ROLE OF XIST RNA AND ITS INTERACTING PROTEIN PARTNERS IN GENE SILENCING

by

JAKUB MINKS

M.Sc., The Institute of Chemical Technology Prague, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2012

ABSTRACT

X-chromosome inactivation ensures equal expression of mammalian male and female X-linked genes by transcriptionally silencing one X chromosome in each female cell. The pivotal molecule responsible for the silencing is a long non-coding RNA XIST; however, an all-encompassing model explaining how XIST induces silencing of the whole X chromosome is yet to emerge. This thesis aims to broaden our understanding of XIST action in humans by leveraging an inducible *XIST* transgene capable of silencing downstream reporters to identify sequences within XIST and XIST-interacting proteins critical for gene silencing.

First, we demonstrate that the repeat A region of XIST is necessary and sufficient to induce gene silencing, at least locally, irrespective of the makeup of the surrounding chromatin, and that XIST induces silencing of a distal gene in one of the HT1080 cell lines. Second, we show that individual repeats of a consensus repeat A sequence contribute additively to silencing. Mutations within a construct consisting of two repeat A units both demonstrate that the two palindromic sequences within the repeat A units spanning 'ATCG' and 'ATAC' tetranucleotides are critical for repeat A function and add to the evidence that the first palindrome forms a hairpin, rather than engaging in pairing between repeat A units.

Third, we explore which proteins are critical for XIST-induced silencing. We show that histone deacetylation, an early mark of an X-chromosome inactivation, is likely a consequence, and not the cause of XIST-induced silencing. We next demonstrate that in the transgenic HT1080 system, gene silencing is not accompanied by recruitment of the H3K27me3 repressive histone mark and XIST induces silencing independently of its previously reported associations with the polycomb repressive complex 2 (PRC2). Finally, we performed siRNA-mediated knock-down of 31 proteins previously implicated to play a role in X-chromosome inactivation. Our results show that proteins involved in XIST RNA localization (YY1), chromatin organization (SATB2, HNRNPU), and cell cycle (ATM), as well as an E3 ubiquitin ligase (SPOP) contribute to XIST-induced gene silencing in the HT1080 system. Thus, we demonstrate that the repeat A alone induces gene silencing and identify candidate pathways critical for its function.

PREFACE

Parts of this thesis were previously published in:

Minks, J and Brown, CJ, Getting to the center of X-chromosome inactivation: the role of transgenes. *Biochem Cell Biol*, **87**(5): 759-766 (2009).

- The modified text published in this paper is contained in section 1. The candidate (J. Minks) wrote the manuscript.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DOX – doxycycline

FCS – fetal calf serum

FRET – fluorescent resonance energy transfer

hnRNA – heterogeneous nuclear RNA

IP – immunoprecipitation

lncRNA – long non-coding RNA

mRNA – messenger RNA

miRNA – microRNA

NMR – nuclear magnetic resonance

PBS – phosphate-buffered saline

PCR – polymerase chain reaction

piRNA – Piwi-interacting RNA

PRC1 – polycomb repressive complex 1

PRC2 – polycomb repressive complex 2

qPCR – quantitative PCR

qRT-PCR – reverse transcription followed by quantitative PCR of the cDNA

rRNA – ribosomal RNA

s.d. – standard deviation

siRNA – small interfering RNA

snRNA – small nuclear RNA

tRNA – transfer RNA

TSA – trichostatin A

VPA – sodium valproate

Xa – active X chromosome

Xi – inactive X chromosome

XIC/*Xic* – human / mouse X-inactivation centre

LIST OF GENE NAMES

The following table lists gene symbols and the corresponding full names of all genes mentioned in this thesis. Unless noted otherwise, human genes are described. Names of other organisms are abbreviated as follows: *C. e.* - *Caenorhabditis elegans*; *D. m.* - *Drosophila melanogaster*; *G. g. Gallus gallus*; *M. m.* - *Mus musculus*.

Gene symbol	Gene name	Organism
ACTB	actin, beta	
AGPAT5	1-acylglycerol-3-phosphate O-acyltransferase 5	
AIR	antisense of IGF2R RNA (non-protein coding)	
AOF2	lysine (K)-specific demethylase 1A	
ASH2L	ash2 (absent, small, or homeotic)-like (Drosophila)	
ATM	ataxia telangiectasia mutated	
ATR	ataxia telangiectasia and Rad3 related	
ATRX	alpha thalassemia/mental retardation syndrome X-linked	
BBS9	Bardet-Biedl syndrome 9	
BRCA1	breast cancer 1, early onset	
CAPG-1	CAP-G condensin subunit-1	*C. e.*
CARM1	coactivator-associated arginine methyltransferase 1	
CBX1 – CBX8	chromobox homolog 1 – 8	
CHEK1	checkpoint kinase 1	
CHEK2	checkpoint kinase 2	
CLDN1	claudin 1	
CLDN16	claudin 16	
CTCF	CCCTC-binding factor (zinc finger protein)	
CUL3	cullin 3	
DAXX	death-domain associated protein	
DCHS2	dachsous 2 (Drosophila)	
DICER1	dicer 1, ribonuclease type III	
DNMT1	DNA (cytosine-5-)-methyltransferase 1	
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	
DNMT3B	DNA (cytosine-5-)-methyltransferase 3 beta	
DPY-21	DumPY-21	*C. e.*
DPY-26	DumPY-26	*C. e.*
DPY-27	DumPY-27	*C. e.*
DPY-28	DumPY-28	*C. e.*
DPY30	dpy-30 homolog (C. elegans)	

Gene symbol	Gene name	Organism
DPY-30	DumPY-30	*C. e.*
DXPas34	DNA segment, Chr X, Pasteur Institute 34	*M. m.*
EED	embryonic ectoderm development	
EHMT1/GLP	euchromatic histone-lysine N-methyltransferase 1	
EHMT2/G9a	euchromatic histone-lysine N-methyltransferase 2	
EZH2	enhancer of zeste homolog 2 (Drosophila)	
FAM222A	family with sequence similarity 222, member A	
FNDC3B	fibronectin type III domain containing 3B	
FRMD4A	FERM domain containing 4A	
FTX	FTX transcript, XIST regulator (non-protein coding)	
H19	H19, imprinted maternally expressed transcript (non-protein coding)	
H2AFY	H2A histone family, member Y	
H2AFY2	H2A histone family, member Y2	
HBA	hemoglobin, alpha [gene cluster]	
HBB	hemoglobin, beta [gene cluster]	
HDAC1 – HDAC11	histone deacetylase 1 – 11	
HNRNPK	heterogeneous nuclear ribonucleoprotein K	
HNRNPU / SAF-A	heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)	
HOTAIR	HOX transcript antisense RNA (non-protein coding)	
HOTAIRM1	HOXA transcript antisense RNA, myeloid-specific 1 (non-protein coding)	
HOTTIP	HOXA distal transcript antisense RNA (non-protein coding)	
Hprt	hypoxanthine phosphoribosyltransferase	*M. m.*
HTR2C	5-hydroxytryptamine (serotonin) receptor 2C, G protein-coupled	
IGF2	insulin-like growth factor 2 (somatomedin A)	
IGF2R	insulin-like growth factor 2 receptor	
IL1RAP	interleukin 1 receptor accessory protein	
JPX	JPX transcript, XIST activator (non-protein coding)	
KCNQ1OT1	KCNQ1 opposite strand/antisense transcript 1 (non-protein coding)	
KDM1A / LSD1	lysine (K)-specific demethylase 1A	
Kdm2	Lysine (K)-specific demethylase 2	*D. m.*
LEPREL1	leprecan-like 1	
LNX3	ligand of numb-protein X 3 [annotated as LOC422320]	*G. g.*
MACF1	microtubule-actin crosslinking factor 1	
MECP2	methyl CpG binding protein 2 (Rett syndrome)	

Gene symbol	Gene name	Organism
MIX-1	MItosis and X associated-1	*C. e.*
MLL	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)	
MTERF	mitochondrial transcription termination factor	
MTF2 / M96	metal response element binding transcription factor 2	
MYT1	myelin transcription factor 1	
NANOG	Nanog homeobox	
NCAPH / BRRN1	non-SMC condensin I complex, subunit H	
NCAPD2 / CNAP1	non-SMC condensin I complex, subunit D2	
NCAPG	non-SMC condensin I complex, subunit G	
NXF1 / TAP	nuclear RNA export factor 1 [tip associating protein]	
PARP1	poly (ADP-ribose) polymerase 1	
Pc	Polycomb	*D. m.*
Pcl	Polycomblike	*D. m.*
Pgk1	phosphoglycerate kinase 1	*M. m.*
PGK1	phosphoglycerate kinase 1	
Ph	Polyhomeotic	*D. m.*
PHF1	PHD finger protein 1	
PHF19	PHD finger protein 19	
PHF8	PHD finger protein 8	
Pol II	polymerase (RNA) II (DNA directed) [a multiprotein complex]	
PRMT1	protein arginine methyltransferase 1	
PRMT5	protein arginine methyltransferase 5	
Psc	Posterior sex combs	*D. m.*
RBBP5	retinoblastoma binding protein 5	
RBP2	retinol binding protein 2, cellular	
RBX1	ring-box 1, E3 ubiquitin protein ligase	
RCOR1 / CoREST	REST corepressor 1	
REST	RE1-silencing transcription factor	
Ring / Sce	Really interesting new gene / Sex combs extra	*D. m.*
RNF2	ring finger protein 2	
roX1	RNA on the X 1	*D. m.*
roX2	RNA on the X 2	*D. m.*
SATB1	SATB homeobox 1	
SATB2	SATB homeobox 2	

Gene symbol	Gene name	Organism
SDC1	syndecan 1	
SDC-1 – SDC-3	Sex determination and Dosage Compensation effect 1 – 3	*C. e.*
SETD7	SET domain containing (lysine methyltransferase) 7	
SETD8 / PR-SET7	SET domain containing (lysine methyltransferase) 8	
SIRT1 - SIRT7	sirtuin 1 - 7	
SLC22A2	solute carrier family 22 (organic cation transporter), member 2	
SLC22A3	solute carrier family 22 (extraneuronal monoamine transporter), member 3	
SMC2	structural maintenance of chromosomes 2	
SMC4 / SMC4L1	structural maintenance of chromosomes 4	
SMCHD1	structural maintenance of chromosomes flexible hinge domain containing 1	
SPOP	speckle-type POZ protein	
SRSF1 / ASF / SF2	serine/arginine-rich splicing factor 1	
SUV39H1	suppressor of variegation 3-9 homolog 1 (Drosophila)	
SUV39H2	suppressor of variegation 3-9 homolog 2 (Drosophila)	
SUV420H1	suppressor of variegation 4-20 homolog 1 (Drosophila)	
SUV420H2	suppressor of variegation 4-20 homolog 2 (Drosophila)	
SUZ12	suppressor of zeste 12 homolog (Drosophila)	
TP53	tumor protein p53	
TSIX	TSIX transcript, XIST antisense RNA (non-protein coding)	
WDR5	WD repeat domain 5	
XIST	X (inactive)-specific transcript (non-protein coding)	
Xite	X-inactivation intergenic transcription elements	*M. m.*
YY1	YY1 transcription factor	

ACKNOWLEDGEMENTS

The path to this thesis was much more exciting then I could have imagined at the beginning. My greatest thanks go to my supervisor Dr. Carolyn Brown for being an approachable advisor, a true role model and a caring person. I also wish to thank my advisory committee members, Dr. LeAnn Howe, Dr. Louis Lefebvre and Dr. Wendy Robinson for their encouragement, novel insights and time.

Many colleagues kindly offered their help, advice and support when I needed it. I wish to thank the members of the Brown lab that I was fortunate to meet and work with. In particular, Sarah Baldry contributed to many of the experiments that form my thesis. My sincere thanks go to Samuel Chang, Danny Leung and many other bright and inspiring colleagues.

I am indebted to my family. To my parents for supporting me in whatever I have done, my wife Adela for giving up so much by joining me in Vancouver, and my daughter Emma, the greatest source of joy.

What I learned in course of my studies culminating in this thesis goes much beyond its scientific topic and I am grateful to all of you who helped me on this journey.

DEDICATION

To Adéla and Emma.

1 INTRODUCTION

The candidate (Jakub Minks) co-authored three review papers with other members of the Brown lab that contributed to the views presented in this section. However, with the exception cited in the Preface, the text of this review is novel.

1.1 Thesis overview

The transcriptional silencing of a whole X chromosome in the process of X inactivation is a classic example of epigenetic regulation of gene expression, a set of mechanisms responsible for setting up of chromatin features that ensure correct utilization of genetic information in each cell in the temporal and spatial context, and their maintenance through cell divisions. The critical importance of epigenetic regulation in development, genomic imprinting, disease and cancer has been recognized in recent years (reviewed in [1-4]).

The discovery of the pivotal role of the *XIST* (X-Inactive Specific Transcript) gene more than twenty years ago was a breakthrough in X inactivation research. However, despite the tremendous progress in understanding *XIST* gene regulation, XIST RNA function and the protein componentry involved in XIST-mediated silencing of the X, the precise mechanism of XIST action is still missing. With the expanding knowledge of the composition and expression of mammalian genomes, many other regulatory long non-coding RNAs (lncRNAs) have been identified. Strikingly, explorations of the function of these RNAs show a common theme - their XIST-like ability to interact with one or multiple proteins, often chromatin modifying complexes, and thus facilitate locus-specific regulation of gene expression.

In this thesis, we use a doxycycline-inducible transgenic system to probe which region of the 17 kb-long XIST RNA is essential to induce gene silencing by creating a series of truncated *XIST* constructs and testing their ability to silence fluorescent reporters. We further dissect which sequence features within this region are responsible for XIST-induced silencing. Finally, to identify XIST-interacting partners responsible for gene silencing, we probe in detail the role of histone deacetylases and polycomb repressive complex 2, and perform a siRNA-mediated knock-down of 31 candidate proteins.

Our findings not only expand the knowledge of critical elements within the XIST sequence, its structure and interacting proteins in humans, but also provide insights into the mechanisms of action of an expanding family of regulatory lncRNAs.

1.2 X-chromosome inactivation and the X-inactivation centre

X-chromosome inactivation ensures equal dosage of X-linked genes in XY males and XX females in placental mammals. One of the two female Xs is randomly committed to silencing early in development, and all subsequent daughter cells inherit the same silent X [5]. The precise timing of initiation of X inactivation varies among species and while limited data on early events of human X inactivation exist [6, 7], the timing has only been well explored in mouse (reviewed in [8]).

X inactivation has far-reaching implications in clinical genetics. Skewed X inactivation, a relatively common and benign phenotype defined by predominant inactivation of either paternal or maternal X, can dramatically influence the severity of clinical outcome of X-linked diseases in female heterozygotes (reviewed in [9]). Skewed X inactivation has also been associated with increased frequency of recurrent spontaneous abortions, premature ovarian failure, and trisomic pregnancies (reviewed in [10]).
X inactivation also dramatically affects the clinical manifestation of X-chromosome aneuploidies, aneusomies and X:autosome translocations (reviewed in [11]). Thus, the knowledge gained in process of elucidating the molecular workings of X inactivation not only sheds light on this fascinating biological phenomenon, but also contributes substantially to understanding of manifestation of X-linked disorders. Similarly, studies of chromatin changes that accompany the transcriptional silencing of the X chromosome by XIST/Xist in the course of X inactivation have been at the fore front of epigenetics generally, and more specifically, the research of transcriptional regulation by lncRNAs.

The restriction of X inactivation to a single chromosome implies a *cis*[1]-regulated process. Studies of X/autosome translocations defined the X-inactivation centre (*XIC*/*Xic*) as a region of the chromosome capable of inducing inactivation [12]. Subsequent molecular analysis of X-chromosome rearrangements refined the location of the human *XIC* to Xq13 [13]; the *Xic* maps to the syntenic region in mouse [14].

As detailed in the following section, the key insight into the function of *XIC/Xic* came with the discovery of the lncRNA XIST/Xist[2]. Apart from *XIST*/*Xist*, the *XIC/Xic* harbors several other non-coding genes that are involved in regulation of proper *XIST*/*Xist* expression (reviewed in [15]).

Much of our knowledge of the *cis*-acting regulators of *Xist* expression that are found within the *Xic* have come from gene deletion studies (reviewed in [16]). In mice, it has been shown that *Tsix*, a gene antisense to *Xist* is expressed only early in development and is regulated by two enhancer elements, *Xite* and *DXPas34* [17, 18]. These elements suppress *Xist* up-regulation in *cis* [19]. The role of human TSIX in *XIST* regulation is however unclear as unlike in mice, the human TSIX transcript does not reach the *XIST* promoter [20]; moreover *XIST* and *TSIX* are expressed from the same X chromosome [21]. Adding to the complex regulation of *XIST*/*Xist* are the recent discoveries of positive regulators of the *Xist* transcription, non-coding RNAs Ftx and Jpx, that are transcribed from the region located 5' of *Xist* promoter [22, 23] and repA, a transcript originating from within 5' end of *Xist* [24].

---

[1] We use *"cis*" or "in *cis*" to denote regulation restricted to the same chromosome, while "*trans*" or "in *trans*" denotes regulation either on both copies of a chromosome, or when the regulating transcript originates from a chromosome other than the chromosome of the regulated gene.
[2] We use the following convention when discussing XIST. We do not italicize XIST (as RNA), as it is the actual gene product. We use *Xist*/Xist when referring specifically to the mouse gene/RNA. When capital letters are used (*XIST*/XIST), we refer to either specifically the human, or in general a mammalian gene/RNA.

Ftx (five prime to *Xist*) was shown to up-regulate *Xist* expression through a yet-unknown mechanism which may involve local changes to chromatin structure [23]. Interestingly an intron of *Ftx* harbors two microRNAs: miR-374 and miR-421 [23]. miR-421 was shown to target a cell cycle-regulating kinase ATM [25]. *Jpx* encodes a trans-acting non-coding RNA that is induced at the onset of X inactivation [22]. Similar to Ftx, how Jpx expression leads to *Xist* up-regulation remains to be elucidated. RepA is a 1.6 kb-long non-coding RNA that spans the 5' region of *Xist*, recruits PRC2 via repeat A sequences (see secton 1.3.1) and up-regulates *Xist* expression. The mechanism of *Xist* up-regulation by RepA is also currently unknown, but the up-regulation is accompanied by local recruitment of H3K27me3.

The long-anticipated *trans*-regulatory elements were also mapped to the *Tsix* region when the X chromosomes were demonstrated to pair at the onset of X inactivation in a process that requires the chromatin regulator CTCF and transcription factor YY1 [26-28]. In addition, another region located 200 kb 5' to *Xist* and termed *Xpr* (X-pairing region) also showed inter-allelic pairing [29]. Thus, X inactivation in mammals requires expression of *Xist* which is delicately balanced by a surrounding group of non-coding RNAs, serving as either activators or repressors to ensure that the dosage of X-linked genes remains constant between males, females or even cells with an aberrant count of X chromosomes.

X-chromosome inactivation in placental mammals is not the only example of dosage compensation of sex chromosomes. Recently, a long non-coding RNA Rsx (RNA specific to X) with XIST-like properties has been described in a marsupial *Monodelphis domestica* [30]. Similar to XIST, Rsx harbors a 5' repeat-rich region, is able to localize in *cis* and induces gene silencing upon expression [30]. In *D. melanogaster*, the process involves two fold upregulation of X-linked genes in males carried out by a complex consisting of X-linked non-coding RNAs expressed only in males, roX1 and roX2 (RNA on X 1 and 2) and a set of proteins responsible for deployment of H3K16ac, an active chromatin mark (reviewed in [31]). In *C. elegans*, dosage compensation is achieved by two fold repression of X-linked genes in XX females by a condensin-like dosage compensation complex (reviewed in [32]). In conclusion, despite the diverse approaches to dosage compensation, different organisms frequently employ non-coding RNAs to equalize sex-linked gene dosage.

1.3 XIST/Xist

XIST is exclusively transcribed from the inactive X (Xi) [33], coats the whole X chromosome from which it is transcribed and is indispensable for its transcriptional silencing [34, 35]. *XIST* encodes an approximately 17 kb-long, Pol II-transcribed, spliced and polyadenylated RNA that contains no open

reading frames of significant length and is thus presumed to be non-coding. Here, we review the previously published reports on *XIST* evolutionary conservation, as well as structure and function of sequences within XIST/Xist and other lncRNAs. We also discuss another intriguing feature of XIST/Xist – its nuclear localization.

1.3.1 Evolutionary sequence conservation of *XIST*

Sequencing of human *XIST* [34] and mouse *Xist* [36] revealed that while the overall structure of the gene is similar (diagrammed in Figure 1.1), the primary sequence in general is not very well conserved (49 percent identity) [37]. Human *XIST* consists of eight exons. The 11 kb-long exon 1 and 4.6 kb-long exon 6 account for the most of *XIST* cDNA sequence; the remaining exons amount to approximately 1 kb in total [34]. Splicing variants of full-length XIST RNA that exclude exon 3, 4 or 7 have been described, along with truncated transcripts that lack fragments of exon 6 or that are terminated within exon 6 [34]. The limited degree of homology between exons 4 and 5 of *XIST* (66 and 59 percent identity, respectively) and chicken protein coding gene *Lnx3,* a member of *LNX* (Ligand of Numb Protein) E3 ubiquitin ligase gene family, suggests that *XIST* is a derivative of *Lnx3* [38, 39].

A distinct feature of the *XIST* gene is its enrichment for tandem repeat sequences named repeat A-F (Figure 1.1); exon 1 harbors all of *XIST*'s repeats, with the exception of repeat E located in exon 6 [36-38, 40, 41]. Repeat A, located approximately 1 kb 3' from XIST transcription start site, consists of extremely well conserved CG-rich core palindromes separated by stretches of T-rich sequence and its composition is extensively discussed in Section 4. Repeat B is a small microsatellite C-rich tract broken by an inserted sequence in primates. In dog, the order of repeat B and C is inverted. Repeat C, a 14-fold repeat of an 11-bp-long monomer is murine specific; humans only contain 1 copy and mole and cow lack repeat C altogether. Repeat D has a monomer length of 290 bp and its sequence is only moderately conserved (64 percent identity) [40]. It takes up a substantial part of exon 1 in many species, but is short in rodents when compared to other mammals. The repeat E monomer is 14-30 bp-long and consists of a CT-rich tandem repeat, a simple TG dimer repetition and a species-specific sequence. While it has been found in all species surveyed so far, repeat E shows the lowest sequence conservation. A 16 bp-long repeat F located downstream of repeat A is present in 5 copies in voles, but only two copies in mouse and human.

While the tandem repeat-rich structure may plausibly be important for XIST function, with the exception of repeat A, substantial differences in size and composition of these repeats exist among species with apparently functional XIST. Comparison of overall *XIST* conservation in 10 mammalian species

uncovered three exceptionally conserved regions: the *Lnx3*-derived exon 4, a region spanning the very 3' end of exon 1 and the repeat A (96, 94 and 91 percent identity, respectively) [40].

1.3.2 Functional sequences within XIST/Xist

In order to achieve chromosome-wide silencing, the XIST RNA is able to perform two remarkable tasks. It is able to cover the whole X chromosome in *cis* without interacting with other chromosomes or delocalizing from the X and it is able to induce silencing of the vast majority of X-linked genes. Clearly, both of these properties have to be embedded in the XIST sequence and analyses of truncated *XIST*/*Xist* constructs have been instrumental in delineating these sequences.

Wutz *et al*. constructed an extensive panel of 50 *Xist* cDNA fragments with both internal and terminal deletions whose expression was controlled by a DOX inducible promoter [42]. The constructs were integrated into the *Hprt* locus in male embryonic stem (ES) cells and thus all the inducible transgenes were single copy and in the same, known, X-linked site. Upon differentiation of the transgenic ES cells containing the full-length *Xist* cDNA in the presence of DOX, Xist was shown to localize to the X and silence the distant *Pgk1* gene, demonstrating a long-range silencing effect. The different constructs were assayed for their ability to localize Xist to the single X and for silencing efficiency, measured by cell lethality rate. The *Xist* transgene lacking only 900 bp of 5' sequence encompassing the repeat A was able to localize, yet silencing was completely abolished. When the 900 bp was moved to the 3' end of *Xist*, silencing was restored, which might suggest that secondary, rather than tertiary structure is important for proper Xist function – or more specifically, that XIST consists of semiautonomous protein binding domains connected by linker RNA sequence. This is in line with a recently proposed hypothesis that the function of many lncRNAs depends upon their ability to interact with various protein complexes through sequence modules that appear in different combinations to target appropriate components of the chromatin modifying machinery to specific chromatin loci [43]. Remarkably, a similar deletion of the repeat A region in human showed normal transcript levels but failed to localize [44]. While this discrepancy may suggest that a yet unknown protein or proteins take part in Xist RNA aggregation by binding to sequences 3' of repeat A in mice, and that such binding does not occur in the human cells, it is currently not clear whether the differing observations indicate a species- or cell-type specific effect. A mouse model of the repeat A deletion validated the need for the repeat A region for X inactivation; notably, *Xist* expression was greatly down-regulated [45]. This observation is in line with a previous report on a 1.6 kb-long non-coding transcript, named RepA, which originates from the repeat A region, interacts with PRC2 and facilitates *Xist* up-regulation [24].

Domains involved in localization were not as clearly delimited [42]. The 5' sequence showed some localization and silencing activity, both of which were greatly enhanced in the presence of at least two out of the three more distal regions proposed to be important for localization. In the absence of repeat A, most of the 3' sequence was needed for proper localization. Therefore the interaction of Xist with chromatin is achieved in part by the repeat A, and partly by redundant regions further downstream. The construct lacking the repeat A has shown that a surprising number of features of the Xi heterochromatin are recruited independently of genic transcriptional silencing, implying that Xist has multiple roles in the establishment of a silent domain. In the absence of the repeat A, Xist is still able to recruit macroH2A, H2AK119ub1, H3K27me3, H4K20me1 and form a transcriptionally repressed domain [42, 46-49].

1.3.3 Sequence and structure of other regulatory long non-coding RNAs

Two decades after the discovery of the first lncRNAs, H19 and XIST [33, 50], it is now becoming evident that non-coding RNAs are in fact abundant. Aside from the non-coding structural (*e. g.* tRNAs, rRNAs) and small regulatory RNAs (miRNAs, piRNAs), there are many non-coding transcripts whose function is not yet clear (reviewed in [51, 52]). Of particular interest for the study of XIST function are other nuclear lncRNAs, which can function either in *cis*, *i. e.* in the genomic region of their transcription (XIST, AIR, Kcnq1ot1 and HOTTIP) or in *trans*, *i. e.* elsewhere in the genome (HOTAIR) (reviewed in [53]). Interestingly, while the *cis*-limited effect of XIST/Xist is long-established, a recent report showed that endogenous Xist transcripts can trans-migrate and localize to a multi-copy *Xist* transgene integrated on an autosome [54]. It is however not clear to what extent the trans-migration also occurs during normal X inactivation. Regulation of *Xist* itself involves, at least in mice, an antisense non-coding RNA Tsix [55]. A transcript overlapping the repeat A region and dubbed repA has also been reported to regulate *Xist* expression in mice [24]. The current view of events leading to proper monoallelic expression of *XIST*/*Xist* in females has recently been extensively reviewed for the 50th anniversary of Lyon's hypothesis (*e. g.* [15]).

HOTAIR is a lncRNA expressed from the *HOXC* (homeotic genes cluster C) gene cluster that acts in *trans* to suppress *HOXD* genes [56]. The approximately 2 kb-long RNA interacts with two distinct protein complexes via domains located at the opposite ends of HOTAIR: the PRC2 component EZH2 interacts with the 5' end, while the 3' end binds REST/CoREST via LSD1 [56, 57]. In cancer cells, an increase in HOTAIR expression correlates with stronger cell invasiveness and poorer prognostic outcomes, caused by epigenetic chromatin reprogramming due to altered PRC2 targeting [58].

HOTTIP is a novel member of the *HOX* gene-regulating lncRNAs [59]. Located at the distal, 5'end of *HOXA* cluster and transcribed from the opposite DNA strand than the *HOXA* genes, HOTTIP forms a 3.8 kb-long, spliced and polyadenylated lncRNA. HOTTIP maintains expression of the neighboring *HOXA* genes with a distance-dependent decrease in effect over the span of 40 kb. A 1 kb-long region within the 5'end of HOTTIP directly interacts with WDR5 to recruit MLL-containing complexes that, in turn, deploy the H3K4me3 activating histone mark. Chromosome conformation capture experiments and the use of transgenes showed that HOTTIP function requires physical interaction with the genes it activates. Apart from HOTTIP, the *HOXA* genomic region harbors another lncRNA, a myeloid lineage-specific HOTAIRM1, which is transcribed in antisense orientation from the CpG island within the *HOXA1* promoter [59, 60].

Air is a more than 100 kb-long unspliced nuclear RNA expressed in antisense orientation from an intronic CpG island near the 3' end of a maternally expressed imprinted gene, *Igf2r*, in mice [61, 62]. Air induces gene silencing by two distinct mechanisms (reviewed in [53]): in the embryo proper, *Air* is transcribed across the *Igf2r* promoter and induces its silencing by DNA methylation. In extraembryonic tissues of the placenta, in addition to *Igf2r* repression, Air silences *Slc22a2* and *Slc22a3*, two other maternally expressed *cis*-linked genes located several hundred kb upstream of *Air*. Air was shown to silence *Slc22a3* by recruiting the G9a histone methyltransferase that deploys the H3K9me3 histone mark [63].

The Kcnq1ot1 lncRNA regulates monoallelic expression of a cluster of maternally expressed genes surrounding *Kcnq1* in mice [64]. Kcnq1ot1 is thought to establish a chromatin compartment lacking RNA Pol II and recruit G9a, PRC2, PRC1 and DNMT1, as reviewed in [53]. A recent study showed that *Kcnq1ot1* transcription *per se*, and not the transcript might be sufficient to maintain the imprinting within the Kcnq1 domain [64]. The Kcnq1ot1 RNA was demonstrated to span 500 kb, although stable shorter transcripts have previously been documented [65]. Another study suggests a substantially extended size of the imprinted domainunder Kcnq1ot1 regulation [66]. Similar to XIST/Xist, Kcnq1ot1 harbors a 5' silencing domain and a localization domain that contains an evolutionarily conserved motif predicted to form a stem-loop and facilitate nucleolar localization of the Kcnq1 domain [67]. Interestingly, attaching the sequences involved in Xist localization 3' to the Kcnq1ot1 silencing domain fully reconstituted the Kcnq1ot1 silencing ability [67]. Thus it is likely that Xist/XIST and Kcnq1ot1 share some mechanisms to induce gene silencing and advances in understanding of their function will therefore be mutually informative.

1.3.4 Nuclear localization of XIST/Xist in the context of RNA metabolism

mRNA precursors and some snRNAs and microRNAs in metazoa are transcribed by RNA Pol II. As Pol II transcription proceeds, the nascent RNAs are spliced in spliceosomes to excise introns and capped at the 5' end, as well as cleaved and polyadenylated at the 3' end to prevent degradation. The capping and splicing promotes export of Pol II transcripts to the cytoplasm, as some of the components of the mRNA processing machinery remain bound to transcripts and recruit TAP/NXF1 RNA export receptor which facilitates interaction with nuclear pores. mRNAs that fail to correctly process the 3' end are degraded in exosomes (reviewed in [68, 69]). Following the export to cytoplasm, mRNAs interact with ribosomes to serve as a template for translation and are eventually degraded via several ribonucleolytic pathways [70].

Experiments utilizing 3' RACE (rapid amplification of cDNA ends) and cDNA sequencing revealed that similar to other Pol II transcripts, XIST/Xist RNA is spliced and polyadenylated, features typical for protein-coding mRNAs [34]. However unlike protein coding RNAs, XIST/Xist is localized exclusively in the nucleus [34, 36, 71]. Thus, XIST/Xist is either actively retained in the nucleus, or it must be efficiently shuttled back to the nucleus from the cytoplasm; it is however unclear how *cis* localization would be regulated in the latter scenario. Two approaches were used to exclude the possibility of transient cytoplasmic XIST presence [72]. First XIST RNA did not shuttle through cytoplasm between nuclei in a heterokaryon assay, which utilizes cells that harbor more than one nucleus. Second, transcription of a fusion XIST-*GFP* RNA did not produce GFP protein, suggesting that the fusion RNA was not present in cytoplasm to allow for translation. The apparent lack of XIST export from the nucleus was corroborated by an immunoprecipitation assay that demonstrated attenuated interaction of XIST with complexes involved in mRNA splicing and export [72]. Thus, rather than harboring a specific signal for nuclear localization, XIST/Xist may avoid export from the nucleus by using an alternative RNA processing mechanism. However, it is not currently known whether XIST/Xist is also actively retained in the nucleus and how this is achieved.

1.4 Proteins implicated in X inactivation

The conserved size and the repetitive nature of *XIST/Xist* suggest that it may serve as an adaptor that links multiple components of the gene-silencing machinery and the Xi. Indeed, a number of proteins and protein complexes that interact with XIST/Xist RNA have been identified to date [8], but direct interaction with Xist has only been observed for the components of PRC2 [24, 73], splicing factor SRSF1 (previously known as ASF/SF2) [74] and a transcriptional repressor YY1 [54]. While it is likely that multiple factors are involved in the complex act of spreading XIST/Xist RNA along the X and

silencing most, but not all, X-linked genes, we predominantly focus on those that have shown an evidence of specific interaction with the Xi or XIST/Xist (Figure 1.1).

1.4.1 Polycomb complexes

1.4.1.1 PRC2

Polycomb repressive complex 2 is intimately involved in X-chromosome inactivation. Not only is the H3K27me3 histone modification it deploys enriched on the Xi, multiple experimental approaches have shown that repeat A of Xist/XIST in both mouse and human directly interacts with PRC2 components. Given its prominent place in X inactivation research, and our own studies of PRC2 involvement in *XIST*-induced gene silencing, we discuss the role of PRC2 in X inactivation in detail in Section 5.

1.4.1.2 PRC1

The PRC1 core complex in *Drosophila*, where it was first discovered, consists of PC, PH, PSC and RING proteins, each of which has 2-6 known mammalian homologs, and functions as a ubiquitin E3 ligase in deployment of a repressive chromatin mark H2AK119ub1 (reviewed in [75]). Classically, the PRC1 is thought to be recruited to chromatin in *Drosophila* by a chromodomain of the *polycomb* protein, which recognizes H3K27me3 mark deployed by PRC2.

Recently, a newly discovered RYBP-PRC1 complex was shown to deploy H2AK119ub1 to a largely overlapping set of genomic sites in PRC2-deficient mouse ES cells [76]. Another PRC1-like complex, dRAF, consists of RING, PSC and KDM2, the latter of which is a demethylase of the H3K36me3 active histone mark. This complex has been shown to be responsible for the most of the H2A ubiquitinylation activity in flies [77].

In *Drosophila*, mutation of the chromodomain of the polycomb gene leads to body segment transformation due to an aberrant expression of *HOX* genes. In mouse ES cells, PRC1 occupies the promoters of more than 1000 genes, most of which are involved in the regulation of development, and the majority of which harbor CpG islands and show ‘bivalent’ H3K4me3/H3K27me3 marks early in development [78]. The precise mechanism of PRC1 action is not fully understood, but involves chromatin compaction [79] and inhibition of transcriptional elongation [80, 81], as reviewed in [75] and [82].

The role of PRC1 in X inactivation was discovered through observations that PRC1 components show transient enrichment on the emerging Xi in mouse trophoblast stem cells, differentiating ES cells, embryos and embryonic fibroblasts as well as differentiated human HEK293 cells [83, 84]. The global loss of H2AK119ub1 enrichment in RING1B-deficient cells and similar loss of Xi-specific H2AK119ub1 enrichment in RING1A/B double knock-outs demonstrated for the first time that PRC1 is the complex responsible for H2A ubiquitinylation [83]. PRC1 deficient mouse ES cells are capable of initiation and maintenance of X inactivation [85]. Mouse ES cells lacking functional PRC2 can recruit RING1B capable of ubiquitinylating H2AK119 on the Xi [49]. While this finding was surprising at the time, it is in line with the discovery of the RYBP-PRC1 complex that is recruited to chromatin independent of PRC2 [76].

The interaction of *Drosophila* PRC1 complex with chromatin is facilitated by the chromodomain of the PC protein, which binds exclusively to chromatin marked by H3K27me3 [86]. Polycomb has five homologs in mammals (CBX2, 4, 6, 7 and 8), all of which were shown to co-localize with the Xi; a fusion CBX4-EGFP protein failed to localize to the Xi in mouse ES cells, however endogenous CBX4 showed Xi accumulation in HEK293 cells [84, 87]. Other CBX proteins, CBX1, CBX3 and CBX5 were previously known as HP1β, HP1γ, and HP1α, respectively, and interact with H3K9me3 (reviewed in [88]). Mammalian homologs of *polycomb* differ in their binding preferences. In an *in vitro* peptide-binding assay, CBX2 and CBX7 bound equivalently H3K9me3 and H3K27me3, while CBX4 preferentially interacted with H3K9me3. CBX6 and CBX8 failed to bind the methylated histone H3 tails [87]. Chromodomains of all CBX proteins were able to non-specifically bind single-stranded RNAs, including an Xist fragment, with the exception of CBX2 which was proposed to bind nucleic acids via a different domain [87]. The role of RNAs in PRC1 recruitment to the Xi was further strengthened by the observation that depletion of single-stranded RNA in cells resulted in a loss of CBX7 enrichment on the Xi [87].

1.4.1.3 PCL2

Polycomblike (PCL), a well-conserved *Drosophila* protein and its mammalian homologues (PCL1/PHF1, PCL2/MTF2 and PCL3/PHF19) were shown to interact with the PRC2 complex and facilitate its localization to target genes [89-93]. The analysis of *Pcl2* deletion transgenes established that one of the two PHD domains is responsible for the PRC2 targeting [93]. PCL2 transiently co-localized with the Xi in differentiating mouse ES cells and embryos at early stages of X inactivation coinciding

with PRC2 enrichment on the Xi. Knock-down of PCL2 impaired recruitment of PRC2 both to its target loci in undifferentiated ES cells and to the Xi upon differentiation [93].

1.4.2 Writers, readers and erasers of chromatin marks

1.4.2.1 ASH2L

ASH2L is a core component, along with WDR5, DPY30 and RbBP5, of several complexes associating with SET (Su(var)3-9, Enhancer of zeste, Trithorax) domain-containing methyltransferases [94], as well as a non-SET domain multi-subunit methyltransferase WRAD [95], that are required for trimethylation of histone H3K4, a mark associated with actively transcribed promoters [96]. Unexpectedly, ASH2L has also been shown to associate with the Xi in mouse [97]. The recruitment of ASH2L requires *Xist* expression, but is independent of repeat A or the presence of functional polycomb complexes [97]. Intriguingly, a lncRNA HOTTIP, which is critical for recruiting H3K4me3 and establishing upregulation of several genes within HOXA cluster was shown to directly interact with WDR5 [59].

1.4.2.2 H3R17 histone methyltransferase CARM1

CARM1 is one of the three known mammalian methyltransferases catalyzing mono- and dimethylation of histone arginine residues (reviewed in [98]).While asymmetric dimethylation of arginine deployed by class I arginine methyltransferases CARM1 and PRMT1 is associated with transcriptional activation, symmetric dimethylation by class II enzyme PRMT5 leads to gene repression. CARM1 predominantly methylates H3R17, and to a lesser extent H3R2 and H3R27, while PRMT1 targets the H4R3 residue. PRMT5 symmetrically methylates H3R8 and H3R3. CARM1 has been shown to play a role in nuclear receptor signal transduction and chromatin remodeling [99, 100] and in the TP53-mediated DNA damage response pathway [101]. Immunofluorescence has shown that H3R17 methylation is depleted from the Xi in mouse embryonic fibroblasts [102], however no H3R17 demethylase has been described to date.

1.4.2.3 H3K9 histone methyltransferases

EHMT1 (GLP) and EHMT2 (G9a) are histone methyltansferases that form heteromeric complexes to catalyze mono- and dimethylation of H3K9 [103, 104], histone marks associated with transcriptionally silent euchromatin. H3K9me2 is enriched on the Xi [105, 106] and G9a was shown to interact with Kcnq1ot1 lncRNA in mouse placenta [65] and to be implicated in the placenta-specific imprinting of distal genes within the Kcnq1 domain [107]. Notably, the de-repression of Kcnq1ot1-silenced targets

was not observed in all G9a-defficient progeny, suggesting that alternative pathways are also at play. While the cited findings make G9a a potential candidate for an XIST-interacting protein, X-inactivation maintenance was unperturbed in G9a-deficient mouse embryos [108].

In contrast, to the H3K9me1- and H3K9me2-enriched facultative heterochromatin, constitutive heterochromatin is marked by H3K9 trimethylation, which is in mammals carried out by SUV39H1 and SUV39H2 and facilitates silencing by recruiting HP1 proteins CBX1, CBX3 and CBX5 (reviewed in [88]). Immunofluorescence microscopy in somatic human cells revealed that the Xi is compartmentalized into H3K9me3-enriched regions and H3K9me3-poor, but H3K27me3-rich regions [109]. The H3K9me3-rich regions were also enriched with H4K20me3 and CBX3 (HP1γ) and replicated relatively late compared to H3K27me3-enriched regions, which in turn associated with Xist RNA accumulation and enrichment with macroH2A. Ectopic *XIST* expression was shown to induce CBX3 recruitment to a reporter gene promoter, further supporting a role for H3K9me3 in X inactivation [44].

1.4.2.4 H4K20 histone methyltransferase PR-SET7

The H4K20 monomethylation in mammals is deployed by the PR-SET7 (SETD8/KDM5a) histone methyltransferase [110-112] and removed by the PHF8 demethylase. SUV420H1/H2 are responsible for di- and trimethylation of H4K20me; enzymes removing the higher methylation degrees of H4K20 are not known [113]. Levels of both PR-SET7 and consequently H4K20me1 oscillate during the cell cycle and are induced in late S and early G2/M, respectively [112]. PR-SET7-null mouse embryos die between 2-4 cell stage. PR-SET7-null mouse ES cells show defects in cell cycle and DNA damage repair [114].

An immunofluorescence screen utilizing an array of antibodies against histone modifications showed enrichment of H4K20me1, but not -me2 or -me3, following induction of an ectopically expressed transgenic *Xist* in undifferentiated mouse ES cells [47, 49]. Recruitment of H4K20me1 is in part dependent on PRC2 [49]. Importantly, expression of a repeat A-lacking *Xist* transgene that is unable to induce silencing also caused H4K20me1enrichment, demonstrating that H4K20me1 recruitment occurs independently of silencing [47]. Chromatin IP analysis confirmed H4K20me1 enrichment over the coding region of puromycin selection marker upon ES cell differentiation when the transgenic *XIST* was expressed [47], a phenomenon also observed for H3K27me3 mark deposited by PRC2 [48, 115].

1.4.2.5 LSD1

LSD1 is a histone demethylase that acts as a co-repressor by demethylating H3K4 [116]; it has also been shown to act as a co-activator by demethylating H3K9 [117]. The only other known mammalian H3K4 demethylase RBP2 interacts with PRC2 complex, and is thus involved in coordinated increase of H3K27me3 and removal of H3K4 methylation [118]. As LSD1 directly interacts with HOTAIR lncRNA [57] and the H3K4me3 histone mark is depleted from the Xi [106, 119], LSD1 may potentially be involved in the X-chromosome inactivation.

1.4.2.6 macroH2A

macroH2A bears similarity to histone H2A, but contains a unique C-terminal sequence comprising approximately 2/3 of the protein. There are three variants of macroH2A in humans: macroH2A1.1 and macroH2A1.2 are encoded by *H2AFY*, macroH2A2 is a product of *H2AFY2*. Although the macroH2A is involved in a context-dependent up- and down-regulation of autosomal gene expression and regulation of cell cycle and cell proliferation [120], its role in X inactivation has been explored more extensively. Shortly after its discovery, macroH2A was shown to form prominent foci in female cell nuclei that were dubbed macrochromatin bodies [121]. Successful chromatin IP of XIST RNA in human HEK293 with an antibody against macroH2A demonstrated physical proximity of macroH2A and XIST [122]. A functional relationship was demonstrated when the lack of *Xist* expression was shown to result in loss of macroH2A recruitment to the Xi in mouse embryonic fibroblasts; the silencing of genes on the Xi was however unperturbed [123]. Conversely, ectopic expression of an inducible *Xist* results in macrochromatin body formation in differentiating mouse ES cells and embryonic fibroblasts. In contrast, lack of macrochromatin body recruitment in undifferentiated ES cells suggests that the environment permissive to macroH2A recruitment is absent prior to the initiation of X inactivation [124]. The role of macroH2A in silencing is further supported by chromatin IP experiments showing depletion of macroH2A from active genes and its enrichment on the CpG methylated alleles of imprinting control regions [125, 126].

1.4.2.7 DNMTs

In mammals, DNA methylation of cytosine in CpG dinucleotides serves as a chromatin mark that is associated both with transcriptionally silent promoters of CpG island-containing genes and with gene bodies of transcribed genes. Three DNMT enzymes are active in mammalian cells; DNMT1 maintains CpG methylation through cell division by binding hemimethylated CpG sites and methylating the newly

synthesized strand of DNA, while DNMT3A and DNMT3B are *de novo* methyltransferases with only partially overlapping functions that establish methylation patterns during development (reviewed in [127]).

