XIST RNA and Architecture of the Inactive X Chromosome
Implications for the Repeat Genome

L.L. HALL AND J.B. LAWRENCE
Department of Cell Biology, University of Massachusetts Medical School, North Worcester, Massachusetts 01655
Correspondence: Jeanne.Lawrence@umassmed.edu

XIST RNA paints and induces silencing of one X chromosome in mammalian female cells, providing a powerful model to investigate long-range chromosomal regulation. This chapter focuses on events downstream from the spread of XIST RNA across the interphase chromosome, to consider how this large noncoding RNA interacts with and silences a whole chromosome. Several lines of evidence are summarized that point to the involvement of repeat sequences in different aspects of the X-inactivation process. Although the “repeat genome” comprises close to half of the human genome, the potential for abundant repeats to contribute to genome regulation has been largely overlooked and may be underestimated. X inactivation has the potential to reveal roles of interspersed and other repeats in the genome. For example, evidence indicates that XIST RNA acts at the architectural level of the whole chromosome to induce formation of a silent core enriched for nongenic and repetitive (Cot-1) DNA, which corresponds to the DAPI-dense Barr body. Expression of repeat RNAs may contribute to chromosome remodeling, and evidence suggests that other types of repeat elements may be involved in escape from X inactivation. Despite great progress in decoding the rest of the genome, we suggest that the repeat genome may contain meaningful but complex language that remains to be better studied and understood.

The inactivation of one X chromosome in mammalian females is a powerful model for epigenetic regulation in early development, involving formation of facultative heterochromatin that sweeps across an entire chromosome. Remarkably, a large noncoding RNA from the X-linked human XIST/mouse Xist gene (Brockdorff et al. 1992; Brown et al. 1992) “paints” its parent chromosome (Brockdorff et al. 1992; Brown et al. 1992; Clemson et al. 1996; Chow et al. 2007) and induces a silencing cascade throughout the chromosome territory (for review, see Hall and Lawrence 2003; Heard and Disteche 2006). A central question has now become: How does XIST RNA localize across its chromosome and impact its architecture and expression? Importantly, XIST RNA is strictly localized to its parent chromosome in cis, yet how it binds and what conforms susceptibility (or resistance) of particular genes to silencing remain poorly understood.

In recent years, substantial progress has been made in defining the series of heterochromatin chromatin modifications that spread across the chromosome shortly after XIST RNA first paints the chromosome (for review, see Heard 2005). Evidence suggests that XIST RNA can “recruit” Polycomb group proteins that induce heterochromatin modifications, such as EZH2, which methylates histone H3K27 (Schoeftner et al. 2006), or RING-1, which ubiquinates H2A (Fang et al. 2004). However, it is increasingly appreciated that genome regulation involves not only “local” changes, at the level of chromatin modifications or transcription factor/repressor binding, but also higher-order changes to nuclear and chromosome/chromatin architecture. In fact, the inactive X chromosome exemplifies a now well-established principle of nuclear structure, that the nuclear interior is highly “compartmentalized” into non-membrane-bound compartments devoted to particular functions (Moen et al. 1995; Cremer and Cremer 2001; Chubb and Bickmore 2003). The inactive X chromosome (Xi) consistently resides in the heterochromatic compartment at the nuclear or nucleolar periphery (Fig. 1), which electron microscopic studies have long shown is enriched for more densely packaged chromatin less active in transcription. Thus, the inactive X is largely excluded from the internal euchromatic compartment that is punctuated by a number of smaller domains enriched in RNA metabolic factors, forming “hubs” of higher activity (Lamond and Spector 2003; for review, see Hall et al. 2006; Meaburn and Misteli 2007).

The organization of the Xi within overall nuclear structure, as well as most biochemical modifications across the chromosome, can be readily visualized by light microscopy (Fig. 1). However, a greater challenge is to investigate what organizational changes may be occurring to particular sequence elements within the chromosome itself. Because some genes are known to escape X inactivation, particularly in humans, a key question is whether those sequences are organized differently at a cytological level within the interphase chromosome territory, or whether escape from silencing is controlled at a more local level. Irrespective of how some genes escape silencing, the XIST-RNA-coated chromosome undergoes large-scale structural changes to form the DNA-dense Barr body, which is readily seen in human cells by light microscopy with DAPI staining (Fig. 1) and corresponds to a condensed chromatin mass by electron microscopy (Rego et al. 2008). Our laboratory recently took two different approaches to investigate the relationship between human Xi regulation and sequence elements, one that used molecular cytology and another using a bioinformatics word-count...
analysis. In both cases, the findings led us to the potential roles of repeat elements, of different types, in chromosome structure and regulation.

The human genome is not a linear entity, but a complex three dimensional (3D) structure with several levels of higher-order packaging, likely influenced by what we refer to as the “fabric” of sequences underlying it. In addition, the overall perspective of the human genome is changing dramatically as it becomes evident that the meaningful information occurs at least as much in the “noncoding” DNA as in protein-coding “genes.” Although the importance of noncoding DNA as a source of small and large noncoding RNAs (microRNAs and LincRNAs) is now being intensely investigated (for review, see Koziol and Rinn 2010), what we would term the “repeat genome” is still largely unexplored. Highly repetitive sequences comprise close to half of the human genome; in addition to the satellite sequences at centromeres, the bulk of the repeat genome is comprised of SINE elements (Alu in human, B1 or B2 in mouse), LINE elements, and simple sequence repeats (for review, see Smit 1996; Lander et al. 2001). We share the perspective stated by Britten and Kohne (1968) more than 40 years ago, “A concept that is repugnant to us is that about half of the DNA of higher organisms is trivial or permanently inert.” Several years ago, we began using the Cot-1 (rapidly reannealing) fraction of the genome, not as a means to mask repeats from consideration, but, rather, as a labeled probe to broadly survey the organization and/or expression of repeats throughout nuclei and chromosomes (Hall et al. 2002a; Clemson et al. 2006). As we consider how the noncoding XIST RNA paints a chromosome or how some genes escape the silencing cascade induced by XIST, we keep in mind the possible roles for the presumptive junk of the genome in chromosome structure and regulation.

