Two ‘Detailed views’ on X chromosome inactivation

Hong, Ru

2016

Hong, R. (2016). Two ‘Detailed views’ on X chromosome inactivation. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/69087>

<https://doi.org/10.32657/10356/69087>

“Two ‘Detailed Views’ on X Chromosome Inactivation”

Hong Ru

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy of Biological Sciences

2016

Acknowledgement

I'm heartily thankful to my supervisor, Assistant Professor Li-Feng Zhang, for his valuable guidance and advice during this project. Also, I would like to thank my co-supervisor, Assistant Professor Xin Chen from SPMS (School of physics and mathematics science) for his help in the computation part of the project and the other members of my Thesis Advisory Committee, Professor Daniela Rhodes, and Adjunct Professor Huck Hui Ng, for their valuable suggestions in my TAC meeting, Assistant Professor Sara Sandin and Dr. Alexander Ludwig, for their help on the EM work. Finally, I want to thank all my lab mates and my family for their help and support.

Table of contents

Acknowledgement	1
Table of contents	3
Abstract.....	7
List of figures	8
Abbreviation.....	12
1 Introduction	15
1.1 Epigenetic modifications.....	15
1.2 X chromosome inactivation	19
1.3 Formation of the <i>Xist</i> cloud and the <i>LINE-1</i> hypothesis.....	23
1.4 X-linked genes which escape XCI	30
1.5 Sporadic gene reactivation along the Xi after conditional deletion of <i>Xist</i> gene in the female fibroblasts	31
1.6 Padlock probe SNP capture: a high throughput digital RNA allelotyping method.....	33
1.7 Non-coding DNA.....	36
1.8 Observation of the Xi through EM with conventional methods.....	39
1.9 The miniSOG labeling method for EM	47
1.10 macroH2A	48
2 Objective	52
3 Materials and Methods.....	53
3.1 Cell Culture:.....	53
3.2 Lipofectamine Transfection:	54
3.3 DNA Constructs	54

3.4	Total RNA isolation:.....	56
3.5	Reverse transcription:.....	57
3.6	Gateway Recombination:.....	58
3.7	Photo-oxidation:	63
3.8	Genomic DNA Extraction:	64
3.9	Colony PCR	65
3.10	RNA FISH:	66
3.11	Amplification of padlock probe library:	67
3.12	Padlock probe SNP capture:.....	69
3.13	KAPA Quantification.....	72
3.14	RNA Pol III inhibition	73
3.15	Isolation of bromouridine (BRU) labelled nascent RNA and reverse transcription.....	73
3.16	Immunostaining	77
3.17	Chromosome conformation capture carbon copy (5C)	78
3.18	shRNA knockdown	79
3.19	Allele specific RT-PCR and southern blotting.....	80
3.20	Data analysis	82
4	Results	86
4.1	Generation of mutant fibroblast cell lines with allele-specific deletion of <i>Xist</i> on the Xi	86
4.2	The design of individual padlock probes.....	94
4.3	The design of the padlock probe library	97
4.4	Padlock probe library synthesis and amplification	102
4.5	SNP capture and sequencing samples preparation	106

4.6	Reactivation pattern study.....	112
4.7	SINEs are enriched along the upstream regions of the reactivated genes in comparison to the genes remain silenced.....	123
4.8	Gene reactivation pattern in long term cultured cell lines.....	127
4.9	Non-coding transcription study on XCI	129
4.10	RNA Polymerase III transcription shows chromatin restructuring function.....	141
4.11	Pol III inhibition impairs XCI during in vitro differentiation	158
4.12	Labeling the Xi in live cells using miniSOG and fluorescent protein	162
5	Conclusion and Discussion	171
6	Reference	178

Abstract

Xist (X-inactive specific transcript), a prototype long non-coding RNA, is essential for establishing X chromosome inactivation (XCI) in early embryos. However, its role in maintaining XCI in soma is unclear. When *Xist* was conditionally deleted in somatic cells from the inactive X ($\text{Xi}^{\Delta\text{Xist}}$), only rare and sporadic gene reactivation could be detected. We carried out high-resolution RNA allelotyping by padlock SNP capture on a panel of wild type and mutant mouse cell lines. We confirm that low level (~4%) gene reactivation occurs along $\text{Xi}^{\Delta\text{Xist}}$, but the pattern is not sporadic. A gene with its upstream region enriched for SINEs (a class of retrotransposon transcribed by RNA polymerase III) is prone to be reactivated. Intriguingly, Pol III transcription from non-coding regions is largely independent of *Xist*-mediated transcriptional silencing. Pol III inhibition alters chromatin looping and impairs XCI suggesting a role of Pol III transcription in regulating chromatin architecture.

In the second project of my graduate research, plasmid fusion constructs of mH2A and miniSOG were established. miniSOG catalyzes polymerization of diaminobenzidine (DAB) which is osmiophilic. Therefore miniSOG can be used to label protein targets in live cells for downstream electron microscope (EM) studies. macroH2A (mH2A) is a histone variant enriched on Xi. One mH2A-miniSOG fusion construct successfully labeled Xi for fluorescent microscopy and photo-oxidation (DAB polymerization). It is ready to be used as the EM tag for future studies on the chromatin structures of the Barr body.

Table of figures

Figure 1 Model for chromosome territory re-organization during XCI adopted from (Chaumeil, Le Baccon et al. 2006).....	26
Figure 2 An EM picture of immunogold labeling of endogenous LRRK2 from NIH-3T3 cells.	46
Figure 3 An EM picture of H3K27me3 immunogold labeling of an inactive X chromosome located at the edge of a nucleus.	46
Figure 4 Comparison of H2A and macroH2A peptides.	51
Figure 5 Experimental designs of generating female fibroblast cell lines in which <i>Xist</i> is conditionally and allele-specifically deleted from the Xi.	90
Figure 6 PCR genotyping results of the established $X^{WT}X^{\Delta Xist}$ cell lines.....	91
Figure 7 $Xa^{WT}Xi^{\Delta Xist}$ cell lines identified by <i>Xist</i> RNA FISH.	92
Figure 8 Allele-specific RT-PCR on <i>Xist</i> RNA of the 6 clones generated by limiting dilution of 2loxT cells (2loxTS1-S6). The genomic DNA of castaneous genetic background was used as the control.....	93
Figure 9 The principle of SNP capture using Padlock probes.	96
Figure 10 The established padlock probe library.	101
Figure 11 Schematic diagram of padlock probe library preparation process.	104
Figure 12 Amplification curves obtained from the heterogeneous padlock probe oligo library in a test PCR run.....	104
Figure 13 Amplification curves obtained from the heterogeneous padlock probe oligo library in a production PCR run.....	105
Figure 14 Gel analysis of single-stranded padlock probes after double nikase digestions.	105
Figure 15 Schematic diagram of SNP capture process, followed by amplification using solexa sequencing primers.	109
Figure 16 Amplification of a padlock probe SNP capture products for allelotyping X-linked genes.	110

Figure 17 Gel analysis of padlock probe SNP capture amplicons using sso7d fusion polymerase.	111
Figure 18 Eleven X-linked genes which escape XCI.	114
Figure 19 Reactivated genes detected in 1lox cell lines.	117
Figure 20 The summary of allelotyping read counts of the gene reactivation study.	117
Figure 21 The 25 frequently reactivated genes.	118
Figure 22 The alignment of reactivated genes with padlock probes targeting genes and other genomic features along mouse X chromosome.	121
Figure 23 The spearman coefficient values between reactivated genes and different genomic features.	122
Figure 24 Distribution of SINE elements along the X-linked genes.	125
Figure 25 Distribution of sense/antisense SINE elements along the X-linked genes.	126
Figure 26 The distribution of gene reactivation of HR5, HR6, HR5.D25 and HR6.D25.	128
Figure 27 The summary of allelotyping read counts of the non-coding transcription study.	130
Figure 28 The calculations and thresholds used to define allelotype and expression pattern.	133
Figure 29 The definitions of allelotype and expression pattern, which are used to describe the allele-specific expression profile of X-linked non-coding SNPs.	134
Figure 30 The non-coding SNPs and the X-linked genes allele-specific expression profile of “129” allele to “Castaneous” allele in Xi ¹²⁹ Xa ^{Cast} cells.	135
Figure 31 A venn diagram showing the expression patterns between the four combos of data sets described in Fig.29.	136
Figure 32 Non-coding SNPs with <i>Xist</i> -dependent expression pattern from the data analysis of four different combinations of data sets.	136

