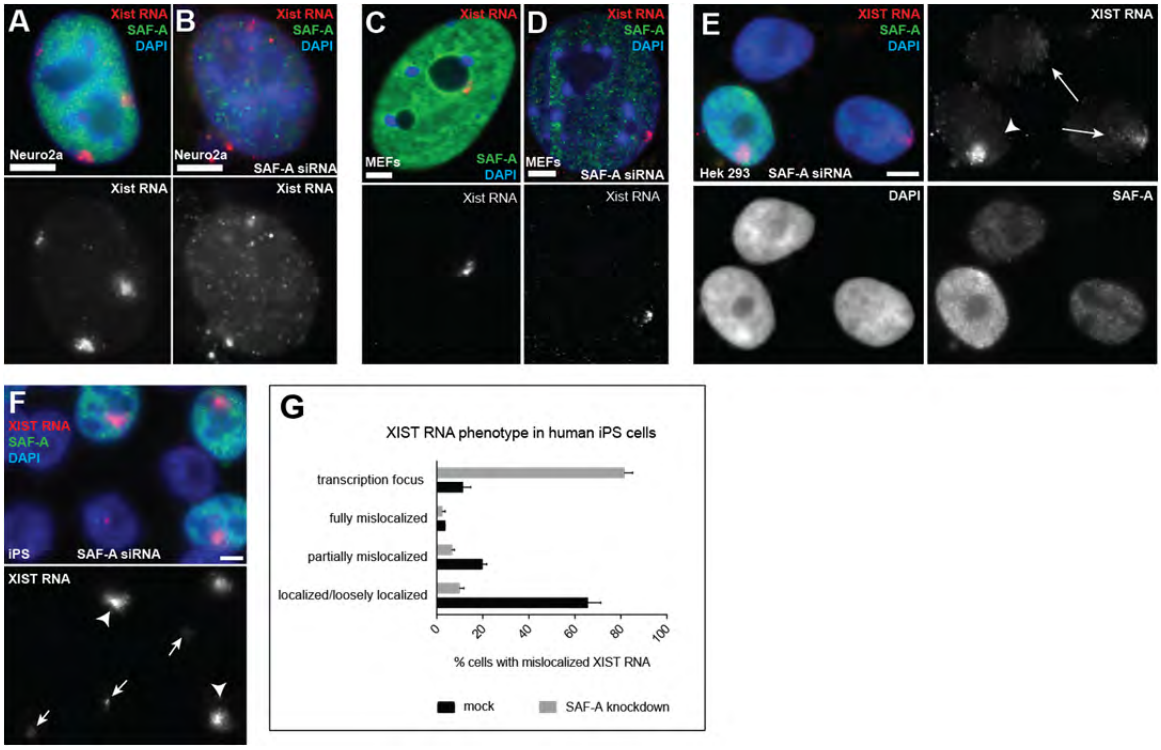

Figure 3.3 SAF-A is required to localize XIST RNA in some transformed or immortal cell types.

(A-B) Xist RNA localization in control (A) and SAF-A siRNA-treated (B) Neuro2a cells. (C-D) Xist RNA localization in control (C) and SAF-A siRNA-treated (D) MEFs. (E) XIST RNA localization in control (arrowhead) and SAF-A siRNA-treated (arrows) Hek 293 cells. (F) XIST RNA localization in control (arrowheads) and SAF-A siRNA-treated (arrows) human iPS cells. (G) Quantification of XIST RNA phenotypes in iPS cells. Error bars: standard error of the mean. Scale bars, 5 μm.

arrowheads). We suggest that once cells differentiate and X inactivation becomes stable, additional proteins contribute to anchoring XIST RNA, making it more tightly localized as the epigenetic state is defined, and more resistant to mislocalization after knockdown of a single protein.

A dominant negative form of SAF-A disrupts XIST RNA localization in normal cells

We considered that SAF-A depletion mislocalizes XIST RNA in transformed or pluripotent but not in normal cells because normal cells have multiple anchors. Therefore I tested whether XIST RNA is disrupted in primary Tig-1 cells expressing a dominant negative SAF-A, which more broadly disrupts nuclear scaffold proteins. I used a plasmid encoding the C-terminal 280 amino acids of SAF-A fused to a GFP tag (C280-GFP), which causes changes to endogenous SAF-A localization and chromatin packaging (visible by DAPI and IF) (Figure 3.4A-B). I found that dominant negative SAF-A caused XIST RNA mislocalization in 73% of Tig-1 cells (Figure 3.4C-D) in contrast to RNAi, suggesting the SAF-A mutant disrupts other XIST RNA localization factors.

Consistent with this, I found that C280-GFP expression affects the localization patterns of several other known nuclear matrix proteins, including FUS, hnRNP C, SAFB1, and NuMA but, interestingly, not LaminB1 (Figure 3.5A-E). These results highlight that SAF-A is embedded with XIST RNA in a complex structure, and they support the concept that XIST RNA depends on multiple proteins and scaffold integrity for proper localization in normal cells.

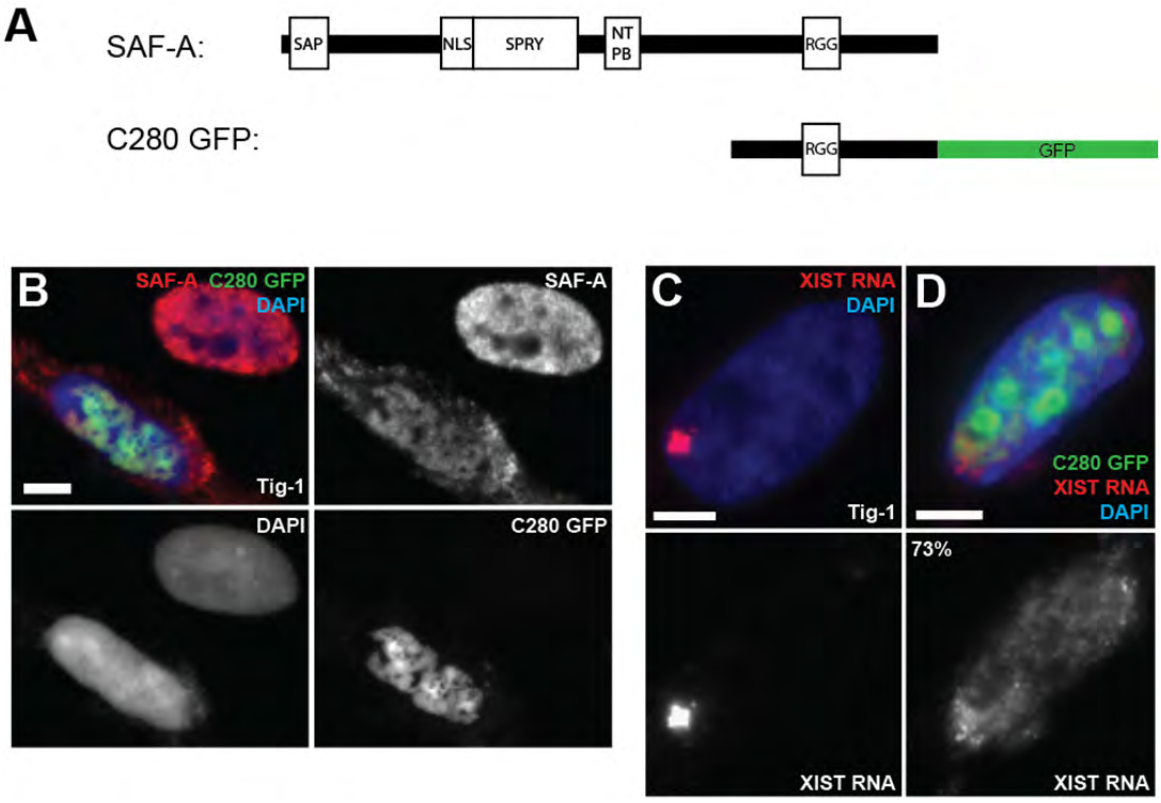

Figure 3.4 Dominant negative SAF-A expression mislocalizes XIST RNA.

(A) Map of full length SAF-A and dominant negative C280-GFP. (B) Dominant negative SAF-A (C280-GFP) disrupts endogenous SAF-A (red). (C-D) XIST RNA localizes to a discreet territory in Tig-1 cell (C) and mislocalizes in Tig-1 cell expressing C280-GFP (D). Scale bars, 5 µm.

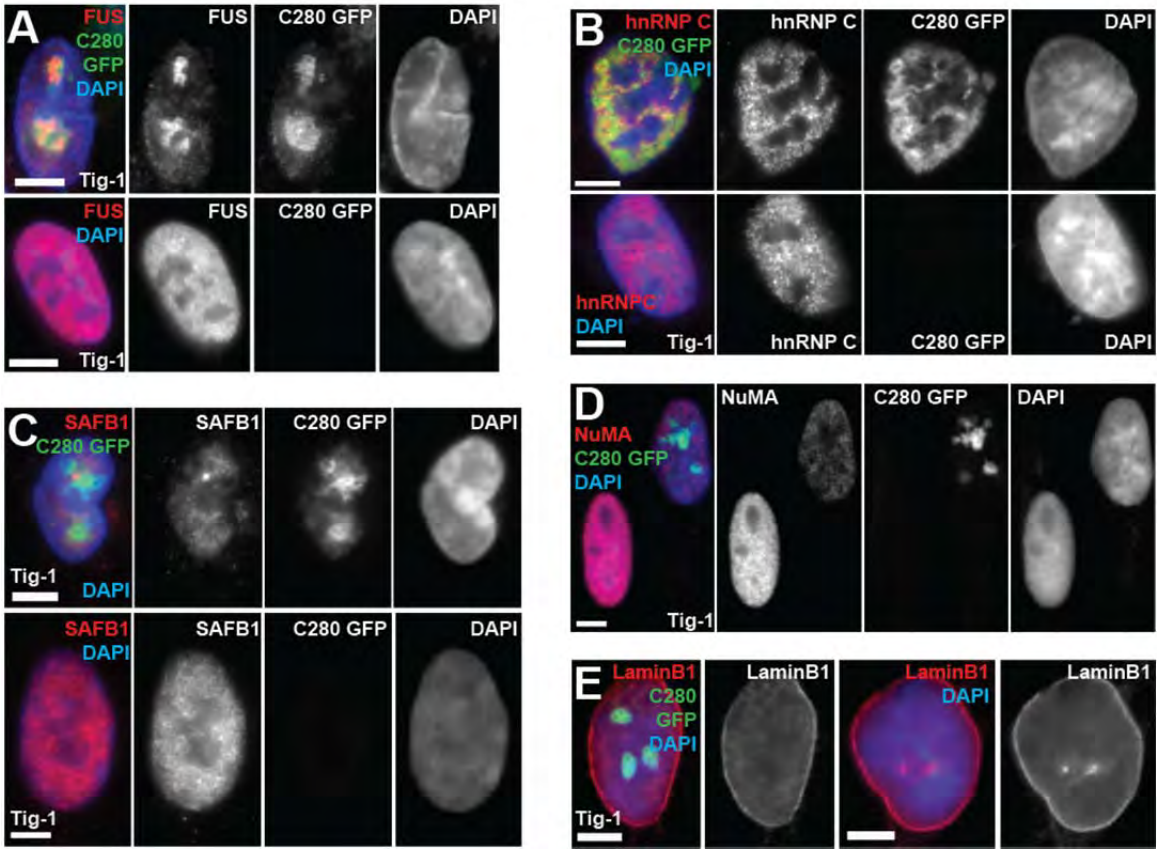

Figure 3.5 Dominant negative SAF-A expression mislocalizes other nuclear matrix proteins.

(A-E) C280-GFP expression affects localization of nuclear matrix proteins FUS (A) hnRNP C (B) SAFB1 (C) and NuMA (D) but not LaminB1 (E). Scale bars, 5 μm.

A point mutation in SAF-A's DNA binding domain releases XIST RNA from the chromosome

To more specifically probe SAF-A's role and the specific domains of SAF-A required to localize human XIST RNA, I transiently expressed tagged SAF-A deletion mutants in Hek 293 cells. I used Hek 293 cells since above results indicate they rely more on SAF-A exclusively to anchor XIST RNA than primary cell types. Importantly, I found that overexpression of full length SAF-A causes XIST RNA release (Figure 3.6A), suggesting that the stoichiometric relationships of SAF-A to other nuclear components are important for its role. Hence, variable and uncontrolled expression levels of transient SAF-A constructs makes them unreliable for analysis of effects on XIST RNA (Figure 3.6B-D). Therefore, to control expression levels, I made stable cell lines expressing doxycycline-inducible Flag-tagged SAF-A or SAF-A deletion mutants lacking either the DNA or RNA binding domains (see Materials and Methods) (Figure 3.6E). Flag-tagged full-length SAF-A expression did not affect XIST RNA localization in the stable cell line (Figure 3.6B, F).

While a protein could contribute to XIST RNA localization without directly binding DNA, a distinguishing feature of a "primary" anchor would be its capacity to bind DNA directly. To test whether SAF-A must bind DNA in order to properly localize human XIST RNA, I used a mutant containing a G29A point mutation in the DNA binding site (Figure 3.6E). I analyzed XIST RNA localization after endogenous SAF-A knockdown in cells expressing stable, inducible G29A-Flag (which contained silent mutations to prevent siRNA targeting of the Flag-tagged transcript, see Materials and

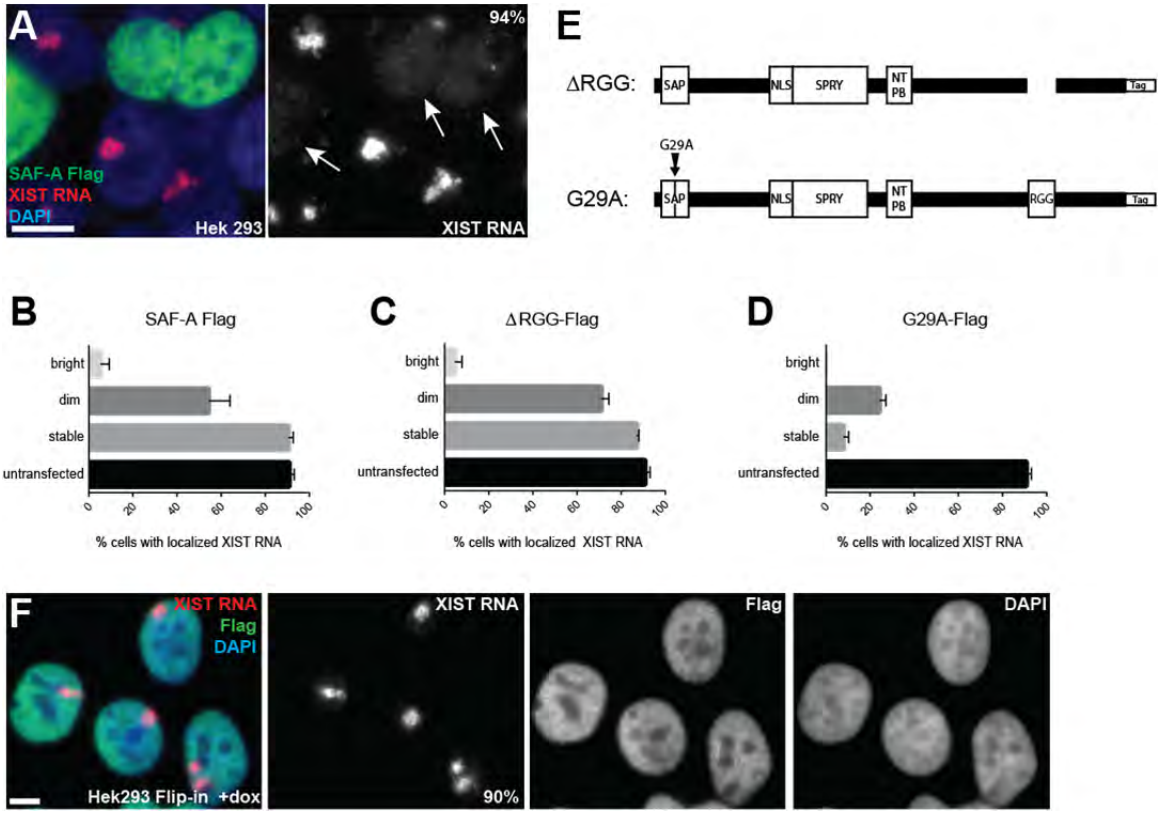

Figure 3.6 Effect of SAF-A and SAF-A deletion mutants on nuclear organization depends on expression level.

(A) Transient overexpression of full-length SAF-A mislocalizes XIST RNA (arrows). (B-D) Quantification of XIST RNA localization in cells expressing stable or transient SAF-A Flag, ΔRGG-Flag, or G29A-Flag. Transiently transfected cells were scored according to whether they had high (bright) or low (dim) levels of Flag-tagged protein. Error bars: standard error of the mean. (E) Map of ΔRGG and G29A SAF-A mutants with C-terminal GFP or Flag tag. (F) XIST RNA territories are not affected in cells expressing stable Flag-tagged SAF-A. Scale bars, 5 μm.

Methods). Whereas full length flag-tagged-SAF-A "rescued" XIST RNA localization after endogenous SAF-A knockdown (Figure 3.7A, D), the G29A mutant did not (Figure 3.7B, D).

Further, I found that XIST RNA mislocalization was more striking and prevalent in cells both lacking endogenous SAF-A and expressing G29A-Flag (55% fully mislocalized) compared to cells only lacking endogenous SAF-A (18% fully mislocalized) (P=.002) (Figure 3.8A). In fact, the G29A mutant had an even stronger disruptive effect in cells that still expressed endogenous SAF-A (81% fully mislocalized) (Figure 3.8B). Taken together, results suggest that SAF-A binding to DNA is required for XIST RNA localization, and further, that the SAF-A protein without the normal DNA binding domain is more disruptive to XIST RNA localization than absence of SAF-A, likely through its impact on other interactors (see model in Figure 3.8C).

SAF-A's RGG domain is not required to support XIST RNA localization

I repeated the above experimental strategy in cells expressing a mutant containing a short deletion of the C-terminal RGG box, ΔRGG (Figure 3.6E). SAF-A's RGG domain was previously shown to bind poly U and poly G RNA, as well as single-stranded DNA *in vitro* (Kiledjian and Dreyfuss, 1992). In contrast to the DNA binding domain mutant, in the same experiments, the ΔRGG mutant was able to substitute for full-length SAF-A, such that XIST RNA was fully or loosely localized in cells expressing ΔRGG-Flag despite endogenous SAF-A knockdown (Figure 3.7C-D). These findings contrast with published results that suggest the RGG domain is required for Xist RNA localization in mouse Neuro2a cells (Hasegawa et al., 2010). However, the interpretation of such

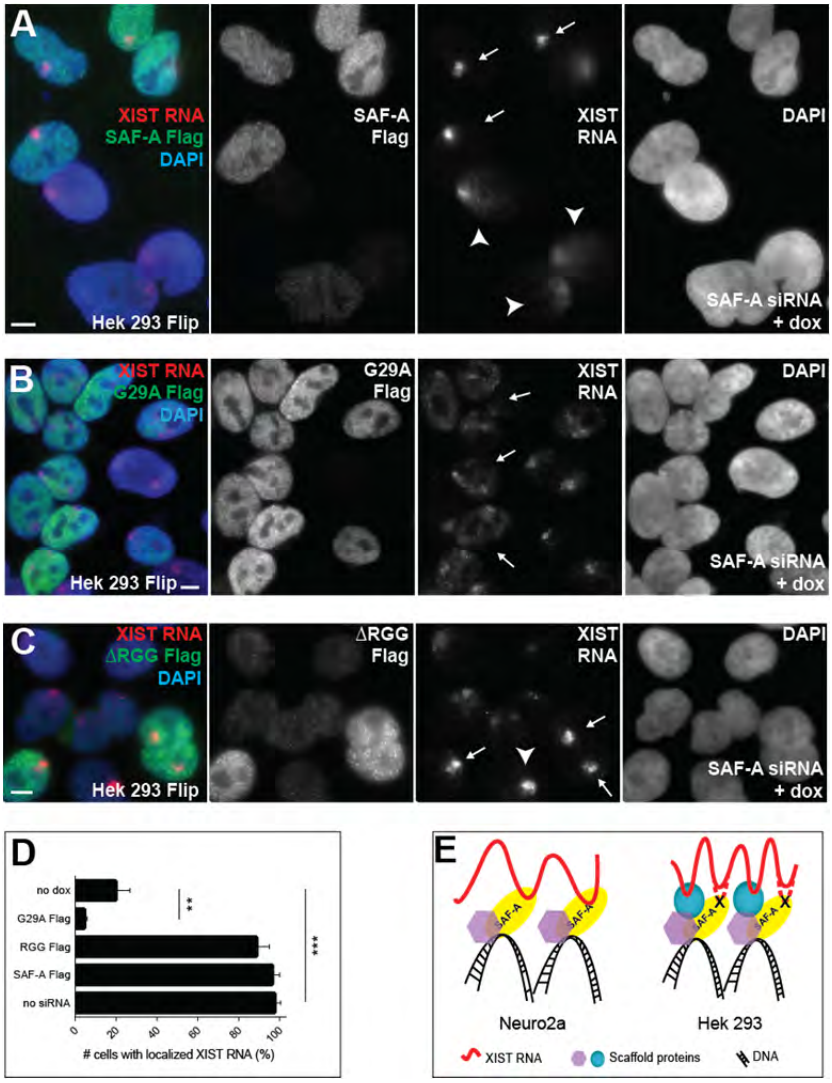

Figure 3.7 The DNA but not the RNA binding domain of SAF-A is required for XIST RNA localization in normal cells.

(A) Flag-tagged SAF-A rescues XIST RNA localization after SAF-A knockdown (arrows). XIST RNA is mislocalized in cells lacking SAF-A Flag (arrowheads). (B) G29A-Flag does not rescue XIST RNA localization after SAF-A knockdown (arrows). (C) ΔRGG-Flag rescues XIST RNA localization after SAF-A knockdown (arrows). Arrowhead points to cell with endogenous SAF-A only. Other cells in the field received SAF-A siRNA. (D) Quantification of XIST RNA localization in cells with SAF-A knockdown and Flag-tagged SAF-A expression. ***P* value = .0021, *** *P* value < .001 by unpaired two-tailed t-test. Error bars: standard error of the mean. (E) A model for localization of XIST/Xist RNA in Neuro2a cells (left) and Hek 293 cells (right). Neuro2a cells require SAF-A's RGG domain for Xist RNA localization and Hek 293 cells do not. Scale bars, 5 μm.

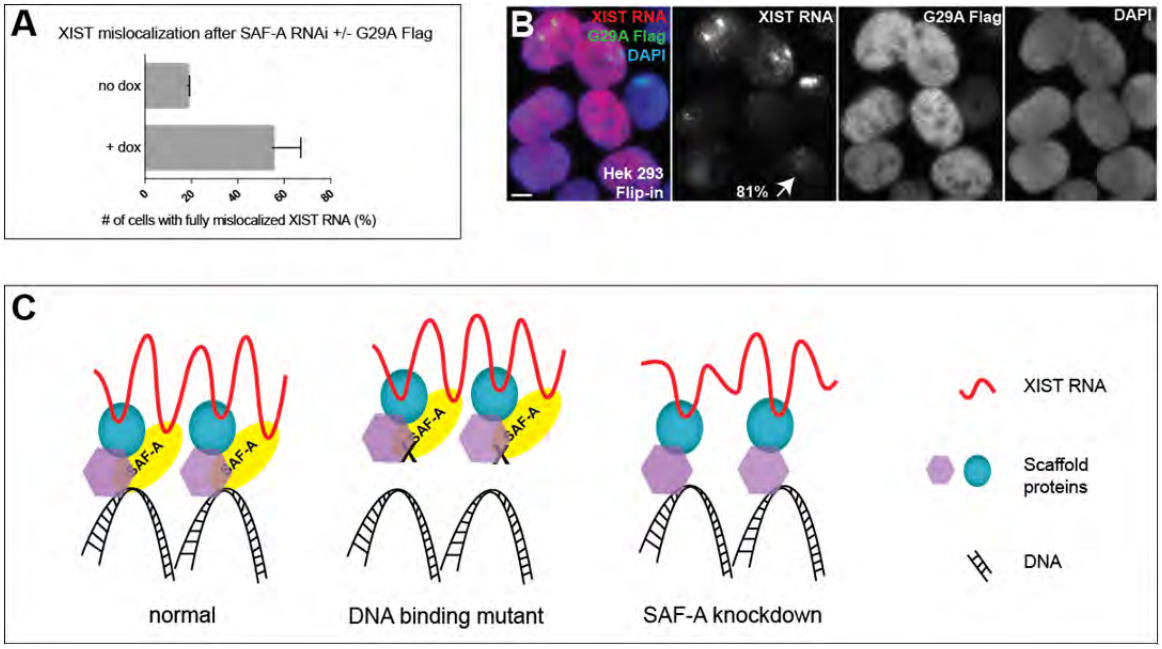

Figure 3.8 SAF-A DNA binding mutant impacts XIST RNA localization in mouse and human cells.

(A) Quantification of XIST RNA mislocalization after SAF-A siRNA depletion in Hek 293 cells expressing stable G29A-Flag (+ dox) or no G29A-Flag (no dox). Error bars: standard error of the mean. (B) XIST mislocalization in Hek 293 cells expressing stable G29A-Flag. (XIST RNA was fully mislocalized in 81% of Flag positive cells). Scale bars, 5 µm. (C) Models of XIST RNA localization in human cells with multiple anchor proteins, suggesting that additional XIST RNA anchor proteins have greater affinity for the SAF-A DNA binding mutant than for DNA, but they rescue XIST RNA localization in the absence of SAF-A.

experiments may be difficult due to over-expression of transiently transfected SAF-A. Supporting this, our transient transfections of mouse or human SAF-A constructs frequently mislocalized Xist RNA in Neuro2a cells, even when there was no siRNA depletion of endogenous SAF-A (Figure 3.9A-B).

While I was surprised that in our studies of human cells, the RGG domain was not required to maintain XIST RNA localization after SAF-A RNAi, the RGG domain may still be required for XIST RNA to *enrich* SAF-A concentration on the Xi (see Discussion), and for binding RNA in other capacities. For example, we have evidence that the RGG domain is required for binding of a different chromosomal RNA, CoT-1 (Figure 3.9C-D) (further explored in Chapter V). I also saw evidence the ΔRGG mutant affects XIST RNA transcription and splicing, which are among SAF-A's varied roles in the nucleus (Vizlin-Hodzic et al., 2011; Huelga et al., 2012; Xiao et al., 2012) (Figure 3.9E-G). It is also possible that SAF-A *does* contribute to XIST RNA anchoring through its RGG domain, but this is not required for XIST RNA localization in Hek 293 cells. For example, ΔRGG-Flag may anchor XIST RNA to DNA indirectly via a redundant XIST RNA binding protein(s), as shown by the model in Figure 3.7E. In contrast, above results indicate that Neuro2a cancer cells are deficient in other proteins that contribute to anchoring Xist RNA, hence SAF-A is completely required in that context (Figure 3.7E).

In vivo manipulation of XIST RNA localization shows SAF-A is not its only anchor

Above findings show that SAF-A contributes to XIST RNA's faithful chromosomal localization in a variety of human and mouse cell lines and types, but also that this biology is more complex than currently realized, and SAF-A is likely not the

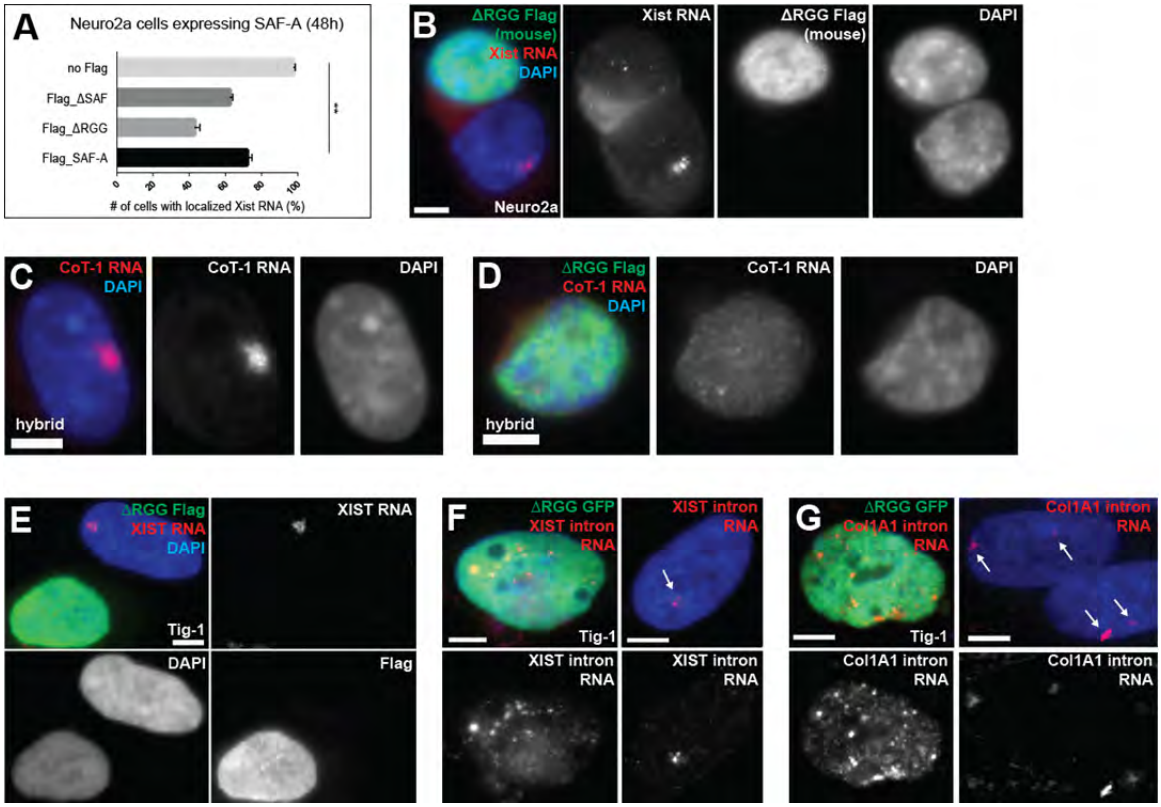

Figure 3.9 SAF-A's RGG domain potentially has multiple functions.

(A-B) Transient overexpression of mouse full length (A) and ΔRGG (A-B) SAF-A constructs releases Xist RNA in Neuro2a cells. Xist RNA was mislocalized in 48% of cells expressing low/moderate levels of ΔRGG-Flag. Error bars: standard error of the mean. (C-D) ΔRGG-Flag expression mislocalizes chromosome-associated CoT-1 RNA (C) from the human chromsome in mouse human hybrid cells (D). (E) 36% of transiently transfected Tig-1 cells expressing ΔRGG-Flag lacked any detectable XIST RNA, even the transcription focus. Many of these cells had recently gone through mitosis, and since XIST RNA needs to be re-synthesized in G1 daughter cells (Clemson et al., 1996), this implicates the RGG domain in XIST RNA transcription. (F-G) Transient ΔRGG-Flag expression causes aberrant accumulation of (F) XIST (47% of Flag-positive cells) and (G) Collagen (Col1A1) (10% of Flag-positive cells) intron-containing RNAs, supporting a role for SAF-A's RGG domain in the processing of nascent RNAs. Scale bars, 5 μm.

sole XIST RNA anchor in most normal cell types. The Lawrence lab earlier demonstrated an approach to investigate what controls XIST RNA localization by manipulating the binding versus release of XIST RNA from the chromosome during different phases of the cell cycle *in vivo*. This approach revealed that inhibition (by RNAi or hesperadin) of Aurora B Kinase, which phosphorylates histone H3 in early prophase, was sufficient to force *retention* of XIST RNA on metaphase chromosomes, when it normally detaches in early prophase. Results also indicated it was more difficult to force *release* of XIST RNA in interphase. Hall et al. (2009) speculated that multiple anchors may be present, but one or more was not controlled by AURKB. Since SAF-A has several known phosphorylation sites and is phosphorylated in mitosis (Douglas, Ye et al. 2015), I investigated if SAF-A normally detaches from chromosomes at prophase and whether AURKB controls its release. If AURKB inhibition forcibly retains SAF-A on all chromosomes, this would show it is the putative "AURKB-dependent" anchor. Alternatively, SAF-A detachment while XIST RNA is retained would support the existence of at least two anchors: SAF-A and an "AURKB-dependent" anchor.

I first addressed whether SAF-A normally detaches from chromatin at mitosis, in early prophase contemporaneously with XIST RNA. There are conflicting reports of whether SAF-A remains on chromosomes at mitosis (Pullirsch et al., 2010; Ma et al., 2011), so I looked at *endogenous* SAF-A using two specific SAF-A antibodies. This showed that SAF-A detaches from early mitotic chromosomes and disperses in the cytoplasm (Figure 3.10A). Interestingly, in many cells, SAF-A-GFP stayed on chromosomes throughout mitosis, suggesting the bulky GFP tag (and/or expression level)

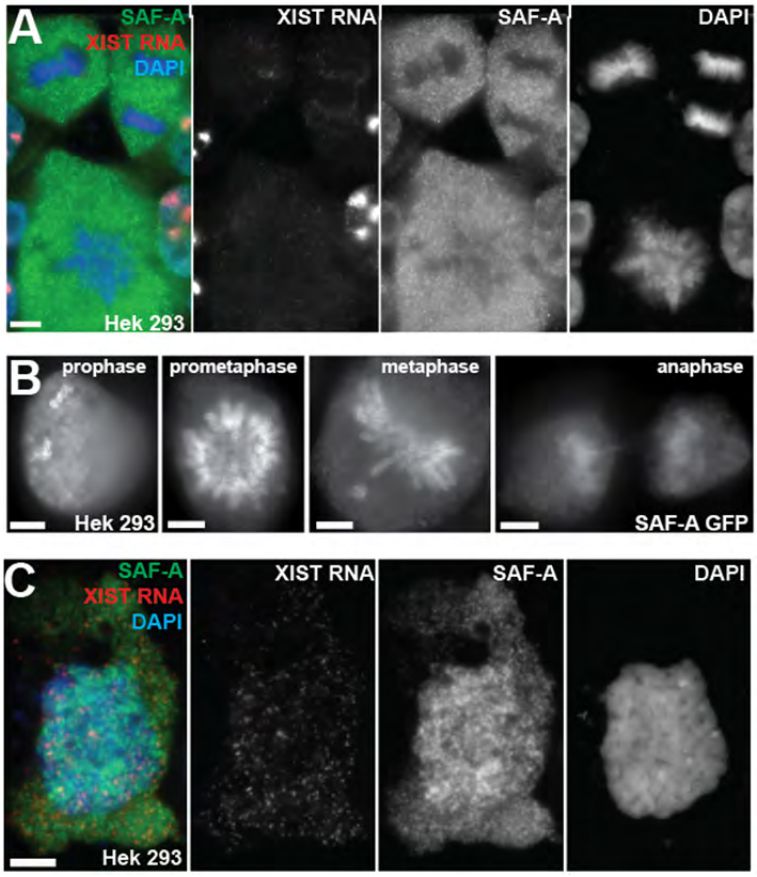

Figure 3.10 SAF-A releases from chromosomes in mitosis.