**XIST RNA BINDS AND INDUCES SILENCING OF ONE X CHROMOSOME IN MAMMALIAN FEMALE CELLS**

To balance gene dosage between male and females, one X chromosome in female cells is almost entirely silenced, providing the preeminent model of facultative heterochromatin, with biochemical and structural changes manifest across the entire chromosome. Although the mouse trophectoderm shows imprinted inactivation of the paternal chromosome, human extraembryonic tissues do not, and in both species X-chromosome inactivation (XCI) within the inner cell mass and embryo proper is random (for review, see Payer and Lee 2008). The timing may differ slightly between human and mouse, but this random XCI initially occurs as cells of the inner cell mass begin to differentiate to specific lineages and involves mechanisms that specify counting and “choice” of which X chromosome will be silenced in each cell, although these are only partially understood. Recent work suggested that the X-linked gene Rlm/Rnf12 has a role in regulating Xist/X inactivation (Jonkers et al. 2009), and this transcription factor was further shown to be essential for imprinted paternal X inactivation and thus viability of the embryo (Shin et al. 2010). Whatever the mechanism whereby XIST is initially expressed from just one homolog, the biology of how this novel chromosomal RNA interacts with and impacts a whole chromosome is fascinating and will likely reveal fundamental insights into the structure and regulation of chromosomes in general.

Human XIST RNA is polyadenylated and spliced to generate an ~19-kb RNA expressed exclusively from the Xi (Brown et al. 1992), and silenced by methylation on Xa. XIST RNA forms a large accumulation tightly restricted to the Xi chromosome territory in interphase.
(Clemson et al. 1996), and in mouse, an antisense termed Tsix regulates Xist expression itself (Payer and Lee 2008). Most studies of the initiation of chromosome silencing have been done in mouse embryonic stem (ES) cells, where it was shown that shortly following Xist-RNA accumulation across the chromosome, a series of chromatin modifications occur, including increased H3K27me3, H3K9me, and H2AK119ub and decreased H3K4me and H3K4ace. XIST RNA is maintained on the Xi in all adult somatic cells, where it is associated with the late-replicating DAPI-dense Barr body and contributes to faithful maintenance of the silent state. Although Xist/XIST RNA is essential to initiate X inactivation in the embryo (Penny et al. 1996), silencing is maintained in somatic cells by the synergistic effects of multiple heterochromatic changes to the chromosome (Csankovszki et al. 2001). How this long noncoding RNA binds to and interacts with the chromosome is not clear, but at some level this must involve the sequence/structure of XIST RNA itself and its affinity for components of chromatin. However, a major question is whether XIST RNA is indifferent to the sequence of the underlying DNA. Below, we briefly consider current knowledge regarding RNA and chromatin protein factors, but focus primarily on evolving evidence that suggests interplay between XIST RNA and underlying chromosomal DNA, leading to the potential role of repeat sequences in chromosome regulation.

**AFFINITY OF A LONG NONCODING RNA FOR ITS PARENT CHROMOSOME AND POTENTIAL ROLES FOR TANDEM REPEATS WITHIN XIST RNA**

XIST RNA has now been studied for almost 20 years, yet only recently has there been movement toward understanding how or what regulates binding of this large RNA to the chromosome. Because XIST RNA binds across the chromosome, the underlying DNA logically has some role in XIST-RNA localization, either directly or indirectly. We have previously shown that the RNA remains surprisingly tightly bound and well localized to the chromosome territory in nuclei after histone extraction and digestion of the bulk (~95%) of chromosomal DNA (Clemson et al. 1996; Hall and Lawrence 2003). This indicates that the RNA is not localized via hybridization to chromosomal DNA, but suggests that the RNA likely bridges chromatin with the residual nonsoluble structural elements (scaffold) of the nucleus (model) (Fig. 2). In addition, as the RNA spreads from its site of transcription on the X and coats the rest of the chromosome, it must somehow recognize some archi-
tectural boundary of the interphase chromosome territory, because it does not promiscuously associate with the neighboring chromatin despite the intimate packaging of chromosomes within nuclei (Hall et al. 2002b). Thus, the chromosome territory likely has an underlying scaffold that somehow limits spread of Xist RNA.

Because the RNA has significant capacity to silence autosomal chromatin, as detailed below, it is essential that its binding be strictly cis-limited. We interpret this to likely require multiple redundant anchor points and cooperative players to ensure strict and stable XIST-RNA binding. Both human XIST and mouse Xist RNAs contain multiple regions of tandem repeats (identified as repeats A, B, C, D). Studies of Xist-RNA transgenes in mouse ES cells suggest that the A-repeat region is required for Xist RNA to silence chromatin (Wutz et al. 2002) and is important for PRC2 (EZH2) recruitment to enact H3K27 methylation (Zhao et al. 2008). With respect to RNA localization, however, Wutz et al. (2002) reported that a number of different regions along the length of the transcript contribute to its proper chromosomal localization. These sites do not exhibit a common motif and are functionally redundant, suggesting multiple low-affinity cooperative binding sites for XIST-RNA binding. This finding is consistent with other evidence from our laboratory (below) that there are likely multiple anchor points involving distinct modifications to chromosomal proteins (Hall et al. 2009).