Consistent with the association of DNA methylation with silent promoters and transcribed gene bodies, gene promoters on the Xi show DNA hypermethylation while the Xi is relatively hypomethylated overall [128, 129]; genes that escape from X inactivation and thus remain transcribed on the Xi accordingly lack promoter CpG methylation [130]. DNA methylation, alongside macroH2A recruitment, is acquired relatively late in X inactivation [131] and is dependent on SMCHD1 [132]. Disruption of DNA methylation leads to partial upregulation of genes on the Xi and this effect is substantially compounded by inhibition of *XIST*/*Xist* expression or by blocking of histone deacetylation [133, 134]. A study performed in DNMT3B-defficienct cells from patients with ICF syndrome, revealed that DNMT3B is responsible for methylation of LINE-1 (long interspersed nuclear element-1) repeats on the inactive X, but not on the active X [135]. DNMT1-deficcient mouse embryos showed partial re-activation of X-linked lacZ and EGFP transgenes, suggesting that unlike DNMT3B, DNMT1 may predominantly play a role in gene repression [134, 136].

1.4.3 Chromatin-remodeling and nuclear ultrastructure proteins

1.4.3.1 ATRX

A member of the helicase family, ATRX is a chromatin remodeling protein involved in heterochromatin formation and maintenance, as well as proper chromosome segregation in meiosis and mitosis [137, 138]. Several domains within ATRX are responsible for direct interaction with DAXX, HP1α, MeCP2 and EZH2, while the C-terminal sequence of ATRX encodes a domain involved in ATRX targeting to PML bodies (reviewed in [137, 138]). An immunofluorescence experiment showed that in the nucleus, ATRX associates with telomeric, rDNA and heterochromatic repeats, as well as PML bodies [139-141]. Chromatin IP followed by massively parallel sequencing in mouse and human cells elucidated that ATRX associates with G-rich tandem repeats and CpG islands both in the previously observed heterochromatic regions (telomeres) and in euchromatin [142]. Mutations in *ATRX* cause Alpha-thalassemia mental retardation syndrome [143] or myelodysplasia [144]; both conditions share a common symptom, alpha thalassemia, caused by suppression of alpha globin (*HBA*).

Interestingly, ATRX was shown to co-localize both with the Xi in mouse embryonic and somatic cells [145] and with the Y chromosome in mouse spermatogonia [146]. In differentiating mouse ES cells, ATRX is recruited to the Xi relatively late during mouse ES cells' differentiation [145], suggesting that it

may be involved in maintenance, rather than initiation of X inactivation. Further, ATRX was shown to bind upstream of *XIST* [145], as well as the unmethylated allele of the gene encoding H19 non-coding RNA [147]; a group of imprinted genes, including *H19* showed increased expression in ATRX-null mouse brains [147].

1.4.3.2 YY1 and CTCF

YY1 is a ubiquitous transcription factor that modulates activation or repression of gene expression by multiple direct and indirect mechanisms (reviewed in [148]). YY1 has both DNA and RNA binding capacity and is indispensable for Xist localization in mouse [54]. Specifically, YY1 directly interacts with *Xist* DNA via three binding sites upstream of repeat F in *Xist* exon 1 and with the Xist RNA via the repeat C region. Importantly, YY1 does not decorate the Xi, suggesting that it is not the factor responsible for recruitment of Xist RNA in *cis* along the whole Xi, rather, the authors propose that YY1 facilitates nucleation of Xist particles at *Xist* locus [54], from which these particles spread along the Xi via a yet unknown mechanism.

CTCF (CCCTC-binding factor) plays a central role in two aspects of chromatin regulation: globally, by maintaining chromatin architecture through regulation of chromatin looping, and locally by serving as a chromatin insulator, (reviewed in [149]). In mammals, between 14 000 and 20 000 CTCF binding sites have been identified genome-wide. Chromatin IP followed by massively parallel sequencing data combined with chromosome conformation capture analyses showed that CTCF mediates intra- and interchromosomal interactions. Chromatin looping is pivotal in insulating gene promoters from being upregulated by enhancers, a phenomenon best described at the *H19-IGF2* and betaglobin (*HBB*) loci [150, 151]. The current model presumes that CTCF regulates chromatin organization and insulation by recruiting the cohesin complex [152-155]. In X inactivation, CTCF is necessary for *XIC* pairing during initiation [28, 156]. CTCF and YY1 were further identified to regulate expression of *Xist*, both directly and via *Tsix* and its enhancer *Xite* (reviewed in [157]).

1.4.3.3 SAF-A / HNRNPU

Scaffold attachment protein A, also known as HNRNPU harbors an N-terminal dsDNA-binding domain and a C-terminal RGG domain that facilitates interaction with RNA [158] and has been implicated in various processes including gene expression and RNA metabolism [159-162] and telomere length regulation [163]. SAF-A has been shown to associate with the Xi in mouse and human HEK293 cells

[97, 164]. Like Ash2l, Saf-A recruitment to the Xi also requires *Xist* transcription, but not the repeat A region nor polycomb complexes [97]. Intriguingly for its potential role in regulation of X inactivation, deletion of either the DNA or the RNA binding domains results in the loss of SAF-A localization to the Xi [97, 164, 165]. The evidence of interaction between XIST/Xist and Saf-A is further strengthened by the observation that knock-down of Saf-A results in a loss of Xist localization to the Xi in mouse Neuro2a cell line and a failure to inactivate the X in differentiating mouse ES cells [165]. RNA immunoprecipitation suggests that Saf-A binds Xist in a region within exon 1 between repeats C and D that has previously been shown be involved in Xist localization [42]. Saf-A was also proposed to play a structural role in the formation of a repressive Xi compartment consisting of non-genic chromatin, into which genes on the Xi are relocated in the course of X inactivation [46].

1.4.3.4 SATB1 and SATB2

SATB1 and SATB2 proteins bind to AT-rich DNA sequences within matrix attachment regions and mediate their interaction with the nuclear matrix, thus ensuring the proper organization of chromatin. *SATB1* and *SATB2* are expressed in largely non-overlapping subsets of cell lines [166-168]. Loss of SATB1 in mouse cells impedes Xist's ability to induce silencing of the X, without affecting Xist localization [169]. In SATB1 expressing cells, Xist was frequently observed to form 'rings' around SATB1 foci, instead of showing the typical overlap with the Xi. Further, SATB1 is expressed in undifferentiated ES cells but is silenced within 3 days following differentiation. This timing coincides with the window in which XIST can induce silencing. SATB2 can substitute for SATB1, with depletion of either protein resulting in partial upregulation of a reporter gene. Furthermore ectopic expression of SATB1 in mouse embryonic fibroblasts, normally resistant to Xist-induced silencing, results in gene silencing upon Xist induction [169].

1.4.3.5 HNRNPK

As recently reviewed [170], HNRNPK binds C-rich RNA and single-stranded DNA regions via its three KH domains that are also present in HNRNPE, another member of the broadly defined and structurally divergent HNRNP protein family. HNRNPK also contains a K-protein-interactive region that facilitates binding with multiple kinases and transcription regulators. Thus, HNRNPK was proposed to act as a 'docking platform', mediating interaction between nucleic acids and multiple signaling pathways [171].

Mass spectroscopy analysis identified more than 100 proteins interacting with HNRNPK [172], in keeping with its involvement in a host of cellular processes including cell cycle regulation, DNA damage control and regulation of mRNA metabolism. HNRNPK interacts with the 5' end of TP53-activated lincRNA-p21 that is indispensable for repression of approximately 750 TP53-regulated genes [173]. While HNRNPK has not been shown to be involved in X inactivation, its proposed role in facilitating non-coding RNA – protein interaction substantiates its position among potential XIST-interacting partners.

1.4.4 Other proteins implicated in X inactivation

1.4.4.1 BRCA1

The *BRCA1* tumor suppressor gene encodes a RING-domain containing ubiquitin ligase that is involved in control of the cell cycle, maintenance of genomic integrity and transcriptional regulation (reviewed in [174, 175]). BRCA1 was reported to associate with XIST and to be essential for XIST RNA localization to the Xi, and knock-down of BRCA1 resulted in partial reactivation of an Xi-linked *EGFP* transgene [176, 177]. Other reports, however, did not confirm these observations [178, 179] and the role of BRCA1 in X inactivation remains unresolved. BRCA1 was shown to be necessary for recruitment of ATR kinase to the XY body during meiotic sex chromosome inactivation in mouse spermatocytes [180].

1.4.4.2 DICER1

DICER1 is an RNase III enzyme responsible for generation of siRNA and miRNA small RNA species from larger RNA templates [181]. These small RNAs are then loaded onto the RISC complex and, after removal of the 'passenger' RNA strand, the 'guide' strand directs the RISC complex to the complementary mRNA which either triggers mRNA degradation, or prevents its translation (reviewed in [182]). In *S. pombe*, an alternative RITS complex maintains the transcriptionally silent centromeric heterochromatin [183].

DICER1 has been implicated in the regulation of the initiation of X inactivation, but conflicting data about its precise role exist (reviewed in [184]). Briefly, dsRNA template was proposed to be formed by low-abundance Xist and Tsix transcripts on the active X in mouse. DICER1 processing of this dsRNA has been suggested to suppress *Xist* by CpG methylation at *Xist* promoter [185]. Later studies have however observed that DICER1 is dispensable for X inactivation [186] and that the effect of DICER1 on

Xist expression may be secondary, as DICER1 regulates expression of *de novo* methyltransferase DNMT3A, which in turn regulates *Xist* expression via DNA methylation [187].

1.4.4.3 ATM and ATR

ATM and ATR kinases play pivotal roles in ATM-CHEK2 and ATR- CHEK1 kinase signalling pathways that respond to DNA damage (reviewed in [188]). ATM-deficient mice further show disruption in meiosis due to the lack of XY body chromosome crossover and XY synapsis, however meiotic sex-chromosome inactivation is not affected [189]. Inhibition of ATM and ATR in mouse embryonic fibroblasts leads to partial hypoacetylation of the Xi and reactivation of an Xi-linked *EGFP* reporter, while Xist localization and macroH2A recruitment are not affected. Knock-down experiments show that depletion of either ATM or ATR alone also induces partial *EGFP* reactivation [190]. Interestingly, Ftx, a lncRNA transcribed from a region located 5' of *Xist* and upregulated at the onset of random X inactivation, was shown to positively regulate *Xist* transcription, possibly via *miR-421* located within *Ftx* intron and implicated in *ATM* regulation [25].

1.4.4.4 PARP1

PARP1 is a ubiquitous chromatin-associated poly(ADP-ribose) polymerase. While it seems to be responsible for the bulk of PARP activity in mammalian cells, up to 17 members of the PARP family have been identified [191]. PARP1 is recruited to chromatin via its interaction with histones, various DNA structures (*e. g.* single- and double-stranded breaks) and gene promoters, as well as a host of chromatin proteins. In keeping with its abundance and binding promiscuity, PARP1 also serves a number of roles, ranging from DNA repair, modulation of chromatin structure and gene transcription, DNA methylation and histone deacetylation, either through its enzymatic activity or competition for binding sites (reviewed in [192, 193]). PARP1 was reported to bind macroH2A1.2 and co-localize with the Xi. Knock-down of PARP1 results in partial reactivation of an *EGFP* transgene in the presence of histone deacetylase and DNA methylation inhibitors, suggesting that PARP1 may be involved in maintenance of the Xi [194].

1.4.4.5 REST and CoREST

REST is a transcriptional repressor that regulates silencing of neuron-specific genes in non-neuronal cells. REST recognizes a 23 bp-long conserved DNA motif via its zinc finger domain and recruits

histone deacetylases HDAC1 and HDAC2 via its N-terminal domain. The C-terminal domain of REST binds CoREST, which in turn recruits a wide range of silencing factors including HDAC1/2, LSD1 and H3K9 methyltransferases [195]. As discussed earlier, a complex consisting of REST, CoREST and LSD1 are recruited by the HOTAIR lncRNA which represses HOX genes [57].

1.4.4.6 SMCHD1

SMCHD1 is a protein of unknown function that contains an ATPase domain and a SMC hinge domain [132], shared among SMC proteins involved in sister chromatid cohesion, chromosome condensation, and DNA repair [196]. The homozygous mutation of *SmcHD1* is embryonic lethal in female, but not male mice, suggestive of its involvement in X inactivation [132]. Indeed, homozygous mutation of *SmcHD1* led to a specific loss of promoter DNA methylation and transcriptional upregulation of genes regulated by X inactivation [132]. Prominent localization of SMCHD1 to the Xi provides additional support for the major, yet undefined role of SMCHD1 in X inactivation [132].

1.4.4.7 SPOP and CUL3

CUL3 is one of seven mammalian cullin proteins that recruits a RING-family protein RBX1 to form an E3 ubiquitin ligase [197]. CUL3 directly interacts with Speckle-type POZ protein (SPOP) [198], which in turn interacts with the histone variant H2AFY (macroH2A1) [199] and PRC1 complex protein BMI1 [200] via its MATH domain. Knockdown of either SPOP or CUL3 in human HEK293 cell line was shown to disrupt recruitment of macroH2A1 to the Xi while XIST localization was unperturbed [200]. siRNA knock-down of either CUL3, SPOP or macroH2A1 resulted in partial de-repression of an Xi-integrated *EGFP*. Interestingly, these knock-downs resulted in partial re-activation of *EGFP* only upon concurrent treatment of cells with DNA methylation and histone deacetylase inhibitors. This observation is consistent with the currently-prevailing model which assumes that multiple, at least partially autonomous, mechanisms ensure silencing of chromatin of the Xi.

1.4.5 Condensins

In *C. elegans*, the equal dosage of X-linked genes between XX hermaphrodites and X0 males is achieved by hermaphrodite-specific downregulation of their two X chromosomes mediated by a condensin-containing dosage compensation complex (DCC) [201]. The worm DCC consists of ten proteins: SDC-1, -2, -3, DPY-21, -30, and five other proteins forming a complex homologous to condensin I and called condensin I^{DC}: MIX-1, DPY-27, DPY-28, CAPG-1 and DPY-26 [201]. The corresponding proteins forming human condensin complex I are SMC2, SMC4, NCAPD2, NCAPG and CAPH [202].

Recruitment of the DCC to the worm Xs is facilitated by two classes of DNA sequences: the autonomous *rex* (recruitment element on X) and *dox* (dependent on X), which recruits the complex only when located on the X [203]. Downregulation of gene expression is likely the result of the DCC-induced changes to the higher order chromatin structure, however recruitment of the MLL/COMPASS H3K4 methyltransferase complex, which shares DPY-30 protein with the DCC was also reported [204].

1.4.6 Genetic evidence for indispensability of the implicated proteins for X inactivation

In *Drosophila*, a number of genes comprising the dosage compensation complex have been identified due to the male-specific lethality of their disruption (reviewed in [31]). Hypothetically, a similar screen for genes that induce female-specific lethality due to aberrations in initiation or maintenance of X-inactivation could be performed in developing mammalian embryos in order to identify protein factors that are critical for X-inactivation. Indeed, identification of SmcHD1 as a previously unknown component of X-inactivation machinery in mouse is a notable example of such a strategy [132]. Female-specific lethality was not described for deficiency of any other protein discussed in section 1.4, however many of the proteins (e. g. Dicer1, PRC1 and PRC2 components, YY1, CTCF, HNRNPU and CUL3 [205-211]) are critical for development of both male and female embryos. However, while depletion of many proteins involved in X inactivation does not cause female-specific defects in embryonic development, this does not eliminate them as candidates for essential X-inactivation factors. Rather, it merely demonstrates that the gene-silencing machinery employed in X inactivation also performs other functions that are critical to set up correct transcriptional patterns in developing embryos.

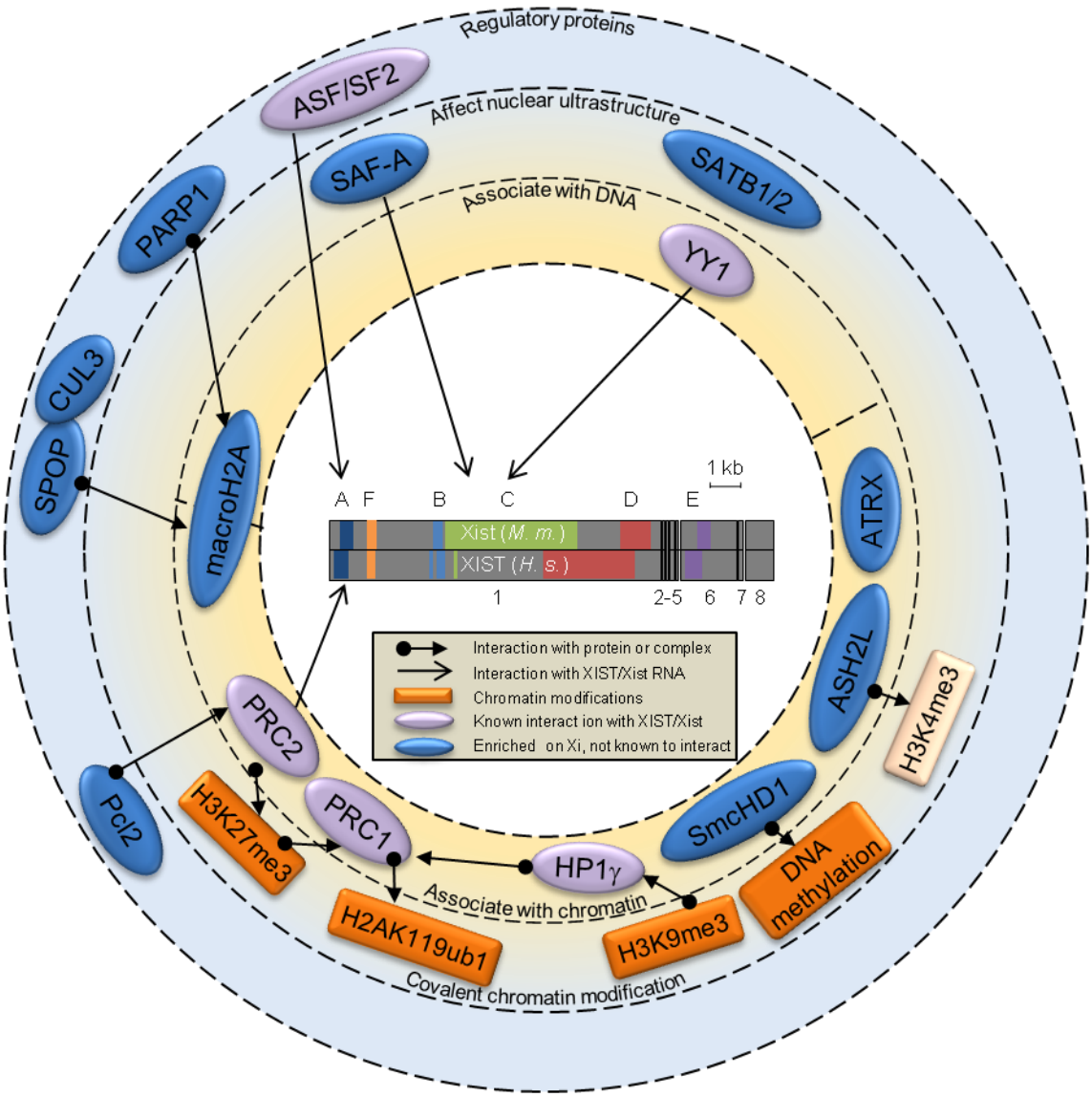


Figure 1.1: Proteins enriched on the inactive X chromosome or interacting with XIST.

A number of proteins have been implicated to play role in X-chromosome inactivation, either due to their enrichment on the Xi, or due to their binding to XIST/Xist RNA. The interactions described in this section are summarized. Chromatin modifications shown in dark orange are enriched on the Xi. H3K4me3, shown in pale orange, is depleted from the Xi, however ASH2L, the enzyme catalyzing H3K4 trimethylation, is enriched. The scaled diagram represents human XIST exon structure and positions of repeat A-F within the human and mouse XIST/Xist. For direct comparison of the relative position of repeat sequences, the same exon structure for XIST/Xist is shown. The simple arrows connect proteins which were demonstrated to directly bind XIST/Xist with their binding regions in XIST/Xist. The dashed lines dividing the inner circle separate proteins associated with DNA from proteins associated more broadly with chromatin.

1.5 Model systems for study of X inactivation

Studying X-chromosome inactivation *in vivo* is difficult because it occurs early in embryonic development. Moreover, *in vivo* studies as well as experiments on fertilized oocytes in humans raise ethical questions. For these reasons, there is a clear need for an *in vitro* system that would enable studies of the early events of X inactivation. An ideal system to study X inactivation would mimic the developmental state at which X inactivation normally occurs, allow for the rapid and controllable manipulation of the DNA, RNA and protein componentry of X inactivation and induce X inactivation gradually so that the intermediate steps of the process can be observed. Finally, an ideal system would employ human cells, as human-focused studies of X inactivation specifically and epigenetic regulation by lncRNAs in general will greatly improve our understanding of the role of lncRNAs in health and disease.

Mouse ES cells recapitulate X inactivation *in vitro*, as XX ES cells retain two active Xs in the undifferentiated state and undergo X inactivation upon differentiation [212]. Mouse ES cells allow both loss of function (knockout) and gain of function (transgene) studies; both approaches have been informative for exploring the processes involved in X inactivation, for example discovery of *Xist* regulators [19, 24, 55], X-chromosome pairing [27-29], functional sequences within *Xist* [42, 54] and chromatin changes occurring in the course of X inactivation (*e. g*. [102]).

By analogy, human ES cells could provide a similar model system. However, extending such studies into humans has been challenging. Surveys of an array of undifferentiated human female ES cells showed varying extents of X inactivation both among cell lines, and within the same cell line. While some human ES cell lines retain two Xa's prior to differentiation and induce *XIST* only when differentiated, the majority of clones have apparently already undergone X inactivation, and a subset of undifferentiated XaXi ES cell lines failed to express *XIST* [213-217]. Since an errant epigenetic regulation of the *XIC* may reflect an overall epigenetic instability and partial ES cell differentiation, the presence of two Xa's has been suggested as a hallmark of healthy human ES cell culture. Interestingly, precocious X inactivation can be triggered by a variety of factors inducing cellular stress, including derivation and maintenance of human ES cell lines under atmospheric, rather than physiological, oxygen concentration [218].

When X inactivation normally occurs during early human development however remains an outstanding question. A report combining RNA/DNA fluorescent *in situ* hybridization analysis of six human blastocysts showed an accumulation of XIST on a single X in 90 percent of cells [6]; early studies suggested *XIST* expression in both male and female embryos at around the 8 cell stage [219, 220].

A study comparing human, mouse and rabbit X inactivation revealed that in contrast with mouse, early *XIST* expression in human is not immediately associated with gene silencing [221]. Given the lack of a human ES cell system that would reliably model features of X inactivation, studies of human X inactivation have relied upon human *XIC* transgenes in mouse ES cells and human somatic cells.

Integrating the human *XIC* into mouse ES cells showed that the human *XIC* was recognized by the murine cells and triggered silencing of the single X in transgenic male mice [222]. However, only some aspects of normal X inactivation were recapitulated in low copy-number (1-2) transgenes, as gene silencing, or expression of the endogenous *Xist* was not induced [223, 224]. Similarly, only transgenic cell lines carrying multi-copy integrations of the mouse *Xic* were able to trigger X inactivation from the endogenous *Xic* [225].

The experiments testing the ability of a mouse *Xist* transgene to trigger X inactivation when induced at different time points during ES cell differentiation had shown that Xist can only induce inactivation during an early developmental window [226]. More recently, however, it has been shown that Xist can recapitulate inactivation for a brief period during hematopoiesis [227], in lymphoma cells [169] and in mouse embryonic fibroblasts when *SATB1* is ectopically expressed [169]. In contrast to the previous reports, transgenic *Xist* was able to form Xist foci and recruit H3K27me3 in mouse embryonic fibroblasts [54]. These results challenge the previously accepted paradigm that X inactivation can only be induced in early developmental stages and demonstrate that at least some features of X inactivation can be recapitulated in more differentiated cells lines. In fact, imperfect X inactivation may help to uncover yet unknown mechanisms that are overlooked in the more robust model systems.

Several transgenic systems have been developed in human somatic cells. Multi-copy *XIST*-containing transgenes were able to induce XIST accumulation in *cis* in HT1080 male fibrosarcoma cells [228]. The autosomal region coated by XIST showed nucleolar localization, histone H4 hypoacetylation and was devoid of CoT-1 RNA hybridization. Further, the neomycin resistance gene was silenced and new heterochromatic foci were established in *cis*, demonstrating both short and long range XIST action, respectively. Similar data were obtained with HT1080 cells carrying an *XIST*-containing PAC (P1-bacteriophage-derived artificial chromosome) clone [229]. In addition, an inducible human *XIST* construct in HeLa cells was able to localize and recruit chromatin marks, although silencing was not examined [48]. Overall, a number of human transformed differentiated cells seem to be capable of recapitulating at least some of the process of X inactivation. Such random integrations, however, were still subject to the variability of integration site and copy number. Therefore, Chow *et al.* [44] combined

the availability of human somatic cells with demonstrated responsiveness to ectopic *XIST* and the ability to target and to regulate expression and created an inducible single copy *XIST* transgene.

1.6 Thesis objective

The precise mechanism involved in XIST-induced gene silencing is not fully understood, in particular in human, where a comprehensive, well-controlled model system to study X inactivation akin to differentiating mouse ES cells is lacking. In pursuit of elucidating the molecular pathways of XIST RNA function, we took advantage of an inducible human transgenic system that enables us to focus on XIST's role in local gene silencing, deliberately isolating this critical facet of XIST action both from the regulation of *XIST* expression that ensures only one active X is retained per nucleus, and from the ability of XIST RNA to spread along the X chromosome.

We utilized a human HT1080 fibrosarcoma cell line in which an inducible XIST cDNA transgene is able to efficiently silence a proximally located fluorescent reporter. To uncover how the silencing is achieved, we first created a series of truncations to determine the minimal region of XIST responsible for silencing. Then, we designed a set of mutations to probe how the sequence and structure of this region influences its ability to silence. Finally, we tested which proteins are indispensable for the XIST-induced reporter silencing by utilizing histone deacetylase inhibitors and siRNA-mediated knock-downs of PRC2 and 31 other proteins previously implicated in X inactivation. Collectively, the data on sequences within XIST that are critical for proximal gene silencing and their secondary structure, as well as the proteins involved in XIST-induced silencing both expand the understanding of molecular pathways that lead from XIST/Xist expression to transcriptional silencing of the whole X chromosome in placental mammals and allow us to draw comparisons between the results obtained in human and mouse systems that model X inactivation.

2 MATERIALS AND METHODS

2.1 Construct generation and creation of the transgenic HT1080 cells

Truncated *XIST* constructs (dPFlMI dNC, del 5'A + 5'A, del 5'A and 5'A) were derived from the pre-existing full-*XIST* cDNA construct. The artificial repeat A construct, its shorter derivatives and mutants were synthesized by GeneArt (now Invitrogen). The constructs were subsequently cloned into the pcDNA5/FRT/TO plasmid (Invitrogen) using standard techniques and transfected into previously created single-copy FRT-harboring HT1080 cells.

The Flp-In T-Rex system (Invitrogen) was used by Sarah Baldry (Brown laboratory) according to the manufacturer's recommendations to generate the transgenic HT1080 cell lines. Briefly, HT1080 cells were first transfected with pcDNA6/TR plasmid which carries the Tet repressor (*TetR*) driven by the CMV promoter, grown in the presence of Blasticidin to allow for positive selection of cells in which pcDNA6/TR was successful integrated and two clones showing strong *TetR* expression, 2-3 and 2-12, were selected. Subsequently, the *TetR*-containing HT1080 cells were transfected with pFRT/LacZeo plasmid harboring a FRT integration site and a SV40 promoter-driven gene for Zeocin resistance which serves as a positive selection marker. After selection with Zeocin, single-cell colonies with random FRT integration sites were expanded and assayed by Southern blotting to select for single-copy FRT integration clones. At this point, the established clones can integrate the pcDNA5/FRT/TO plasmid when co-transfected with a Flp recombinase-containing pOG44 plasmid.

Successful pcDNA5/FRT/TO integration detaches the Zeocin resistance gene from the CMV promoter, and brings a Hygromycin resistance gene (*Hyg*) directly 3' of the CMV. Thus, cells with properly integrated pcDNA5/FRT/TO are Hygromycin resistant and Zeocin sensitive. The FRT-integrated *XIST* constructs cloned into the pcDNA5/FRT/TO can be induced by addition of tetracycline or doxycycline (DOX).

In the absence of DOX, two TetR homodimers occupy the two $TetO_2$ sequences within a modified CMV promoter and block transcription of *XIST*. Upon addition into culturing media, DOX binds the TetR homodimers in 1:1 stoichiometry, which results in TetR conformation change, prevents TetR from binding TetO2 sequences and allows the *XIST* to be expressed.

The HT1080 F55 cell line harboring a single copy FRT site integration on the X chromosome [230] was a kind gift of Dr. Chunhong Yan. Genomic localization of the FRT integration sites in the utilized HT1080 clones is listed in Table 2.1.

Table 2.1: Genomic localization of FRT integration sites in the described HT1080 cell lines

Cell line name	Genomic localization	Approximate distance to the nearest gene
2-3-0.5+3#1	7q21.2	215 kb (*MTERF*)
2-3-0.5+3#4	3q28	10 kb (*CLDN1*)
2-3-0.5a	8p23	in an intron (*AGPAT5*)
2-3-1.0#5	7p14.3	20 kb (*BBS9*)
2-3-1.0d	1p34.3	in an intron (*MACF1*)
2-12-0.5#3	3q26	in an intron (*FNDC3B*)
2-12-0.5#8	4q32	55 kb (*DCHS2*)
2-12-0.5+3#11	unknown	unknown
2-12-0.5+3#2	10p13	in an intron (*FRMD4A*)
2-12-1.0#14	12q24	in an intron (*FAM222A*)
2-12-4.0#9	unknown	unknown
F 55 DsRED #1	Xq23 [230]	150 kb (*HTR2C*)

2.2 Identification of transgene integration sites by inverse PCR

The ends of linearized plasmids are subject to exonuclease activity. Thus the actual integrated transgene often lacks several hundred of base pairs on each end (Figure 2.1A). A series of PCR assays was first used to identify the 5'- and 3'-most transgene sequences that are still intact (Figure 2.1B). *Pst*I and *Rsa*I restriction endonucleases with a known restriction site several hundred bp internally from the identified transgene ends were used to digest genomic DNA isolated from the 2-3-0.5+3#4 and HEK293 cell lines, respectively. The use of frequently-cutting restriction endonucleases, typically those that recognize a tetranucleotide sequence, yields a DNA fragment that on one end contains the plasmid sequence fragment and on the other end several hundred bp to several kb-long genomic sequence fragment (Figure 2.1C). T4 DNA ligase (Invitrogen) was used to create circular DNA molecules and the entrapped genomic DNA was amplified by nested PCR with primers facing outward from the plasmid fragment (Figure 2.1D, E). Finally the PCR product was gel purified, and the DNA was either directly sequenced or, if PCR did not yield DNA that was suitable for sequencing, cloned into the pGEM-T easy vector (Promega) and amplified in *E. coli* prior to sequencing (Figure 2.1F). Finally, the results of DNA sequencing were compared to the plasmid sequence, and the genomic location of remaining genomic sequence was identified using the BLAT algorithm (http://genome.ucsc.edu/cgi-bin/hgBlat).

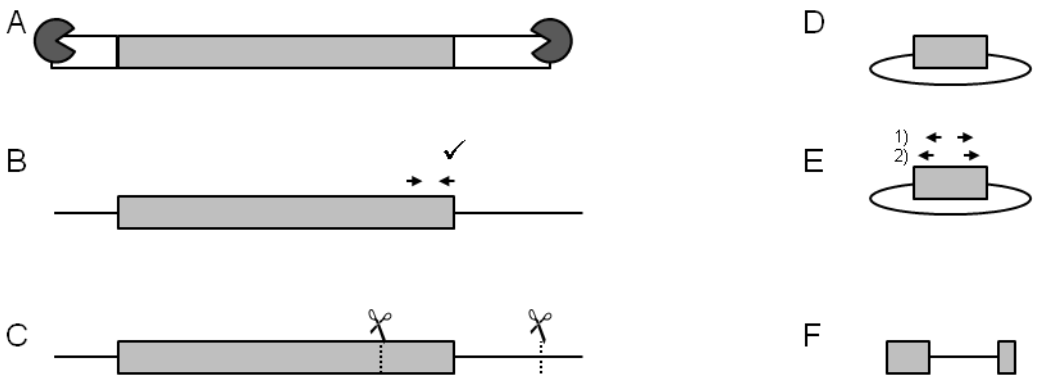


Figure 2.1: A general approach to identify transgene integration sites by inverse PCR.

(A), (B) PCR was used to identify 5'- and 3'-most plasmid sequences that were not degraded by exonucleases.

(C) Genomic DNA was digested by frequently-cutting restriction enzymes to obtain DNA fragments that contain a portion of plasmid sequence and a genomic DNA bordering with the integration site.

(D) T4 DNA ligase was used to obtain circular DNA fragments.

(E) The entrapped genomic fragment was amplified by nested PCR.

(F) The resulting PCR product was gel purified and either sequenced directly, or cloned into the pGEM-T easy vector and amplified in *E. coli*.

2.3 Flow cytometry

HT1080 cell pellets were washed with PBS and resuspended in 0.5 mL of PBS with 10% FCS. LSRII flow cytometer (BD) was used to record 30 000 events; 10 000 events were recorded in the siRNA-mediated knock-down experiments, as less cells were used per experiment. Mean fluorescence intensity of EGFP was assesed by using a combination of 488 nm laser excitation and 530/30 nm bandpass filter; 561 nm laser and 582/15 nm filter were used for DsRED-Express2.

2.4 RNA isolation and reverse transcription

RNA was isolated from frozen cell pellets by TRIZOL (Invitrogen) and treated with DNase I (Roche) according to the manufacturers' recommendations. Following phenol-chloroform extraction, RNA concentration was assessed by spectrophotometer and 0.5–2.5 μg of RNA was reverse-transcribed by M-MLV reverse transcriptase (Invitrogen) in a 20 μL total reaction volume.

2.5 Quantitative PCR

HS Taq (Fermentas) and EvaGreen (Biotium) were used in the quantitative PCR reactions under the following conditions: 5 min. 95 °C, 40x [15 sec. 95 °C, 30 sec. 60 °C, 60 sec. 72 °C]; composition of the reaction mix is shown in Table 2.2. Primer Express 3.0 (Applied Biosystem) software was used to design the primers, which are listed in Table 2.3. The software's algorithm consistently designed primers that performed optimally under the standard cycling conditions described above. In the rare instances when PCR primers did not perform, an alternative primer pair was designed. The use of one set of standard qPCR conditions allowed maximum flexibility in combining multiple PCR reactions into one 96-well plate run. The standard curve method (6 times 1:4 dilution series) was used to quantify sample concentration and 'blank' reactions lacking the PCR template were included in each experiment to detect any potential primer-dimer products. Unless specified otherwise, standards, samples and blank reactions were assayed in triplicate.

Table 2.2: Quantitative PCR reaction mix composition

Reaction mix component	**Volume per 1 well [μL]**
dNTP mix (25 mM)	0.16
$MgCl_2$ (25 mM)	2
HS reaction buffer (10x)	2
For + Rev primer mix (25 μM)	0.2
Template	1.5
HS Taq	0.16
EvaGreen (20x)	1
deionized H_2O	12.98
TOTAL	20

Table 2.3: List of PCR primers

Primer name	Sequence	Notes
qXIST_-1kb F	CTGCTCTGATGCCGCATAGTT	p1 in Figure 3.7
qXIST_-1kb R	TTTTGCTCGCGCACTACTCA	
qXIST 5 F	TCAGCCCATCAGTCCAAGATC	p2 in Figure 3.7
qXIST 5 R	CCTAGTTCAGGCCTGCTTTTCAT	
qpFRT_4719 F	GCTCAGAAGAAATGCCATCTAGTG	p3 in Figure 3.7
qpFRT_4790 R	TTTTTTGGAGGAGTAGAATGTTGAGA	
qpFRT_5921 F	CCACCAACAGCAAAAAAATGAA	p4 in Figure 3.7
qpFRT_5986 R	ACTCATGAAAATGGTGCTGGAA	
qpcDNA5 F3	CGCCATCCACGCTGTTTT	qRT-PCR of XIST expression, p5 in Figure 3.7
qpcDNA5 R3	CCGGAGGCTGGATCGGT	
qEGFP594 F	AGCGCTACCGGACTCAGAT	qRT-PCR, ChIP
qEGFP649 R	GTACCGTCGACTGCAGAATTC	
qACTB 1	TTGCCGACAGGATGCAGAA	qRT-PCR
qACTB 2	GCCGATCCACACGGAGTACTT	
qSUZ12 F	GGGAGACTATTCTTGATGGGAAGAG	
qSUZ12 R	TCCAACGAAGAGTGAACTGCAA	
qEZH2 F	GGTAAATCCAAACTGCTATGCAAA	
qEZH2 R	GGATGGCTCTCTTGGCAAAA	
qHyg F	CAGCGAGAGCCTGACCTATTG	
qHyg R	CAGGCAGGTCTTGCAACGT	
qDsRED_Exp2 F	TGAAGCTGCCCGGCTACTA	
qDsRED_Exp2 R	TCCTCGTTGTGGGAGGTGAT	
qPgk1 1F	GGCACTTGGCGCTACACAA	qPCR – ChIP
qPgk1 1R	CCTACCGGTGGATGTGGAAT	
qPgk1 3F	AGCGGCCAATAGCAGCTTT	
qPgk1 3R	CCCCTTCCCAGCCTCTGA	
qPgk1 4F	TCTGCCGCGCTGTTCTC	
qPgk1 4R	GATGGATGCAGGTCGAAAGG	
qMYT1 F	GCTACAGCAGCTACCAGGGAAT	
qMYT1 R	CTCTTCCACCAGGGTCTCTTCA	
qAPRT F	GCCTTGACTCGCACTTTTGT	
qAPRT R	TAGGCGCCATCGATTTTAAG	

Primer name	Sequence	Notes
qBRRN1_F	TCTCGAGTTGCCAGAGTTAGGTT	qRT-PCR to assay knock-down efficiency in section 5.2.3
qBRRN1_R	TCTGGCGATCTTCTGCACACT	
qSMC4L1_F	AGAATGGGTTCCTCACTTGTTATTG	
qSMC4L1_R	TTAGAGTCGTTTTGCAACTGTGATT	
qCNAP1_F	ACTGCTTGCCAAAGCTAGTTACAA	
qCNAP1_R	AGGGTTCGGACTCCTGGAAGT	
qAOF2_F	GGGATTTGGCAACCTTAACAAG	
qAOF2_R	CATGCCCGAACAAATTGACA	
qPARP1_F	AACACTCATGCAACCACACACAA	
qPARP1_R	GCTGGCATTCGCCTTCAC	
qASH2L_F	GGCTGACACATTTGGCATAGATAC	
qASH2L_R	GATGGCAGACGTTGCAATGA	
qSDC1_F	CGAGAGGGCTGCTGAGGAT	
qSDC1_R	ATTCTCCCCCGAGGTTTCAA	
qM96_F	GAGGCCCTGGAGACTGGTATT	
qM96_R	GCATTGCACACAAGCCTCAT	
qCUL3_F	TCAGTCAGCCACACCAAAGTG	
qCUL3_R	CACTGTGTTTGGCTAAGTAGAACCTT	
qSPOP_F	TTCCAGGCTCACAAGGCTATC	
qSPOP_R	TTGCTCTCCTCCATTTCATGTTC	
qATM_F	AATGCTTGCTGTTGTGGACTACA	
qATM_R	ATCCAGCCAGAAAGCATCATTAA	
qATRX_F	ACAAGGCGTTCAAGCGAAAA	
qATRX_R	GTGCAAGGAAGTCATGAAGCTTCT	
qSMCHD1_F	CGGCTACCACTTTTATCAAGAACCT	
qSMCHD1_R	TGTTGCTGCTTCTTAACATCATTG	
qBRCA1_F	GGCAAACTTGTACACGAGCATAA	
qBRCA1_R	CAGAAAGGGTCAACAAAAGAATGTC	
qH2AFY_F	TTGAGGTGGAGGCCATAATCA	
qH2AFY_R	TTTCTTCTCCAGCGTGTTTCC	
qH2AFY2_F	GATAGCCCCGAGACACATCTTG	
qH2AFY2_R	TGGCGATGGTCACTCCTTTT	
qHNRPU_F	GCGAAATTTTATTCTGGATCAGACA	
qHNRPU_R	GCTGGAAGCCTGCAAACAG	
qSATB1_F	GTTATTTATGTGCTGTCAAGTTTTGAAGT	
qSATB1_R	TGAGTTGCCTCGTTCAAATGAT	

Primer name	Sequence	Notes
qSATB2_F	CTGTCCGAGGGTCTTCTTCCT	qRT-PCR to assay knock-down efficiency in section 5.2.3
qSATB2_R	TGTCTTTGCAAGAGTGGCATTC	
qSET7_F	TGCAAGGCATCATCCACATAA	
qSET7_R	GGGAACTTTGTTCACGGAGAAA	
qCARM1_F	CTGATGGCCAAGTCTGTCAAGTA	
qCARM1_R	AATGGGATTTCTATCCTGTGCAA	
qRNF2_F	CAGCCCTTAGAAGTGGCAACA	
qRNF2_R	TGGGTCTGGCCTTAGTGATCTT	
qYY1_F	ACCTGGCATTGACCTCTCAGA	
qYY1_R	TTTTTCTTGGCTTCATTCTAGCAA	
qCTCF_F	CATCTCTGTGGCAGGGCATT	
qCTCF_R	TTGTGAGGACGAGTACCTGTGTGT	
qCBX4_F	AGCTGATGGGATATCGGAAGAG	
qCBX4_R	ATTGGAACGACGGGCAAAG	
qCBX7_F	ATCGGCGAGCAGGTGTTC	
qCBX7_R	CACTTCACCAGATACTCGACTTTACC	
qHNRPK_F	GCCCCGAGCGCATATTG	
qHNRPK_R	TTCCAAGGTAGGGATGATTTTCTT	
qEHMT2_F	GGACGACTGCTCTAGCTCCAA	
qEHMT2_R	GGAGCAATCGCCCATCCT	
qEHMT1_F	GTCCAGTACCTGCTTTCAAATGG	
qEHMT1_R	TTGTACTCTGTGGCCCAGATCAT	
qREST_F	TCCTTACTCAAGTTCTCAGAAGACTCA	
qREST_R	CCACATAACTGCACTGATCACATTT	
qRCOR1_F	GCATGGGTACAACATGGAACAG	
qRCOR1_R	GGCAAATCAGCCAATGACTTTT	
qDICER1_F	CATGAGGGCCGCCTTTC	
qDICER1_R	CCATGCGGCTGGGTAGTC	
qCLDN1-F	AGCACCGGGCAGATCCA	qRT-PCR in section 3.2.2
qCLDN1-R	CACGGGTTGCTTGCAATGT	
qCLDN16-F	CGCACCTGTGATGAGTACGATT	
qCLDN16-R	TCGAGTTACCACCAGCTTCAAG	
qLERPREL1_F	TATGGAGGACGACAGGATGAGA	
qLERPREL1_R	AACTTCTGCTCCCTCTACGTTCA	
qIL1RAP_F	GGCCCACTCTCCTCAATGAC	
qIL1RAP_R	TTTGCTGCAATATGTAGTGTTCCTT	

2.6 Cell culture

Clones harboring single-copy integration of *XIST* constructs into HT1080 fibrosarcoma cell lines were generated and cultured as described previously [44]. The *XIST* transgenes were induced by doxycycline (1 μg / mL) and cell culture medium was changed every 24 hours.

2.7 siRNA-mediated knock-down

siRNA-mediated knock-down was performed according to the manufacturer's protocol. A 0.5-1 μL aliquot of Dharmafect 4 transfection reagent (Thermo Scientific) and 2.5-5 μL of 5 μM siGenome SMARTpool siRNA (Table 2.4; Thermo Scientific) were used per 500 μL of medium in each well. Cells were seeded in a 24-well plate at 30 000 cells per well density. Timelines for the DOX and siRNA treatments varied and are always depicted for the individual experiments.

Table 2.4: List of siRNAs used

Target gene (human)	Accession number	Product number
SUZ12	NM_015355	M-006957-00
EZH2	NM_152998	M-004218-03
BRRN1	NM_015341	M-012853-01
SMC4L1	NM_001002800	M-006837-01
CNAP1	NM_014865	M-021198-00
AOF2	NM_015013	M-009223-01
PARP1	NM_001618	M-006656-01
ASH2L	NM_004674	M-019831-01
SDC1	NM_002997	M-010621-01
M96	NM_007358	M-012796-02
CUL3	NM_003590	M-010224-02
SPOP	NM_001007228	M-017919-02
ATM	NM_138292	M-003201-04
ATRX	NM_138270	M-006524-01
SMCHD1	NM_015295	M-032684-00
BRCA1	NM_007298	M-003461-02
H2AFY	NM_004893	M-011964-00
H2AFY2	NM_018649	M-010913-01
HNRPU	NM_004501	M-013501-01
SATB1	NM_002971	M-011771-00
SATB2	NM_015265	M-023161-00
SET7	NM_030648	M-014643-01
CARM1	NM_199141	M-004130-00
RNF2	NM_007212	M-006556-01
YY1	NM_003403	M-011796-02
CTCF	NM_006565	M-020165-02
CBX4	NM_003655	M-008356-01
CBX7	NM_175709	M-009561-02
HNRPK	NM_002140	M-011692-00
EHMT2	NM_025256	M-006937-01
EHMT1	NM_024757	M-007065-00
REST	NM_005612	M-006466-02
RCOR1	NM_015156	M-014076-01
DICER1	NM_030621	M-003483-00

2.8 Chromatin immunoprecipitation

All steps were performed as published previously [231]; incubation with micrococcal nuclease for 8 minutes provided an ideal size of chromatin fragments. Antibodies used were: 5 μg (per reaction) of anti-H3K27me3 (07-449; Millipore), 7.5 μg of anti-H3 (H9289; Sigma), 10 μg of IgG (I8140; Sigma) and 2.5 μg of anti-panH4acetyl (06-598; Millipore).