Figure 33 The 85 non-coding region SNPs with <i>Xist</i> -independent expression pattern.	137
Figure 34 A different threshold used in analyzing transcription from non-coding regions.	139
Figure 35 Data analysis result using the threshold shown in Fig.34.	140
Figure 36 The distribution of the detectable <i>Xist</i> -independent non-coding regions transcripts, aligned with <i>Xist</i> RNA coating and Alu density.	142
Figure 37 The non-coding transcriptions are Pol III products.....	143
Figure 38 A model of <i>Xist</i> -independent Pol III transcription from “ihoc” nucleosomes/regions regulating the epigenetic architecture of the Xi.	145
Figure 39 Allele-specific RT-PCR on Med14 (uc009sqz.1), a reactivated gene along the Xi ^{129-Δ<i>Xist</i>} allele in HR7 cells.....	145
Figure 40 Co-localization of RNA polymerases and <i>Xist</i> RNA using immuno-RNA FISH.	147
Figure 41 Chromatin IP to check the enrichment of H3K27me3 along the Xi in 2loxT.S2 cells with and without the Pol III inhibitor treatment.....	150
Figure 42 Enrichment of H3K27me3 along the Xi in 2loxT.S1 cells arrested at G1/S boundary before Pol III inhibition.	151
Figure 43 Heatmaps of ChIP-Seq data showing H3K27me3 enrichment along the X chromosome.	154
Figure 44 The alignment of local-viewed ChIP-Seq peaks distribution and SINE density.	155
Figure 45 Pol III inhibition causes chromatin conformation change.....	157
Figure 46 Pol III inhibition impairs XCI in vitro.....	160
Figure 47 Distribution of SINE elements along the X-linked genes.....	161
Figure 48 The labeling of Xi with mH2A-SOG and H2AD-SOG.....	167
Figure 49 The transient transfection result of 129T mouse fibroblast cells with two plasmids, pminiSOG-mCherry-C1 and ptdminiSOG-N1.	167
Figure 50 The labeling of Xi with H2AD-mc, H2AD-SOG-mc and H2AD-tdSOG.	168

Figure 51 The labeling of Xi with mH2A-mc-SOG, H2AD-mc-SOG and mH2A-SOG-mc.	169
Figure 52 A photo-oxidation effect of a mH2A-miniSOG-mCherry labeled cell which is a transient transfected 129T mouse fibroblast cell with p-mH2A-SOG-mc.	170
Figure 53 A photo-oxidation effect of a mH2A-miniSOG-mCherry labeled cell which is a transient transfected tetraploid 293T human fibroblast cell with p-mH2A-SOG-mc.....	170

Abbreviation

Å	Angstrom
129	Mouse strain 129
µg	Microgram
µl	Microliter
Amp	Ampicillin
ATG	above threshold genotype
BSA	Bovine serum albumin
C57Bl	Mouse strain C57 Black
Cast	Mouse strain Castaneus
CDK8	Cyclin-dependent protein kinase 8
DAPI	4', 6-diamidino-2-phenylindole
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle Medium
E5.5	5.5days post-conception embryos
EM	Electron microscopy
ESC	Embryonic stem cell
FISH	Fluorescence <i>in situ</i> hybridization
Fmol	Femto Molar
H2AD	The H2A domain of macroH2A
H2AK119ub	Histone H2A lysine 119 ubiquitination
H3K4me3	Histone H3 tri-methylated on lysine 4
H3K27me3	Histone H3 lysine 27 trimethylation

H4Ac	Acetylated Histone H4
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPRT	Hypoxanthine-guanine phosphoribosyltransferase
hr	Hour
ICM	Inner cell mass
<i>LINE1</i>	Long interspersed nuclear element 1
MCB	Macrochromatin body
MEF	Mouse embryonic fibroblast
mH2A	macroH2A
min	minute
miniSOG	mini Singlet Oxygen Generator
miRNA	Micro Ribonucleic acid
mmol	Milli Molar
ncRNA	non-coding RNA
ng	Nanogram
nt	Nucleotide
°C	Degree centigrade
<i>Oct4</i>	Octamer-binding transcription factor 4
PAR	PolyADP ribosylation
<i>PcG</i>	Polycomb group
Pmol	Pico Molar
<i>PRC1</i>	Polycomb repressive complex 1
<i>Rex1</i>	Reduced expression protein 1

RNase	Ribonuclease
<i>Rnf12</i>	RING finger protein 12
rpm	Revolution per minute
Saf-A	Scaffold attachment factor A
<i>SATB1</i>	Special AT-rich binding protein
SNP	Single-nucleotide polymorphism
snRNA	Small nucleolar RNA
<i>Sox2</i>	Sex-determining region Y-box 2
Xa	Active X chromosome
<i>Xce</i>	X-controlling element
XCI	X chromosome inactivation
XCR	X chromosome reactivation
Xi	Inactive X chromosome
<i>Xic</i>	X chromosome inactivation center
<i>Xist</i>	X-inactive specific transcript
YY1	Yin-Yang1

1 Introduction

1.1 Epigenetic modifications

The essential role of epigenetic mechanism can be seen from the development of an independent organism from a single cell embryo. During this process, different types of cells precisely express a specific subset of genes, even though a complete and identical genome is carried in each cell. Epigenetic mechanism is one important regulatory mechanism of selective gene expression. Epigenetics indicates heritable changes in gene expression or cellular phenotype through mechanisms unrelated to DNA sequence alteration. It refers to a series of functional modifications to the genome, including DNA methylation, histone modifications, histone variants, noncoding RNAs and higher order of chromatin structures. Mounting evidence has demonstrated that epigenetic modifications regulate and/or “annotate” gene expression activities.

DNA methylation takes place mainly in the CpG islands of promoter regions, through converting cytosine into 5-methylcytosine. In general, a

hypermethylated promoter is an inactive promoter. In DNA replication, inheritance of methylation in cytosine to the two daughter DNA molecules is “inherited” because of the methylation function of DNMT1 is in accordance with the parent DNA molecules (Robertson, Uzvolgyi et al. (1999).

Histone modifications are another type of epigenetic modifications. The histone modifications usually happen in the N'-terminus of histone peptides. These modifications include methylation, acetylation, citrullination, ubiquitylation, phosphorylation, ADP-ribosylation and sumoylation.

Since histone acetylation is found to associate with initiation of transcription, it becomes an intensively studied area. Most histone acetylation occurs on lysine residues and is associated with actively transcribed DNA regions (euchromatin). It is believed that comparing to the positively charged unmodified lysine residue, which can bind tightly to the negatively charged DNA phosphate backbone, acetylation of lysine neutralizes its positive charge and consequently relaxes the compactness

of chromatin fibers to create a friendly environment for transcription.

Additional evidence also shows that histone acetylation facilitates transcription by forming binding sites for other chromatin modification enzymes (Tse, Sera et al. 1998). Furthermore, many transcriptional activators have the enzyme activity of histone acetyltransferase (Brownell and Allis 1996, Brownell, Zhou et al. 1996, Wolffe and Pruss 1996), while transcriptional repressors recruit the histone deacetylase (Taunton, Hassig et al. 1996, Pazin and Kadonaga 1997, Wolffe 1997).

Histone methylation is another important histone modification.

Similar to histone acetylation, methylation of histone also frequently occurs on lysine residues. Different from histone acetylation, different histone methylation plays different roles in regulating gene expression.

The methylation of lysine 9 on histone H3 (H3K9me_{2/3}) is associated with constitutive heterochromatin, for example the centromeres. H3K27me₃ is associated with facultative heterochromatin, for example the inactive X chromosome (Xi). H3K4me is associated with euchromatin ((Strahl, Ohba

et al. 1999, Bernstein, Humphrey et al. 2002, Schneider, Bannister et al. 2004, van Dijk, Marley et al. 2005).

Non-coding RNA (ncRNA) refers to the RNA molecules, which are not to be translated into proteins. Commonly found ncRNAs include tRNA, rRNA, snoRNA and microRNAs (Washietl, Pedersen et al. 2007). In current research, more and more long ncRNAs, called lincRNAs (long intergenic ncRNA), are identified as components involved in epigenetic mechanisms (Denisenko, Shnyreva et al. 1998, Katayama, Tomaru et al. 2005, Sanchez-Elsner, Gou et al. 2006, Mazo, Hodgson et al. 2007, Rinn, Kertesz et al. 2007). One of the best known lincRNA is the *Xist* RNA in X chromosome inactivation. The inactivated X chromosome (Xi) carried by each female somatic cell of mammals is an intriguing piece of chromatin heavily decorated by epigenetic modifications. Besides the *Xist* RNA, Xi carries all the foregoing epigenetic modifications including histone variants and higher order of chromatin structures.