(A) Endogenous SAF-A releases from chromosomes during normal mitosis. (B) SAF-A-GFP stays on mitotic chromosomes throughout mitosis. The bulky GFP tag may prevent SAF-A release from DNA, perhaps by blocking a phosphorylation site. Alternatively, the stoichiometry of SAF-A to other nuclear proteins is important, and when the pool of these interaction partners is exhausted due to a high level of SAF-A, the "superfluous" SAF-A-GFP may bind DNA (and maybe RNA) in an uncoordinated, non-physiological way. (C) SAF-A releases from mitotic chromosomes in prophase shortly after XIST RNA. Scale bars, 5 μm.

impacts the protein's behavior (Figure 3.10B). Nonetheless, it is significant that endogenous SAF-A releases from prophase chromosomes at about the same time as XIST RNA (with some suggestion that XIST RNA may detach first (Figure 3.10C)). Thus it was of interest to address whether SAF-A is the AURKB-controlled anchor that forces XIST RNA to remain on Xi after AURKB inhibition.

I used hesperidin to inhibit AURKB in Hek 293 cells and examined localization of SAF-A and XIST RNA. As previously reported, XIST RNA is retained on the Xi under these conditions. Strikingly, SAF-A came off mitotic chromosomes as usual, but with a notable exception: much of it remained with XIST RNA on the Xi (Figure 3.11A). Since SAF-A still detaches from all other chromosomes, this indicates SAF-A's association with chromatin is not controlled by AURKB. Rather, SAF-A likely remains on Xi through its interaction with XIST RNA, which in turn remains on Xi via the anchor controlled by AURKB. Hence, SAF-A is the "AURKB-independent" anchor previously speculated to exist. These findings further validate that SAF-A interacts with XIST RNA *in vivo* and that XIST RNA is tethered by multiple proteins, whose chromatin association is controlled by distinct mechanisms.

While retention of one anchor can tether XIST RNA at metaphase, its release at interphase would require disruption of all anchors. I further tested this hypothesis by activating AURKB in interphase by RNAi to its inhibitor, Protein phosphatase-1 (PP1). I saw little release of XIST RNA (partially or fully mislocalized in 4% of cells) despite clear up-regulation of H3S10 phosphorylation (Figure 3.11B). This is likely due at least in part to the continued presence of SAF-A, an AURKB-independent anchor. Therefore, I

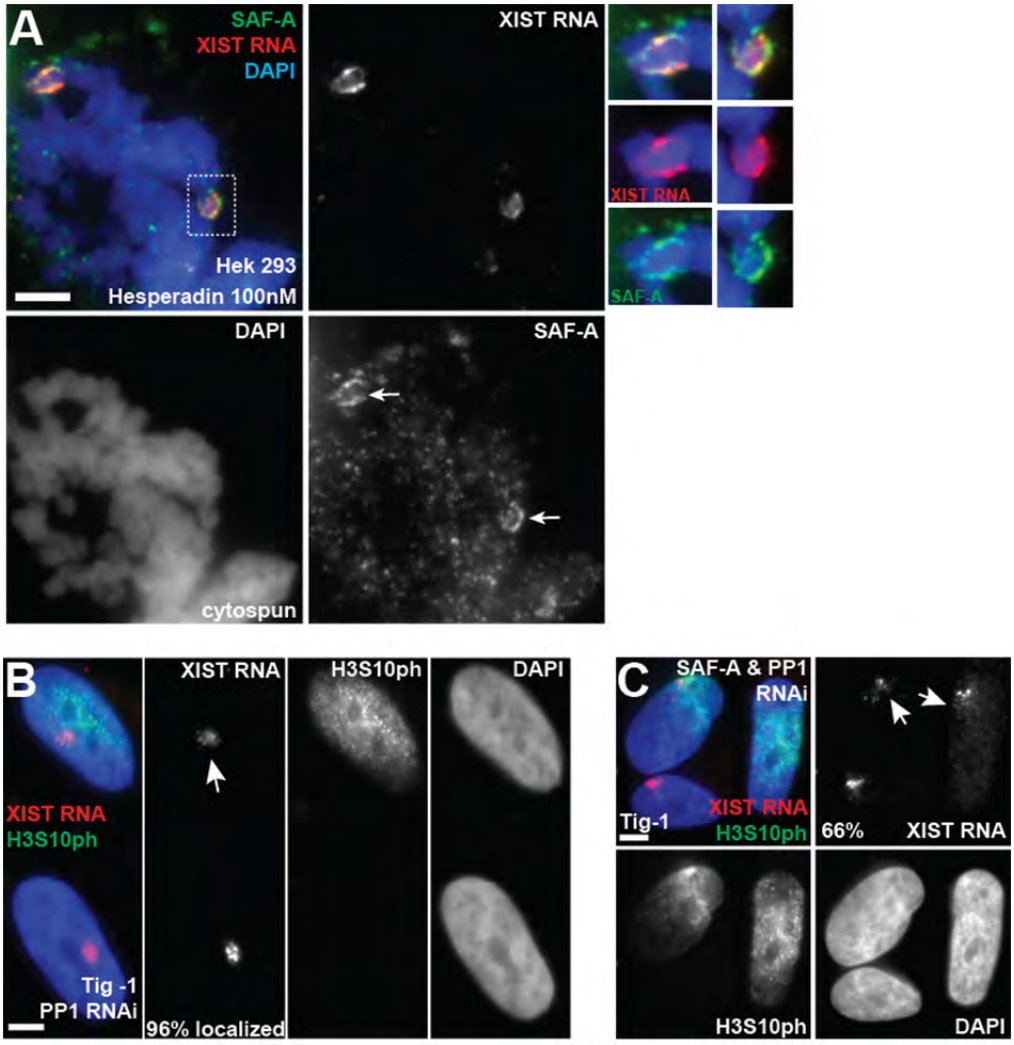

Figure 3.11 SAF-A is one of multiple XIST/Xist RNA anchor proteins.

(A) SAF-A comes off of all chromosomes except Xi when AURKB is inhibited by Hesperadin during mitosis (see arrows and inset). (B) AURKB activation (indicated by H3S10ph) by PP1 siRNA does not affect XIST RNA in interphase cells (96% of cells retain localized or loosely localized XIST RNA, arrow). (C) Simultaneous AURKB activation and SAF-A knockdown releases XIST RNA in the majority of Tig-1 cells (XIST is partially or fully released in 66% of H3S10ph positive cells, arrows). H3S10ph negative cell got no siRNA. Scale bars, 5 µm.

used RNAi to PP1 and SAF-A simultaneously to perturb both anchors in the same cells. Consistent with the hypothesis, 66% of cells had partially or fully mislocalized XIST RNA (Figure 3.11C), more than PP1 RNAi (or SAF-A RNAi, Figure 3.1) alone. Thus in these normal fibroblasts, XIST RNA has at least two anchors, and binding of one (or more) is controlled by AURKB, whereas binding of SAF-A is not. Finally, I note that even after disruption of two anchor proteins, XIST RNA release and dispersal in the nucleoplasm is not as complete as both Hasegawa et al. (2010) and I saw after depletion of only SAF-A in the tumor cell line, Neuro2a (Figure 3.2, 3.3A-B).

Forced retention of XIST RNA at metaphase can identify *bona fide* interacting partners *in vivo,* such as hnRNP K

As described in Chapter I, biochemists have struggled to identify proteins that directly interact with XIST RNA because it is embedded in complex nuclear structure. However, three recent studies used three variations of RNA pull-down assays to identify XIST RNA "interactors" (iDRiP, RAP-MS, and ChIRP-MS) (Chu, Zhang et al. 2015, McHugh, Chen et al. 2015, Minajigi, Froberg et al. 2015). While these studies advance XIST RNA biology by identifying what may best be considered *candidates* for XIST RNA interacting proteins, collectively, they also highlight the limitations of extraction-based biochemical approaches. The lists identified by each group contain striking differences, with only six proteins in common between all three data sets (Table 1.1).

As the approach inhibiting AURKB-mediated chromatin phosphorylation can provide direct *in vivo* evidence that a protein interacts with XIST RNA, I extended this analysis to three hnRNP proteins identified by the recent pull-downs. I treated Hek 293

cells with hesperadin and determined whether each protein remained with XIST RNA on the mitotic Xi. Interestingly, hnRNP K, a factor identified in two studies, also remains exclusively on the mitotic Xi during AURKB inhibition (Figure 3.12A-B), however hnRNP C and hnRNP M, identified by each of the three studies, do not (Figure 3.12C-D). Further, I looked at the localization of these proteins in interphase nuclei and found that hnRNP K, but neither hnRNP C nor hnRNP M substantially overlap with the Barr body (Figure 3.13A-C). While it is possible that hnRNP C and hnRNP M play some role in initiation of X inactivation in pluripotent cells, these findings highlight the importance that potential XIST RNA binding factors be validated *in vivo*. Future studies in pluripotent cells are needed to further test *in vivo* interactions of these and other candidates with XIST RNA.

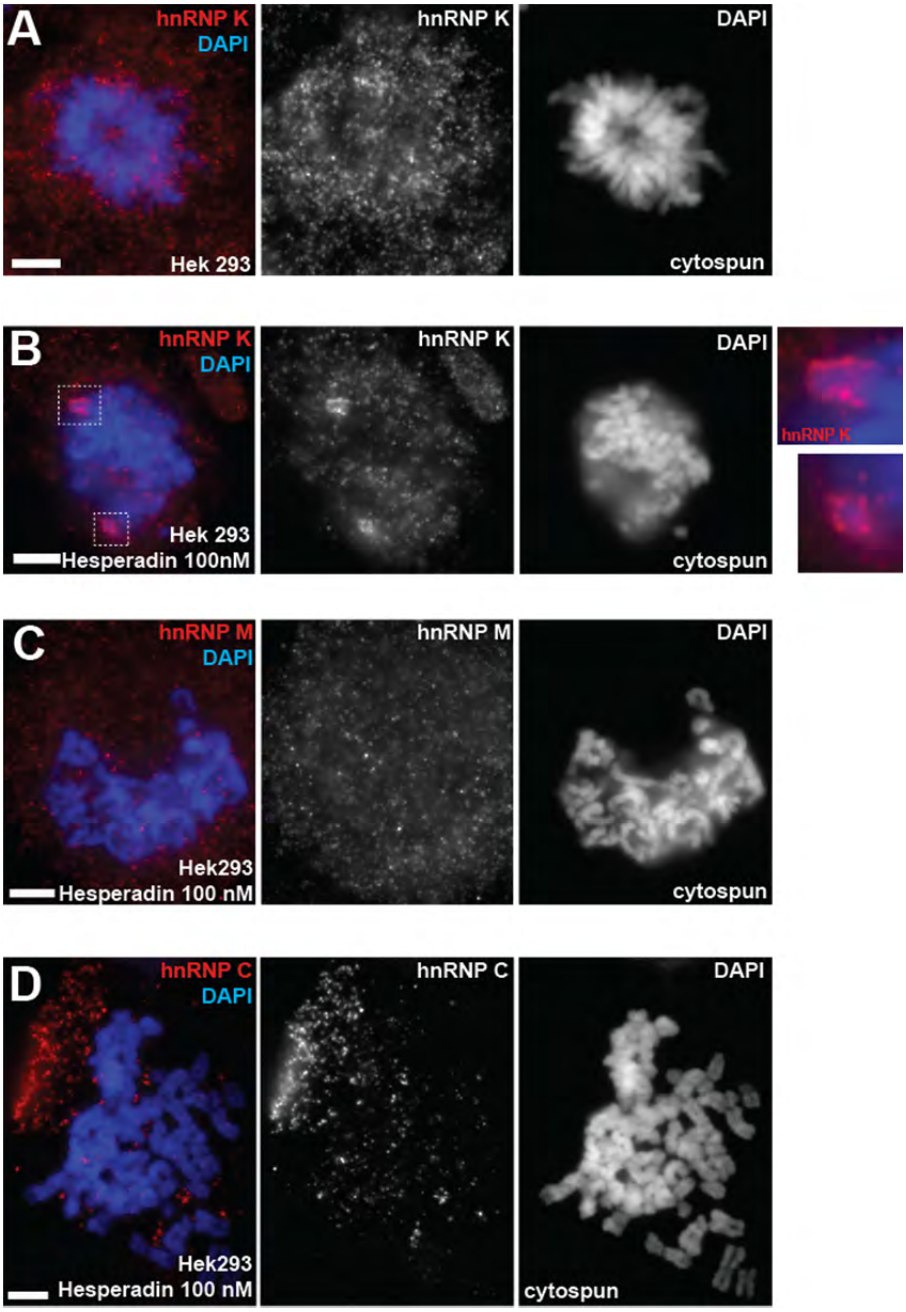

Figure 3.12 hnRNP K interacts with XIST RNA in vivo.

(A-B) hnRNP K normally releases from chromosomes during mitosis (A) but stays exclusively on the Xi when AURKB is inhibited by Hesperadin (see inset, red) (B). (C-D) hnRNP M (C) and hnRNP C (D) are released from all chromosomes including the Xi when AURKB is inhibited by Hesperadin during mitosis. Scale bars, 5 μm.

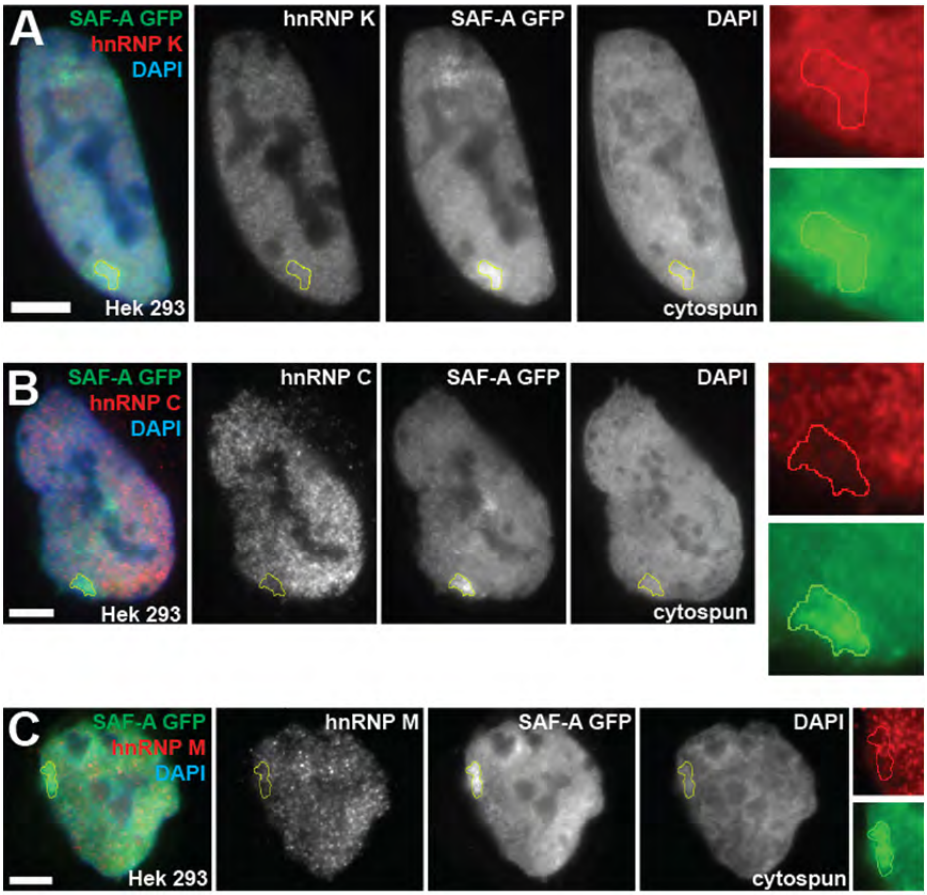

Figure 3.13 hnRNP K localizes with the Barr body during interphase but hnRNP C and M do not.

(A-C) hnRNP K (A) overlaps with the Barr body (enriched with SAF-A-GFP) in interphase Hek 293 cells but hnRNP C (B) and hnRNP M (C) do not. Scale bars, 5 μm.

Discussion

Interactions between chromosomal RNAs and chromatin are increasingly recognized as essential for regulating the epigenome, but how RNAs like XIST/Xist interact with chromatin is not well understood. The prevailing view is that SAF-A/hnRNP U is a single molecule bridge that tethers Xist RNA on the Xi during maintenance of X inactivation ((Hasegawa et al., 2010) cited e.g. in (Nakagawa and Prasanth, 2011; Gendrel and Heard, 2014)). I confirm a requirement for SAF-A to localize Xist RNA in Neuro2a cells; however, I discovered that Neuro2a cells are the only somatic cells tested in which SAF-A depletion fully and consistently releases XIST/Xist RNA from the chromosome. In several other human and mouse cell types, in which SAF-A is as or more effectively depleted, the effect on XIST RNA ranged from significant, but more modest and variable, to negligible. Hence, the biology of XIST RNA and the nuclear scaffold is far more complex than a single protein bridge between DNA and RNA.

The fact that XIST RNA still localized in primary human cells lacking SAF-A might initially suggest that SAF-A is not involved and a different protein anchors XIST RNA in these cells. However, our results support that SAF-A is involved, but redundant anchors together support faithful localization of RNA to the chromosome. For example, our experiments expressing the dominant negative SAF-A mutant in normal fibroblasts and manipulating AURKB support the presence of an additional anchor(s). I find that endogenous SAF-A detaches from all chromosomes in early prophase, as does XIST RNA, but its release is not controlled by AURKB chromatin phosphorylation. Thus, at

least one other protein still tethers XIST RNA to the mitotic chromosome when AURKB is inhibited. Further evidence for multiple anchors comes from experiments showing more extensive XIST RNA delocalization in interphase nuclei with both SAF-A depletion and AURKB activation (perturbing both anchors). Interestingly, even in this condition, XIST RNA was not as fully released as in Neuro2a cells treated with just SAF-A RNAi, suggesting there may be more than two anchor proteins for XIST RNA in normal somatic cells.

Results here demonstrate that the requirement for SAF-A to localize XIST/Xist RNA varies with the cells examined, and generally correlates with cellular transformation or immortality (and not with efficiency of SAF-A RNAi). While it is possible that the proteins binding RNA in the nuclear scaffold vary between all normal differentiated cell types, I favor the hypothesis that differences in XIST RNA phenotypes observed here relate to loss of redundant anchors in neoplastic cells. SAF-A is ubiquitous and essential for early development, and pluripotent stem cells share several properties with de-differentiated neoplastic cells. This includes less defined nuclear architecture and more open chromatin state (Meshorer and Misteli, 2006; Carone and Lawrence, 2013), as well as more loosely localized XIST RNA. Hence, I suggest that the more committed, differentiated state may be associated with tighter, redundant anchoring of chromosomal RNAs, to maintain a more fixed and stable epigenome.

It is a compelling, straightforward model that a single protein molecule bridges DNA to RNA. This may occur in some contexts, but our studies encountered more complexity. Results support that human SAF-A's DNA binding domain localizes the

"XIST RNA complex" to the chromosome, and I identify a single nucleotide required for this function. While SAF-A is often characterized primarily as an RNA binding protein, it was shown to have much greater affinity for DNA *in vitro*, and electron microscopy revealed its remarkable ability to organize DNA into loops (Fackelmayer et al., 1994). Hence, I have compelling evidence to support earlier studies showing the contribution of SAF-A's DNA binding domain to XIST RNA chromosomal association (Hasegawa et al., 2010; Pullirsch et al., 2010) and the higher-order folding of chromatin in response to XIST RNA.

Having obtained initially unexpected results from transient transfections, I used multiple strategies to validate that the RGG domain is not essential for SAF-A's contribution to XIST RNA localization in the human cell types examined. It is possible that SAF-A binds XIST RNA somewhere other than its RGG domain. I note that eight of the ten proteins suggested to directly bind Xist RNA by McHugh et al (2015) contain RRM RNA binding motifs (not RGG), which SAF-A and LBR lack (McHugh et al., 2015). Interestingly, in other work I find the RGG domain is required for localizing a different RNA (described in Chapter V), and the RGG domain may be required for SAF-A *enrichmen*t on Xi (Helbig and Fackelmayer, 2003), but not to support XIST RNA localization. SAF-A does not become *enriched* on Xi until after Xist coats the chromosome during initiation of X inactivation (Pullirsch et al., 2010). Thus evidence indicates there may be two "layers" of SAF-A: ubiquitous SAF-A integral to all chromosomes which supports XIST RNA localization, and an enriched layer on Xi which requires XIST RNA. The enriched layer of SAF-A may be that which remains bound

when XIST RNA is retained at mitosis following AURKB inhibition. Hasegawa et al. (2010) reported that transient expression of neither the SAF-box nor the RGG deletion mutant could rescue Xist RNA localization after SAF-A RNAi in neuroblastoma cells. Despite our finding that transient overexpression of even wild-type SAF-A can cause XIST RNA mislocalization, the RGG domain may not be required in some cells because other SAF-A associated scaffold proteins bind XIST RNA, whereas Neuro2a cells lack these additional proteins.

Here, I have investigated factors that influence XIST RNA anchoring to chromatin directly within the cell. This study supports that SAF-A/hnRNP U is one major player connecting long XIST/Xist transcripts, embedded with scaffold proteins, to chromosomal DNA. These findings point to new, interesting, and impactful directions into how this fundamental biology changes with the epigenetic and neoplastic state of cells, and how the RNA/scaffolding structure is disassembled and reassembled during mitosis.

Chapter IV : Abundant RNA from interspersed repeats is a stable component of euchromatic interphase chromosome structure

Preface

Work presented in this chapter resulted from a joint effort by Lawrence lab members to characterize the heterogeneous CoT-1 RNA signal visualized by RNA FISH, and it produced the following publication:

Hall, L.L., D.M. Carone, A.V. Gomez, **H.J. Kolpa**, M. Byron, N. Mehta, F.O. Fackelmayer, and J.B. Lawrence. 2014. Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell*. **156**:907-919.

This chapter contains a slightly abridged version of the published manuscript and includes my contribution, which was to analyze the association of CoT-1 RNA with the nuclear scaffold. The chapter also includes some unpublished results in which I investigate the behavior of specific repeat families that may be differentially regulated.

This work was supported by NIH grant GM053234 and GM107604 to J.B.L. and NIH/NCI NRSA 1F32CA154086 to D.M.C. The authors thank Jeff Nickerson for helpful discussion of EM results and Lara Strittmatter from the UMMS Electron Microscopy core for the EM fixation and analysis. The EM Core facility is supported by NIH award S10RR027897.

Introduction

Recent studies recognize a vast diversity of noncoding RNAs with largely unknown functions, but few have examined interspersed repeat sequences, which constitute almost half our genome. Although some noncoding RNAs (ncRNAs) are thought to interact with specific loci, RNA is not generally considered a broad component of chromatin or chromosomes. Our interest in exploring this understudied half of the genome is bolstered considering how XIST RNA enacts X chromosome silencing, by propagating across and "painting" the whole chromosome, yet remaining strictly localized within the boundary of the chromosome territory (Clemson et al., 1996).

Here, *in situ* analyses show that RNA broadly and stably associates with euchromatin and that the predominant component of this chromatin-associated RNA is surprisingly abundant CoT-1 RNA from interspersed repetitive elements, including LINE-1. The unusual properties of CoT-1 RNAs are distinct from short-lived nascent transcripts and indicate that CoT-1 repeat RNAs comprise a class of chromosomal RNAs, which persist long after transcriptional inhibition, and remain localized strictly with the interphase chromosome territory in *cis*. Further, the RNA is tightly associated with euchromatin and the nonchromatin scaffold, but it is silenced by XIST RNA on inactivated chromosomes. Finally, we suggest that chromosomal RNA associated with euchromatin, comprised largely of CoT-1 repeat RNA, may help maintain open chromatin packaging.

Materials and Methods

Cell Culture and Treatments: DM1 myoblasts, HT1080 G3 (Hall et al., 2002a), Wi38, Tig-1, NIH 3T3, and GM11687 were grown under conditions recommended by supplier (ATCC, DMI, and myoblasts from Charles Thorton, Wellstone Muscular Dystrophy Cooperative Research Center, Rochester). Inhibitors: Actinomycin D (5–20 μg/ml), 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) (40–80 μg/ml), or α-amanitin (5–50 μg/ml) (dissolved in DMSO) were added to culture media (4–6 hr, also 16–32 hr for α-amanitin). Nocodazole (100 ng/ml) -arrested cells were released into inhibitors for 4–6 hr to score inhibited G1d cells. Cytospin: interphase cells were trypsinized and cytospun at 8,000–10,000 RPM onto glass coverslips prior to fixation. Matrix prep: digestion of DNA and histones was performed as described previously (Clemson et al., 1996). RNase: unfixed cells on coverslips were permeabilized with 0.1% triton-X in CSK buffer (4°C for 3 min) and then treated with 5 μl/ml DNase-free RNase (Roche 11119915001) in cytoskeleton (CSK) buffer (20 min) or in PBS (3 hr, 37°C) for fixed cells. SAF-A: 2 μg/ml GFP-tagged full-length and C280 SAF-A were used to transfect cells with Lipofectamine 2000 (Invitrogen). Cells were fixed after 16, 24, 48, or 72 hr. Highly overexpressed cells were not scored.

Cell Fixation, FISH, and Immunofluorescence: This was done as described in chapter II. Human tissue blocks were cryosectioned and then fixed. The absence of cytoplasmic CoT-1/L1 RNA was confirmed using fixations that preserve cytoplasmic RNAs

(Clemson et al., 1996; Hall et al., 2009). See the Figure 4.9 legend for standardized beads.

FISH Probes: We used human L1.3 ORF2 and L1.3 ORF1 (from J. Moran, University of Michigan), 2 kb of DMPK intron 9 (Table 4.1), the 10 kb human XIST pG1A construct, 10 kb 18S rDNA (from G. Stein, UMMS), the 330 bp Human Alu pPD39 clone (ATTC), human and mouse CoT-1 DNA (Roche and Invitrogen, respectively), and two Chr 4 paints (Qbiogene & Cytocell). Introns (nonrepetitive) for *XIST*, *COL1A1*, and *GAPDH* were PCR-generated and cloned into pSC-A (Stratagene) (Table 4.1). Oligos used were 54-mer Poly-dT, 55-mer 5s rRNA, and 33-mer human Alu (Table 4.1).

Antibodies: Antibodies used were HP-1 Gamma (Chemicon), Histone H3K9me3 (Upstate), and Rad 21 (Abcam).

Northern Blot: Total RNA was extracted from Tig-1 fibroblast cell pellets using Trizol (Invitrogen) per the manufacturer's instructions. 10μg total RNA was DNase-treated (Promega RQ1 RNase-free DNase) then electrophoresed on a 1% agarose denaturing gel and transferred to a nylon membrane (Ambion Brighstar Plus) using a semi-dry electroblotter at 400 mA for 1 hr and UV crosslinked. 10ng Biotin-labeled L1ORF2 plasmid was hybridized overnight in UltraHyb buffer (Ambion Northernmax) in a rotating hyb oven at 37°C and washes were performed according to manufacturer's instructions (Ambion Brightstar Biodetect kit).

3D Colocalization Analysis: 0.5μm Z-stacks were deconvolved using Zeiss Axiovision or Volocity (Perkin Elmer) Restoration package. Images were quick projected for publication quality images or co-localization analysis was performed in 3D using Volocity co-localization analysis. A thresholded Pearson's correlation coefficient was generated from 3D images to take into account background levels for individual fluorophor channels.

L1 qPCR: Trizol-extracted Tig-1 fibroblast RNA was reverse transcribed with IScript Supermix (Bio-Rad) and then RT-qPCR for L1 was performed and analyzed as per (Carone et al., 2013). Expression levels obtained with primer sets for L1 5′ UTR: Fwd: 5′ GAACAGCTCCGGTCTACAGC Rev: 5′TCACCCCTTTCTTTGACTCG and L1 3′UTR: Fwd: 5′TGATGAGTTCATATCCTTTGTAGGG Rev: 5′GATATTCCCCTTCCTGTGTCC were normalized to levels for β-ACTIN: Fwd: 5′AGCGAGCATCCCCCAAAGTT Rev: 5′GGGCACGAAGGCTCATCATT.

DNA/RNA qPCR: DNA or RNA was extracted from $1.4X10^5$ Tig-1 fibroblast cells with phenol-chloroform or Trizol, respectively. RNA was then reverse transcribed as above and qPCR was performed from Tig-1 cDNA and genomic DNA with the primer sets as above. Dilution calculations were taken into account to normalize for the amount of DNA or RNA going into the initial PCR reaction (to account for amount subjected to reverse transcription and dilution of DNA). qPCR analysis was performed as above. Results shown (Figure 4.14) are in log scale.

Transmission Electron Microscopy and Fixation: Mitotic shake-off cell cultures grown for 2 hr in the presence of 40ug/ml DRB or 4uL/ml DMSO (control) in 6 well plates were fixed by adding 2.5% glutaraldehyde in 0.1 M Na Cacodylate buffer (pH 7.2) to the culture plates one drop at a time until the initial volume of media was doubled. The cell cultures were allowed to stabilize in this solution for 10 min, then all the media/glutaraldehyde was removed and fresh 2.5% glutaraldehyde in the same buffer was added and the cells were allowed to fix for 60 min. at room temperature. After this primary fixation, the cells were rinsed three times in fresh fixation buffer for 10 min. each time and were secondarily fix with 1.0% osmium tetroxide in ddH2O for 30 min at room temperature. The cell cultures were then washed again three times in ddH2O and subjected to a tertiary fixation with 1% Uranyl Acetate in ddH2O for another 30 min. After two more washes with ddH2O the cells were dehydrated through a graded series of ethanol (10% to 100%; 3 changes), and then transferred to ethanol 100%: SpiPon 812/Araldite 502 resin (50:50 / V:V) for 12 hr at room temperature. Following infiltration in the resin, the cells were transferred through 3 changes of pure SpiPon 812/Araldite 502 epoxy resin each 1 hr long. The cells were then changed to a final step of embedding resin mixture and were polymerized for two days at 70°C in their original culture dishes. Each individual culture plate was separated from the others, and then the rims of each dish were sawed off. The individual sawn plates were then placed in liquid nitrogen to separate the bottom of the dish from the cells embedded in the resin. Pieces of the embedded cultures were then mounted onto blank stubs with a drop of super glue and the

blocks were trimmed and sectioned (100 nm thick). Sections were collected onto 200 mesh copper support grids and imaged using both a Philips CM 10 and Tecnai 12 BT Transmission under 80Kv accelerating voltage. Images were recorded with a Gatan Erlangshen CCD Digital camera.

Microscopy and Image Analysis: This was done as described in Chapter II/III. Volocity from Perkin Elmer and Metamorph digital imaging software was used to quantify signals.

Table 4.1 Probe Sequences:

Probe Name	**Sequence**	**Label**	**Reference**
Sat2-LNA (Oligo)	5'ATTCCATTCAGATTCCATTCGATC	5' Biotin	Exiqon, Product # 200501-03
Sat3 (Oligo)	5'CCATTCCATTCCATTCCATT	3' Biotin	Prosser et al., 1986 (Invitrogen)
HuAlpha Sat (Oligo)	5' CCT TTT GAT AGA GCA GTT TTG AAA CAC TCT TTT TGT AGA ATC TGC AAG TGG ATA TTT GG	3'& 5' FITC	Biosource & Invitrogen
HuAlu (Oligo)	CCC AAA GTG CTG GGA TTA CAG GCG TGA GCC ACC	3'& 5' FITC	Biosource
DMPK intron 9 (primers)	Fwd: 5'GAGTTGCAGGATCAGTCTTGGA Rev:5'TGCTGATTCTCTGGTGGAGAAC	Biotin or Dig PCR label	
L1 5' UTR (primers)	Fwd: 5'GAACAGCTCCGGTCTACAGC Rev: 5'TCACCCCTTTCTTTGACTCG	Biotin or Dig PCR label	
L1 3' UTR (primers)	Fwd: 5'TGATGAGTTCATATCCTTTGTAGGG Rev: 5'GATATTCCCCTTCCTGTGTCC	Biotin or Dig PCR label	
Poly A (oligo)	T…(54)	5' FITC	
5s rRNA	CCTGCTTAGCTTCCGAGATCAGACGAG ATCGGGCGCGTTCAGGGTGGTATGGCC G	Alexa 594	
COL1A1 Intron 6	Fwd: 5'GTTCTGCTTGTGTCTGGGTTTC Rev: 5'AGTAAGTTCAACCTTCCCCCTCC		
XIST Intron 1	Fwd: 5'TCCCATCACTGTTCAATGCC Rev: 5'ATTGTGGACCCAGCTCTATGC		
GADPH Intron 1	Fwd: 5'ATGGAGGTCCTCTTGTGTCCC Rev: 5'AGACAAGAGGCAAGAAGGCAT		

Results

CoT-1 repeat RNA is abundant in mammalian nuclei

CoT-1 DNA fraction is routinely used to block nonspecific hybridization of genomic probes to repeats; however, here, we use labeled CoT-1 DNA is as a probe to detect RNAs containing high copy repeats (CoT-1 RNA). We previously showed that CoT-1 RNA hybridization provides a convenient assay to identify silent heterochromatic regions within nuclei by the absence of hnRNA hybridization signal (Hall et al., 2002a; Tam et al., 2004; Clemson et al., 2006). CoT-1 hnRNA is broadly distributed throughout the nucleoplasm in all interphase cells but is absent from nucleoli, cytoplasm, and DAPI-dense heterochromatic regions (Figure 4.1). This robust CoT-1 RNA signal is ubiquitous in all primary and cancer cell lines (mouse and human) and frozen tissue sections examined (Table 4.2). Several lines of evidence establish this signal is single-stranded RNA and not DNA: it is detected under nondenaturing conditions (which does not detect highly abundant DNA sequences) (Lawrence et al., 1989), it is eliminated by RNase A (Figure 4.2), and the CoT-1 RNAs detach from chromatin and disperse to the cytoplasm at mitosis (Figure 4.1A).