2.9 RNA structure modeling

Mfold server version 2.3 was used to predict secondary RNA structures (http://mfold.rna.albany.edu).

2.10 Analysis of repeat A core sequences in mammals

Repeat A sequences in a panel of mammalian species were identified using a combination of BLAST, BLAT and *in silico* PCR searches of mammalian genomes available through NCBI (http://blast.ncbi.nlm.nih.gov) and ENSEMBL (http://www.ensembl.org/Multi/blastview) databases, as well as UCSC genome browser (http://genome.ucsc.edu). Accession numbers or genomic locations of repeat A sequences are listed in Table 2.5. Sequences were aligned in clustalw2 (http://www.ebi.ac.uk/Tools/msa/clustalw2) and screened to exclude all non-*bona fide* repeat A CG-rich core sequences from further analyses. CG-rich core sequences that contained bases deviating from the canonical sequence of either stem 1 or stem 2 were identified. Finally, we tested whether such a mutation was reciprocated by a mutation within the same repeat A unit, or in any other repeats of that species.

Table 2.5: List of accession numbers or sequence coordinates of repeat A sequences compared in sequence analyses

Species	Accession number / genomic location
Mus musculus	NR_001463
Rattus norvegicus	chrX:91,467,666-91,468,097 Nov. 2004
Ellobius lutescens	EU086094.1
Equus caballus	U50911.1
Pan troglodytes	chrX:73,645,114-73,645,544 Oct. 2010 assembly
Gorilla gorilla	chrX:71,018,187-71,018,664 May 2011 assembly
Pongo pygmaeus	chrX:71,294,280-71,294,715 Jul. 2007 assembly
Homo sapiens	NR_001564
Macaca mulatta	chrX:72,974,560-72,974,994 Jan. 2006 assembly
Callithrix jacchus	chrX:65,411,057-65,411,432 Mar. 2009 assembly
Echinops telfairi	scaffold_298824:9,733-10,152 Jul. 2005 assembly
Cavia porcellus	scaffold_26:23,393,897-23,394,265 Feb. 2008 assembly
Tursiops truncatus	scaffold_92440:418-831 Jul. 2008 assembly
Oryctolagus cuniculus	U50910.1
Erinaceus europaeus	scaffold_354641:1,200-1,618 Jun. 2006 assembly
Sorex araneus	scaffold_229162:51,879-52,334 Oct. 2005 assembly
Felis catus	chrUn_ACBE01438274:3,390-3,792 Dec. 2008 assembly
Bos taurus	NR_001464.2
Sus scrofa	CU855548.6
Tupaia belangeri	scaffold_148376:1,812-2,307 Jun. 2006 assembly
Microcebus murinus	scaffold_20625:5,197-5,658 Jun. 2007 assembly
Canis lupus familiaris	chrX:60,410,297-60,410,751 May 2005 assembly
Ailuropoda melanoleuca	GL194824.1:76,507-76,993 Dec. 2009 assembly
Vicugna pacos	scaffold_25540:1,149-1,688 Jul. 2008 assembly
Tarsius syrichta	scaffold_135455:2,195-2,478 Aug. 2008 assembly
Myotis lucifugus	GL429771:10,780,415-10,780,834 Jul. 2010 assembly
Pteropus vampyrus	scaffold_7187:74,526-74,917 Jul. 2008 assembly

2.11 Statistical analyses

When shown, error bars represent ±1 standard deviation. Two-tailed Student's t-test was used to probe whether differences in gene expression levels were significant. Correlations are calculated using Pearson correlation coefficient.

3 THE REPEAT A IS SUFFICIENT TO INDUCE GENE SILENCING IN MULTIPLE INTEGRATION SITES IN THE HT1080 TRANGENIC SYSTEM

The candidate (Jakub Minks) designed, performed and analyzed all experiments presented in this section with the following exceptions:

Sarah Baldry, a member of the Brown laboratory, has performed all experiments required to transfect *XIST* transgenes into the HT1080 cells.

Transgene integration sites in HT1080 cells lines shown in Figure 3.5C were previously identified by Dr. Jennifer Chow, Sarah Baldry, Jackie Goyns and Christine Yang of the Brown laboratory.

3.1 Introduction

X inactivation has been most thoroughly studied in human and mouse. The key principles of X inactivation are shared in both organisms, and indeed all placental mammals studied so far. However, humans and mice differ in many aspects of X inactivation, ranging from different regulation of *XIST*/*Xist* expression [20] to differences in XIST/Xist structure (see section 1.3.2) and the proportion of genes that escape X inactivation (reviewed in [11]). As X inactivation is an early developmental event, human *in vivo* studies are very limited due to ethical and practical considerations. Therefore, development of human-specific model systems that recapitulate all or some features of X inactivation *in vitro* is critical to further our understanding of the molecular underpinnings of human X inactivation.

As discussed in section 1.5, human ES cells have, so far, not proven as useful for the study of human X inactivation as their mouse counterparts. Therefore, several human models employing differentiated cell lines have been developed. With an intention to avoid the shortcomings of previously utilized systems, and in the absence of an ES cell or a similar system that would mimic the events taking place during normal X-chromosome inactivation in humans, our laboratory has previously developed a transgenic system consisting of two main components. First, we created a set of HT1080 male fibrosarcoma cell lines harboring randomly integrated single-copy FRT sites. Second, we cloned a series of plasmid constructs containing either a full-length cDNA of human XIST or truncated *XIST* cDNAs driven by a doxycycline inducible promoter [44]. These constructs were then transfected into the FRT site-containing HT1080 cell lines.

In one HT1080 cell line (HT1080 2-3-0.5+3#4), the FRT site-containing plasmid was co-transfected with an *EGFP*-containing plasmid, which resulted in *EGFP* integration directly downstream of FRT site. The presence of fluorescent reporter gene allowed for convenient assessment of XIST's silencing ability. For this reason, the 2-3-0.5+3#4 cell line was used in the majority of experiments presented in this thesis. Fluorescence *in situ* hybridization mapped the transgene to the 3q chromosome arm. Upon doxycycline-mediated induction, full-length XIST transcripts localized in *cis* and silenced an adjacent *EGFP* reporter. An approximately 80% decrease in EGFP signal was observed by flow cytometry after 4 days of *XIST* expression. Notably, *EGFP* repression required continuous *XIST* expression. A subset of the epigenetic modifications associated with X inactivation was also observed after *XIST* induction, however in comparison with normal mouse X inactivation, the changes occurred at slower rate. Upon *XIST* induction, the CMV promoter driving *EGFP* showed a decrease in H3K4 di- and trimethylation and H4 acetylation, accompanied by an increase of CBX3 (HP1γ) and H4K20me1. No recruitment of H3K9me2 or DNA CpG methylation was observed in the course of *EGFP* silencing. The *XIST* signal co-localized with a nuclear territory depleted for hnRNA transcription, forming a so called 'CoT hole' [228]. While

this suggests that the transgenic XIST is able to form a silent compartment, chromosome-wide gene silencing was presumably not induced, as haploinsufficiency for multiple *cis*-linked genes would have likely resulted in cell death.

Taking advantage of the ability to re-target different *XIST* constructs into the same FRT integration site and directly compare the function of various *XIST* constructs, Chow *et al.* tested the impact of three *XIST* deletions [44]. First, a deletion of a central portion of *XIST* exon 1 had no effect on XIST localization or *EGFP* silencing. Second, a deletion truncating the cDNA from the 3' section of exon 1 downstream resulted in slightly less localized XIST accumulation, but did not affect *EGFP* silencing. Third, a deletion of the repeat A region resulted in loss of XIST's ability to silence *EGFP*; consistent with the data from mouse *Xist* constructs [42]. However in contrast to the mouse repeat A deletion, XIST localization was also lost [44]. The transgenic XIST was also integrated into a commercially available HEK293 cell line with a single FRT integration site. Compared to the HT1080 cell line, the HEK293 showed robust recruitment of chromatin marks associated with the Xi [44].

In summary, the transgenic HT1080 cells offer an exciting model for the study of human X-chromosome inactivation because *XIST* induction leads to gene repression with only a subset of chromatin changes observed in normal X inactivation and because it allows us to reproducibly probe the function of multiple *XIST* constructs by inducing *XIST* expression with DOX and using flow cytometry or qRT-PCR to measure the extent of gene silencing induced by XIST. The work presented in this section aimed to refine the minimal sequence of XIST necessary and sufficient to induce gene silencing in the HT1080 2-3-0.5+3#4 cell line. Further we explored whether the extent of XIST-induced gene silencing differs between cell types and HT1080 cell lines, as well as in multiple single cell-clones of the same cell line. Finally, we aimed to test whether the transgenic XIST has capacity to silences other transgenes and endogenous genes.

3.2 Results

3.2.1 Repeat A is sufficient to induce *EGFP* silencing

To further explore which regions of XIST are critical for silencing, we first assayed the relative level of transcription along the transgenic *XIST*. The ‘full-length’ *XIST* cDNA transgene corresponds to a common splicing variant which retains *XIST* exons 1-7, but lacks approximately 2/3 of exon 6 [34, 44]. qRT-PCR data show approximately equal amount of transcripts along *XIST*, suggesting that most of the transcripts span the whole length of the transgenic *XIST* (Figure 3.1). Importantly, the expression of the

induced transgenic *XIST* in the HT1080 cell line is comparable to the *XIST* expression in normal female lymphoblasts (Figure 3.1).

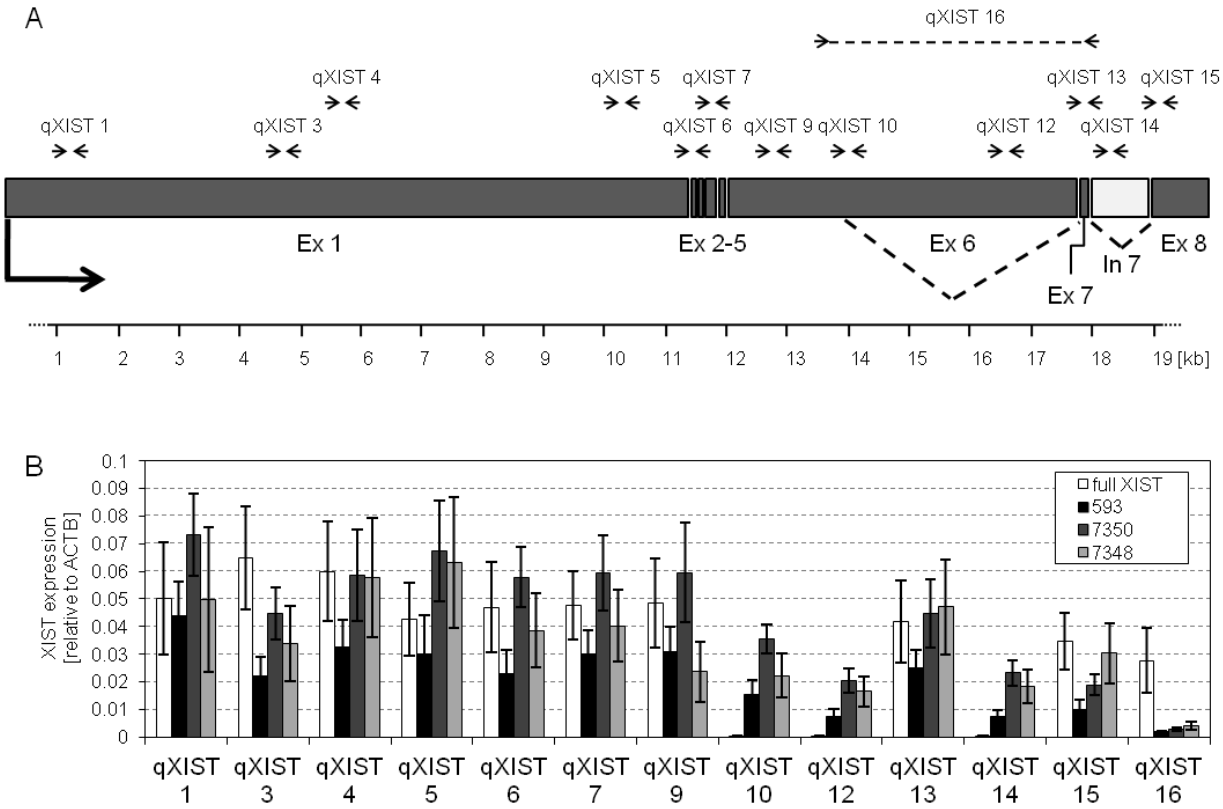


Figure 3.1: Transcription along *XIST*.

(A) Schematic of full-length *XIST* cDNA transgene depicting position of qPCR primer pairs qXIST 1 – qXIST 16 relative to *XIST* exons and intron 7.

(B) Expression of *XIST* in the HT1080 cell line harboring full-length *XIST* cDNA transgene and three female lymphoblast cell lines (593, 7050 and 7348), relative to *ACTB* expression. The error bars represent ±1 s.d. In order to compare the transcript abundance along the *XIST* transgene, genomic DNA from the HT1080 transgenic cell line was used as a template for standard curve samples to normalize for variation in qPCR efficiency. Since the HT1080 transgenic cell line is male, it contains one copy of endogenous *XIST* and one copy of the transgenic *XIST* cDNA, which contains exons 1-5, the spliced variant of exon 6, and exons 7-8. Therefore, we normalize for this variation in copy number of different XIST regions.

Having confirmed that *XIST* transcription in the full-length cDNA transgene mimics *XIST* transcription in normal female cell lines, we focused on delineating the XIST sequence that is critical for transgene silencing. Our laboratory has previously shown that the ability of full-length XIST to silence *EGFP* was retained in both a construct lacking the 3.8 kb region 3' of repeat A sequences and in a construct lacking exons 2-8 and the 3'-most portion of exon 1 [44]. We created a construct that combines these deletions

(Figure 3.2A) and tested its silencing potential. Flow cytometry showed that even this construct consisting only of the 5'-most fragment of exon 1 which includes repeat A and a further 3.5 kb of exon 1 sequence induced strong *EGFP* silencing upon *XIST* expression (Figure 3.2B). To probe whether the overall structure of XIST impacts its ability to silence, we created a transgene in which repeat A region is located at the 3', instead of at the 5' end of *XIST*. Flow cytometry survey of three single-cell colonies again showed strong *EGFP* silencing upon *XIST* expression (Figure 3.2C).

Our laboratory has previously reported that the repeat A-lacking *XIST* construct failed to induce gene silencing in the HT1080 cells [44]. A similar repeat A-lacking Xist also did not induce gene silencing in mouse ES cells [42]. However, in contrast to the observations in the mouse system, the repeat A-lacking construct was unable to form XIST foci in the HT1080 cells. To explore whether the failure of the silencing-deficient human XIST to localize is accompanied by an accumulation of XIST in the cytoplasm, we assayed nuclear and cytoplasmic concentrations of the truncated XIST. qRT-PCR analysis of nuclear and cytoplasmic fractions revealed that both the full-length and the repeat A-lacking *XIST* transgenes are predominantly localized to the nucleus (Figure 3.2D).

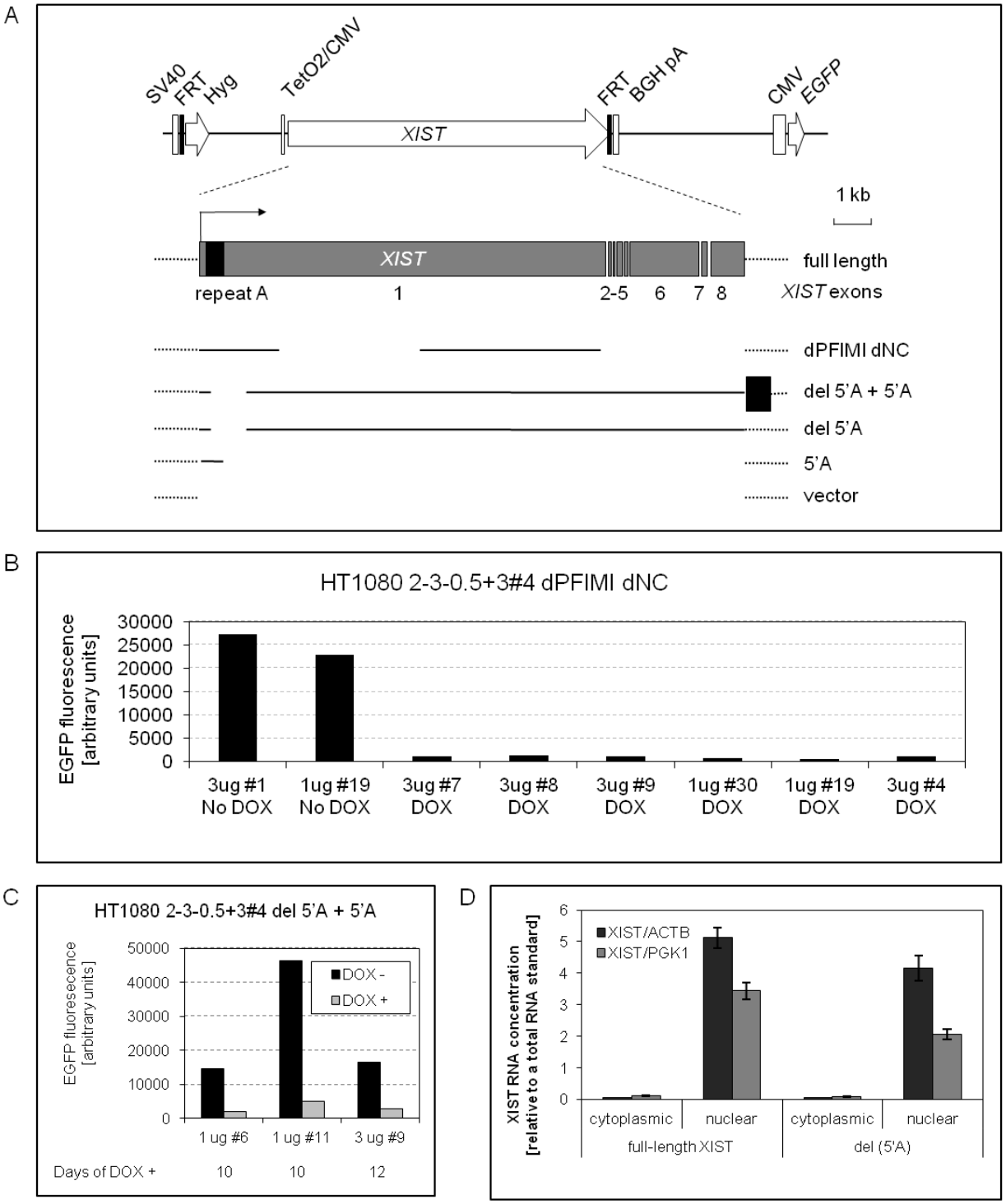


Figure 3.2: *XIST* transgenes containing repeat A are capable of gene silencing.

All data on shown in this figure were measured in the *EGFP*-containing HT1080 2-3-0.5+3#4 cell line.

(A) Schematic of full-length *XIST* cDNA transgene depicting *XIST* exons and regions included in shorter *XIST* constructs. Deletion in the del 5'A construct spans units 2-9 of repeat A, as well as approximately 450 bp immediately downstream of repeat A.

(B) Expression of *EGFP* in six single cell clones harboring the dPFlMI dNC construct was measured by flow cytometry following *XIST* induction for 14 days. *EGFP* expression was also measured in the absence of *XIST* expression in two clones (3 μg #1 and 1 μg #19). The clone names reflect the amount of plasmid DNA that was transfected into the HT1080 cells (1 μg or 3 μg) and the order of the individual single-cell colony isolated.

(C) As in (B). *XIST* was induced for 10-12 days as indicated.

(D) A representative qRT-PCR analysis of cytoplasmic and nuclear concentrations of the full-length *XIST* cDNA and del (5'A) constructs. The amount of XIST RNA in nuclear or cytoplasmic fractions was normalized, respectively, to nuclear or cytoplasmic concentration of ACTB or PGK1 RNA. The error bars represent ±1 s.d. of qRT-PCR technical triplicate.

Since all transgenic cell lines tested up to this point that contained repeat A sequences were able to silence *EGFP* and since the *XIST* transgene that lacked repeat A but was otherwise intact failed to repress *EGFP* [44], we wished to test whether a transgenic cell line containing only a repeat A fragment is sufficient for proximal gene silencing. Indeed, multiple single cell clones showed strong *EGFP* silencing following induction of a construct containing only repeat A sequence (Figure 3.3A). The extent and dynamics of *EGFP* silencing by repeat A mimicked that of full-length XIST over the first 5 days following induction by doxycycline, suggesting that the ability of XIST to silence *EGFP* is solely attributable to the repeat A region (Figure 3.3 B, C).

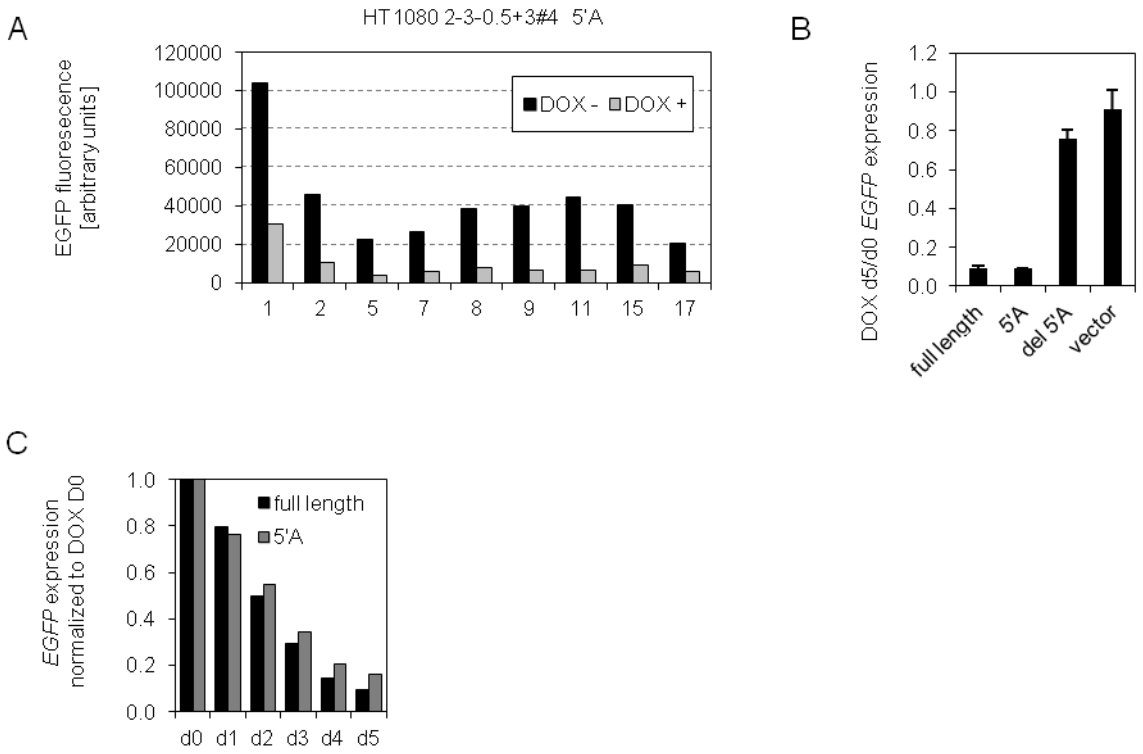


Figure 3.3: Repeat A region of XIST is sufficient to induce gene silencing.

All data on shown in this figure were measured in the *EGFP*-containing HT1080 2-3-0.5+3#4 cell line.

(A) Expression of *EGFP* measured in 9 different single-cell clones (1 – 17) by flow cytometry. *XIST* was induced for 10 days.

(B) *EGFP* expression following 5 days of transgene induction, measured by qRT-PCR. Values are normalized to *EGFP* expression in uninduced cells and to *ACTB* expression. The error bars represent ±1 s.d. of two biological replicates.

(C) *EGFP* expression following 1–5 days (d1–d5) of full-length *XIST* and 5'A induction, measured by flow cytometry and normalized to *EGFP* expression in uninduced cells (d0).

3.2.2 Genes both up- and downstream of transgenic *XIST* undergo silencing in multiple integration sites

Efficient XIST-induced silencing requires not only functional XIST RNA and its protein partners, but also proper genomic context, as has been demonstrated by limited spread of X inactivation into autosomal portions of translocated X:A chromosomes (reviewed in [11]). Having established that repeat A is sufficient for *EGFP* silencing, we aimed to better describe the HT1080 transgenic system and explore both whether this effect extends to other genes and whether XIST functions in various integration sites. The Flp-In system (Invitrogen) that was used to create the HT1080 transgenic cell lines

consists of a plasmid that carries the gene for the tetracycline repressor (*TetR*) and an FRT-harboring plasmid. The *TetR*-containing plasmid was integrated in two different, but unmapped, genomic sites in the 2-3 and 2-12 'parental' cell lines. Subsequently, numerous cell lines with differing random FRT integration sites were created using either the 2-3 or the 2-12 'paternal' HT1080 cell lines. Finally, the pcDNA5/FRT/TO plasmid was employed to integrate the transgenes into a FRT site by transient co-transfection of a Flp integrase construct. Successful pcDNA5/FRT/TO integration results in expression of the hygromycin resistance gene (*Hyg*) by the SV40 promoter, which is located approximately 2.3 kb upstream of the transgene-driving inducible promoter and serves as a positive selection marker (detailed in section 2.1). A commercially available HEK293 cell line (Invitrogen) containing FRT and *TetR* was also used to probe the effects of ectopic *XIST* expression.

Using qRT-PCR, we first tested the expression levels of the *TetR* (Figure 3.4A). *TetR* was expressed in all clones tested. Overall, the expression in the 2-3 -derived cell lines was stronger compared to 2-12-derived cell lines ($p < 0.0013$, two-tailed t-test); we however note that *TetR* expression shows clone-to-clone variation. As expected, *XIST* expression is upregulated in cells treated with DOX (Figure 3.4B). Although the expression levels of *XIST* varied considerably in the six different cell lines tested, the clones that showed higher *XIST* expression also tended to express *Hyg* more strongly. Previously, a qRT-PCR analysis suggested hygromycin is not subject to XIST-induced silencing [44]. However, in all cell lines tested here, *XIST* induction lead to *Hyg* repression (Figure 3.4C).

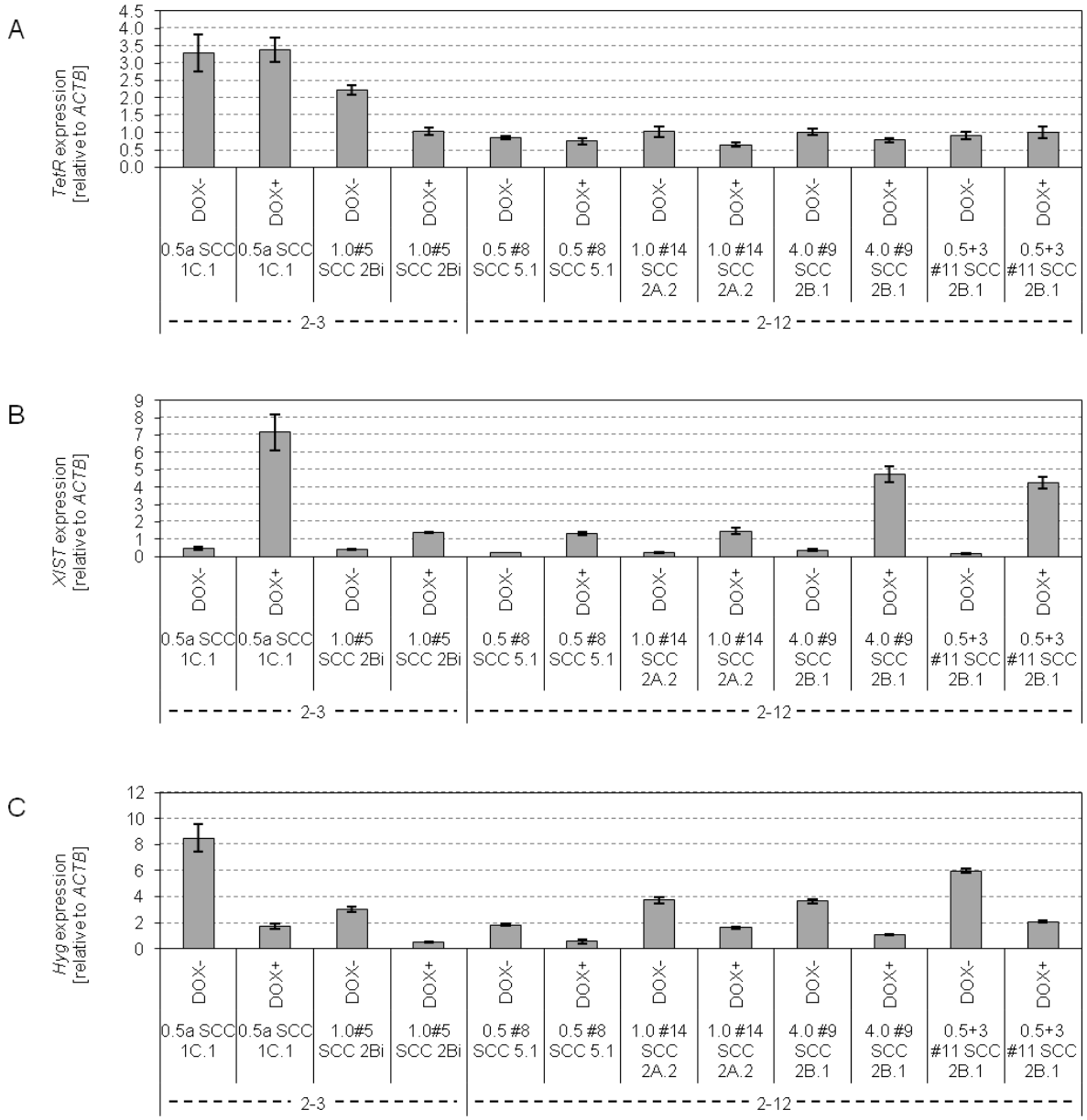


Figure 3.4: Expression levels of *TetR*, *XIST* and *Hyg* in multiple integration sites.

Expression of *TetR*, *XIST* and *Hyg* was assayed by qRT-PCR with qTetR S:AS, qXIST5 S:AS and qHyg S:AS primer pairs, respectively, and normalized to *ACTB*. In the samples labeled DOX +, *XIST* expression was induced for 7-14 days. The error bars represent ±1 s.d. of qRT-PCR triplicate.

To further explore how the surrounding chromatin influences XIST-induced silencing, we took advantage of the set of existing transgenic cell lines with known genomic locations of the integrated FRT site. As fluorescent reporters allow for efficient screening, we created a plasmid that carries both the inducible repeat A and a DsRED Express2, driven by the mouse *Pgk1* promoter (Figure 3.5A).

Upon induction, repeat A efficiently silenced the reporter gene in the HT1080 2-3-1.0d (Figure 3.5B). The silencing was less prominent in the HEK293 cell line and in the HT1080 2-3-0.5+3#4, the cell line that harbors *EGFP* downstream of the FRT integration site (Figure 3.5B). Interestingly, the silencing of *EGFP* in the HT1080 2-3-0.5+3#4 #3 cell line was also less prominent when compared to the original 5'A transgene (Figure 3.3C). In this cell line, we observed that the cells clustered into two populations in DOX-free media. One population showed *EGFP* and DsRED expression comparable to the other cell lines, while the other population transcriptionally silenced both the *EGFP* and the DsRED express2 transgenes. After we isolated the transcriptionally silent and active cell populations by fluorescence-activated cell sorting (FACS), the cultures reverted to the mixed populations of cells with transcribed and silent reporters (data not shown). In the cell lines tested, removal of DOX from the culture media led to reversal of the transgene silencing within 5 days (Figure 3.5B); DsRED express2 was fully expressed after the cell lines were maintained for 30 days in DOX-free media (not shown). Similar lack of XIST's ability to induce stable (XIST-independent) silencing was also previously reported for the CMV promoter-driven *EGFP*.

To test whether the ability of repeat A to silence the reporter depends on the genomic integration site, we inserted the repeat A – DsRED express2 transgene into six HT1080 cell lines with a known chromosomal location of the FRT integration site. Flow cytometry revealed that repeat A induced efficient reporter gene silencing in multiple clones within all integration sites tested (Figure 3.5C).

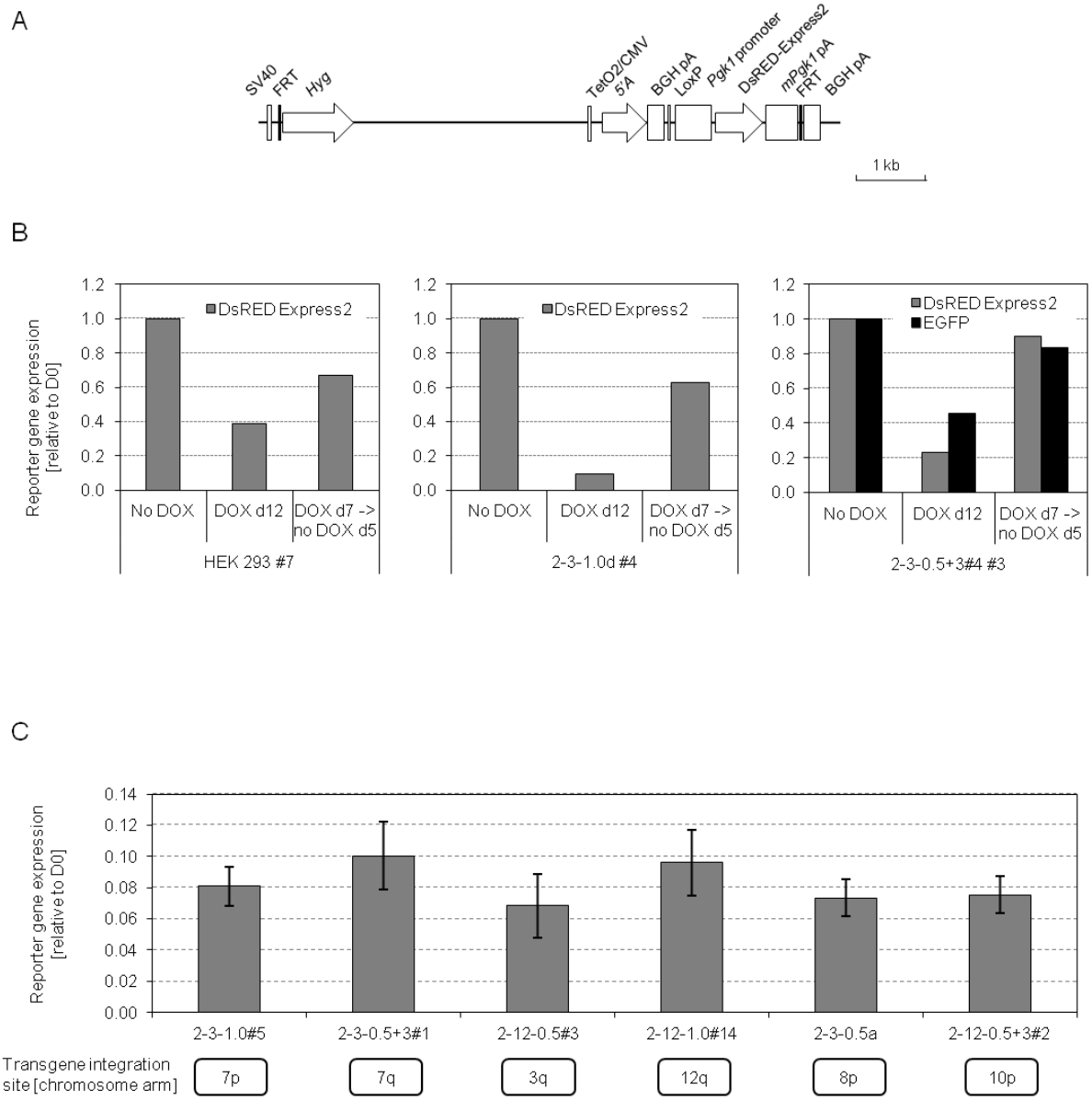


Figure 3.5: Repeat A silences the *Pgk1*-driven fluorescent reporter irrespective of the surrounding genomic context.

(A) Map of the repeat A – DsRED express2 construct.

(B) DsRED express2 and *EGFP* expression, measured by flow cytometry, following induction and subsequent repression of repeat A.

(C) Repeat A expression for 5 days results in robust silencing of DsRED express2. The error bars represent ±1 s.d. of the silencing levels of the individual single-cell clones (N = 8-11). Chromosome arms on which the transgenes are integrated are shown; precise genomic localization of the probed integration sites is detailed in Table 2.1.

Our laboratory has previously identified FRT integration sites in 9 cell lines. However, the precise integration of the FRT sites, and therefore *XIST* transgenes, in two extensively used cell lines was unknown. We have identified the integration sites in both the HEK293 and the HT1080 2-3-0.5+3#4 cell lines by inverse PCR and DNA sequencing (Figure 3.6A, B). Using a combination of conventional PCR and DNA sequencing, we have also established the precise transgenic DNA sequence resulting from the integration of the FRT- and *EGFP*-containing plasmids in the 2-3-0.5+3#4 (Figure 3.6C).

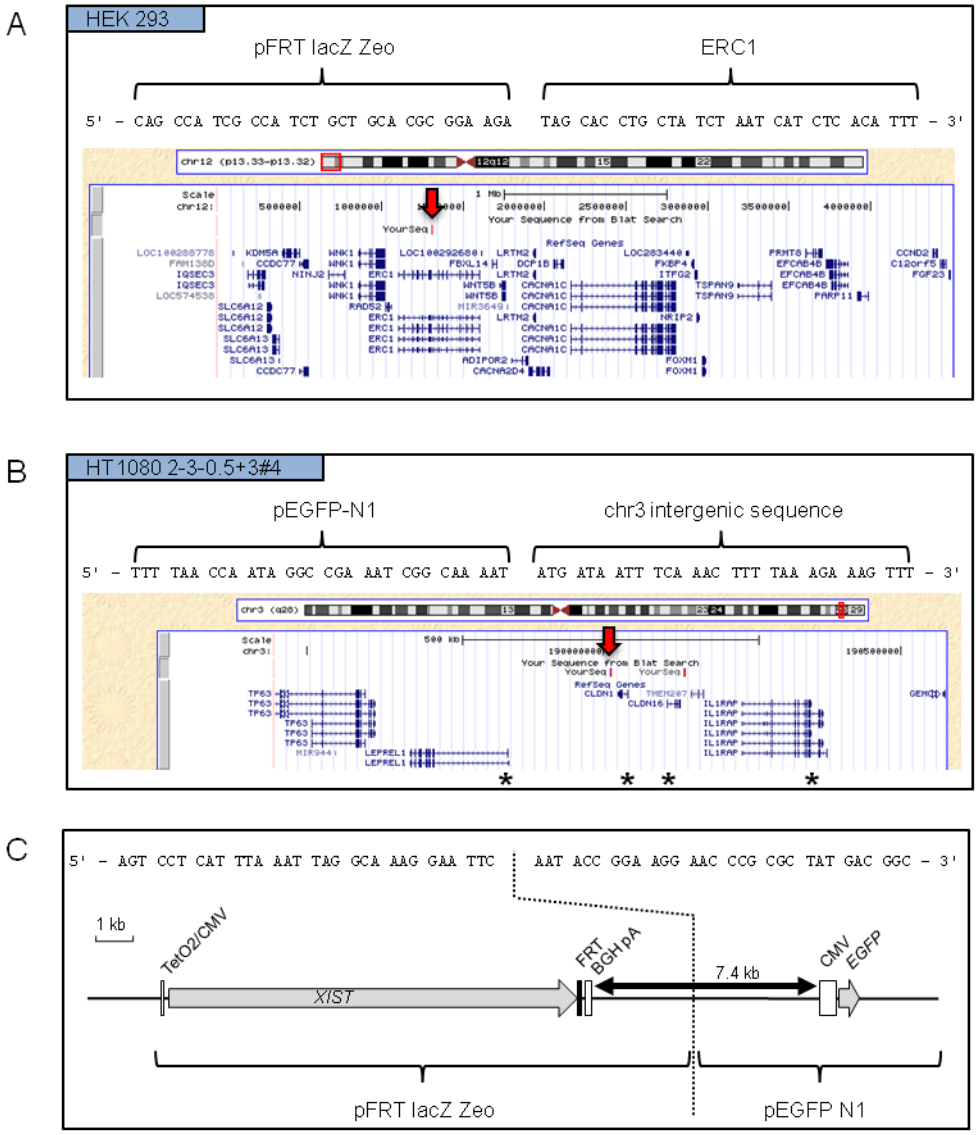


Figure 3.6: Genomic location of FRT sites in HEK293 and HT1080 2-3-0.5+3#4 cell lines.

(A) A view of the UCSC genome browser depicting the FRT integration site location in the HEK293 cell line. The red arrow points to the region of transgene integration. The portion of DNA sequence obtained by inverse PCR which was not part of the plasmid was submitted to BLAT query against human genome assembly hg19 (Feb. 2009). 30 bp of sequence flanking the breakpoint are shown.

(B) As in (A). The integration site in the HT1080 2-3-0.5+3#4 cell line is depicted. Asterisks highlight the genes surrounding the transgene integration site that were tested for XIST-induced silencing.

(C) The map of the transgene in the HT1080 2-3-0.5+3#4 cell line showing 30 bp of the flanking sequences surrounding the breakpoint between pFRT lacZ Zeo and pEGFP N1 integrated plasmids.

To further confirm that silencing results from an XIST RNA-related, sequence-specific effect, we also demonstrated the absence of XIST transcripts which would proceed through the *EGFP* reporter construct. Although some transcripts were present downstream of the polyadenylation site, transcription was completely absent at a site approximately 2 kb 5' of the *EGFP* promoter (Figure 3.7).While the absence of transcripts may also be explained by transcript instability, our conclusion that silencing is not due to transcription interference is further supported by XIST-dependent attenuation of the expression of the *Hyg* gene located upstream of *XIST* (Figure 3.4C) and absence of silencing with vector, or the transgene deleted for repeat A (Figure 3.3B).

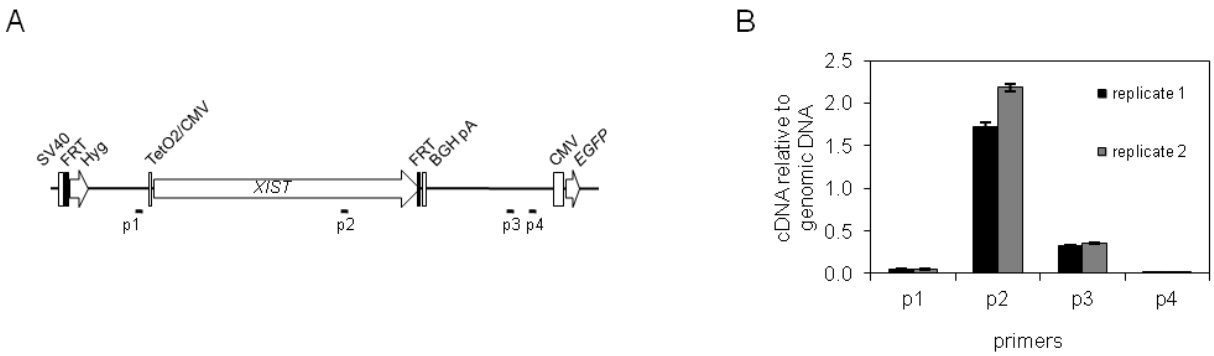


Figure 3.7: Stable *XIST* transcripts do not overlap with *EGFP* gene.

(A) Position of the primer pairs p1-p4 relative to the transgenic *XIST*.

(B) qRT-PCR analysis of expression within full-length *XIST* transgene (p2) and upstream (p1) and downstream (p3, p4) of *XIST* sequence following 5 days of DOX-induced *XIST* expression. Genomic DNA standard was used to normalize for amplification efficiency. The error bars represent ±1 s.d. of the qRT-PCR technical triplicates.

Finally, we explored whether XIST is able to induce silencing of the genes flanking the integration site in the 2-3-0.5+3#4 cell line. qRT-PCR analysis showed that expression of full-length *XIST* induced silencing of *CLDN16*, a gene located approximately 100 kb downstream of *XIST*. The approximately 50% decrease in *CLDN16* transcription is consistent with the complete silencing of the *cis*-located allele (Figure 3.8A). The silencing was however not observed when a construct containing only the repeat A sequence was induced, nor in the construct lacking repeat A. As expected, no silencing of *CLDN16* by the transgene entirely lacking *XIST* sequence was observed. *IL1RAP*, a more robustly expressed gene with a promoter located a further 120 kb downstream (*i. e.* 220 kb from *XIST*) was not subject to XIST-induced silencing (Figure 3.8B). Very low expression levels of *CLDN1* and *LEPREL1* genes prevented a reliable analysis by qRT-PCR.