1.2 X chromosome inactivation

In a female mammalian somatic cell, one X chromosome is transcriptionally silenced. This phenomenon, called X chromosome inactivation (XCI), is the strategy that mammals use to achieve the X-linked gene dosage compensation between males (XY) and females (XX). Since the two Xs in each female cell cannot be distinguished by their DNA sequences, the opposite transcription status of the two identical chromosomes are solely controlled by epigenetic mechanisms. Therefore, XCI is considered as an absolute epigenetic event.

XCI occurs in the early embryonic development. In the mouse, the two Xs in each female zygote carry special epigenetic modifications that are pre-deposited by the parents onto the chromosomes during gametogenesis. These epigenetic modifications, called the genetic imprints, determine that the paternal X chromosome is specifically inactivated in the early cleavage stage embryos. This form of XCI is called imprinted XCI (Singer-Sam, Chapman et al. 1992, Kay, Penny et al. 1993). When the embryo develops into a blastocyst (E3.5-4.5), the cells in the

inner cell mass (ICM) undergo X chromosome reactivation (XCR), which results in two active X (Xa) chromosomes per cell. The mouse embryonic stem cells (ESCs) isolated from the inner cell mass at this stage also carry two Xa per cell (Hajkova, Erhardt et al. 2002). Meanwhile, the imprinted XCI continues in the cell lineages of extraembryonic tissues (the trophoctoderm lineage and primitive endoderm lineage) (West, Frels et al. 1977). As the embryonic development continues, implantation and gastrulation occur (E5.5-E6.5). Also at this time, XCI occurs again. In the second round of XCI, each cell independently and randomly selects one of the two Xs to inactivate. Different from the imprinted XCI, the second round of XCI is called random XCI (Monk and Harper 1979, Kay, Penny et al. 1993). The inactive state of one particular X chromosome, once established, is faithfully maintained during each subsequent mitotic cell division along the entire life span in all but one type of somatic cells. The Xi is only reactivated again in the primordial germ cells (PGCs) around E7.5-E11.5.

The initiation of XCI is controlled by a special X-linked DNA region, the X chromosome inactivation center (*Xic*), which carries the *Xist* gene and several other genes related to XCI regulation. *Xist* stands for Xi specific transcript, which encodes a 17kb non-coding RNA. In each undifferentiated female embryonic stem (ES) cells, both Xa alleles express *Xist* gene at low level. Upon differentiation, the *Xist* expression is allele-specifically up-regulated along the chosen Xi. Gene knock-out and knock-in studies have demonstrated that Xi-specific up-regulation of *Xist* is the key event to trigger XCI (Brockdorff, Ashworth et al. 1991, Brockdorff, Ashworth et al. 1992, Kay, Penny et al. 1993). The up-regulated *Xist* RNA transcripts coat the whole X chromosome territory *in cis*, and mediate many epigenetic processes to establish multiple layers of epigenetic modifications on the X chromosome. The chromosome-wide gene silencing is then achieved (Penny, Kay et al. 1996, Sheardown, Duthie et al. 1997). Coating of *Xist* RNAs on the Xi chromosome can be detected in RNA fluorescence *in situ* hybridization (FISH) experiment as a bright RNA cloud signal (the *Xist* cloud). Known epigenetic modifications recruited by the

Xist cloud include the enrichment of heterochromatin marks, such as histone H3 lysine 27 trimethylation (H3K27me3) and histone H2A lysine 119 ubiquitination (H2AK119ub), dearth of euchromatic histone modification, enrichment of histone H2A variants (macroH2A) and DNA methylation on CpG islands.

Based on the current knowledge, the process of XCI can be genetically parsed into three phases (Pereira and Vasques 2000). Counting: the process that each cell independently determines its X chromosome copy number; and XCI is only initiated when more than one copy of X is detected per diploid chromosome set. Choice: the mechanism that cells use to select/determine the future X_a/X_i . Silencing: the process of establishing the gene silencing during XCI and the process of maintaining the gene silencing in somatic cells. While many interesting questions are still waiting for the answers and intensive research is being carried out on all aspects of XCI, the questions that are specifically investigated in my graduate research are all related with the third aspect: how the chromosome-wide gene silencing is established and maintained in XCI.

1.3 Formation of the *Xist* cloud and the *LINE-1* hypothesis

One interesting behavior of the *Xist* RNA is that the RNA transcripts form a RNA cloud covering the chromosome territory *in cis*. The RNA cloud does not “travel” to the other X chromosome territory *in trans* (Brockdorff, Ashworth et al. 1991). It also should not affect its neighboring chromosomes. How do the *Xist* RNA transcripts spread out along the Xi? The answer to this question remains elusive. In 1998, Mary Lyon, the British scientist who first put forth the XCI hypothesis in 1951, proposed another hypothesis that the long interspersed nuclear elements (*LINE-1*) serve as the “way stations” to facilitate *Xist* RNA spreading (*LINE-1* hypothesis) (Lyon 1998).

LINE-1, an autonomous retrotransposon, is a major type of repetitive genetic element widely distributed in the mammalian genomes. In humans, *LINE-1* distribution showed significantly enriched on the X chromosome and 79% recent *LINE-1* insertions in the human genome are insertions into the X chromosome (Ostertag and Kazazian 2001). However, the enriched distribution of *LINE-1* is only observed in humans but not in

mice. In a number of studies, the gene-silencing status or the epigenetic landscape along the Xi has been compared with the *LINE-1* distribution pattern along the X chromosome. Some results are consistent with the *LINE-1* hypothesis, providing evidences for the relationship between *LINE-1* elements and XCI (Bailey, Carrel et al. 2000, Carrel, Park et al. 2006, Chaumeil, Le Baccon et al. 2006, Chow, Ciaudo et al. 2010, Tang, Mishra et al. 2012). Other results are not, suggesting *LINE-1* activity is dispensable for XCI (Chow, Yen et al. 2005, Cantrell, Carstens et al. 2009). In cytogenetic analysis, it was observed that, during XCI, X-linked genes to be silenced are relocated from the periphery of the chromosome territory into a gene silencing core compartment, which is formed by intergenic DNA sequences (Fig.1) (Chaumeil, Le Baccon et al. 2006). Interestingly, in a recent study (Chow, Ciaudo et al. 2010), it was observed that the *LINE-1* repeats transcribes *LINE-1* RNA during the *in vitro* differentiation of mouse ES cells. RNA FISH results show that the *LINE-1* RNAs form “*LINE-1* clouds”. In the cells of day 4 *in vitro* differentiation, the *LINE-1* clouds are adjacent to the *Xist* clouds, while by day 10 of differentiation, *LINE-1* clouds overlap

with the *Xist* clouds. In fully differentiated cells, *LINE-1* cloud cannot be detected. However, it is still unclear whether *LINE-1* transposon activities are involved in this process. It is also unclear whether *LINE-1* RNAs behave similarly during XCI in humans. In conclusion, the mechanism of *Xist* cloud formation is still unclear. The *LINE-1* hypothesis still awaits confirmation by future studies.

Figure 1 Model for chromosome territory re-organization during XCI

adopted from (Chaumeil, Le Baccon et al. 2006).

The timeline of differentiation and main XCI events are shown. Day 0: in undifferentiated female ES cells, both Xs are active. Genes subjected to XCI are represented as small circles. Day 1-2: Xist RNA starts to accumulate on the X chromosome chosen for inactivation. Day 2-3: a repressive compartment is formed in Xist RNA cloud territory, which is associated with the exclusion of the transcription machinery. > day3: gene silencing on Xi starts. These genes are then relocated into the silencing compartment. This locus internalization may help the genes to maintain the silent state. Genes escaping XCI, such as *Jarid1c*, are resistant to the internalization due to certain mechanisms. These genes are located outside the silencing compartment, which may help them to escape.

In addition to LINE-1, there are some other factors identified to involve in *Xist* RNA cloud formation: Yin-Yang1 (YY1) (Jeon and Lee 2011), special AT-rich binding protein (SATB1) , and scaffold attachment factor A (Saf-A, hnRNP U or SP120) (Helbig and Fackelmayer 2003).