Studies from the Strauss and Lee laboratories have found that introduction of PNA or LNA oligos, respectively, to Xist in living cells can dislodge Xist RNA from the chromosome (Memili et al. 2001; Sarma et al. 2010). Interestingly, both studies report that oligos that target the C-repeat of Xist are more effective in disrupting its localization, although this happens most rapidly when the LNA oligos are used. This suggests that the C-repeat is important in Xist-RNA binding; however, an involvement of other sequences cannot be ruled out because some parts of Xist RNA may be inaccessible to oligos in live cells. Nonetheless, this is an intriguing approach for future studies.

AURKB REGULATES XIST-RNA LOCALIZATION INDEPENDENT OF MITOSIS, ALLOWING MANIPULATION OF XIST-RNA INTERPHASE BINDING OR MITOTIC RELEASE

Although many laboratories have attempted to identify proteins that specifically interact with XIST RNA using standard biochemical approaches, identification of proteins that bind and localize XIST RNA has long frustrated researchers in this field. Evidence indicates that Xist RNA can recruit certain chromatin modifiers to Xi, as shown for EZH2/PRC2 (Fang et al. 2004; Schoefneter et al. 2006), but identification of proteins that localize or regulate Xist RNA has been a challenge. One study implicated the tumor suppressor BRCA1 as localizing across the Xi to support XIST-RNA binding (Ganesan et al. 2002); however, our group found that BRCA1 associates with replicating satellite heterochromatin on many chromosomes but does not “paint” the Xi or have a direct role in XIST localization (Pageau and Lawrence 2006; Pageau et al. 2007b). Interestingly, loss of the Barr body and XIST RNA is common in several types of aggressive breast and other tumors, and there is now consensus that this primarily reflects a broader effect of BRCA1 or other tumor-suppressor loss on genetic and epigenetic instability of cells (Pageau et al. 2007a). Very recently, Hasegawa et al. (2010) appear to have had success in identifying an Xist-RNA-binding factor, implicating the matrix protein hnRNPA/SAF-A in Xist-RNA binding (Hasegawa et al. 2010). This finding followed an earlier report showing that SAF-A localized to Xi with XIST RNA (Helbig and Fackelmayer 2003).

Because biochemical analysis of such a large, tightly chromatin-bound RNA has been difficult by standard extraction-based approaches, we pursued a different approach to manipulate the binding or release of XIST RNA directly within living cells. This strategy built on a significant clue from our earlier demonstration that XIST RNA releases from the mitotic Xi during early prophase in human cells, and slightly later in mouse (Clemson et al. 1996; Hall et al. 2009). Following mitotic release, the RNA is visible as bright punctate dots throughout the cytoplasm (Fig. 2), and it is resynthesized in early G1 daughter cells (Clemson et al. 1996). Tests of inhibitors that mimic mitotic chromatin modifications implicated an indirect role of protein phosphatase 1 (PP1), where only PP1 inhibitors released XIST RNA to disperse throughout the nucleoplasm at interphase, similar to its punctate distribution at mitosis (Fig. 2). PP1 normally inhibits AURKB at interphase, and AURKB normally localizes across the chromosome arms at prophase (when XIST RNA detaches), where it phosphorylates H3 and other chromatin proteins and releases HP1. Our results indicated that the improper activation of AURKB in interphase caused H3Ser10 phosphorylation of interphase chromatin concomitant with release of the RNA, whereas RNAi depletion of AURKB prevented the normal release of XIST RNA from condensed mitotic chromosomes (Fig. 2). Thus, the release of XIST RNA at mitosis is independent of chromosome condensation but appears to involve chromatin modifications controlled directly or indirectly by AURKB-mediated phosphorylation. AURKB, which is highly overexpressed in many cancers, is known for its role at the mitotic spindle and centromeres, but Hall et al. (2009) surprisingly identified AURKB as a key player in regulation of RNA binding to heterochromatin. Although H3Ser10 and HP1 (regulated by AURKB phosphorylation) may be involved, phosphorylation of H3Ser10 alone at interphase was not sufficient to release XIST RNA. However, as shown in the model in Figure 2, these and the above results suggest that XIST-RNA binding likely involves multiple distinct anchor points. As summarized in Figure 2, the ability to manipulate XIST RNA in vivo (to remain bound at metaphase or to release at interphase) provides a new approach to further advance the thorny question of what regulates XIST-RNA binding and identify the biochemical changes most closely linked to XIST-RNA behavior (see Fig. 2).
XIST RNA can repress most autosomal chromatin but is not indifferent to DNA sequence content

A remarkable aspect of Xist/XIST chromosome biology is that this novel RNA can silence autosomal chromatin, as indicated by silencing of autosomal chromatin in X;autosome translocations (White et al. 1998) that is associated with XIST RNA (Hall et al. 2002a), as well as silencing by Xist/XIST transgenes inserted into autosomes (Lee and Jaenisch 1997; Hall et al. 2002a). Although it was suggested that mouse ES cells that had begun to differentiate were no longer competent to enact chromosome silencing in response to Xist transgenes (Wutz et al. 2002), we and our collaborators found that human XIST transgenes could inactivate autosomal chromosomes in somatic cell lines (Hall et al. 2002a; Chow et al. 2003, 2007). In fact, the most “robust” example of a condensed Barr body currently in our laboratory is actually a chromosome 4 Barr body, carrying a cosmid X transgene (Fig. 1) in HT1080 fibrosarcoma cells. However, because the HT1080 and 293 cell lines used have neoplastic origins, we suggested that such cells may have higher epigenetic plasticity (Hall et al. 2002a), which the Wutz laboratory (Agrelo et al. 2009) recently reported involves a requirement for the SATB1 protein.