The abundance of CoT-1 repeat RNA is indicated by a bright signal broadly distributed through most of the nucleus, the intensity of which is close to or exceeds that of highly abundant nuclear RNAs such as rRNA or XIST RNA (Figure 4.3). For example, the linescan in Figures 4.3C-D illustrates that the intensity of CoT-1 RNA signal is similar to that of XIST RNA on the X chromosome territory, but CoT-1 RNA is much more widely distributed (and thus abundant). Likewise, by measuring the average

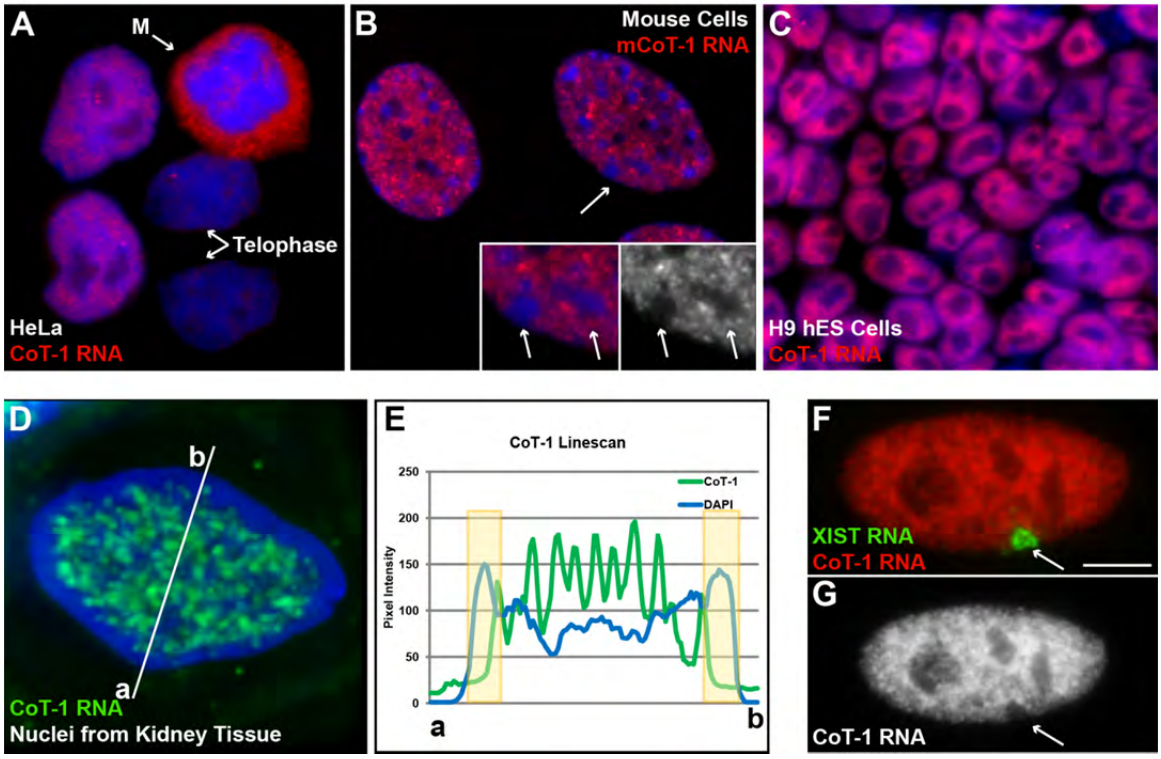

Figure 4.1 CoT-1 RNA is expressed in all mammalian cells examined.

(A and B) All blue signals are DAPI DNA. CoT-1 RNA is expressed in 100% of interphase cells, e.g., HeLa (M = metaphase) (A) and NIH 3T3 (B). Inserts: CoT-1 RNA exclusion from chromocenters (inset arrows) of indicated cell (arrow). (C and D) (C) Human ES cells and (D) frozen human tissue sections. (D and E) Linescan across nucleus shows CoT-1 RNA absence from peripheral heterochromatic compartment (yellow zones). (F and G) XIST RNA (arrow) paints the inactive X chromosome in female cells, while CoT-1 RNA is silenced. Scale bar, 5 μm.

Table 4.2 Cell types positive for nucleoplasmic CoT-1 RNA.

HUMAN CANCER LINES
U2OS - Osteosarcoma
SAOS-2 - Osteosarcoma
PC3 - Prostate Adenocarcinoma
HEP-G2 - Hepatocellular carcinoma
HT-1080 - Fibrosarcoma
MCF-7 - Breast Adenocarcinoma
HCC-1937 - Breast Ductal Carcinoma
T47D - Breast Ductal Carcinoma
MDA-MB-436 - Breast Adenocarcinoma
MDA-MB-231 - Breast Adenocarcinoma
SUM-149pt - Inflammatory breast cancer
JAR - Choriocarcinoma
HELA - Cervical Adenocarcinoma
HUMAN NON-CANCER LINES
MCF-10A - Breast Fibrocystic Disease
IMR-90 – Primary Lung Fibroblast
WS1 – Primary Embryonic Skin Fibroblast
TIG-1 – Primary Fetal Lung Fibroblast
Wi38 –Primary Fetal Lung Fibroblast
HFF – Primary Foreskin Fibroblast
HSMM – Myoblasts & Myotubes
HUMAN NORMAL TISSUES
1880N – Renal
2081N – Fallopian Tube
2312N – Pancreas
2632N – Thyroid
MOUSE LINES
NIH 3T3 - Fibroblasts
MEFs – Embryonic Fibroblasts
3X female line
GM11667 – Mouse-Human Hybrid cells

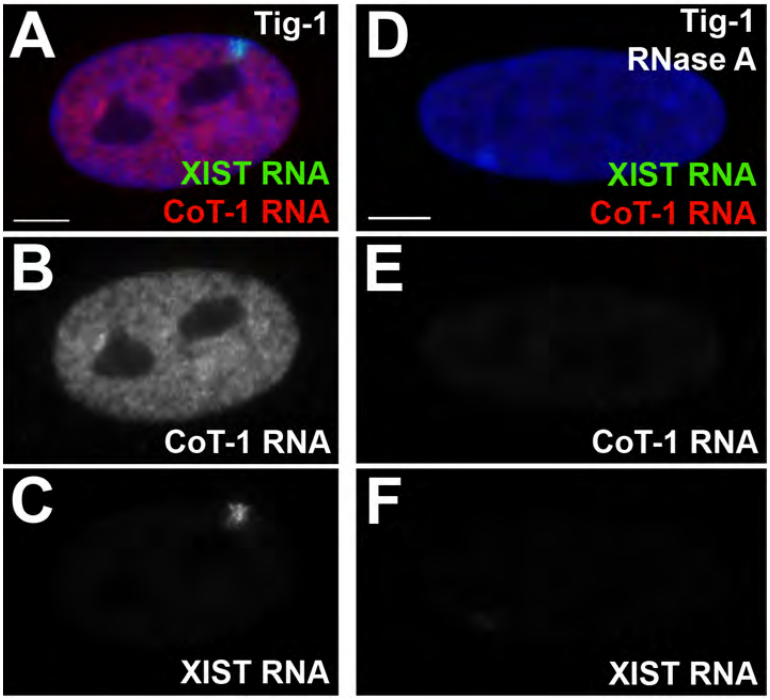

Figure 4.2 CoT-1 RNA signal is removed with RNase treatment.

All blue signals are DAPI DNA. (A-F) RNase A treatment removes the majority of CoT-1 RNA and XIST RNA signals in all treated cells (D-F) compared to controls (A-C). The RNA signals are separated in images B-C & E-F. All images were taken at same exposure. All scale bars are 5 μm.

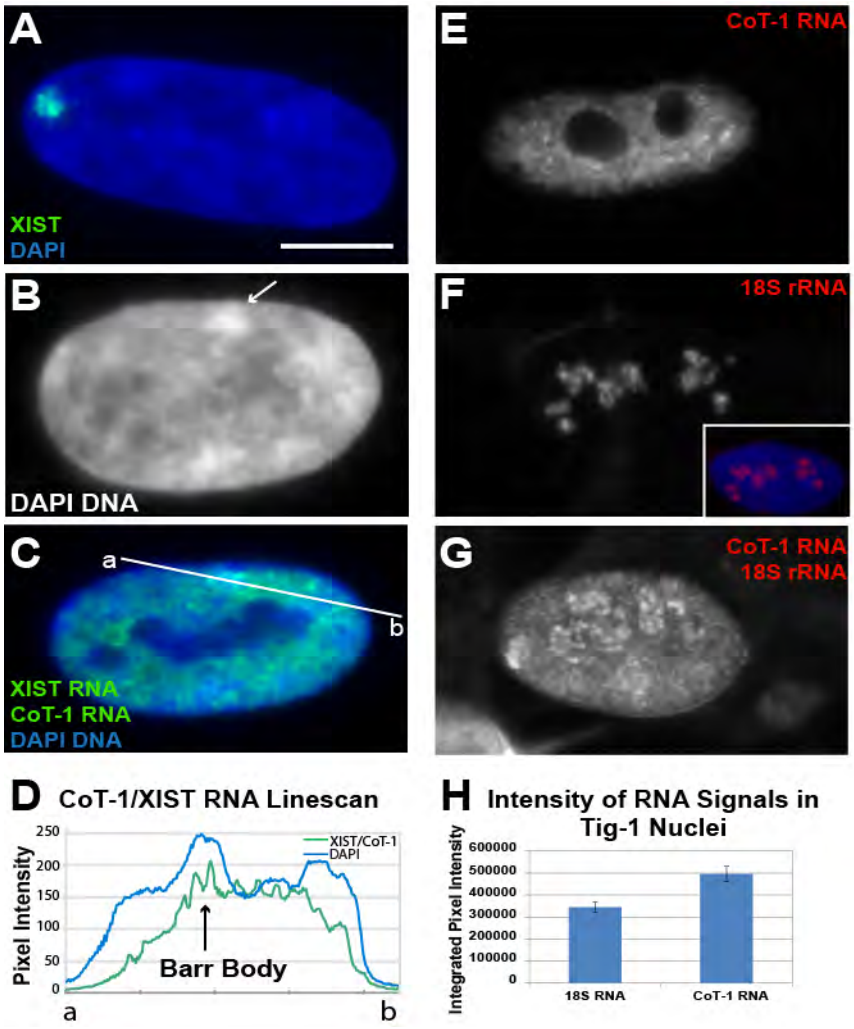

Figure 4.3 CoT-1 RNA appears highly abundant.

(A) XIST RNA paints interphase X chromosome territory. (B) The DAPI dense Barr body (arrow) defines the inactive X chromosome painted by XIST RNA. (C) Both CoT-1 and XIST RNAs are detected simultaneously with the same fluorochrome. (D) Linescan (white line in C) through the Barr body (arrow) illustrates similar intensities for both RNAs (green). (E–G) CoT-1 RNA is more abundant than rRNA. All images are the same exposure. (E) CoT-1 RNA alone, (F) nucleolar 18S rRNA alone (insert is same cell with DAPI DNA included), and (G) CoT-1 RNA and 18S rRNA detected with same fluorochrome. (H) Average total intensities were measured for each RNA signal (n = 40). Error bars represent 95% confidence interval. Scale bar, 5 μm.

total RNA signal for rRNA and CoT-1 RNA per nucleus, we found that the CoT-1 RNA signal exceeds that of abundant nuclear rRNA (Figures 4.3E–H). The surprising abundance and distribution of this nuclear repeat RNA signal led us to investigate its stability and relationship to chromosome structure.

CoT-1 RNA localization is tightly restricted to the parent chromosome territory

XIST RNA is unique in that it coats and is restricted to its parent chromosome territory in interphase nuclei and does not disperse to the surrounding nucleoplasm (Figure 4.4A), unlike mRNAs or many excised introns (Figure 4.4C-D) (Clemson et al., 1996; Johnson et al., 2000; Smith et al., 2007). This finding was key to understanding that an ncRNA could have an unanticipated role in regulating chromatin structure and function. To investigate whether CoT-1 RNA showed a structural relationship to the chromosome of origin, we needed a strategy to examine ubiquitous repeats expressed from just an individual chromosome.

To this end, we studied the localization and behavior of human CoT-1 RNA in mouse somatic hybrid cells (GM11687) containing a single human Chromosome 4. Although mouse and human genomes have similar families and patterns of interspersed repeats, the primary sequences differ sufficiently such that human CoT-1 does not hybridize to mouse DNA (or RNA) and vice versa. RNA FISH in the hybrid cells showed a striking result: CoT-1 RNA is restricted to a tightly defined nuclear territory (Figure 4.4B). This is highly reminiscent of the XIST RNA territory over the inactive X chromosome (Figure 4.4A). Sequential hybridization to CoT-1 RNA and DNA (in two different colors) demonstrates that both the RNA and DNA territories appear as sharply

bordered structures with coincident boundaries (Figures 4.4E-F) and are highly co-localized by 3D imaging analysis (Figures 4.4G-I), which is similar to XIST RNA and its parent chromosome (Figures 4.5A-B). Additionally, human CoT-1 RNA does not overlap with mouse CoT-1 RNA (Figure 4.5C), indicating that they remain associated with their parent chromosomes.

Although XIST is a structural RNA tightly bound to its parent chromosome in *cis*, it is released from chromatin in mitotic cells and resynthesized in early G1d cells (Figures 4.6A-C). CoT-1 RNA behaves similarly and disassociates from the chromosome at prophase, where it disperses in the cytoplasm and is resynthesized in G1d cells (Figures 4.6D-F).

Most RNA stably associated with interphase chromosomes is composed of repeats

The abundance and localization of repeat RNA to its parent chromosome suggests that it may not be simply a byproduct of "genic" transcription. To further evaluate this, we compared the abundance of repeat RNAs associated with the interphase chromosome to the collective contribution of unique RNA from a whole Chr 4 library.

Chromosome libraries are designed to detect specific sequences throughout a particular chromosome. Although libraries are depleted of repeats, they still contain some, requiring competition with cold CoT-1 DNA to enhance the library specificity (Figures 4.6G and 4.5D-E). Using two commercially available Chr 4-specific library probes to detect RNA in hybrid cell nuclei, we found that the RNA territory over the human chromosome was only weakly defined (Figures 4.6H-I) in contrast to the well-defined, bright CoT-1 RNA territory over the same chromosome. This was not simply

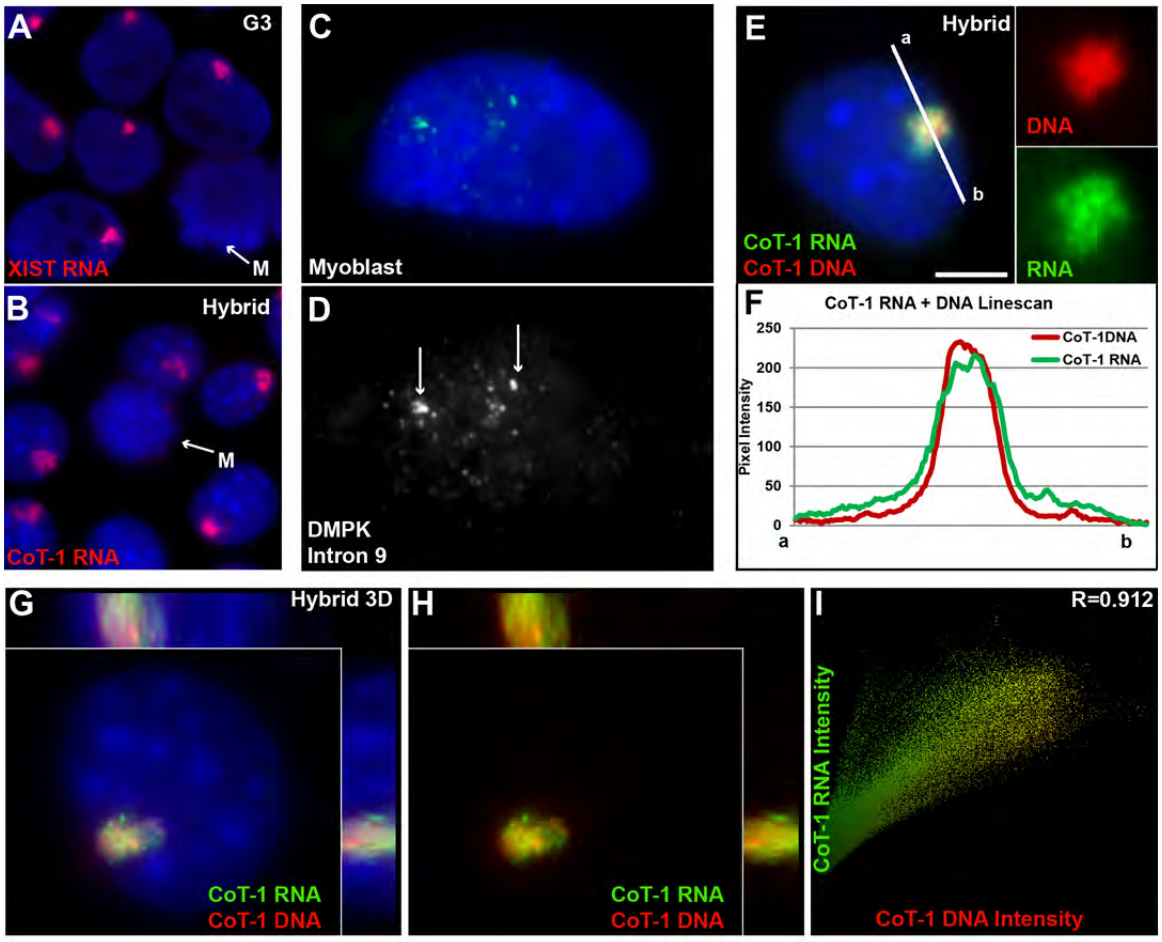

Figure 4.4 CoT-1 RNA localizes to the chromosome similar to XIST RNA.

(A) XIST RNA is strictly localized to the inactive chromosome in interphase. (B) Human CoT-1 RNA is also strictly localized to human Chr 4 in all interphase hybrid cells. Both RNAs are released at mitosis (M) (arrows). (C–D) Excised introns do not localize to chromatin and drift away from their transcription sites. RNA channel separated in (D) and shown in black and white. Transcription foci are denoted by arrows. (E) CoT-1 DNA identifies human Chr 4 in hybrid cells painted by CoT-1 RNA (channels separated at right). (F) A linescan (white line in E) shows that the RNA has a sharp border at the edge of the chromosome territory. (G) 3D image of CoT-1 DNA and RNA on human Chr4 in a single hybrid nucleus, with side views. (H) Same image as (G) with DAPI DNA removed. (I) CoT-1 DNA and RNA signals are co-localized. Scale bar, 5 μm.

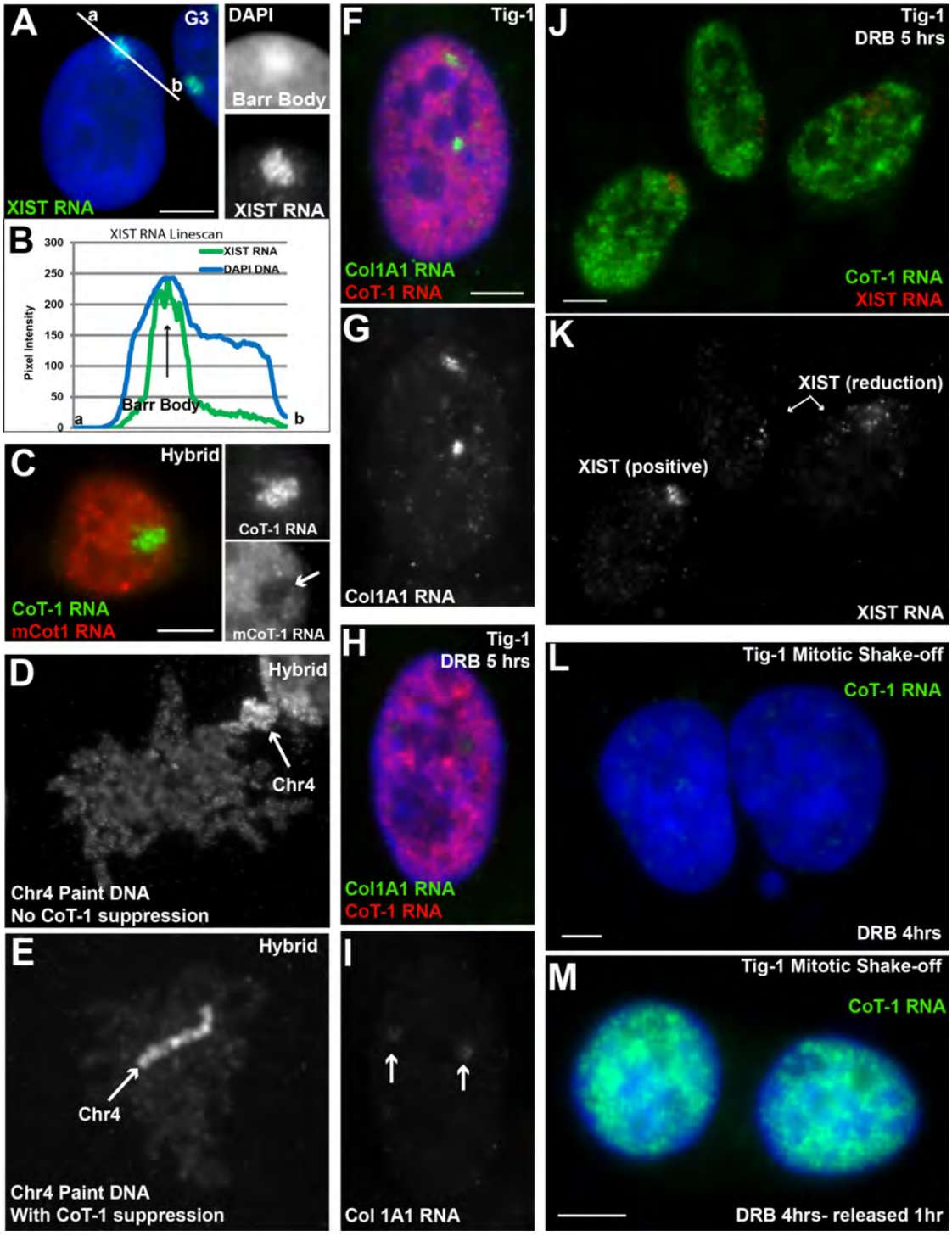

Figure 4.5 CoT-1 RNA remains localized, is more stable than XIST or Collagen in interphase, and is resynthesized after release from DRB.

Figure 4.5 CoT-1 RNA remains localized, is more stable than XIST or Collagen in interphase, and is resynthesized after release from DRB.

All blue signals are DAPI DNA. (A and B) (A) XIST RNA is localized to the inactive chromosome and delineates the DAPI dense Barr body (BB). Close up of XIST-covered chromosome separated into DNA and XIST RNA channels at right. White line is region in linescan. (B) A linescan shows a defined XIST RNA border over the DAPI dense BB. (C–E) (C) Mouse and human CoT-1 RNA do not co-localize, and remain associated with their respective chromosomes, making a "hole" in the mouse CoT-1 RNA signal (arrow) over the human chromosome territory. Close up of region separated into mouse CoT-1 and human CoT-1 RNA channels at right D-E) Human Chr 4 library probe non-specifically cross-hybridizes with other chromosomes in a metaphase spread (DNA stain of chromosomes not shown), due to the presence of residual CoT-1 repeats in the probe (D), which are effectively repressed with unlabeled CoT-1 competition (E). (F and G) Col1A1 makes very large transcription foci in many fibroblast nuclei. (H and I) Most cells lack collagen RNA after 5 hr of transcriptional inhibition, although a low number (18%) still show a residual signal (arrows). Images F-I are at the same exposure. (J and K) CoT-1 RNA signal is fairly uniform in interphase Tig-1 cells treated with DRB for 5 hr, but the XIST RNA signal is highly variable in this same treatment, with over half the population showing significant reduction or no XIST signal remaining (arrows). XIST was more variable with DRB than we had seen before. (L and M) All G1 daughter cells from mitotic shake-off show full repression of CoT-1 re-synthesis in the transcriptional inhibitor DRB, and within 1 hr of DRB removal CoT-1 expression has recovered in all cells. Scale bars, 5 μm.

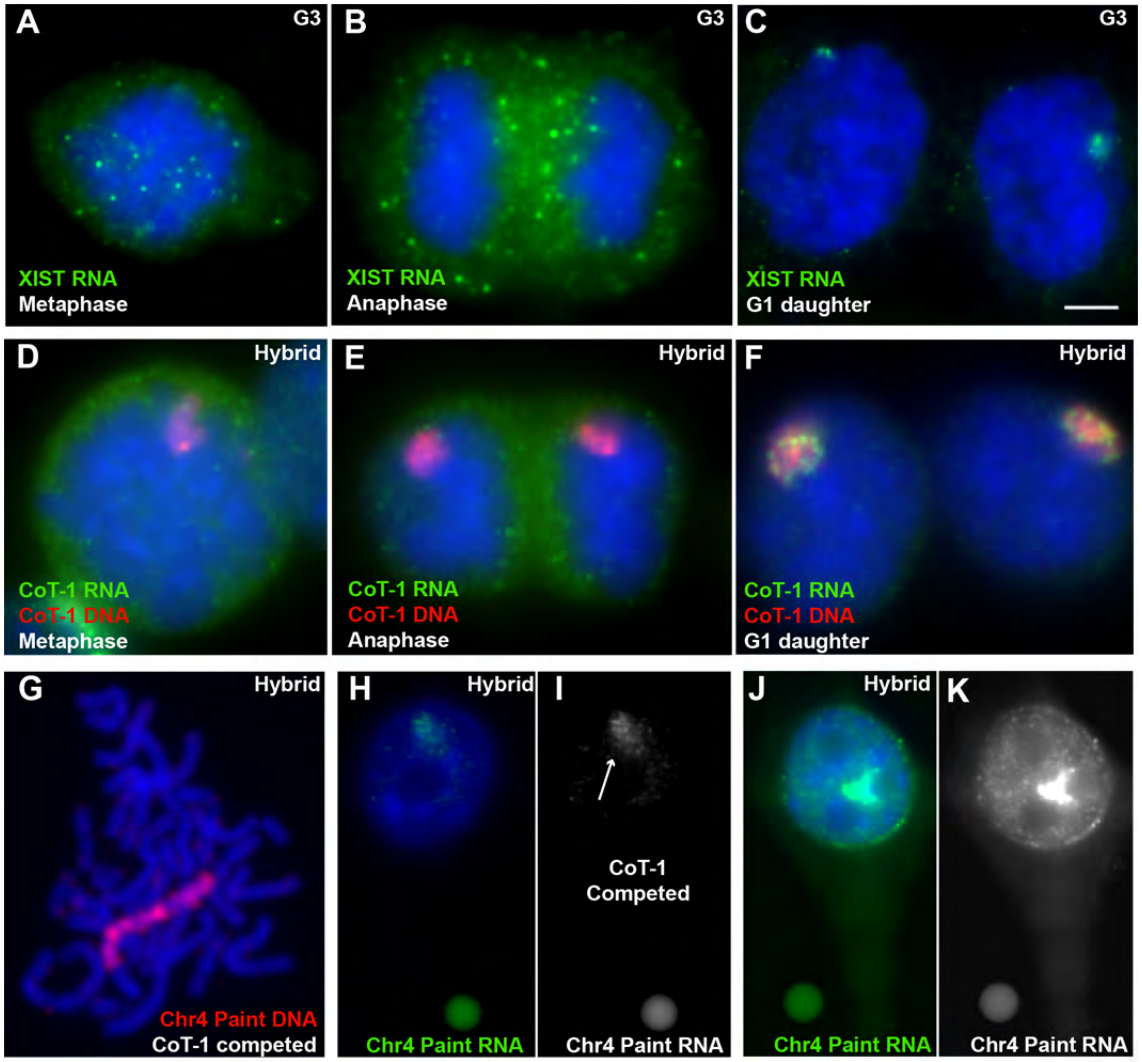

Figure 4.6 The majority of RNA associated with interphase chromosomes is repeat RNA, which is released at mitosis and resynthesized in G1.

(A–C) XIST RNA is released at mitosis and resynthesized in early G1d cells. (D–F) CoT-1 RNA is also released at mitosis and resynthesized in G1d. (G) A Chr 4 library probe effectively detects the unique single-copy DNA sequences across the chromosome (with CoT-1 competition). DAPI DNA in blue. Green channel separated at right (I). (H and I) The same Chr 4 library probe detects only a weak RNA signal (arrow) over the Chr 4 territory (with CoT-1 competition). (J and K) More RNA is detected on the Chr 4 territory when CoT-1 competition is removed. DAPI DNA in blue. Green channel separated at right (K). (H)-(K) are taken at the same exposure with standardized fluorescent beads (small round objects in images). Beads are 2.5 μm. Scale bar, 5 μm.

due to the complexity of the Chr 4 library because it labels the unique sequences along the whole chromosome well by DNA FISH (Figure 4.6G). Instead, this suggests that the total amount of steady-state RNA that accumulates over the chromosome may be substantially less from unique sequences than from repetitive sequences. To further examine this, we removed the cold CoT-1 competitor from the hybridization reaction, and this identical Chr 4 library yielded a large increase in RNA signal over the territory (Figures 4.6J-K). This supports that the vast majority of RNA sequences associated with an interphase chromosome territory are repetitive in nature.

CoT-1 transcripts are stable following transcriptional inhibition

The relative prevalence of repeat-containing RNA with interphase chromosomes could be due to either a high transcription rate or accumulation of stable RNA. Although transcriptional inhibitors can have complex effects on RNA metabolism (Bensaude, 2011), most nascent transcripts or introns are reported to be relatively short-lived with transcriptional inhibition. For example, the small nascent transcription focus of XIST intron RNA is no longer detected 1 hr after inhibition, whereas the mature XIST RNA has a longer (~5–6 hr) half-life, which was part of the initial evidence for a novel type of chromosomal RNA (Clemson et al., 1996). To assess CoT-1 RNA stability when transcription is arrested (by multiple distinct mechanisms), we used three different transcription inhibitors: 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB), Actinomycin-D (ActD), and α-amanitin (α-aman) in hybrid and human cells. Numerous experiments with the different inhibitors consistently demonstrated exceptional stability of CoT-1 RNA in interphase nuclei (see Materials and Methods). In fact, because of this

unusual stability, we relied on G1d cells to confirm that the inhibitors were working (i.e., preventing resynthesis). As detailed in Figures 4.5, 4.7 and 4.8, CoT-1 RNA was not resynthesized in 90%–100% of G1d cells in all three inhibitors but remained robust in 93%–100% of nuclei that had not divided.

Detailed analysis was performed with DRB in human fibroblasts to compare interphase CoT-1 RNA stability with mRNA transcription (COL1A1 and GAPDH) and with the relatively long-lived XIST RNA (Figures 4.5, 4.7, and 4.8). Five hours in DRB was sufficient to essentially eliminate COL1A1 RNA transcription foci in interphase Tig-1 nuclei (Figures 4.7I–L), with only 18% retaining a barely visible signal (Figures 4.5H-I), and this was also seen using intron probes. In contrast, the CoT-1 RNA, though somewhat reduced, remained in 100% of these same nuclei and persisted longer than XIST RNA (Figures 4.5J-K). CoT-1, XIST, COL1A1, and GAPDH RNA all were absent and not resynthesized in inhibited G1d cells (Figures 4.7K-L and data not shown). Upon removal of the reversible DRB inhibitor, 100% of G1d cells re-expressed CoT-1 across the nucleus within an hour (Figures 4.5L-M). Taken together, the persistence of CoT-1 RNA in these transcriptionally inhibited interphase cells is due to stability, not continued synthesis.

We used highly extended treatments with α-amanitin to further examine the stability of the RNA and were surprised to see that a bright RNA signal remained after 16–32 hr. In fact, comparison of the RNA signal to a standard fluorescent bead showed

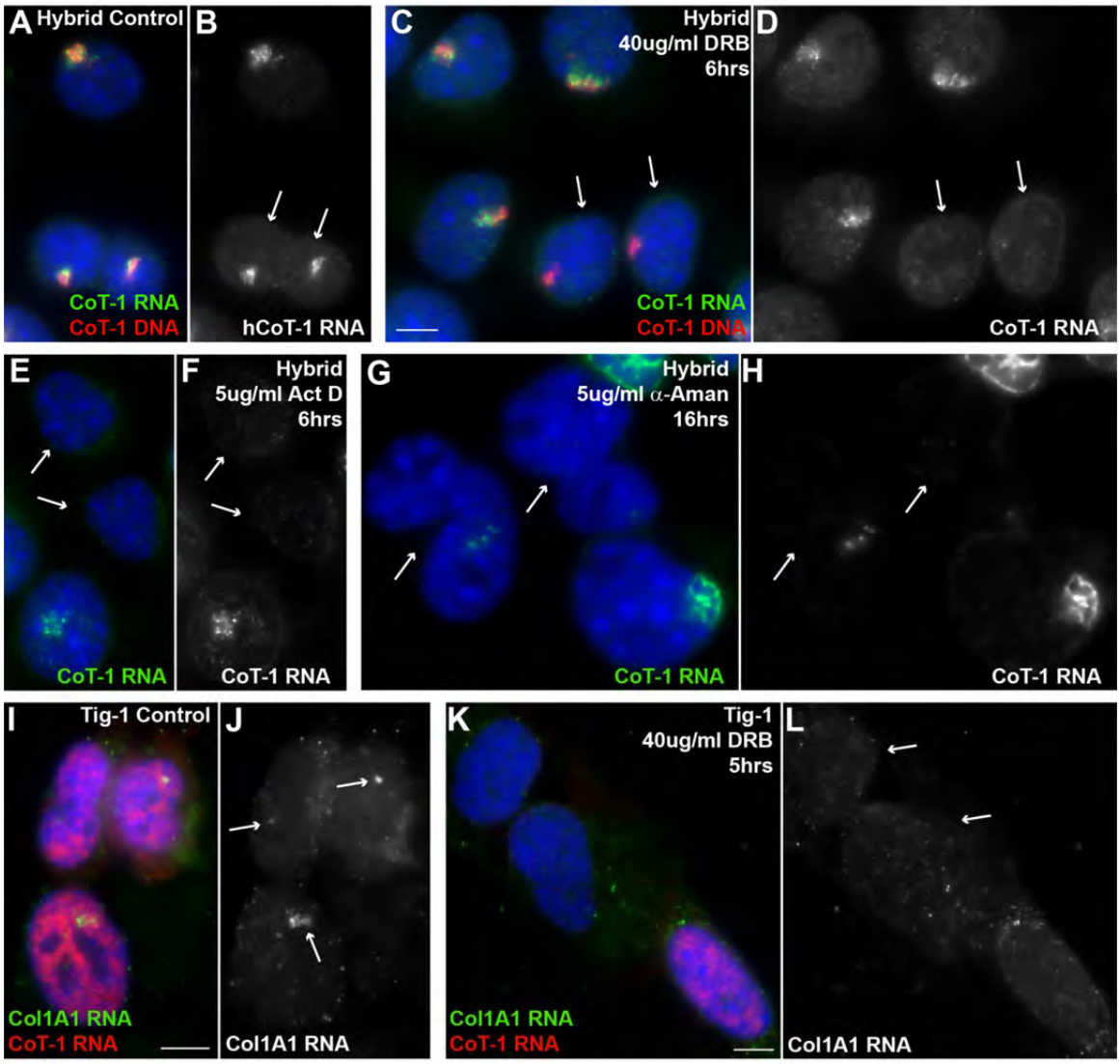

Figure 4.7 CoT-1 RNA localization is very stable under transcriptional inhibition.

All blue signals are DAPI DNA. (A and B) CoT-1 RNA is resynthesized early in G1d (arrows). (C–H) Transcriptional inhibition using three inhibitors (C–H) prevents CoT-1 resynthesis in 95%–100% of G1d cells (arrows) but does not significantly affect CoT-1 RNA levels in 93%–95% of interphase cells. The RNA channel is separated and shown in black and white to the right of the merged color image. (I–L) COL1A1 transcription foci are detected in 95% of interphase and G1d cells (arrows). (I and J) COL1A1 RNA foci are eliminated in 82% of cells after 5 hr in DRB, but only G1d cells (arrows) also lack CoT-1 RNA. Scale bars, 5 μm.