YY1 is a ubiquitously distributed DNA-binding transcription factor, which is able to both repress and activate a bunch of genes (Tang, Mishra et al. 2012). Previous studies found YY1 is also involved in transcription regulation of the *LINE-1* elements (Minakami, Kurose et al. 1992, Becker, Swergold et al. 1993, Kurose, Hata et al. 1995, Jeon and Lee 2011). shRNA knockdown of YY1 disrupts the *Xist* cloud coating along Xi. YY1 binds to both *Xist* RNA and X chromosome, thus suggesting that YY1 may play a role in tethering *Xist* RNA to the Xi (Minakami, Kurose et al. 1992, Kurose, Hata et al. 1995, Jeon and Lee 2011).

In most differentiated cell types, artificially induced *Xist* expression cannot initiate gene silencing. However, induced *Xist* expression in lineage-restricted progenitor cells in the hematopoietic system of adult mice can cause XCI (Savarese, Flahndorfer et al. 2006). This special

developmental context, which provides the permissiveness to XCI, is found to be related with SATB1 (Agrelo, Souabni et al. 2009). In normal thymocytes, SATB1 forms a cage-like structure in the nucleus, functions as gene expression regulator and chromatin-remodeling factor recruiter (Cai, Han et al. 2003). It is observed that SATB1 interferes *Xist* RNA localization. In more than half of the thymocytes, *Xist* RNA only localized along the *SATB1* positive areas (the SATB1 “cage”), and after *SATB1* expression is down regulated, *Xist* RNAs start to coat the Xi (Agrelo, Souabni et al. 2009). The same study also shows knockdown both SATB1 and the closely related SATB2 proteins lead to the decreasing of gene silencing efficiency by *Xist* in ESCs (Agrelo, Souabni et al. 2009). However, there are other results which may be against the essential role SATB1 played in XCI process. First, SATB1 does not co-localize with *Xist* RNA in ESCs and is not expressed in all somatic cells (Alvarez, Yasui et al. 2000, Agrelo, Souabni et al. 2009). Also, recently, a study shows that double knockout of SATB1 and SATB2 does not affect the initiation of XCI in mouse female ES cells (Nechanitzky, Davila et al.). These evidences suggest a dispensable role of *SATB1* in XCI.

Saf-A is a nuclear matrix protein, which is found to be concentrated on the inactive X chromosome with low protein mobility (Helbig and Fackelmayer 2003). This enrichment is *Xist* dependent. On the other hand, knocking down of Saf-A by RNAi also results in disruption of the *Xist* RNA cloud formation. Knocking down of Saf-A also demonstrates that Saf-A is required for *de novo* XCI, but the enrichment of saf-A is dispensable for XCI maintenance (Pullirsch, Hartel et al. 2010).

In the recent two years, the distribution of *Xist* RNA coating along mouse X chromosome was also profiled by studying the genomic DNA sequences interacting with the *Xist* RNA (Simon, Wang et al. 2011) (Engreitz, Pandya-Jones et al. 2013). It was found that *Xist* RNA preferentially binds to gene-rich areas. Genes escaped from XCI are found to have less *Xist* RNA coated and the low *Xist* RNA occupancy is strictly limited to the escaped gene bodies, in comparison with their flanking regions. Although *LINE-1s* were speculated to serve as “way stations” in *Xist* RNA spreading process, the studies reveal a poor correlation between *Xist* RNA coating and the distribution of *LINE-1*, as *LINE-1* is known to

reside in gene-poor areas. *Xist* RNA coating is positively correlated with the distribution of heterochromatic mark H3K27me3 in mouse lung fibroblast (MLFs) and in mouse embryonic fibroblast (MEFs). Moreover, spreading of H3K27me3 and enhancer of zeste homolog 2 (Ezh2) shows high consistency with *Xist* RNA spreading in differentiating mouse ES cells at time points of day 0 and day 7, supporting the previous finding that Eed(embryonic ectoderm development)-EZH2 complex is recruited by *Xist* RNA and it conducts the histone modification H3K27me3 occurrence (Plath, Fang et al. 2003, Silva, Mak et al. 2003, Zhao, Sun et al. 2008).

1.4 X-linked genes which escape XCI

Although most of X-linked genes are subjected to XCI, it has long been known that there are some X-linked genes which escape XCI. By comparing the features of the escaper genes and those involved in silencing may help us to better understand the mechanism of XCI gene silencing. It has been found that these escaper genes have DNA hypomethylation at their promoters (Carrel and Willard 2005). However,

the number, distribution and conservation of genes which escape XCI are different between mice and humans. Quite fewer escaping genes are found in the mouse, while in humans, there are at least 15% of X-linked genes reported to escape XCI, and most of these genes are located on the p arm of human X chromosome which carries the pseudo-autosomal region. Pseudo-autosomal region is a short chromosomal region homologous between X and Y chromosomes. It is the “left-over” region of the ancient autosome, from which X and Y were evolved (Carrel and Willard 2005).

1.5 Sporadic gene reactivation along the Xi after conditional deletion of *Xist* gene in the female fibroblasts

The gene silencing of Xi is established during early embryonic development. Once established, the gene silencing status of Xi is tightly maintained in somatic cells. Conditional deletion of *Xist* gene from the Xi in fibroblast cells only caused rare and sporadic gene reactivation along the Xi (Brown and Willard 1994, Csankovszki, Panning et al. 1999, Zhang,

Huynh et al. 2007). Upon *Xist* removal, the reactivation rate of a Xi-linked GFP reporter is as low as 0.05% (Csankovszki, Nagy et al. 2001). Allele-specific RT-PCR results also confirmed that reactivation of endogenous X-linked genes does occur in a rare and sporadic manner (Zhang, Huynh et al. 2007). These results show that the *Xist* RNA plays a minimal, if not dispensable, role in maintaining the Xi gene silencing in somatic cells. However, such a view was generated by conventional techniques, such as allele-specific PCR, which can only be carried out on a small sampling of a few selected genes. For example in the study of (Zhang, Huynh et al. 2007), 4 gene reactivation events are detected among 7 cell lines on 7 selected genes. There are more than 2000 X-linked genes in the mouse. Based on current available results, we still do not know how many genes are reactivated along the entire X chromosome after *Xist* removal. Is the gene reactivation pattern truly sporadic along the entire X chromosome? Is there any X-linked gene, which is consistently or frequently reactivated after *Xist* removal? If so, is there any consensus DNA sequence that we can recognize among these genes. Such DNA motifs not only can be used

to predict a reactivation-prone gene, but also can be used to study the mechanism of *Xist* RNA mediated gene silencing. These are important questions, which can be answered by a chromosome-wide comprehensive gene reactivation profile. However, such a comprehensive profile is impossible to be achieved by conventional methods, such as allele-specific RT-PCR.

1.6 Padlock probe SNP capture: a high throughput digital RNA allelotyping method

Recently, a new technique based on high throughput DNA sequencing was developed for RNA allelotyping. This method, called “padlock sequencing” or “padlock probe SNP capture”, uses specially designed DNA oligonucleotides probes to capture SNPs of interest (Fig. 9). One probe is designed for one SNP (single nucleotide polymorphism). Each probe contains two “capture arms”, which form complementary base pairs with the cDNA regions flanking the target SNP. After the two capture arms anneal with their cDNA template, the probe forms a “padlock” like

structure. Subsequently, a single round of DNA chain elongation is carried out to “capture” the target SNP. After that, one round of ligation reaction is carried out to “lock up” the “padlock” into a single-stranded circular structure. Other than the two capture arms, which are uniquely designed for each SNP, each padlock probe also contains a common linker which provides two PCR priming sites for downstream sequencing library construction. Since the PCR priming sites are common for all probes, the SNPs captured by a padlock probe library can be PCR amplified for solexa sequencing. In the initial attempt of padlock probe SNP capture, a padlock probe library is designed to capture 27,000 SNPs within 10,345 genes of the human genome (Zhang, Li et al. 2009). Since the specific SNPs are selectively captured and amplified by pad-lock probes before being subjected to deep sequencing, this method can concentrate the sequencing power onto a selected pool of SNPs, and therefore is much more quantitative and less expensive compared to direct RNA-sequencing method.

The padlock probe SNP capture method is a perfect fit for the XCI studies. In my research, I used this method to generate a chromosome-wide comprehensive gene reactivation profile of the Xi after *Xist* removal. The female mouse fibroblast cell lines involved in this research carry two X chromosomes from two different mouse strains: 129 (001137; 129P1/ReJ) and Castaneus (000928). To distinguish them, there are plenty of available SNPs collected in the current SNP database (<http://www.sanger.ac.uk/resources/mouse/genomes/>). The *Xist* gene on the X¹²⁹ allele is mutated, so that the conditional *Xist* gene deletion specifically occurs on the X¹²⁹ allele. In our experiment, we engineered female cell lines in which the mutated X¹²⁹ allele is non-randomly silenced in all cells. Therefore high throughput RNA allelotyping on such a cell line can generate a chromosome-wide comprehensive gene reactivation profile.