Despite the finding that XIST RNA is retained in nuclear structure after digestion of chromosomal DNA, numerous observations indicate that XIST/XIST RNA is not indifferent to the chromosomal DNA sequence. Foremost among them is that in humans, many genes escape X inactivation, to varying degrees (Carrel and Willard 2005). Additionally, in mouse mitotic cells, before Xist RNA detaches, it appears that the mouse does not coat the centromeric region, and the RNA selectively detaches to form a discrete banding pattern (Smith et al. 2004), similar to observations on vole chromosomes (Nesterova et al. 2004). In fact, the most “robust” example of a condensed Barr body currently in our laboratory is actually a chromosome 4 Barr body, carrying a cosmid X transgene (Fig. 1) in HT1080 fibrosarcoma cells. However, because the HT1080 and 293 cell lines used have neoplastic origins, we suggested that such cells may have higher epigenetic plasticity (Hall et al. 2002a), which the Wutz laboratory (Agrelo et al. 2009) recently reported involves a requirement for the SATB1 protein.

It is notable that Migeon et al. (1999) and others have reported that human XIST RNA is capable of anchoring to mouse chromosomes, even though most of the primary sequence of human and mouse XIST RNA is not well conserved (Brockdorff et al. 1992; Brown et al. 1992; Migeon et al. 1999). As the above studies on AURKB strongly indicate that protein modifications regulate XIST-RNA binding, this suggests that human XIST RNA can recognize mouse chromatin proteins on a mouse chromosome. However, two studies (Clemson et al. 1998; Hansen et al. 1998) showed that, in mouse human hybrid cells carrying just one human chromosome, human XIST RNA cannot localize to the human X chromosome in the otherwise mouse nucleus, indicating that mouse chromosomal proteins do not properly interact with the human chromosomal DNA to stabilize XIST-RNA binding. As recently reviewed elsewhere (Minks and Brown 2009), Xist/XIST transgene studies to date are limited by differences in random insertion sites, copy numbers, and epigenetic factors of different cells. However, use of targeted, single-copy XIST transgenes in the future has great potential to further elucidate fundamental aspects of XIST RNA and chromosome biology.

Bioinformatics analysis implicates different types of repeat elements in chromosome silencing and escape

X-chromosome inactivation provides a singular opportunity to investigate the potential relationship between sequence elements and the structural and functional transformation of essentially a whole chromosome. Because at least hundreds or thousands of XIST transcripts bind across the chromosome, sequence motifs involved would likely be highly represented throughout the genome, potentially making it difficult to discriminate them from putative “junk.” It was long ago suggested that repetitive sequences may be involved in promoting chromosome inactivation (Gartler and Riggs 1983), with LINE elements a suspect, as noted above (for review, see Lyon 2003). Using bioinformatics sequence analysis, Bailey et al. (2000) reported that the human X chromosome has an ~1.7-fold higher level of L1 LINE elements than the autosomal average except in the region at Xp22 that escapes chromosome inactivation. Although this correlation is consistent with an involvement of L1 elements in promoting X inactivation, L1 may also have accumulated on the X because of its lower meiotic recombination in all but the pseudoautosomal region (which also lacks L1 enrichment). Other studies of canonical repeats concluded that L1 elements are either not likely involved (Chureau et al. 2002; Ke and Collins 2003) or may not be solely responsible (Ross et al. 2005).

Rather than focus on candidate elements, we took an open, unbiased bioinformatics search for any motifs that are abundant, widely distributed, and specifically enriched on the X chromosome. Additionally, analyses were performed with and without masking for known interspersed repeat families (e.g., LINEs, SINEs, and LTRs), because these copious elements may well contribute, but would...
also likely obscure other repeated motifs. Using a linguistic approach, the number and distribution of all 9-bp words in the genomic sequence of all individual human chromosomes were examined. In addition, we focused on the X chromosome as comprising two distinct segments: XE, an ~10-Mb region at Xp22.3 (Figs. 3 and 4) that includes the pseudoautosomal region and more fully escapes X inactivation (Carrel and Willard 2005); and XS, the remainder of the chromosome, which is largely silenced on Xi. Although there are some genes scattered throughout XS that partially escape silencing in some cell types, XE is a large, unique chromosomal domain that is wholly resistant to X inactivation, unlike autosomal chromatin, which has substantial capacity for inactivation.

This analysis revealed several substantial new features of X-chromosome sequence content. First, we confirmed that L1 is enriched on XS but extended this to show that this enrichment on X is distinct from all individual autosomes (not just the autosomal average) (Fig. 3). This was important because individual chromosomes can vary substantially; for example, gene-rich Chr 19 is especially depleted in L1 elements in contrast to Chr 4. Notably, the rest of our findings identified differences in simple sequence repeats, which are typically excluded from such analyses. Results showed that the dinucleotide repeats [AT], [AC], and [AG], are significantly enriched across the X chromosome compared to autosomes. Intriguingly, these repeats have the property of being able to form unusual DNA structures, which potentially could contribute to the regulation of facultative heterochromatin.