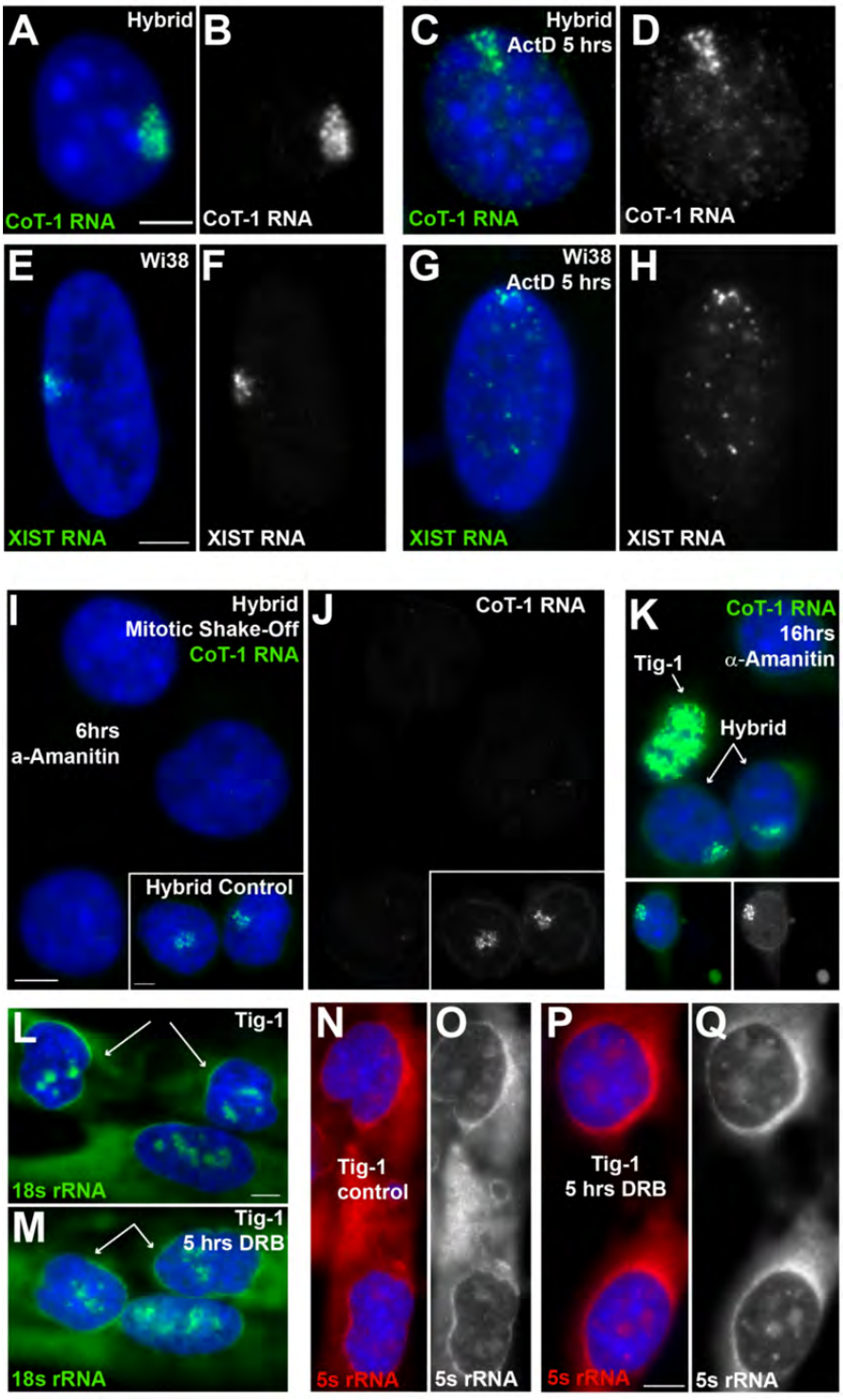

Figure 4.8 CoT-1 RNA is stable under long-term transcriptional inhibition, some is mislocalized by ActD, and evidence suggests that RNAPII may transcribe some of it.

Figure 4.8 CoT-1 RNA is stable under long-term transcriptional inhibition, some is mislocalized by ActD, and evidence suggests that RNAPII may transcribe some of it.

All blue signals are DAPI DNA. (A–H) The DNA intercalating inhibitor, Actinomycin-D, sometimes affected localization of both CoT-1 RNA (A-D) (A–B is a control cell) and XIST RNA (E-H) (E-F is a control cell). RNA channels are separated in black & white. (I and J) α-amanitin was effective in preventing re-synthesis of CoT-1 RNA in 92% of mitotic shake-off (G1d enriched) cells. Control cells in insert. (K) Even high concentrations of α-amanitin for 16 hr did not affect the localization of CoT-1 RNA in interphase Tig-1 and hybrid cells (~93% of cells had localized CoT-1 RNA). Both cell types were cultured and hybridized together on the same slide. Insert shows the intensity of the CoT-1 RNA territory in treated cells compared to the bead standard. Bead is 2.5 μm. (L and M) RNAP I transcribed 18 s rRNA is present in nucleoli of all control (L) and DRB-treated (M) G1 daughter cells (arrows). (N–Q) RNAP III transcribed 5 s rRNA is present in nucleoli of all control (N-O) and DRB treated (P-Q) post-mitotic, G1 daughter cells. Scale bars, 5 μm.

that the signal actually became brighter in most cells at both concentrations (5 and 20 μg/ml), which was seen in multiple experiments (Figure 4.8K). Although this may relate to the extraordinary stability of the RNA, as considered in the Discussion, it is possible that this is due to increased synthesis of some repeat RNAs in response to stress. Because 18 s rRNA (RNAPI) and 5 s rRNA (RNAPIII) were seen in G1d nuclei under conditions in which CoT-1 RNA was not (Figures 4.8L–Q), this suggests that much of the CoT-1 RNA signal could be RNAPII regulated. However, the increased interphase expression with prolonged α-amanitin potentially implicates the involvement of RNAPIII. These results are consistent with other recent evidence that there is a complex interplay between RNAPII and RNAPIII transcription (Raha et al., 2010).

An important observation is that the repeat RNA consistently maintained its tight localization to the chromosome territory in *cis*, unlike poly-A and other nuclear RNAs, which typically redistribute if they persist following transcription arrest (discussed in (Hall et al., 2006)). CoT-1 RNAs did not significantly disperse during short or long periods of DRB or α-amanitin (Figures 4.7A–H), suggesting a structural association with the chromosome territory. The only exception was that, in ActD, some portion of the CoT-1 RNA drifted from the parent chromosome in hybrid cells (Figures 4.8A–B, E-F). This was also seen with XIST RNA in female fibroblasts (Figures 4.8E–H), suggesting the interesting possibility that this DNA intercalating drug displaces some chromosome-bound RNAs.

Collectively, these data indicate that the abundance of the repeat RNA signal reflects a steady-state accumulation of stable RNAs on the chromosome territory, rather than a high rate of their transcription.

5′ Truncated L1 sequences are a prominent component of CoT-1 RNA

Because CoT-1 DNA will detect a mixture of repeat-containing transcripts, we sought to identify specific components of CoT-1 RNA and evaluate the relative expression of different repeat families. Using specific probes to several different repeat families (Table 4.1), we determined that satellite RNAs (α, SatII and SatIII) and several simple sequence repeats did not contribute appreciably to the broadly distributed nucleoplasmic CoT-1 RNA signal. In contrast, both L1 and Alu RNAs showed nucleoplasmic signal, although initial observations suggested that their levels may differ. Comparison of overall expression of repeat families is complicated by differences in size, number, and divergence of genomic repeats. To circumvent both this and technical differences in probes, we developed a strategy based on the ratio of nuclear RNA to DNA signal as an indication of the amount of RNA detected per unit of DNA. Thus RNA:DNA ratios for L1s (5′ and 3′) and Alu (Table 4.1) were compared (see Materials and Methods). We introduced fluorescent beads as an intensity standard (Figure 4.9) for more rigorous quantitative comparisons within and between slides.

Full-length (transposable) LINEs (6 kb) have a canonical 5′ promoter transcribing two open reading frames (ORF1 and ORF2), encoding a reverse transcriptase and endonuclease that can “copy and paste” the retrotransposon elsewhere in the genome.

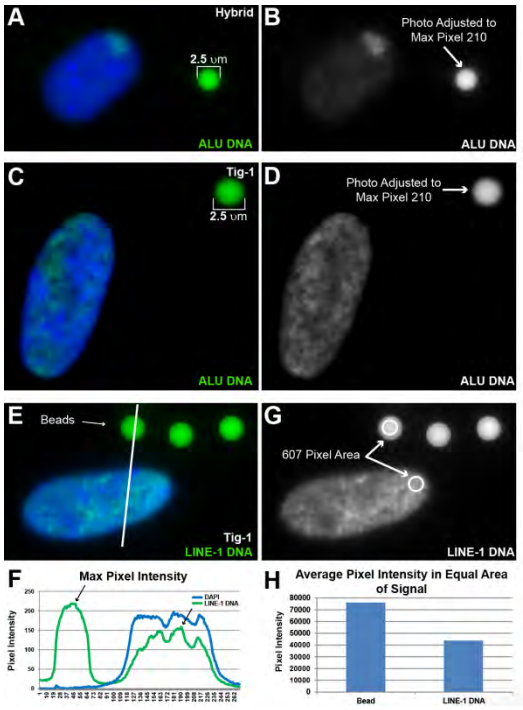

Figure 4.9 Use of fluorescent beads for *in situ* standardization/quantitative comparison.

(A–H) We use standardized microscope intensity calibration beads of uniform intensity (~3% relative intensity) and size (2.5μm) (Invitrogen/Molecular Probes InSpeck Green (505/515) Microscope Image Intensity Calibration Kit) to provide standardization for quantitative comparisons between different samples hybridized under identical conditions in the same color. These kits contain a series of well-defined fluorescent intensity levels covering the range of intensities and fluorescence spectra (red, green & blue) commonly encountered in microscopy applications. Appropriately colored beads (same Fluorescence Ex/Em spectra as labeled probed to be measured) are included in the mounting media on the slides with the cells. Calibration of different sample images is done at the time of photographing (all exposure settings adjusted to bead intensity), or can be accomplished later by manipulating image enhancement with software such as Photoshop, ensuring that all beads in the images are of equal intensity (A-D). This allows you to compare signal intensities across different experiments that may not have been taken at the same exposure or on the same day. It also provides an internal “scale bar” of 2.5 μm. The beads allow us to measure relative probe signal intensities more reliably between different samples. For example, linescans (E-F) or area measurements (G-H) (including whole nuclei area measurements as in Figures 4.3E–H) can be used to measure the signal intensities of the same labeled probe on different samples, normalized to the bead. White line in (E) is region measured in linescan (F). Max pixel intensity for green bead and cell signal indicated by arrows. Signal intensities measured in nucleoli (if signal to be measured is not present there) or on the glass slide are subtracted as background.

Most of the ~500,000 human copies are truncated and lack the 5′ promoter, with the few full-length L1 elements suppressed by promoter methylation in somatic cells (Lander et al., 2001; Goodier and Kazazian, 2008). Thus, the vast majority of L1 are thought to be silent in normal cells. We first used a 4 kb probe to ORF2 of human L1 (L1.3) to broadly assess for presence of L1 RNAs (Figure 4.10A).

L1 ORF2 RNA signal was abundant through the nucleoplasm but was not in the cytoplasm, nucleoli, or heterochromatin, similar to CoT-1 RNA (Figures 4.10B–G). In hybrid cells, L1 RNA formed a bright, well-defined RNA territory, precisely overlapping the corresponding Chr 4 DNA territory (Figures 4.11A-B) and also persisted following transcription arrest. Using the quantification strategy described above, L1 RNA signal was roughly 4-fold more than its abundant DNA signal. The RNA:DNA ratio (range 2.3–4.5) was comparable in hybrid cells and normal fibroblasts (Figures 4.11A–D, I and Table 4.3); hence, L1 RNA level is not an aberration of hybrid cells (Figures 4.10H–M).

Because full-length L1s are 5′ truncated, we measured the RNA:DNA ratios for probes to the 5′ or 3′ ends (Figure 4.10A and Table 4.3). L1 ORF2 RNA was consistently several-fold brighter than the corresponding DNA, whereas the opposite was true for the 5′ ORF1 (RNA:DNA ~1:4) (Figures 4.11E–H and 4.12A). Preference for 3′ L1 RNA was confirmed using smaller probes of identical size (200 nt) to the 5′ versus 3′ UTRs of L1Hs (Figures 4.12B–F) and was further supported by 5′ versus 3′ L1Hs RT-qPCR (Figure 4.11J). Consistent with other evidence that full-length L1Hs are silenced, results

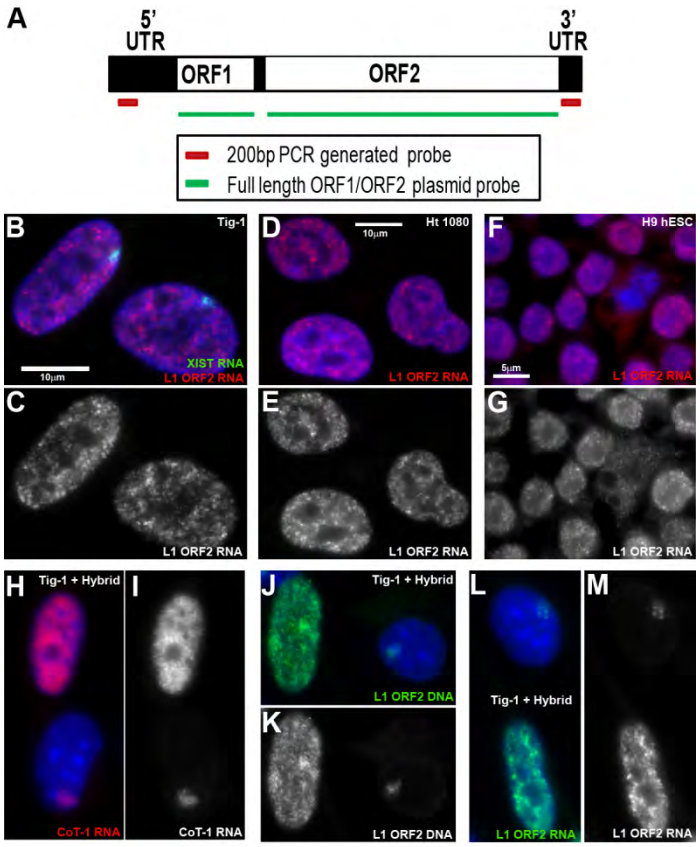

Figure 4.10 L1 ORF2 RNA is expressed in different cell types and is not aberrantly expressed in hybrid cells.

All blue signals are DAPI DNA. (A) Full-length LINE elements are 6 kb long and have two open reading frames (ORF1 and ORF2), encoding a reverse transcriptase and an endonuclease. Four different probes were used to detect L1 RNA or DNA by FISH. Two plasmids containing either the entire ORF1 or ORF2 regions of L1, or two PCR-generated probes to the 3′ or 5′ UTRs (see Table 4.1 for primer sequences). (B–G) L1 ORF2 RNA is expressed as a large component of CoT-1 RNA in all cell types examined, including (B-C) primary female fibroblasts, (D-E) cancer cell lines and (F-G) human ES cells. Exposures are the same in all images, and XIST RNA is included in the fibroblasts for comparison. (H and I) The intensity of CoT-1 RNA signal is similar between normal Tig-1 fibroblasts and mouse/human hybrid cells, suggesting the hybrid cells do not aberrantly overexpress this RNA (cells were hybridized together on the same slide). (J and K) As expected, the intensity of L1 DNA signal between the single human chromosome in the hybrid cells is similar to the average human signal in fibroblasts, although some regions of the human nucleus are brighter (cells were hybridized together on the same slide). (L and M) The L1 ORF2 RNA from the human chromosome in the hybrid cells is not aberrantly overexpressed and may even be slightly lower than the average level in the human nucleus (cells were hybridized together on the same slide).

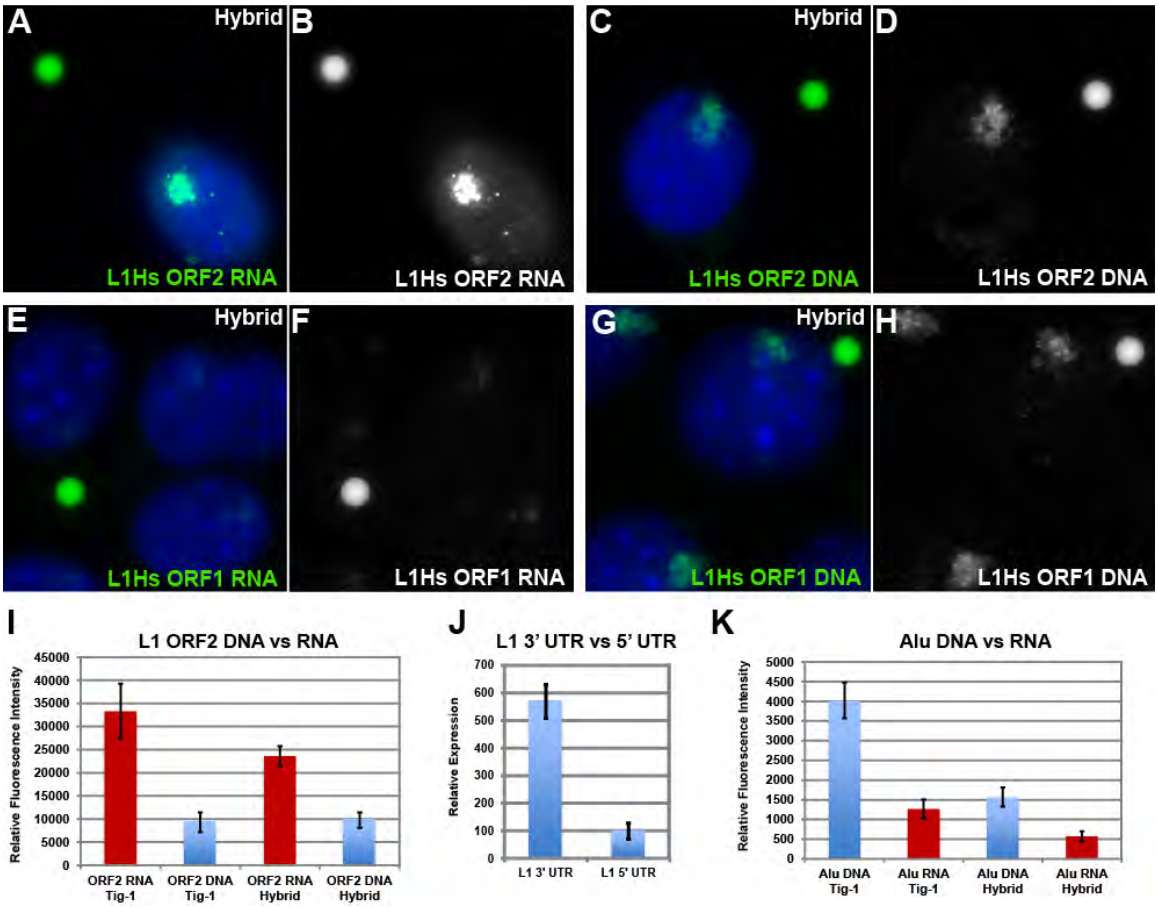

Figure 4.11 RNA from the 3′ end of L1 is a large component of the CoT-1 RNA signal.

All blue signals are DAPI DNA. (A–H) L1 ORF2 RNA signal (A and B) is ~4× brighter than the double-stranded L1 ORF2 DNA signal (C and D), whereas the L1 ORF1 RNA signal (E and F) is ~0.2× the DNA signal (G and H). (I) L1 RNA/DNA signals were quantified using digital fluorimetry (n = 25). (J) qPCR confirms that L1 ORF2 RNA is ~5–7× more expressed than ORF1 RNA. (K) Alu RNA is ~3–4× dimmer than its DNA signal quantified by digital fluorimetry (n = 25). For (I) –(K), all error bars represent 95% confidence interval.

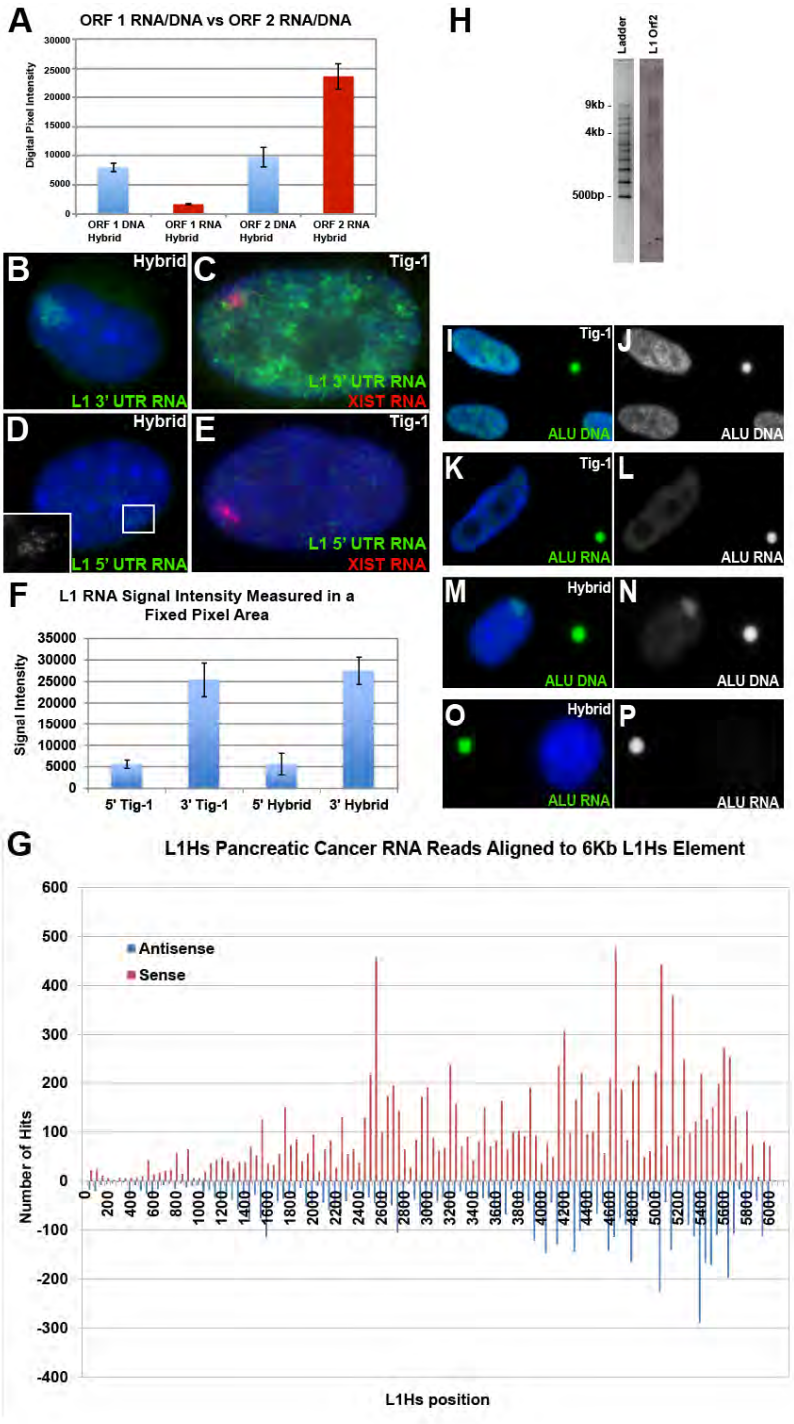

Figure 4.12 L1 RNA arises predominantly from the 3′ end of the element, in the sense direction, and it is heterogeneous in size, whereas Alu RNA is expressed at a low level.

Figure 4.12 L1 RNA arises predominantly from the 3′ end of the element, in the sense direction, and it is heterogeneous in size, whereas Alu RNA is expressed at a low level.

All blue signals are DAPI DNA. (A) Quantification of signal intensity for L1 ORF1 and ORF2 DNA and RNA signal in hybrid cells reveals that very little ORF1 RNA is expressed from its corresponding DNA signal, confirming previous reports that full-length L1 elements are silenced in normal cells. Error bars represent 95% confidence interval. (B–E) Because L1 sequences are divergent, it was difficult to design an oligo to broadly detect L1 sequences as was done for Alu. Instead we used 200 bp PCR-generated probes for the 3′ and 5′ UTR of L1 to confirm more 3′ expression (B-C) compared to 5′ (D-E) in hybrid and fibroblast cells. All images are at the same exposure, except for the insert in D (RNA signal in black & white), which has been increased for visibility. XIST RNA signal is shown for comparison in (C-E). (F) The 3′ and 5′ UTR RNA levels were measured using digital morphometrics and normalized to the bead. Error bars represent 95% confidence interval. (G) Most L1Hs deep sequencing reads from pancreatic cancer (Ting et al., 2011) map to the 3′ end of the element and are predominantly in the sense direction. (H) Northern blot analysis shows L1 ORF2 RNA extracted by Trizol is low in abundance but heterogeneous in size. (I–P) Because Alu shows more conservation than L1, oligo probes succesfully detected Alu DNA sequences comparable to full length (~200nt) Alu plasmid probes. Here, using an oligo probe, Alu DNA (I-J) signal is much brighter than Alu RNA (K-L) in Tig-1 fibroblasts. Similarly, Alu DNA (M-N) signal is brighter than Alu RNA (O-P) in hybrid cells. All images are at the same exposure, normalized to the bead. Alu RNA is visible by eye through the microscope in hybrid cells, but it is too low to be seen in the image (P) when normalized to the bead. All beads are 2.5 μm.

Table 4.3 L1 versus Alu RNA levels normalized to their DNA signal in both cell lines.

	RNA	DNA
L1 Fibroblast	4	1
L1 Hybrid	4.5	1
Alu Fibroblast	.25	1
Alu Hybrid	<0.1	1

Since DNA is double-stranded and the RNA is single stranded, the ratio of signals is ~6-8 times more RNA than DNA per L1 ORF2 sequence element

indicate that 5′ truncated L1s are abundantly and stably associated with chromosome territories.

We also examined L1 expression in several human RNA deep sequencing data sets and found more L1 reads mapping to the 3′ than to the 5′ end, mostly in the sense direction (Figure 4.12G). However, L1 read frequency overall in these extraction-based RNA samples was lower than expected from our FISH analyses. For example, in normal pancreatic tissue, using a database of L1 consensus sequences from Repbase, we find only 100 reads per million mapping to L1. We also attempted to examine these repeat RNAs by dot blot or Northern blot (Figure 4.12H) using standard Trizol RNA extraction and found only a low level of heterogeneously sized L1 transcripts. We suggest that the underrepresentation of these transcripts by most extraction-based methods may be due to the structural nature of this chromatin-associated RNA (discussed further below).

The copious nature of L1 nuclear RNA was reinforced by comparison to a substantially lower RNA:DNA ratio for the major human SINE, Alu, using oligo as well as larger probes (Table 4.1), measured in parallel samples with L1. Although Alu RNA was across the nucleoplasm, the signal was consistently several-fold less than its DNA signal in fibroblasts and was very weak in hybrid cells (Figures 4.11K and 4.12I–P). Assuming roughly similar accessibility of sequences *in situ*, this marked difference in RNA:DNA ratios indicates that representation of L1 sequences in nuclear RNA is substantially higher (Table 4.3). Bioinformatic analysis of L1 and Alu on human Chr 4 shows similar copy number (151 full-length and 58,639 truncated L1s, and 54,993 Alu). Because 68% of Alus and 47% of L1s are in introns, genic transcription would not

account for this difference. This further suggests that the nuclear repeat RNA signal is not mere transcriptional noise, and it raises the possibility of differential regulation/roles of repeat families.

CoT-1 RNA remains structurally associated with the interphase territory but can be released by perturbation of the nuclear scaffold

Strict localization of CoT-1 RNA to the territory in *cis* (even after transcription arrest) suggests that it may be actively tethered to nuclear/chromosome structure, as opposed to passive accumulation of a nontransported RNA (or rapid degradation of released RNA). We found that CoT-1 RNA localization resists mechanical disruption of nuclei by vigorous cytospinning (Figure 4.13A), and even after biochemical fractionation (removing > 90% protein and DNA), the bright, precisely localized CoT-1 RNA territory remains undisturbed (Figures 4.13B–E and 4.14A–D). This suggests that the RNA is bound to or embedded in a nonchromatin nuclear substructure (described above and reviewed in (Nickerson, 2001)), similar to XIST RNA (Figures 4.14E–H).

Because CoT-1 RNA fractionates with the nonchromatin scaffold, we used a dominant-negative mutant of Scaffold Attachment Factor-A, SAF-A/hnRNPU (C280) (Figure 4.14I) (Fackelmayer et al., 1994), to generally disrupt the scaffold. Strikingly, transfection with C280 readily released CoT-1 RNA (in ~80% of hybrid cells), which is then clearly seen dispersed through the nucleoplasm (Figures 4.13H-I, and 4.14J–N). Full-length SAF-A did not have this effect (88% remain localized) (Figures 4.13F-G, and 4.14L), and there was no indication of apoptosis (data not shown). Because this mutant will impact the complex nuclear scaffold, we do not yet know whether the

relationship of SAF-A and repeat RNAs is direct. However, the fact that CoT-1 RNA can be released from the territory shows that its normal localization is not via passive accumulation but involves a reversible mechanism for tethering to the parent chromosome.

Early studies suggested that chromatin-associated RNAs may be more resistant to extraction (Bynum and Volkin, 1980), and our results indicate that repeat RNAs are nuclear embedded and likely underrepresented by traditional methods. Supporting this, we find that the RNA:DNA ratio for L1 when hybridized *in situ* is 4:1 but is less than 1:1,500 following extraction and qRT-PCR (Figure 4.14O and Table 4.3). This may also explain why L1 ORF2 RNA is poorly represented in data sets compared to RNA FISH (Figures 4.14P-Q).

Evidence for expression of specific repeat families

I extended this analysis to specific repeat families to test which potential components of the CoT-1 RNA fraction were mislocalized by dominant negative SAF-A expression. I used probes specific for L1 ORF1, L1 ORF2, Alu, and mouse B1 and B2 RNA and found that with the exception of mouse B1 RNA, these repetitive and RNAs mislocalized from the Chromosome 4 territory, and they also appeared to be substantially upregulated in a fraction of cells (Figure 4.15). I saw similar results in human cells (not shown). This suggests that specific repeat families within CoT-1 may be differentially regulated during stress.

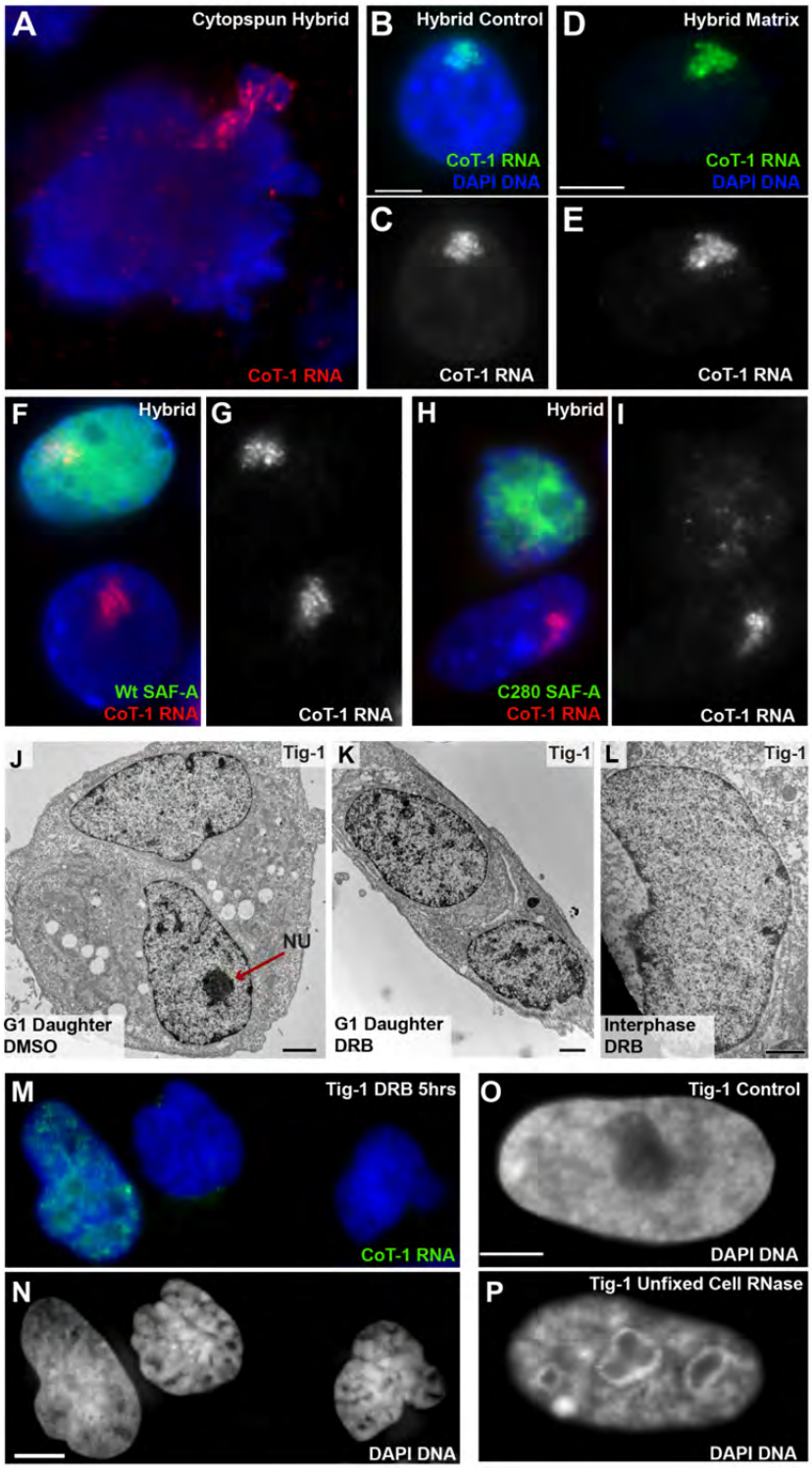

Figure 4.13 CoT-1 RNA associates with the nuclear scaffold, and Cot-1 RNA loss is coincident with chromatin collapse.

Figure 4.13 CoT-1 RNA associates with the nuclear scaffold, and Cot-1 RNA loss is coincident with chromatin collapse.

All blue signals are DAPI DNA. (A) CoT-1 RNA remains bound to human Chr 4 in ruptured nuclei. (B–E) (B and C) CoT-1 RNA is localized in control cells and (D and E) remains localized after removal of ~95% of DNA and histones (Matrix digestion) in 100% of interphase cells. Exposure times are equal for both images. (F–H) (F and G) CoT-1 RNA remains localized in 88% of cells containing wild-type SAF-A-GFP, whereas 80% of cells transfected with C280-GFP mutant (H and I) release CoT-1 RNA. Images include neighboring untransfected cells. (J–L) (J and K) Chromatin collapse seen by EM in transcriptionally inhibited G1d cells compared to similarly inhibited interphase cells (L). (M) Transcriptionally inhibited interphase cells retain stable CoT-1 RNA, whereas postmitotic cells do not. (N) Chromatin does not collapse in interphase cells, but more compacted postmitotic cells were seen in treated (86%) versus control (12%) samples. (O and P) Interphase chromatin collapses when unfixed cells are treated with RNase. Scale bars are 5 μm except for EM images, which are 2 μm.