1.7 Non-coding DNA

The eukaryotic genome is composed of a variety of non-coding regions. The name “non-coding DNA” is used to represent these regions. In human, ~98% genome is made up of these non-coding DNA (Lander, Linton et al. 2001). Some of the non-coding regions can be transcribed into RNA molecules of known or unknown functions. The function of the non-coding DNA has been in the dark for a long time, people called them “junk DNA” at first place. The function of junk DNA has been gradually characterized. Types of non-coding DNA including non-coding functional RNA, Cis- and Trans-regulatory elements, introns, pseudogenes, repetitive sequences, transposons, viral elements and telomeres are classified. Some of these RNAs indispensably participate in essential cellular processes, such as ribosomal RNA (rRNA), transfer RNA (tRNA), Piwi-interacting RNA, long non-coding RNA (lncRNA) and microRNA (miRNA). Cis-regulatory elements refer to the DNA elements that can regulate the expression of a nearby gene and targets of trans-regulatory elements are distant genes. Promoters and enhancers are also classified into this

category. Introns are non-coding regions located inside genes. Introns are transcribed into precursor RNAs and, which are eventually removed by RNA splicing. After splicing, the mRNAs are considered as mature mRNAs, which are translated into proteins. Pseudogenes are those derived from genes but have lost expression ability (Zheng, Frankish et al. 2007).

Repetitive sequences are composed of two classes: tandemly repeated sequences and dispersed sequences. Tandemly repeated sequences are products of gene duplication. Dispersed sequences consist of transposons and the most typical representatives of dispersed sequences are long interspersed nuclear element (LINE) and short interspersed nuclear element (SINE) (Okada 1991). They also possess transposition ability (Rogers 1985, Weiner, Deininger et al. 1986). Telomeres are a special repetitive DNA regions located at the end of chromosomes (Berendes and Meyer 1968). The binding of proteins and telomeres can protect chromosome ends (Chan and Blackburn 2004).

Uncovered functions of non-coding DNA sequences include protecting the genome, regulating gene expression, determining the

binding of transcription factors, repressing, enhancing, silencing and promoting gene expression by binding to different types of proteins and insulating active gene regions from silent gene regions.

Last year, a research studying on C₀t-1 repeat RNA reveals that non-coding DNA transcripts are related to chromosome structure altering (Hall, Carone et al. 2014). C₀t-1 DNA is highly repetitive DNA in genome, including *LINE-1*. It is widely used in RNA FISH to block nonspecific hybridization between FISH probes and genomic repeat areas. In this study, C₀t-1 DNA was fluorescently labeled to detect the transcripts from repetitive genomic DNA regions using RNA FISH. To their surprise, abundant C₀t- 1 RNAs, mainly including LINE-1, are associated with interphase euchromatin and nonchromatin scaffold. Loss of the coating of these repetitive RNAs leads to abnormal chromatin distribution and condensation. However, on the inactive X chromosome, which is coated by *Xist* RNA, the association of C₀t-1 RNA is rarely seen (Hall, Carone et al. 2014). This study suggests that genomic repetitive DNA region transcripts

may participate in altering chromosome structure, which inspired (not true) us to study the non-coding transcription pattern on XCI.

1.8 Observation of the Xi through EM with conventional methods

Another question related with the gene silencing mechanisms of XCI is whether higher order of chromatin structures are involved in gene silencing. It is not only a challenging question for the XCI research field, but also for the epigenetic research field in general.

The Xi is first observed by Barr and colleagues in 1949. They named the structure, which frequently appeared around the nucleoli in female cat neurons as “the nucleolus satellite” (Barr and Bertram 1949). This structure, also known as “the Barr body”, is identified as the Xi later on. Since the opposite gene expression activities of the two Xs in each female cell cannot be determined by their DNA sequences. XCI is considered as a pure epigenetic phenomenon. Indeed, multiple epigenetic modifications specifically enriched on the Xi have been identified. It is believed that these epigenetic modifications heterochromatinize the chromosome to make the Xi chromatin highly “compact”. This is why Xi could be

specifically stained by basic dyes, such as cresyl violet, and be observed by Barr and colleagues in 1949. However, although it has been firmly established that the Xi carries multiple layers of special modifications on its chromatin, no direct evidence is available to show the details of the high order chromatin structure of Xi. Is the Xi chromatin truly compact? Other than the histone modifications, is there any Xi-specific higher order chromatin structure? As the Xi chromatin is heavily modified by epigenetic modifications, it is very difficult to use biochemistry or structure biology approaches to study the Xi *in vitro*. Using the high resolution of electron microscope (EM) to directly observe the Xi in cells is one of the best available tools to study the chromatin structure of the Xi.

The higher order chromatin structure of the Xi can be studied using a transmission electron microscope (TEM), which applies a beam of electrons to transmit through an ultra-thin specimen. The electrons are emitted at the same direction and at a very high speed. When the electron beam passes through a specimen, the electrons that interact with the atoms of the specimen will change their direction, and therefore create

“contrast”. The electrons that pass through the specimen will be detected by an electron sensor. Because TEMs use electrons to scan the specimen, they are able to provide images in a resolution of 2nm, which is a much higher resolution than conventional light microscopies, which typically provide a resolution of 200 nm. Furthermore, the size of a nucleosome is about 11 x 5.5 nm (White, Suto et al. 2001), which fits perfectly with the EM resolution. However, there are at least two major difficulties involved in using EM to study chromatin structure. First, a special piece of chromatin, such as the Xi, is heavily modified by post-translational modifications on the histone tails. Such a chromatin fragment is impossible to be prepared in large amount using biochemistry or biophysics approaches. Therefore, the chromatin structure has to be visualized in cells, which requires labeling method to identify the chromatin specimen. Secondly, as what is discussed in the following section, most of biological samples is difficult to be visualized under the conditions required in electron microscopy.

The electron-dense compounds of heavy metals are typically used for staining in TEM study, as the atoms of heavy metals can easily interact with electrons to create strong contrast. Unlabeled biological samples, however, are not friendly for TEM studies mainly because of two reasons. First, the atoms of a biological specimen can hardly interact with the electron beam, therefore, the biological specimen are very “transparent” under TEM. Secondly, biological samples are usually easy to be crushed within an electron microscope, in which extreme conditions are created for the emission of the electron beam.

There are three conventional approaches of using TEM in biology studies: freeze-etching, negative staining and positive staining.

In the freeze-etching technique, also called “cryo-EM”, samples are frozen at a very high speed that the water molecules do not have time to reorganize into ice, and their forms remain amorphous (frozen-hydrated samples). These samples are kept in liquid nitrogen to avoid the evaporation of water, therefore, in cryo-EM, the sample is hydrated but in solid state. This technique has significant advantages. For example, the

preservation is extremely good and the resolution is very high. However, it also has significant disadvantages. First, the cryo-EM is highly technically challenging. Secondly, the target is not labelled; therefore when cryo-EM is used to study cells, the location of the targets (a protein, or a specific subcellular structure) may not be easily identified. Other problems are the low contrast (biological samples are almost transparent under EM) and the fragility of biological samples even under minimized radiation damage.

In a negative staining approach, the biological sample is stained with a solution of heavy metal compound. This staining solution is supposed to fill up all the “open spaces” surrounding all the “components” within a biological sample. Such a stained sample is then subjected to EM; the biological samples are visualized as the transparent patterns within a dark background (negative stain). This approach can generate high contrast and reduce radiation damage, but it also has apparent shortfalls. The staining only creates a crust covering the surface of the biological specimen. This significantly decreases the image resolution. In addition, the “crust” may

still be crushed by the high energy electron beam because of the fragility of the biological samples.

In the positive staining approach, the biological target is stained directly. The most commonly used technique of positive staining is the immunogold labeling, in which the gold atom is conjugated to an antibody that specifically recognizes an antigen. After a biological sample is stained by such an antibody, the biological sample is “positively” stained for EM work. The disadvantage of this approach is that the size of an antibody (5.6nm) puts a limitation on the resolution of EM, which can achieve a resolution of 2nm. Secondly, in the immunostaining, how well the antibody can penetrate through the biological specimen to bind the antigen is a significant concern. As shown in Fig.2, the gold particles (arrow) in this EM image can only be used to confirm the localization of the target protein, but they cannot at all provide any structure details of the target.