Most importantly, this analysis uncovered a dramatic difference in the content of small simple repeats scattered throughout the whole region. As shown in Figure 4, a striking enrichment (>10-fold) of (GATA)n distinctly marks the 10-Mb segment at Xp22 that escapes inactivation, which is confirmed by FISH with an oligo GATA probe, and which, importantly, is also seen in other eutherians that have a pseudoautosomal region (McNeil et al. 2006). These findings suggested a new paradigm whereby a regional escape from X inactivation in a large chromosomal domain may be due to the presence of elements that prevent heterochromatinization, rather than simply lack of elements that promote it. The GATA repeats are clearly a marked and conserved feature dispersed at many sites throughout the “fabric” of this large chromosomal segment; in fact, further analysis showed that no

Figure 3. L1 LINE density versus gene density across all human chromosomes and XE escape domain. The silenced region of X chromosome (XS) has highest L1 (L1) density compared to each chromosome in the human genome as well as to escape region of the X (XE). Note that autosomes show significant variations in density of L1 elements, but these data show that XS has significantly more than the most enriched autosome (Chr 4), which has similar gene density. Because LINE content is known to correlate with gene density, it was important to show that the difference is not simply explained by gene density. Data also show that Y chromosome is more enriched in L1s than autosomes, consistent with the possibility that some of the evolutionary accumulation of LINE elements on the sex chromosomes may have to do with their lower recombination frequency. (From McNeil et al. 2006; reprinted, with permission, from Genome Research, ©2006.)

Figure 4. Pseudoautosomal region, which fully escapes inactivation on the Xi, exhibits striking 11-fold enrichment in GATA-repeat sequence. (Right) Distribution of GATA repeats (and CTAT) along the X chromosome determined by word-count analysis and expanded view of pseudoautosomal and XE region that escapes inactivation (yellow). Importantly, small repeats of (GATA) were widely dispersed at very many sites throughout this ~10-Mb segment and were not found in large blocks. (Left) DNA FISH using GATA DNA oligo probe labels pseudoautosomal region of X chromosome strongly, confirming that it is unique in the genome for this striking sequence feature.
other 10-Mb chromosomal segment in the genome showed such a striking enrichment for any 9-mer word (J McNeil and JB Lawrence, unpubl.). Thus, this strongly suggests that the GATA repeats are involved in the unique biology of this region, in either escape from silencing or potentially the obligatory meiotic recombination of this region in the XY body, or both (McNeil et al. 2006). Recent literature provides other examples in which gene regulation appears to be coordinate across a chromosomal domain, such as in hESC (Li et al. 2006), or silencing of a tumor-suppressor gene in a band-sized chromatin domain (Frigola et al. 2006). We suggest that the broader sequence suppressor gene in a band-sized chromatin domain such as in hESC (Li et al. 2006), or silencing of a tumor-suppressor gene in a band-sized chromatin domain (Frigola et al. 2006). We suggest that the broader sequence context of a chromosomal region may increasingly prove important in gene regulation, and the chromosomal domain may in part be defined by the repeated motifs or “words” that populate it.

Although (GATA), repeats clearly mark this unique chromosomal domain that more fully escapes silencing, GATA enrichment is not seen for the individual genes that partially or more variably escape inactivation throughout the rest of XS (McNeil et al. 2006). These may be regulated by a distinct mechanism, because in mouse the Jarid1C gene (and two other individual genes that escape silencing) were flanked by CTCF-binding sites (Filippova et al. 2005), and Li and Carrel (2008) further showed this was an intrinsic property of the Jarid1C locus. Carrel et al. (2006) and Wang et al. (2006) each published that computer profiling could recognize motifs that predict genes that escape silencing even outside of the XE region. Although no discrete consensus motif was identified, this provides further evidence that XIST RNA is not indifferent to chromosomal sequence context.


In addition to the bioinformatics studies above, our laboratory also took a molecular cytological approach to address whether genes that escape X inactivation are organized differently at a cytological level. The Xi clearly exhibits a striking “architectural” change manifested by formation of the heterochromatic Barr body, which is most obvious in human cells (Fig. 1), but also visible in mouse cells. We hypothesized that genes that escape silencing might lie outside of the XIST-RNA territory and avoid being engulfed and silenced in the condensed Barr body (Fig. 5, Model B) or, alternatively, be within the Barr body but controlled more locally (Fig. 5, Model A). The Barr body has clear features of heterochromatin and, as further evidenced below, is transcriptionally inactive; therefore it was logically and universally assumed that the Barr body contained the silenced genes of the Xi. Surprisingly, however, through direct analysis of 14 genes (six that escape and eight that are silenced), we found that not only are the escape genes just outside of the Barr body, but the silenced genes are as well, a result we never anticipated (Fig. 5, last model). All 14 genes localized overwhelmingly on the rim of the XIST-RNA territory (Fig. 5) and, moreover, outside of the DNA-dense Barr body. Initially when we saw this result with the first few genes examined, we hesitated to publish these counterintuitive results. However, we analyzed more loci in numerous experiments using different fixation methods, and this ruled out artifacts due to hybridization efficiency, probe penetration, and 3D imaging. This gene organization within the overall Xi architecture became clearer as we better defined three distinct ways to examine the Xi: the DNA-dense Barr body, the XIST-RNA territory, and the X-chromosome DNA territory (Clemson et al. 2006). Although these are all used as ways to visualize Xi, they are not equivalent. As shown by the line-scan analysis in Figure 5, the DAPI-dense Barr body does not, in fact, encompass the entire X-chromosomal DNA territory, but occupies ~65% of its interior. Similarly, as previously indicated (Clemson et al. 1996), the XIST-RNA territory is slightly smaller (80%–85%) than the DNA territory, but larger than the Barr body. Thus, the Barr body is not the Xi, but a densely packaged core within Xi. Quantitative results showed that these gene-coding loci positioned predominantly, although not entirely, at the border or outer rim of the XIST-RNA territory regardless of whether they escape or are silenced. In a minority of cells, the gene signal appears to be more internal within the territory, but these frequently localized to an invagination or gap in the Xi-DNA staining.