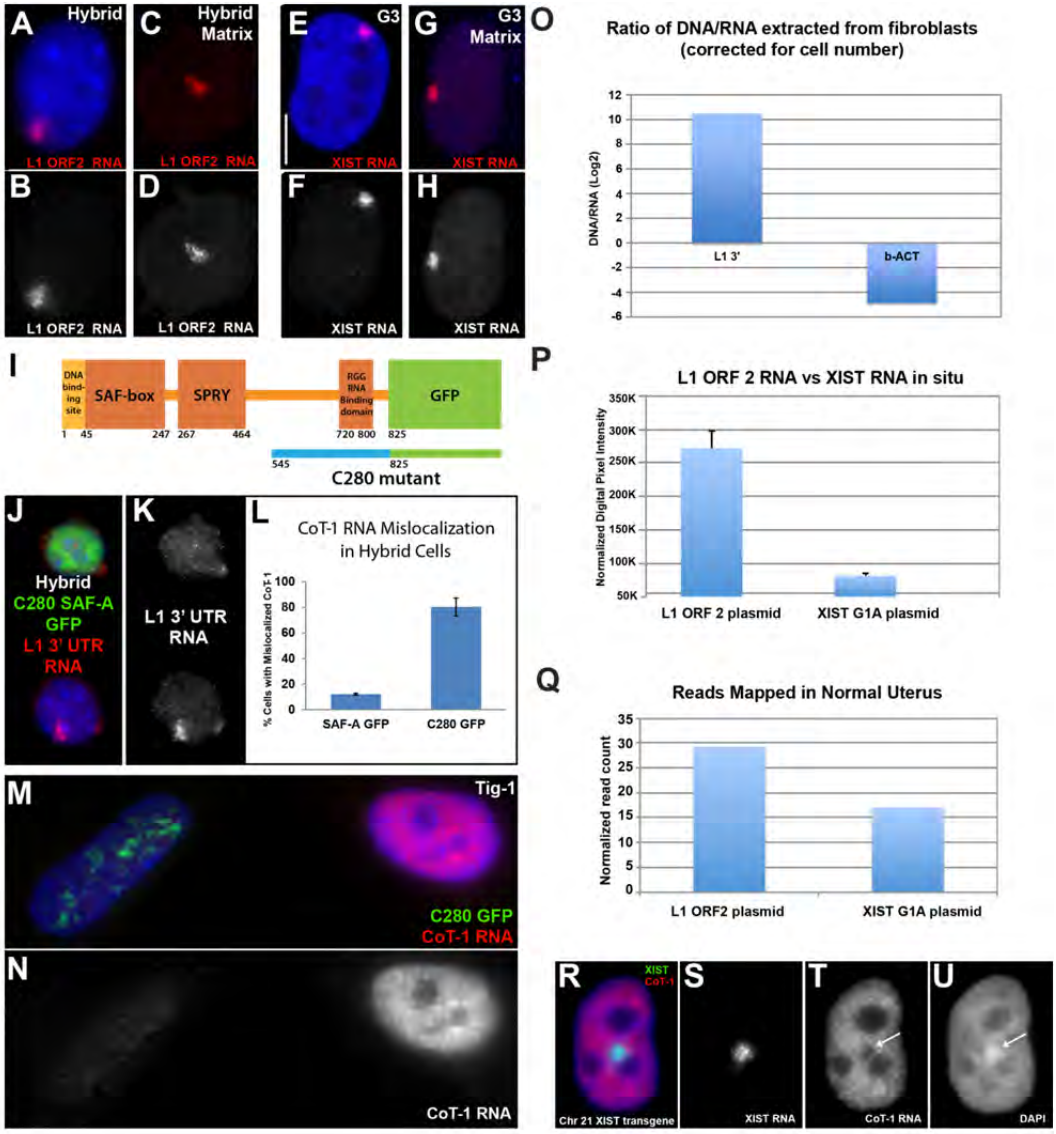

Figure 4.14 XIST and CoT-1 RNA are both retained on the nuclear matrix and released with the SAF-A C280 mutant, and the RNA may resist standard RNA extraction.

Figure 4.14 XIST and CoT-1 RNA are both retained on the nuclear matrix and released with the SAF-A C280 mutant, and the RNA may resist standard RNA extraction.

(A–H) All blue signals are DAPI DNA. (A–D) Like CoT-1 RNA, L1 ORF2 RNA is localized across the interphase chromosome territory of its parent chromosome in control cells (A-B), and remains attached following removal of ~90%–95% of DNA and chromatin proteins (Matrix digestion) using ammonium sulfate and DNase (C-D) (Clemson et al., 1996). Scale bar, 5 μm. (E-H) This is also true for XIST RNA, which remains equally localized in all control (E-F) and matrix-digested (G-H) hybrid interphase nuclei. All treated and control images are taken at the same exposure showing little to no diminishment of the RNA signal after chromatin digestion. Red channels separated below multicolor images. (I) Diagram of full length SAF-A/hnRNP U and the C280 SAF-A deletion mutant. (J and K) RNA from the 3′UTR of L1 is released from Chr 4 in hybrid cells transfected with the C280 mutant, while untransfected cells in the same field are not affected. (L) Number of cells showing mislocalized CoT-1 RNA was scored in cells transfected with C280 or full length SAF-A (n = 100, in triplicate). (M and N) In Tig-1 cells transfected with C280, CoT-1 RNA is released and is degraded over time, while untransfected cells in the same field still contain CoT-1 RNA. (O) L1 ORF2 DNA:RNA ratio and β-actin DNA:RNA ratio by qPCR from fibroblasts (corrected for cell input number, see Materials and Methods). Shown in log scale. (P and Q) Comparison of L1 and XIST RNA levels in normal cells measured either by digital fluorimetry of RNA FISH in fibroblast cell line (P) or by RNA deep-sequencing reads in normal uterine tissue (Q). Error bars represent 95% confidence interval, and data were normalized to account for size of the probe. Deep-seq reads from normal uterus (Ting et al., 2011) were mapped to the same sequence used as probes for RNA FISH (L1Hs ORF2 and XIST G1A) by local BLAT. (R–U) A chromosome 21 transgene (Jiang et al., 2013) demonstrates that XIST RNA can act on, and silence, Cot-1 repeat RNA on an autosome (see Cot-1 RNA “hole” arrow), which is coincident with Barr body formation (arrow). Separated channels are shown in white (S-U).

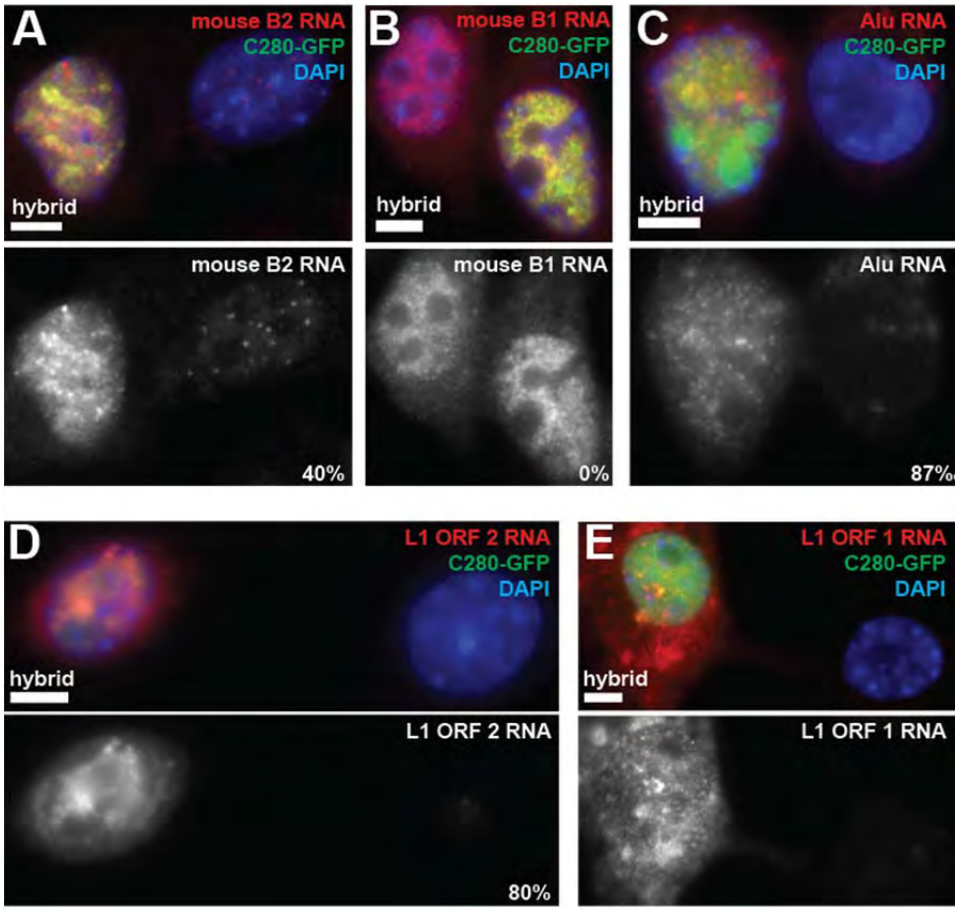

Figure 4.15 Dysregulation of specific repeat families in hybrid cells expressing dominant negative SAF-A.

(A-E) mouse B2 (A), human Alu (C), and human LINE-1 (D-E) transcripts are mislocalized and/or upregulated in hybrid cells expressing dominant negative SAF-A C280-GFP, but mouse B1 RNA (B) is not. Percentages indicate fraction of GFP-positive cells with mislocalized phenotype. Scale bars, 5µm.

Loss of CoT-1 RNA is associated with chromatin condensation

In contrast to XIST RNA's association with Xi heterochromatin, CoT-1 RNA distributes across euchromatic regions but is excluded from heterochromatin, as is apparent for mouse pericentromeric heterochromatin (Figure 4.1B), and the peripheral heterochromatic compartment (Tam et al., 2002) (Figures 4.1D-E). We previously showed that exclusion of CoT-1 RNA provides a convenient assay for chromosome silencing (Hall et al., 2002a) and that this CoT-1 RNA "hole" overlapped the condensed Barr body, which we showed was a repeat-rich silent core of the chromosome (Hall et al., 2002a; Clemson et al., 2006) (Figures 1F -G). Although CoT-1 RNA has been assumed to reflect nascent pre-mRNAs, results here indicate that repeat RNAs are more widespread and associated with active chromosomes. Thus, chromosome inactivation by XIST involves widespread silencing of CoT-1 RNA and overall chromosome condensation.

This suggests the possibility that CoT-1 RNA loss may be associated with chromatin condensation in heterochromatin. To further investigate this, we examined post-mitotic G1d cells, which lack CoT-1 RNA following transcriptional inhibition, compared to two types of controls: inhibited interphase cells (which retain the stable CoT-1 RNAs, but not short-lived or pre-mRNAs; Figures 4.7K-L) and untreated normal G1d cells (which resynthesize CoT-1 RNA). Using mitotic shake-off to enrich for G1d cells, we noted more condensed DNA regions within treated nuclei and examined them by both fluorescence and electron microscopy. Electron microscopy confirmed a consistent increase in dense chromatin clumps in most DRB-treated G1d cells (Figures

4.13J-K, 4.16B, D) compared to similarly treated interphase nuclei (Figure 4.13L) and untreated G1d controls (Figures 4.16A, C); additionally, we noted two types of unusual condensed structures not seen in controls (Figures 4.16E–F). We quantified these same cultures by light microscopy, where 77% of DRB-treated G1d cells exhibited larger, more numerous condensed clumps of chromatin, compared to only 16% of controls. Similar results were seen in inhibited asynchronous cultures, which showed 2-fold more paired G1 daughter cells with condensed clumps of chromatin (Figures 4.13M-N, 4.16G-H). The chromatin of interphase cells remained more uniform in topography, with fewer regions of high-density DNA compared to inhibited G1d nuclei (Figures 4.16M-O), and this was also evidenced by less homogeneous distribution of chromatin proteins (Figures 4.16I–L). Examination of the Chr 4 DNA territory in inhibited G1d hybrid cells showed the territory partially decondensed but retained a large mass of condensed chromatin within it (Figures 4.16P-Q).

The lack of chromatin collapse in interphase cells (which retain CoT-1 RNA) (Figures 4.13L–N) indicates that condensation is not due to transcriptional inhibition per se or to the absence of nascent or short-lived nuclear RNAs. Alternatively, it could be due to lack of resynthesis of stable nuclear RNAs or proteins necessary to open chromatin. To address this, we used RNase in unfixed cells to determine whether removal of CoT-1 and other nuclear RNA could cause interphase chromatin to condense (Figures 4.16R-S). Chromatin collapse was immediately visible (within 20 min); ~60% of cells were markedly affected as seen in Figures 4.13O-P. This is consistent with earlier observations of RNase (Nickerson et al., 1989), but collective findings here indicate that abundant,

stable, repeat-rich RNAs are widely associated with euchromatic chromosome territories and likely promote more open chromatin packaging.

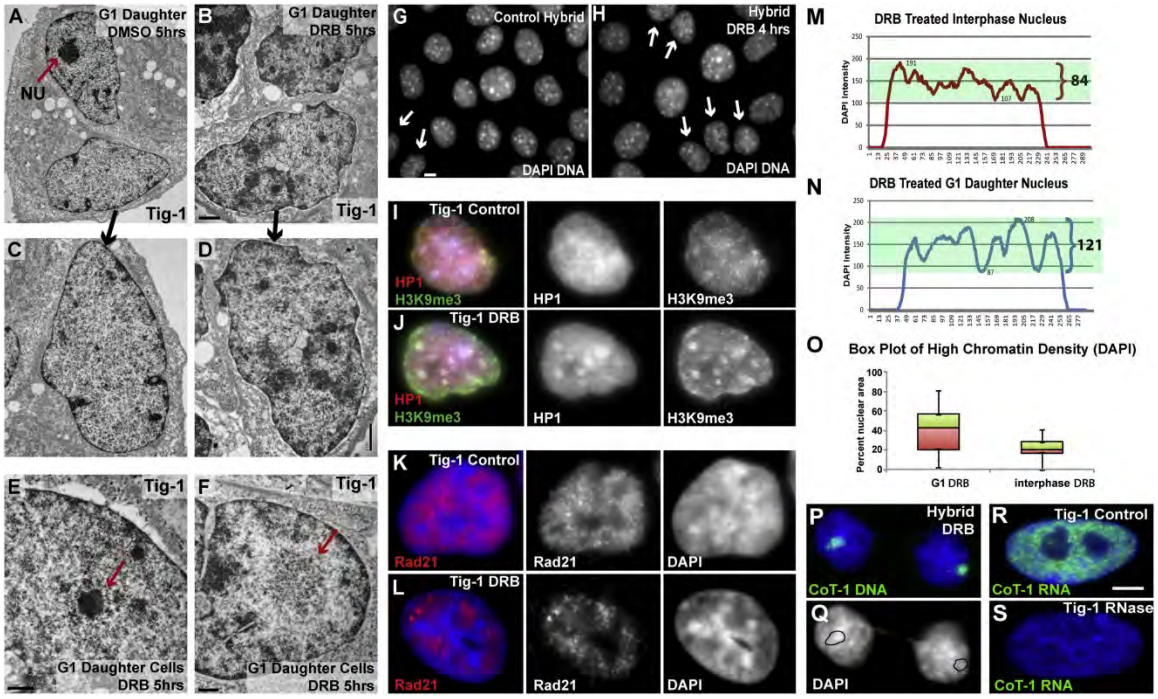

Figure 4.16 RNase digestion and transcriptional inhibition causes changes in chromatin structure.

Figure 4.16 RNase digestion and transcriptional inhibition causes changes in chromatin structure.

(A and B) A) Control DMSO-treated G1d cells exhibit a normal nuclear phenotype B) DRB-treated G1 daughter cells show condensed chromatin clumps by EM. Scale bar, 1 μm. (C and D) C) Enlarged normal G1d cell to show chromatin distribution D) Enlarged DRB-treated G1d cell to show aberrant chromatin clumps. (E) Example of aberrant darkly staining structure with spindle-like fibers radiating outward (arrow), which are found in all DRB-treated G1d cells (2-8 per nucleus). Scale bar, 0.5 μm. (F) High resolution EM analysis shows large clumps of dense material in DRB-treated G1d cells (arrow), present in ~70% of G1d nuclei. Scale bar, 0.5 μm. (G and H) Paired nuclei with condensed chromatin clumps (arrows) were about twice as common in transcriptionally inhibited cultures (H) compared to controls (G). Control = 11%, DRB = 23% & ActD = 25%. (I–L) Nucleoplasmic chromatin proteins showed a more "lumpy" distribution in DRB-treated G1d nuclei than in control G1d nuclei. Some proteins delineated the condensed chromatin (HP1 and H3K9me3), while others were excluded from the chromatin clumps (Rad21). (M and N) Linescan measurements through the nuclear DNA of two representative cells show a broader range of DAPI DNA pixel intensity in a transcriptionally inhibited G1d nucleus compared to a similarly treated interphase nucleus. (O) The DNA pixel intensity was measured in ten 3D images of transcriptionally inhibited G1d and neighboring interphase cells and the highest 3/5ths (pixel intensity > 129) of the DAPI signal was compared. Box and whisper plot of high intensity DAPI staining calculated as percent area of nuclei for DRB-treated G1d and interphase cells (n = 10). Horizontal lines on boxes indicate lower, median, and upper quartile values. (P and Q) Human Chr 4 is delineated by CoT-1 DNA hybridization (P) and the regions of the chromosome territories (black circles) are transferred to the DAPI DNA channel (Q) to determine if individual chromatin clumps comprise a single chromosome. (R and S) All CoT-1 RNA is removed with RNase A treatment in unfixed cells. All blue signals are DAPI DNA. Scale bars, 5 μm.

Discussion

These findings impact two “black boxes” of genome biology: the unexplained prevalence of repetitive elements interspersed throughout higher-order genomes and the poorly understood diversity of ncRNAs. *In situ* visualization of RNA using CoT-1 DNA as a probe reveals that abundant RNA is broadly distributed with chromatin and remains stable and localized to the parent chromosome long after transcriptional arrest. Recent evidence indicates that specific ncRNAs may bind a subset of chromosomal loci. Findings here support a distinct but potentially related concept that RNA generally and broadly associates with nuclear chromosome territories, where it remains stable and localized independent of ongoing transcription. This suggests that RNA is a fundamental component of chromosome biology, rather than only a product of it. Like XIST RNA, CoT-1 RNA detaches from mitotic chromosomes, yet it tightly adheres to the interphase chromosome structure in *cis*, even after nuclear fractionation. Hence, CoT-1 RNAs and XIST RNA can be considered “chromosomal RNAs” that likely bridge chromatin with insoluble nonchromatin structural elements. Although speculative, repetitive sequences would be well suited to form an intermolecular lattice or structural element because lncRNAs can have an “architectural” role, and intermolecular RNA duplexes were long ago noted in nuclear RNA (Fedoroff et al., 1977). Such RNAs may also play a role in higher-order chromatin packaging linked to regulation. In contrast to XIST RNA, which triggers chromosome condensation, CoT-1 transcripts specifically distribute across euchromatin, where they may promote an open chromatin state.

These findings point to the import of interspersed repeat sequences in chromatin-associated RNA, raising many next questions about this poorly studied fraction of the genome. Britten and Davidson long ago hypothesized that repeats function in genome regulation, stating “a concept that is repugnant to us is that about half of the DNA of higher organisms is trivial or permanently inert…” (Britten and Kohne, 1968; Britten and Davidson, 1971). Our motivation to study repeats stemmed from their poorly explained abundance and wide distribution in conserved patterns, suggesting to us a potential role in coordinate regulation of the genome. We show that repeat sequences are not only prevalent in chromosome-associated RNA but that they are the predominant component of hnRNA stably associated with chromosomes. Retention on the parent chromosome and longer nuclear half-life likely contribute to the greater level of repeat RNAs detected on the chromosome versus the collective nonrepetitive sequences in the whole chromosome library. Hence, the level of an RNA on the chromosome can be disproportionate to the transcription rate and cytoplasmic abundance. The *in situ* analyses here revealed both unexpected nuclear abundance and unusual cellular properties of CoT-1 RNAs distinct from those expected for nascent pre-mRNAs or “genic” transcripts, including localization to the chromosome territory, which persisted for more than a day after transcription arrest. To minimize any chance that an inhibitor induced stabilization, we tested inhibitors with three distinct mechanisms with similar results. Interestingly, it has been reported that the many thousands of uncharacterized lncRNAs are rich in (or related to) interspersed repeat elements (Hadjiargyrou and Delihas, 2013; Kapusta et al., 2013), and many show prolonged stability (Clark et al., 2012).

Full characterization of the heterogeneous class of CoT-1 repeat RNAs will require improved nuclear RNA extraction methods and RNA sequencing; however, we were able to identify L1 RNA as a major component. The prevalence of 5′ truncated L1 is consistent with evidence that full-length transposable L1s are silenced in normal cells (Goodier and Kazazian, 2008) and is distinct from the reported transient expression of a few full-length L1s during mouse X chromosome silencing (Chow et al., 2010). A recent study found that highly abundant, truncated L1 elements contain internal and 3′ promoters, are expressed at low levels in a cell-type-specific manner, and some may be resistant to extraction (Faulkner and Carninci, 2009). Our *in situ* analysis with large probes and preliminary Northern analyses (also limited by extraction) showed that most L1 transcripts are large (>4 kb); thus, these could include lncRNAs or even intron-derived sequences. Importantly, the vast majority of introns contain interspersed repeats, and most introns still have no known function. Although excised introns are generally thought to rapidly degrade, a very recent study in *Xenopus* oocytes found that highly stable RNAs representing excised introns accumulate in the germinal vesicle (Gardner et al., 2012). Although the *Xenopus* oocyte is a distinct biological system, it is an interesting possibility that repeat-rich sequences in excised introns could accumulate on chromosomes.

To minimize the technical obstacles to studying repeats, we needed to develop new approaches, such as the use of RNA:DNA ratios (and fluorescent beads as intensity standards) to assess the relative expression of repeat families and control for differences between probes. It has been reported that SINE RNAs (mouse B2 and human Alu) are

upregulated by RNAP III upon heat shock and then bind RNAP II to broadly repress transcription (Espinoza et al., 2004; Mariner et al., 2008). Studies here used normal unstressed cells; however, under prolonged transcriptional inhibition (with α-amanitin), CoT-1 RNA signal intensity increased consistently. Thus, we are investigating whether some component(s) of CoT-1 RNA are upregulated upon inhibition of RNAP II. Interestingly, HERV-H retroviruses, a component of CoT-1 DNA, have been reported to be highly expressed specifically in pluripotent stem cells (Santoni et al., 2012). Thus, we believe the collective evidence supports that studies of genome biology should include the "repeat genome" in which distinct repeat families likely will show differential regulation and different functions in a cell-type-specific manner. This is also supported by our results showing differential regulation of repeat families in response to disruption of the nuclear scaffold by dominant negative SAF-A expression.

Our demonstration that the CoT-1 RNA "territories" resist mechanical (or transcriptional) disruption but can be released to disperse indicates that their accumulation is not passive but involves a reversible mechanism(s). That the release was triggered by perturbation of a nuclear scaffold factor further supports the biochemical fractionation results, which indicate that repeat RNAs are tethered with a nonchromatin scaffold. Because the latter is likely a complex structure (not simple polymers), we do not yet have a full understanding of how the dominant negative C280 mutant disrupts CoT-1 RNA localization. Preliminary evidence indicates that the relationship may not be direct and that multiple factors may anchor the RNA; thus, the role of SAF-A with RNA and chromatin will be examined in Chapter V. Findings here identify *cis*-localized CoT-1

RNA as a major component of the undefined RNAs previously suggested to be part of an insoluble nonchromatin structure (e.g., (Fey et al., 1986)). The structural nature of repetitive RNAs may well make them more difficult to extract by methods designed for the cytoplasm. We are working to identify improved extraction methods for further analyses; however, molecular hybridization *in situ* avoided this issue and was essential to study the relationship of these RNAs to the nuclear chromosomes.

In sum, the repetitive "junk" genome is worthy of study, not only as DNA, but also as a major part of the "dark matter" transcriptome. As only 5% of promoters are with canonical genes (Venters and Pugh, 2013), we are far from understanding genome expression and regulation. This work contributes to what may emerge as a next revolution in genome science: the relationship of chromosomal RNAs and repetitive elements to each other and to genome packaging, regulation, and evolution.

Chapter V : CoT-1 RNA is integral to the structure of euchromatic interphase chromosome territories

Preface

This work was supported by NIH grant GM107604 to J.B.L. I wish to thank Frank Fackelmayer for the gift of SAF-A GFP constructs and helpful discussions, and Rick Adshead from the UMMS MAPs department for training and time on the SIM microscope. A modified version of this chapter will be submitted for publication.

Introduction

In Chapter IV, I showed that abundant non-coding CoT-1 repeat RNA localizes in *cis* to euchromatin, and the RNA territory persists after transcriptional inhibition, reminiscent of XIST RNA painting the heterochromatic Xi chromosome (Clemson et al., 1996; Hall et al., 2014). Evidence suggests this CoT-1 chromosomal RNA may localize to chromatin via proteins of the nuclear scaffold, defined as an insoluble fibrillogranular structure that resists extraction of histones and genomic DNA (Nickerson, 2001; Hall et al., 2014). While the initial findings were hotly debated, some early studies suggested that most nuclear RNA (hnRNA) is in an insoluble state bound to scaffold proteins (e.g., (Fey et al., 1986; Nickerson et al., 1989; He et al., 1990)). Experiments here investigated whether the class of repeat-rich CoT-1 RNAs are bound to and potentially integral structural components of the nuclear chromosome territory.

I used SAF-A knockdown and a series of SAF-A deletion mutants in human cells to investigate whether SAF-A has a role in localizing CoT-1 RNA to euchromatin, and to address what domains of the protein may be required. In doing so, I manipulate CoT-1 RNA's association with chromatin to examine the hypothesis that this RNA supports the more open, decondensed packaging of euchromatin. I ask whether these chromosomal RNAs are not only bound to the scaffold, but are integral to scaffold structure, and thus required to maintain SAF-A or specific nuclear proteins. I found that SAF-A is not required for CoT-1 RNA localization to chromatin, but instead, SAF-A associates with chromatin in an RNA-dependent manner. Association between CoT-1 RNA and the

nuclear matrix is required for CoT-1 RNA stability and maintenance of euchromatin structure.

Materials and Methods

Cell culture and drug treatments: Tig-1 human fibroblasts (Coriell) and GM11687 mouse human hybrid cells (ATCC) were grown under supplier-recommended conditions. Transcriptional inhibition: DRB (40–80µg/ml dissolved in DMSO) was added to culture media and left on for 4–6 hours prior to fixation. Cell synchronization: Cells were arrested in Nocodazole (100ng/mL) for 5 hours and collected by mitotic shakeoff prior to transfection. RNase: We permeabilized unfixed cells on coverslips with 0.1% triton-X in CSK buffer (4°C for 3 min) then treated with 5µL/mL DNase-free RNase (Roche #11119915001) or RNasin Plus RNase inhibitor (Promega), in CSK (10-30 min, 37°C).

Fixation: Our standard cell fixation protocols have been previously described in Chapter II.

RNAi and Plasmid transfections: For RNAi, cells (70-80% confluent) were transfected with SMARTpool siRNA and DharmaFECT 1 transfection reagent (GE Dharmacon) according to the manufacturer's instructions. Cells were fixed after 72 hours. Control siGLO siRNA (GE Dharmacon) had no impact on CoT-1 RNA localization. For plasmid transfections, 2 ug DNA was mixed with Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions, and added to 80-90% confluent cells in a 6-well plate. Cells were fixed and assayed 6, 24, 48 or 72 hours after transfection.

IF: IF and epitope retrieval was done as described in Chapter II. Anti-SAF-A antibodies were from Abcam (ab20666 and ab10297).

Fluorescence in situ hybridizations: RNA hybridization was performed under non-denaturing conditions as previously Chapters II/III. The DNA probe was made using CoT-1 DNA (Roche).

Microscopy and image analysis: Digital imaging and analysis was performed as described in Chapters II and III. I used Huygens to measure Chr 4 territory voxel volumes. Super-resolution 3D-SIM images were acquired on a DeltaVision OMX V4 (GE Healthcare) equipped with a 60x/1.42 NA PlanApo oil immersion lens (Olympus), 405, 488, 568, and 642 nm solid state lasers and sCMOS cameras (pco.edge). Image stacks of 7-9 µm with 0.125 µm thick z-sections and 15 images per optical slice (3 angles and 5 phases) were acquired using immersion oil with a refractive index of 1.518. Images were reconstructed using Wiener filter settings of 0.001 and optical transfer functions (OTFs) measured specifically for each channel with SoftWoRx 6.1.3 (GE Healthcare) to obtain super-resolution images with a two-fold increase in resolution both axially and laterally. Images from different color channels were registered using parameters generated from a gold grid registration slide (GE Healthcare) and SoftWoRx 6.5.2 (GE Healthcare). I used Image J/Fiji for signal intensity quantification and, when necessary, Fiji or Photoshop (Adobe) to enhance images for brightness and contrast.

Results

Endogenous SAF-A localizes primarily to chromatin and not to nuclear speckles

Originally classified as an hnRNP protein, SAF-A/hnRNPU was widely implicated to have a role in pre-mRNA splicing (Kiledjian and Dreyfuss, 1992). SAF-A was pulled down as part of a major splicing complex (C complex) (Jurica et al., 2002), and purified from mouse interchromatin granule clusters (SC-35 domains), which contain pre-mRNA splicing factors (Saitoh et al., 2004). However, a similar study failed to identify SAF-A as part of the spliceosome (Zhou et al., 2002). Hence, I carefully examined the nuclear distribution of SAF-A relative to chromatin and non-chromatin nuclear compartments using IF, FISH, and structured illumination microscopy (SIM). SAF-A and CoT-1 RNA both release from chromosomes in early prophase (similar to XIST RNA and most other nuclear structural proteins). Although some residual SAF-A can be seen on condensed chromatin prior to full release (Figure 5.1A), to evaluate the extent to which SAF-A associates with chromatin during interphase, I carefully scrutinized its distribution relative to DNA and SC-35, a protein marker for non-chromatin SC-35 domains (speckles or interchromatin granule clusters) in which most splicing factors and spliceosomal components concentrate (Hall et al., 2006; Spector and Lamond, 2011). Results using both widefield microscopy and SIM showed that the vast majority of SAF-A co-localizes over DAPI stained DNA, including euchromatin and heterochromatin. In contrast, SAF-A was clearly not enriched in SC-35 domains, from which it appears largely excluded (Figure 5.1B), as is DNA (Carter et al., 1991). These results do not preclude that SAF-A has dual roles in pre-mRNA metabolism and

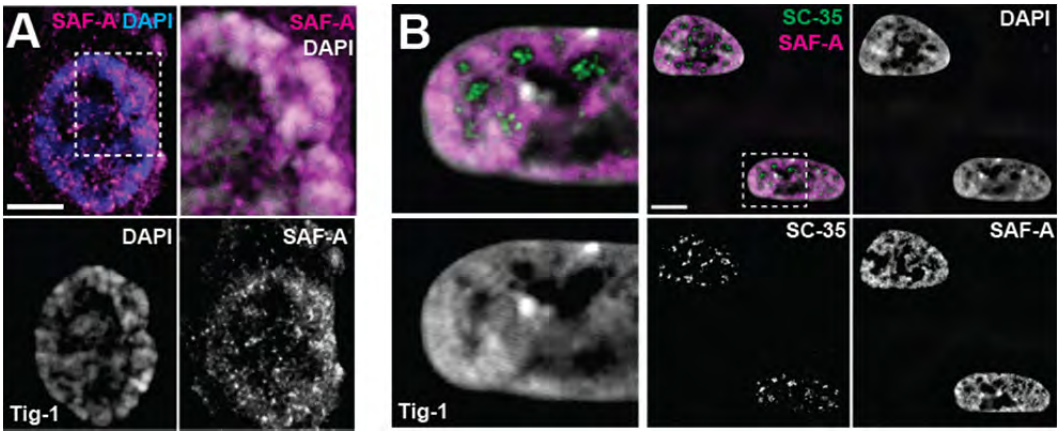

Figure 5.1 Endogenous SAF-A localizes to chromatin.

(A) SAF-A (red) localizes to prophase chromosomes (blue, DAPI). (B) SAF-A (red) localizes primarily with heterochromatin and euchromatin (gray, DAPI) but not splicing-factor-rich SC-35 domains (SC-35 protein, green, arrows).

chromatin packaging, but they indicate that the greatest proportion of SAF-A associates with chromatin.

A highly truncated SAF-A protein releases CoT-1 RNA from chromatin whereas SAF-A depletion does not

As shown in Chapter IV, repeat-rich CoT-1 RNAs show unusually strict localization to the parent chromosome territory in *cis*, remaining abundant and tightly localized after prolonged transcriptional inhibition, as well as after digestion of nuclear DNA and removal of most nuclear proteins (Hall et al., 2014). Yet, evidence indicated that CoT-1 RNAs could be rapidly dispersed upon disruption of the nuclear scaffold by expression of truncated SAF-A (C280-GFP) (Figure 5.2A-C) in mouse hybrid cells containing one human chromosome that allow examination of CoT-1 RNA localization relative to the individual chromosome territory (Figure 5.2A). This indicates that SAF-A plays some role in CoT-1 RNA's stable association with chromatin, however, it does not address whether SAF-A directly localizes CoT-1 RNA.

To investigate whether SAF-A is required to maintain CoT-1 RNA localization, I used siRNA (Dharmafect SMARTpool Thermo/Fisher) to deplete SAF-A in mouse-human hybrid cells, and affirmed by microfluorimetry in individual cells that SAF-A depletion was highly effective, reduced by 97% compared to untransfected cells. In marked contrast to expression of the truncated protein (C280-GFP), SAF-A depletion had no visible effect on either the CoT-1 RNA territory (or overall DNA distribution, as further discussed below) (Figure 5.2D).