The inactive X chromosomes in interphase nuclei were observed by TEM in one previous study (Rego, Sinclair et al. 2008). Several different

fixation approaches were adopted to prepare the sample for the EM observation. The unlabeled Xi was identified in the nuclei only by its unique “dark color”, which sometimes helps to distinguish the bar body from its surrounding neighborhood. H3K27me3 immunogold staining was used in this study (Fig.3). However, the immunogold staining pattern does not provide any structure details of the Barr body. To improve the penetration of the antibody in immunostain, this study applies detergent extraction of nucleoplasm before fixation. However, the immunogold staining is still of poor resolution, which is only used to confirm the identification of the bar body. The results of this study suggest that unlike the solid heterochromatin structure of centromeres, Xi heterochromatin showed a loose package of chromatin fibers/domains.

Figure 2 An EM picture of immunogold labeling of endogenous LRRK2 from NIH-3T3 cells.

This figure is adopted from (Civiero, Vancraenenbroeck et al.).
 Arrowhead: immunogold particles. Scale bar: 100nm.

Figure 3 An EM picture of H3K27me3 immunogold labeling of an inactive X chromosome located at the edge of a nucleus.

This figure is adopted from (Rego, Sinclair et al. 2008). Black arrow: Xi; Arrowhead: an immunogold particle; Red arrow: nuclear membrane.

1.9 The miniSOG labeling method for EM

MiniSOG (mini singlet oxygen generator) is a genetically engineered protein generated by modifying a light-absorbing protein in plants (Shu, Lev-Ram et al.). When miniSOG is illuminated by blue light, it generates sufficient singlet oxygen, which catalyzes the reaction of diaminobenzidine (DAB) polymerization. DAB polymer is osmiophilic, which can be easily stained by osmium tetroxide and becomes electron-opaque. This reaction enables miniSOG to serve as a genetically-coded label for electron microscope. In addition to its ability of catalyzing oxidative polymerization of DAB, miniSOG can also function as a green fluorescent protein. Taking advantages of the green fluorescence of miniSOG, highly localized DAB polymerization can be achieved by piping oxygen to a target area which is identified under fluorescent microscope. Compared to the conventional labeling method of EM, the miniSOG labeling method provides significant advantages. First, it is a positive labeling method, therefore, it provides high-contrasted images with better resolution; and the labeled specimen is more resistant to the damage caused by the electron beam. Secondly,

the target protein is labeled in live cells by a fusion protein; therefore, poor antibody penetration is not a concern.

1.10 macroH2A

macroH2A (mH2A), one of the histone variants, is found to play important roles in many epigenetic processes. Two genes encode macroH2A. macroH2A1, which has four isoforms (mH2A1.1, mH2A1.2, mH2A1.3, mH2A1.4) in different splicing patterns, is encoded by gene *H2afy*, and macroH2A2 is encoded by a separate gene *H2afy2* (Rasmussen, Huang et al. 1999). The macroH2A protein can be dissected into two domains, the conserved H2A domain, which shares high conservation sequence with the core histone H2A (Pehrson and Fuji 1998) and the macro domain (Muthurajan, McBryant et al. 2011) (Fig.4).

The enrichment of macroH2A1.2 on the Xi can be recognized as a macrochromatin body (MCB) in female cells (Costanzi and Pehrson 1998). However macroH2A plays a dispensable role in XCI. The establishment of the enrichment of macroH2A on Xi is a late event of XCI process. Double

knockout of the two genes encoding macroH2A does not affect XCI initiation (Tanasijevic and Rasmussen 2011). However, nuclear transfer experiments showed that, when a mouse somatic nucleus is transferred into an *Xenopus* oocyte, macroH2A plays a role in the resistance to reprogramming of the Xi, indicating that macroH2A might be involved in the epigenetic mechanism maintaining the gene silencing of Xi (Pasque, Gillich et al. 2011).

The enrichment of macroH2A, a histone variant, on the Xi makes macroH2A a good marker protein for live-cell labeling of Xi. Live-cell labeling of Xi has been achieved using a macroH2A-GFP fusion protein. The results show that the Xi could be successfully labeled by fusing full length macroH2A with GFP (mH2A-GFP). The Xi labeling could also be achieved by fusing the H2A domain of macroH2A with GFP (H2AD-GFP). However, the Xi labeling is not successful by fusing the macro domain of mH2A with GFP (Chadwick, Valley et al. 2001). In my research, we plan to label the Xi by mH2A-miniSOG and H2AD-miniSOG fused proteins and we

hope that this study can provide us a more detailed view on the Xi chromatin structure under the electronic microscope.

Figure 4 Comparison of H2A and macroH2A peptides.

The H2A domain, 116 a.a on the N terminal of macroH2A labeled in blue, is homologous with histone H2A. The C-terminal domain of mH2A, labeled in orange, is called “macro domain”.

2 Objective

- Use Padlock probe SNP capture to generate a comprehensive gene reactivation profile of the X-linked genes in mouse fibroblast cells in which *Xist* is conditionally and allele-specifically deleted from Xi.
- Examine whether long-term cell culture can affect the gene reactivation when *Xist* is conditionally deleted from Xi.
- Using Padlock probe SNP capture to generate a comprehensive expression profile of the X-linked non-coding regions in mouse ES cells, fibroblast cells and the fibroblast cells in which *Xist* is conditionally and allele-specifically deleted from Xi.
- Fuse miniSOG with macroH2A to label Xi in fibroblast cells and attempt to get an improved EM observation on the chromatin structure of Xi.

3 Materials and Methods

3.1 Cell Culture:

Mouse fibroblast cells were cultured in DMEM (Gibco 12800-082) with 10% FBS (Hyclone SV30160.03), 1% Penicillin/Streptomycin (Gibco 15140), 25mM HEPES (Gibco, 15630-080), 1 x MEM Non-Essential Amino Acids Solution (Gibco, 11140-050) and 1% sodium bicarbonate at 37°C and 5% CO₂.

Mouse embryonic stem cells were cultured in DMEM (Gibco 12800-082) with 15% FBS characterized (Hyclone SH30071.03), 1% Penicillin/Streptomycin (Gibco, 15140), 25mM HEPES (Gibco, 15630-080), 1 x MEM Non-Essential Amino Acids Solution (Gibco, 11140-050), 1% sodium bicarbonate and 500 units/ml Leukemia inhibitory factor (Merck, ESG1107) at 37°C and 5% CO₂.

Cell lines used:

129T cells: a transformed female fibroblast cell line established from the mouse strain 129 (129P1/ReJ).

2loxT 129/Cast cells: a transformed female mouse fibroblast cell line containing two X chromosomes from two different mouse strains: 129 (001137; 129P1/ReJ) and Castaneous (000928; CAST/EiJ). The X¹²⁹ allele was genetically engineered. Two loxP sites were inserted into the *Xist* gene body on the X¹²⁹ allele (Csankovszki, Panning et al. 1999) (Fig.5A).

HR1,HR5-HR9 cells: 2loxT 129/Cast cells with *Xist* conditionally deleted from the inactive 129 strain X chromosome through transient transfection of p-SALK-CRE. The cells were isolated by limiting dilution and the deletion was confirmed by PCR genotyping and *Xist* RNA FISH. They are called 1lox cells in this study.

HR5.D25 and HR6.D25 cells: two long term cultured cells derived from HR5 and HR6, respectively.

3F1 cells: Mouse female embryonic stem cells, genotype as $Xa^{129}Xa^{Cast}$, in which *Tsix* gene on Xa^{129} is permanently mutated (Lee and Lu 1999).

HR7.shRNA: HR7 cells with stable integration of shRNA against mouse general transcription factor IIIC, polypeptide 4 (Gtf3c4, gene ID: 269252).

3.2 Lipofectamine Transfection:

Transfection of mammalian cells was performed using Lipofectamine2000 as described in the product's user manual (Life Technologies 11668-019). To transfect the transformed mouse fibroblast cells in a 9.4 cm² culture vessel, 2 µg plasmid DNA and 10 µl Lipofectamine were use.