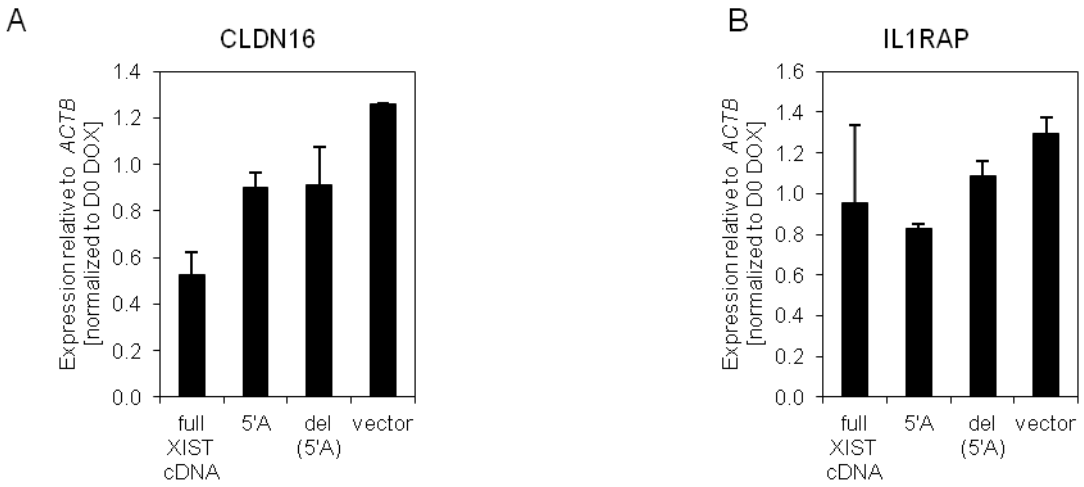


Figure 3.8: *CLDN16* is silenced upon transgenic *XIST* induction.

qRT-PCR analysis of *CLDN16* (A) and *IL1RAP* (B) expression following the induction of *XIST* for 5 days. Gene expression was normalized to *ACTB*. The error bars represent ±1 s.d. of the normalized expression levels for four (full-length *XIST* cDNA) or two (the remaining constructs).

Together, our data show that the previously observed XIST-induced silencing of the reporter gene in the HT1080 human somatic cells [44] can be recapitulated solely by the repeat A sequence in multiple integration sites, in two cell lines (HT1080 and HEK293) and irrespective of the genomic context surrounding the integration site. In total, three different promoters: SV40, CMV and mouse *Pgk1*, located either directly upstream or directly downstream of the *XIST* transgenes responded to XIST-induced silencing. Moreover, we show that *CLDN16*, a gene located approximately 100 kb downstream of the *XIST*, is also subject to XIST-induced silencing.

3.3 Discussion

The purpose of the work presented in this section was to validate and extend our previous observations that the transgenic XIST is able to induce reporter silencing in a differentiated human cell line [44]. The HT1080 model of human X-chromosome inactivation is attractive in several aspects. First, HT1080 is a male cell line, thus endogenous X inactivation does not interfere with the effects of the transgene. The use of single copy FRT sites integrated in the HT1080 genome allows targeting of multiple *XIST* constructs into the same genomic region, thus eliminating different and/or multiple integration sites as a factor influencing XIST's ability to silence. Therefore it is possible to directly compare the effects of multiple constructs at the same integration site as well as the effects of the same construct and reporter at

different integration sites. The DOX-inducible promoter allows for control over *XIST* expression and prevents any negative selection of cells in which XIST-induced gene silencing may be lethal. By inducing *XIST* expression only after stable cell lines were established, the possibly-detrimental effects of ectopic *XIST* expression can be observed. The transgenic model recapitulates some, but not all aspects of endogenous X inactivation. Only a subset of changes associated with X inactivation is observed upon induced *XIST* expression. Broadly, a loss of active chromatin marks akin to that seen in normal X inactivation, but less rapid is observed, however recruitment of inactive marks is substantially impaired. As the transgenic *XIST* in the HT1080 cells is still able to induce gene silencing, the components of X inactivation absent in this system are dispensable at least for a local XIST-induced gene silencing. Last, but not least, the HT1080 cells show rapid growth and good viability under a broad range of culturing conditions.

We have shown that the inducible full-length *XIST* cDNA transgene is transcribed at approximately equal amount as endogenous *XIST* in female lymphoblasts (Figure 3.1). The extent of transcription along the transgene was constant between the 5' end of exon 1 and exon 6; the *XIST* transcription dropped to approximately 50% by exon 7. As the expression of the transgenic *XIST* depends on, and thus must originate solely from, the inducible CMV/TetO2 promoter, we conclude that most transcripts span the majority of *XIST* cDNA. These results demonstrate that only a single XIST isoform is needed to induce gene silencing and suggest that the multiple spliced isoforms of XIST (discussed in Section 1.3.1) either lack functional significance, or have other roles, possibly in the course of initiation of X inactivation.

We next focused on delineating the sequence of XIST that is critical for silencing. Previously, our laboratory has showed the repeat A-lacking *XIST* transgene fails to form XIST foci and to induce gene silencing [44]. We have now confirmed the lack of gene silencing by qRT-PCR (Figure 3.3B) and shown that despite losing the ability to localize to chromatin in *cis*, the del (5'A) transcript remained localized in the nucleus. This finding contrasts with the observation in mouse ES cells, where a repeat A-lacking construct was able to cover the *cis*-linked chromosome [42]. Recently, YY1 was shown to be indispensable for Xist localization in mouse ES cells which requires YY1 binding both to the *Xist* DNA, via YY1 binding sites in repeat F and to the Xist RNA, via the murine-specific repeat C [54]. The del (5'A) construct lacks not only the repeat A, but also the region harboring YY1 binding sites that is syntenic to mouse repeat F. Moreover, as human XIST practically lacks repeat C [34], either YY1 interacts with different regions of XIST, or the localization of XIST to the Xi is in human controlled by an entirely independent mechanism. Thus, the difference between the mouse and the human repeat A-lacking transgenes in their capacity to localize may either be caused by deletion of regions outside of the repeat A in the human construct, or may be species- or cell type-specific. For example, the delocalization

of the repeat A-less XIST in the HT1080 cells may be due to a reduced amenability of chromatin to epigenetic modification and chromatin remodeling in the differentiated cells in comparison with the plastic chromatin of ES cells, or due to an absence of a protein factor that is critical for engaging sequences 3' of repeat A in forming of the transcriptionally silent XIST domain.

The limited understanding of why some spliced and polyadenylated lncRNAs, including XIST, show nuclear localization leaves us to speculate why the del (5'A) transcript remains nuclear after losing the ability to localize in *cis*. Current models assume that XIST interacts with an array of proteins to form a heterogeneous complex which comprises a nuclear compartment not permissible to transcription (reviewed in [8]). We propose that these interactions *per se* are the reason why XIST is not exported to the cytoplasm and that the repeat A-lacking transcript retains some of these interactions to remain nuclear, but lacks the interactions mediated by the repeat A region to from a nuclear compartment, and thus cannot induce gene silencing. Alternatively, XIST RNA, full length or truncated, may remain nuclear simply because it avoids interaction with proteins involved in RNA exporting from the nucleus [72].

Truncated *Xist* transgenes that retained the repeat A region were previously shown to induce gene silencing [42] and a similar observation was made by Chow *et al.* [44] in the HT1080 transgenic system. We have extended these analyses by showing that a construct combining the previously tested deletions [44] also retained the ability to silence (Figure 3.2B). Transposing the repeat A region from its normal 5' end location to the 3' end also did not affect gene silencing (Figure 3.2C). While the previously published studies showed that repeat A is necessary for gene silencing, constructs consisting of only repeat A were not tested. Our results show that transcripts consisting of only the human 5'A repeat maintain nuclear localization (Figure 3.2D). We propose that similar to the repeat A-lacking transcript, the nuclear localization of repeat A is due to binding of XIST RNA with proteins that interact with chromatin; however, which sequence features of repeat A are required to maintain its nuclear localization or what proteins are involved is currently not known. Finally, the 5'A repeat construct retains the full silencing potential of the whole XIST and thus, that the repeat A is both necessary and sufficient to induce gene silencing (Figure 3.3A-C).

As the ectopic *XIST* expression in the HT1080 cells does not likely lead to a chromosome-wide silencing, we were unable to systematically test whether repeat A alone can induce gene silencing outside of the region from which it is transcribed. However, our finding that expression of the full-length *XIST* cDNA, but not of the repeat A construct in the 2-3-0.5+3#4 cell line induces silencing of *CLDN16* (Figure 3.8A), a gene located approximately 100 kb downstream of the transgenic site (Figure 3.6B),

suggests that XIST requires the sequences 3' of repeat A in order to form a functional silencing compartment.

The ability of XIST to spread in *cis* and induce transcriptional silencing is modulated by yet unknown chromatin features, as demonstrated by the limited inactivation of autosomal chromatin in X:autosome translocations (reviewed in [11]). We used two approaches to test a hypothesis that different *XIST* integration sites in the HT1080 cells will show varying degree of reporter gene silencing. First, we surveyed existing cell lines with transgenic full-length *XIST* integrated in different genomic loci. In the absence of DOX, *XIST* expression is suppressed by tetracycline repressor (*TetR*). The HT1080 cell lines described here originate from two different HT1080 cell lines harboring *TetR* transgene (2-3 and 2-12). We observed that while *TetR* expression was higher in the 2-3-derived cell lines, the lower TetR expression in the 2-12-derived cell lines was fully sufficient to suppress *XIST* expression in the absence of DOX (Figure 3.4A, B). In all the clones we examined, *XIST* expression was induced following DOX treatment, however the extent of upregulation and the absolute levels of XIST RNA (normalized to *ACTB*) varied considerably among the clones, and to some extent may be dictated by the chromatin structure surrounding the integration sites (Figure 3.4B). Similarly, *XIST* expression led to silencing of *Hyg* in all tested clones, although *Hyg* expression and the extent of silencing showed substantial clone to clone variability (Figure 3.4C).

The variable *XIST* expression levels and differences in transgene silencing may be intrinsic to the integration site or alternatively, they may be a result of stochastic events, as each cell line is derived from a single-cell colony. We took advantage of the finding that a short fragment of XIST, the repeat A, is sufficient for the gene silencing and created a construct that carries both the repeat A and a fluorescent reporter. Based on our experience with the HT1080 transgenic system, we modified several features to create a system that is better equipped to answer how repeat A induces gene silencing (Figure 3.5A). Namely, we used the DsRED express2 fluorescent reporter for its comparatively shorter half-life and low toxicity. Avoiding the use of EGFP allows greater flexibility in future experiment designs as EGFP is often utilized as a marker of expression (*e. g.* in shRNA screens). Further, the DsRED express2 is driven by the mouse *Pgk*1 promoter in the construct we generated. This avoids the complication in designing specific qPCR primers that we encountered in the *EGFP* transgene, where the inducible promoter was derived from CMV which also drives *EGFP*. Compared to the CMV promoter of viral origin, a mammalian promoter may be a more biologically relevant target to acquire epigenetic changes induced by repeat A. Finally, the new construct also harbors a LoxP site which will facilitate insertion of various DNA elements to test their role in suppressing X inactivation.

DsRED express2 was efficiently silenced following 12 days of repeat A induction in a 2-3-1.0d cell line (Figure 3.5B). The silencing was less efficient when the transgene was integrated into the HEK293 cell line. While the absolute extent of repeat A expression or its relative up-regulation was not tested, in our hands, a wide range of absolute *XIST* transgene expression levels and even as low as 5-fold up-regulation following treatment with DOX is sufficient to induce reporter gene silencing. This is in contrast with the immunofluorescence-based experiments which showed that compared to the HT1080 transgenes, XIST was better able to recruit marks of inactive chromatin in the HEK293 cells [44]. Silencing of both *EGFP* and DsRED express2 was attenuated in the HT1080 2-3-0.5+3#4 cell line. This is probably due to a transgene silencing observed in a subset of cells which occurs independently of *XIST* expression and is likely an X inactivation unrelated artifact. The reporter silencing required continuous repeat A expression. The inability to induce stable gene silencing was also observed for the full-length *XIST* transgene [44] and is consistent with the absence of recruitment of DNA methylation and other chromatin marks acquired late in X inactivation.

Finally, a flow cytometry screen of variation of DsRED express2 silencing revealed that despite a moderate variation among the individual clones, repeat A induced strong gene silencing in the six different integration sites we tested (Figure 3.5C). While our results indicate that the surrounding chromatin environment does not modulate the ability of repeat A to induce local gene silencing, we note that the FRT-containing plasmid may preferentially integrate into accessible chromatin regions and thus, that our screen of the existing clones with randomly integrated FRT sites may not capture the whole spectrum of chromatin states.

Overall, we showed that the repeat A region of XIST alone is sufficient to induce robust silencing of two different transgenic reporters, the CMV-driven *EGFP* and the *Pgk1*-driven DsRED express2 in multiple cell lines, as well as an endogenous gene located downstream of the transgene in one of the cell lines. These results demonstrate the robustness of the HT1080 transgenic system that has been critical for dissection of the functional elements within the repeat A region.

4 MINIMAL SEQUENCE OF XIST RNA AND STRUCTURAL REQUIREMENTS FOR GENE SILENCING

The candidate (Jakub Minks) designed, performed and analyzed all experiments presented in this section with the following exception:

Sarah Baldry, a member of the Brown laboratory, has performed all experiments required to transfect repeat A-derived transgenes into the HT1080 cells.

4.1 Introduction

The repeat A region of XIST/Xist was previously shown to be necessary for gene silencing in both mouse and human [42, 44]. As detailed in section 3, we have also demonstrated that repeat A alone is sufficient to induce gene silencing. Understanding which sequences within the repeat A are critical for its function would add a new layer of resolution to the ongoing effort to describe, on a molecular level, how XIST achieves X inactivation. A number of sequence and structural elements within repeat A may be important in this process.

Repeat A consists of 24 bp-long CG-rich core sequences that are the best conserved XIST sequences amongst eutherians. These sequences are separated by approximately 20–50 bp-long T-rich spacers (Figure 4.2A). The CG-rich core is formed by two palindromes, each of which is broken by 4 bp-long sequences. In contrast with other repeat sequences within XIST, the number of repeat A monomers is also well conserved (see section 1.3.1). The lack of variation in repeat A sequence both among species and between individual repeats of each species suggests that they are critical for XIST function, and are likely involved in protein binding. The conservation of the number of repeat A units suggests either that the repeat A functions as the whole and higher or lower number of units would interfere with the proper repeat A structure, or that the putative protein cooperatively binds to repeat A in several copies and 8-9 repeat A units allow for optimal dynamics and/or extent of silencing. The number of repeat A units was previously reported to correlate with the ability of Xist to induce silencing in differentiating mouse ES cells [42].

The palindromic nature of the repeat A core sequences strongly suggests their involvement in forming a distinct secondary RNA structure. Several alternative but mutually exclusive structures were previously suggested (Figure 4.1). The first model proposed that each of the two palindromes forms a hairpin and thus, the repeat A region of XIST RNA folds into a two-hairpin 8- or 9-mer [42]. However, an *in vitro* analysis of repeat A structure by fluorescent resonance energy transfer (FRET), as well as sensitivity to RNases that specifically digest single- or double-stranded RNA regions, proposed an alternative structure. The first palindrome encompassing the ‘ATCG’ tetraloop was suggested to engage in pairing between two separate monomers, and not within repeat A monomers, and the model proposed that the second palindrome does not form a defined structure [73]. Nuclear magnetic resonance (NMR) analyses of repeat A monomer and dimer structures revealed that under *in vitro* conditions, the first palindrome formed a hairpin, while the second palindrome engaged in pairing between repeat A units [232, 233]. Thus, the primary sequence of repeat A shows a strong propensity to form secondary structures. The exact nature of the repeat A structure is still unclear and its solution may require *in vivo* experiments, as repeat A-binding proteins likely stabilize the RNA structure.

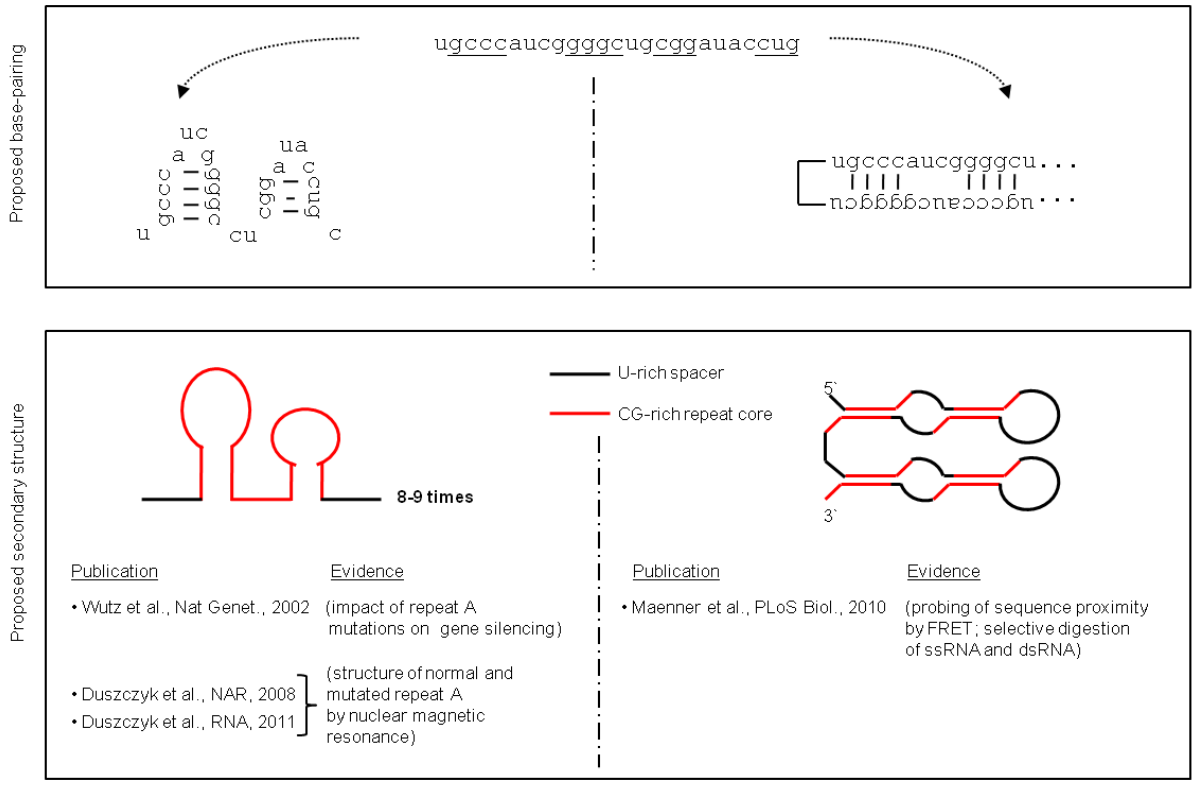


Figure 4.1: Competing models of repeat A structure

Simplified representation of two models of repeat A structure. Canonical sequence of repeat A core is shown in the top panel. The two palindromic sequences within the repeat A core are underlined and their engagement in forming intra- versus inter-unit base pairs is shown. For simplicity, the intra-unit pairing is also shown for the second palindrome, however experimental data does not fully support its formation [233]. The bottom panel illustrates the overall structure of repeat A and lists the evidence in support as of either of the two models.

So far, two protein partners of repeat A have been identified. A splicing factor ASF/SF2, which is critical for Xist RNA accumulation and processing, interacts with repeat A [74], however it is likely not directly involved in gene silencing. The other known repeat A partner is PRC2, the protein complex responsible for deployment of H3K27me3, a histone modification associated with transcriptionally repressed chromatin and enriched on the Xi. While it has been shown to interact with repeat A, we demonstrate in section 5 that PRC2 is not necessary for repeat A-induced gene silencing.

Therefore, while repeat A is known to be critical for XIST-induced silencing, neither the necessary sequence and structure of the repeat A nor the repeat A-interacting proteins are known. In this section, we take advantage of the transgenic HT1080 system to identify the sequence and structural elements within repeat A that are critical for its function by creating a series of truncated and mutated repeat A

constructs, transfecting them into the 2-3-0.5+3#4 cell line and testing the ability of these repeat A transgenes to induce silencing of the *EGFP* reporter.

4.2 Results

4.2.1 Repeat A monomers additively contribute to silencing

In order to dissect the link between repeat A sequence and its silencing ability, we generated new constructs that eliminated the potential confounding effect of sequence variations in the individual monomers, particularly in the T-rich linker regions. We created an artificial repeat A consisting of a nine-fold repetition of a 46 bp consensus monomer sequence, and containing restriction enzyme sites in the T-rich stretches to further allow for the creation of constructs with reduced numbers of repeats (Figure 4.2A). Flow cytometry data showed that the artificial repeat A silences *EGFP* to the same extent as full-length XIST or human repeat A constructs (Figure 4.2B). Since variability within the individual repeats and spacer regions did not contribute to silencing we were able to test the silencing ability of constructs with fewer repeats. Transgenes harboring 2–6 repeat A monomers were functional, with a linear relationship between the number of repeats and their silencing ability (Figure 4.2B). While we observed as high as a 3.3-fold difference in expression levels of the individual artificial repeat A constructs (data not shown), expression levels appear to fluctuate randomly, do not correlate with the ability to silence and often vary for the same construct between individual biological replicates. A time course experiment showed that silencing induced by the repeat A 2-mer gradually increased between day 2 and approximately day 8, however longer induction of the repeat A 2-mer did not promote further *EGFP* silencing (Figure 4.2C).

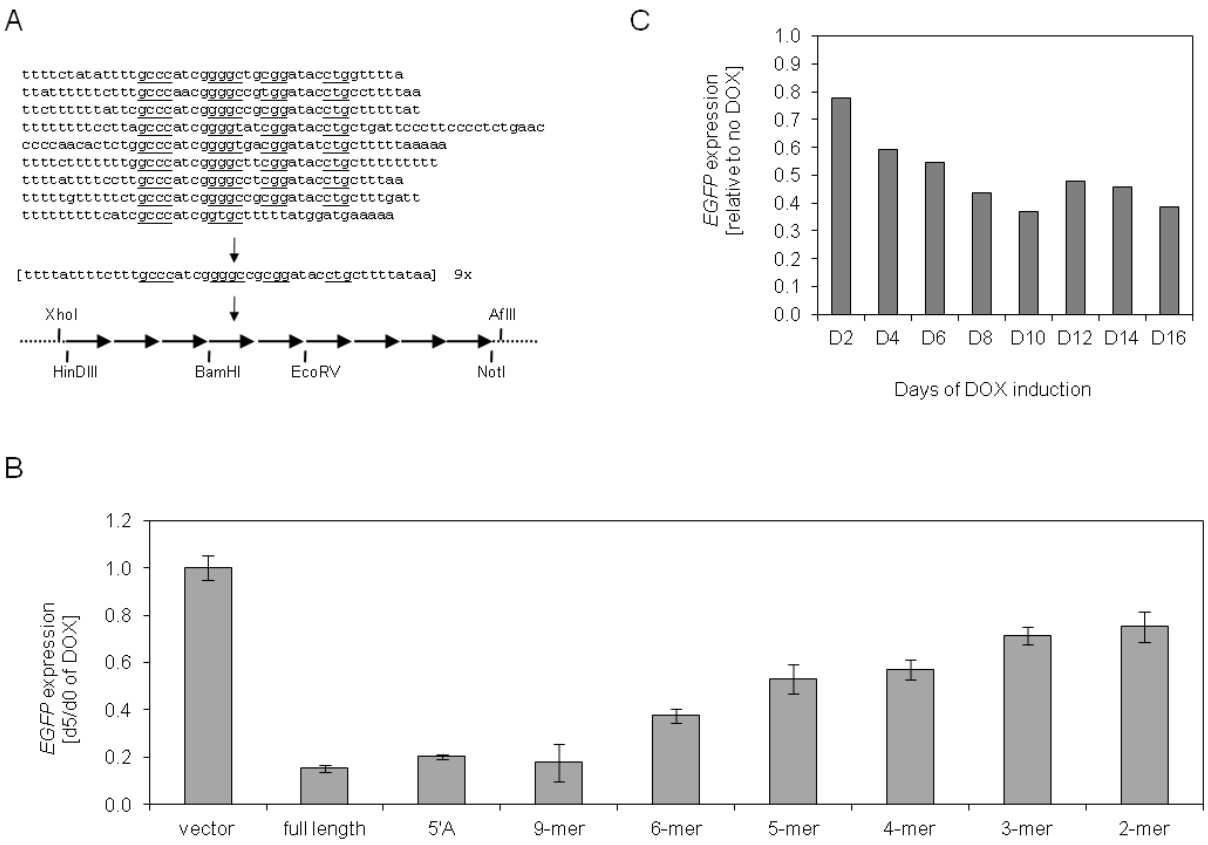


Figure 4.2: Repeat A monomers additively contribute to silencing.

(A) Human repeat A sequence consists of 8.5 copies of a well-conserved CG-rich core and T-rich spacer sequences. Palindromic sequences hypothesized to form a secondary structure are underlined. Artificial repeat A was constructed as a 9-mer repetition of consensus monomer sequence and restriction enzyme sites were introduced to allow for the creation of shorter constructs.

(B) *EGFP* expression following 5 days of transgene induction as measured by qRT-PCR, relative to d0 and normalized to changes in expression caused by induction of the vector alone and to *ACTB* expression (N = 2). Error bars indicate ±1 s.d.

(C) *EGFP* expression was measured by flow cytometry every 2 days for 16 days following induction of repeat A 2-mer. *EGFP* expression in cells that were not induced with DOX served as a control.

Remarkably, even the 2-mer repeat A construct partially silenced *EGFP*, providing us with a well-defined template for further dissection of the relationship between repeat A sequence and its silencing ability. Repeat A monomers were previously predicted to form 2-hairpin CG-rich structures with the T-rich stretches serving as spacers [42]. While alternative structures have since been proposed, for simplicity, we refer to the four components of CG-rich consensus core as stem 1, loop 1, stem 2 and loop 2 (Figure 4.3A). We created four variants of the 2-mer repeat A to probe the role of these elements (Figure 4.3A). All mutations completely ablated the transgenes' ability to silence *EGFP* compared to a

canonical repeat A 2-mer, as measured by flow cytometry of two representative clones for each mutation (Figure 4.3B), and analysis by qRT-PCR showed the same trends (Figure 4.3C). Thus, the most conserved regions of *XIST* both among the individual repeats in human (Fig. 2A) and among different species (Figure A.1), the CG-rich palindromes and their intervening 'ATCG' and 'ATAC' sequences, are critical for XIST function.

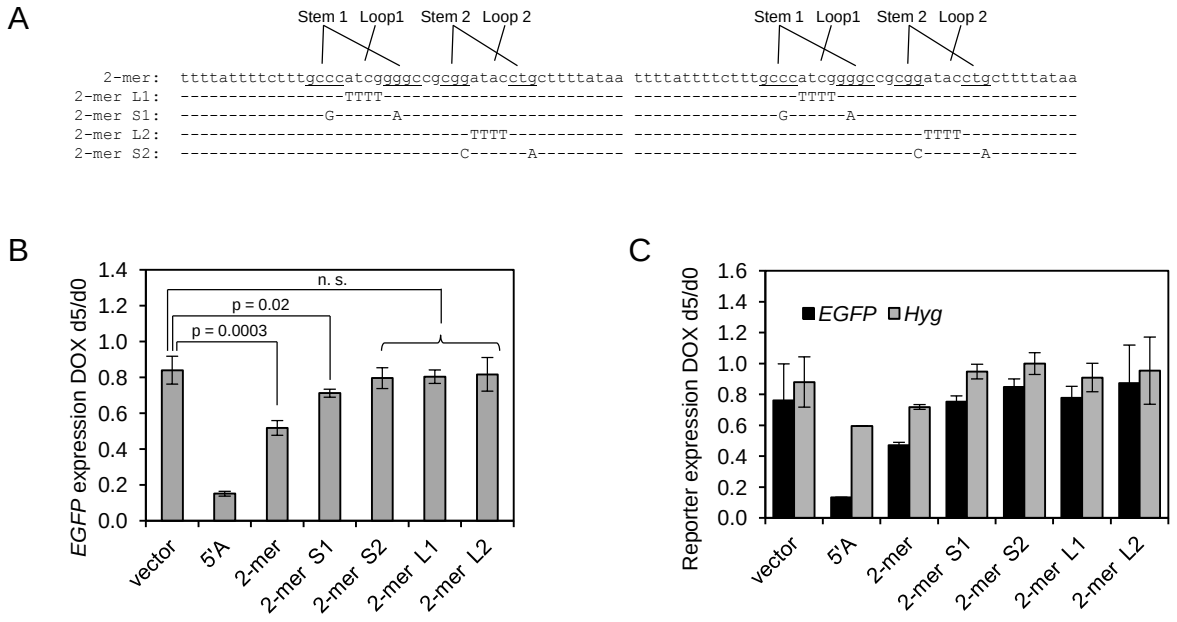


Figure 4.3: Mutation of the core repeat A sequences abrogates its silencing ability.

(A) Sequence of the canonical repeat A monomer and four mutant constructs created to target the hypothesized repeat A hairpins. Underlined sequences correspond to stem 1 and stem 2. Dashes indicate no sequence change.

(B) Mean *EGFP* expression following 5 days of transgene induction, measured by flow cytometry and normalized to *EGFP* expression in uninduced cells (d0) (N = 2).

(C) *EGFP* and hygromycin resistance gene (*Hyg*) expression following 5 days of transgene induction, measured by qRT-PCR and normalized to *ACTB* and relative to *EGFP* expression in uninduced cells (d0) (N = 2).

Taking advantage of the well-defined 2-mer repeat A transgene, we used mfold [234] to design a quartet of mutations that were predicted to enforce pairing either within (A1, A2) or between (B1, B2) each monomer (Figure A.2 and Figure 4.4A). Measured by flow cytometry, the mutants that were predicted to enforce the pairing within each monomer functioned better than those enforcing the interaction between

the monomers; although, none of the four mutants silenced *EGFP* as efficiently as the canonical repeat A 2-mer (Figure 4.4B).

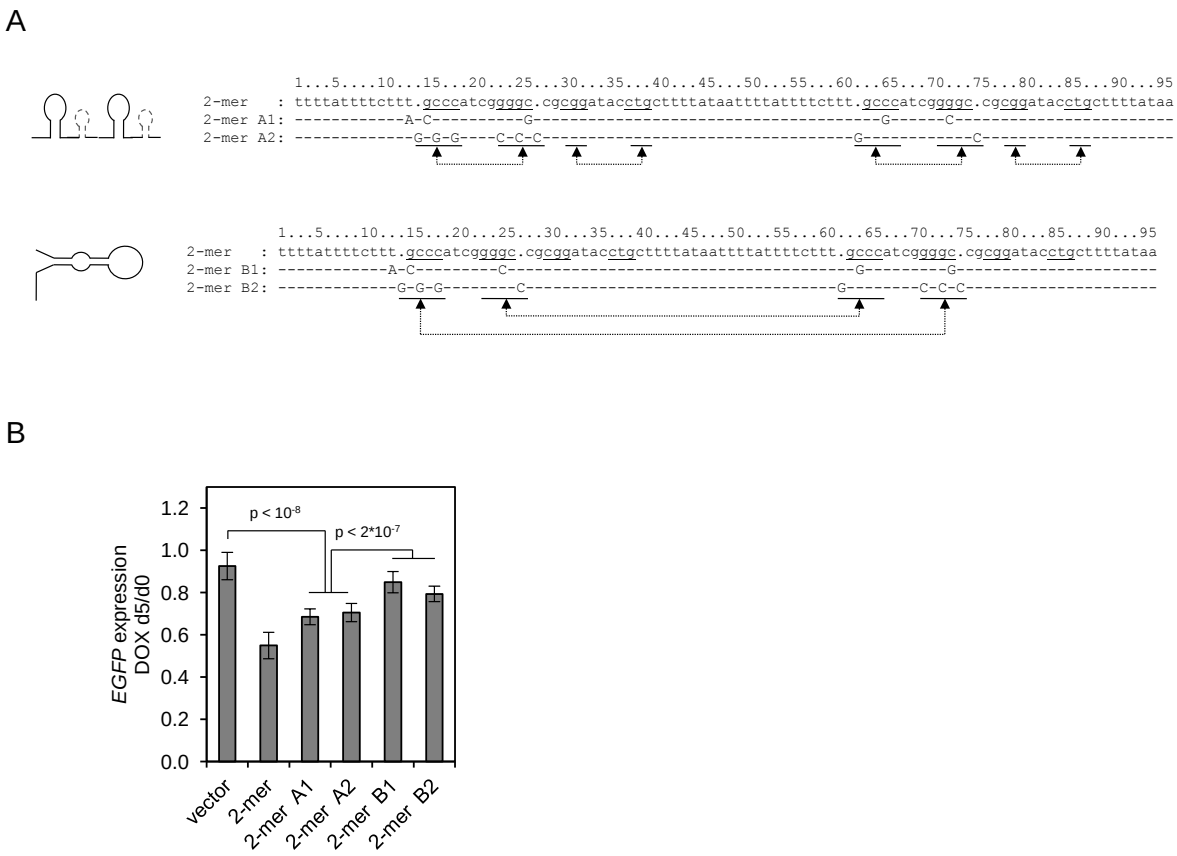


Figure 4.4: Silencing ability of 2-mer repeat A construct is retained when forced to form the stem-loop 1 structure but abrogated when the alternative structure is enforced.

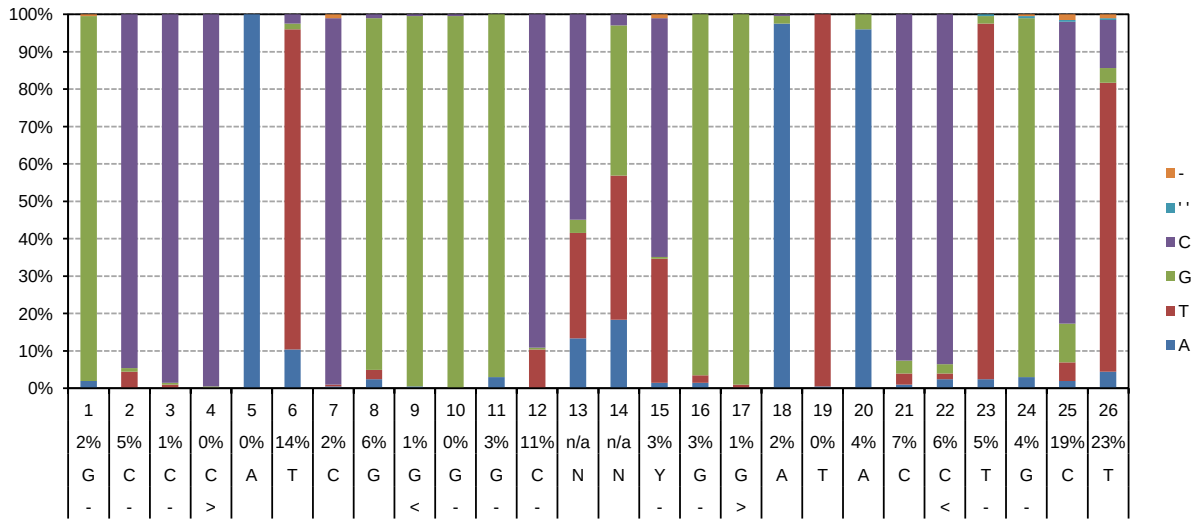
(A) Sequence of the canonical repeat A 2-mer and four mutant constructs that either enforce formation of stem-loop 1 (A1, A2) or an alternative folding (B1, B2) of repeat A sequences, as indicated by schematics. Dashes indicate no change in sequence.

(B) Mean *EGFP* expression following 5 days of transgene induction, measured by flow cytometry and normalized to *EGFP* expression in uninduced cells (d0) (N=7, two-tailed paired t-test).

4.2.2 Survey of repeat A mutations shows strong preference for stem 1 and mild preference for stem 2 formation

To leverage the increasing number of sequenced mammalian genomes, we created a repeat A alignment of 27 mammalian species (Figure A.1). As expected, repeat A was well conserved, in particular within the CG-rich core sequences (Figure 4.5A). Of the defined stem-loop structures, loop 1 showed the highest frequency of deviation from the canonical ‘ATCG’ sequence (Figure 4.5B), with approximately 10% (20/202) of repeat A units harboring an ‘AACG’ tetraloop instead.

A



B

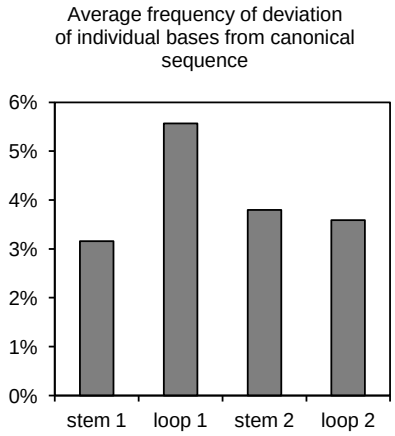


Figure 4.5: Sequence conservation of repeat A units among 27 mammalian species.

(A) Sequence conservation of 202 core repeat A units among 27 mammalian species. Lines on the X axis depict (from top to bottom) position of bases, percent of units that deviate from canonical sequence, the canonical sequence and arrows corresponding to bases forming the hypothesized stem 1 and stem 2.

(B) Average frequency of deviation from canonical sequence in the two putative stem-loops.

Taking advantage of the wealth of natural mutations, we asked whether a reciprocal mutation exists in the same species that would re-create a fully complementary double stranded sequence either within the same unit, or with another unit. Of the 50 stem 1 mutations we analyzed, 24 could not be linked with a reciprocal mutation, suggesting that while the uncompensated deviations from the canonical loop 1

sequence are not frequent, they are viable as, presumably, the overall ability of repeat A to induce silencing is retained via the remaining repeat A units. Twelve of the remaining 26 mutations were accompanied by a reciprocal mutation exclusively within the same unit, and further 10 could pair either within the same unit, or with another unit (Figure 4.6A, C). These findings strongly argue in favor of the predicted stem-loop 1 formation. Survey of stem 2 mutations uncovered 46 deviating repeat A units, 28 of which could not pair with any reciprocal mutation. Of the remaining 18 mutants, 8 could exclusively form a stem-loop by pairing within each unit, with a further 3 allowing for pairing both within a unit and with other units (Figure 4.6B, D). While the propensity of stem 2 region to harbor reciprocal mutations that would allow for stem-loop 2 formation is less striking than that of stem 1, it is still remarkably high. If the rate of reciprocal mutations was stochastic, mutation in any repeat A unit would occur with the same frequency and therefore only about 11% of reciprocal mutations would be expected to occur within the same unit in a 9-unit-long repeat A sequence. However, the frequency of reciprocal mutations that occur solely within the same unit was 46% for stem 1 and 44% for stem 2. This argues either that stem 2 indeed forms a stem-loop by pairing within each unit, or that repeat A structure is species-specific and may involve a combination of both modes of pairing.

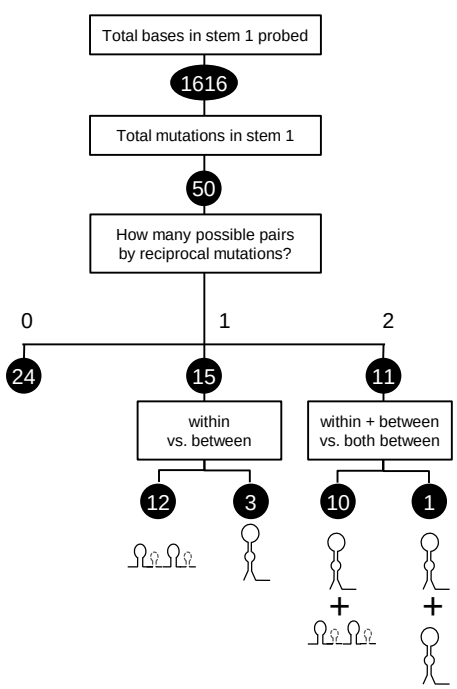
A

Stem 1		Creates pair with unit								
		1	2	3	4	5	6	7	8	-
Mutation in unit	1	0	0	0	0	0	0	0	0	2
	2	0	8	0	0	2	2	0	0	4
	3	0	1	0	0	0	1	0	0	1
	4	0	0	0	0	0	0	0	0	5
	5	0	3	0	0	10	0	0	1	10
	6	0	3	1	0	0	4	0	0	0
	7	0	0	0	0	0	0	0	0	1
	8	0	0	0	0	1	0	0	0	1

B

Stem 2		Creates pair with unit								
		1	2	3	4	5	6	7	8	-
Mutation in unit	1	0	2	1	1	0	0	1	0	7
	2	1	8	4	3	0	1	3	3	6
	3	0	1	0	0	0	0	0	0	0
	4	1	0	0	0	0	0	0	0	3
	5	0	0	0	0	0	0	0	1	3
	6	0	0	0	0	0	0	0	0	1
	7	0	0	0	0	0	0	0	0	1
	8	0	1	0	0	0	0	0	4	7

C



D

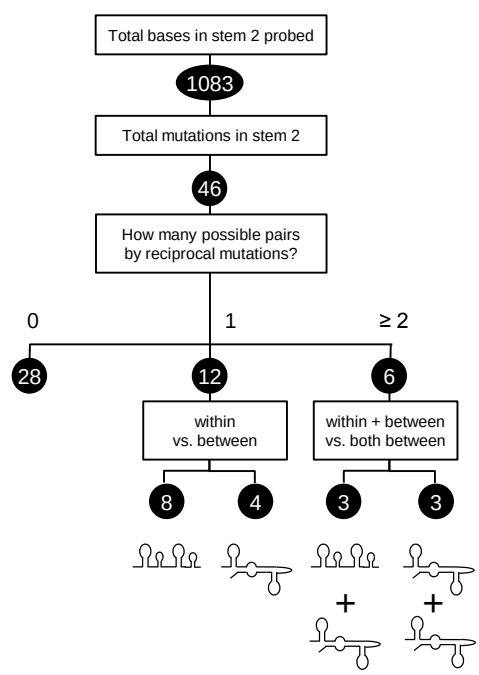


Figure 4.6: Frequency of reciprocal mutations within stem 1 and stem 2 suggests preference for intra-unit pairing.

(A) Analysis of reciprocal mutations in the stem 1 of individual repeat A units. The table depicts the number of occurrences when mutation in a repeat A unit would allow pairing due to the existence of a reciprocal mutation within the same unit (highlighted in gray), in a different unit, or when no reciprocal mutation exists in the species' repeat A (listed in the last column).

(B) As in (A), but stem 2 is analyzed.

(C) Analysis of individual deviations from canonical repeat A within stem 1. Individual repeat A units are categorized by their theoretical ability pair within the units or with another unit to form a complete stem.

(D) As in (C), but stem 2 is analyzed.

4.3 Discussion

The silencing of an adjacent *EGFP* reporter is achieved through an additive effect of repeat A monomers, with even a 2-mer repeat A inducing partial *EGFP* silencing. The repeat A 2-mer also induced partial gene silencing of a gene located approximately 100 kb downstream of the transgene integration site (Figure 3.8A). These observations provide strong evidence that repeat A functions through additive action of the individual units. The ability of a mere repeat A 2-mer to induce partial gene silencing not only locally but also over a long distance is surprising and supports the model of additive repeat A action. Notably, the extent of *EGFP* silencing reached equilibrium after 8-10 days of repeat A 2-mer expression, while the full *XIST* cDNA transgene was previously shown to induce further silencing beyond this time frame [44]. Thus, the number of repeat A monomers affects both dynamics and extent of silencing and may explain why the number of repeat A units remains essentially constant in species that show otherwise substantial variation in repeat structure of *XIST*. In agreement with a previous report on mouse *Xist* [42], artificial repeat A retains full silencing potential when compared to human repeat A. This suggests that neither the sequence variations within the CG-rich core nor the varying length of the T-rich spacers in individual repeat A monomers is essential for *XIST* function.

The core repeat A sequence consists of two palindromes; the first allowing for perfect C-G pairing broken by ‘ATCG’ and the second involving a G-U pair broken by ‘ATAC’. Several secondary structures of repeat A have been proposed based on analysis of repeat A mutants [42], NMR data, [232, 233] and RNase footprinting and FRET data [73]. All of the proposed structures predict the existence of an ‘ATCG’ loop. Indeed, mutation to ‘TTTT’ (Figure 4.3) completely abolishes repeat A function and mutation to ‘TAGC’ in mouse partially abolishes Xist function [42]. The first palindrome was suggested to form either a hairpin by pairing within each monomer [42, 232, 233] or alternatively, between monomers [73]. The ability of repeat A 2-mer to induce gene silencing allowed us to use mfold, an RNA structure prediction algorithm [234], to design repeat A mutants that would address which of the two structures is functional. In our hands, modeling of larger than 2-mer repeat A structures was highly unreliable as multiple structures of similar minimum free energies (ΔG) were predicted. We note that while we designed the 2-mer mutants so that only one structure would be favored by mfold predictions,

we have not experimentally confirmed (e. g. by employing NMR, FRET or RNase footprinting) the structure of any of the described repeat A constructs.

Our experimental data (Figure 4.4) and assessment of evolutionary sequence conservation (Figure 4.6A, C), support the intra-repeat pairing model, consistent with outcomes observed in mice [42], that the first palindrome indeed forms a stem to expose the 'ATCG' tetraloop. Notably, neither of the mutants designed to enforce the intra-repeat pairing silenced *EGFP* as efficiently as the repeat A 2-mer (Figure 4.4). This may indicate either that the specific sequence, rather than the structure is critical for the function of repeat A units, or that the predicted structures are inaccurate. Indeed, the mfold algorithm is unable to predict some secondary RNA structures (e. g. pseudoknots) and comparison of mfold results with known RNA structures shows that the ability of mfold to predict the correct RNA structure is greatly dependent on the RNA length and sequence [235].

The mutations we introduced to the second palindrome resulted in a complete loss of silencing ability (Figure 4.3), supporting the importance of these sequences; however, these mutations did not directly address the precise secondary structure. While the second palindrome was proposed to pair within each monomer to form the second stem-loop [42], recent studies suggest that the secondary structure may rather involve pairing between individual repeat A monomers [232, 233] or with the T-rich spacers [73]. Our assessment of evolutionary sequence conservation provides evidence in favor of second stem-loop formation, though the frequency of compensatory mutations is less striking in comparison with stem-loop 1 (Figure 4.6B, D). In conclusion, we and others have now shown that the core CG-rich repeat A sequences are central to XIST/Xist silencing function. While the overall secondary structure of mammalian repeat A remains to be solved, current models favor formation of at least one stem-loop exposing 'AUCG' sequence, inviting speculations that this array of 9 stem-loops may serve as a multimerization platform for binding of XIST/Xist partners.