A study from the Heard laboratory included analysis of six X-linked genes in the mouse; although these findings do not disagree with those of Clemson et al. (2006), initially it may appear that they do (Chaumeil et al. 2006). Chaumeil et al. (2006) studied Xi genes during the initial formation of the Xi in mouse ES cells, and reported that the genes moved inward from a more distal location, relative to the Xist-RNA territory, as X inactivation proceeds. In this case, shift to a more “internal” location was apparently in comparison to the much looser chromosome packaging in ES cells (where genes can reside outside even the visible X DNA territory). In the discussion, Chaumeil et al. (2006) refer to the propensity of genes to be more peripheral in the Xi territory as shown by Clemson et al. (2006) and then state, “It should be noted that all of the X-linked genes we have examined so far tend to be fairly peripherally located on the X-chromosome territory, whatever their status, as in the study by Clemson et al. (2006).” Thus, in this study of the mouse, six genes (regardless of silencing) remained in the periphery of the territory and not in the center, consistent with our demonstration of gene organization on the human Xi in somatic cells.

In addition to this surprising organization of silenced genes, Clemson et al. (2006) also concluded that, at the cytological levels examined, there was not a clear pattern of differential localization between silenced versus escape genes. Clemson et al. specifically state that a more subtle difference in packaging at the cytological level could not be ruled out, but the results indicate that escape from silencing appears to be controlled at a more “local” level of chromatin packaging or gene regulation, likely below the resolution of light microscopy. For example, it remains possible that a higher-resolution (ultrastructural) analysis of gene position relative to the dense heterochromatin of
the Barr body would show some difference in proximity of escape versus silenced genes. Chaumeil et al. reported that the *Jarid1C* gene, which escapes inactivation, had a slightly more external position in differentiating ES cells; however, any difference was quite subtle, and the consistency of this distinction remains to be established. Because many more genes escape inactivation in the human and the Barr body is more readily seen, the human Xi provided an advantageous system to examine this for a larger sampling of genes, which in our hands showed for the "steady state" of Xi of somatic cells no consistent difference in the organization of silent and escape genes within the chromosome territory.

Finally, as further evidenced below, although the silenced X-linked human genes were typically not within the core of the Barr body, they were still positioned abutting this large block of repressed chromatin, which likely is key to their silencing. This is also reminiscent of the reported positioning of inactive loci adjacent to the large heterochromatic mouse chromocenters (Brown et al. 1997), suggested to contribute to their silencing. In addition, the localization of genes outside of the Barr body may place them in an environment more permissive to escape from inactivation, dependent on whether local sequence context supports this, such as the possible abundance of GATA repeats (McNeil et al. 2006) or other elements discussed above.

**XIST RNA IMPACTS CHROMOSOME ARCHITECTURE AND COATS “JUNK”-RICH CORE WITH GENE-RICH OUTER RIM**

These unanticipated findings on gene organization on the human Xi led to two important related concepts, as forwarded by Clemson et al. (2006). First, because the numerous genes studied in human (with detection efficien-
cies of 90%–99%) localized at the XIST-RNA periphery and just outside of the Barr body, this begged the question: What sequences are within the Barr body? Clemson et al. took a significant step toward addressing this by showing hybridization to the “repeat genome,” using Cot-1 DNA as a probe; Cot-1 DNA was detected within the central regions of the XIST-RNA territory and within the Barr body (Fig. 6), as was the X centromere in many cells. Because Cot-1 DNA is composed mostly of interspersed LINE (L1) and SINE (Alu) sequences, this indicated that such abundant repeats may have a distinct role in Xi chromosome structure, facilitating the formation of a heterochromatic inner core that is disproportionately “noncoding” and repeat-rich. Thus, Clemson et al. suggested that structural elements of the chromosome likely provide a framework that positions most protein-coding genes at or near the surface of the chromosome DNA territory. A recent study by Chow et al. (2010), using a probe to LINE elements, corroborated that LINE repeats are found within the central regions of the mouse Xi DNA territory.

Moreover, the surprising organization of genes at the peripheral rim provided an important new insight into how XIST RNA acts in chromosome silencing, in that these findings strongly indicated that XIST RNA does not act just at the local level on individual genes, as might have been anticipated, but has a more architectural relationship with the chromosome as a whole. For example, although a priori it would be plausible for XIST RNA to silence by binding near gene promoters, much (~87%) of the interphase chromosome painted by XIST RNA is the DAPI-dense inner core. As XIST RNA detaches from mouse mitotic chromosomes, it produces a banding pattern because it is retained longest on gene-rich R bands, suggesting that it could have a different affinity for gene-rich regions (Smith et al. 2004). However, it is the repeat-rich inner core that is converted into a DAPI-dense heterochromatic ball on the Xi (similar to a mouse chromocenter), and this appears to be one of the early steps in initiation of inactivation. Thus, this indicates that as XIST RNA paints the Xi during initiation, its interaction with these copious repeats results in generation of a large (on a molecular scale) nuclear compartment in which the whole ~1–1.5-µm region is silent and devoid of RNA metabolic factors, as further discussed below.