3.3 DNA Constructs

Plasmid constructs involved in my studies are listed as follows:

p-mH2A-SOG: fusion construct of mH2A fused with a single copy miniSOG tag driven by a CMV promoter;

p-H2AD-SOG: fusion construct of H2A domain of mH2A fused with a single copy miniSOG tag driven by a CMV promoter;

p-H2AD-mc: fusion construct of H2A domain of mH2A fused with mCherry fluorescent tag driven by a CMV promoter;

p-H2AD-SOG-mc: fusion construct of H2A domain of mH2A fused with a single copy miniSOG tag and mCherry tag driven by a CMV promoter;

p-H2AD-tdSOG: fusion construct of H2A domain of mH2A fused with a tendon dimer miniSOG tag driven by a CMV promoter;

p-mH2A-mc-SOG: fusion construct of mH2A fused with mCherry fluorescent tag followed by a single copy miniSOG tag driven by a CMV promoter;

p-H2AD-mc-SOG: fusion construct of H2A domain of mH2A fused with fluorescent mCherry tag followed by a single copy miniSOG tag driven by a CMV promoter;

pSUPER-hygro-sv40: a pSUPER vector containing a hygromycin resistant cassette.

Plasmid constructs obtained commercially or from other laboratories are listed as follows:

p-SALK-CRE: for expression of Cre recombinase;

p-miniSOG-mCherry-C1: a plasmid obtained from Roger Tsien's laboratory (<http://www.tsienlab.ucsd.edu/>) containing the miniSOG-mCherry sequence;

ptdminiSOG-N1: a plasmid obtained from Roger Tsien's laboratory (<http://www.tsienlab.ucsd.edu/>) containing the tandem dimer miniSOG sequence.

Sx9 probe: a P1 DNA construct containing about 40kb genomic DNA fragment covering the *Xist* gene.

3.4 Total RNA isolation:

Total RNA was isolated using TRIzol reagent (Life Technologies, 15596-026) according to the manufacture's instruction. In brief, 1 ml TRIzol reagent was used to lyse cells per 10 cm² growth area. The cell lysate was incubated at room temperature for 5 minute (min). After incubation, 0.2 ml Chloroform was added. The reaction mixture was vortexed vigorously and incubated at room temperature for 3 min. The mixture was then centrifuged at 12,000 g at 4°C for 15 min. The upper aqueous phase containing RNA was then transferred into a fresh tube. 0.5 ml isopropyl alcohol was added. The mixture was incubated at room temperature for 10min followed by centrifuge at 12,000 g at 4°C for 10 min. The

supernatant was discarded. The pellet was wash twice with 1 ml 75% ethanol. After air drying, the pellet was dissolved in appropriate amount of DEPC water. To remove DNA, the dissolved RNA was incubated with 3µl RQ1 RNase-Free DNase (Promega, M6101, 1 U/µl) in 1 x DNase I buffer at 37°C for 30 min and with 3 µl newly added RQ1 DNase at 37°C for another 30 min. After incubation, equal volume of phenol chloroform (Invitrogen 15593) was added to the RNA solution. The mixture was strongly vortex for 20 second (sec) and centrifuged at RT at 14,000 rpm for 5 min to separate into two liquid layers. The upper aqueous layer was transferred to a fresh tube and was added with 01 volume of 5 M LiCl and 3 volume of 100% ethanol. The solution was incubated at -80°C overnight. The next day the solution was centrifuged at 12,500 g at 4°C for 15 min, followed by 75% ethanol washing and centrifuge at 12,500 g at 4°C for 5 min. The RNA pellet was dissolved in 100 µl DEPC water for at least 30 min on ice and it was ready for the consequent use.

3.5 Reverse transcription:

Reverse transcription of X-linked genes was carried out using SuperScript III First Strand Synthesis System (Invitrogen, 18080-051) by following the manufacture's instruction. Oligo(dT)₂₀ primer was used to synthesize first-strand cDNA. cDNA was then cleaned up with a Sephadex G-25 column. In brief, 10 µg total RNA was mixed with 2 µl 50 µM

oligo(dT)₂₀ and 2 µl 10 mM dNTP mix. The reaction mix was incubated at 65°C for 5min, then, on ice for at least 1 min. The reaction mix was then mixed with 4 µl 10 x buffer, 8 µl 25mM MgCl₂, 4 µl 0.1 M DTT, 2 µl RnaseOUT (40U/µl) and 2 µl SuperScript III RT (200U/µl). After incubating at 50°C for 50min, 85°C for 5min and put on ice for at least 1 min, 1 µl RNase H was added to the reaction mix and the reaction was carried on at 37°C for 20 min.

To generate cDNA for none-coding regions padlock probe targeting, iScript Reverse Transcription Supermix kit (Bio-Rad, 170-8841) was used by following the manufacture's instruction. 1 µg total RNA was mixed with 4 µl 5x iScript reverse transcription supermix, making the total volume as 20 µl, which was followed by incubating at 25°C for 5 min, 42°C for 30 min and 85°C for 5 min. 10 parallel reactions were done for each cell line. Ethanol precipitation was applied to concentrate the first-strand cDNA.

The single-stranded cDNA was purified using Sephadex G-25 column (GE Healthcare, 27-5325-01) and the concentration was measured using QubitssDNA Assay Kit (Invitrogen Q10212), and stored at -20°C.

3.6 Gateway Recombination:

The DNA fragments of mH2A and miniSOG were combined using multiple gateway reaction kit (Life Technologies, 12537-100) by following

the manufacture's instruction. In brief, the total volume of each BP recombination reaction was 8 µl and the reaction reagent contained 1-7 µl attB PCR product (15-150 ng), 1 µl pDONRTMvector (150 ng/µl) and 1 x TE Buffer (pH8). The reaction was carried out by 2 µl BP ClonaseTM II enzyme mix (Life Technologies 11789-020) and was incubated at 25°C for 1 hour (hr). The reaction was then stopped by 1 µl proteinase K solution (Life Technologies, AM2546) at 37°C for 10min.

For LR reaction, the volume of reaction reagent was 8µl contains 1-7 µl BP reaction product (15-150 ng), 1 µl pcDNATM3.2/V5-DESTvector (Life Technologies, 12489-019, 150 ng/µl) and 1 x TE Buffer (pH8). 2 µl LR ClonaseTM Plus enzyme mix (Life Technologies 12538-120) was then added. The reaction mix was incubated at 25°C for 1 hour. After that, 1 µl proteinase K solution (Life Technologies AM2546) was added, and the reaction mix was incubated at 37°C for 10 min to stop the reaction.

PCR primers used are shown as follows:

miniF:

5' –

GGGGACAACCTTTGTATACAAAAGTTGCCATGGAGAAAAGTTTCGTGATAACT

GATCCA–3' ;

miniR:

5' -

GGGGACCACTTTGTACAAGAAAGCTGGGTATTATCCATCCAGCTGCACTCCG

-3' ;

mH2AF:

5' -

GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGTCGAGCCGCGCGGGAAG

A-3' ;

mH2AR:

5' -

GGGGACAACTTTGTATACAAAGTTGTGTTGGCGTCCAGCTTGGCCATT-3' ;

mH2AR_H2Adomain:

5' -GGGGACAACTTTGTATACAAAGTTGTCTTCTTCGCTAGCAACTCAGG

ATGG-3' ;

td-mChrF:5' -

GGGGACAACTTTGTATACAAAAGTTGCCATGGAGAAGAGCTTCGTGATCACC

G-3' ;

mChrR:

5' -

GGGGACCACTTTGTACAAGAAAGCTGGGTATTACTTGTACAGCTCGTCCATG

CCGC-3' ;

mChrF:

5' -

GGGGACAACTTTGTATACAAAAGTTGCCATGGTGAGCAAGGGCGAGGAG-3' ;

mH2ARB4:

5' -

GGGGACAACCTTTGTATAGAAAAGTTGGGTGGTTGGCGTCCAGCTTGGCC-3' ;

H2ADRB4:

5' -

GGGGACAACCTTTGTATAGAAAAGTTGGGTGCTTCTTCGCTAGCAACTCAGGA

TG-3' ;

mChrF4r:

5' -

GGGGACAACCTTTTCTATACAAAAGTTGCCATGGTGAGCAAGGGCGAGGAG-3' ;

mChrR3r:

5' -

GGGGACAACCTTTATTATACAAAAGTTGTCTTGTACAGCTCGTCCATGCCG-3' ;

miniFB3:

5' -

GGGGACAACCTTTGTATAATAAAGTTGCCATGGAGAAGAGCTTCGTGATCACC

-3-3

miniRB2:

5' -

GGGGACCACTTTGTACAAGAAAGCTGGGTATTAGCCGTCCAGCTGCACG-3' ;

mH2Ain8-FB1:

5' -

GGGGACAAGTTTGTACAAAAAGCAGGCTCCAACTCAAAGCAGTCAGCAGTG

A-3' ;

mH2Ain8-RB4 :