5 PROTEINS CRITICAL FOR XIST-INDUCED SILENCING

The candidate (Jakub Minks) designed, performed and analyzed all experiments presented in this section with the following exception:

Angela Kelsey, a member of the Brown laboratory, has performed all fluorescent *in situ* hybridization experiments.

5.1 Introduction

A number of proteins have been implicated to play a part in X-chromosome inactivation, including proteins involved in chromatin compaction and organization, writers, readers and erasers of histone marks, transcription factors, and cell cycle regulators. Although there is no explicit evidence that the XIST transcript *per se* cannot induce gene silencing, proteins and protein complexes have been shown to interact directly with Xist RNA. Similar interaction between proteins and RNA was observed in other lncRNAs that induce gene silencing (reviewed in sections 1.4 and 1.3.3).

While much is known about the differences between the active X and the inactive X, how *XIST* transcription leads to the *cis*-linked transcriptional silencing remains elusive. In fact, the complex and highly dynamic epigenetic changes that occur in the course of X inactivation pose a difficulty in separating the changes that drive X inactivation from those that are merely a consequence of the XIST-induced transcriptional repression. The HT1080 transgenic inducible *XIST*, combined with fluorescent reporters, provides an excellent system to study the chromatin changes associated with the observed gene repression. Indeed, our laboratory has previously shown that ectopic *XIST* expression induces some, but not all chromatin changes associated with X inactivation [44]. Specifically, chromatin IP analysis has shown a decrease in H4 acetylation followed by a decrease in H3K4 dimethylation and trimethylation and recruitment of HP1γ and H4K20me1 at the *EGFP* promoter. H3K9me2 was not increased following 7 days of *XIST* expression and even prolonged *XIST* expression did not induce DNA methylation at the *EGFP* promoter. Combined immunofluorescence and fluorescent *in situ* hybridization studies showed that while expression of transgenic *XIST* in HEK293 cells leads to an accumulation of macroH2A, H3K27me3 and H4K20me1 foci that co-localized with the XIST signal, no such foci were observed to co-localize with XIST foci in HT1080 cells (unpublished).

Histone deacetylation is among the earliest chromatin changes occurring at the onset of X inactivation; approximately 50% of mouse ES cells show Xi-specific histone deacetylation 2–3 days after the onset of differentiation [102]. In mammals, histone acetylation at lysine residues of N-terminal core histone tails is associated with transcriptional activation (reviewed in [236-238]). The removal of acetyl groups from histone tails is carried out by histone deacetylases (HDACs). HDACs show differences in sub-cellular localization [239] and apart from their role in histone deacetylation, HDACs, as well as their counterparts, histone acetyltransferases, also regulate acetylation of a number of non-histone proteins. HDACs function within multiprotein complexes; isolated HDACs typically show low substrate specificity [237]. The 18 HDACs described in human are categorized in four families. Class I (HDAC1, 2, 3 and 8), class III (HDAC4, 5, 6, 7, 9 and 10), and class IV (HDAC11) are Zn^{2+}-dependent. Class II includes sirtuins SIRT1-7, HDACs homologous to yeast Sir2 which require NAD+ as a cofactor. Histone

deacetylases can be blocked by HDAC inhibitors. Despite their wide-reaching and unpredictable effect on gene expression, HDAC inhibitors have been successfully used as therapeutic agents (*e. g.* in psychiatry [240] and cancer treatment [241]). As histone deacetylases within each class show structural similarities, the commonly used HDAC inhibitors typically affect catalytic activity of multiple HDACs [242]. Thus, studies attempting to test whether histone deacetylation is leading, rather than following, transcriptional silencing are inherently impacted by HDACs' multiple and overlapping functions, as well as limited selectivity of HDAC inhibitor treatment. Moreover, while HDACs are strong candidates for proteins that are tethered by XIST to the Xi in early steps of X inactivation, the interaction with XIST may not be direct.

As described in the preceding section, the repeat A region of XIST can autonomously induce gene repression. This observation posits that either repeat A alone induces gene silencing, or that repeat A binds proteins that carry out the gene repression. To date, apart from the splicing factor ASF/SF2, which is involved in Xist RNA processing [74], PRC2 is the only known complex shown to interact with repeat A via its components SUZ12 [73, 243] or EZH2 [24, 73]. H3K27me3, a mark deployed by PRC2 is enriched on the Xi [48, 115], and dissociation of Xist from the Xi results in delocalization of PRC2 [90]. Given the strong data pointing to PRC2 as a likely candidate for repeat A effector, we wished to explore whether H3K27me3 is enriched at the *EGFP* promoter and whether PRC2 is necessary for XIST-induced silencing of the *EGFP* reporter. While the HT1080 cell line did not previously show gross enrichment for H3K27me3 by immunofluorescence, local enrichment was not tested. Indeed, despite the lack of wide-spread enrichment of H4K20me1 upon *XIST* induction observed by immunofluorescence, chromatin IP showed H4K20me1 was in fact recruited to the silenced *EGFP* promoter [44].

Both HDACs and PRC2 are strong candidates for XIST-interacting partners. The former because *XIST* expression is rapidly followed by histone deacetylation, and the latter because PRC2 components interact with repeat A which can autonomously induce gene silencing. Alternative approaches to a hypothesis-driven search for novel factors involved in X-chromosome inactivation employed various screening approaches. N-ethyl-N-nitrosourea (ENU)-induced random mutagenesis yielded a mutation that was embryonic-lethal only in females [244] and was later shown to disrupt SMCHD1, an Xi interacting protein [132]. SATB1 was identified as an indispensable factor for X inactivation in mice when a subpopulation of transgenic male lymphoma cells carrying an X-linked inducible *Xist* became resistant to inactivation of the single X chromosome, which otherwise caused cell death [169]. Several groups screened human autoimmune sera for an enrichment of signal over the Xi using immunofluorescence [97, 245, 246]. Finally, results of two RNAi-mediated gene knock-down screens have been published. A knock-down screen which used a siRNA library designed to knock-down 174 RNA-interacting proteins

showed that HNRNPU is required for Xist localization in differentiated mouse cells [165]. In a separate genome-wide study, a shRNA-mediated knockdown of 32 genes caused partial re-activation of an Xi-linked *EGFP* reporter in mouse embryonic fibroblasts [247].

In this section, we show that despite its known interaction with repeat A, PRC2 is not responsible for the transcriptional silencing observed in the HT1080 transgenic system. We further demonstrate that while histone deacetylation accompanies XIST-induced silencing of *EGFP*, inhibiting histone deacetylases does not abolish the silencing potential of XIST. Finally, we extend our search for factors critical for XIST action by performing a siRNA-mediated knock-down screen of proteins that were previously implicated to play part in X inactivation to test whether they are critical for XIST-induced silencing in the HT1080 system.

5.2 Results

5.2.1 XIST-induced histone deacetylation is not critical for gene silencing

Histone deacetylation closely follows *XIST* expression both at the *EGFP* promoter in the HT1080 transgene [44] and in the course of normal X inactivation [102]. We utilized a set of histone deacetylase inhibitors to determine whether histone deacetylation is the cause or the consequence of *EGFP* silencing. First, we employed qRT-PCR to test the effect of increasing concentrations of sodium valproate (VPA) (1.5 mM – 6 mM) on XIST's ability to silence *EGFP* expression (Figure 5.1). We observed that VPA causes dose-dependent up-regulation of *EGFP* prior to induction of *XIST* with DOX (Figure 5.1A). While the expression of *EGFP* in the VPA-treated cells was higher than in the untreated control following 1 and 2 days of *XIST* induction, relative to the *EGFP* expression in the absence of *XIST* induction, silencing by XIST was stronger in the VPA-treated cells. Native chromatin IP followed by qPCR to assay histone acetylation levels at *EGFP* promoter confirmed that treatment with VPA results in an increase of histone H4 acetylation prior to *XIST* induction (Figure 5.1B). However, even when the cells were treated with the highest tested VPA concentration (6 mM), induction of *XIST* resulted in proportional loss of histone H4 acetylation at *EGFP* promoter.

Next, we tested whether trichostatin (TSA), a more potent HDAC inhibitor [242], would disrupt XIST's ability to silence *EGFP*. As with VPA, increasing amounts of TSA (100 nM – 400 nM) resulted in an increase of *EGFP* expression prior to *XIST* induction (Figure 5.1C). Unlike in the VPA-treated cells, greater than 300 nM concentration of TSA resulted in continuous *EGFP* upregulation despite *XIST* expression; of note, while the 300 nM concentration of TSA did not have dramatic effects on the cells' phenotype, the highest TSA concentration resulted in poor cell growth.

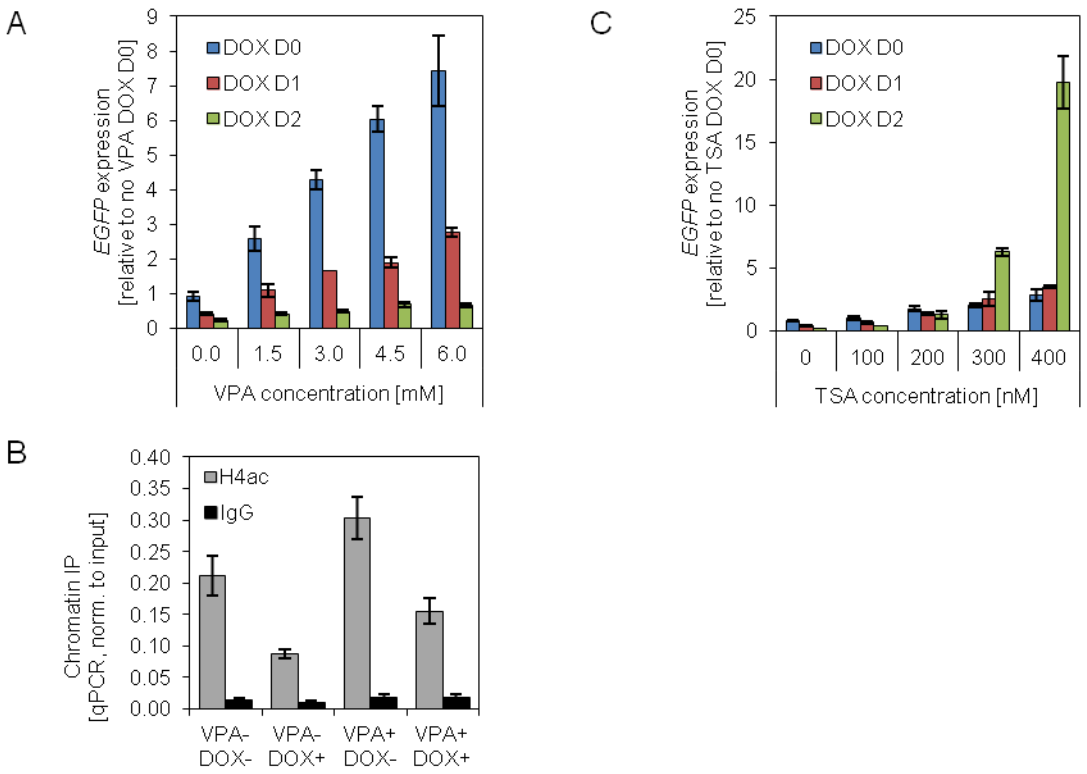


Figure 5.1: Effects of VPA and TSA on *EGFP* silencing by full-length XIST.

(A) Flow cytometry analysis of *EGFP* expression in cells treated with increasing concentration of VPA. VPA was added to the media 24 hours prior induction of full-length *XIST* with DOX and the VPA concentration was maintained for the duration of the experiment. Error bars indicate ±1 s.d. of the qPCR triplicate.

(B) Native chromatin IP using panH4acetyl antibody (Millipore 06-598, 2.5 µg per pull-down) and IgG (Sigma I8765, 10 µg) as an unspecific control. Cells were treated with 6 mM VPA for 6 days (samples labeled VPA+) and/or DOX for 5 days (samples labeled DOX+). Error bars indicate ±1 s.d. of the qPCR triplicate.

(C) As in (A), but the impact of TSA treatment on *EGFP* expression is depicted.

While XIST was unable to silence *EGFP* after efficient inhibition of HDACs, this did not in principle exclude the possibility that XIST is, at least in part, able to counteract the *EGFP* up-regulation caused by histone hyperacetylation. To test this possibility, we treated the cells with VPA, TSA and two other HDAC inhibitors – apicidin and MS-275, which was used in two different concentrations (Figure 5.2A-E). In general, the highest concentrations of HDAC inhibitors that did not grossly affect the cells' physiology were used. Based on a previously published report [242], the used concentrations of HDAC inhibitors that should selectively inhibit different HDACs (Figure 5.2F). However the results of an *in vitro* assay may not entirely reflect which HDACs are silenced in the cell culture. Following the treatment with HDAC inhibitors for 24 hours, *XIST* was induced with DOX and flow cytometry was

used to compare *EGFP* expression in DOX-treated versus untreated cells after a further 1 or 2 days. *EGFP* expression increased in all samples treated with HDAC inhibitors (Figure 5.2A-E, blue lines). When HDAC inhibitor treatment was combined with *XIST* induction, *EGFP* expression was markedly lower in all instances (red lines). Finally, we probed whether the XIST will be able to attenuate *EGFP* expression in HDAC inhibitor-treated cells that were previously treated with DOX for 48 hours to induce partial *EGFP* silencing (Figure 5.2G). In accord with the previous observations, while *EGFP* expression markedly increased in the presence of the HDAC inhibitor (red line), the continuing *XIST* expression substantially reduced the *EGFP* up-regulation (purple line). We therefore conclude that silencing of *EGFP* by XIST is not dependent on histone deacetylation and therefore, that histone deacetylation is a consequence of transcriptional silencing.

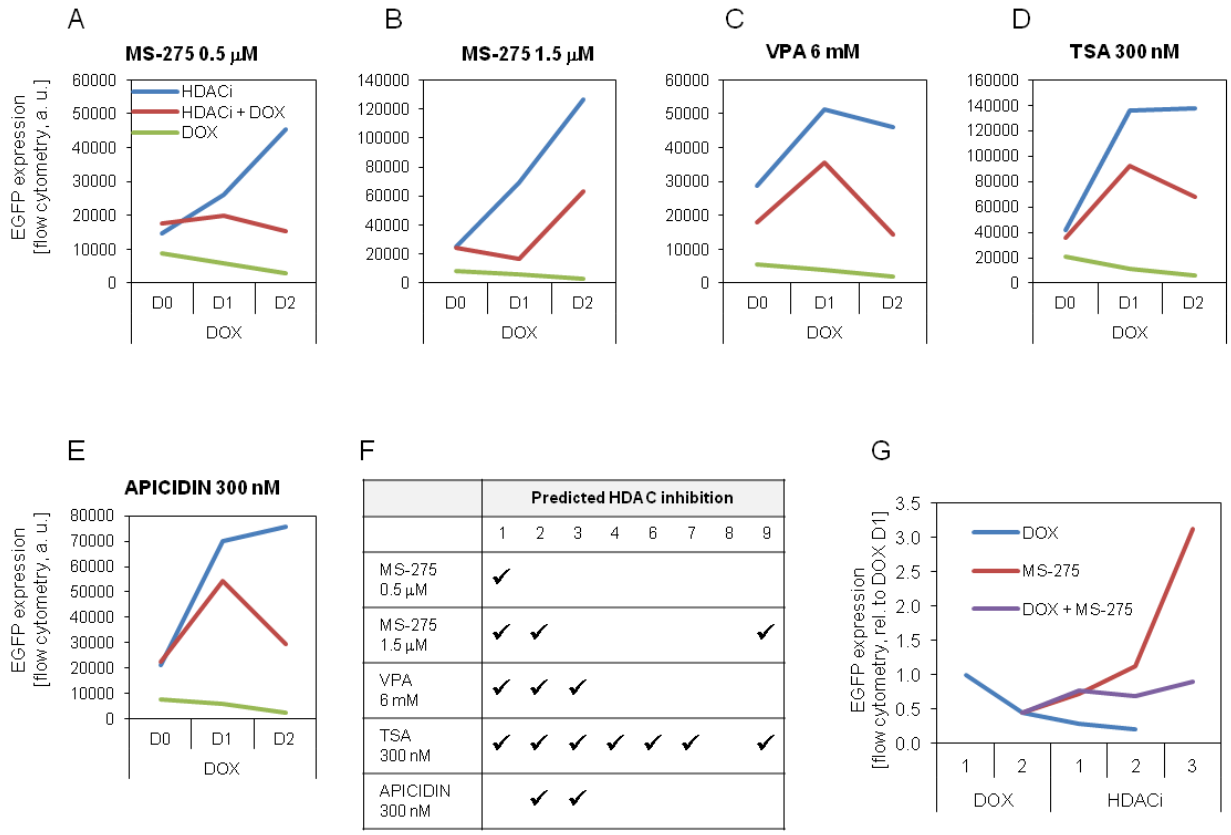


Figure 5.2: XIST partially counteracts the effect of HDAC inhibitors.

(A)–(E) *EGFP* expression measured by flow cytometry. HDAC inhibitors were added to the media 24 hours prior induction of *XIST* with DOX.

(F) The HDAC inhibitors at concentrations used in this experiment inhibited a specific subset of HDACs in an *in vitro* assay [242].

(G) *EGFP* expression measured by flow cytometry. After inducing *XIST* expression with DOX for 48 hours, the cell culture was re-plated and treated with DOX, MS-275 (0.5 µM), or both.

5.2.2 PRC2 is not necessary for XIST action

One of the hallmarks of the Xi is enrichment for H3K27me3 [48, 115]; indeed, EZH2 and SUZ12, components of PRC2, have previously been shown to directly interact with repeat A [24, 73]. As the repeat A core is necessary for XIST-induced silencing in our transgene (Figure 3.3B), we wished to explore if PRC2 is the effector responsible for *EGFP* silencing. Native chromatin IP following five days of induction of the full-length *XIST* transgene showed accumulation of histone H3 but not H3K27me3 at a site immediately 3' of the *EGFP* promoter (Figure 5.3A). To test whether H3K27me3 is accumulated within the promoter of an X-linked gene normally subject to X inactivation, we generated a separate construct harboring the inducible human repeat A together with the DsRED-Express2 reporter driven by the promoter of mouse *Pgk1* (Figure 5.3B). Similar to the results obtained with the *EGFP* reporter, the DsRED-Express2 reporter is effectively silenced upon 5 days of repeat A expression (Figure 3.3B, C) without any increase in H3K27me3 at three loci spread across the *Pgk1* promoter (Fig. 5B).

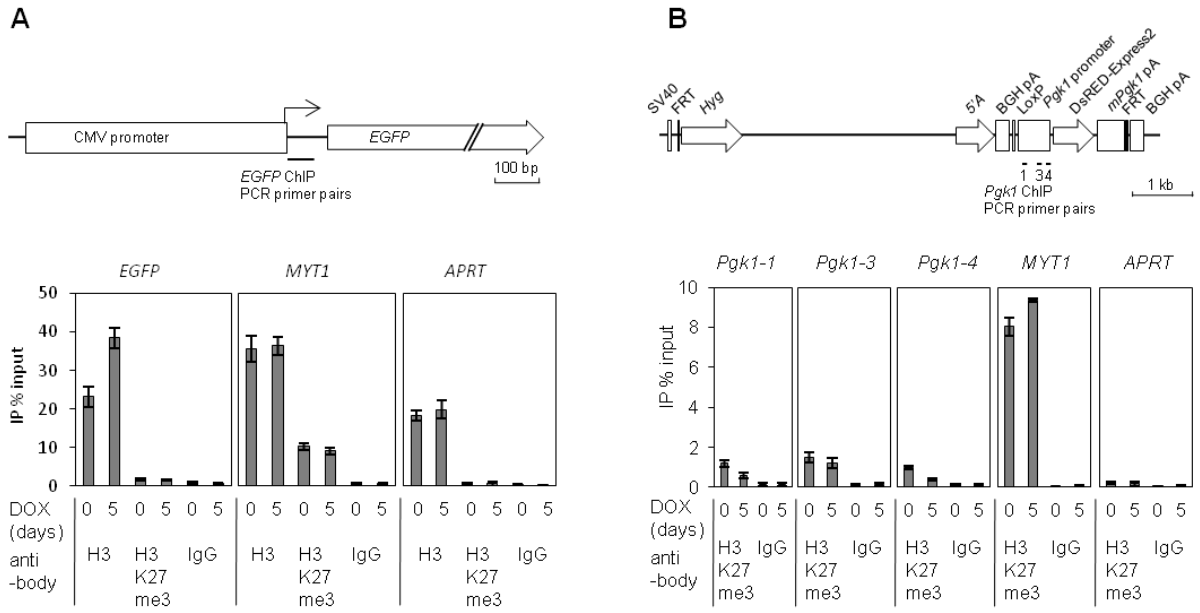


Figure 5.3: H3K27me3 is not recruited to reporter promoters upon XIST-induced silencing.

(A) Chromatin IP followed by qPCR was used to assess H3K27me3 levels at the *EGFP* promoter in cells where *XIST* was induced for 5 days versus in uninduced cells. Antibodies against histone H3 and IgG were used as positive and negative chromatin IP controls, respectively, as listed below panels. Primers targeting *MYT1* and *APRT* promoters were used as positive and negative controls, respectively, for H3K27me3 occupancy [248]. Position of *EGFP* qPCR primers used in the chromatin IP experiment is indicated. Error bars indicate ±1 s.d. of the qPCR technical triplicate.

(B) Map of an inducible 5'A transgene with the mouse *Pgk1* promoter driving DsRED-Express2 reporter showing positions of qPCR primer pairs 1, 3 and 4. Chromatin IP was performed as in (A). Error bars indicate ±1 s.d. of the qPCR technical triplicate.

To further examine whether PRC2 plays a role, perhaps distinct from H3K27me3 recruitment, in XIST-induced silencing in the HT1080 system, we performed siRNA knock-down of SUZ12 and EZH2 for 36 hours combined with *XIST* induction for the last 24 hours (Figure 5.4A). Despite an effective SUZ12 (-91%) or EZH2 (-77%) mRNA down-regulation, measured by qRT-PCR (Figure 5.4B), the ability of XIST to silence *EGFP* was unaffected. We observed relative upregulation of *XIST* in the PRC2 knock-down cells (Figure 5.4B); however, in our experience, the approximately 2 fold difference in transgenic *XIST* expression does not affect the extent of reporter silencing. To exclude the possibility that the short timeframe of the knock-down experiment did not allow for sufficient depletion of the PRC2 complex, we performed siRNA double knock-down of SUZ12 and EZH2 for 6 days, followed by qRT-PCR to assess the impact on DsRED-Express2 silencing (Figure 5.4C). In accord with our previous observations, depletion of the PRC2 complex did not abolish repeat A's ability to induce silencing of the reporter (Figure 5.4D). Taken together, we conclude that repeat A is able to cause silencing of multiple reporters in the absence of PRC2 and without recruitment of H3K27me3.

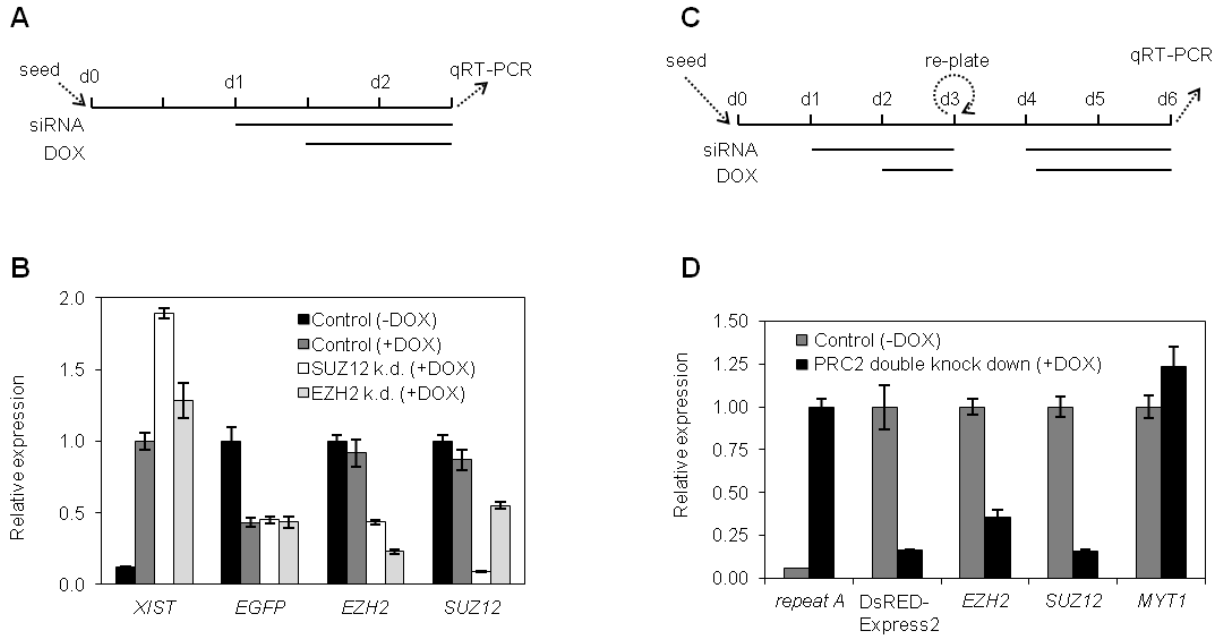


Figure 5.4: PRC2 is dispensable for repeat A-induced reporter gene silencing.

(A) The timeline of *SUZ12* and *EZH2* knock-down experiment.

(B) Knock-down of SUZ12 (-91%) and EZH2 (-77%) has no effect on *EGFP* repression by full-length XIST. qRT-PCR results were normalized to *ACTB* and set to 1 in uninduced cells (*EGFP*, *EZH2* and *SUZ12*), or cells induced with DOX, but untreated with siRNA (*XIST*). Error bars indicate ±1 s.d. of the qPCR triplicate.

(C) The timeline of double knock-down of *SUZ12* and *EZH2* followed by qRT-PCR, employing the alternative construct that harbors repeat A and DsRED-Express2 reporter.

(D) Double knock-down of *SUZ12* (-84%) and *EZH2* (-64%) does not abolish DsRED-Express2 repression by repeat A construct. Repeat A was induced by DOX in the double knock-down cells, but not in the control cells treated only with the transfection agent. qRT-PCR results were normalized to *ACTB* and set to 1 in control cells (DsRED Express2, *EZH2* and *SUZ12* and *MYT1*), or the DOX-induced, double knock-down cells (repeat A). *MYT1*, a gene normally repressed by H3K27me3 in HT1080 cell line shows slight upregulation upon following the double knock-down.

5.2.3 Identification of proteins involved in XIST-induced silencing

We have shown that neither histone deacetylation, nor PRC2 recruitment are responsible for XIST-induced silencing in the HT1080 transgenic system. To extend the search for proteins affecting XIST silencing ability, we have designed a siRNA library targeting 31 proteins that were previously implicated to play a role in X-chromosome inactivation (reviewed in section 1.4) and *SDC1*, a gene encoding

syndecan-1, a proteoglycan facilitating interaction of cells with the interstitial matrix and not known to be involved in X inactivation.

To test the effect of protein knock-downs on the ability of XIST to induce gene silencing, the 2-3-0.5+3#4 HT1080 transgenic cell line was transfected with the siRNAs and after 24 hours, the full-length *XIST* transgene was induced with DOX. Further 48 hours later, the samples were collected for flow cytometry analysis of *EGFP* expression and qRT-PCR analysis to assay the efficiency of each gene knock-down, as well as expression of *EGFP* and *XIST* (Figure 5.5A and Figures A.3 -A.6).

By comparing *EGFP* expression levels in the siRNA-treated samples and in the transfection reagent-treated controls, we observed that the siRNA treatment affected the *EGFP* expression levels even when *XIST* was not induced (Figure A.3). This is perhaps not surprising as many of the proteins targeted in the knock-down screen are involved in chromatin regulation. Predictably, two days of *XIST* induction led to partial *EGFP* silencing in the control cell line and in the siRNA-treated cells. To assess whether the knock-down affected XIST's ability to silence we introduced "relative loss of silencing ability", a measure that quantifies how many fold less silencing occurred in the siRNA-treated cells compared to the transfection reagent-treated control cells (Figure 5.5B).

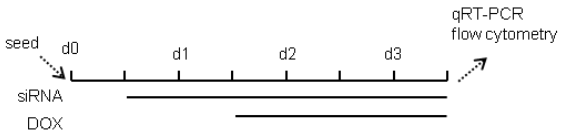
Depletion of seven proteins, ASH2L, ATM, DICER1, SPOP, SATB2, YY1 and HNRNPU caused a substantial reduction of XIST's silencing ability in two independent experiments, as is reflected by high relative loss of silencing ability ratios when EGFP protein levels were assayed by flow cytometry (Figure 5.5C). CARM1 depletion also resulted in high relative loss of silencing ability ratio in one replicate, however cell viability was dramatically reduced in the other replicate and we therefore could not confirm the effect. The high relative loss of silencing ability was strongly correlated ($r = -0.79$, Pearson correlation coefficient, first replicate of the flow cytometry assay) with reduced expression of *EGFP* in the absence of *XIST* induction (Figure A.3).

Overall, the results showed high correlation between the two replicates and between flow cytometry and qRT-PCR for both replicates (Figure 5.5C). In the first replicate, knock-down of SDC1, a protein that is unlikely involved in X inactivation, showed a relative loss of silencing ability of 1.08, which closely corresponded to the theoretical value, *i. e.* 1.00 if SDC1 had no influence on XIST's ability to silence. In the second replicate, the SDC1 knock-down showed a relative loss of silencing ability of 1.22, which was unexpectedly high. To further validate the results of the second replicate, we have performed a knock-down of SUZ12 and EZH2, components of the PRC2 complex that, as we have previously shown (section 5.2.2), is not involved in *EGFP* silencing in the HT1080 system. Indeed, SUZ12 and EZH2 showed relative loss of silencing ability 0.95 and 0.87, respectively (data not shown).

We further repeated the screen using the repeat A – DsRED Express2-containing HT1080 2-3-0.5a #8 cell line (Figure 5.5C). Based on the flow cytometry and qRT-PCR analyses, the protein knock-downs that disrupted DsRED expression only partially overlapped with the candidates identified in the full-length *XIST* cDNA – *EGFP* cell line. While the differences in the relative loss of silencing ability that we observed following protein knock-downs in the full-length XIST cDNA versus the repeat A cell lines may provide insights into the mechanism by which these proteins interact with XIST, we noted that the DsRED expression levels showed overall weaker response to the treatment with the siRNAs, and we also observed that cell growth and morphology were generally affected less in the 2-3-0.5a #8 cell line. Therefore, we used a different repeat A-harboring HT1080 cell line in the subsequent knock-down experiments.

As part of her thesis work, Angela Kelsey used fluorescent *in situ* hybridization to test how the siRNA-mediated down-regulation of the proteins included in our panel affects XIST localization. XIST showed tightly localized signal in the transfection reagent-treated control cells and following the knock-down of the majority of the proteins. Interestingly, in 6 of the 32 siRNA-treated samples, XIST showed delocalized, or ‘speckled’ signal. Proteins identified by the siRNA screen to be important for XIST localization largely overlapped with those that were also shown to contribute to XIST-mediated *EGFP* silencing (Figure 5.5C).

A



B

$$\text{Relative loss of silencing ability} = \frac{\dfrac{\text{Reporter expression in siRNA treated cells (DOX+)}}{\text{(DOX-)}}}{\dfrac{\text{Reporter expression in transfection reagent treated cells (DOX+)}}{\text{(DOX-)}}}$$

C

	Relative loss of silencing ability							
	Flow cytometry			qRT-PCR				
	Full-length XIST #1	Full-length XIST #2	Repeat A	Full-length XIST #1	Full-length XIST #2	Repeat A	**Average** (full-length XIST flow cytometry)	XIST delocalization
DICER1	1.37	1.55	XK	1.39	2.65	XK	**1.46**	
ATM	1.37	1.36	1.09	2.36	1.72	1.52	**1.37**	
SATB2	1.37	1.33	X	1.17	1.37	X	**1.35**	
CARM1	D	1.35	1.05	D	1.75	0.81	**1.35**	
ASH2L	1.30	1.37	1.06	2.74	1.29	1.43	**1.33**	✓
YY1	1.39	1.26	1.02	1.46	1.70	1.19	**1.33**	✓
HNRPU	1.28	1.29	1.11	1.46	1.38	1.08	**1.29**	✓
SPOP	1.24	1.25	1.15	1.53	2.05	1.18	**1.25**	✓
PARP1	1.08	1.34	0.98	1.01	1.49	1.59	**1.21**	
SDC1	1.08	1.22	1.02	1.11	1.52	1.14	**1.15**	
SET7	1.11	1.15	X	1.15	0.95	X	**1.13**	
HNRPK	1.14	1.12	X	0.91	1.54	X	**1.13**	
SMCHD1	1.12	K	0.88	1.67	K	1.07	**1.12**	
RCOR1	1.06	1.14	0.76	0.89	1.94	0.74	**1.10**	
CTCF	X	1.07	1.22	X	1.11	0.88	**1.07**	
BRCA1	1.05	K	1.04	0.79	K	1.00	**1.05**	
CBX7	1.03	1.03	X	1.25	1.49	X	**1.03**	✓
H2AFY	0.89	1.07	0.98	0.58	0.99	1.05	**0.98**	
BRRN1	0.90	1.01	1.08	0.86	1.07	1.35	**0.96**	
SMC4L1	0.89	0.97	1.02	0.68	0.83	1.02	**0.93**	
CNAP1	0.94	0.87	1.00	0.61	0.56	1.04	**0.91**	
CUL3	0.94	0.88	0.72	0.72	1.43	0.59	**0.91**	✓
SATB1	D	0.87	1.06	D	0.58	0.87	**0.87**	
RNF2	0.85	0.86	1.03	0.43	0.42	0.91	**0.86**	
REST	X	0.80	X	X	0.91	X	**0.80**	
EHMT2	0.80	0.78	X	0.95	1.11	X	**0.79**	
ATRX	0.78	0.76	1.17	0.25	0.49	0.78	**0.77**	
AOF2	D	X	0.99	D	X	1.34		
CBX4	X	X	X	X	X	X		
EHMT1	K	K	X	K	K	X		
H2AFY2	N/E	N/E	N/E	N/E	N/E	N/E		
M96	K	K	X	K	K	X		

Pearson correlation coefficient

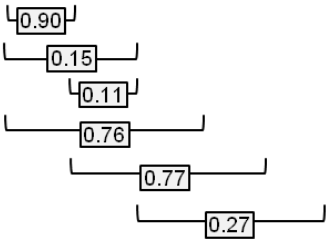


Figure 5.5: siRNA knock-down screen identifies proteins involved in XIST-induced silencing.

(A) The timeline of single knock-down experiments.

(B) A formula used to calculate the relative loss of silencing ability. A ratio of *EGFP* or DsRED Express2 expression in cells expressing *XIST* (DOX +) versus in cells not expressing *XIST* (DOX -) was calculated for all siRNA-treated cells and divided by an identically calculated ratio for the transfection regent-treated controls cells.

(C) Flow cytometry and qRT-PCR was used to survey the relative loss of silencing ability. The full-length *XIST* cDNA construct was screened in duplicate, along with the single screen utilizing the Repeat A-DsRED Express2 transgene. Samples are sorted in the order of a descending average relative loss of silencing ability as measured by flow cytometry in the full-length *XIST* cDNA screen. The color coding corresponds to samples with highest (green) to lowest (red) relative loss of silencing ability in each column. Samples in which excessive cell death occurred (D), in which *XIST* failed to up-regulate at least 5-fold following DOX induction (X) or in which siRNA-mediated knock-down failed to reduce the expression by at least 40% (K) were excluded from the analysis. H2AFY2 is not expressed in the HT1080 cells we tested (N/E). The tick marks denote proteins that are indispensable for XIST localization, as assayed by fluorescence *in situ* experiments that were performed by Angela Kelsey.

To further validate the candidates identified in the siRNA screen, we performed an extended analysis of the impact of ATM, DICER1, SPOP and YY1 protein knock-down on the reporter gene silencing. Instead of the single knock-down approach used in the original screen, two consecutive rounds of siRNA transfection and 3 days of DOX treatment were used to allow for stronger phenotype manifestation (Figure 5.6A).

In comparison with the single knock-down treatment, the relative loss of silencing ability, measured by flow cytometry, was more prominent after the double knock-down in all four candidate proteins (Figure 5.6B). We have noted that *EGFP* expression was attenuated in the siRNA-treated samples in comparison with the transfection reagent-treated control cells, (*i. e.* samples that show relative fluorescent reporter expression < 1.0 in Figure 5.6B). This ‘pre-silencing’ phenomenon, which was also observed in the original screen (Figure A.3 and Figure A.4), was particularly apparent prior to DOX induction, but persisted also in the DOX-treated cells (Figure 5.6B).

Next, we used the same experimental setup to assay the effects of the double knock-down on a cell line harboring the repeat A - DsRED Express2 transgene. As the DsRED Express2 expression in the cell line originally used was less modulated by the siRNA screen then the *EGFP* expression in the full-length *XIST* cDNA cell line, we tested whether a different repeat A - DsRED Express2 cell line would respond more strongly. Indeed, the HT1080 F55 #1 repeat A - DsRED cell line in which the transgene is integrated on the X chromosome showed strong relative loss of silencing ability (shown above the ascending arrows). Moreover, siRNA treatment in the HT1080 F55 #1 repeat A - DsRED cell line had less impact on DsRED expression. While the ‘pre-silencing’ effect persisted to some degree in the cells where *XIST* was not induced, as documented by values < 1.0 in Figure 5.6B, 3 days after *XIST* expression DsRED expression was higher in the siRNA-treated cells, as documented by values > 1.0 in Figure 5.6B.

To ascertain that the effects observed are specific to the XIST-induced silencing, we performed a single knock-down of the four tested candidate proteins in a cell line in which a plasmid harboring DOX-inducible CMV promoter, but no *XIST* sequence is integrated upstream of the *EGFP* ('vector' in Figure 3.2A). The siRNA treatment resulted in attenuated *EGFP* expression, in accord with the results obtained in the full-length *XIST* cDNA transgene. While the relative loss of silencing ability observed following *XIST* induction was also in part present in the *XIST* sequence-lacking control cell line (Figure 5.6B), the effect was consistently stronger in the full-length *XIST*-cDNA cell line (Figure 5.5C).

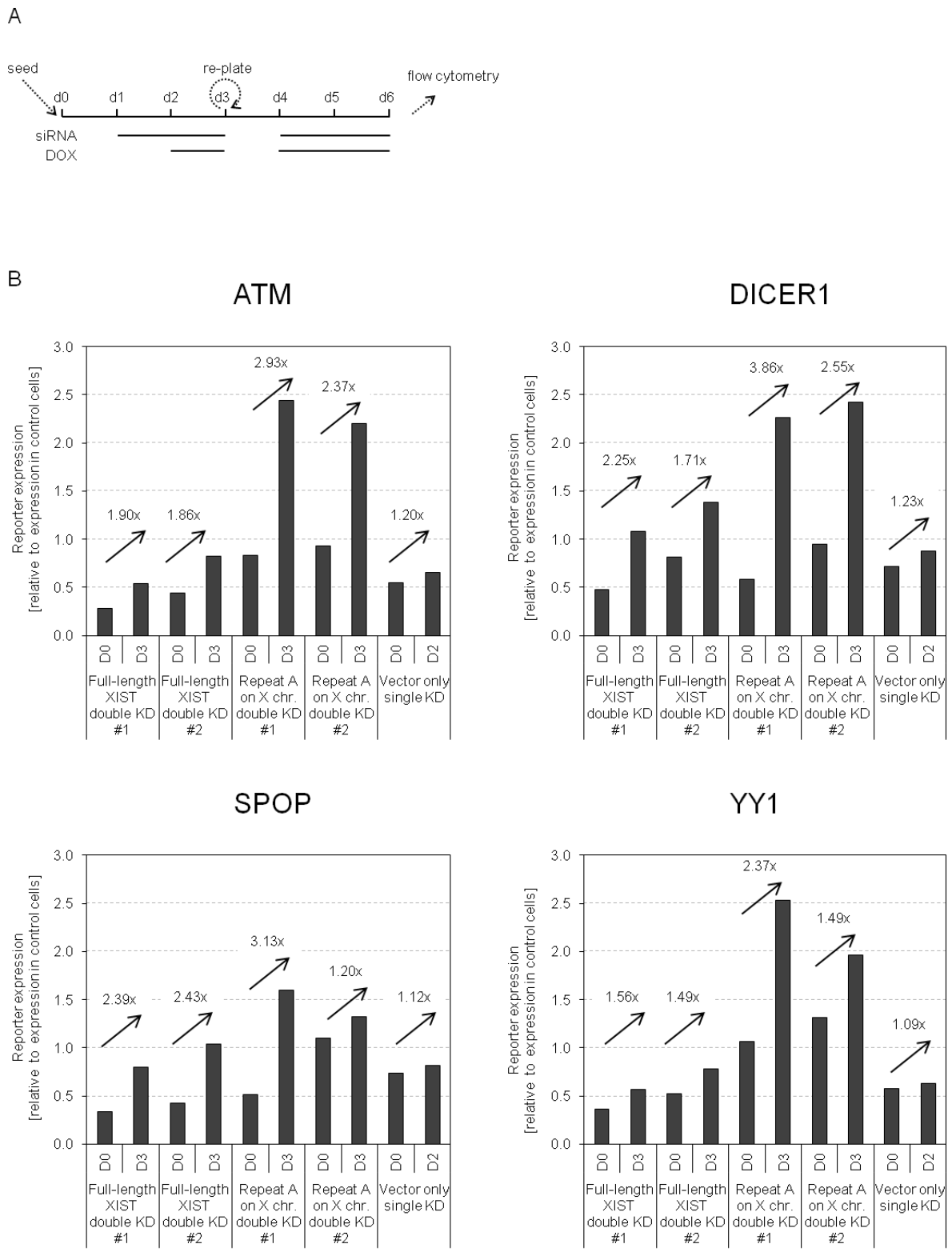


Figure 5.6: Multiple cell lines and experimental setups validate the candidate proteins.

(A) The timeline of double knock-down experiments.

(B) siRNA knock-down of ATM, DICER1, SPOP and YY1 lead to the relative loss of XIST's silencing ability. Duplicate double knock-down experiments are shown for both the full-length *XIST* cDNA - *EGFP* cell line and the F55 #1 repeat A - DsRED Express2 cell line. A control single knock-down experiment utilizing a cell line in which an empty vector, instead of *XIST* is integrated upstream of the *EGFP* is also shown. The individual bars represent the reporter expression in the siRNA-treated cells as measured by flow cytometry. The data are normalized to the reporter expression in transfection reagent-treated control cells. The relative loss of silencing ability is depicted above the arrows.

5.3 Discussion

We have used the transgenic HT1080 system to explore which proteins affect the ability of XIST to induce gene silencing and present evidence that neither histone deacetylation, nor H3K27 trimethylation are the cause of XIST-induced gene silencing. Histone deacetylation closely follows *Xist* expression in mouse [102] and the *EGFP* reporter is depleted of histone acetylation upon *XIST*-induced silencing in the transgenic HT1080 system [44]. These observations can be explained by two different chains of events. Either XIST/Xist recruits histone deacetylases or histone deacetylation is a secondary effect of gene silencing achieved by other means. To address whether histone deacetylation is the cause or the consequence of XIST-induced silencing, we have employed an array of histone deacetylase inhibitors and tested the ability of the inducible *XIST* transgene to induce silencing of the *EGFP* reporter.

In an ideal experiment, treatment with HDAC inhibitors would ensure that *EGFP* acetylation levels remain constant following *XIST* expression, which would rule out histone deacetylation as a factor in XIST-induced silencing. If *EGFP* was still subject to silencing by XIST under such conditions, the results would suggest that histone deacetylation indeed is a consequence of XIST-induced gene silencing. However our results demonstrate that HDAC inhibitor treatment results in an increase of *EGFP* expression (Figure 5.1). To control for this effect, we compared expression of *EGFP* in HDAC inhibitor-treated cells in the presence and absence of *XIST* expression. In total, the cells were treated with four different HDAC inhibitors, one of which, MS-275, was tested in two different concentrations (Figure 5.2). In all cases, XIST was able to induce partial *EGFP* silencing, despite the presence of HDAC inhibitors. The HDAC inhibitors we employed were previously shown to selectively inhibit some of the histone deacetylases under the used conditions in an *in vitro* assay [242], as summarized in Figure 5.2F. None of the HDAC inhibitors suppresses Sir2 deacetylase homologs and HDAC8. In conclusion, our data suggest that XIST does not require histone deacetylation mediated by HDAC1-7 or HDAC9 to induce gene silencing.

Prominent among the potential partners critical for gene silencing is the well-established chromatin silencing complex PRC2. We however present evidence that repeat A-induced silencing occurs

independently of PRC2. The *XIST* transgene we employed silenced two distinct reporters without recruiting H3K27me3 (Figure 5.3) and despite siRNA-mediated knock-down of PRC2 components (Figure 5.4), although it remains to be confirmed by western blot analysis that the SUZ12 and EZH2 proteins were also depleted and that levels of H3K27me3 were decreased. Silencing of *EGFP* is accompanied by a substantial increase of histone H3 occupancy, suggesting that the transgenic XIST induces chromatin compaction (Figure 5.3A). DsRED-Express2 under the control of the mouse *Pgk1* promoter is also silenced without recruiting H3K27me3 (Figure 5.3B).