Because most autosomal chromatin has substantial capacity to be silenced by XIST RNA, it is important to note that XIST-silenced autosomal material also showed a similar organization of genes at the border of the XIST-RNA territory (Clemson et al. 2006). Thus, it is likely that the structural principles of chromosome organization that allow for formation of facultative heterochromatin may be shared between XIST-mediated facultative heterochromatin and other developmental mechanisms that control cell-type-specific heterochromatin formation on autosomes.

Figure 6. Results and model for relationship of Cot-1 repeat RNA to active and inactive X chromosomes. (A) Cot-1 repeat RNA (green) is present throughout the nucleoplasm of active chromosomes but is essentially absent within interior core of Xi, even though Cot-1 repeat DNA (red) is detectable within DAPI-dense Barr bodies (white). Thus, Xi has a core of silent repeat elements. Line scans through Xi in normal female fibroblast quantifies the sharp distinction between Cot-1 DNA enrichment and lack of Cot-1 RNA. (B) Model for the loosely peripheral organization of genes on Xa, which on Xi becomes more striking and surrounds a compacted and silent inner core. We suggest that Cot-1 RNA may be expressed throughout central regions of euchromatic interphase chromosome territories, even though evidence indicates that central regions have fewer genes.
A POTENTIAL ROLE OF REPEAT RNA EXPRESSION AND SILENCING IN CHROMOSOME REGULATION

In 2002, we demonstrated a new method to assay silencing across an XIST-transgenic chromosome by assessing what we referred to as hnRNA (Hall et al. 2002a), but that we long suspected suggests something more profound regarding repeat RNAs. Instead of using cold Cot-1 DNA as a competitor to suppress repeat hybridization, we used labeled Cot-1 DNA as a probe hybridized to nonadenatured nuclei, thereby detecting single-stranded RNAs containing repeats, which would include introns in pre-mRNAs. Rather than laborious expression assays for individual genes along the chromosome, this new approach allowed a more rapid and convenient overview of long-range transcriptional activity, which we found nicely distinguishes the Xa from Xi (Hall et al. 2002a) and delineates heterochromatin at centromeres or the nuclear or nucleolar periphery (Tam 2004). We discussed above the evidence for involvement of repetitive DNA in chromosome structure and regulation; however, analysis of “Cot-1 RNA” led us to suggest that interspersed repeat RNAs may be involved as well (Clemson et al. 2006). Following the spread of XIST RNA across the chromosome, the repetitive DNA at the “core” of the territory is compacted into the Barr body, which we showed lacked Cot-1 RNA despite containing copious Cot-1 DNA. In contrast, Cot-1 RNA is robustly detected throughout the nucleoplasm associated with the autosomal and X chromosome territories. Clearly, the repeats within the Barr body are silent (Figs. 1, 6), and XIST RNA acts on the chromosome to generate this silent compartment. Because most LINE elements are truncated and generally believed silent, one possibility is that this already silent DNA is reorganized into a recognizable structure (the Barr body). However, we suggest that an alternative hypothesis may be true: that XIST RNAs’ first function is not only to structurally reorganize this nongenic DNA in this core, but also to silence it. In either case, a nuclear compartment lacking in Cot-1 RNA, RNA Pol II (Chaumeil et al. 2006; Clemson et al. 2006), and other RNA metabolic factors (Clemson et al. 1996) is formed. Creating a large Barr body to shut off at most a few hundred genes would seem “overkill,” but this suggests that formation of this structural compartment is key to the subsequent genic silencing, which involves regulation at the level of the whole chromosome and its nuclear organization.

These observations likely have relevance to the structure and regulation of euchromatic chromosomes. Studies of several autosomal genes reported a tendency to be loosely associated with the euchromatin or centromeres (e.g., clustered heterochromatin) may be near an XIST-RNA focus (Hall et al. 2002a) and delineates heterochromatin at centromeres or the nuclear or nucleolar periphery (Tam 2004). We discussed above the evidence for involvement of repetitive DNA in chromosome structure and regulation; however, analysis of “Cot-1 RNA” led us to suggest that interspersed repeat RNAs may be involved as well (Clemson et al. 2006). Following the spread of XIST RNA across the chromosome, the repetitive DNA at the “core” of the territory is compacted into the Barr body, which we showed lacked Cot-1 RNA despite containing copious Cot-1 DNA. In contrast, Cot-1 RNA is robustly detected throughout the nucleoplasm associated with the autosomal and X chromosome territories. Clearly, the repeats within the Barr body are silent (Figs. 1, 6), and XIST RNA acts on the chromosome to generate this silent compartment. Because most LINE elements are truncated and generally believed silent, one possibility is that this already silent DNA is reorganized into a recognizable structure (the Barr body). However, we suggest that an alternative hypothesis may be true: that XIST RNAs’ first function is not only to structurally reorganize this nongenic DNA in this core, but also to silence it. In either case, a nuclear compartment lacking in Cot-1 RNA, RNA Pol II (Chaumeil et al. 2006; Clemson et al. 2006), and other RNA metabolic factors (Clemson et al. 1996) is formed. Creating a large Barr body to shut off at most a few hundred genes would seem “overkill,” but this suggests that formation of this structural compartment is key to the subsequent genic silencing, which involves regulation at the level of the whole chromosome and its nuclear organization.