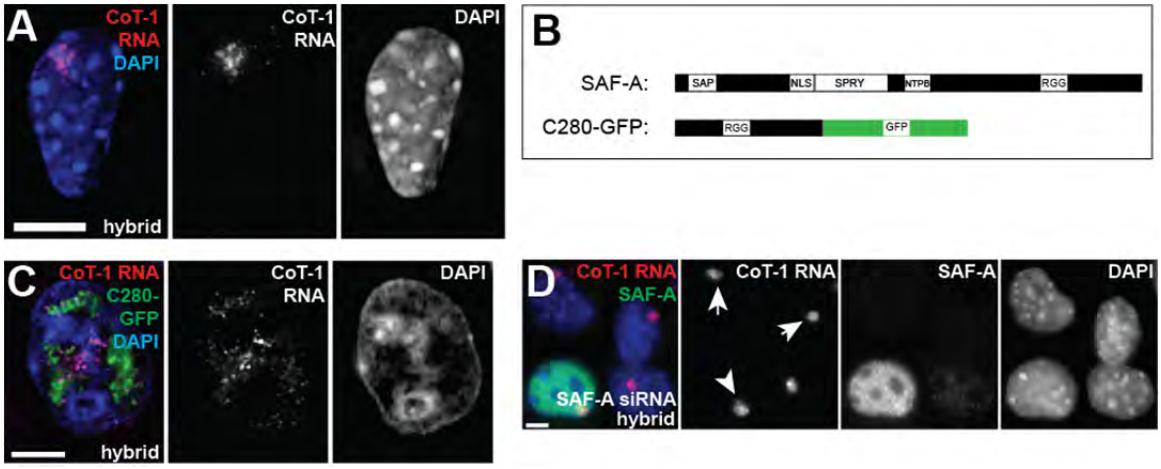

Figure 5.2 Dominant negative SAF-A mislocalizes CoT-1 RNA in human cells.

(A) CoT-1 RNA (red) localizes to discreet chromosome territory in mouse-human hybrid cells. (B) Map of SAF-A domains and truncated C280-GFP mutant. (C) Truncated SAF-A (C280-GFP) expression causes CoT-1 RNA (red) release in mouse human hybrid cells. (D) SAF-A RNAi has no effect on CoT-1 RNA localization in mouse/human hybrid cells. CoT-1 RNA (red) remains in condensed territory in cells lacking endogenous SAF-A (green). Scale bars, 5 μm.

The fact that the C280 mutant has a much greater effect than SAF-A depletion indicates it acts as dominant negative, likely through perturbation of other nuclear structural proteins that support Cot-1 RNA localization. As described in Chapter III, the C280 mutant causes general disruption to the nuclear localization patterns of NuMA, SafB1, FUS, and hnRNP C. Here I additionally show that both hnRNP C and FUS co-localized with C280-GFP aggregates in a fraction of cells (Figure 5.3A-C). This supports that hnRNP C and FUS were perturbed by and likely interact with the normal and mutant SAF-A protein. Collectively these results suggest that SAF-A is part of a complex scaffold that binds CoT-1 RNA, but that it may not anchor CoT-1 RNA directly as a single molecule bridge.

Mutations in either SAF-A's RNA or DNA binding domains mislocalizes CoT-1 RNA from chromatin

Previously I showed that SAF-A's DNA but not its RNA binding domain is required to localize XIST RNA (Chapter III). Here I examined the effect of expressing the RNA and DNA binding domain deletion mutants to more specifically characterize the effect on CoT-1 RNA localization. I used the precise SAF-A mutant with a point mutation in its DNA binding domain (G29A-GFP) (Figure 3.6E) to test whether CoT-1 RNA localization is impacted by SAF-A's ability to bind DNA. I found that expression of this mutant results in release of CoT-1 RNA from the chromosome territory in most hybrid cells expressing G29A-GFP (~60% P=.0016) (Figure 5.4A-B). I repeated this experiment in primary Tig-1 fibroblasts. Strikingly, in 82% of cells, CoT-1 RNA was no longer uniformly distributed throughout the nucleus, and it was localized with G29A-

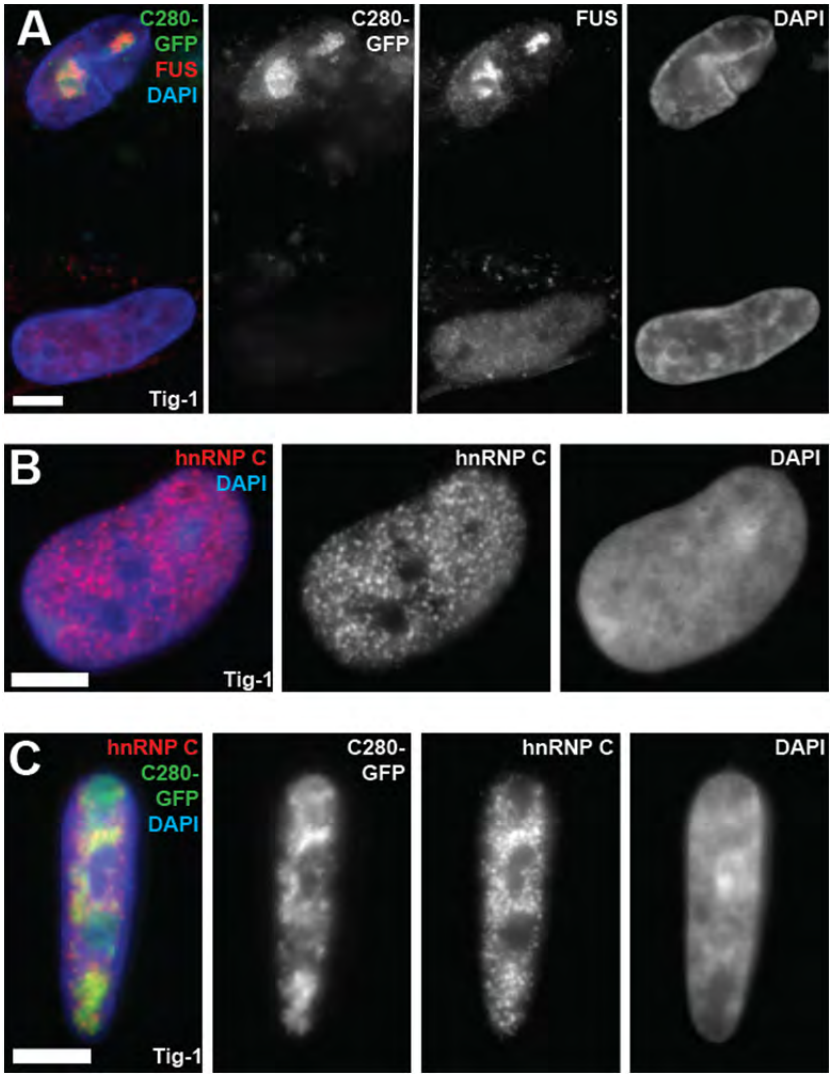

Figure 5.3 Scaffold proteins FUS and hnRNP C localize with dominant negative C280-GFP.

(A) FUS (red) localizes in C280-GFP aggregates. (B-C) hnRNP C co-localizes with C280-GFP (C) in contrast to its more uniform distribution on chromatin in control cell (B). Scale bars, 5 uM.

GFP in DAPI-depleted nuclear regions (Figure 5.4C-D). This indicates that the G29A mutant SAF-A cannot bind DNA, but suggests it can still bind CoT-1 RNA, potentially through its RGG domain, and thus this G29A mutant appears to "pull" CoT-1 RNA off of chromatin. Most importantly, these results support that SAF-A's ability to bind DNA impacts CoT-1 RNA's ability to localize to chromatin.

I next examined the ΔRGG SAF-A mutant (Figure 3.6E), which can bind DNA but not RNA. I found that CoT-1 RNA was mislocalized in most hybrid cells expressing ΔRGG-GFP (~60% P=.004) (Figure 5.5A). The CoT-1 RNA signal was largely diminished and did not localize with the ΔRGG mutant, supporting that CoT-1 RNA normally binds to SAF-A's RGG domain (Figure 5.5B). I tested whether CoT-1 RNA was similarly affected in human Tig-1 fibroblasts. Unlike in normal cells where CoT-1 RNA clearly localizes over chromatin (Figure 5.5C), CoT-1 RNA was diminished and restricted to DAPI-depleted areas in 53% of cells expressing the ΔRGG mutant for 24 hours, while the ΔRGG mutant protein remained localized with chromatin (Figure 5.5D-E). I also examined the localization of endogenous SAF-A with an antibody that does not recognize the ΔRGG mutant. In 96% of GFP-positive cells exhibiting abnormal CoT-1 RNA distribution, the majority of *endogenous* SAF-A localized with the released CoT-1 RNA in DAPI holes and not with chromatin and the ΔRGG mutant (Figure 5.5F and Figure 1B). Hence, ΔRGG-GFP appears to largely displace endogenous SAF-A on chromatin, suggesting that CoT-1 RNA requires SAF-A's RGG domain to localize to its parent chromosome. Since SAF-A has a long (~48h) half-life, it further suggests that a cell may have to go through mitosis before the ΔRGG mutant can gain access to

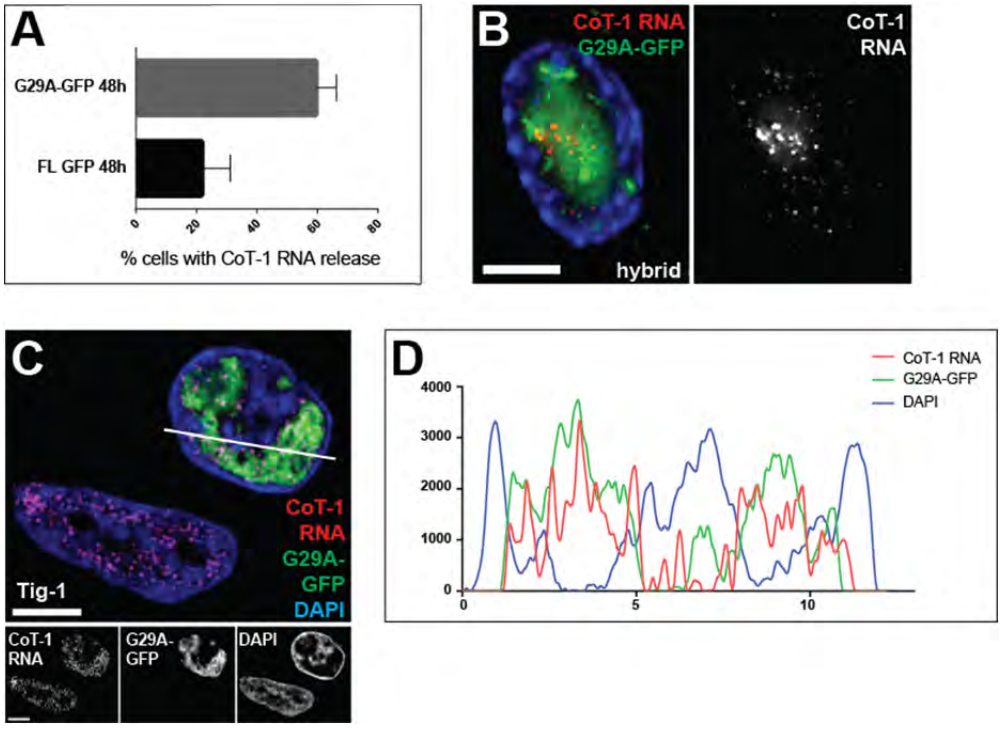

Figure 5.4 SAF-A DNA binding domain mutant causes CoT-1 RNA release and chromatin condensation.

(A) SAF-A DNA binding domain mutant (G29A) expression releases CoT-1 RNA from the human chromosome in mouse-human hybrid cells (59%, *P*=.001). (B) CoT-1 RNA (red) releases from its parent chromosome in mouse-human hybrid cell. (C-D) G29A-GFP causes chromatin condensation and CoT-1 RNA release in human Tig-1 cell. Linescan shows CoT-1 overlaps with G29A-GFP and not with DAPI. Images represent single slice of z-stack deconvolved using Softworx. Scale bars, 5 µm.

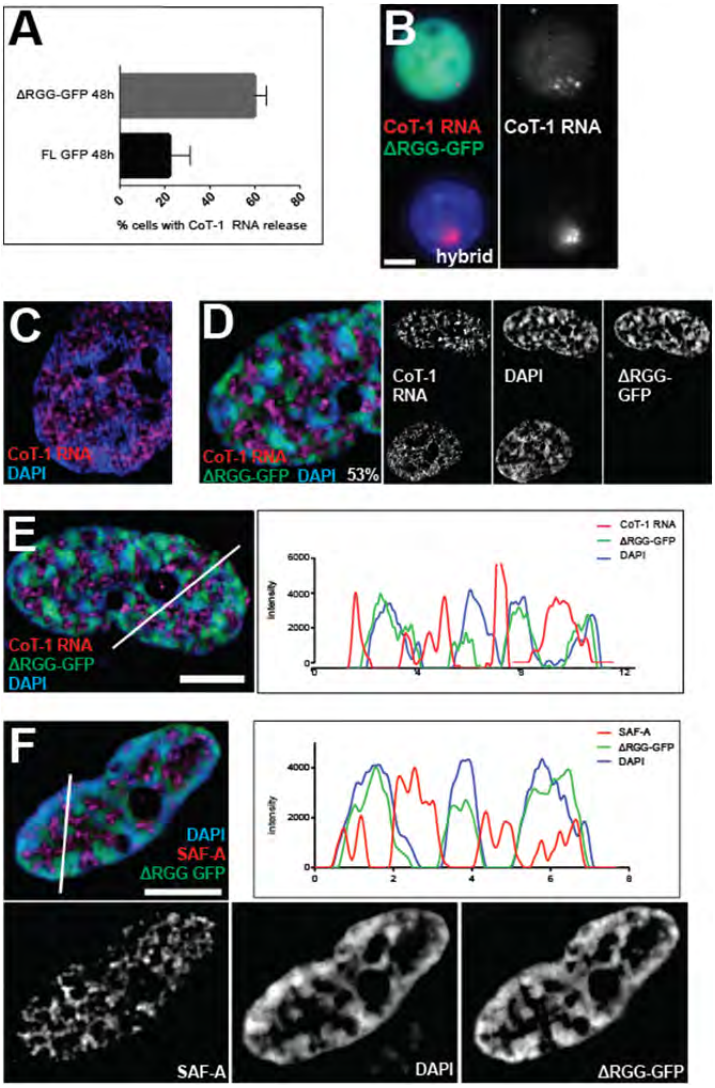

Figure 5.5 SAF-A's RNA binding domain is required for CoT-1 RNA localization.

(A-B) CoT-1 RNA releases from the human chromosome in mouse-human hybrid cells expressing the RGG domain deletion mutant (58% *P*=.004). Error bars, standard deviation of the mean (A). CoT-1 RNA (red) is largely diminished compared to GFP-negative cell (B). (C-E) CoT-1 RNA releases from chromatin in Tig-1 cell expressing ΔRGG-GFP (D) but not in normal cell (C). GFP mutant co-localizes with chromatin, while CoT-1 RNA (red) localizes in DAPI holes, see linescan (E). CoT-1 RNA signal was boosted for visibility. (F) Endogenous SAF-A (red) localizes in DAPI holes and not with chromatin and ΔRGG-GFP (see linescan). Scale bars, 5 μm.

chromatin and displace CoT-1 RNA. This may also explain the penetrance to only ~50-60% of transfected cells.

These results show a complex relationship between SAF-A and CoT-1 RNA. SAF-A appears to bind CoT-1 RNA through its RGG domain, and it appears to play a role in localizing CoT-1 RNA to DNA by a complex mechanism, but not as a direct bridge. Collective results suggest that both mutants act through a dominant negative mechanism, since SAF-A knockdown had no effect on CoT-1 localization.

Lack of CoT-1 RNA, and not lack of SAF-A, consistently causes chromatin condensation

These perturbations also facilitate investigation of Cot-1 RNA and/or SAF-A's role on the overall packaging and condensation of chromatin. In this context, expression of all three SAF-A mutants markedly perturbs normal chromatin distribution, including complete loss of DAPI-dense mouse chromocenters (Figure 5.2A, C; 5.4C-D; 5.5D-F). To further examine this relationship, I tested the effect of the C280-GFP mutant in normal human cells. I found that ~50% of C280-GFP positive human cells contained large clumps of chromatin, indicated by large peaks in staining intensities compared to more uniform staining in control cells (compare the blue lines in Figures 5.6A-B). Moreover, these DNA clumps lacked both C280-GFP protein and CoT-1 RNA (Figure 5.6A, C). This suggests that as C280-GFP expression causes CoT-1 RNA release, it also causes widespread chromatin perturbation, in particular the formation of condensed "clumps". Since C280-GFP still contains SAF-A's RNA binding RGG domain (Fackelmayer et al., 1994) (Figure 5.2B), it is plausible that the SAF-A mutant binds to

CoT-1 RNA, pulling it away from chromatin and causing chromatin to condense. This is further suggested by the overlap of Cot-1 RNA with large C280-GFP foci in 71% of cells with chromatin collapse (Figure 5.6A, C).

DAPI staining showed similar abnormal clumps of DNA in 82% of cells expressing G29A-GFP, indicated by large peaks in signal intensity (Figure 5.4C-D).To more rigorously test for chromatin collapse caused by this mutant, I measured the impact on the human chromosome 4 DNA territory in hybrid cells expressing G29A-GFP, delineated by hybridization with the human CoT-1 DNA probe (which does not hybridize to mouse repetitive DNA) (Figure 5.6 D, F). As measured in deconvolved z-stacks (Huygens, see Methods), the total volume of the human chromosomal DNA territory decreased in cells expressing the mutant SAF-A, as compared to untreated control cells (Figure 5.6E). These findings are further consistent with the possibility that loss of CoT-1 RNA from chromatin promotes chromatin collapse. However, it remains unclear whether chromatin condenses due to loss of CoT-1 RNA, SAF-A and/or potentially other scaffold proteins, or both.

Importantly, chromatin also condenses in cells expressing ΔRGG-GFP, in which the mutant remains with chromatin but CoT-1 RNA releases (Figure 5.5D-E). This supports that the impact on chromatin condensation and overall chromatin packaging noted above for the C280 and G29A mutants likely is not due to the loss of SAF-A, but to the release of RNA from chromatin. Supporting this, siRNA depletion of endogenous SAF-A does not impact chromatin structure in the majority of cells tested (with the potential exception of Neuro2a cells examined in Chapter III), thus lack of CoT-1 RNA,

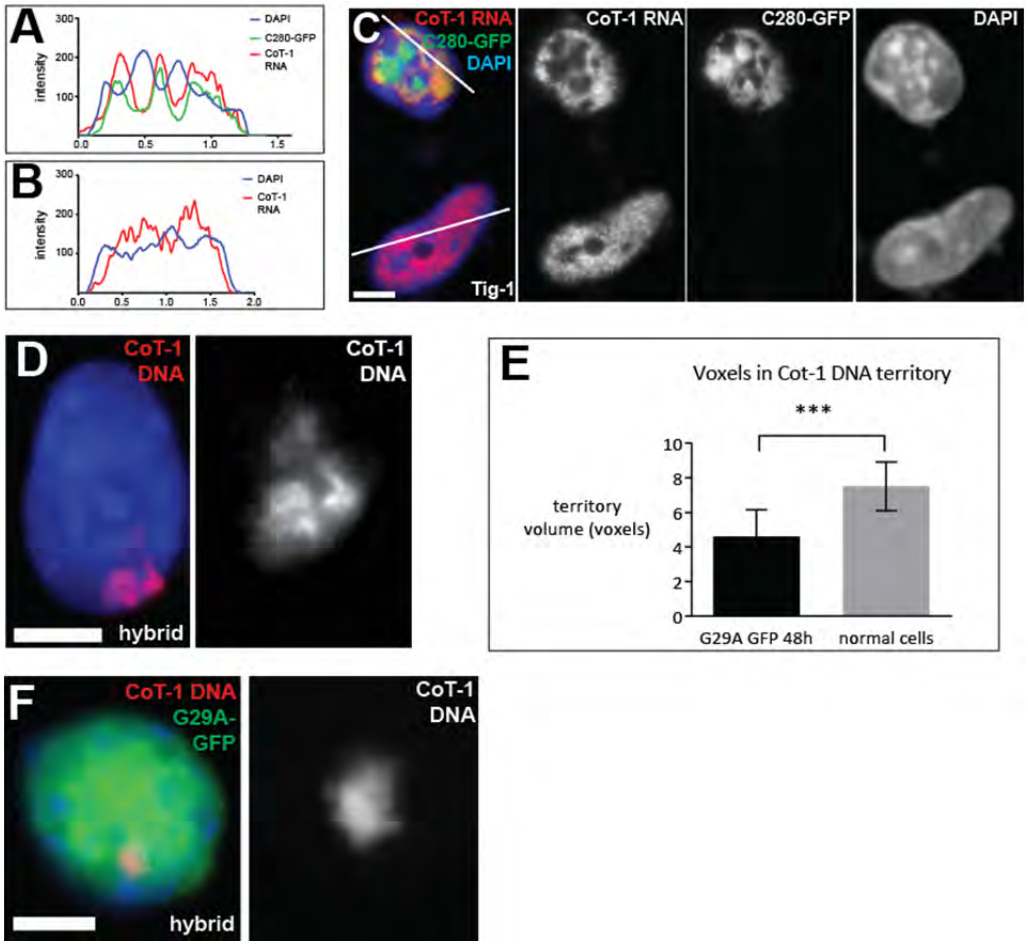

Figure 5.6 SAF-A mutants cause chromatin collapse.

(A-C) C280-GFP expression causes chromatin condensation in human Tig-1 fibroblasts. (A) Linescan shows higher DAPI peaks indicating large variations in staining intensity in GFP-positive cell compared to lower, more uniform DAPI intensity in control cell (B). CoT-1 RNA localizes with C280-GFP and not with DAPI (A). (D-E) The human Chr 4 territory collapses in mouse human hybrid cells expressing SAF-A's DNA binding domain mutant. The territory (hybridized to human CoT-1 DNA) in normal cell (D) is larger than in cell expressing G29A-GFP (F). The total volume of the chromosome territory decreases by 38% (*P*=.0002, n=11) (E). Error bars, standard deviation of the mean. Z-stacks deconvolved and voxel volume measured using Huygens software. Scale bars, 5 µm.

and not lack of SAF-A, is the only condition I observed to *consistently* cause chromatin condensation.

RNA is an integral component of the nuclear scaffold required to retain SAF-A but not Lamin B1

Earlier literature describing evidence for a non-chromatin nuclear scaffold or matrix shows that ~70% of nuclear RNA remained after extensive fractionation removes over 90% of protein and DNA. While the concept was controversial, significant evidence suggested that processes such as splicing occurred in a structurally bound context, hence most pre-mRNAs undergoing metabolism would be bound to the matrix (Xing and Lawrence, 1991; Nickerson, 2001). In addition, some views suggested that certain RNAs were not just bound to the matrix, but that unidentified RNA(s) may comprise an integral component of that structure (Nickerson et al., 1989; He et al., 1990). Hence, several properties of CoT-1 RNA led us to question whether it was bound to the scaffold by one or more proteins, or whether it may be a component of the scaffold itself, such that it would support the retention of specific proteins in the scaffold. For example, results described above indicate this abundant CoT-1 RNA fraction promotes open chromatin structure, since its release consistently is associated with abnormal chromatin condensation and/or chromocenter loss. I therefore tested whether CoT-1 RNA depletion affects the distribution of known nuclear matrix proteins. Changes to matrix structure and composition would implicate CoT-1 RNA as an architectural component of the nuclear matrix.

Since CoT-1 RNA is heterogeneous, stable, and highly abundant, it is not feasible to knock down using traditional methods. As reported in Chapter IV, prolonged transcriptional inhibition eliminates pre-mRNA but does not eliminate stable repeat-rich CoT-1 RNAs from interphase nuclei (Hall et al., 2014). However, CoT-1 RNA is released from chromosomes during mitosis (as is SAF-A), and cells that divide in the presence of transcriptional inhibitors cannot resynthesize it (Hall et al., 2014).

I examined SAF-A's distribution after transcriptional inhibition and found it was substantially diminished and aberrantly distributed in G1d cells lacking CoT-1 RNAs. In contrast to its more homogeneous granular distribution on chromatin in normal G1d cells (Figure 5.7A), or in DRB-treated interphase cells that retain CoT-1 RNA but lack mRNAs (Figure 5.7B), remaining SAF-A was prominently localized in small rings roughly 1μM in diameter that localized in DAPI holes (Figure 5.7C, arrows). This suggests that endogenous SAF-A largely requires long-lived nuclear RNAs for its chromatin association.

To test the possibility that SAF-A may require an RNA component to maintain its relationship to nuclear structure, I treated unfixed, permeabilized Tig-1 fibroblasts with RNAse and visualized SAF-A by SIM. Strikingly, within ten minutes of RNAse treatment, the SAF-A signal effectively disappeared (Figure 5.8A-B). This indicates that the widely-held assumption that SAF-A functions as an anchor for chromosomal RNA may in fact be backwards, and that instead, RNA may anchor SAF-A. It is consistent with results in Chapter III showing that XIST RNA may retain SAF-A on the mitotic Xi when it is retained during AURKB inhibition.

I next tested whether RNAse treatment affects two other nuclear matrix proteins, NuMA and Lamin B1. I found that while the NuMA signal was reduced in a fraction of cells after RNAse treatment, it still retained a relatively homogeneous distribution (Figure 5.8C-D), while the Lamin B1 signal was unaffected (Figure 5.8E-F). I reduced the RNAse treatment to one minute and found similar results for SAF-A, NuMA, and Lamin B1 (not shown).

The impact of RNAse treatment on SAF-A and NuMA but not Lamin B1 suggests RNA removal does not simply cause non-specific devastation of nuclear structure, but rather suggests there are “layers” to nuclear substructure, one that requires RNA for integrity and one that does not. Collectively, these data suggest that CoT-1 repeat RNA is a significant fraction of the RNA comprising an integral part of the nuclear scaffold. As suggested by the model in Figure 5.9, SAF-A requires this RNA to form a scaffold for heterochromatin and euchromatin, but, importantly, other proteins do not.

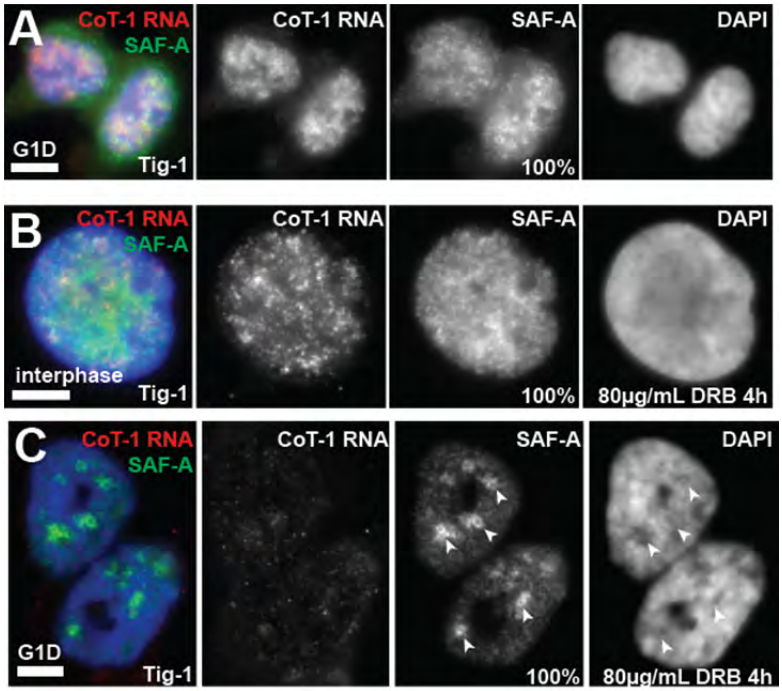

Figure 5.7 CoT-1 RNA depletion impacts SAF-A's association with chromatin.

(A) Normal G1 daughter cells express both SAF-A (green) and CoT-1 RNA (red). (B) CoT-1 RNA and SAF-A remain on chromatin in 100% of transcriptionally inhibited interphase cells. (C) Some SAF-A is recycled during mitosis but it does not localize properly to chromatin in transcriptionally-inhibited G1 daughter cells lacking CoT-1 RNA. SAF-A (green) is largely excluded from DAPI (arrows) in 100% of transcriptionally inhibited G1 daughter cell that lack CoT-1 RNA (red). Scale bars, 5 µm.

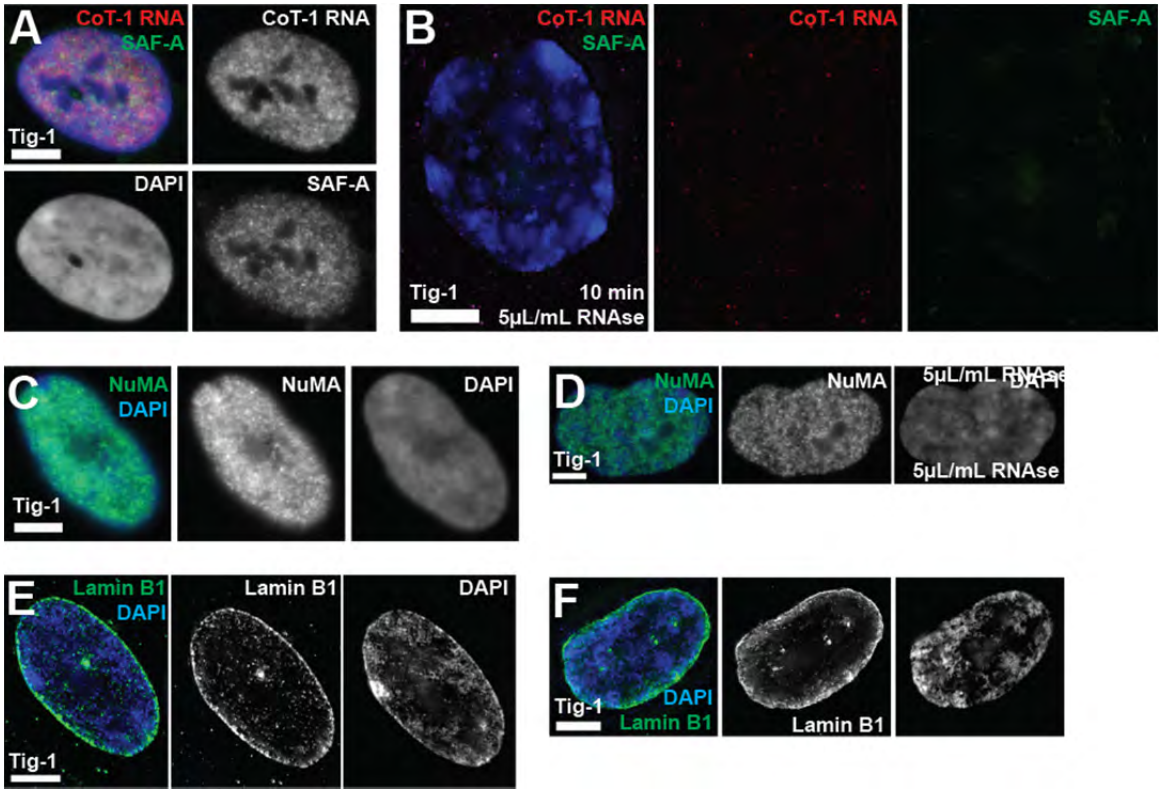

Figure 5.8 SAF-A association with chromatin is RNA-dependent.

(A) CoT-1 RNA (red) and SAF-A (green) in normal Tig-1 cells. (B) SAF-A evacuates the nucleus after 10 minutes of live-cell RNAse treatment (100% of cells). (C-D) NuMA (green) is marginally depleted in some cells but largely unaffected by RNAse treatment. (E-F) Lamin B1 (green) is unaffected by RNAse treatment (100% of cells). SIM images in E-F are single slices from z-stacks deconvolved in Softworx (see Materials and Methods). Scale bars, 5 µm.

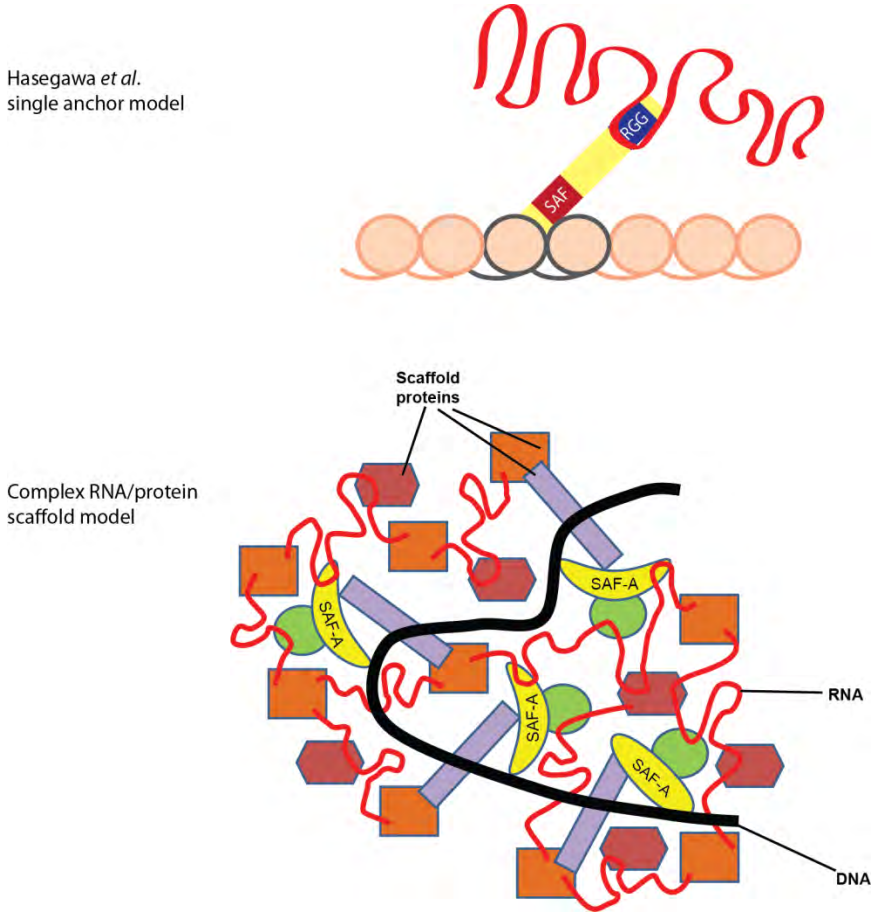

Figure 5.9 Contrasting two models of chromosomal RNA localization.

Above: Model depicting SAF-A as a single molecule bridge anchoring chromosomal RNA to chromatin, adapted from graphical abstract of Hasegawa et al. (2010).
Below: Our revised model showing SAF-A and chromosomal RNA woven into a "fabric" of the interphase chromosome, in which multiple proteins bind to chromosomal RNA.

Discussion

The release of CoT-1 RNA by expression of truncated SAF-A may initially appear to indicate that SAF-A is directly involved in localizing CoT-1 RNA. However, results here indicate that this and potentially other SAF-A mutants have dominant negative effects on other chromosomal architectural proteins with which SAF-A and chromosomal RNAs interact. Results indicate that CoT-1 RNA and SAF-A associate with a common structure, referred to here as the nuclear scaffold. The results also suggest a possibility with broad fundamental implications for genome and chromosome biology: that the abundant presence of RNA on chromatin counters an inherent propensity to condense. This possibility is consistent with various results in this Chapter, where I have attempted to characterize the association between CoT-1 RNA and SAF-A. I found that expressing several specific SAF-A deletion mutants, but not SAF-A RNAi, perturbed CoT-1 RNA's relationship to chromatin, and loss of CoT-1 RNA from chromatin was consistently accompanied by widespread chromatin condensation, visible in human cells by formation of large DAPI clumps.