In conclusion, while there is strong evidence for a role of PRC2 in X-chromosome inactivation, our data argue that it is not necessary to induce proximal gene silencing and therefore other XIST/Xist-interacting partner(s) are likely involved in silencing. That silencing can occur without PRC2 is supported by observations that female embryos and ES cells lacking functional Eed (embryonic ectoderm development, a core component of PRC2) are capable of initiation and maintenance of random X inactivation [49, 115, 249]. Although EED is essential for maintenance of imprinted X inactivation in extraembryonic tissues, Xist is able to coat the Xi when both the PRC2 components and H3K27me3 were absent prior to, and in course of, random X inactivation [249, 250]. In addition, H3K27me3 can be recruited to in *cis* by constructs lacking repeat A that are silencing-defective [48]. Furthermore, a knock-down screen for genes involved in maintenance of X inactivation in mouse embryonic fibroblasts failed to identify PRC2 components amongst the candidates and an EZH2 knock-down confirmed that PRC2 was dispensable for silencing of the X-linked *EGFP* reporter [247].

As neither histone deacetylation nor H3K27 trimethylation by PRC2 are necessary for XIST's ability to induce silencing, we have broadened the search for XIST-interacting partners. We have compiled a list of 31 proteins that were previously shown to affect X inactivation and employed a siRNA-mediated knock-down screen to identify proteins that affect XIST's ability to silence fluorescent reporters (Figure 5.5). The knock-down of seven proteins consistently attenuated the extent of XIST-induced silencing. Of these seven proteins, four (ASH2L, SPOP, YY1 and HNRNPU) were also indispensable for XIST localization. The remaining three proteins (ATM, DICER1 and SATB2) contributed to XIST-induced silencing, but their knock-down did not cause XIST delocalization. Conversely, knock-down of CBX7 and CUL3 caused XIST delocalization, but did not substantially disrupt gene silencing.

The seven proteins identified by the knock-down screen showed high relative loss of silencing ability ratios largely because the *EGFP* expression was down-regulated prior to XIST induction. In other words, rather than preventing XIST from silencing *EGFP* to the level observed in the transfection regent-treated control cells, the knock-down of these proteins reduced the expression of *EGFP* and the extent of further

silencing induced by XIST was relatively lower. The established function of some of the seven identified proteins provides plausible explanation of the 'pre-silencing' effect. ASH2L is a subunit of protein complexes involved in H3K4 trimethylation [94, 95], a chromatin mark associated with transcriptional activation [96]. SATB2 and HNRNPU play role in maintaining higher order chromatin structure and their knock-down may affect the accessibility of *EGFP* for expression. The strong correlation between high relative loss of silencing ability ratios and the pre-silencing effect, as well as the slight, but consistently observed relative loss of silencing ability in the cells lacking *XIST* sequence (Figure 5.6B) require further probing. The pre-silencing effect was also was also present in the 2-3-0.5+3#4 cell line lacking *XIST* sequence and it was substantially less prominent in the F55 #1 cell line harboring the repeat A - DsRED Express2 transgene (Figure 5.6B). Therefore the pre-silencing may be a specific effect observed only in some integration sites. Alternatively, the differences in the extent of pre-silencing may be caused by the use of two different promoters driving the fluorescent reporter genes (CMV versus *Pgk1*). While the F55 #1 and the 2-3-0.5+3#4 cell lines also harbor different *XIST* constructs (repeat A versus full-length cDNA, respectively), it is unlikely that the pre-silencing is an XIST-specific effect as down-regulation was also observed with vector only in the 2-3-0.5+3#4 cell line (Figure 5.6B). DICER1 was previously implicated to regulate establishment of *Xist* expression [185], however the interaction is likely indirect via down-regulation of DNMT3A [187]. Our novel finding that down-regulation of DICER1 impacts XIST-induced silencing opens the possibility that DICER1 may also have either a direct or, more likely, an indirect role in XIST-induced silencing. Given the broad-ranging effects of DICER1 knock-down on mis-regulation of gene expression, further studies are needed to uncover the functional relationship between DICER1 and XIST-induced silencing.

Down-regulation of ATM was previously shown to cause partial re-activation of the Xi, but did not affect Xist localization [190]. Similarly, we also demonstrated that ATM knock-down partially disrupts XIST-induced gene silencing, without affecting localization. ATM is known to regulate the DNA damage response (reviewed in [188]), however the previously examined knock-down induced Xi reactivation was not accompanied by increase in DNA damage [190] and ATM knock-down does not impact cell cycle progression [251].

SATB1 or SATB2 were previously shown to be indispensable for Xist-induced silencing in a redundant fashion, but expression of *Satb1* or *Satb2* was not required for Xist localization [169]. In the HT1080 cells, knock-down of SATB1 or SATB2 had no effect on XIST localization. However knock-down of SATB2 alone, but not of SATB1 disrupted gene silencing. Interestingly, SATB2 positively regulates a pluripotency factor NANOG [252]. It is thus possible that the cancer-cell-derived HT1080 cells are able

to undergo ectopic XIST-induced silencing because they retain chromatin structure and/or gene expression profile that confers them partial pluripotent-cell-like qualities.

Knock-down of RNF2, one of the core components PRC1 complex did not disrupt silencing or localization. Combined with our previous results, this observation shows that the canonical polycomb complexes are not recruited by XIST in the HT1080 cells. Interestingly, knock-down of CBX7 caused XIST delocalization, without disrupting gene silencing. Recently, knock-down of CBX7 has been shown to promote ES cell differentiation in mouse. Similarly to knock-down of SATB2, CBX7 down-regulation may cause changes in chromatin structure and gene expression patterns which result in disruption of XIST localization. The extent of delocalization may however be less prominent and does not prevent the transgenic XIST to locally induce gene silencing.

ASH2L was previously shown to localize to the Xi in mouse ES cells, however knock-down of ASH2L did not affect gene silencing or localization of XIST [97]. In contrast, we have observed disruption of both localization and silencing following ASH2L knock-down. While the 'pre-silencing' of *EGFP* in the ASH2L knock-down cells is consistent with its H3K4 methyltransferase role, the delocalization of XIST following ASH2L knock-down supports its role in formation of the Xi-associated chromatin. Thus, we and others have shown that ASH2L is intimately involved in X inactivation in mouse and humans and further work is needed to decipher this seemingly paradoxical involvement of a gene-activating complex in X inactivation.

YY1 was recently reported to mediate the initial loading of Xist RNA to the *Xist* gene and thus facilitate nucleation of the Xi compartment in differentiating mouse ES cells [54]. Our results show that YY1 is also indispensable for XIST localization and XIST-induced silencing in human cells. YY1 in mouse binds Xist via repeat C [54], which is essentially absent in human [34]. It thus remains to be shown which regions of XIST in human and other non-rodents are involved in YY1 binding.

HNRNPU/SAF-A was previously shown to localize to the Xi in both mouse and human cells [97, 164] and is required for X inactivation in differentiating mouse ES cells and Xist localization in mouse neuroblastoma cells [165]. Our results now extend these finding to the HT1080 cell line and allows us to utilize the inducible XIST to address the precise mechanism by which XIST triggers changes in chromatin organization that are critical for X inactivation.

CUL3 and SPOP interact to form an E3 ubiquitin ligase that interacts with macroH2A and PRC1. Knock-down of either CUL3 or SPOP disrupts gene silencing on the Xi and recruitment of macroH2A, but does not affect XIST localization [200]. In contrast, in the HT1080 cells, knock-down of either

CUL3 or SPOP led to XIST delocalization. Surprisingly, while XIST delocalization was also accompanied by disruption of XIST-induced gene silencing when we down-regulated SPOP, CUL3 down-regulation had no effect on gene silencing. It is possible that the residual expression of CUL3 is sufficient for local XIST-induced silencing, but did not allow for establishment of a fully-developed XIST body.

In summary, the siRNA-mediated screen has identified several proteins that contribute to XIST-induced silencing in the HT1080 model of human X inactivation. We have demonstrated that the siRNA-mediated knock-down screen, in combination with fluorescent *in situ* hybridization to assay XIST localization, is a powerful tool to identify XIST-interacting factors. However, the approach requires careful use of control experiments to identify false positive candidates, as a number of XIST-interacting proteins play a more general role, for example in cell proliferation and chromatin composition.

Knock-down of the candidate proteins never resulted in complete abolition of XIST-induced silencing. While this may simply reflect that siRNA-mediated knock-down only achieves partial depletion of the candidate proteins, alternatively, the candidate proteins we identified may contribute to gene silencing only partially and in parallel with other proteins or protein pathways. Importantly, in combination with our data on HDACs and PRC2, the siRNA screen has also demonstrated that a number of proteins and protein complexes that were previously shown to affect X inactivation in other cell types are likely not involved in XIST-induced silencing and/or XIST localization in HT1080 cells. These findings will thus contribute to our understanding of which of the many processes that lead to the establishment of the Xi during normal embryonic development are absolutely critical the primordial function of XIST RNA, the *cis*-linked gene silencing.

6 DISCUSSION

6.1 Summary of the experiments and future directions

The single-copy DOX-inducible transgenic *XIST* provides a tractable system for dissection of the relationship between the sequence and structure of the *XIST* transgene and XIST-induced silencing. In contrast with normal course of X inactivation initiated from the *XIC in vivo*, the gene silencing by the transgenic XIST is reversible and not accompanied by robust recruitment of inactive chromatin marks or DNA methylation [44]. Using the transgenic system, we showed that the repeat A region of *XIST* is sufficient and necessary for gene silencing and demonstrated that XIST is able to silence reporter genes located directly upstream and downstream of *XIST* in multiple genomic integration sites. The bi-directional effect of XIST on neighboring gene expression argues against a simple transcriptional interference mechanism for the XIST transgene-dependant silencing.

As induction of the transgenic *XIST* is not lethal, arguing either that the HT1080 cells are tolerant to partial functional uneuploidies, or that XIST fails to induce wide-spread silencing in the HT1080 cells. The latter possibility is more probable as induction of *XIST* in the F55 cell line, in which XIST is integrated on the single X chromosome also does not cause cell death. However, we demonstrated that an endogenous gene located approximately 100 kb downstream of the transgene integration site is subject to XIST-induced silencing in the HT1080 2-3-0.5+3#4 cell line. This is surprising, as many of the features normally associated with the spread of heterochromatin across the Xi are absent upon XIST-induction in the HT1080 cell line. If such a broader-range effect of the transgenic XIST can be confirmed in other HT1080 integration sites, expression status of these genes will provide a means to assay XIST's ability to form a silent compartment beyond the proximally-located fluorescent reporter genes and to characterize the postulated way stations implicated in XIST spreading (reviewed in [11]).

The search for the endogenous genes silenced by the transgenic XIST will require use of methods that are sensitive enough to detect relatively small changes in expression, as complete allele silencing will only lead to 50% or 33% decrease in transcription of genes located in diploid or triploid regions, respectively. qRT-PCR is robust, and time and cost effective, but allows for only a limited set of genes to be probed. While genes directly neighboring with the integration sites are strong candidates, it may be misleading to assume that the XIST-induced silencing spreads linearly from the transgenic site. Genome-wide approaches will circumvent the need to select the candidate genes. In our hands, expression microarrays did not provide enough resolution to observe partial gene silencing and are limited by the defined set of probes they carry. Therefore, whole genome RNA sequencing would be a method of choice to uncover distant genes silenced by the transgenic XIST. As several lines of evidence suggest that XIST-induced silencing involves changes to chromatin loop structure, whole genome chromatin

conformation capture assays that probe for physical proximity between chromatin regions will provide novel insights into the mechanism of XIST action.

We have leveraged the reproducibility of the silencing achieved by the single-copy DOX-inducible transgenic XIST to examine *in vivo* the relationship between sequence and structure of the repeat A, the region of XIST critical for silencing. In agreement with a previous report on mouse Xist [42], artificial repeat A retained full silencing potential when compared to human repeat A, suggesting that neither sequence variation within the CG-rich core nor the varying length of the U-rich spacers separating individual repeat A monomers is essential for *XIST* function. We have further reduced the complexity of deciphering one of the critical roles of XIST by showing that repeat A monomers act additively to induce silencing and that the mere 92 bp-long repeat A 2-mer can induce silencing. Curiously, the 2-mer and 3-mer showed similar silencing efficiency. This phenomenon was also observed for 4-mer and 5-mer. It is thus plausible that repeat A 2-mer, and not a monomer is the smallest functional element of XIST capable of inducing gene silencing. Our system offers several ways to test this hypothesis. First, as we have shown that even the 2-mer induces substantial silencing over longer time frame, repeat A monomer should be constructed and its effects compared to the 2-mer. Second, a time course experiment can be performed to assay whether the silencing of even-numbered (N)-mers shows the same extent and dynamics of silencing as odd-numbered (N+1)-mers. Third, repeat A 7-mer can be constructed and tested for silencing efficiency along with 6-mer. Other constructs could be created to test the relationship between sequence, structure and function, for example to address whether the order and distance of the two repeat A palindromes is critical for silencing. Similarly, the effects of repeat A mutations, in particular when tested on the repeat A 2-mer-derived constructs, are likely to be more pronounced over time frames longer than the 5 days of induction by DOX that we employed, as we demonstrate by a time course experiment that tracked *EGFP* silencing by the repeat A 2-mer over the first 16 days Figure 4.2C. However, elucidating which proteins interact with XIST, and specifically repeat A, is ultimately biologically more relevant.

In an ongoing search for XIST-interacting proteins, we have demonstrated that gene silencing by the XIST transgene is PRC2-independent and not induced by histone deacetylation. We have further performed a siRNA-mediated knock-down screen to assay the involvement of 31 other proteins that were previously implicated in various aspects of X inactivation. Seven of the 31 proteins that affected silencing are involved in a surprisingly broad range of functions. SATB2 and HNRNPU regulate chromatin organization. YY1 is critical for loading XIST RNA onto the Xi. A subunit of an E3 ubiquitin ligase, SPOP, and a DNA damage response protein, ATM, contribute to gene silencing on the Xi. ASH2L is enriched on the Xi, but its role is unknown and, given its canonical involvement in gene

activation, not intuitive. And importantly, DICER1 is a subunit of small-RNA-processing complexes known to indirectly regulate *Xist* expression at the onset of X inactivation [187], but has not been implied to contribute to XIST/Xist-induced silencing. In parallel, Angela Kelsey (a graduate student in Brown laboratory) assessed the siRNA-treated cells for XIST localization by fluorescent *in situ* hybridization and found that XIST signal was still present, but showed more dispersed localization upon knock-down of ASH2L, SPOP, YY1, HNRNPU, CBX7 or CUL3. Thus, a subset of the identified proteins was only involved in localization, or silencing.

In contrast to a similar construct in mouse ES cells [42], the repeat A-lacking XIST failed to localize in the HT1080 cells. The *XIST* sequence removed in the del 5'A construct however extends approximately 450 bp downstream of the repeat A and therefore not only eliminates the repeat A sequences, but also disrupts some of the YY1 binding sites located downstream of repeat A. As YY1 knock down disrupts XIST/Xist localization in the HT1080 cells (Figure 5.5C) and mouse embryonic fibroblasts [49], a more refined repeat A deletion construct is needed to test whether the loss-of-localization effect can be attributed to the loss of YY1 binding. It is also plausible that the proteins responsible for Xist localization in mouse ES cells may not interact with XIST RNA in human, or the interaction of XIST with chromatin requires intronic sequences not present in the cDNA transgene. Unlike in differentiating mouse ES cells, *XIST* expression in the HT1080 transgenic system does not induce chromosome-wide silencing. Furthermore, the local gene silencing in the HT1080 cells is reversible (*i. e.* XIST-dependent) [44]. In mouse, sequences 3' of repeat A were previously shown to be responsible for *Xist* localization , as well as for recruitment of several chromatin marks characteristic for the Xi (reviewed in [8, 253]).

The successful execution of this directed siRNA screen and the methodology we established to achieve reproducible protein knock-downs, combined with a robust readout of fluorescent reporter gene silencing provides a strong platform for the search of XIST-interacting proteins that are critical for gene silencing. One line of experiments will focus on exploring precisely how the seven proteins we identified contribute to the silencing. To that end, if a likely biological mechanism is lacking (*e. g.* for ASH2L and ATM), we will use siRNAs to knock-down proteins that are known interacting partners of the identified proteins. If previous studies are indicative of the candidates' function in X inactivation, we will expand these results in the HT1080 cell line. For example, YY1 binds Xist RNA via mouse-specific repeat C sequences [54], and it is currently not clear how it interacts with human XIST. Similarly, how exactly HNRNPU and SATB1/2 contribute to organization of chromatin that undergoes silencing by XIST/Xist is unknown. Of highest importance for elucidating the cascade of events that are triggered by XIST are the proteins that directly interact with XIST RNA. Therefore, all novel candidates for proteins involved

in XIST-induced silencing will be tested by electromobility shift assay (EMSA) for their direct interaction with XIST.

An alternative line of experiments will employ genome-wide screening techniques to broaden our search beyond the known or likely candidates for XIST-interacting proteins. These techniques however carry several potential caveats that need to be considered. If time- and cost-effective pooled screens, where multiple proteins are targeted in each sample, are employed, no protein can formally be excluded as a contributor to XIST-induced silencing, because the extent of individual proteins' knock-down cannot be assayed. Further, as our data exemplify, a number of potential candidate proteins are known to modulate gene expression and thus, a simple experimental setup in which cells that show highest *EGFP* expression following XIST induction would be recovered will be inherently biased to yield candidate proteins that are involved in gene silencing (*e. g.* HDACs). Therefore an experimental design that controls for these effects is needed. One such more laborious, but easier to track, approach would simply be to substantially expand the screen while using the setup presented in this thesis. Such an experiment will allow for control of effective gene knock-down and eliminate the effects unrelated to XIST-induced silencing. However false positive and false negative results are still to be expected as knock-down of some proteins that may be involved in X inactivation will be lethal and conversely, some proteins may be identified because they prevent down-regulation of the reporter RNA or protein (*e. g.* if the knock-down increased the reporter's half-life), but are not directly involved in X inactivation.

6.2 Concluding remarks

We used the tetracycline inducible system to advance the understanding of events that follow *XIST* induction and lead to transcriptional silencing of the *cis*-linked genes. Our results for the first time provide evidence that repeat A alone is sufficient to induce gene silencing (Figure 3.3). This result extends the previous reports both in mice [42] and in humans [44] that repeat A is necessary for gene silencing. Moreover, we also show that an artificial repeat A, a 9-mer of a consensus repeat A sequence, induces silencing to the same extent as the human repeat A sequence (Figure 4.2).

Similarly, oligomers of mouse consensus repeat A sequence were previously shown to be able to replace the canonical mouse repeat A sequence [42]. As repeat A sequence is very well conserved between mouse and humans (Figure A.1), it is not surprising that sequences tested in our study and in the mouse model system behave identically with only two exceptions. First, in mouse, the two palindromic

sequences within repeat A core were spaced by a ‘CT’ dinucleotide, in humans by a ‘CG’ dinucleotide (Figure 4.2A). Second, the T-rich spacers separating the repeat A core sequences in the two studies differed in sequence and length; while the human artificial repeat A had a 22 bp-long spacer (Figure 4.2A), two spacer lengths, 9 bp and 21 bp, were tested to be functional in mouse.

These results allow as to formulate two conclusions. First, the sequence deviations both among the individual repeat A units and between human and mouse repeat A sequences, as well as the spacer length, at least in the tested range, do not contribute substantially to the functionality of repeat A in gene repression. Second, as almost identical repeat A sequences are functional in both mouse and human, it is likely that repeat A forms identical structures in both species and that these structures are recognized by homologous proteins. Given the high degree of repeat A sequence conservation among the species sequenced to date, it is likely that an identical protein, or proteins, is responsible for the direct interaction with repeat A in all XIST-carrying mammals.

Our detailed and quantitative analysis of the silencing ability of repeat A fragments ranging from the full repeat A 9-mer to a 2-mer shows that the extent of silencing decreases with the number of repeats, suggesting that individual repeat A units contribute to additively to gene silencing (Figure 4.2B). A similar relationship between the number of repeats, ranging from 4 to 12, and the extent of gene silencing was previously observed in mouse [42]. Further, our results with the series of mutations of repeat A 2-mer shows that the two palindromic sequences within the repeat A core, as well as the tetranucleotides that are spanned by each of the palindromes, are critical for its function (Figure 4.3).

In line with our data, previous results showed that disruption of the first palindrome ablates Xist’s ability to induce gene silencing [42]. Interestingly, disruption of the ‘ATCG’ tetranucleotide sequence spanned by the first palindrome led only to partial loss of repeat A silencing function in mouse [42], but completely abolished the repeat A 2-mer function in human (Figure 4.3). While the transgenic systems, as well as the assays used in the two studies differed, we attribute the different outcomes of the two experiments to the difference in sequence to which the ‘ATCG’ tetranucleotides were mutated: ‘TTTT’ in our study versus ‘TAGC’ in the study of mouse repeat A [42]. To reconcile these differences, a more extensive series of repeat A core mutants is required.

Our data for the first time demonstrates that mutations in the second palindromic sequence within repeat A core disrupt the ability of the repeat A 2-mer to induce gene silencing (Figure 4.3). Further, we demonstrate that the mutation of the ‘ATAC’ tetranucleotide sequence spanned by the second palindrome ablates the silencing ability of the repeat A sequence (Figure 4.3). This result corroborates a

previous report, in which a single T>C mutation within the ‘ATAC’ sequence completely abolished repeat A function.

Because the repeat A core is rich in palindromic sequences, repeat A has long been proposed to form a secondary structure; two competing models have emerged (Figure 4.1): one predicting that repeat A units form a double stem-loop structure [42], the other suggesting that the palindromes interact not within the same unit, but with palindromes and spacer sequences of another repeat A unit [73]. The two models have recently been reconciled and a current model compatible with all previously published reports suggests that the first palindrome within each repeat A core forms a hairpin, while the second palindrome engages in inter-unit binding, thereby providing a double-stranded ‘backbone’ that presents the individual repeat A hairpins, which are in turn presumed to facilitate repeat A-protein interactions via the ‘ATCG’ tetraloop sequence [233].

Our data largely support this model, as mutations to the first palindrome that enforce pairing between units resulted in a dramatic loss of repeat A 2-mer silencing ability (Figure 4.4). However, mutations of the first palindrome that enforced the hairpin formation, but changed the hairpin sequence also attenuated the silencing ability of the repeat A 2-mer. This data contrasts with a previous finding in a mouse transgenic system, which showed that repeat A function was not affected when the sequence, but not the ability to form a hairpin of the first palindrome was modified [42] and further work is needed to clarify whether only the structure, or both the structure and the sequence of the repeat A hairpin stalk are critical for its gene silencing function.

While the transgenic systems have been instrumental elucidating that repeat A is sufficient and necessary for gene silencing by XIST, uncovering sequence and structural requirements for repeat A function and gathering substantial knowledge about the sequences involved in XIST’s ability to ‘coat’ the X chromosome, our knowledge about the proteins that are involved in XIST’s ability to spread along the Xi-elect and induce chromosome-wide silencing is still limited.

Given the importance of the hairpin structure formed by repeat A core sequences for silencing (Figure 4.4), it is likely that the hairpin interacts with one, or multiple proteins. However, it is difficult to envision, given the steric restrictions of protein-RNA interactions, that more than one protein interacts with a repeat A unit at any given time. So far, both EZH2 and SUZ12, components of PRC2, as well as ASF/SF2, a splicing factor, have been implicated to interact with repeat A. However, we (Figure 5.3 and Figure 5.4) and others have shown that PRC2 is not necessary for XIST-induced silencing. Because ASF/SF2 is critical for RNA processing, a study that would test the ability of XIST to induce silencing in

the absence of ASF/SF2 is not feasible and therefore, ASF/SF2 is currently the only known candidate for a repeat A-binding protein necessary for XIST-induced silencing.

We have identified several proteins that modulate XIST's ability to silence a proximally-located gene (Figure 5.5). While the results of the siRNA knock-down screen require further validation, based on the current knowledge about the role of these proteins, the heterogeneous set may be categorized into proteins involved in: regulation of nuclear ultrastructure (HNRNPU, SATB2), *cis*-targeting of XIST (YY1), transcriptional activation (ASH2L, CARM1) and finally, proteins that likely influence XIST's ability to silence indirectly, through regulation of cell cycle (ATM), gene expression (DICER1) and protein pathway regulation (SPOP). We also showed that histone deacetylation (Figure 5.2), and components of polycomb repressive complexes PRC2 (SUZ12 and EZH2, Figure 5.4) and PRC1 (RNF2, Figure 5.5) are not required for XIST-induced silencing of a proximally-located reporter gene. A previous report also demonstrated that XIST-induced silencing in the HT1080 transgenic system is reversible and does not involve DNA methylation [44]. Based on these findings, we propose a speculative model that is compatible with the current knowledge about our system and highlight several unanswered questions.

Upon transcription, the transgenic XIST RNA interacts with YY1 which facilitates its localization in *cis*. The transgenic XIST then induces silencing of genes in a region that spans several hundred kb, but is unable to transcriptionally silence the whole chromosome. The spread of silencing is facilitated by proteins regulating chromatin ultrastructure of which at least HNRNPU is directly recruited by XIST. These proteins translocate the neighboring chromatin loops to the proximity of XIST and the chromatin is transcriptionally silenced by a yet unknown mechanism, which requires the presence of CG-rich core sequences within repeat A that fold to project a series of hairpins to which an unidentified protein binds. This protein alone, or through interaction with other proteins and/or protein complexes induces transcriptional silencing. The initial silencing also requires expression of ASH2L and CARM1, which may serve to maintain a chromatin structure that is amenable to silencing [169]. Thus, by providing binding sites for proteins with diverse roles, XIST serves as a signalling molecule, identifying the chromosome to be silenced, providing a reference point to which chromatin is reeled and finally, locally increasing concentration of proteins that silence chromatin in their vicinity. The role of XIST as a scaffold ensuring that the right mixture of proteins is recruited to induce X inactivation also explains the nuclear localization of XIST, which can be observed both for repeat A and for the sequences downstream of repeat A.

While the transgenic XIST is able to induce chromatin silencing that is accompanied by chromatin compaction and histone deacetylation, it fails to trigger a stable, XIST-independent and chromosome-wide silencing and induce many changes to chromatin that are observed in normal X inactivation [44]. Experiments that utilized differentiating mouse ES cells to recapitulate the events of normal X inactivation demonstrated that XIST can only trigger X inactivation in a specific stage of embryonic development [47]. The inability of the XIST RNA to induce widespread and stable gene silencing in the HT1080 cells may indicate the deficiency of the system in forming the repressive compartment upon *XIST* induction or in translocating genes into the repressive compartment, either because the chromatin structure is no longer amenable to epigenetic modifications that impose the stable silencing, or because the proteins inducing these changes are not expressed in the HT1080 cells. Indeed, while the transgenic XIST was able to deplete hnRNA transcription, forming a so called 'CoT hole', accumulation of the H3K27me3, macroH2A and H4K20me1 inactive chromatin marks, at least at the level of resolution achieved by immunocytochemistry, was not observed in the HT1080 cell line [44].

Also unknown is the mechanism that prevents XIST transcripts from silencing *XIST* itself. The simplest explanation assumes that *XIST* indeed is in part silenced by XIST; however the expression of *XIST* reaches equilibrium through this feed-back loop. This explanation is compatible with our observation that *XIST* expression level fluctuates in the several first days following XIST induction (data not shown). Alternatively, XIST may actively 'refrain' from silencing its own promoter; however this process would need to be sequence-independent, as multiple promoters have been previously used to drive *XIST*/*Xist* expression.

In summary, we have shown that the transgenic *XIST* in HT1080 male fibrosarcoma cells induces gene silencing in multiple integration sites and across at least a 100 kb region. By recapitulating XIST-induced gene silencing, but stopping short of full-featured X inactivation, our *XIST* transgene exposes the most basal aspects of XIST function. Further data on the relationship of repeat A sequence, function and proteins that it recruits will provide foundation for elucidating the yet unclear connection between sequence of lncRNAs like XIST/Xist and their ability to silence chromatin.

REFERENCES

1. Wood, AJ and Oakey, RJ, Genomic imprinting in mammals: Emerging themes and established theories. *PLoS Genet*, **2**(11): 1677-1685 (2006).
2. Baylin, SB and Jones, PA, A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer*, **11**(10): 726-734 (2011).
3. Hirst, M and Marra, MA, Epigenetics and human disease. *Int J Biochem Cell Biol*, **41**(1): 136-146 (2009).
4. Saitou, M, Kagiwada, S, and Kurimoto, K, Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development*, **139**(1): 15-31 (2012).
5. Lyon, MF, Gene action in the X-chromosome of the mouse *(Mus musculus L.). Nature*, **190**: 372-373 (1961).
6. van den Berg, IM, Laven, JSE, Stevens, M, Jonkers, I, Galjaard, RJ, Gribnau, J, and van Doorninck, JH, X Chromosome Inactivation Is Initiated in Human Preimplantation Embryos. *Am J Hum Genet*, **84**(6): 771-779 (2009).
7. van den Berg, IM, Galjaard, RJ, Laven, JSE, and van Doorninck, JH, XCI in preimplantation mouse and human embryos: first there is remodelling ... *Hum Genet*, **130**(2): 203-215 (2011).
8. Wutz, A, Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet*, **12**(8): 542-553 (2011).
9. Orstavik, KH, X chromosome inactivation in clinical practice. *Hum Genet*, **126**(3): 363-373 (2009).
10. Bretherick, K, Gair, J, and Robinson, WP, The association of skewed X chromosome inactivation with aneuploidy in humans. *Cytogenet Genome Res*, **111**(3-4): 260-265 (2005).
11. Yang, C, Chapman, AG, Kelsey, AD, Minks, J, Cotton, AM, and Brown, CJ, X-chromosome inactivation: molecular mechanisms from the human perspective. *Hum Genet*, **130**(2): 175-185 (2011).
12. Russell, LB, Mammalian X-chromosome action: inactivation limited in spread and in region of origin. *Science*, **140**: 976-978 (1963).
13. Brown, CJ, Lafreniere, RG, Powers, VE, Sebastio, G, Ballabio, A, Pettigrew, AL, Ledbetter, DH, Levy, E, Craig, IW, and Willard, HF, Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature*, **349**(6304): 82-84 (1991).
14. Rastan, S, Non-random X-chromosome inactivation in mouse X-autosome translocation embryos-location of the inactivation centre. *J Embryol Exp Morph*, **78**: 1-22 (1983).
15. Lee, JT, Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nat Rev Mol Cell Biol*, **12**(12): 815-826 (2011).
16. Starmer, J and Magnuson, T, A new model for random X chromosome inactivation. *Development*, **136**(1): 1-10 (2009).
17. Debrand, E, Chureau, C, Arnaud, D, Avner, P, and Heard, E, Functional analysis of the DXPas34 locus, a 3' regulator of Xist expression. *Mol Cell Biol*, **19**: 8513-8525 (1999).
18. Ogawa, Y and Lee, JT, *Xite*, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol Cell*, **11**: 731-743 (2003).
19. Lee, JT, Regulation of X-chromosome counting by Tsix and Xite sequences. *Science*, **309**(5735): 768-771 (2005).
20. Migeon, BR, Chowdury, AK, Dunston, JA, and McIntosh, I, 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 (2001).
21. Migeon, BR, Lee, CH, Chowdury, AK, and Carpenter, H, Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. *Am J Hum Genet*, **71**: 286-293 (2002).
22. Tian, D, Sun, S, and Lee, JT, The Long Noncoding RNA, Jpx, Is a Molecular Switch for X Chromosome Inactivation. *Cell*, **143**(3): 390-403 (2010).

23. Chureau, C, Chantalat, S, Romito, A, Galvani, A, Duret, L, Avner, P, and Rougeulle, C, Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum Mol Genet*, **20**(4): 705-718 (2011).
24. Zhao, J, Sun, BK, Erwin, JA, Song, JJ, and Lee, JT, Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, **322**(5902): 750-756 (2008).
25. Hu, HL, Du, LT, Nagabayashi, G, Seeger, RC, and Gatti, RA, ATM is down-regulated by N-Myc-regulated microRNA-421. *Proc Natl Acad Sci USA*, **107**(4): 1506-1511 (2010).
26. Bacher, CP, Guggiari, M, Brors, B, Augui, S, Clerc, P, Avner, P, Eils, R, and Heard, E, Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol*, **8**(3): 293-299 (2006).
27. Xu, N, Tsai, CL, and Lee, JT, Transient homologous chromosome pairing marks the onset of X inactivation. *Science*, **311**(5764): 1149-1152 (2006).
28. Xu, N, Donohoe, ME, Silva, SS, and Lee, JT, Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet*, **39**(11): 1390-1396 (2007).
29. Augui, S, Filion, GJ, Huart, S, Nora, E, Guggiari, M, Maresca, M, Stewart, AF, and Heard, E, Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science*, **318**(5856): 1632-1636 (2007).
30. Grant, J, Mahadevaiah, SK, Khil, P, Sangrithi, MN, Royo, H, Duckworth, J, McCarrey, JR, VandeBerg, JL, Renfree, MB, Taylor, W, et al., Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature*, **487**(7406): 254-258 (2012).
31. Conrad, T and Akhtar, A, Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet*, **13**(2): 123-134 (2012).
32. Meyer, BJ, Targeting X chromosomes for repression. *Curr Opin Genet Dev*, **20**(2): 179-189 (2010).
33. Brown, CJ, Ballabio, A, Rupert, JL, Lafreniere, RG, Grompe, M, Tonlorenzi, R, and Willard, HF, A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, **349**: 38-44 (1991).
34. Brown, CJ, Hendrich, BD, Rupert, JL, Lafreniere, RG, Xing, Y, Lawrence, J, and Willard, HF, 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 (1992).
35. Penny, GD, Kay, GF, Sheardown, SA, Rastan, S, and Brockdorff, N, Requirement for Xist in X chromosome inactivation. *Nature*, **379**: 131-137 (1996).
36. Brockdorff, N, Ashworth, A, Kay, GF, McCabe, VM, Norris, DP, Cooper, PJ, Swift, S, and Rastan, S, The product of the mouse Xist gene is a 15kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, **71**: 515-526 (1992).
37. Nesterova, TB, Slobodyanyuk, SY, Elisaphenko, EA, Shevchenko, AI, Johnston, C, Pavlova, ME, Rogozin, IB, Kolesnikov, NN, Brockdorff, N, and Zakian, SM, Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res*, **11**(5): 833-849 (2001).
38. Elisaphenko, EA, Kolesnikov, NN, Shevchenko, AI, Rogozin, IB, Nesterova, TB, Brockdorff, N, and Zakian, SM, A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable Elements. *PLoS One*, **3**(6) (2008).
39. Duret, L, Chureau, C, Samain, S, Weissenbach, J, and Avner, P, The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science*, **312**(5780): 1653-1655 (2006).
40. Horvath, JE, Sheedy, CB, Merrett, SL, Diallo, AB, Swofford, DL, Green, ED, Willard, HF, and Progra, NCS, Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate XIST locus. *Genome Res*, **21**(6): 850-862 (2011).
41. Yen, ZC, Meyer, IM, Karalic, S, and Brown, CJ, A cross-species comparison of X-chromosome inactivation in Euteria. *Genomics*, **90**(4): 453-463 (2007).
42. Wutz, A, Rasmussen, TP, and Jaenisch, R, Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet*, **30**(2): 167-174 (2002).

43. Guttman, M and Rinn, JL, Modular regulatory principles of large non-coding RNAs. *Nature*, **482**(7385): 339-346 (2012).
44. Chow, JC, Hall, LL, Baldry, SE, Thorogood, NP, Lawrence, JB, and Brown, CJ, Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. *Proc Natl Acad Sci U S A*, **104**(24): 10104-10109 (2007).
45. Hoki, Y, Kimura, N, Kanbayashi, M, Amakawa, Y, Ohhata, T, Sasaki, H, and Sado, T, A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development*, **136**(1): 139-146 (2009).
46. Chaumeil, J, Le Baccon, P, Wutz, A, and Heard, E, A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev*, **20**(16): 2223-2237 (2006).
47. Kohlmaier, A, Savarese, F, Lachner, M, Martens, J, Jenuwein, T, and Wutz, A, A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol*, **2**(7): E171 (2004).
48. Plath, K, Fang, J, Mlynarczyk-Evans, SK, Cao, R, Worringer, KA, Wang, H, de la Cruz, CC, Otte, AP, Panning, B, and Zhang, Y, Role of histone H3 lysine 27 methylation in X inactivation. *Science*, **300**(5616): 131-135 (2003).
49. Schoeftner, S, Sengupta, AK, Kubicek, S, Mechtler, K, Spahn, L, Koseki, HH, Jenuwein, T, and Wutz, A, Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J*, **25**(13): 3110-3122 (2006).
50. Brannan, CI, Dees, EC, Ingram, RS, and Tilghman, SM, The product of the *H*19 gene may function as an RNA. *Mol Cell Biol*, **10**(1): 28-36 (1990).
51. Ponting, CP, Oliver, PL, and Reik, W, Evolution and functions of long noncoding RNAs. *Cell*, **136**(4): 629-641 (2009).
52. Hung, T and Chang, HY, Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol*, **7**(5): 582-585 (2010).
53. Wutz, A, RNA-mediated silencing mechanisms in mammalian cells. *Prog Mol Biol Transl Sci*, **101**: 351-376 (2011).
54. Jeon, Y and Lee, JT, YY1 tethers Xist RNA to the inactive X nucleation center. *Cell*, **146**(1): 119-133 (2011).
55. Lee, JT, Davidow, LS, and Warshawsky, D, *Tsix*, a gene antisense to Xist at the X-inactivation center. *Nat Genet*, **21**: 400-404 (1999).
56. Rinn, JL, Kertesz, M, Wang, JK, Squazzo, SL, Xu, X, Brugmann, SA, Goodnough, LH, Helms, JA, Farnham, PJ, Segal, E, et al., Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**(7): 1311-1323 (2007).
57. Tsai, MC, Manor, O, Wan, Y, Mosammaparast, N, Wang, JK, Lan, F, Shi, Y, Segal, E, and Chang, HY, Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science*, **329**(5992): 689-693 (2010).
58. Gupta, RA, Shah, N, Wang, KC, Kim, J, Horlings, HM, Wong, DJ, Tsai, MC, Hung, T, Argani, P, Rinn, JL, et al., Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464**(7291): 1071-U1148 (2010).
59. Wang, KC, Yang, YW, Liu, B, Sanyal, A, Corces-Zimmerman, R, Chen, Y, Lajoie, BR, Protacio, A, Flynn, RA, Gupta, RA, et al., A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, **472**(7341): 120-124 (2011).
60. Zhang, XQ, Lian, Z, Padden, C, Gerstein, MB, Rozowsky, J, Snyder, M, Gingeras, TR, Kapranov, P, Weissman, SM, and Newburger, PE, A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood*, **113**(11): 2526-2534 (2009).
61. Seidl, CIM, Stricker, SH, and Barlow, DP, The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J*, **25**(15): 3565-3575 (2006).

62. Lyle, R, Watanabe, D, Vruchte, DT, Lerchner, W, Smrzka, OW, Wutz, A, Schageman, J, Hahner, L, Davies, C, and Barlow, DP, The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat Genet*, **25**(1): 19-21 (2000).
63. Nagano, T, Mitchell, JA, Sanz, LA, Pauler, FM, Ferguson-Smith, AC, Feil, R, and Fraser, P, The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. *Science*, **322**(5908): 1717-1720 (2008).
64. Golding, MC, Magri, LS, Zhang, L, Lalone, SA, Higgins, MJ, and Mann, MR, Depletion of Kcnq1ot1 non-coding RNA does not affect imprinting maintenance in stem cells. *Development*, **138**(17): 3667-3678 (2011).
65. Pandey, RR, Mondal, T, Mohammad, F, Enroth, S, Redrup, L, Komorowski, J, Nagano, T, Mancini-DiNardo, D, and Kanduri, C, Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Mol Cell*, **32**(2): 232-246 (2008).
66. Jones, MJ, Bogutz, AB, and Lefebvre, L, An extended domain of Kcnq1ot1 silencing revealed by an imprinted fluorescent reporter. *Mol Cell Biol*, **31**(14): 2827-2837 (2011).
67. Mohammad, F, Pandey, RR, Nagano, T, Chakalova, L, Mondal, T, Fraser, P, and Kanduri, C, Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol Cell Biol*, **28**(11): 3713-3728 (2008).
68. Kohler, A and Hurt, E, Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol*, **8**(10): 761-773 (2007).
69. Hocine, S, Singer, RH, and Grunwald, D, RNA Processing and Export. *Cold Spring Harb Perspect Biol*, **2**(12) (2010).
70. Ghosh, S and Jacobson, A, RNA decay modulates gene expression and controls its fidelity. *Wiley Interdiscip Rev RNA*, **1**(3): 351-361 (2010).
71. Clemson, CM, McNeil, JA, Willard, HF, and Lawrence, JB, XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol*, **132**(3): 259-275 (1996).
72. Cohen, HR and Panning, B, XIST RNA exhibits nuclear retention and exhibits reduced association with the export factor TAP/NXF1. *Chromosoma*, **116**(4): 373-383 (2007).
73. Maenner, S, Blaud, M, Fouillen, L, Savoye, A, Marchand, V, Dubois, A, Sanglier-Cianferani, S, Van Dorsselaer, A, Clerc, P, Avner, P, et al., 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association. *PLoS Biol*, **8**(1) (2010).
74. Royce-Tolland, ME, Andersen, AA, Koyfman, HR, Talbot, DJ, Wutz, A, Tonks, ID, Kay, GF, and Panning, B, The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. *Nat Struct Mol Biol*, **17**(8): 948-955 (2010).
75. Morey, L and Helin, K, Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci*, **35**(6): 323-332 (2010).
76. Tavares, L, Dimitrova, E, Oxley, D, Webster, J, Poot, R, Demmers, J, Bezstarosti, K, Taylor, S, Ura, H, Koide, H, et al., RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell*, **148**(4): 664-678 (2012).
77. Lagarou, A, Mohd-Sarip, A, Moshkin, YM, Chalkley, GE, Bezstarosti, K, Demmers, JA, and Verrijzer, CP, dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev*, **22**(20): 2799-2810 (2008).
78. Ku, M, Koche, RP, Rheinbay, E, Mendenhall, EM, Endoh, M, Mikkelsen, TS, Presser, A, Nusbaum, C, Xie, X, Chi, AS, et al., Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet*, **4**(10): e1000242 (2008).
79. Francis, NJ, Kingston, RE, and Woodcock, CL, Chromatin compaction by a polycomb group protein complex. *Science*, **306**(5701): 1574-1577 (2004).
80. Zhou, W, Zhu, P, Wang, J, Pascual, G, Ohgi, KA, Lozach, J, Glass, CK, and Rosenfeld, MG, Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol Cell*, **29**(1): 69-80 (2008).

81. Stock, JK, Giadrossi, S, Casanova, M, Brookes, E, Vidal, M, Koseki, H, Brockdorff, N, Fisher, AG, and Pombo, A, Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol*, **9**(12): 1428-1435 (2007).
82. Muller, J and Verrijzer, P, Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr Opin Genet Dev*, **19**(2): 150-158 (2009).
83. de Napoles, M, Mermoud, JE, Wakao, R, Tang, YA, Endoh, M, Appanah, R, Nesterova, TB, Silva, J, Otte, AP, Vidal, M, et al., Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Dev Cell*, **7**(5): 663-676 (2004).
84. Plath, K, Talbot, D, Hamer, KM, Otte, AP, Yang, TP, Jaenisch, R, and Panning, B, Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. *J Cell Biol*, **167**(6): 1025-1035 (2004).
85. Leeb, M and Wutz, A, Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J Cell Biol*, **178**(2): 219-229 (2007).
86. Czermin, B, Melfi, R, McCabe, D, Seitz, V, Imhof, A, and Pirrotta, V, Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal polycomb sites. *Cell*, **111**(2): 185-196 (2002).
87. Bernstein, E, Duncan, EM, Masui, O, Gil, J, Heard, E, and Allis, CD, Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol*, **26**(7): 2560-2569 (2006).
88. Martin, C and Zhang, Y, The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol*, **6**(11): 838-849 (2005).
89. Nekrasov, M, Klymenko, T, Fraterman, S, Papp, B, Oktaba, K, Kocher, T, Cohen, A, Stunnenberg, HG, Wilm, M, and Muller, J, Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J*, **26**(18): 4078-4088 (2007).
90. Sarma, K, Margueron, R, Ivanov, A, Pirrotta, V, and Reinberg, D, Ezh2 requires PHF1 to efficiently catalyze h3 lysine 27 trimethylation in vivo. *Mol Cell Biol*, **28**(8): 2718-2731 (2008).
91. Cao, R, Wang, HB, He, J, Erdjument-Bromage, H, Tempst, P, and Zhang, Y, Role of hPHF1 in H3K27 methylation and Hox gene silencing. *Mol Cell Biol*, **28**(5): 1862-1872 (2008).
92. Boulay, G, Rosnoblet, C, Guerardel, C, Angrand, PO, and Leprince, D, Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes. *Biochem J*, **434**: 333-342 (2011).
93. Casanova, M, Preissner, T, Cerase, A, Poot, R, Yamada, D, Li, X, Appanah, R, Bezstarosti, K, Demmers, J, Koseki, H, et al., Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development*, **138**(8): 1471-1482 (2011).
94. Dou, YL, Milne, TA, Ruthenburg, AJ, Lee, S, Lee, JW, Verdine, GL, Allis, CD, and Roeder, RG, Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol*, **13**(8): 713-719 (2006).
95. Patel, A, Vought, VE, Dharmarajan, V, and Cosgrove, MS, A Novel Non-SET Domain Multi-subunit Methyltransferase Required for Sequential Nucleosomal Histone H3 Methylation by the Mixed Lineage Leukemia Protein-1 (MLL1) Core Complex. *J Biol Chem*, **286**(5): 3359-3369 (2011).
96. Mikkelsen, TS, Ku, M, Jaffe, DB, Issac, B, Lieberman, E, Giannoukos, G, Alvarez, P, Brockman, W, Kim, TK, Koche, RP, et al., Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, (2007).
97. Pullirsch, D, Hartel, R, Kishimoto, H, Leeb, M, Steiner, G, and Wutz, A, The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. *Development*, **137**(6): 935-943 (2010).
98. Wysocka, J, Allis, CD, and Coonrod, S, Histone arginine methylation and its dynamic regulation. *Front Biosci*, **11**: 344-355 (2006).