These observations likely have relevance to the structure and regulation of euchromatic chromosomes. Studies of several autosomal genes reported a tendency to be loosely localized within an outer zone of the chromosome territory (for review, see Cremer and Cremer 2001; Chubb and Bickmore 2003). Although the number of genes in most studies was small and the extent of this was debated, Clemson et al. (2006) examined this for 15 loci on Xa as well as Xi, and showed a much less pronounced organization of Xa versus Xi genes, but nonetheless a loose propensity for genes to reside in the peripheral zone on Xa as well (Fig. 6). Although this indirectly suggests that nongenic and repeat DNA may be preferentially enriched within the central regions, Cot-1 repeat RNA appears copiously expressed throughout euchromatic chromosome territories (Clemson et al. 2006; A Gomez, LL Hall, and JB Lawrence, unpubl.). Thus, we hypothesize that Cot-1 RNA may include substantial repeat RNA apart from that in introns of pre-mRNAs (Clemson et al. 2006), which we interpret to be most consistent with the results of Chaumeil et al. (2006) showing that the Cot-1 RNA–depleted compartment forms while gene transcription foci are still expressed. Currently it is widely believed that intergenic LINE elements are mostly inert, as only a small fraction is “active” and believed capable of transcription. However, years ago, early evidence suggested expression of repeated sequences in cell-type-specific patterns (Britten and Kohne 1968; Davidson and Britten 1979), as further evidenced recently by Faulkner et al. (2009), and by earlier evidence that murine B1 and B2 RNAs are involved in the stress response (Allen et al. 2004; Williams et al. 2004).

More recently, Chow et al. (2010) reported that specific active full-length LINE elements are transiently up-regulated as X inactivation initiates in mouse ES cells. They interpret this to suggest that particular active L1 elements influence the regulation of a nearby gene that escapes silencing. These interesting findings suggest some role of particular L1 loci in regulation of specific neighboring genes, but this is distinct from the concept discussed above regarding the broader set of abundant repeats distributed throughout the chromosome contributing en masse to the regulation of the heterochromatic versus euchromatic status of a chromosome or chromosomal domain.

WORD OF CAUTION REGARDING INTERPRETING COT-1 RNA ASSAY AS INDICATOR OF CHROMOSOME SILENCING

The COT-1 RNA assay (Hall et al. 2002a) has been widely adopted as an effective tool to delineate Xi or other heterochromatin, and in many cases, this assay has been used to good effect. However, we believe that it should be applied cautiously with at least two major caveats in mind: (1) As per the discussion above, silencing of Cot-1 RNA might be distinct from silencing of protein-coding genes, or even specific noncoding RNAs. If so, as we suggest, this would be both interesting and important, but nonetheless should be considered in interpreting results. Cot-1 RNA does not necessarily equal genic transcription. (2) The smaller the Cot-1 RNA “hole,” the more subject to interpretation and misinterpretation the analysis is. We have noted in recent literature that quite small Cot-1 RNA “holes” are interpreted to support important points, but there are many small areas of the nucleoplasm that have diminished Cot-1 RNA expression, and neighboring structures that may lack Cot-1 RNA (e.g., clustered heterochromatin or centromeres) may be near an XIST-RNA focus but be unrelated to XIST or chromosome silencing.

LANGUAGE OF THE REPEAT GENOME

As evident throughout the above review, the power of the X-chromosome inactivation model has begun to reveal new insights into the potential roles of repeat sequences in the genome, at multiple levels. Several different types of repeats are implicated in several distinct aspects of
XIST RNA/chromosome regulation, and in addition to a structural role of DNA elements, their expression in RNA may undergo complex, possibly cell-type-specific changes. Although this area of research is still at a very early stage and the import and mechanisms of most of these findings remain to be established, the overall lesson in our view is that the “repeat genome” may contain substantial meaningful information well beyond the widely held view of most repeats as evolutionary debris. Repeats have physical properties that make them especially well suited for contributing to chromosomal regulation. For example, McNeil et al. (2006) discuss the potential for GATA and other simple sequence repeats to form unusual DNA structures (triplex and Z-DNA). Interspersed repeats, such as LINES and SINES, have the intriguing potential to form interstrand molecular hybrids (with DNA or RNA) that could impact higher-order chromatin folding. In addition, owing to their abundance and broad “interspersed” distribution, they have greater potential to regulate chromosomal domains, not just individual genes, by recruiting or sequestering substantial quantities of chromatin-modifying RNAs or enzymes. On the basis of the various examples evidenced above and other theoretical considerations beyond the scope of this chapter, we speculate that there is a “language of the repeat genome” that we have yet to learn how to read. Ultimately, however, the drive to understand the regulation of protein-coding genes will require greater attention to this very major part of the genome long masked from most genomic analyses.

ACKNOWLEDGMENTS

Our research is supported by the National Institutes of Health grants GMGM053234 and GM085548 to J.B.L. from the Institute of General Medicine. We thank Dawn Carone and Meg Byron for reading the manuscript and Charlene Baron for help with figure processing.

REFERENCES

Britten RJ, Kohne DE. 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.


Pageau GJ, Hall LL, Lawrence JB. 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. J Cell Biochem 100: 835–850.


XIST RNA and Architecture of the Inactive X Chromosome: Implications for the Repeat Genome

L.L. Hall and J.B. Lawrence

Cold Spring Harb Symp Quant Biol published online March 29, 2011
Access the most recent version at doi:10.1101/sqb.2010.75.030

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to Cold Spring Harbor Symposia on Quantitative Biology go to:
http://symposium.cshlp.org/subscriptions