Experiments treating live cells with RNase further indicate that RNA is required to prevent visible, gross changes to DNA condensation suggestive of chromatin collapse. Furthermore, these experiments also reveal that RNA is required for SAF-A to bind to chromatin, and therefore for the integrity of what is an *RNA-dependent* chromosome scaffold. However, RNAse experiments do not address whether the collapse is specific to loss of CoT-1 repeat-rich RNAs or to RNA in general. While CoT-1 RNA may reflect a diversity of repeat-rich lincRNAs and intron-containing RNAs in nuclei, we show in

Chapter IV that these sequences appear to be the most prominent component of euchromatin-associated RNA, therefore the CoT-1 component likely plays a large role in the collapse.

Experiments using transcriptional inhibitors more convincingly show that collapse is due to loss of CoT-1 RNA specifically (not loss of all RNA) since only G1d cells lacking both CoT-1 RNA and short-lived transcripts are affected. Here I also show evidence that while some SAF-A is clearly recycled after cell division, it is not properly localized in DRB-treated G1d cells, further suggesting the requirement for RNA on chromatin to replicate the matrix following cell division. The failure of G1d nuclei to rebuild open chromatin structure despite the recycled SAF-A suggests that an RNA-dependent matrix plays a fundamental role as an integral component of interphase chromosome structure.

Work in this chapter advances understanding of how abundant nuclear CoT-1 RNAs associate with the chromosome territory and how RNA impacts overall chromosome packaging. Findings provide renewed support for the earlier contested concept of a nuclear scaffold or “matrix,” which was reported to bind most nuclear RNA. Evidence that that SAF-A deletion mutants, but not SAF-A RNAi, impact CoT-1 RNA association with interphase chromosomes supports that chromosomal RNAs are not just tethered by one or two scaffold proteins, but are intimately intertwined in a complex structure (see model in Figure 5.9). As suggested by the model, results show that specific proteins and RNA’s association with the chromosome is mutually dependent.

Finally, findings here contribute an additional, unanticipated insight into the fundamental nature of nuclear chromosome structure. In marked contrast to the release of SAF-A by RNAse, removal of RNA had no effect on the retention of Lamin B1, not only at the nuclear periphery but throughout the nucleoplasm. This finding merits and will prompt further investigations into the hypothesis that nuclear chromosome architecture involves at least two distinct and separable layers: one, which involves SAF-A and is RNA-dependent, and the other, involving lamins, which appears to be RNA-independent. In either case, these structural elements are not simply proteins or RNA bound to canonical chromatin (DNA wrapped in histones), but components of the complex fabric which packages and regulates the genome.

Chapter VI : Final Summary and Conclusions

Overall Summary

The large non-coding XIST RNA is the premier example of an RNA that binds an interphase chromosome to modify its structure and epigenetic state. We now have evidence for a heterogeneous but novel class of chromosome-associated repeat-rich CoT-1 RNA that like XIST, binds to the interphase chromosome territory in *cis* and may function as an architectural RNA. However, even for the heavily studied XIST RNA, we lack a clear understanding of how chromosomal RNAs structurally associate with chromosomes. This study advances knowledge of how XIST and CoT-1 chromosomal RNAs integrate into chromosome scaffolds for heterochromatin and euchromatin, respectively, to support epigenetic genome regulation. I demonstrate unrecognized complexity to the anchoring of chromosomal RNA embedded in what we suggest is a complex interconnected “fabric” of interphase chromosomes.

At the onset of this work, we anticipated initial experiments would validate and extend the current predominant model, which depicts Scaffold Attachment Factor-A tethering chromosomal RNA to chromosomal DNA as a single molecule bridge. I began to examine the role of SAF-A for XIST and CoT-1 RNAs in human cells. Through series of experiments using RNAi and SAF-A mutants in numerous cell types, I found that SAF-A is not required for XIST/Xist RNA localization in normal human or mouse cells. Evidence leads to a different model, in which SAF-A is one of multiple proteins in nuclear scaffold that is mutually dependent on the presence of RNA to associate with chromatin. Evidence supports that changes to the nuclear scaffold during transformation

may compromise anchors in certain cancer cells, as most clearly illustrated by mouse Neuro2a cells. In addition, results expressing a series of SAF-A deletion mutants demonstrated that the DNA binding domain was required to support XIST RNA localization, but, unexpectedly, the RGG domain capable of binding RNA was not. This contradicts the assumption that a single protein tethers chromosomal RNAs to the chromosome as a direct RNA-DNA bridge.

Similarly, we find that repeat-rich CoT-1 RNA associates with euchromatic interphase chromosome territories but releases upon disruption of the nuclear scaffold. Therefore I examined the potential role of SAF-A in the chromosomal association of CoT-1 RNA. Highly effective SAF-A depletion had no effect on CoT-1 RNA localization to the human chromosome, whereas a single point mutation in the DNA binding domain releases CoT-1 RNA, resulting in abnormal DNA condensation. Unlike results for XIST RNA, SAF-A's RGG domain appears required to localize CoT-1 RNA, and CoT-1 RNA loss is associated with euchromatin condensation, also supported by RNAse experiments.

Importantly, results support a model in which repeat-rich CoT-1 RNAs are not merely tethered to architectural proteins, but instead are integral to nuclear architecture as components of a complex nuclear scaffold/matrix. Remarkably, SAF-A is completely lost from RNAse-treated nuclei, indicating the potential for mutually dependent chromosomal association of SAF-A and RNA. In contrast, Lamin B1 is unaffected by RNAse treatment, suggesting two "layers" to the architectural underpinnings of interphase chromosomes, one that requires the RNA structural component and one that does not. The RNAse experiment indicates that RNA in the scaffold does not function via

recruitment of epigenetic factors, but suggests the possibility that the physical presence of this diverse set of repeat-rich RNA on the chromosome serves to keep it open. It is thus tempting to speculate that the default state of nucleosomal chromatin is condensation.

Caveats

Reproducibility of scientific experiments is a major problem plaguing researchers today. One recent study estimated that >50% of pre-clinical research is not reproducible, at a cost of roughly 28 billion dollars in NIH funds (Freedman et al., 2015). While this figure is almost certainly inflated (Kaiser, 2015), it does highlight the degree of the problem. Three major factors contributing to lack of reproducibility are widespread in the literature: (1) Transient overexpression of proteins and/or mutant proteins (2) introduction of protein tags (GFP, Flag, Myc, etc.) and (3) use of transformed cells or inconsistent cell types. Each of these widely-used experimental approaches can introduce off-target effects that highly complicate data analysis and interpretation.

Throughout this work, I have struggled with complications from all three of these factors and made an effort to control for them as much as possible. For example, I created stable cell lines to avoid effects of protein overexpression, and when I had to use transient transfections, I carefully scored cells according to whether they expressed “high” or “low/moderate” levels of the transfected protein in the most objective manner possible.

I additionally examined endogenous SAF-A by IF (rather than using tagged proteins) whenever possible. When I did use tagged proteins, I compared both GFP and Flag-tagged constructs to the behavior of endogenous SAF-A. Our Flag and GFP-tagged mutants produced similar phenotypes, however, the GFP-tagged version often affected a

higher percentage of cells at a lower expression level, and this depended on the specific mutation or the cell type.

Results presented in this thesis also highlight the problem of using transformed cells to test principles of basic biology. In Chapter III, I attempt to correct a misunderstanding in the field that stems from use of transient transfections in transformed cells. In testing the effect of SAF-A RNAi on XIST RNA localization, I show that phenotypic differences arise in different cell types. In particular, transformed cells exhibit heterogeneity not only between tissue types, but also cell types, cell lines, and individual cells. In many cases I saw significant variation in my results that may have resulted from epigenetic differences within a cultured cell line.

While these studies presented many challenges, they also highlight the importance of using *in situ* experiments to validate findings from extraction-based techniques. For example, as described above, hnRNP M and C were identified as top candidate Xist RNA interacting proteins in recent RNA pull downs, however I show that neither protein appears to localize to the Barr body. Thus, although scoring can be subjective, looking directly at cells, rather than simply extracting their various components, is an essential part of understanding biology.

Major results, implications and future directions

XIST RNA recognizes the autosome territory boundary in a differentiated cell

To initiate chromosome inactivation for dosage compensation, XIST RNA spreads across the inactivating X chromosome and localizes strictly in *cis*, limiting its

spread to neighboring chromosome territories which are densely packaged together in the 10-20 μM nucleus. As I show in Chapter II, transgenic XIST RNA can also spread and localize across an autosome in *cis* in a fully differentiated cell. This has several implications for both the mechanism of XIST RNA spread and the required cellular context for initiation of chromosome silencing.

Although some studies suggest that L1 DNA elements may facilitate XIST RNA propagation across the chromosome (Lyon, 1998; Bailey et al., 2000), as described above, XIST RNA appears to localize through interactions with the nuclear scaffold rather than by interacting directly with DNA. Engreitz et al. (2013) proposes a model in which Xist RNA spreads via proximity in 3D. In this model, Xist RNA localizes to regions of the chromosome that are nearby in 3D space by binding to the nuclear scaffold proteins associated with that DNA, then actively modifying the chromosome conformation to "pull" more DNA into the silencing compartment. While this model does not rely on specific DNA sequences and should theoretically work on any chromosome, it does not account for how Xist RNA avoids spreading to neighboring chromosomes.

XIST RNA recognition of both the Xi and an autosome boundary is consistent with the hypothesis that individual chromosome scaffolds are distinct and separable from those of neighboring chromosomes. In Chapter V, I show evidence for two distinct nuclear scaffolds, one that depends on an RNA component, and one that does not. These findings provide support for a hypothesis that RNA-dependent scaffold proteins could somehow delineate the boundaries between individual chromosome territories to limit XIST RNA spread to neighboring chromosomes. It is possible that XIST RNA

recognizes and even becomes an integral part of the RNA-dependent single-chromosome scaffold, while the RNA-independent scaffold (the lamina) functions to fix all the chromosome territories together in space.

Results in Chapter II also imply that any chromosome can transform from euchromatin into a heterochromatic Barr body. Further, this transformation is not strictly limited to a pluripotent setting, which is the natural setting for XCI. Anton Wutz and others suggested that only pluripotent cells have capacity to initiate chromosome silencing (Surralles and Natarajan, 1998; Wutz and Jaenisch, 2000; Savarese et al., 2006), though one study showed that viral expression of Special AT-rich Binding protein (SATB1) confers the capacity for silencing post-differentiation (Agrelo et al., 2009). However, these studies are limited by the use of transgenic *Xist* targeted to an autosome in a diploid cell, thus creating a deleterious monosomy upon *Xist* activation, and in some cases, the read-out for chromosome silencing is cell death (Wutz and Jaenisch, 2000; Savarese et al., 2006). Some prior evidence suggests differentiated HT1080 cancer cells have capacity for silencing (Hall et al., 2002a; Chow et al., 2007). However, as discussed in Chapter III, cancer cells often resemble pluripotent cells epigenetically, and may therefore exhibit epigenetic plasticity that normal somatic cells lack. Importantly, our system targeting human *XIST* to a trisomic chromosome 21 provides further indication that *XIST*-mediated human chromosome silencing may be possible in a fully differentiated primary cell.

Proper XIST RNA localization requires specific scaffold proteins and DNA packaging, not specific DNA sequence

The capacity for transgenic XIST RNA to spread to and silence an autosome also might suggest that XIST RNA could spread across and silence a chromosome from a different species. However, evidence presented in this thesis and prior evidence suggests that chromosome packaging may be more important than DNA content, therefore, the chromosome's species of origin may matter less than the cellular context. For example, human XIST RNA will not spread to a human chromosome in a mouse cell (Clemson et al., 1998). This is particularly interesting given that SAF-A, shown to be involved in localizing XIST/Xist RNA in both species, is highly conserved, with only ~20 amino acids different between mouse and human. Therefore, this provides further support that SAF-A is not the sole XIST/Xist RNA localization factor in either species, and it supports the hypothesis that the proteins packaging chromosomal DNA are essential for XIST RNA localization, and the human RNA and DNA alone are not sufficient.

Additional support that proper DNA packaging is required for XIST RNA spread comes from studies showing human XIST RNA can paint a *mouse* chromosome in *mouse* ES cells. Although this only occurred when *XIST* was expressed from a yeast artificial chromosome (YAC) containing the entire XIC (Heard et al., 1999; Migeon et al., 1999), and not in clones with low-copy number *XIST* integrations (in which the RNA may have been unstable) (Heard et al., 1999; Migeon et al., 2001b) and (Lawrence lab, unpublished observations), this supports that scaffold proteins are more important for RNA localization than the DNA (or RNA) by itself. As further illustrated by the table below,

the human RNA can't localize unless the DNA and scaffold proteins are species compatible.

It would be interesting to test whether human XIST RNA would spread to a mouse chromosome in a *human* cell, where the RNA and proteins are species-compatible, but the DNA is not. Evidence that DNA sequence is not involved in XIST RNA spread suggests that perhaps it would. However the human packaging proteins may not properly fold the foreign DNA, making it unrecognizable to the RNA.

RNA	**DNA**	**Proteins**	**Localized RNA?**
Human	Human	Mouse	no
Human	Mouse	Mouse	yes (copy-number dependent)
Human	Mouse	Human	?

Similar studies have been done in other organisms. *XIST*-mediated XCI is a eutherian mechanism, however marsupials also inactivate an X chromosome through the action of a recently-discovered noncoding RNA that like XIST RNA, paints the inactive chromosome territory (Grant et al., 2012; Whitworth and Pask, 2016). RNA on the silent X (*RSX*) is a ~27 kb spliced and polyadenylated RNA enriched in several tandem repeats, however it has no sequence homology with *Xist/XIST* (Grant et al., 2012). Notably, it was shown that a *single-copy* Rsx transgene inserted into mouse chromosome 18 produces an RNA that paints the mouse chromosome and induces partial silencing. We also have evidence that under some circumstances mouse Xist RNA paints a human chromosome in

a human cell (unpublished observations), therefore, human XIST RNA may be unique in its more limited ability to localize to foreign or improperly packaged DNA.

Other species with X and Y sex chromosomes have mechanisms of dosage compensation that differ more significantly from eutherians, and these include upregulation of the X chromosome in male *Drosophila* (through the MSL complex which includes *Rox* ncRNAs), and partial silencing of both X chromosomes in *C. elegans* hermaphrodites (through the multi-protein DCC complex which contains several condensin proteins) (for review: (Chery and Larschan, 2014)). It is interesting to speculate whether XIST RNA would paint and silence a *C. elegans* chromosome, for example, in either a human or *C. elegans* cell. Evidence suggests that cellular context is more important than chromosome sequence suggests that this would be more likely to occur in the human cell than in the *C. elegans* cell.

Multiple anchors localize XIST RNA

As described in Chapter I, identification of XIST RNA binding and localization factors has been difficult. XIST RNA is embedded in nuclear structure and has eluded most extraction-based methods of study. Hasegawa et al. (2010) reported that SAF-A depletion consistently released Xist RNA in mouse neuroblastoma cells, suggesting that SAF-A, an abundant nuclear matrix protein that binds both DNA and RNA, is the singular factor required for its localization. However, this finding contradicts prior evidence manipulating XIST RNA during the cell cycle, which suggested multiple proteins would anchor the long RNA molecule to the human chromosome, and that one of these proteins is a substrate of the mitotic kinase Aurora B (Hall et al., 2009).

Therefore, despite decades of work, we still lacked a clear picture of the proteins involved in XIST RNA localization.

Described in Chapter III, I investigated the role of SAF-A in human XIST RNA localization and provide evidence that the factors binding XIST RNA to the chromosome exhibit complex biology, which varies with cell type. I affirm that mouse Xist RNA is completely delocalized in the Neuro2a neuroblastoma cells studied by Hasegawa et al. (2010), but its localization is at most mildly perturbed in normal MEFs, consistent with our results in primary human cells. These findings in MEFs directly contradict the 2010 study, which relied on MEFs to support that the observations in Neuro2a cells were generally applicable and not due to anomalies of the Xi in the Neuro2a line, which are known to lack specific Xi hallmarks. This thesis contributes a needed depth of perspective and correction of an overly simplified view of the biology of Xist RNA localization and erroneous conclusion that SAF-A is singularly required for Xist RNA localization (Hasegawa et al., 2010).

I uncovered multiple lines of evidence supporting that XIST RNA is not tethered to the chromosome by a single anchor protein. The Neuro2a cells used by Hasegawa et al. (2010) proved the only cell line, even among several transformed cell lines, to show a complete requirement for SAF-A to localize Xist RNA. Moreover, in various normal cells tested, effective SAF-A RNAi simply did not release XIST/Xist RNA from the chromosome. Importantly, this included the same cells and RNAi approach Hasegawa et al. used, further suggesting it is not a technical difference. In addition, I found evidence that SAF-A is not required to keep XIST RNA on the chromosome using a very different

experimental strategy, by manipulating the normal release of XIST RNA during mitosis. Our lab previously showed that AURKB activity is one factor controlling XIST RNA release (Hall et al., 2009). Here, I further show that AURKB phosphorylation is not required for SAF-A release. Thus, SAF-A is not the "AURKB-dependent" XIST RNA anchor. Supporting this, simultaneous knockdown of SAF-A and activation of AURKB during interphase increased XIST RNA mislocalization in normal cells. Therefore, while SAF-A may be involved in localizing XIST RNA, it is one of multiple factors.

Most recently, I found preliminary evidence that SAF-A may not be the only Xist RNA anchor in Neuro2a cells. Human SAF-A consistently releases from the mitotic chromosome after XIST RNA (Chapter III), but mouse Xist RNA hangs on to the mitotic chromosome slightly longer. I frequently observed Neuro2a cells in which Xist RNA remained on the mitotic Xi after SAF-A had been released (data not shown). While this may be transient and/or fleeting, it is possible that another protein anchors Xist RNA after SAF-A has been released. If, like all other cell types tested, Neuro2a cells have multiple anchors for Xist RNA, SAF-A RNAi may affect Xist RNA in these cells in a more dominant negative way, for example, by disrupting the overall nuclear scaffold. This is supported by visible changes to chromatin structure that occur after SAF-A knockdown in these cells that do not appear to occur in other cells types examined (not shown).

It will be important for future studies to examine the potential role of putative Xist RNA binding proteins recently identified in several RNA pull downs described above. I have demonstrated a tool to examine relationships *in situ*, to determine whether specific

proteins localize with XIST RNA when it is forcibly retained on the mitotic Xi. I show evidence that one of these proteins, hnRNP K, binds XIST RNA directly within the cell, but hnRNP M and C may not. Future studies examining additional antibodies in both somatic and pluripotent cell types may help to narrow down the long list of candidate Xist RNA binding proteins. This may also lead to identification of the putative "AURKB-dependent" XIST RNA anchor protein, if one of these candidates is found to be retained on all mitotic chromosomes after AURKB inhibition when it is normally released.

Epigenetic evolution of the chromatin-RNA interactome in transformed and cancer cells

This study provides evidence for a new concept: that the normal faithful binding of an RNA that epigenetically regulates chromatin may be commonly weakened or compromised in cancer. Cancer progression almost invariably involves epigenetic as well as genetic instability (for review, see (Timp and Feinberg, 2013)). Any compromise in the binding of epigenetic RNAs to chromatin could contribute to this instability.

While SAF-A RNAi minimally impacted XIST RNA localization in normal cells, it had a moderate effect in several immortalized and transformed cell lines. Compromise of RNA's interaction with chromatin may be common and important to the deregulation of the epigenome in cancer, and it likely relates to other perturbations to nuclear structure during transformation and cancer progression, which commonly includes loss of organized heterochromatin. Earlier studies suggesting that neoplasia can involve changes to the nuclear scaffold (Getzenberg et al., 1991; Konety and Getzenberg, 1999) support the suggestion that RNA mislocalization is underpinned by loss or compromise of certain

nuclear architectural proteins. It appears such changes can contribute to heterogeneity, as I observed the XIST RNA phenotype shows marked variability in clusters of cells of some transformed cell lines (before and after treatment), consistent with epigenetic evolution of the chromatin-RNA "interactome."

Further supporting this, XIST RNA localization is lost or compromised in many cultured cancer cell lines (Pageau et al., 2007a; Pageau et al., 2007b). I also noted that the XIST RNA territory often appears "decondensed" in transformed cells prior to SAF-A knockdown, similar to fibroblasts after SAF-A knockdown. Hence, this leads credibility to the conclusion that cancer cells can commonly lack certain factors that normally contribute to the packaging of XIST RNA territories, making XIST RNA more susceptible to release by knockdown of a single nuclear scaffold factor (i.e. SAF-A).

As described in Chapter III, one XIST RNA anchor is a substrate of Aurora B Kinase, which phosphorylates H3 Serine 10 along the chromosome arms at the onset of mitosis. Overexpression of AURKB is another common property of cancer cells (Hegyi and Mehes, 2012). AURKB activity was shown to be essential for proper chromosome condensation and segregation, and failure of H3 phosphorylation can lead to polyploidy (Hauf et al., 2003). However, through an unknown mechanisms, hyper phosphorylation of H3S10 caused by AURKB overexpression can lead to aneuploidy (Ota et al., 2002), which is both a symptom and a cause of epigenetic instability that can lead to cancer.

Since AURKB activity is required for the mitotic release XIST RNA from the chromosome (Hall et al., 2009), it is possible that excess active AURKB contributes to the loss of XIST RNA commonly seen in cancer cells. It may also abrogate the AURKB-

dependent XIST RNA anchor, making cancer cells more susceptible to XIST RNA release by SAF-A RNAi alone. For example, in Chapter III I show evidence that simultaneous activation of AURKB (via RNAi to its inhibitor, PP1) and depletion of SAF-A in primary cells releases more XIST RNA than SAF-A or PP1 RNAi alone, perhaps mimicking the situation already existing in transformed cell lines that overexpress AURKB.

Although my data provides evidence that SAF-A is not required for XIST RNA localization in somatic cells, it supports prior evidence that SAF-A is required for formation of the XIST RNA territory during XCI initiation in pluripotent cells (Hasegawa et al., 2010). Pluripotent cells share several properties with cancer cells including a relative lack of heterochromatin and “open” nuclear structure (Meshorer and Misteli, 2006; Carone and Lawrence, 2013). Changes occurring during cancer development often return cells to a relatively de-differentiated state. Prior evidence has shown that the nuclear scaffold changes during differentiation, as it does during transformation (Dworetzky et al., 1990). We hypothesize that as cells differentiate, the nuclear scaffold evolves to contain many more proteins, including proteins that contribute to XIST RNA localization. In somatic cells, these proteins act in a semi-redundant fashion to maintain XIST RNA localization and heterochromatin organization.

Collective evidence supports a model in which chromosomal XIST RNA attaches to a nuclear scaffold that increases in complexity as cells differentiate, and loses complexity in cancer. However, future studies are needed to further characterize the relationship of chromosomal RNA to nuclear scaffold proteins. While work presented

here suggest a strong correlation between transformation and release of XIST RNA from the Barr body, similar analyses in additional cell types (both primary and transformed) are needed to further examine this potentially broadly important point for cancer biology. In addition, proteomic analysis after nuclear matrix fractionation in cancer compared to normal cells would be compelling to more precisely characterize changes to the nuclear matrix that occur with transformation.

SAF-A may not function as a single molecule RNA-DNA bridge

Having found that the nuclear scaffold is involved in localizing both XIST and CoT-1 chromosomal RNAs, and that SAF-A likely plays an important role, in Chapters III and V, I tested the requirement for SAF-A's separate RNA and DNA binding domains. These experiments were complicated by the fact that transient overexpression of the SAF-A deletion mutants and of the full-length protein mislocalized XIST and CoT-1 RNAs, while also inducing dramatic changes to overall chromatin structure (visible by DAPI staining). This suggests the stoichiometry of SAF-A must be carefully regulated for maintaining nuclear structure, and overexpression may disrupt other scaffold proteins in a dominant negative manner.

I found that regardless of expression level, SAF-A containing a single point mutation in the DNA binding SAP domain (G29A) delocalizes XIST RNA in multiple cell types, and it was unable to rescue XIST RNA localization after SAF-A RNAi (in transformed cells in which SAF-A depletion has an effect). In contrast, expression of the putative RNA binding domain deletion mutant (ΔRGG) had a minimal effect on XIST RNA localization, and, surprisingly, it was able to rescue XIST RNA localization

following SAF-A knockdown. Since this result did not affirm the conclusions from rescue experiments with the mouse ΔRGG mutant by Hasegawa et al. (2010), I examined this question multiple times, first by transient transfection as in Hasegawa et al., and then in stable cell lines expressing dox-inducible SAF-A (and deletion mutants) to avoid the overexpression effects of SAF-A constructs on XIST RNA localization. Hence, my results further argue against a simple model that SAF-A forms a single molecule bridge necessary to connect XIST RNA to the Xi DNA.

I was not able to perform the same rescue experiments with CoT-1 RNA, since SAF-A RNAi had no effect on CoT-1 RNA in the hybrid cells used to visualize CoT-1 RNA localization. Nevertheless, at lower expression levels, both the SAP and RGG domain mutants mislocalized CoT-1 RNA in over 50% of cells (human fibroblasts and mouse-human hybrid cells), and the ΔRGG mutant had a greater effect on CoT-1 RNA than on XIST RNA. The fact that the distribution of the ΔRGG mutant no longer closely coincides with CoT-1 RNA, which becomes displaced from chromatin, suggests that SAF-A normally binds CoT-1 RNA through its RGG domain. It is interesting that when the ΔRGG mutant localizes over chromatin, both endogenous SAF-A and CoT-1 RNA become excluded. Although this could fit with a requirement of SAF-A's RGG domain to localize CoT-1 RNA to chromatin, our collective findings suggest that SAF-A mutants can have dominant negative effects on other chromosomal proteins and more generally perturb the nuclear scaffold, and the total lack of a CoT-1 RNA phenotype after SAF-A RNAi argues against a model in which SAF-A is a single molecule bridge between CoT-1 RNA and DNA.

It is generally assumed that scaffold proteins bind or tether chromosomal RNAs to chromatin (histones wrapped in DNA). SAF-A has been considered a prime candidate for this with its separate RNA and DNA binding domains. Results in Chapters III and V support a model in which SAF-A is fundamental in attaching S/MAR regions of DNA to the nuclear scaffold. This is consistent with early evidence that SAF-A not only binds AT-rich DNA with high affinity *in vitro*, but also organizes DNA into loops (Fackelmayer et al., 1994). However, direct binding between SAF-A and chromosomal RNA is less clear. Differences in the XIST and CoT-1 RNA phenotypes in cells expressing the ΔRGG domain mutant suggest that the two RNAs may differ in how they interact with the nuclear matrix and SAF-A. This is plausible given the divergent nature of euchromatin and heterochromatin. Further experiments are required to truly know the nature of the relationship between SAF-A's RGG domain and chromosomal XIST and CoT-1 RNAs. Despite this, collective results using both RNAi and deletion constructs support a model in which SAF-A is a major player in forming a "basement" layer of nuclear architecture, supporting that interconnected DNA, RNA, and protein are woven into a "fabric" of the interphase chromosome (see model in Figure 5.9)

Chromosomal association of SAF-A and RNA may be mutually dependent

We were surprised to find that RNAse treatment which effectively eliminates RNA from nuclei also eliminates SAF-A. It is curious that disruption of SAF-A's association with RNA would also impact its association with DNA, especially given that SAF-A was shown to have greater affinity for DNA than for RNA *in vitro* (Fackelmayer et al., 1994). This finding also is not compatible with the prevailing model in which SAF-

A anchors chromatin-associated RNAs to the scaffold. Instead, this suggests the potential for a mutually dependent chromosomal association of SAF-A and RNA. The idea that a protein may require RNA for association with chromatin is not unprecedented, as it was recently shown that noncoding RNAs transcribed from regulatory elements may tether the transcription factor YY1 to those sites in a positive feedback loop stabilizing transcription (Sigova et al., 2015).

Other evidence supports the concept that SAF-A associates with chromatin in an RNA-dependent manner. As described in Chapter III and above, XIST RNA is normally released from the Xi in prophase after activation of AURKB. AURKB inhibition (with the drug Hesperadin) leads to XIST RNA retention on the mitotic Xi, and this leads to Xi-specific retention of SAF-A (which releases from all other chromosomes as usual). While it is possible that AURKB phosphorylation impacts SAF-A specifically on the Xi but not on all other chromosomes, a more likely explanation is that XIST RNA retention (by the AURKB substrate) leads to retention of the entire XIST RNA localization complex which includes SAF-A. In this situation, XIST RNA tethers SAF-A rather than SAF-A tethering XIST, as one might initially expect. It is noteworthy that mitotic release of CoT-1 RNA from euchromatic chromosomes is not affected by Hesperadin treatment. It would be interesting to see if SAF-A would be retained on all chromosomes with CoT-1 RNA if we had a way to test this.

In contrast to results with SAF-A, RNAse has no effect on the nuclear matrix protein Lamin B1. Lamin B1 is an intermediate filament protein that polymerizes into a network of nuclear fibers that concentrate in the nuclear lamina abutting the nuclear

envelope. The lamins also distribute throughout the nucleoplasm, thus providing mechanical stability and structure to the nucleus, analogous to the cytoskeleton (Goldman et al., 2002). Similar to SAF-A, lamins have been shown to play a role in higher order genome organization, binding to chromatin and to a large number of nuclear regulatory proteins (Dittmer and Misteli, 2011). The finding that two major structural components of the nucleus are differentially impacted by RNA depletion supports a new model in which the architectural fabric of chromosomes comprises two distinct "layers," one that requires an RNA component to maintain its structural relationships, and one that does not. While experiments with RNAse and the ΔRGG SAF-A mutant indicate that the RNA-dependent component is required to maintain "open" chromatin structure, other aspects of nuclear structure may depend on the lamina. These include overall nuclear shape, which is not altered in cells after RNA depletion, and/or the organization of chromosome territories.

Potential for competition between XIST and CoT-1 RNAs binding to the scaffold on the inactivating Xi

Comparing/contrasting CoT-1 RNA with XIST RNA offers an interesting perspective from which to evaluate the role of each RNA in determining chromatin structure. XIST and CoT-1 RNAs have several similarities, but CoT-1 RNA distributes specifically over euchromatin, rather than condensed heterochromatin like XIST. This mutually exclusive relationship presents an enticing model: that XIST RNA may out-compete or somehow displace CoT-1 RNA during the initial silencing of the X chromosome. Further, the nature of the RNA may modify the architectural arrangement of other components to shape chromatin packaging into heterochromatin or euchromatin.

The Barr body is enriched in CoT-1 DNA but lacks CoT-1 RNA (Clemson et al., 2006). While these sequences are what traditionally would have been considered silent junk DNA, evidence in Chapter IV suggests at least some may contribute to the CoT-1 RNA signal on the X chromosome prior to its inactivation during development (Hall and Lawrence, 2010). CoT-1 RNA silencing is one of the earliest events in the initial silencing of the X chromosome (Lawrence lab, unpublished observations). Prior evidence showed that formation of the CoT-1 RNA "hole" on the Barr body is spatially (Clemson et al., 2006) and temporally (Chaumeil et al., 2006) separable from gene silencing, and results described above show that CoT-1 RNA loss is coincident with DNA condensation. This suggests CoT-1 silencing may be a fundamental step in transformation of chromatin from active to inactive, as it is required for DNA condensation into the Barr body. It further raises the possibility that competition for XIST RNA binding contributes to CoT-1 RNA silencing and Barr body formation during initiation of XCI. This may also be a mechanism that allows for XIST-mediated silencing to occur on any chromosome (as shown in Chapter II), not just the X. Whereas XIST RNA may function to promote chromatin condensation, CoT-1 RNA may conversely promote decondensation. *In vitro* competition studies analyzing XIST and CoT-1 RNAs binding to SAF-A would further test this hypothesis.

Concluding remarks

Collective results presented in this thesis support a model in which SAF-A does not function as a bridge tethering RNA directly to DNA, and instead they suggest a more complex relationship. Supporting earlier evidence for an RNA component to the "core

filaments” of a nuclear scaffold (He et al., 1990), the data seems to fit a model in which chromosomal RNA serves as a virtual “nuts and bolts” for scaffold proteins and chromatin in the RNA-dependent fraction of the nuclear scaffold. This sheds light on how nuclear RNA that resists extraction contributes to interphase nuclear architecture and influences chromatin condensation. I further show that an additional component to the nuclear scaffold does not require the presence of this RNA. The recent discovery of a vastly underestimated transcriptome containing numerous previously unknown noncoding RNAs raises the possibility that other RNAs may also interact with and regulate chromatin loci as part of the nuclear scaffold. Such a discovery would improve our understanding of the epigenetic state and how it is altered in cancer.

REFERENCES

Agrelo, R., A. Souabni, M. Novatchkova, C. Haslinger, M. Leeb, V. Komnenovic, H. Kishimoto, L. Gresh, T. Kohwi-Shigematsu, L. Kenner, and A. Wutz. 2009. SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells. *Developmental cell*. 16:507-516.

Allen, T.A., S. Von Kaenel, J.A. Goodrich, and J.F. Kugel. 2004. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nature structural & molecular biology*. 11:816-821.

Bailey, J.A., L. Carrel, A. Chakravarti, and E.E. Eichler. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*. 97:6634-6639.

Barakat, T.S., F. Loos, S. van Staveren, E. Myronova, M. Ghazvini, J.A. Grootegoed, and J. Gribnau. 2014. The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. *Molecular cell*. 53:965-978.

Barboro, P., E. Repaci, C. D'Arrigo, and C. Balbi. 2012. The role of nuclear matrix proteins binding to matrix attachment regions (Mars) in prostate cancer cell differentiation. *PloS one*. 7:e40617.

Baumann, C., and R. De La Fuente. 2009. ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells. *Chromosoma*. 118:209-222.

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

Berezney, R., and K.W. Jeon. 1995. Nuclear Matrix: Structural and Functional Organization. Academic Press.

Blewitt, M.E., A.V. Gendrel, Z. Pang, D.B. Sparrow, N. Whitelaw, J.M. Craig, A. Apedaile, D.J. Hilton, S.L. Dunwoodie, N. Brockdorff, G.F. Kay, and E. Whitelaw. 2008. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nature genetics*. 40:663-669.

Boggs, B.A., P. Cheung, E. Heard, D.L. Spector, A.C. Chinault, and C.D. Allis. 2002. Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nature genetics*. 30:73-76.

Boisvert, F.M., S. van Koningsbruggen, J. Navascues, and A.I. Lamond. 2007. The multifunctional nucleolus. *Nature reviews. Molecular cell biology*. 8:574-585.

Britten, R.J., and E.H. Davidson. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol*. 46:111-138.

Britten, R.J., D.E. Graham, and B.R. Neufeld. 1974. Analysis of repeating DNA sequences by reassociation. *Methods Enzymol*. 29:363-418.

Britten, R.J., and D.E. Kohne. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science*. 161:529-540.

Brockdorff, N., A. Ashworth, G.F. Kay, V.M. McCabe, D.P. Norris, P.J. Cooper, S. Swift, and S. Rastan. 1992. The product of the mouse Xist gene is a 15 kb inactive

X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*. 71:515-526.