99. Choi, HK, Choi, KC, Oh, SY, Kang, HB, Lee, YH, Haam, S, Ahn, YH, Kimi, KS, Kim, K, and Yoon, HG, The functional role of the CARM1-SNF5 complex and its associated HMT activity in transcriptional activation by thyroid hormone receptor. *Exp Mol Med*, **39**(4): 544-555 (2007).
100. Xu, W, Cho, H, Kadam, S, Banayo, EM, Anderson, S, Yates, JR, Emerson, BM, and Evans, RM, A methylation-mediator complex in hormone signaling. *Genes Dev*, **18**(2): 144-156 (2004).
101. An, W, Kim, J, and Roeder, RG, Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell*, **117**(6): 735-748 (2004).
102. Chaumeil, J, Okamoto, I, Guggiari, M, and Heard, E, Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. *Cytogenet Genome Res*, **99**(1-4): 75-84 (2002).
103. Tachibana, M, Sugimoto, K, Nozaki, M, Ueda, J, Ohta, T, Ohki, M, Fukuda, M, Takeda, N, Niida, H, Kato, H, et al., G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes & Devel*, **16**: 1779-1791 (2002).
104. Tachibana, M, Ueda, J, Fukuda, M, Takeda, N, Ohta, T, Iwanari, H, Sakihama, T, Kodama, T, Hamakubo, T, and Shinkai, Y, Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev*, **19**(7): 815-826 (2005).
105. Heard, E, Rougelle, C, Arnaud, D, Avner, P, Allis, CD, and Spector, DL, Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell*, **107**: 727-738 (2001).
106. Okamoto, I, Otte, AP, Allis, CD, Reinberg, D, and Heard, E, Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science*, **303**(5658): 644-649 (2004).
107. Wagschal, A, Sutherland, HG, Woodfine, K, Henckel, A, Chebli, K, Schulz, R, Oakey, RJ, Bickmore, WA, and Feil, R, G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Mol Cell Biol*, **28**(3): 1104-1113 (2008).
108. Ohhata, T, Tachibana, M, Tada, M, Tada, T, Sasaki, H, Shinkai, Y, and Sado, T, X-inactivation is stably maintained in mouse embryos deficient for histone methyl transferase G9a. *Genesis*, **40**(3): 151-156 (2004).
109. Chadwick, BP and Willard, HF, Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc Natl Acad Sci U S A*, **101**: 17450-17455 (2004).
110. Fang, J, Feng, Q, Ketel, CS, Wang, HB, Cao, R, Xia, L, Erdjument-Bromage, H, Tempst, P, Simon, JA, and Zhang, Y, Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Curr Biol*, **12**(13): 1086-1099 (2002).
111. Nishioka, K, Rice, JC, Sarma, K, Erdjument-Bromage, H, Werner, J, Wang, YM, Chuikov, S, Valenzuela, P, Tempst, P, Steward, R, et al., PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol Cell*, **9**(6): 1201-1213 (2002).
112. Rice, JC, Nishioka, K, Sarma, K, Steward, R, Reinberg, D, and Allis, CD, Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes Dev*, **16**(17): 2225-2230 (2002).
113. Beck, DB, Oda, H, Shen, SS, and Reinberg, D, PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev*, **26**(4): 325-337 (2012).
114. Oda, H, Okamoto, I, Murphy, N, Chu, JH, Price, SM, Shen, MM, Torres-Padilla, ME, Heard, E, and Reinberg, D, Monomethylation of Histone H4-Lysine 20 Is Involved in Chromosome Structure and Stability and Is Essential for Mouse Development. *Mol Cell Biol*, **29**(8): 2278-2295 (2009).
115. Silva, J, Mak, W, Zvetkova, I, Appanah, R, Nesterova, TB, Webster, Z, Peters, AH, Jenuwein, T, Otte, AP, and Brockdorff, N, Establishment of histone H3 methylation on the inactive X

chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell*, **4**(4): 481-495 (2003).
116. Shi, Y, Lan, F, Matson, C, Mulligan, P, Whetstine, JR, Cole, PA, Casero, RA, and Shi, Y, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, **119**(7): 941-953 (2004).
117. Metzger, E, Wissmann, M, Yin, N, Muller, JM, Schneider, R, Peters, AHFM, Gunther, T, Buettner, R, and Schule, R, LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature*, **437**(7057): 436-439 (2005).
118. Pasini, D, Hansen, KH, Christensen, J, Agger, K, Cloos, PAC, and Helin, K, Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev*, **22**(10): 1345-1355 (2008).
119. Brinkman, AB, Roelofsen, T, Pennings, SW, Martens, JH, Jenuwein, T, and Stunnenberg, HG, Histone modification patterns associated with the human X chromosome. *EMBO Rep*, **7**(6): 628-634 (2006).
120. Gamble, MJ and Kraus, WL, Multiple facets of the unique histone variant macroH2A From genomics to cell biology. *Cell Cycle*, **9**(13): 2568-2574 (2010).
121. Costanzi, C and Pehrson, JR, Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature*, **393**: 599-601 (1998).
122. Gilbert, SL, Pehrson, JR, and Sharp, PA, XIST RNA associates with specific regions of the inactive X chromatin. *J Biol Chem*, **275**(47): 36491-36494 (2000).
123. Csankovszki, G, Panning, B, Bates, B, Pehrson, JR, and Jaenisch, R, Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat Genet*, **22**: 323-324 (1999).
124. Rasmussen, TP, Wutz, A, Pehrson, JR, and Jaenisch, R, Expression of Xist RNA is sufficient to initiate macrochromatin body formation. *Chromosoma*, **110**: 411-420 (2001).
125. Changolkar, LN and Pehrson, JR, macroH2A1 histone variants are depleted on active genes but concentrated on the inactive X chromosome. *Mol Cell Biol*, **26**(12): 4410-4420 (2006).
126. Choo, JH, Kim, JD, Chung, JH, Stubbs, L, and Kim, J, Allele-specific deposition of macroH2A1 in imprinting control regions. *Hum Mol Genet*, **15**(5): 717-724 (2006).
127. Chen, ZX and Riggs, AD, DNA Methylation and Demethylation in Mammals. *J Biol Chem*, **286**(21): 18347-18353 (2011).
128. Weber, M, Davies, JJ, Wittig, D, Oakeley, EJ, Haase, M, Lam, WL, and Schubeler, D, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet*, **37**(8): 853-862 (2005).
129. Hellman, A and Chess, A, Gene body-specific methylation on the active X chromosome. *Science*, **315**(5815): 1141-1143 (2007).
130. Cotton, AM, Lam, L, Affleck, JG, Wilson, IM, Penaherrera, MS, McFadden, DE, Kobor, MS, Lam, WL, Robinson, WP, and Brown, CJ, Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Hum Genet*, **130**(2): 187-201 (2011).
131. Lock, LF, Takagi, N, and Martin, GR, Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell*, **48**: 39-46 (1987).
132. Blewitt, ME, Gendrel, AV, Pang, Z, Sparrow, DB, Whitelaw, N, Craig, JM, Apedaile, A, Hilton, DJ, Dunwoodie, SL, Brockdorff, N, et al., SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet*, **40**(5): 663-669 (2008).
133. Hansen, RS, Stoger, R, Wijmenga, C, Stanek, AM, Canfield, TK, Luo, P, Matarazzo, MR, D'Esposito, M, Feil, R, Gimelli, G, et al., Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. *Hum Mol Genet*, **9**(18): 2575-2587 (2000).
134. Csankovszki, G, Nagy, A, and Jaenisch, R, Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J Cell Biol*, **153**: 773-783 (2001).

135. Hansen, RS, X inactivation-specific methylation of LINE-1 elements by DNMT3B: implications for the Lyon repeat hypothesis. *Hum Mol Genet*, **12**(19): 2559-2567 (2003).
136. Sado, T, Fenner, MH, Tan, SS, Tam, P, Shioda, T, and Li, E, X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol*, **225**(2): 294-303 (2000).
137. De La Fuente, R, Baumann, C, and Viveiros, MM, Role of ATRX in chromatin structure and function: implications for chromosome instability and human disease. *Reproduction*, **142**(2): 221-234 (2011).
138. Berube, NG, ATRX in chromatin assembly and genome architecture during development and disease. *Biochem Cell Biol*, **89**(5): 435-444 (2011).
139. Gibbons, RJ, McDowell, TL, Raman, S, O'Rourke, DM, Garrick, D, Ayyub, H, and Higgs, DR, Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet*, **24**(4): 368-371 (2000).
140. McDowell, TL, Gibbons, RJ, Sutherland, H, O'Rourke, DM, Bickmore, WA, Pombo, A, Turley, H, Gatter, K, Picketts, DJ, Buckle, VJ, et al., Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci USA*, **96**(24): 13983-13988 (1999).
141. Xue, YT, Gibbons, R, Yan, ZJ, Yang, DF, McDowell, TL, Sechi, S, Qin, J, Zhou, SL, Higgs, D, and Wang, WD, The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proc Natl Acad Sci USA*, **100**(19): 10635-10640 (2003).
142. Law, MJ, Lower, KM, Voon, HPJ, Hughes, JR, Garrick, D, Viprakasit, V, Mitson, M, De Gobbi, M, Marra, M, Morris, A, et al., ATR-X Syndrome Protein Targets Tandem Repeats and Influences Allele-Specific Expression in a Size-Dependent Manner. *Cell*, **143**(3): 367-378 (2010).
143. Gibbons, RJ, Picketts, DJ, Villard, L, and Higgs, DR, Mutations in a Putative Global Transcriptional Regulator Cause X-Linked Mental-Retardation with Alpha-Thalassemia (Atr-X Syndrome). *Cell*, **80**(6): 837-845 (1995).
144. Gibbons, RJ, Pellagatti, A, Garrick, D, Wood, WG, Malik, N, Ayyub, H, Langford, C, Boultwood, J, Wainscoat, JS, and Higgs, DR, Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the alpha-thalassemia myelodysplasia syndrome (ATMDS). *Nat Genet*, **34**(4): 446-449 (2003).
145. Baumann, C and De La Fuente, R, ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells. *Chromosoma*, **118**(2): 209-222 (2009).
146. Baumann, C, Schmidtmann, A, Muegge, K, and De La Fuente, R, Association of ATRX with pericentric heterochromatin and the Y chromosome of neonatal mouse spermatogonia. *BMC Mol Biol*, **9** (2008).
147. Kernohan, KD, Jiang, Y, Tremblay, DC, Bonvissuto, AC, Eubanks, JH, Mann, MRW, and Berube, NG, ATRX Partners with Cohesin and MeCP2 and Contributes to Developmental Silencing of Imprinted Genes in the Brain. *Dev Cell*, **18**(2): 191-202 (2010).
148. Gordon, S, Akopyan, G, Garban, H, and Bonavida, B, Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene*, **25**(8): 1125-1142 (2006).
149. Millau, JF and Gaudreau, L, CTCF, cohesin, and histone variants: connecting the genome. *Biochem Cell Biol*, **89**(5): 505-513 (2011).
150. Bell, AC and Felsenfeld, G, Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**(6785): 482-485 (2000).
151. Bell, AC, West, AG, and Felsenfeld, G, The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, **98**(3): 387-396 (1999).

152. Wendt, KS, Yoshida, K, Itoh, T, Bando, M, Koch, B, Schirghuber, E, Tsutsumi, S, Nagae, G, Ishihara, K, Mishiro, T, et al., Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, **451**(7180): 796-U793 (2008).
153. Stedman, W, Kang, H, Lin, S, Kissil, JL, Bartolomei, MS, and Lieberman, PM, Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J*, **27**(4): 654-666 (2008).
154. Parelho, V, Hadjur, S, Spivakov, M, Leleu, M, Sauer, S, Gregson, HC, Jarmuz, A, Canzonetta, C, Webster, Z, Nesterova, T, et al., Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, **132**(3): 422-433 (2008).
155. Yao, HJ, Brick, K, Evrard, Y, Xiao, TJ, Camerini-Otero, RD, and Felsenfeld, G, Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev*, **24**(22): 2543-2555 (2010).
156. Donohoe, ME, Zhang, LF, Xu, N, Shi, Y, and Lee, JT, Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell*, **25**(1): 43-56 (2007).
157. Navarro, P and Avner, P, An embryonic story: Analysis of the gene regulative network controlling Xist expression in mouse embryonic stem cells. *Bioessays*, **32**(7): 581-588 (2010).
158. Kim, MK and Nikodem, VM, hnRNP U inhibits carboxy-terminal domain phosphorylation by TFIIH and represses RNA polymerase II elongation. *Mol Cell Biol*, **19**(10): 6833-6844 (1999).
159. Ameyar-Zazoua, M, Souidi, M, Fritsch, L, Robin, P, Thomas, A, Hamiche, A, Percipalle, P, Ait-Si-Ali, S, and Harel-Bellan, A, Physical and Functional Interaction between Heterochromatin Protein 1 alpha and the RNA-binding Protein Heterogeneous Nuclear Ribonucleoprotein U. *J Biol Chem*, **284**(41): 27974-27979 (2009).
160. Martens, JHA, Verlaan, M, Kalkhoven, E, Dorsman, JC, and Zantema, A, Scaffold/Matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated Nucleosomes. *Mol Cell Biol*, **22**(8): 2598-2606 (2002).
161. Yugami, M, Kabe, Y, Yamaguchi, Y, Wada, T, and Handa, H, hnRNP-U enhances the expression of specific genes by stabilizing mRNA. *FEBS Lett*, **581**(1): 1-7 (2007).
162. Kukalev, A, Nord, Y, Palmberg, C, Bergman, T, and Percipalle, P, Actin and hnRNP U cooperate for productive transcription by RNA polymerase II. *Nat Struct Mol Biol*, **12**(3): 238-244 (2005).
163. Fu, D and Collins, K, Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation. *Mol Cell*, **28**(5): 773-785 (2007).
164. Helbig, R and Fackelmayer, FO, Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma*, **112**(4): 173-182 (2003).
165. Hasegawa, Y, Brockdorff, N, Kawano, S, Tsutui, K, Tsutui, K, and Nakagawa, S, The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. *Dev Cell*, **19**(3): 469-476 (2010).
166. Galande, S, Purbey, PK, Notani, D, and Kumar, PP, The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr Opin Genet Dev*, **17**(5): 408-414 (2007).
167. Dobreva, G, Dambacher, J, and Grosschedl, R, SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin 11 3048 gene expression. *Genes Dev*, **17**(24): 3048-3061 (2003).
168. Britanova, O, Akopov, S, Lukyanov, S, Gruss, P, and Tarabykin, V, Novel transcription factor Satb2 interacts with matrix attachment region DNA elements in a tissue-specific manner and demonstrates cell-type-dependent expression in the developing mouse CNS. *Eur J Neurosci*, **21**(3): 658-668 (2005).
169. Agrelo, R, Souabni, A, Novatchkova, M, Haslinger, C, Leeb, M, Komnenovic, V, Kishimoto, H, Gresh, L, Kohwi-Shigematsu, T, Kenner, L, et al., SATB1 Defines the Developmental Context

for Gene Silencing by Xist in Lymphoma and Embryonic Cells. *Dev Cell*, **16**(4): 507-516 (2009).
170. Han, SP, Tang, YH, and Smith, R, Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J*, **430**: 379-392 (2010).
171. Bomsztyk, K, Denisenko, O, and Ostrowski, J, HnRNP K: One protein multiple processes. *Bioessays*, **26**(6): 629-638 (2004).
172. Mikula, M, Dzwonek, A, Karczmarski, J, Rubel, T, Dadlez, M, Wyrwicz, LS, Bomsztyk, K, and Ostrowski, J, Landscape of the hnRNP K protein-protein interactome. *Proteomics*, **6**(8): 2395-2406 (2006).
173. Huarte, M, Guttman, M, Feldser, D, Garber, M, Koziol, MJ, Kenzelmann-Broz, D, Khalil, AM, Zuk, O, Amit, I, Rabani, M, et al., A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. *Cell*, **142**(3): 409-419 (2010).
174. Venkitaraman, AR, Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell*, **108**(2): 171-182 (2002).
175. Roy, R, Chun, J, and Powell, SN, BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer*, **12**(1): 68-78 (2012).
176. Ganesan, S, Silver, DP, Greenberg, RA, Avni, D, Drapkin, R, Miron, A, Mok, SC, Randrianarison, V, Brodie, S, Salstrom, J, et al., BRCA1 supports XIST RNA concentration on the inactive X chromosome. *Cell*, **111**(3): 393-405 (2002).
177. Silver, DP, Dimitrov, SD, Feunteun, J, Gelman, R, Drapkin, R, Lu, SD, Shestakova, E, Velmurugan, S, Denunzio, N, Dragomir, S, et al., Further evidence for BRCA1 communication with the inactive X chromosome. *Cell*, **128**(5): 991-1002 (2007).
178. Xiao, C, Sharp, JA, Kawahara, M, Davalos, AR, Difilippantonio, MJ, Hu, Y, Li, W, Cao, L, Buetow, K, Ried, T, et al., The XIST noncoding RNA functions independently of BRCA1 in X inactivation. *Cell*, **128**(5): 977-989 (2007).
179. Pageau, GJ, Hall, LL, and Lawrence, JB, 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*, (2006).
180. Turner, JMA, Aprelikova, O, Xu, XL, Wang, RH, Kim, SS, Chandramouli, GVR, Barrett, JC, Burgoyne, PS, and Deng, CX, BRCA1, histone H2AX phosphorylation, and male meiotic sex chromosome inactivation. *Curr Biol*, **14**(23): 2135-2142 (2004).
181. Li, L and Liu, Y, Diverse small non-coding RNAs in RNA interference pathways. *Methods Mol Biol*, **764**: 169-182 (2011).
182. Valencia-Sanchez, MA, Liu, JD, Hannon, GJ, and Parker, R, Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev*, **20**(5): 515-524 (2006).
183. Volpe, TA, Kidner, C, Hall, IM, Teng, G, Grewal, SIS, and Martienssen, RA, Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**(5588): 1833-1837 (2002).
184. Kota, SK, RNAi in X inactivation: contrasting findings on the role of interference. *Bioessays*, **31**(12): 1280-1283 (2009).
185. Ogawa, Y, Sun, BK, and Lee, JT, Intersection of the RNA interference and X-inactivation pathways. *Science*, **320**(5881): 1336-1341 (2008).
186. Kanellopoulou, C, Muljo, SA, Dimitrov, SD, Chen, X, Colin, C, Plath, K, and Livingston, DM, X chromosome inactivation in the absence of Dicer. *Proc Natl Acad Sci USA*, **106**(4): 1122-1127 (2009).
187. Nesterova, TB, Popova, BC, Cobb, BS, Norton, S, Senner, CE, Tang, YA, Spruce, T, Rodriguez, TA, Sado, T, Merkenschlager, M, et al., Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. *Epigenetics & Chromatin*, **1** (2008).
188. Smith, J, Tho, LM, Xu, NH, and Gillespie, DA, The ATM-Chk2 and ATR-Chk1 Pathways in DNA Damage Signaling and Cancer. *Adv Cancer Res*, **108**: 73-112 (2010).

189. Barchi, M, Roig, I, Di Giacomo, M, de Rooij, DG, Keeney, S, and Jasin, M, ATM promotes the obligate XY crossover and both crossover control and chromosome axis integrity on autosomes. *PLoS Genet*, **4**(5) (2008).
190. Ouyang, Y, Salstrom, J, Diaz-Perez, S, Nahas, S, Matsuno, Y, Dawson, D, Teitell, MA, Horvath, S, Riggs, AD, Gatti, RA, et al., Inhibition of Atm and/or Atr disrupts gene silencing on the inactive X chromosome. *Biochem Biophys Res Commun*, **337**(3): 875-880 (2005).
191. Ame, JC, Spenlehauer, C, and de Murcia, G, The PARP superfamily. *Bioessays*, **26**(8): 882-893 (2004).
192. Krishnakumar, R and Kraus, WL, The PARP Side of the Nucleus: Molecular Actions, Physiological Outcomes, and Clinical Targets. *Mol Cell*, **39**(1): 8-24 (2010).
193. Ji, YBA and Tulin, AV, The roles of PARP1 in gene control and cell differentiation. *Curr Opin Genet Dev*, **20**(5): 512-518 (2010).
194. Nusinow, DA, Hernandez-Munoz, I, Fazzio, TG, Shah, GM, Kraus, WL, and Panning, B, Poly(ADP-ribose) polymerase 1 is inhibited by a histone H2A variant, MacroH2A, and contributes to silencing of the inactive X chromosome. *J Biol Chem*, **282**(17): 12851-12859 (2007).
195. Ballas, N and Mandel, G, The many faces of REST oversee epigenetic programming of neuronal genes. *Curr Opin Neurobiol*, **15**(5): 500-506 (2005).
196. Jessberger, R, The many functions of SMC proteins in chromosome dynamics. *Nat Rev Mol Cell Biol*, **3**(10): 767-778 (2002).
197. Sarikas, A, Hartmann, T, and Pan, ZQ, The cullin protein family. *Genome Biol*, **12**(4) (2011).
198. Furukawa, M, He, YZJ, Borchers, C, and Xiong, Y, Targeting of protein ubiquitination by BTB-Cullin 3-Roc1 ubiquitin ligases. *Nat Cell Biol*, **5**(11): 1001-1007 (2003).
199. Takahashi, I, Kameoka, Y, and Hashimoto, K, MacroH2A1.2 binds the nuclear protein Spop. *Biochim Biophys Acta*, **1591**: 63-68 (2002).
200. Hernandez-Munoz, I, Lund, AH, van der Stoop, P, Boutsma, E, Muijrers, I, Verhoeven, E, Nusinow, DA, Panning, B, Marahrens, Y, and van Lohuizen, M, Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. *Proc Natl Acad Sci U S A*, **102**(21): 7635-7640 (2005).
201. Meyer, BJ, X-Chromosome dosage compensation. *WormBook*: 1-14 (2005).
202. Csankovszki, G, Condensin function in dosage compensation. *Epigenetics*, **4**(4): 212-215 (2009).
203. Jans, J, Gladden, JM, Ralston, EJ, Pickle, CS, Michel, AH, Pferdehirt, RR, Eisen, MB, and Meyer, BJ, A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes Dev*, **23**(5): 602-618 (2009).
204. Pferdehirt, RR, Kruesi, WS, and Meyer, BJ, An MLL/COMPASS subunit functions in the C. elegans dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression. *Genes Dev*, **25**(5): 499-515 (2011).
205. Bernstein, E, Kim, SY, Carmell, MA, Murchison, EP, Alcorn, H, Li, MZ, Mills, AA, Elledge, SJ, Anderson, KV, and Hannon, GJ, Dicer is essential for mouse development. *Nat Genet*, **35**(3): 215-217 (2003).
206. Faust, C, Schumacher, A, Holdener, B, and Magnuson, T, The eed mutation disrupts anterior mesoderm production in mice. *Development*, **121**(2): 273-285 (1995).
207. Voncken, JW, Roelen, BA, Roefs, M, de Vries, S, Verhoeven, E, Marino, S, Deschamps, J, and van Lohuizen, M, Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A*, **100**(5): 2468-2473 (2003).
208. Donohoe, ME, Zhang, X, McGinnis, L, Biggers, J, Li, E, and Shi, Y, Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol Cell Biol*, **19**(10): 7237-7244 (1999).
209. Fedoriw, AM, Stein, P, Svoboda, P, Schultz, RM, and Bartolomei, MS, Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science*, **303**(5655): 238-240 (2004).

210. Roshon, MJ and Ruley, HE, Hypomorphic mutation in hnRNP U results in post-implantation lethality. *Transgenic Res*, **14**(2): 179-192 (2005).
211. Singer, JD, Gurian-West, M, Clurman, B, and Roberts, JM, Cullin-3 targets cyclin E for ubiquitination and controls S phase in mammalian cells. *Genes Dev*, **13**(18): 2375-2387 (1999).
212. Rastan, S and Robertson, EJ, X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *J Embryol Exp Morph*, **90**: 379-388 (1985).
213. Hoffman, LM, Hall, L, Batten, JL, Young, H, Pardasani, D, Baetge, EE, Lawrence, J, and Carpenter, MK, X-inactivation status varies in human embryonic stem cell lines. *Stem Cells*, **23**(10): 1468-1478 (2005).
214. Hall, LL, Byron, M, Butler, J, Becker, KA, Nelson, A, Amit, M, Itskovitz-Eldor, J, Stein, J, Stein, G, Ware, C, et al., X-inactivation reveals epigenetic anomalies in most hESC but identifies sublines that initiate as expected. *J Cell Physiol*, (2008).
215. Shen, Y, Matsuno, Y, Fouse, SD, Rao, N, Root, S, Xu, R, Pellegrini, M, Riggs, AD, and Fan, G, X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. *Proc Natl Acad Sci U S A*, **105**(12): 4709-4714 (2008).
216. Silva, SS, Rowntree, RK, Mekhoubad, S, and Lee, JT, X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc Natl Acad Sci U S A*, **105**(12): 4820-4825 (2008).
217. Dvash, T, Lavon, N, and Fan, GP, Variations of X Chromosome Inactivation Occur in Early Passages of Female Human Embryonic Stem Cells. *PLoS One*, **5**(6) (2010).
218. Lengner, CJ, Gimelbrant, AA, Erwin, JA, Cheng, AW, Guenther, MG, Welstead, GG, Alagappan, R, Frampton, GM, Xu, P, Muffat, J, et al., Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell*, **141**(5): 872-883 (2010).
219. Daniels, R, Zuccotti, M, Kinis, T, Serhal, P, and Monk, M, XIST expression in human oocytes and preimplantation embryos. *Am J Hum Genet*, **61**: 33-39 (1997).
220. Ray, PF, Winston, RM, and Handyside, AH, XIST expression from the maternal X chromosome in human male preimplantation embryos at the blastocyst stage. *Hum Mol Genet*, **6**(8): 1323-1327 (1997).
221. Okamoto, I, Patrat, C, Thepot, D, Peynot, N, Fauque, P, Daniel, N, Diabangouaya, P, Wolf, JP, Renard, JP, Duranthon, V, et al., Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature*, **472**(7343): 370-U141 (2011).
222. Migeon, BR, Kazi, E, Haisley-Royster, C, Hu, J, Reeves, R, Call, L, Lawler, A, Moore, CS, Morrison, H, and Jeppesen, P, Human X inactivation center induces random X chromosome inactivation in male transgenic mice. *Genomics*, **59**: 113-121 (1999).
223. Migeon, BR, Winter, H, Kazi, E, Chowdhury, AK, Hughes, A, Haisley-Royster, C, Morrison, H, and Jeppesen, P, 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**(2): 156-162 (2001).
224. Heard, E, Mongelard, F, Arnaud, D, Chureau, C, Vourc'h, C, and Avner, P, Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. *Proc Natl Acad Sci, USA*, **96**: 6841-6846 (1999).
225. Heard, E, Mongelard, F, Arnaud, D, and Avner, P, Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. *Mol Cell Biol*, **19**: 3156-3166 (1999).
226. Wutz, A and Jaenisch, R, A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell*, **5**: 695-705 (2000).
227. Savarese, F, Flahndorfer, K, Jaenisch, R, Busslinger, M, and Wutz, A, Hematopoietic precursor cells transiently reestablish permissiveness for x inactivation. *Mol Cell Biol*, **26**(19): 7167-7177 (2006).

228. Hall, LL, Byron, M, Sakai, K, Carrel, L, Willard, HF, and Lawrence, JB, An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. *Proc Natl Acad Sci U S A*, **99**(13): 8677-8682 (2002).
229. Chow, JC, Hall, LL, Lawrence, JB, and Brown, CJ, Ectopic XIST transcripts in human somatic cells show variable expression and localization. *Cytogenet Genome Res*, **99**(1-4): 92-98 (2002).
230. Yan, CH and Boyd, DD, Histone H3 acetylation and H3K4 methylation define distinct chromatin regions permissive for transgene expression. *Mol Cell Biol*, **26**(17): 6357-6371 (2006).
231. Leung, DC, Dong, KB, Maksakova, IA, Goyal, P, Appanah, R, Lee, S, Tachibana, M, Shinkai, Y, Lehnertz, B, Mager, DL, et al., Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. *Proc Natl Acad Sci U S A*, **108**(14): 5718-5723 (2011).
232. Duszczyk, MM, Zanier, K, and Sattler, M, A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. *Nucleic Acids Res*, **36**(22): 7068-7077 (2008).
233. Duszczyk, MM, Wutz, A, Rybin, V, and Sattler, M, The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. *RNA*, **17**(11): 1973-1982 (2011).
234. Zuker, M, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**(13): 3406-3415 (2003).
235. Doshi, KJ, Cannone, JJ, Cobaugh, CW, and Gutell, RR, Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics*, **5**: 105 (2004).
236. Bannister, AJ and Kouzarides, T, Regulation of chromatin by histone modifications. *Cell Res*, **21**(3): 381-395 (2011).
237. Riester, D, Hildmann, C, Grunewald, S, Beckers, T, and Schwienhorst, A, Factors affecting the substrate specificity of histone deacetylases. *Biochem Biophys Res Commun*, **357**(2): 439-445 (2007).
238. Kouzarides, T, Chromatin modifications and their function. *Cell*, **128**: 693-705 (2007).
239. Aka, JA, Kim, GW, and Yang, XJ, K-acetylation and its enzymes: overview and new developments. *Handb Exp Pharmacol*, **206**: 1-12 (2011).
240. Schroeder, M, Hillemacher, T, Bleich, S, and Frieling, H, The Epigenetic Code in Depression: Implications for Treatment. *Clin Pharmacol Ther*, **91**(2): 310-314 (2012).
241. Song, SH, Han, SW, and Bang, YJ, Epigenetic-Based Therapies in Cancer Progress to Date. *Drugs*, **71**(18): 2391-2403 (2011).
242. Khan, N, Jeffers, M, Kumar, S, Hackett, C, Boldog, F, Khramtsov, N, Qian, XZ, Mills, E, Berghs, SC, Carey, N, et al., Determination of the class and isoform selectivity of small-molecule histone deacetylase inhibitors. *Biochem J*, **409**: 581-589 (2008).
243. Kanhere, A, Viiri, K, Araujo, CC, Rasaiyaah, J, Bouwman, RD, Whyte, WA, Pereira, CF, Brookes, E, Walker, K, Bell, GW, et al., Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol Cell*, **38**(5): 675-688 (2010).
244. Blewitt, ME, Vickaryous, NK, Hemley, SJ, Ashe, A, Bruxner, TJ, Preis, JI, Arkell, R, and Whitelaw, E, An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A*, **102**(21): 7629-7634 (2005).
245. Hong, B, Reeves, P, Panning, B, Swanson, MS, and Yang, TP, Identification of an autoimmune serum containing antibodies against the Barr body. *Proc Natl Acad Sci, USA*, **98**: 8703-8708 (2001).
246. Brooks, WH, Satoh, M, Hong, B, Reeves, WH, and Yang, TP, Autoantibodies from an SLE patient immunostain the Barr body. *Cytogenet Genome Res*, **97**(1-2): 28-31 (2002).
247. Chan, KM, Zhang, H, Malureanu, L, van Deursen, J, and Zhang, ZG, Diverse factors are involved in maintaining X chromosome inactivation. *Proc Natl Acad Sci USA*, **108**(40): 16699-16704 (2011).

248. Schlesinger, Y, Straussman, R, Keshet, I, Farkash, S, Hecht, M, Zimmerman, J, Eden, E, Yakhini, Z, Ben-Shushan, E, Reubinoff, BE, et al., Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet*, **39**(2): 232-236 (2007).
249. Wang, J, Mager, J, Chen, Y, Schneider, E, Cross, JC, Nagy, A, and Magnuson, T, Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat Genet*, **28**(4): 371-375 (2001).
250. Kalantry, S, Mills, KC, Yee, D, Otte, AP, Panning, B, and Magnuson, T, The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nat Cell Biol*, **8**(2): 195-202 (2006).
251. Pizarro, JG, Folch, J, de la Torre, AV, Junyent, F, Verdaguer, E, Jordan, J, Pallas, M, and Camins, A, ATM Is Involved in Cell-Cycle Control Through the Regulation of Retinoblastoma Protein Phosphorylation. *J Cell Biochem*, **110**(1): 210-218 (2010).
252. Savarese, F, Davila, A, Nechanitzky, R, De La Rosa-Velazquez, I, Pereira, CF, Engelke, R, Takahashi, K, Jenuwein, T, Kohwi-Shigematsu, T, Fisher, AG, et al., Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. *Genes Dev*, **23**(22): 2625-2638 (2009).
253. Brockdorff, N, Chromosome silencing mechanisms in X-chromosome inactivation: unknown unknowns. *Development*, **138**(23): 5057-5065 (2011).

APPENDIX

```
                                                  --->.#1.<---  -->.#1.<-
Mus musculus                  ----------TGTTTATATAT-TCTT-GCCCATCGGGGCCACGGATACCT 38
Rattus norvegicus             -----------TTTTACATTT-TTTTTGCCCATCGGGGCCACGGATACCT 38
Ellobius lutescens            -CTTTCTTTTATTCCATCTCTGTTTTTGCCCAT-GGGGCTACAGATGACT 48
Equus caballus                -----------CCTTTTCTGTATTTT-GCCCATCGGGGCTGCGGATACCT 38
Pan troglodytes               -----------TCTTTTCTATATTTT-GCCCATCGGGGCTGCGGATACCT 38
Gorilla gorilla               -----------TCTTTTCTATGTTTT-GCCCATCGGGGCTGCGGATACCT 38
Pongo pygmaeus                -----------TCTCTTTTATATTTT-GCCCATCGGGGCTGCGGATACCT 38
Homo sapiens                  -----------TCTTTTCTATATTTT-GCCCATCGGGGCTGCGGATACCT 38
Macaca mulatta                -----------TCTTTTCTATATTTT-GCCCATCGGGGCTGCGGATACCT 38
Callithrix jacchus            -----------TCTTTTCTATATTTT-GCCCATCGGGGCTGCGGATACCT 38
Echinops telfairi             GGCGGCTCTCTTCTTTCTTGTATTTT-GCCCATCGGGGCTGCGGATACCG 51
Cavia porcellus               ------------------------------------------------    ●
Tursiops truncatus            ------------------------------------------------    ●
Oryctolagus cuniculus         -------------------------------------------------A 1  ●
Erinaceus europaeus           -----------TACTTTTTCTATTTT-GCCCATCGGGGCTGTGGATAACA 38
Sorex araneus                 CTGTGGATAATTGGTATGATTATTT--GCCCAGCGGGGCTGTGGATGCGT 74
Felis catus                   -----------CCTTTTCTACATTTT-GCCCATCGGGGCTGTGGATACCA 38
Bos taurus                    -----------TTTTCTATATTATTT-GCCCATCGCGGCTGTGGATACCT 38
Sus scrofa                    -----------TTTTCTGTGTATTAT-GCCCATCGGGGCTGTGGATACCT 38
Tupaia belangeri              -----------ATTTATGTATATTTT-GCCCATCGGGGCTGAGGATACCT 38
Microcebus murinus            -----------CTTTTTCTATATTTG-GCCCATCGGGGCTGCGGATATCT 38
Canis lupus familiaris        -----------TTTTTTCTATATTTT-GCCCATCGGGGCTGCGGATACCG 38
Ailuropoda melanoleuca        -----------TATATTCTATATTTT-GCCCATCGGGGCTGCGGATACCG 38
Vicugna pacos                 TTTTCCCCCCTCCTTTTCTATATTTT-GCCCATCGGGGTTGTGGATACCT 99
Tarsius syrichta              ------------------------------------------------    ●
Myotis lucifugus              ------------------------------------------------    ●
Pteropus vampyrus             ------------------------------------------------    ●

                              -
Mus musculus                  GTGTGTCCTCC---------------------------------CCGCC 54
Rattus norvegicus             GTGTGTCCTCC---------------------------------CAGCC 54
Ellobius lutescens            GCATGTACTCCTT-------------------------CCCCGTCCCTTC 73
Equus caballus                G---ATCTTCT---------------------------------TATTA 51
Pan troglodytes               GGTTT---------------------------------------TATTATTT 51
Gorilla gorilla               GGTTT---------------------------------------TATTATTT 51
Pongo pygmaeus                GGTTT---------------------------------------TATTATTT 51
Homo sapiens                  GGTTT---------------------------------------TATTATTT 51
Macaca mulatta                GGTTT---------------------------------------TATTATTT 51
Callithrix jacchus            GGTCT---------------------------------------TA---TTT 48
Echinops telfairi             GATTT---------------------------------------TTTCATTC 64
Cavia porcellus               ----------------------------------------CATTTTAT 8
Tursiops truncatus            ----------------------------------------CCTTTTCT 8
Oryctolagus cuniculus         CATTTCCATATTTTGCCCATCGGGGCTATGGATACCTGGTTTTATTATTT 51
Erinaceus europaeus           GGTTT-----GAT-TATTAACATATT-----GCCCA---ACGGGGCTGTG 74
Sorex araneus                 GACTTTTAGAGAT-TTTATATATATTTGGATGTACATGTACATGTATGTA 123
Felis catus                   GGTTTT-----AT-TATCATTTTTTTC----ACCCA---ACGGGGCTGTG 75
Bos taurus                    GGTTTTAGTGTGTATATAGATAGGTAGATAGATATATTATTATTGTTATT 88
Sus scrofa                    GGTTTTAATA-------------------------ATTGTTATT--TATT 61
Tupaia belangeri              GTTTTTTAATTTTTTTTTCACTTTTTG----TTTTTGGTTTCTTTTTTTC 84
Microcebus murinus            AGTTTCA-----------------------------------TTATTGTTT 54
Canis lupus familiaris        GGTTTTATTATTA-------------------------------TTAT 55
Ailuropoda melanoleuca        GGTTTCACTATTAGT------------------------AGTAGTATCAT 64
Vicugna pacos                 GGTTTTAATATTT-------------------------TTTGTTTTGTGT 124
Tarsius syrichta              ----------------------------------------AGTTGA 6
Myotis lucifugus              -------------------TGATTTAGTTCTCTATTTCCTCCACTTTTCT 31
Pteropus vampyrus             ----------------------------------------TTTTTTCT 8

                                     --->.#2.<---   -->.#2.<--
Mus musculus                  ATTCCATGCCCAACGGGGT-TTTGGATACTTA-CC--------TGCCTTT 94
Rattus norvegicus             ATTCCATGTCCAGCTGGGC-TTGGGATACTTAACC--------TGCCTTT 95
Ellobius lutescens            ACTCCGTGCCCAGTGGGGC-TGTGGATACTTACCCGCTTTTAATTCGTTT 122
Equus caballus                TTTTTTTGCCCAACGGGGC-TGTGGATACCTG-------------CCTTT 87
Pan troglodytes               TTTCTTTGCCCAACGGGGC-CGTGGATACCTG-------------CCTTT 87
Gorilla gorilla               TTTCTTTGCCCAACGGGGC-CGTGGATACCTG-------------CCTTT 87
Pongo pygmaeus                TTTCTTTGCCCAACGGGGC-CGTGGATACCTG-------------CCTTT 87
Homo sapiens                  TTTCTTTGCCCAACGGGGC-CGTGGATACCTG-------------CCTTT 87
Macaca mulatta                TTTCTTTGCCCAACGGGGC-TGTGGATACCTA-------------CCTTT 87
Callithrix jacchus            TTTCTTTGCCCAACGGGGC-CGTGGATACCTG-------------CCTTT 84
Echinops telfairi             TTCCTTAGCCCATCGGGGT-TGTGGATAGCTG-------------CCTTA 100
Cavia porcellus               ATATTTTGCCCATCGGGGC-CGTGGATACCTG-------------CTTTA 44
Tursiops truncatus            ATATTTTGCCCATCGGGGC-TGTGGATACCTG-------------GTTTT 44
Oryctolagus cuniculus         TTTCTTTGCCCAACAGGGT-TCTAGCTACCTG-------------TCTTA 87
Erinaceus europaeus           GATACCTGC-CTTATAGAT-------CACGAC-------------TATCT 103 ●
Sorex araneus                 CATATGTATGCATATACATACGTATGTATGTG-------------TATAT 160
Felis catus                   GATACCTGCGTTTTAATTC-------TTTTCT-------------TTTAT 105 ●
Bos taurus                    ATTTTTTACCCAACGGGGT-CATGGATACCTG-------------CCTTT 124
Sus scrofa                    --TTTTTGCCCAACGGGGC-CGTGGATACCTG-------------CCTTT 95
Tupaia belangeri              CCCCTTTGCCCAACGGGGC-CGTGGATACCTG-------------CCTCT 120
Microcebus murinus            TTTCTTTGCCCAACGGGGC-TGTGGATACCTG-------------CCTTT 90
Canis lupus familiaris        TATTATTGCCCAACGGGGC-TGTGGATACCTG-------------CCTTT 91
Ailuropoda melanoleuca        TATTTTTGCCCAACGGGGC-TGTGGATACCTG-------------CCTTT 100
Vicugna pacos                 ATTTTTTGCCCAACGGGGC-CATGGATACCTG-------------CTTTT 160
Tarsius syrichta              CATTTCTGCTTCTTGTAGT--TTAGTTTTCTA-------------TTCCA 41 ●
Myotis lucifugus              ATATTTTGCCCATCAGGGT-TGCGGATACCTG--------------ATTT 66
Pteropus vampyrus             ATATTTTGCCCATCGGGGC-CGCGGATACCTG--------------CTTT 43
```

```
                                                                  --->.#3.
Mus musculus                TCATTCTTTTTTTTTCTTATTATTTTTTTT---TCTAAACTTGCCCATCT 141
Rattus norvegicus           TAATCCTTTTTTCTTCTTCTTACTTTTCTTCT-TCTAAACTTGCCCATCT 144
Ellobius lutescens          TGGTTTTTCCCTCCTTCTCCTCTTCTTTTTCT-TCTAAATTTGCCCATCT 171
Equus caballus              TAATTTTTTTTTTTTTT--------------------AATTTGCCCATCG 117
Pan troglodytes             TAATTCTTTTTT-ATTC-------------------------GCCCATCG 111
Gorilla gorilla             TAATTCTTTTTT-ATTT-------------------------GCCCATCG 111
Pongo pygmaeus              TAATTCTTTTTT-ATTC-------------------------GCCCATCG 111
Homo sapiens                TAATTCTTTTTT-ATTC-------------------------GCCCATCG 111
Macaca mulatta              TAATTCCTTTTT-ATTC-------------------------GCCCATCG 111
Callithrix jacchus          TAATTCTTTTTTTATTT-------------------------GCCCATCG 109
Echinops telfairi           TTATATTTTTCCTTTTCA----------------------TCGCCCATCG 128
Cavia porcellus             AAT-------------TTA--------------------TTTGCCCATCG 61
Tursiops truncatus          AATATTGTTAT-----TTT--------------------TTTGCCCATCG 69
Oryctolagus cuniculus       ATTCCTTTTTTAAAGAAAA--------------------TTAGCCCACCG 117
Erinaceus europaeus         G-ATCATTATTCTTTT-----------------------TTTGTCCATCG 129
Sorex araneus               GCATTAATGTGTATCT-----------------------TTTGCCCATCG 187
Felis catus                 ---TTATTTTTTTTAA-----------------------TTTGCCCATCG 129
Bos taurus                  TATTTTATTTTTTTTTAA---------------------TTTGCCCATCG 153
Sus scrofa                  TAATTCTTTTTTTAAAGAC--------------------TTTGCCCATCG 125
Tupaia belangeri            TAACTTTTCTCTTTTTAA---------------------TTTGCCCATCG 149
Microcebus murinus          TAATTCTTTTTCTTTTTTAAA------------------ATTTGCCCATCG 123
Canis lupus familiaris      TAATTCTTTTTTTTTGTTTGTTTGTTTTT-------AAATTTGCCCATCG 134
Ailuropoda melanoleuca      TAATTCTTTTCTTTTCTTTTCTTTCTTTCTTTTTTAAAATTTGCCCATCG 150
Vicugna pacos               TAATTCTTTTTTTTTTTTTAA------------------TTGCCCATCG 190
Tarsius syrichta            TCACCCCTTTTTCTATAT---------------------TTTGCCCATCG 70
Myotis lucifugus            TATTTATTTACTTTAAAAAA-------------------TGTGCCCATCG 97
Pteropus vampyrus           TAATTTTTTTCCCCCTTAA--------------------GCCCATCG 70

                            <---  -->.#3.<--
Mus musculus                GGGCTGTGGATACCTGCT---TTTATTCTTTTTTTCTTCT---CCT---- 181
Rattus norvegicus           GGGTTGTGGATACCTGCT---TTTATTCTTTTTTTCTTCTTCTCCT---- 187
Ellobius lutescens          GGGCTGTGGATACCTGCT---CTTCTTTTTTT-----------CCT---- 203
Equus caballus              GGGCCACGGATACCTGCG---TTTATTTTTTTTTTCCC-----CCT---- 155
Pan troglodytes             GGGCCGCGGATACCTGCT---TTTTATTT--TTTTTTT-----CCT---- 147
Gorilla gorilla             GGGCCGCGGATACCTGCT---TTTTATTT--TTTTTTT------CCT---- 146
Pongo pygmaeus              GGGCCGCGGATACCTGCT---TTTTATTT--TTTTTTT----TCCT---- 148
Homo sapiens                GGGCCGCGGATACCTGCT---TTTTATTT--TTTTTT------CCT---- 146
Macaca mulatta              GGGCCGCGGATACCTGCT---TTTTATTT--TTTTTTT----TCCT---- 148
Callithrix jacchus          GGGCCGCGGATACCTGCT---TTTTATTA--TTTTTTT----TCCT---- 146
Echinops telfairi           GGGCTGCGGATACCTGCT---TTTAATTTCATTTTTTT----CCCT---- 167
Cavia porcellus             GGGCTGCGGATACCTGCC---TTTTTTCTTTTTTTC------CTCT---- 98
Tursiops truncatus          GGGCCGCGGATACCTGCT---TTTAATTTTTTTTTC------C-CT---- 105
Oryctolagus cuniculus       GGGCTGCGGATACCTGCT---TTTAATTTTTTTTTC------C--T---- 152
Erinaceus europaeus         GGGCCTCGGATACCTGTT---TTTATTTTTTCT-CC------CCCT---- 165
Sorex araneus               GGGCCATGGATACGTGCG---CTAAAAACTTTTATC------CCCT---- 224
Felis catus                 GGGCAGCGGATACCTGCT---TTTAATTTTTTTTTTTTTTCACCCT---- 172
Bos taurus                  GGGCCACGGATACCTGCT---TTTAATTTTTTTTTT----CCCCCT---- 192
Sus scrofa                  GGGCCGCGGATACCTGCT---TTTAATTTTTTTTT------CCCCT---- 162
Tupaia belangeri            GGGCCACGGATACCTGCT---TTTTTATTATTATTA----TTTTCT---- 188
Microcebus murinus          GGGCAGTGGATACCTGCT---TTTTATTTTTTTTTTTT--CCCCCT---- 164
Canis lupus familiaris      GGGCCACGGATACCTGCT---TTTATTTTTTTTT-C------CCCT---- 170
Ailuropoda melanoleuca      GGGCCGCGGATACCTGCT---TTAATTTTTTTTTTC------CCCT---- 187
Vicugna pacos               GGGCCGCGGATACCTGCT---TTTAATTTTTTTTCC------TCCT---- 227
Tarsius syrichta            GGGCTGCGGATACCTGGT---TTTATTATTATTTTTG------TCA---- 107
Myotis lucifugus            GGGCCGCGGATACCTGCT---TTTAATTTTTTTCCCC-----CCCT---- 135
Pteropus vampyrus           GGGCGCGGATACCTGCTGTGTCCCCTCTTCATCCCCAA-TCCCCTAACT 119

                                 --->.#4.<---  -->.#4.<--
Mus musculus                ---TAGCCCATCGGGGCCATGGATACCTGCTTTTTGTAAA-AAAAAAAAA 227
Rattus norvegicus           ---TAGCCCATCGGGGCCATGGATACCTGCTTTTTACCAA-AAAACGCCG 233
Ellobius lutescens          ---TAGCCCATCGGGGCCATGGATACCTGCCTTTTTTTAACAAGAAAACG 250
Equus caballus              ---CAGCCCATCGGGGCCTCGGATACCTGCTTTTTTCCTT-----AAATT 197
Pan troglodytes             ---TAGCCCATCGGGGTATCGGATACCTGCTGATTCCCTTCCCCTCTGAA 194
Gorilla gorilla             ---TAGCCCATCGGGGTATCGGATACCTGCTGATTCCCTTCCCCTCTGAA 193
Pongo pygmaeus              ---TAGCCCATCGGGGTATCGGATACCTGCTGATTCCCTTCCCCTCTGAA 195
Homo sapiens                ---TAGCCCATCGGGGTATCGGATACCTGCTGATTCCCTTCCCCTCTGAA 193
Macaca mulatta              ---TAGCCCATCGGGGCATCGGATACCTGCTGATTCCCTTCCCCTCTGAC 195
Callithrix jacchus          ---TAGCCCATCGGGGCATCGGATACCTGCTGATTCCCTTCCCCTCTGAC 193
Echinops telfairi           ---TAGCCCATCGGGGCCATGGATGCGTGC-GATCTCCTCTCCACCCGA- 212
Cavia porcellus             ---TAGGCCATCGGGGCCT-GGATACCTGCTG-GTACACTAACCCTCC-- 141
Tursiops truncatus          ---TAGCCCATCGGGGCCTCGGATACCTGCTGTGTACCCCCTCTCTCCCC 152
Oryctolagus cuniculus       ---TAGCCCATCGGGGCCTCGGATACCTGCAGTGCCCCCTTTTGCCCCCC 199
Erinaceus europaeus         ---TAGCCCATCGGGGCCTTGGATAGCTGCTGTCCACTTCCTCCCCCTCAA 212
Sorex araneus               ---TCGCCCATCCGGGCCTCGGATACCTGCAGTCCGCTCCCTAACCTTAT 271
Felis catus                 ---TAGCCCATCGGGGCCTCGGATACCTGCTGT--GTCTCCTTCCCTC-- 215
Bos taurus                  ---TAGCCCATCGGGGCCTCGGATACCTGCTGTGTACCCCCTCTCTCC-- 237
Sus scrofa                  ---TAGCCCATCGGGGCCTCGGATACCTGCTGTGT-CCCCCTCTTTCT-- 206
Tupaia belangeri            ---TAGCCCATCGGGGCCTTGGATACCTGCTGTGT-CTCCCCCTCCCT-- 232
Microcebus murinus          ---TAGCCCAT-GGGGCAGTGGATGTGT-CCACGCGCCCCCCCCCCCAAC 209
Canis lupus familiaris      ---TAGCCCATCGGGGCCTCGGATACCTGCTGTGCCCCCCCCC------- 210
Ailuropoda melanoleuca      ---TAGCCCATCGGGGCCTCGGATACCAGCTGTGTCCCCCCTCTCTCT-- 232
Vicugna pacos               ---TAGCCCATCGGGGCCTCGGATACCTGCTGTGGCCCCCTTTTCTCCCC 274
Tarsius syrichta            ---TTGCCCAACGGGGCTGTGGATACCTGCCTTATAATTATTATTATTTT 154
Myotis lucifugus            ---TAGCCCATCGGGGCCTCGGATAGCTGCTGTGTCCTCCTTTTCCCC-- 180
Pteropus vampyrus           CTGTAGCCCATCGGGGCCACGGATACCTGCTATTTTTTTTTTTTC--T-- 165
```

```
                                                  --->.#5.<---  -->.#5.<--
Mus musculus                    AAAAAAAAAAAAACCTTTCTCGGTCCATCGGGACCTCGGATACCTGCGTT 277
Rattus norvegicus               TA-------------TTTCTCGGTCCATCGGGACCTCGGATACCTGCGTT 270
Ellobius lutescens              TA-------------TTTCTTGGTCCATCGGGACCTCGGATACCTGCGTT 287
Equus caballus                  TA-------------TTTTTTGGCCCATCGGGGCCTCGGATACCTGCTTC 234
Pan troglodytes                 C--------CCCCAACACTCTGGCCCATCGGGGTGACGGATACCTGCTTT 236
Gorilla gorilla                 C--------CCCCAACACTCTGGCCCATCGGGGTGACGGATACCTGCTTT 235
Pongo pygmaeus                  C--------CCCCAACACTCTGGCCCATCGGGGTGACGGATACCTGCTTT 237
Homo sapiens                    C--------CCCCAACACTCTGGCCCATCGGGGTGACGGATATCTGCTTT 235
Macaca mulatta                  C--------CCCCAACACTCTGGCCCATCGGGGTGGCGGATACCTGCTTT 237
Callithrix jacchus              C--------CCCCAACACTCTGGCCCATCGGGGTGACGGATATCTGCTTT 235
Echinops telfairi               C--------CCATGGCCCACGAAGCAAC--AAATAGCAGTACTTTATTAC 252●
Cavia porcellus                 ------------ATTTT--CTGGGCCATCGGGGCAATGGATACCTGCTTT 177
Tursiops truncatus              A-----------ACCCT--CTGGCCCAACGGGGCAATGGATACCTGCCTT 189
Oryctolagus cuniculus           ------------AAATT--CTGGCCCATCGGGGCAACGGATACCTGTTTT 235
Erinaceus europaeus             A---AATTAACAAATTT--TTGACCCACCGGGGTAACGGATACCTGTTTT 257
Sorex araneus                   AC--AAACATTTAAATGAGCCGGCCCATCAGGGCAACGGATACCTGCTTT 319
Felis catus                     -------CCCTTAACCCT-CTGGCCCATCGGGGCAATGGATACCAGCTTT 257
Bos taurus                      ----------CTAACC----TGGCCCATCGGGGCAATGGATACCTGCCTC 273
Sus scrofa                      ----------CCAACCCC-TTGGCCCATCAGGGTAATGGATACCTGCTTT 245
Tupaia belangeri                ------------AAACTCTCTGGCCCATCGGGGCATAGGATACCAGCTTT 270
Microcebus murinus              T-------------CCTCTCTGGCTGATCGGGGCAACGGATACCTGGTTT 246
Canis lupus familiaris          ----------CCCAACTCCCTGGCCCATCGGGGCAATGGATACCTGCTTA 250
Ailuropoda melanoleuca          ---------CCCCAACCCCCTGGCCCATCGGGGCAATGGATACCTGCTTA 273
Vicugna pacos                   A-------------ACCTCCTGGCCCATCGGGGCAACGGATACCTGCTTT 311
Tarsius syrichta                C---------TTTTTGAAATTCACCCATCGGGGTCACGGATACCTGCTTT 195
Myotis lucifugus                -------------AAACCACTGGCCCACCGGGGCCTAGGATACCTGCTTT 217
Pteropus vampyrus               -------------ATTTTTCTTGCCCATCGGGGCCTCGGATACCTGCTTT 202

                                                                              ---
Mus musculus                    T------AGTCTTT--------------TTTTCCCAT----------GCC 297
Rattus norvegicus               T------AGT--TT--------------TTTTCCCAT----------GCC 288
Ellobius lutescens              T------AGT-------------------CTTTCCTT----------GCC 302
Equus caballus                  T------ATTTTTT--------------TTTTTCCTT----------GCC 254
Pan troglodytes                 T-----TAAAAATT-------------TTCTTTTTTT---------GGCC 259
Gorilla gorilla                 T-----TAAAAATT-------------TTCTTTTTTT---------GGCC 258
Pongo pygmaeus                  T-----TAAAAATT-------------TTCTTTTTT----------GGCC 259
Homo sapiens                    T-----TAAAAATT-------------TTCTTTTTTT---------GGCC 258
Macaca mulatta                  T-----TAAAAATT-------------TTGTTTTTTT---------GGCC 260
Callithrix jacchus              T-----TAAAAATT-------------TTCTTTTTTT---------GGCC 258
Echinops telfairi               C-----CATCATAT-------------ATATTTTTTT---------GGCC 275
Cavia porcellus                 TTTTTTTTTTTTTTTTTAATTATGTGTGTGTTTGTTTGTTTGTTTTGGCC 227
Tursiops truncatus              TTTTTTTTTTTTT--------------AAATTTATTT------TTTGGCC 219
Oryctolagus cuniculus           TGTGGTTTTTTTTTGTTTTGTTTT----GTTTTTGAAT-TGGCTTTTGGCC 280
Erinaceus europaeus             TTTTTTTAAAATGT--------------TTTTTTTTTTT-------GGTC 286
Sorex araneus                   TCTTTTCATATAAT--------------TGTTTTCTT---------GGTC 346
Felis catus                     A------AAAAAAG--------------TTCCTTTTT----------GGCC 278
Bos taurus                      TTTTTTAAATGTGT--------------TGTTTTTTTTTTTTTCCTTGCC 309
Sus scrofa                      TTTATTTAAAAAAAA--------------AATTTTTTTTG--------GCC 273
Tupaia belangeri                T-----AAAAATAC--------------AATAAAATAAAAAGATTTTGCC 301
Microcebus murinus              T-----TAAAAATC--------------TTTGTTTTT----------GGCC 268
Canis lupus familiaris          A------AAAAAA----------------TTACTTTT----------GGCC 269
Ailuropoda melanoleuca          A------AAAAAAA--------------TTTACTTTTT--------GGCC 295
Vicugna pacos                   TTTTAAAAAAAAAA----------------TTTTTTT----------GCC 335
Tarsius syrichta                T------ATTTTTA--------------TTTTTAATT----------AGCC 216
Myotis lucifugus                T------TTTTTTT--------------TTTCTTCTTTT----TATAGCC 243
Pteropus vampyrus               C------ACTATTT--------------TTTTTCTT-----------GCC 221

                                >.#6.<---  -->.#6.<--
Mus musculus                    CAACGGGGCCTCGGATACCTGCTGTTATTATTTTTT-----TTTCTTTTT 342
Rattus norvegicus               CAACGGGGCCTCGGATACCTGCT-TTA--ATTTTTT-----TTTCTTTTC 330
Ellobius lutescens              CATCGGGGCCTTGGATACCTGCT-TCA-----TTTT-----TTTTCCCTC 341
Equus caballus                  CATCGGGGCCTCGGATACCTGCTCTCA--------------TTTTTTTTT 290
Pan troglodytes                 CATCGGGGCTTCGGATACCTGCTTTTTTTTTTTTTT---------ATTTT 300
Gorilla gorilla                 CATCGGGGCTTCGGATACCTGCTTTTTTTTTTTTTTT---------ATTTT 299
Pongo pygmaeus                  CATCGGGGCTTCGGATACCTGCTTTTTTTTTTTTTTT--------ATTTT 301
Homo sapiens                    CATCGGGGCTTCGGATACCTGCTTTTTTTTTTTTTTA-----------TTT 297
Macaca mulatta                  CATCGGGGCTTCGGATACCTGCTTTTATTTTTTATT------------TT 298
Callithrix jacchus              CATCGGGGCTTCGGATACCTGCTTTTATTTTTTATT------------TT 296
Echinops telfairi               CACCCGGGCTGTGGATACCTGATTTTATTTTGTTTTTTTAAAAAAATGTT 325
Cavia porcellus                 CACCGGGGCCCTGGATACCTGCTT-----TA----------TTATTTTTT 262
Tursiops truncatus              CATCGGGGCCTTGGATACCTGCTT-----TGATTTT-----TCTTTTTTT 259
Oryctolagus cuniculus           CATCGGGGCCTCGGATACCTGCTTA----TATTTTT-----TTTTTAAAT 321
Erinaceus europaeus             CATCGGGACCCTGGATATATGATTT-----TATT--------AAAACTTT 323
Sorex araneus                   CATCGGGACCCTGGATACCTGCTCG-----GTTT--------TTTTTTTT 383
Felis catus                     CATCGGGGCCTCGGATACCTGCTTT-----TATTA-------TTTTTTTT 316
Bos taurus                      CATCGGGGCCTCGGATACCTGCTTT-----AATT--------TTTTTTTT 346
Sus scrofa                      CATCGGGGCCTCGGATACCTGCTTT-----AATT--------TTTTTTTT 310
Tupaia belangeri                CATCGGGGCCTCGGATACCTGCTTTTCAAAAATTAAAAAAAAAATTTTTT 351
Microcebus murinus              CATCGGGGCTTCGGATACCTGCTTT-----TATTGT--------TTTTTT 305
Canis lupus familiaris          CATCGGGGCTTCGGATACCTGCTTT----TATTTT--------TTTTTTC 307
Ailuropoda melanoleuca          CATCGGGGCCTCGGATACCTGCTTT----TTTTTT--------TTTTTTC 333
Vicugna pacos                   CATCGGGGCCTCGGATACCTGCTTT-----AATTTT--------TTTTTC 372
Tarsius syrichta                CATCGGGGCATCGGATACCTGCTGTGTCCCCCCCAC---------CAACC 257
Myotis lucifugus                CATCGGGGCTTCGGATACCTGCTTTAA------------------TTTTT 275
Pteropus vampyrus               CATCGGGGCCTCGGATACCTGCTTTAAA-------------TTTTTTTTT 258
```

```
                                  --->.#7.<---  -->.#7.<--
Mus musculus                      CTTTTGCCCATCGGGGCTGTGGATACCTGCTTTAAATTTTTTTTTTC--- 389
Rattus norvegicus                 CTTTTGCCCATCGGGGCTGTGGATACCTGCTTTAA-TTTTTTTTTTC--- 376
Ellobius lutescens                TTTTTGCCCATCGGGGCTGTGGAAACCTGCTTCA---TTTTTTTTTC--- 385
Equus caballus                    TCCTTGCCCATCGGGGCCACGGATACCTGCTTAGATTTTTTTTTTTC--- 337
Pan troglodytes                   TCCTTGCCCATCGGGGCCTCGGATACCTGCTTTAATTTTT------GTTT 344
Gorilla gorilla                   TCCTTGCCCATCGGGGCCTCGGATACCTGCTTTAATTTTT------GTTT 343
Pongo pygmaeus                    TCCTTGCCCATCGGGGCCTCGGATACCTGCTTTAATTTT-------GTTT 344
Homo sapiens                      TCCTTGCCCATCGGGGCCTCGGATACCTGCTTTAATTTTT------GTTT 341
Macaca mulatta                    TCCTTGCCCATCGGGGCCTCGGATACCTGCTTTAATTTTT------GTTT 342
Callithrix jacchus                TCCTTGCTCATCGGGGCCTCGGATACCTGCTTTAATTTT---------TT 337
Echinops telfairi                 CCCTTGCCCATCGGGGCCTCGGATACCGGCTCTGATTTTTTTCCCCGTTC 375
Cavia porcellus                   TCATTGCCCATCGGGGCCTCGGATACCTGCCCTAATTTTTTTTTT----- 307
Tursiops truncatus                CCTTTGCCCATCGGGGCCTCGGATACCTGCTTTAATTTTTTTTTTTCC--- 306
Oryctolagus cuniculus             TTTTTGCCCATCAGGGCCTCGGATACCTGCTCTGATTTTTTTTTTTTT--- 368
Erinaceus europaeus               CAACTGCCCATCGGGGCAT-GGATACCTGCTTAATTTTTGT-TTTTC--- 368
Sorex araneus                     TCCTTGCCCATCGGGGCCTCGGATACCTGCTCGCTTTTTTT-TTTTTTTC 432
Felis catus                       CCCTTGCCCATCGGGGCCTTGGATACCTGCTTTATTTATTTATTTTTTTT 366
Bos taurus                        --CTTGCCCATCGGGGCCTCGGATACCTGCTTTAAT---TTTTTTTC--- 388
Sus scrofa                        TCCTTGCCCATCGGGGCCTCGGATACCTGCTTTATTATTTTTTTTTC--- 357
Tupaia belangeri                  TTTTGGCCCATCGGGGCTTCGGATACCTGCTTTAATTTTTTTTTGTTGTT 401
Microcebus murinus                CCCTTGCCCATCGGGGCCTTGGATACCTGCTTTTATTATTTTTTTTT--- 352
Canis lupus familiaris            CC-TTGCCCATCGGGGCCTCGGATACCTGCTTTTAATTTTTTTT------ 350
Ailuropoda melanoleuca            CCCTTGCCCATCGGGGCCTCGGATACCTGCTTTAATTTTTTTTTT----- 378
Vicugna pacos                     CTCTTGCCCATCGGGGCCTCGGATACCTGCTTTAATTTTTTTTTT----- 417
Tarsius syrichta                  CTTTGGCCCATCGGGGCAACGGATACCNNNNNNNNNNNNNNNNNNNNNN 307
Myotis lucifugus                  TTCTTGCCCATCGGGGCCGCGGATACCTGCTTAGATTTTTTTTTTTT--- 322
Pteropus vampyrus                 TCCTTGCCCATCGGGGCCGCGGATACCTGCTTAGATTTTTTTTTTA---- 304

                                  --->.#8.                <---  -->.#8.<--
Mus musculus                      --ACGGCCCAACG----------------GGGCGCTTGGTGGATGGAAAT 421
Rattus norvegicus                 --ACGGCCCATCG----------------GGGCATTTGGTGGATGGAAAT 408
Ellobius lutescens                --CTTGTCCATCG----------------GGGCATTTGGTGGATGGATAT 417
Equus caballus                    --ATTGCCCATCG----------------GGGGTTTTTATGGATAGAAAA 369
Pan troglodytes                   TTCTGGCCCATCG----------------GGGCCGCGGATACCTGCTTTG 378
Gorilla gorilla                   TTCTGGCCCATCG----------------GGGCCGCGGATACCTGCTTTG 377
Pongo pygmaeus                    TTCTGGCCCATCG----------------GGGCCGCGGATACCTGCTTTG 378
Homo sapiens                      TTCTG-CCCATCG----------------GGGCCGCGGATACCTGCTTTG 374
Macaca mulatta                    TTCTGGCCCATCG----------------GGGCCGCGGATACCTGCTTTG 376
Callithrix jacchus                TTCTGGCCCATCG----------------GGGCCGCGGATACCTGCTTTG 371
Echinops telfairi                 TCTGGGCCCATCG----------------GGGCCGCGGATACCTGCTGTG 409
Cavia porcellus                   -TT--GCCCATCG----------------GGGCCTTGGATACCTGCTTTA 338
Tursiops truncatus                -TT--GCCCATCG----------------GGGCCGCGGATACCTGCTTAG 337
Oryctolagus cuniculus             -TCTGGCCCATCG----------------GGGCCCCGGATACCTGCTCTG 401
Erinaceus europaeus               --CTTGCCCATCG----------------GGGCCACGGATACCTGCTTAG 400
Sorex araneus                     CACTTGCCCATCG----------------GGGCCATGGAT---------- 456●
Felis catus                       CCTTTGCCCATCG----------------GGGCTGTGGATACCTGCTTAG 400
Bos taurus                        --CTTGCCCATCG----------------GGGCCGCGGATACCTGCTTAG 420
Sus scrofa                        --CTTGCCCATCG----------------GGGCCGTGGATACCTGCTTAG 389
Tupaia belangeri                  TTTTGGCCCATCG----------------GGGCCACGGATACCTGCTCTC 435
Microcebus murinus                TCCTTGCCCATCG----------------GGGCCTCGGATACCTGCTCTC 386
Canis lupus familiaris            -CCTTGCCCATCG----------------GGGCTGTGGATACCTGTTTAG 383
Ailuropoda melanoleuca            -CCTTGCCCATCG----------------GGGCTGTGGATACCTGTTTAG 411
Vicugna pacos                     -CCTTGCCCATCG----------------GGGCCGCGGATACCTGCTTAG 450
Tarsius syrichta                  NNNNNNNNNNNNN----------------NNNNNNNNNNNNNNNNNNNNN 341●
Myotis lucifugus                  -AATCGCTC--------------------GAAATTTTGTTGTTTTCTGTG 351●
Pteropus vampyrus                 -AATCGCCCATCGCGGTCTTTTATGCACGGAAAATTGGTTGTCATCAGTG 353●

                                                  --->.#9.<---
Mus musculus                      A---TGGTTTT-GTGAGTTATTGCACTACCTGGA---------------- 451●
Rattus norvegicus                 AA--TGGTTTT-GTGAGTTATTGAACT----------------------- 432●
Ellobius lutescens                ATGTTGGTTTT-GTGAGTTATTGCACTGCCTGGAATATCCATAACTTTTT 466●
Equus caballus                    TTGTTGGTTTTTGTGGTTCGTTGTACTATCTGGA---------------- 403●
Pan troglodytes                   ATTTTTTTTTT--TCATCGCCCATCGGTGCTTTTTATGGATGAAAAAATG 426●
Gorilla gorilla                   ATTTTTTTTTT--TCATCGCCCATCGGTGCTTTTTATGGATGAAAAAATG 425●
Pongo pygmaeus                    ATTTTTTTT----TCATCGCCCATCGGTGCTTTTTATGGATGAAAAAATG 424●
Homo sapiens                      ATTTTTTTTTT--TCATCGCCCATCGGTGCTTTTTATGGATGAAAAAATG 422●
Macaca mulatta                    ATTTTTTTTT---TCATCGCCCGTCGGTGCTTTTTATGGATGAAAAAATG 423●
Callithrix jacchus                ATTTT--------------------------------------------- 376●
Echinops telfairi                 ATTCTTGTTTT--------------------------------------- 420●
Cavia porcellus                   ACTTCTTTGAT--TTCTTG----TAGAAGCTCTTTAT------------- 369●
Tursiops truncatus                ATTTTTTTTTT--TCATCGCCCATCGGGGCTTTTTATGGATGGAAAAGTG 385●
Oryctolagus cuniculus             ATTTTTTTTTT--CCATCGCCCATCGGGGCCTTTCATGGATGAAAATGTG 449●
Erinaceus europaeus               ATTTTTTTTCC--ATCGTCCA----------------------------- 419●
Sorex araneus                     ------------------------------------------------   ●
Felis catus                       ATT----------------------------------------------- 403●
Bos taurus                        ATTTTTGTTTT--ACACCACCCATCGGGGCTTTATATGGTTGGAAAAGTG 468●
Sus scrofa                        ATTTTTTTTT---------------------------------------- 399●
Tupaia belangeri                  ATTTTTGTTTT--TATCTCAATCGCTCATCGGGGCTTTTTATGGATGAAA 483●
Microcebus murinus                ATTTTTTTTTT--CCATCGCCCATCGGGGCTTCCCATGGATGAAAGGCGT 434●
Canis lupus familiaris            ATTTTTTTTTT-CTCATTGCCCATCGGGGCCTTTTATGGATGGAAGTGTT 432●
Ailuropoda melanoleuca            ATTTTTTTTTTTTCTCATTGCCCATCGGGGCCTTTTATGGATGGAAGTGTT 461●
Vicugna pacos                     ATTTTTTTTTTCATCGCCCATCGGGGCTTTTTGATGGATGGAGAAGTGTT 500●
Tarsius syrichta                  NNNNNNNNNNNNNNNNNNN------------------------------- 360●
Myotis lucifugus                  GTTCGTTATACTATCTGGAATGTCTACAAATTTTTGCTGCTAATCTTTGG 401●
Pteropus vampyrus                 GTTCGTTGTACTATCTGATATGTCT----CTTTTCGCCGTTAA------- 392●
```

Figure A.1: Analysis of repeat A sequences in 27 mammals.

Sequence alignment of repeat A region in 27 mammalian species. Black circles mark sequences that were not considered *bona fide* repeat A units and were thus excluded from further analyses.

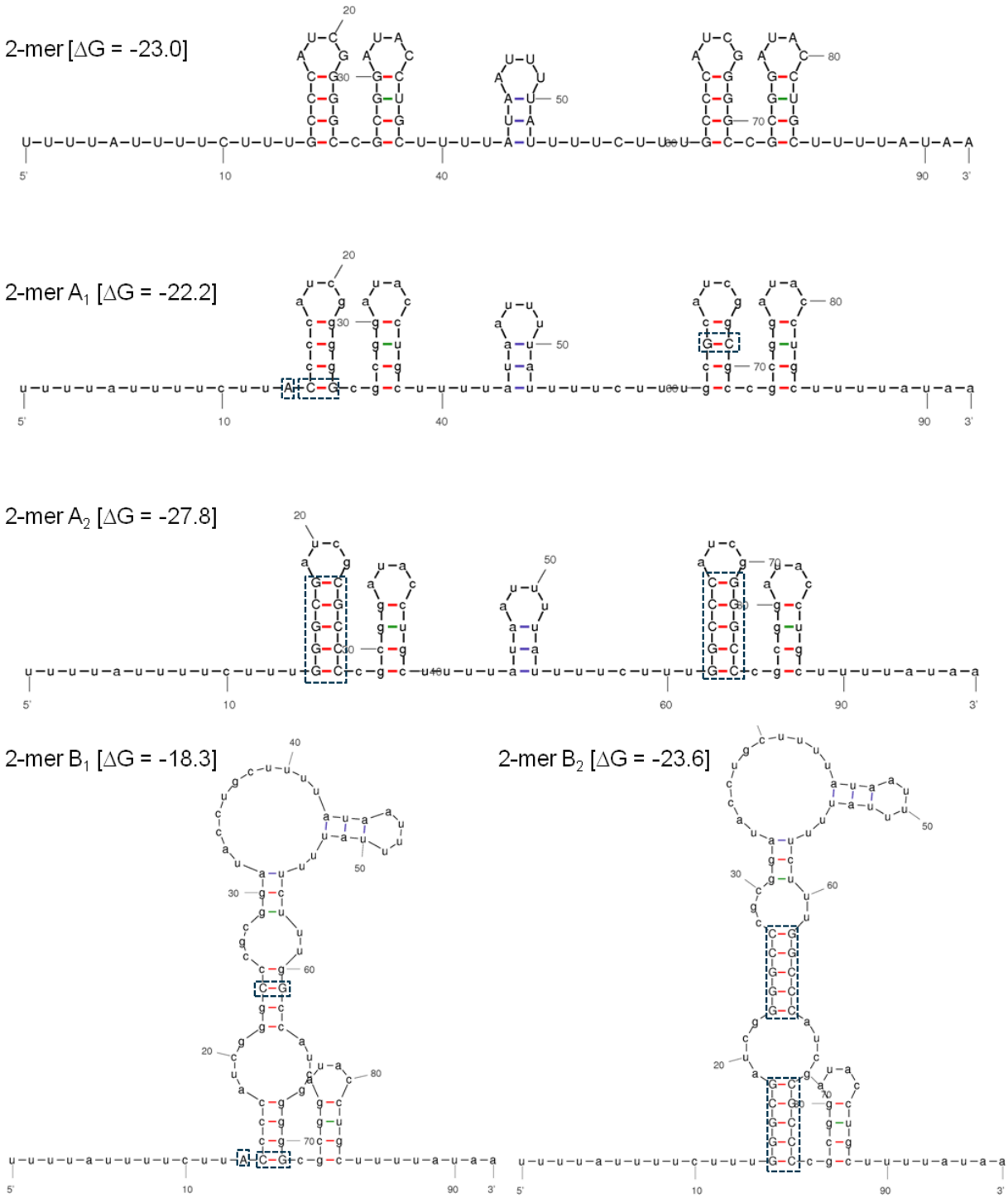


Figure A.2: *In silico* prediction of repeat A mutant structure.

Structures and free energies of 2-mer repeat A and its mutants created to enforce pairing within each monomer (A1, A2) or between the two monomers (B1, B2) predicted by mfold. Bases diverging from the canonical repeat A sequence are capitalized and highlighted. ΔG values represent minimal free energy for the structures shown [kcal/mol].


A
Flow cytometry EGFP expression D0 (replicate 1)
Flow cytometry EGFP expression D2 (replicate 1)
B
Flow cytometry EGFP expression D0 (replicate 2)
Flow cytometry EGFP expression D2 (replicate 2)
C
Flow cytometry DsRED expression D0
Flow cytometry DsRED expression D2
D
Flow cytometry EGFP expression D2/D0 (replicate 1)
Flow cytometry EGFP expression D2/D0 (replicate 2)
Flow cytometry DsRED expression D2/D0
(1) BRRN1
(1) SMC4L1
(1) CNAP1
(1) AOF2
(1) PARP1
(1) ASH2L
(1) SDC1
(1) M96
(4) CUL3
(2) SPOP
(2) ATM
(2) ATRX
(2) SMCHD1
(2) BRCA1
(2) H2AFY
(2) H2AFY2
(3) HNRPU
(3) SATB1
(3) SATB2
(3) SET7
(3) CARM1
(3) RNF2
(2) YY1
(3) CTCF
(3) CBX4
(4) CBX7
(4) HNRPK
(4) EHMT2
(4) EHMT1
(4) REST
(4) RCOR1
(4) DICER1
lipo B1R1
lipo B1R1-2
lipo B2R1
lipo B2R1-2
lipo B3R1
lipo B3R1-2
lipo B4R1
lipo B4R1-2

Figure A.3: siRNA screen - raw data; flow cytometry analysis of reporter gene expression.

(A-C) Flow cytometry analysis of *EGFP* (A, B) or DsRED Express2 (C) expression in cells expressing *XIST* for 2 days (D2) versus in cells not expressing *XIST* (D0). The values represent the mean amount of fluorescence, in arbitrary units. The screen was performed in four batches and samples within each batch are identified by the numbers 1–4 in parentheses. Samples labeled 'lipo' were treated with transfection reagent but not with siRNA. B1-B4 refers to batches 1–4. R1 denotes the first replicate of the screen. Each batch contained two pairs of transfection reagent-treated control cells (*e. g.* 'lipo B1R1' and 'lipo B1R1-2').

(D) The data from panels (A–C) are shown as a ratio of fluorescent reporter expression after XIST-induced silencing (D2) compared to the cells that were not expressing *XIST* (D0).

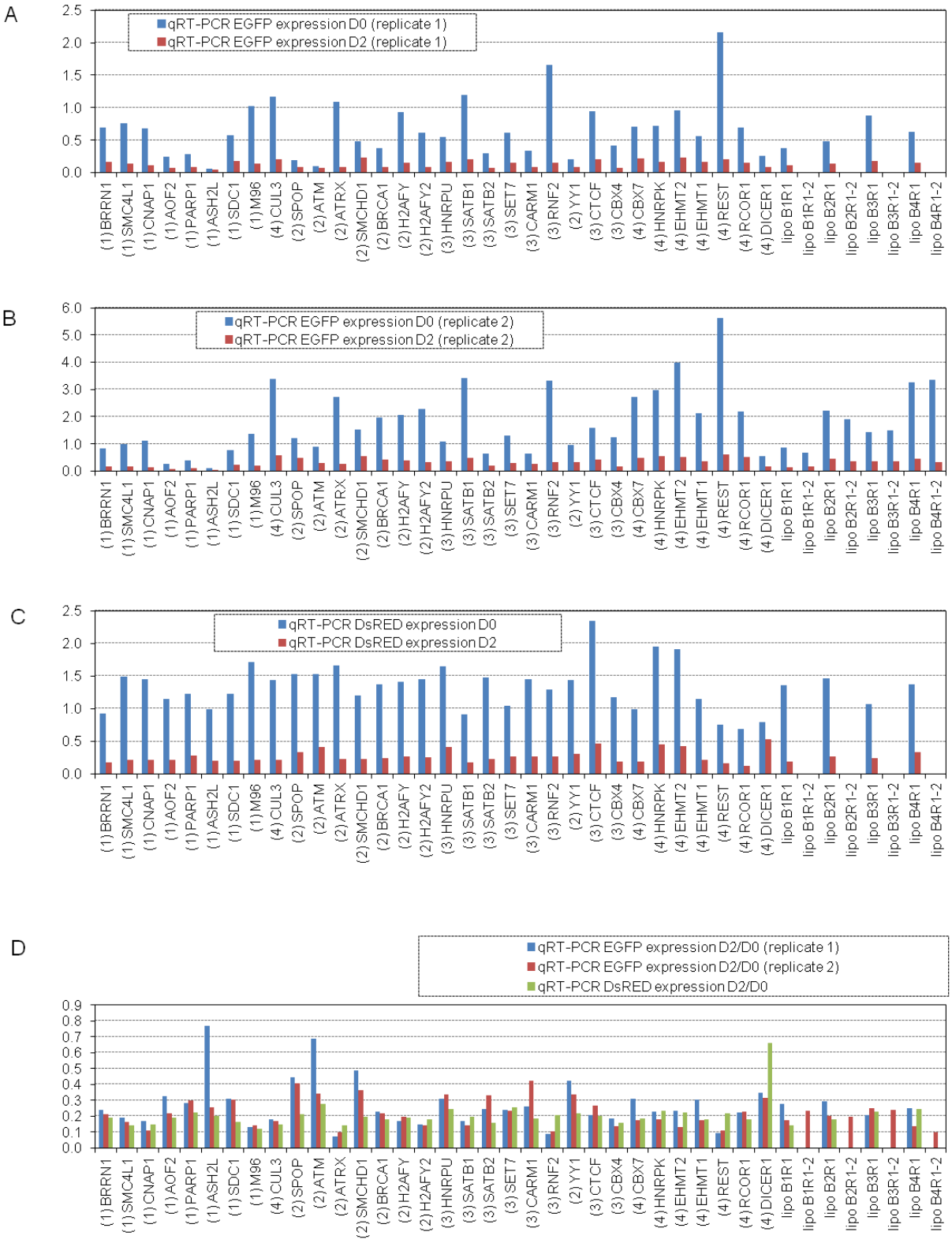

A
qRT-PCR EGFP expression D0 (replicate 1)
qRT-PCR EGFP expression D2 (replicate 1)
B
qRT-PCR EGFP expression D0 (replicate 2)
qRT-PCR EGFP expression D2 (replicate 2)
C
qRT-PCR DsRED expression D0
qRT-PCR DsRED expression D2
D
qRT-PCR EGFP expression D2/D0 (replicate 1)
qRT-PCR EGFP expression D2/D0 (replicate 2)
qRT-PCR DsRED expression D2/D0
(1) BRRN1
(1) SMC4L1
(1) CNAP1
(1) AOF2
(1) PARP1
(1) ASH2L
(1) SDC1
(1) M96
(4) CUL3
(2) SPOP
(2) ATM
(2) ATRX
(2) SMCHD1
(2) BRCA1
(2) H2AFY
(2) H2AFY2
(3) HNRPU
(3) SATB1
(3) SATB2
(3) SET7
(3) CARM1
(3) RNF2
(2) YY1
(3) CTCF
(3) CBX4
(4) CBX7
(4) HNRPK
(4) EHMT2
(4) EHMT1
(4) REST
(4) RCOR1
(4) DICER1
lipo B1R1
lipo B1R1-2
lipo B2R1
lipo B2R1-2
lipo B3R1
lipo B3R1-2
lipo B4R1
lipo B4R1-2

Figure A.4: siRNA screen - raw data; qRT-PCR analysis of reporter gene expression.

(A-C) qRT-PCR analysis of *EGFP* (A, B) or DsRED Express2 (C) expression in cells expressing *XIST* for 2 days (D2) versus in cells not expressing *XIST* (D0). All data are normalized to *ACTB* expression and in arbitrary units. The screen was performed in four batches and samples within each batch are identified by the numbers 1–4 in parentheses. Samples labeled 'lipo' were treated with transfection reagent but not with siRNA. B1–B4 refers to batches 1–4. R1 denotes the first replicate of the screen. Each batch contained two pairs of transfection reagent-treated control cells, *e. g.* 'lipo B1R1' and 'lipo B1R1-2'.

(D) The data from panels (A–C) are shown as a ratio of fluorescent reporter expression after *XIST*-induced silencing (D2) compared to the cells that were not expressing *XIST* (D0).

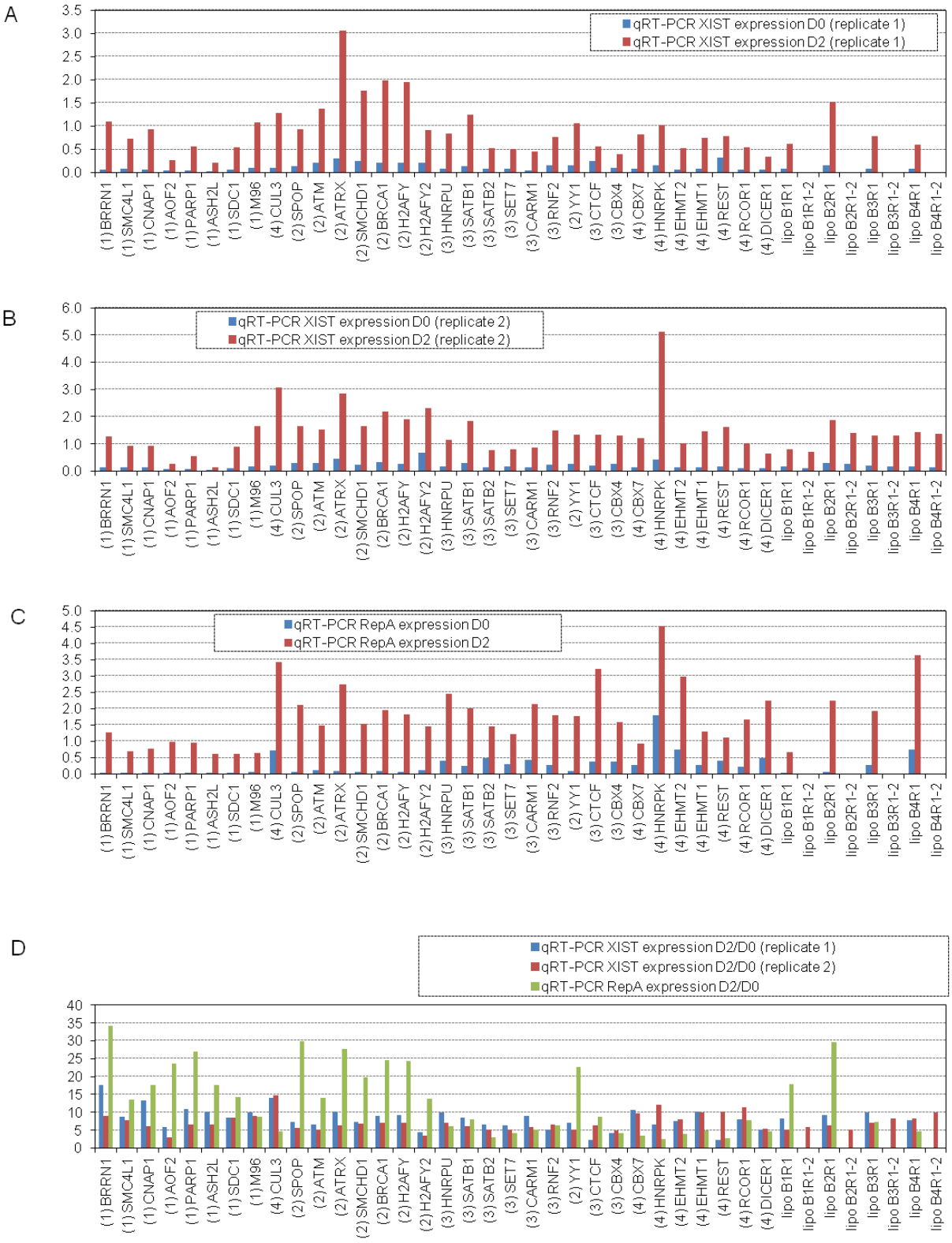

A
qRT-PCR XIST expression D0 (replicate 1)
qRT-PCR XIST expression D2 (replicate 1)
B
qRT-PCR XIST expression D0 (replicate 2)
qRT-PCR XIST expression D2 (replicate 2)
C
qRT-PCR RepA expression D0
qRT-PCR RepA expression D2
D
qRT-PCR XIST expression D2/D0 (replicate 1)
qRT-PCR XIST expression D2/D0 (replicate 2)
qRT-PCR RepA expression D2/D0
(1) BRRN1
(1) SMC4L1
(1) CNAP1
(1) AOF2
(1) PARP1
(1) ASH2L
(1) SDC1
(1) M96
(4) CUL3
(2) SPOP
(2) ATM
(2) ATRX
(2) SMCHD1
(2) BRCA1
(2) H2AFY
(2) H2AFY2
(3) HNRPU
(3) SATB1
(3) SATB2
(3) SET7
(3) CARM1
(3) RNF2
(2) YY1
(3) CTCF
(3) CBX4
(4) CBX7
(4) HNRPK
(4) EHMT2
(4) EHMT1
(4) REST
(4) RCOR1
(4) DICER1
lipo B1R1
lipo B1R1-2
lipo B2R1
lipo B2R1-2
lipo B3R1
lipo B3R1-2
lipo B4R1
lipo B4R1-2

Figure A.5: siRNA screen - raw data; qRT-PCR analysis of *XIST* and repeat A expression.

(A-C) qRT-PCR analysis of *XIST* (A, B) or repeat A (C) expression in cells treated with DOX for two days (D2) versus in cells not treated with DOX (D0) is shown. All data are normalized to *ACTB* expression and in arbitrary units. The screen was performed in four batches and samples within each batch are identified by the numbers 1–4 in parentheses. Samples labeled 'lipo' were treated with transfection reagent but not with siRNA. B1–B4 refers to batches 1–4. R1 denotes the first replicate of the screen.

(D) The data from panels (A–C) are shown as a ratio of *XIST* or repeat A induction in DOX treated cells (D2) versus in cells not treated with DOX (D0).

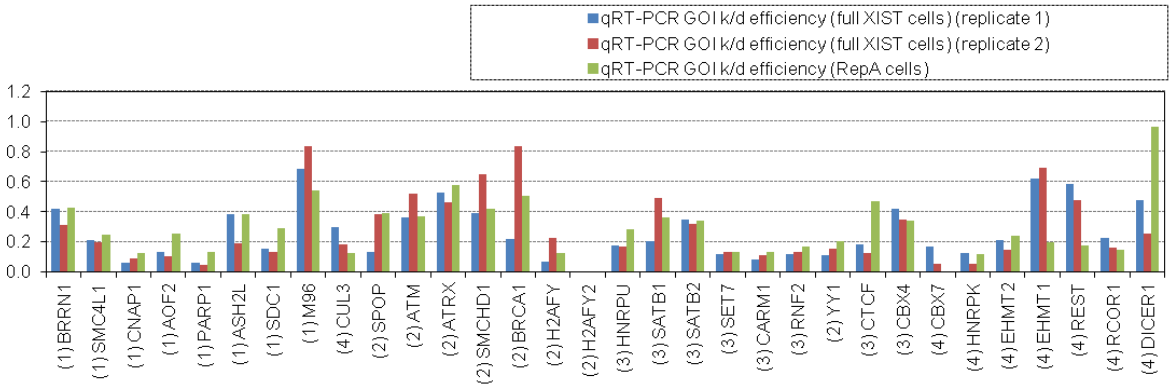


Figure A.6: siRNA screen - raw data; qRT-PCR analysis of mRNA knock-down efficiency.

The efficiency of siRNA-mediated mRNA knock-down measured by qRT-PCR is shown for the full-*XIST* cDNA cell line or the repeat A - DsRED express2 cell line. The expression of each gene in the cells treated with the respective siRNA is shown relative to expression in the transfection reagent-treated control cells and normalized to *ACTB*. *H2AFY2* was not expressed in the HT1080 cells.