Brown, C.J., L. Carrel, and H.F. Willard. 1997. Expression of genes from the human active and inactive X chromosomes. *Am J Hum Genet*. 60:1333-1343.

Brown, C.J., B.D. Hendrich, J.L. Rupert, R.G. Lafreniere, Y. Xing, J. Lawrence, and H.F. Willard. 1992. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*. 71:527-542.

Bynum, J.W., and E. Volkin. 1980. Chromatin-associated RNA: differential extraction and characterization. *Biochimica et biophysica acta*. 607:304-318.

Byron, M., L.L. Hall, and J.B. Lawrence. 2013. A multifaceted FISH approach to study endogenous RNAs and DNAs in native nuclear and cell structures. *Curr Protoc Hum Genet*. Chapter 4:Unit 4 15.

Capco, D.G., K.M. Wan, and S. Penman. 1982. The nuclear matrix: three-dimensional architecture and protein composition. *Cell*. 29:847-858.

Capy, P., G. Gasperi, C. Biemont, and C. Bazin. 2000. Stress and transposable elements: co-evolution or useful parasites? *Heredity (Edinb)*. 85 (Pt 2):101-106.

Carone, D.M., and J.B. Lawrence. 2013. Heterochromatin instability in cancer: from the Barr body to satellites and the nuclear periphery. *Seminars in cancer biology*. 23:99-108.

Carone, D.M., C. Zhang, L.E. Hall, C. Obergfell, B.R. Carone, M.J. O'Neill, and R.J. O'Neill. 2013. Hypermorphic expression of centromeric retroelement-encoded small RNAs impairs CENP-A loading. *Chromosome Res*. 21:49-62.

Carrel, L., and H.F. Willard. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*. 434:400-404.

Carter, K.C., K.L. Taneja, and J.B. Lawrence. 1991. Discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus. *The Journal of cell biology*. 115:1191-1202.

Cerase, A., G. Pintacuda, A. Tattermusch, and P. Avner. 2015. Xist localization and function: new insights from multiple levels. *Genome biology*. 16:166.

Chaumeil, J., P. Le Baccon, A. Wutz, and E. Heard. 2006. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev*. 20:2223-2237.

Chery, J., and E. Larschan. 2014. X-marks the spot: X-chromosome identification during dosage compensation. *Biochimica et biophysica acta*. 1839:234-240.

Chow, J.C., C. Ciaudo, M.J. Fazzari, N. Mise, N. Servant, J.L. Glass, M. Attreed, P. Avner, A. Wutz, E. Barillot, J.M. Greally, O. Voinnet, and E. Heard. 2010. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell*. 141:956-969.

Chow, J.C., L.L. Hall, S.E. Baldry, N.P. Thorogood, J.B. Lawrence, and C.J. Brown. 2007. Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. *Proceedings of the National Academy of Sciences of the United States of America*. 104:10104-10109.

Chow, J.C., Z. Yen, S.M. Ziesche, and C.J. Brown. 2005. Silencing of the mammalian X chromosome. *Annual review of genomics and human genetics*. 6:69-92.

Chu, C., K. Qu, F.L. Zhong, S.E. Artandi, and H.Y. Chang. 2011. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell*. 44:667-678.

Chu, C., R.C. Spitale, and H.Y. Chang. 2015a. Technologies to probe functions and mechanisms of long noncoding RNAs. *Nature structural & molecular biology*. 22:29-35.

Chu, C., Q.C. Zhang, S.T. da Rocha, R.A. Flynn, M. Bharadwaj, J.M. Calabrese, T. Magnuson, E. Heard, and H.Y. Chang. 2015b. Systematic discovery of Xist RNA binding proteins. *Cell*. 161:404-416.

Chureau, C., M. Prissette, A. Bourdet, V. Barbe, L. Cattolico, L. Jones, A. Eggen, P. Avner, and L. Duret. 2002. Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome research*. 12:894-908.

Clark, M.B., R.L. Johnston, M. Inostroza-Ponta, A.H. Fox, E. Fortini, P. Moscato, M.E. Dinger, and J.S. Mattick. 2012. Genome-wide analysis of long noncoding RNA stability. *Genome research*. 22:885-898.

Clemson, C.M., J.C. Chow, C.J. Brown, and J.B. Lawrence. 1998. Stabilization and localization of Xist RNA are controlled by separate mechanisms and are not sufficient for X inactivation. *The Journal of cell biology*. 142:13-23.

Clemson, C.M., L.L. Hall, M. Byron, J. McNeil, and J.B. Lawrence. 2006. The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proceedings of the National Academy of Sciences of the United States of America*. 103:7688-7693.

Clemson, C.M., J.A. McNeil, H.F. Willard, and J.B. Lawrence. 1996. XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear chromosome structure. *Journal of Cell Biology*. 132:259-275.

Cohen, D.E., L.S. Davidow, J.A. Erwin, N. Xu, D. Warshawsky, and J.T. Lee. 2007. The DXPas34 repeat regulates random and imprinted X inactivation. *Developmental cell*. 12:57-71.

Consortium, E.P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489:57-74.

Costanzi, C., and J.R. Pehrson. 1998. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature*. 393:599-601.

Cremer, M., K. Kupper, B. Wagler, L. Wizelman, J. von Hase, Y. Weiland, L. Kreja, J. Diebold, M.R. Speicher, and T. Cremer. 2003. Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. *The Journal of cell biology*. 162:809-820.

Cremer, T., M. Cremer, S. Dietzel, S. Muller, I. Solovei, and S. Fakan. 2006. Chromosome territories--a functional nuclear landscape. *Curr Opin Cell Biol*. 18:307-316.

Croft, J.A., J.M. Bridger, S. Boyle, P. Perry, P. Teague, and W.A. Bickmore. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of cell biology*. 145:1119-1131.

Dang-Nguyen, T.Q., and M.E. Torres-Padilla. 2015. How cells build totipotency and pluripotency: nuclear, chromatin and transcriptional architecture. *Curr Opin Cell Biol*. 34:9-15.

Darnell, R.B. 2010. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA*. 1:266-286.

Dechat, T., K. Pfleghaar, K. Sengupta, T. Shimi, D.K. Shumaker, L. Solimando, and R.D. Goldman. 2008. Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev*. 22:832-853.

Dittmer, T.A., and T. Misteli. 2011. The lamin protein family. *Genome biology*. 12:222.

Doyon, J.B., B. Zeitler, J. Cheng, A.T. Cheng, J.M. Cherone, Y. Santiago, A.H. Lee, T.D. Vo, Y. Doyon, J.C. Miller, D.E. Paschon, L. Zhang, E.J. Rebar, P.D. Gregory, F.D. Urnov, and D.G. Drubin. 2011. Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nat Cell Biol*. 13:331-337.

Dundr, M. 2012. Nuclear bodies: multifunctional companions of the genome. *Curr Opin Cell Biol*. 24:415-422.

Dworetzky, S.I., E.G. Fey, S. Penman, J.B. Lian, J.L. Stein, and G.S. Stein. 1990. Progressive changes in the protein composition of the nuclear matrix during rat osteoblast differentiation. *Proceedings of the National Academy of Sciences of the United States of America*. 87:4605-4609.

Eils, R., S. Dietzel, E. Bertin, E. Schrock, M.R. Speicher, T. Ried, M. Robert-Nicoud, C. Cremer, and T. Cremer. 1996. Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. *The Journal of cell biology*. 135:1427-1440.

Engreitz, J.M., A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E.S. Lander, K. Plath, and M. Guttman. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. 341:1237973.

Espinoza, C.A., T.A. Allen, A.R. Hieb, J.F. Kugel, and J.A. Goodrich. 2004. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nature structural & molecular biology*. 11:822-829.

Fackelmayer, F.O., K. Dahm, A. Renz, U. Ramsperger, and A. Richter. 1994. Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro. *European journal of biochemistry / FEBS*. 221:749-757.

Faulkner, G.J., and P. Carninci. 2009. Altruistic functions for selfish DNA. *Cell cycle*. 8:2895-2900.

Fedoroff, N., P.K. Wellauer, and R. Wall. 1977. Intermolecular duplexes in heterogeneous nuclear RNA from HeLa cells. *Cell*. 10:597-610.

Fey, E.G., D.A. Ornelles, and S. Penman. 1986. Association of RNA with the cytoskeleton and the nuclear matrix. *J Cell Sci Suppl*. 5:99-119.

Fey, E.G., and S. Penman. 1988. Nuclear matrix proteins reflect cell type of origin in cultured human cells. *Proceedings of the National Academy of Sciences of the United States of America*. 85:121-125.

Freedman, L.P., I.M. Cockburn, and T.S. Simcoe. 2015. The Economics of Reproducibility in Preclinical Research. *PLoS Biol*. 13:e1002165.

Galupa, R., and E. Heard. 2015. X-chromosome inactivation: new insights into cis and trans regulation. *Current opinion in genetics & development*. 31:57-66.

Ganesan, S., D.P. Silver, R.A. Greenberg, D. Avni, R. Drapkin, A. Miron, S.C. Mok, V. Randrianarison, S. Brodie, J. Salstrom, T.P. Rasmussen, A. Klimke, C. Marrese, Y. Marahrens, C.-X. Deng, J. Feunteun, and D.M. Livingston. 2002. BRCA1 Supports XIST RNA Concentration on the Inactive X Chromosome. *Cell*. 111:393-405.

Gardner, E.J., Z.F. Nizami, C.C. Talbot, Jr., and J.G. Gall. 2012. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of Xenopus tropicalis. *Genes Dev*. 26:2550-2559.

Gendrel, A.V., and E. Heard. 2014. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annual review of cell and developmental biology*. 30:561-580.

Germain, P.L., E. Ratti, and F. Boem. 2014. Junk or functional DNA? ENCODE and the function controversy. *Biol Philos*. 29:807-831.

Getzenberg, R.H., K.J. Pienta, E.Y. Huang, and D.S. Coffey. 1991. Identification of nuclear matrix proteins in the cancer and normal rat prostate. *Cancer research*. 51:6514-6520.

Gilbert, N., S. Gilchrist, and W.A. Bickmore. 2005. Chromatin organization in the mammalian nucleus. *Int Rev Cytol*. 242:283-336.

Gohring, F., and F.O. Fackelmayer. 1997. The scaffold/matrix attachment region binding protein hnRNP-U (SAF-A) is directly bound to chromosomal DNA in vivo: a chemical cross-linking study. *Biochemistry*. 36:8276-8283.

Goldman, R.D., Y. Gruenbaum, R.D. Moir, D.K. Shumaker, and T.P. Spann. 2002. Nuclear lamins: building blocks of nuclear architecture. *Genes Dev*. 16:533-547.

Gomez, J.A., O.L. Wapinski, Y.W. Yang, J.F. Bureau, S. Gopinath, D.M. Monack, H.Y. Chang, M. Brahic, and K. Kirkegaard. 2013. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell*. 152:743-754.

Goodier, J.L., and H.H. Kazazian, Jr. 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*. 135:23-35.

Grant, J., S.K. Mahadevaiah, P. Khil, M.N. Sangrithi, H. Royo, J. Duckworth, J.R. McCarrey, J.L. VandeBerg, M.B. Renfree, W. Taylor, G. Elgar, R.D. Camerini-Otero, M.J. Gilchrist, and J.M. Turner. 2012. Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature*. 487:254-258.

Graur, D., Y. Zheng, N. Price, R.B. Azevedo, R.A. Zufall, and E. Elhaik. 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol*. 5:578-590.

Gregory, T.R., and P.D. Hebert. 1999. The modulation of DNA content: proximate causes and ultimate consequences. *Genome research*. 9:317-324.

Gupta, R.A., N. Shah, K.C. Wang, J. Kim, H.M. Horlings, D.J. Wong, M.C. Tsai, T. Hung, P. Argani, J.L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R.B. West, M.J. van de Vijver, S. Sukumar, and H.Y. Chang. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 464:1071-1076.

Guschin, D.Y., A.J. Waite, G.E. Katibah, J.C. Miller, M.C. Holmes, and E.J. Rebar. 2010. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol*. 649:247-256.

Guttman, M., J. Donaghey, B.W. Carey, M. Garber, J.K. Grenier, G. Munson, G. Young, A.B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J.L. Rinn, D.E. Root, and E.S. Lander. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 477:295-300.

Hadjiargyrou, M., and N. Delihas. 2013. The intertwining of transposable elements and non-coding RNAs. *Int J Mol Sci*. 14:13307-13328.

Hall, L.L., M. Byron, G. Pageau, and J.B. Lawrence. 2009. AURKB-mediated effects on chromatin regulate binding versus release of XIST RNA to the inactive chromosome. *The Journal of cell biology*. 186:491-507.

Hall, L.L., M. Byron, K. Sakai, L. Carrel, H.F. Willard, and J.B. Lawrence. 2002a. An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. *Proceedings of the National Academy of Sciences of the United States of America*. 99:8677-8682.

Hall, L.L., D.M. Carone, A.V. Gomez, H.J. Kolpa, M. Byron, N. Mehta, F.O. Fackelmayer, and J.B. Lawrence. 2014. Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell*. 156:907-919.

Hall, L.L., C.M. Clemson, M. Byron, K. Wydner, and J.B. Lawrence. 2002b. Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. *Human molecular genetics*. 11:3157-3165.

Hall, L.L., and J.B. Lawrence. 2010. XIST RNA and architecture of the inactive X chromosome: implications for the repeat genome. *Cold Spring Harb Symp Quant Biol*. 75:345-356.

Hall, L.L., K.P. Smith, M. Byron, and J.B. Lawrence. 2006. Molecular anatomy of a speckle. *Anat Rec A Discov Mol Cell Evol Biol*. 288:664-675.

Hasegawa, Y., N. Brockdorff, S. Kawano, K. Tsutui, K. Tsutui, and S. Nakagawa. 2010. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Developmental cell*. 19:469-476.

Hasegawa, Y., and S. Nakagawa. 2011. Revisiting the function of nuclear scaffold/matrix binding proteins in X chromosome inactivation. *RNA biology*. 8:735-739.

Hathaway, N.A., O. Bell, C. Hodges, E.L. Miller, D.S. Neel, and G.R. Crabtree. 2012. Dynamics and memory of heterochromatin in living cells. *Cell*. 149:1447-1460.

Hauf, S., R.W. Cole, S. LaTerra, C. Zimmer, G. Schnapp, R. Walter, A. Heckel, J. van Meel, C.L. Rieder, and J.M. Peters. 2003. The small molecule Hesperadin reveals

a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *The Journal of cell biology*. 161:281-294.

He, D.C., J.A. Nickerson, and S. Penman. 1990. Core filaments of the nuclear matrix. *The Journal of cell biology*. 110:569-580.

Heard, E. 2005. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Current opinion in genetics & development*. 15:482-489.

Heard, E., and C.M. Disteche. 2006. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev*. 20:1848-1867.

Heard, E., F. Mongelard, D. Arnaud, C. Chureau, C. Vourc'h, and P. Avner. 1999. Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*. 96:6841-6846.

Hegyi, K., and G. Mehes. 2012. Mitotic failures in cancer: Aurora B kinase and its potential role in the development of aneuploidy. *Pathol Oncol Res*. 18:761-769.

Helbig, R., and F.O. Fackelmayer. 2003. Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma*. 112:173-182.

Herman, R., L. Weymouth, and S. Penman. 1978. Heterogeneous nuclear RNA-protein fibers in chromatin-depleted nuclei. *The Journal of cell biology*. 78:663-674.

Huelga, S.C., A.Q. Vu, J.D. Arnold, T.Y. Liang, P.P. Liu, B.Y. Yan, J.P. Donohue, L. Shiue, S. Hoon, S. Brenner, M. Ares, Jr., and G.W. Yeo. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell reports*. 1:167-178.

Jeon, Y., and J.T. Lee. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell*. 146:119-133.

Jeppesen, P., and B.M. Turner. 1993. The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell*. 74:281-289.

Jiang, J., Y. Jing, G.J. Cost, J.C. Chiang, H.J. Kolpa, A.M. Cotton, D.M. Carone, B.R. Carone, D.A. Shivak, D.Y. Guschin, J.R. Pearl, E.J. Rebar, M. Byron, P.D. Gregory, C.J. Brown, F.D. Urnov, L.L. Hall, and J.B. Lawrence. 2013. Translating dosage compensation to trisomy 21. *Nature*. 500:296-300.

Johnson, C., D. Primorac, M. McKinstry, J. McNeil, D. Rowe, and J.B. Lawrence. 2000. Tracking COL1A1 RNA in osteogenesis imperfecta. splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *The Journal of cell biology*. 150:417-432.

Johnson, C.V., R.H. Singer, and J.B. Lawrence. 1991. Fluorescent detection of nuclear RNA and DNA: implications for genome organization. *Methods in cell biology*. 35:73-99.

Jurica, M.S., L.J. Licklider, S.R. Gygi, N. Grigorieff, and M.J. Moore. 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*. 8:426-439.

Kaiser, J. 2015. Study claims $28 billion a year spent on irreproducible biomedical research. Vol. 2016. AAAS, Science.

Kapusta, A., Z. Kronenberg, V.J. Lynch, X. Zhuo, L. Ramsay, G. Bourque, M. Yandell, and C. Feschotte. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS genetics*. 9:e1003470.

Keene, J.D., J.M. Komisarow, and M.B. Friedersdorf. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc*. 1:302-307.

Kibar, Y., S. Goktas, S. Kilic, H. Yaman, O. Onguru, and A.F. Peker. 2006. Prognostic value of cytology, nuclear matrix protein 22 (NMP22) test, and urinary bladder cancer II (UBC II) test in early recurrent transitional cell carcinoma of the bladder. *Ann Clin Lab Sci*. 36:31-38.

Kiledjian, M., and G. Dreyfuss. 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *The EMBO journal*. 11:2655-2664.

Konety, B.R., and R.H. Getzenberg. 1999. Nuclear structural proteins as biomarkers of cancer. *Journal of cellular biochemistry*. Suppl 32-33:183-191.

Konety, B.R., T.S. Nguyen, R. Dhir, R.S. Day, M.J. Becich, W.M. Stadler, and R.H. Getzenberg. 2000. Detection of bladder cancer using a novel nuclear matrix protein, BLCA-4. *Clin Cancer Res*. 6:2618-2625.

Kupper, K., A. Kolbl, D. Biener, S. Dittrich, J. von Hase, T. Thormeyer, H. Fiegler, N.P. Carter, M.R. Speicher, T. Cremer, and M. Cremer. 2007. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma*. 116:285-306.

Lamond, A.I., and D.L. Spector. 2003. Nuclear speckles: a model for nuclear organelles. *Nature reviews. Molecular cell biology*. 4:605-612.

Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al. 2001. Initial sequencing and analysis of the human genome. *Nature*. 409:860-921.

Lawrence, J.B., R.H. Singer, and L.M. Marselle. 1989. Highly localized tracks of specific transcripts within interphase nuclei visualized by in situ hybridization. *Cell*. 57:493-502.

Lee, J.T., and N. Lu. 1999. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell*. 99:47-57.

Lee, J.T., W.M. Strauss, J.A. Dausman, and R. Jaenisch. 1996. A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell*. 86:83-94.

Li, T., J. Spearow, C.M. Rubin, and C.W. Schmid. 1999. Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. *Gene*. 239:367-372.

Lichter, P., T. Cremer, J. Borden, L. Manuelidis, and D.C. Ward. 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet*. 80:224-234.

Liu, W.M., W.M. Chu, P.V. Choudary, and C.W. Schmid. 1995. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic acids research*. 23:1758-1765.

Lyon, M.F. 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature*. 190:372-373.

Lyon, M.F. 1998. X-chromosome inactivation: a repeat hypothesis. *Cytogenetics and cell genetics*. 80:133-137.

Ma, N., S. Matsunaga, A. Morimoto, G. Sakashita, T. Urano, S. Uchiyama, and K. Fukui. 2011. The nuclear scaffold protein SAF-A is required for kinetochore-microtubule attachment and contributes to the targeting of Aurora-A to mitotic spindles. *Journal of cell science*. 124:394-404.

Malhas, A., C.F. Lee, R. Sanders, N.J. Saunders, and D.J. Vaux. 2007. Defects in lamin B1 expression or processing affect interphase chromosome position and gene expression. *The Journal of cell biology*. 176:593-603.

Mariner, P.D., R.D. Walters, C.A. Espinoza, L.F. Drullinger, S.D. Wagner, J.F. Kugel, and J.A. Goodrich. 2008. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Molecular cell*. 29:499-509.

Martelli, A.M., E. Falcieri, M. Zweyer, R. Bortul, G. Tabellini, A. Cappellini, L. Cocco, and L. Manzoli. 2002. The controversial nuclear matrix: a balanced point of view. *Histol Histopathol*. 17:1193-1205.

McHugh, C.A., C.K. Chen, A. Chow, C.F. Surka, C. Tran, P. McDonel, A. Pandya-Jones, M. Blanco, C. Burghard, A. Moradian, M.J. Sweredoski, A.A. Shishkin, J. Su, E.S. Lander, S. Hess, K. Plath, and M. Guttman. 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*. 521:232-236.

McHugh, C.A., P. Russell, and M. Guttman. 2014. Methods for comprehensive experimental identification of RNA-protein interactions. *Genome biology*. 15:203.

McNeil, J.A., K.P. Smith, L.L. Hall, and J.B. Lawrence. 2006. Word frequency analysis reveals enrichment of dinucleotide repeats on the human X chromosome and [GATA]n in the X escape region. *Genome research*. 16:477-484.

Meaburn, K.J., E. Cabuy, G. Bonne, N. Levy, G.E. Morris, G. Novelli, I.R. Kill, and J.M. Bridger. 2007. Primary laminopathy fibroblasts display altered genome organization and apoptosis. *Aging Cell*. 6:139-153.

Meshorer, E., and T. Misteli. 2006. Chromatin in pluripotent embryonic stem cells and differentiation. *Nature reviews. Molecular cell biology*. 7:540-546.

Meshorer, E., D. Yellajoshula, E. George, P.J. Scambler, D.T. Brown, and T. Misteli. 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental cell*. 10:105-116.

Migeon, B.R. 2014. Females are Mosaics: X inactivation and Sex Differences in Disease.

Migeon, B.R., A.K. Chowdhury, J.A. Dunston, and I. McIntosh. 2001a. Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation. *Am J Hum Genet*. 69:951-960.

Migeon, B.R., E. Kazi, C. Haisley-Royster, J. Hu, R. Reeves, L. Call, A. Lawler, C.S. Moore, H. Morrison, and P. Jeppesen. 1999. Human X inactivation center induces random X chromosome inactivation in male transgenic mice. *Genomics*. 59:113-121.

Migeon, B.R., C.H. Lee, A.K. Chowdhury, and H. Carpenter. 2002. Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. *Am J Hum Genet*. 71:286-293.

Migeon, B.R., H. Winter, E. Kazi, A.K. Chowdhury, A. Hughes, C. Haisley-Royster, H. Morrison, and P. Jeppesen. 2001b. Low-copy-number human transgene is recognized as an X inactivation center in mouse ES cells, but fails to induce cis-inactivation in chimeric mice. *Genomics*. 71:156-162.

Mili, S., and J.A. Steitz. 2004. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA*. 10:1692-1694.

Miller, J.C., M.C. Holmes, J. Wang, D.Y. Guschin, Y.L. Lee, I. Rupniewski, C.M. Beausejour, A.J. Waite, N.S. Wang, K.A. Kim, P.D. Gregory, C.O. Pabo, and E.J. Rebar. 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*. 25:778-785.

Minajigi, A., J.E. Froberg, C. Wei, H. Sunwoo, B. Kesner, D. Colognori, D. Lessing, B. Payer, M. Boukhali, W. Haas, and J.T. Lee. 2015. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science*.

Moehle, E.A., J.M. Rock, Y.L. Lee, Y. Jouvenot, R.C. DeKelver, P.D. Gregory, F.D. Urnov, and M.C. Holmes. 2007. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America*. 104:3055-3060.

Nakagawa, S., and K.V. Prasanth. 2011. eXIST with matrix-associated proteins. *Trends in cell biology*. 21:321-327.

Nickerson, J. 2001. Experimental observations of a nuclear matrix. *Journal of cell science*. 114:463-474.

Nickerson, J.A., G. Krochmalnic, K.M. Wan, and S. Penman. 1989. Chromatin architecture and nuclear RNA. *Proceedings of the National Academy of Sciences of the United States of America*. 86:177-181.

Nickerson, J.A., G. Krockmalnic, K.M. Wan, and S. Penman. 1997. The nuclear matrix revealed by eluting chromatin from a cross-linked nucleus. *Proceedings of the National Academy of Sciences of the United States of America*. 94:4446-4450.

Ogawa, Y., and J.T. Lee. 2003. Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. *Molecular cell*. 11:731-743.

Ohno, S. 1972. So much "junk" DNA in our genome. *Brookhaven Symp Biol*. 23:366-370.

Ota, T., S. Suto, H. Katayama, Z.B. Han, F. Suzuki, M. Maeda, M. Tanino, Y. Terada, and M. Tatsuka. 2002. Increased mitotic phosphorylation of histone H3 attributable to AIM-1/Aurora-B overexpression contributes to chromosome number instability. *Cancer research*. 62:5168-5177.

Pageau, G.J., L.L. Hall, S. Ganesan, D.M. Livingston, and J.B. Lawrence. 2007a. The disappearing Barr body in breast and ovarian cancers. *Nature reviews. Cancer*. 7:628-633.

Pageau, G.J., L.L. Hall, and J.B. Lawrence. 2007b. BRCA1 does not paint the inactive X to localize XIST RNA but may contribute to broad changes in cancer that impact XIST and Xi heterochromatin. *Journal of cellular biochemistry*. 100:835-850.

Pageau, G.J., and J.B. Lawrence. 2006. BRCA1 foci in normal S-phase nuclei are linked to interphase centromeres and replication of pericentric heterochromatin. *The Journal of cell biology*. 175:693-701.

Pagel, M., and R.A. Johnstone. 1992. Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox. *Proc Biol Sci*. 249:119-124.

Pederson, T. 1998. Thinking about a nuclear matrix. *J Mol Biol*. 277:147-159.

Pederson, T. 2000. Half a century of "the nuclear matrix". *Molecular biology of the cell*. 11:799-805.

Phair, R.D., and T. Misteli. 2000. High mobility of proteins in the mammalian cell nucleus. *Nature*. 404:604-609.

Plath, K., J. Fang, S.K. Mlynarczyk-Evans, R. Cao, K.A. Worringer, H. Wang, C.C. de la Cruz, A.P. Otte, B. Panning, and Y. Zhang. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science*. 300:131-135.

Pollex, T., and E. Heard. 2012. Recent advances in X-chromosome inactivation research. *Curr Opin Cell Biol*. 24:825-832.

Pullirsch, D., R. Hartel, H. Kishimoto, M. Leeb, G. Steiner, and A. Wutz. 2010. The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. *Development*. 137:935-943.

Raha, D., Z. Wang, Z. Moqtaderi, L. Wu, G. Zhong, M. Gerstein, K. Struhl, and M. Snyder. 2010. Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proceedings of the National Academy of Sciences of the United States of America*. 107:3639-3644.

Razin, S.V., O.V. Iarovaia, and Y.S. Vassetzky. 2014. A requiem to the nuclear matrix: from a controversial concept to 3D organization of the nucleus. *Chromosoma*. 123:217-224.

Riley, K.J., and J.A. Steitz. 2013. The "Observer Effect" in genome-wide surveys of protein-RNA interactions. *Molecular cell*. 49:601-604.

Riley, K.J., T.A. Yario, and J.A. Steitz. 2012. Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA*. 18:1581-1585.

Rinn, J.L., M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, L.H. Goodnough, J.A. Helms, P.J. Farnham, E. Segal, and H.Y. Chang. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 129:1311-1323.

Romig, H., F.O. Fackelmayer, A. Renz, U. Ramsperger, and A. Richter. 1992. Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *The EMBO journal*. 11:3431-3440.

Saitoh, N., C.S. Spahr, S.D. Patterson, P. Bubulya, A.F. Neuwald, and D.L. Spector. 2004. Proteomic analysis of interchromatin granule clusters. *Molecular biology of the cell*. 15:3876-3890.

Santoni, F.A., J. Guerra, and J. Luban. 2012. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology*. 9:111.

Sarma, K., C. Cifuentes-Rojas, A. Ergun, A. Del Rosario, Y. Jeon, F. White, R. Sadreyev, and J.T. Lee. 2014. ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell*. 159:869-883.

Savarese, F., K. Flahndorfer, R. Jaenisch, M. Busslinger, and A. Wutz. 2006. Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation. *Molecular and cellular biology*. 26:7167-7177.

Shopland, L.S., and J.B. Lawrence. 2000. Seeking common ground in nuclear complexity. *The Journal of cell biology*. 150:F1-4.

Sigova, A.A., B.J. Abraham, X. Ji, B. Molinie, N.M. Hannett, Y.E. Guo, M. Jangi, C.C. Giallourakis, P.A. Sharp, and R.A. Young. 2015. Transcription factor trapping by RNA in gene regulatory elements. *Science*. 350:978-981.

Smith, K.P., M. Byron, C. Johnson, Y. Xing, and J.B. Lawrence. 2007. Defining early steps in mRNA transport: mutant mRNA in myotonic dystrophy type I is blocked at entry into SC-35 domains. *The Journal of cell biology*. 178:951-964.

Spector, D.L., and A.I. Lamond. 2011. Nuclear speckles. *Cold Spring Harbor perspectives in biology*. 3.

Stavropoulos, N., N. Lu, and J.T. Lee. 2001. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proceedings of the National Academy of Sciences of the United States of America*. 98:10232-10237.

Sun, S., B.C. Del Rosario, A. Szanto, Y. Ogawa, Y. Jeon, and J.T. Lee. 2013. Jpx RNA activates Xist by evicting CTCF. *Cell*. 153:1537-1551.

Surralles, J., and A.T. Natarajan. 1998. Position effect of translocations involving the inactive X chromosome: physical linkage to XIC/XIST does not lead to long-

range de novo inactivation in human differentiated cells. *Cytogenetics and cell genetics*. 82:58-66.

Tam, R., L. Shopland, C. Johnson, J. McNeil, and J. Lawrence. 2002. Applications of RNA FISH for visualizing gene expression and nuclear architecture. *FISH (Fluorescence In Situ Hybridization). BG Beatty, S. Mai, and JA Squire, editors. Oxford University Press, Oxford*:93-110.

Tam, R., K.P. Smith, and J.B. Lawrence. 2004. The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. *The Journal of cell biology*. 167:269-279.

Tanzi, R.E., and L. Bertram. 2005. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell*. 120:545-555.

Tattermusch, A., and N. Brockdorff. 2011. A scaffold for X chromosome inactivation. *Hum Genet*. 130:247-253.

Thandapani, P., T.R. O'Connor, T.L. Bailey, and S. Richard. 2013. Defining the RGG/RG motif. *Molecular cell*. 50:613-623.

Thomas, C.A., Jr. 1971. The genetic organization of chromosomes. *Annu Rev Genet*. 5:237-256.

Timp, W., and A.P. Feinberg. 2013. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nature reviews. Cancer*. 13:497-510.

Ting, D.T., D. Lipson, S. Paul, B.W. Brannigan, S. Akhavanfard, E.J. Coffman, G. Contino, V. Deshpande, A.J. Iafrate, S. Letovsky, M.N. Rivera, N. Bardeesy, S. Maheswaran, and D.A. Haber. 2011. Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science*. 331:593-596.

Tripathi, V., Z. Shen, A. Chakraborty, S. Giri, S.M. Freier, X. Wu, Y. Zhang, M. Gorospe, S.G. Prasanth, A. Lal, and K.V. Prasanth. 2013. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS genetics*. 9:e1003368.

Ule, J., K. Jensen, A. Mele, and R.B. Darnell. 2005. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods*. 37:376-386.

Urnov, F.D., J.C. Miller, Y.L. Lee, C.M. Beausejour, J.M. Rock, S. Augustus, A.C. Jamieson, M.H. Porteus, P.D. Gregory, and M.C. Holmes. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*. 435:646-651.

Urnov, F.D., E.J. Rebar, M.C. Holmes, H.S. Zhang, and P.D. Gregory. 2010. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*. 11:636-646.

Venters, B.J., and B.F. Pugh. 2013. Genomic organization of human transcription initiation complexes. *Nature*. 502:53-58.

Vizlin-Hodzic, D., R. Runnberg, J. Ryme, S. Simonsson, and T. Simonsson. 2011. SAF-A forms a complex with BRG1 and both components are required for RNA polymerase II mediated transcription. *PloS one*. 6:e28049.

Vogelstein, B., D.M. Pardoll, and D.S. Coffey. 1980. Supercoiled loops and eucaryotic DNA replicaton. *Cell*. 22:79-85.

Webb, R.L., and M.P. Murphy. 2012. beta-Secretases, Alzheimer's Disease, and Down Syndrome. *Current gerontology and geriatrics research*. 2012:362839.

White, W.M., H.F. Willard, D.L. Van Dyke, and D.J. Wolff. 1998. The spreading of X inactivation into autosomal material of an x;autosome translocation: evidence for a difference between autosomal and X-chromosomal DNA. *Am J Hum Genet*. 63:20-28.

Whitworth, D.J., and A.J. Pask. 2016. The X factor: X chromosome dosage compensation in the evolutionarily divergent monotremes and marsupials. *Semin Cell Dev Biol*.

Worman, H.J., and G. Bonne. 2007. "Laminopathies": a wide spectrum of human diseases. *Exp Cell Res*. 313:2121-2133.

Wutz, A. 2011. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nature reviews. Genetics*. 12:542-553.

Wutz, A., and J. Gribnau. 2007. X inactivation Xplained. *Current opinion in genetics & development*. 17:387-393.

Wutz, A., and R. Jaenisch. 2000. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell*. 5:695-705.

Wutz, A., T.P. Rasmussen, and R. Jaenisch. 2002. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature genetics*. 30:167-174.

Xiao, C., J.A. Sharp, M. Kawahara, A.R. Davalos, M.J. Difilippantonio, Y. Hu, W. Li, L. Cao, K. Buetow, T. Ried, B.P. Chadwick, C.X. Deng, and B. Panning. 2007. The XIST noncoding RNA functions independently of BRCA1 in X inactivation. *Cell*. 128:977-989.

Xiao, R., P. Tang, B. Yang, J. Huang, Y. Zhou, C. Shao, H. Li, H. Sun, Y. Zhang, and X.D. Fu. 2012. Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Molecular cell*. 45:656-668.

Xing, Y.G., and J.B. Lawrence. 1991. Preservation of specific RNA distribution within the chromatin-depleted nuclear substructure demonstrated by in situ hybridization coupled with biochemical fractionation. *The Journal of cell biology*. 112:1055-1063.

Zhao, J., B.K. Sun, J.A. Erwin, J.J. Song, and J.T. Lee. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*. 322:750-756.

Zhou, Z., L.J. Licklider, S.P. Gygi, and R. Reed. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature*. 419:182-185.

Zink, D., A.H. Fischer, and J.A. Nickerson. 2004. Nuclear structure in cancer cells. *Nature reviews. Cancer*. 4:677-687.