5' -GGGGACAAC TTTGTATAGAAAAGTTGGGTGGTTGGCGTCCAGCTTGGC-

3' ;

SOGmc-FB4r :

5' -

GGGGACAAC TTTTCTATACAAAAGTTGCCATGGAGAAGAGCTTCGTGATCA-

3' ;

SOGmc-RB3r :

5' -

GGGGACAAC TTTATTATACAAAAGTTGTTTACTTGTACAGCTCGTCCATGC-

3' ;

mH2Aex9-FB3 :

5' -GGGGACAAC TTTGTATAATAAAGTTGCCTAGGCCAGTGATCCCTAGAG-

3' ;

mH2Aex9-RB2 :

5' -

GGGGACCACTTTGTACAAGAAAGCTGGGTATGCCACTGGGTGTACGGTA-3' ;

3.7 Photo-oxidation:

The photo-oxidation of miniSOG was achieved by following the protocol based on the previously published (Shu, Lev-Ram et al.). Cells were grown in MatTek dishes (MatTek, P35G-1.5-10-C). The fixation is done using 2% glutaraldehyde (Sigma-Aldrich, EM grade, G5882) in 0.1 M cacodylate buffer (Sigma-Aldrich, C0250), pH7.4 on ice for 1 hr. Then the cells were rinsed in ice-cold 0.1 M cacodylate buffer + 100 mM glycine. To permeabilize fibroblast cells, 0.5% Triton was added to the rinsing buffer at this step. After rinsing for 6 times x 2 min, the cells were blocked in ice-cold 5 mM mersalyl acid (Sigma-Aldrich, M9784), followed by 3 times x 1 min rinsing in ice-cold 0.1M cacodylate buffer +100mM glycine. The cells were then blocked in block2 solution (100 mM glycine, 10 mM aminotriazole (AT) (Sigma-Aldrich, A8056), 10 mM potassium cyanide (KCN) (Sigma-Aldrich, 31252)). After rinsing in ice-cold 0.1 M cacodylate buffer +100 mM glycine for 3 times x 1 min, the miniSOG labeled cells were centralized under a fluorescent microscopy. Rinsing buffer was carefully removed by a dropper and 0.5 mg/ml DAB in cacodylate buffer, pH7.4 was added in dark. After 10 – 15 min incubation, strong light from a 150W Xe light source using a standard FITC filter set (Ex470/40, DM510, BA520) was applied to illuminate the region of interest. The photo-oxidation usually took 2-5min and the reaction had to be stopped once a light brownish reaction product showed up in the miniSOG labeling place. After photo-oxidation, ice-cold cacodylate buffer was used to rinse cells

for 5 times x 3 min on ice. Cells were then post-fixed in 1% OsO₄ in CB for 30 min on ice, followed by rinsing in ice-cold ddH₂O for 5 times x 1 min. Dehydration was performed by rinsing cells in 20-90% ethanol for 2 min each on ice and in 100% ethanol at room temperature for twice. Then cells were infiltrated with Durcupan ACM resin (Electron Microscopy Sciences, 14040) by adding 1:1 Durcupan: ethanol for 30 min at room temperature, followed by adding 100% Durcupan for 3 times x 1 hr. Cells were then left in 100% Durcupan overnight shaking at room temperature. Resin polymerization was achieved at 60 °C for 2 days.

3.8 Genomic DNA Extraction:

For a 96-well plate, cells were washed with 1 x PBS and treated by 100 µl extract solution (components shown as below) per well. The extraction was incubated at 45°C overnight. 100 µl isopropanol was added to each well and the plate was shaken at 250 rpm on a horizontal shaker at room temperature for 4 hrs. The plate was centrifuged at 3000 g at 4°C for 5 min. The pellets were then washed with 75% ethanol twice. After air drying at room temperature for 1 hour, 75µl H₂O was added to dissolve the pellets at 45°C overnight.

For T25 flask or 6-well plate, cells were trypsinized and transferred into an eppendorf tube. The cells were washed by 1 x PBS and treated by 700 µl extract solution. The extraction was incubated at 55°C overnight.

After incubation, the phenol Chloroform (Invitrogen 15593) extraction was performed, followed by ethanol precipitation.

Extract solution:

Protease K (50 mg/ml) [Add when use]	100 µl
SDS (20%)	0.5 ml
NaCl (5 M)	0.2 ml
EDTA (0.5 M)	0.3 ml
Tris (1 M, pH8)	0.5 ml
H ₂ O	to 10 ml

3.9 Colony PCR

After genome DNA isolation, colony PCR was used to select 1lox 129/Cast cell line. The recipe and program are described as follows:

Recipe:

Template	200 ng
ZP370 (25 µM)	0.5 µl
ZP389 (25 µM)	0.5 µl
ZP371 (25 µM)	0.5 µl
dNTP (New England Biolabs, N0447S, 2.5 mM)	4.5 µl
10xBuffer	5.0 µl
Taq enzyme	0.5 µl
dH ₂ O	To 50 µl

Program:

95°C 5 min -> (95°C 30 sec -> 56.5°C 20 sec-> 72°C 30 sec) x 35 -> 72°C
10 min -> 4°C hold

Primers involved in the PCR include:

ZP370 : 5' -CAAGACGTGCCATATCAGTGAGC-3' ;

ZP389 : 5' -TTTATAACCAAGGTTGAGAGAGCAAAG-3' ;

ZP371 : 5' -TCTGTCACCTACCAATGAGAGATCC-3' ;

3.10 RNA FISH:

DNA probe was labeled by nick translation kit (Roche 10976776001) according to the manufacture's instruction. The fluorescent dye used to label DNA probe in this study was Cy3-dUTP. Cells were harvested onto the microscope slides using Cytospin (Thermo WD7020) according to the manufacture's instruction. The RNA FISH was carried out following the previously published methods (Zhang, Huynh et al. 2007). In brief, the labeled probe was denatured at 75°C for 10 min and pre-annealed at 42°C for 10-12 min. The slides were dehydrated sequentially in room temperature with 80%, 90%, 100% ethanol for 2 min each. The slides were air-dried. 5.5 µl denatured and pre-annealed probe was applied onto each slide. The slides were then covered with cover-slips and were incubated in a humidified incubator at 42°C overnight. After incubation, the slides were

washed three times in 50% formamide/2 x SSC (300 mM NaCl, 30 mM Na₃Citrate) at 45°C for 5 min with shaking, followed by three times washing in 2 x SSC at 45°C for 5 min with shaking. 1 X PBS/ 0.2% Tween-20 was used to rinse the slides for 3 min and 6 µl Vectashield (containing DAPI) was applied to the slides. The slides were sealed with cover slip and were observed under fluorescent microscope.

For the *Xist* RNA FISH, Sx9 probe was adopted.

3.11 Amplification of padlock probe library:

Oligonucleotide libraries were synthesized by LC Sciences. The synthesized padlock probes library was amplified as previously published (Shoemaker, Deng et al. 2010).

The primers and oligonucleotide utilized to generate padlock probe library are shown as follows:

eMIP_CA_F: 5' – TGCCTAGGACCGGATCAACT–3'

eMIP_CA_R: 5' – GAGCTTCGGTTCACGCAATG–3'

To do expansion PCR, 10 µl 20 nM LC Sciences oligos was mixed with 0.80 µl 100 µM eMIP_CA_F, 0.80 µl 100 µM eMIP_CA_R and 100 µl 2 x KAPA SYBR MM. Nuclease free water was added to make the total volume as 200 µl.

Then real-time PCR was applied. The program is shown as follows:

95°C 30 sec -> (95°C 5 sec -> 52°C 60 sec-> 72°C 30 sec) x 5 -> (95°C 5 sec -> 60°C 30 sec-> 72°C 30 sec) x 10-12 -> 72°C 2 min -> 15°C hold

The first round PCR product was amplified to obtain more probes by using 0.2 µl 10 nM 1st round amplicons for 100 µl reaction, which contained 0.40 µl 100 µM eMIP_CA_F, 0.40 µl 100 µM eMIP_CA_R and 50 µl 2 x KAPA SYBR MM (KAPA Biosystems).

The 2nd amplification program is as follows:

95°C 30 sec -> (95°C 5 sec -> 60°C 30 sec -> 72°C 20 sec) x 13-15 -> 72°C 2 min ->15°C hold

The amplicons were purified by ethanol precipitation and Qiaquick column. After purification, 2-4 µg amplified amplicon was digested with 10 µl Nt.AlwI (New England Biolabs, R0627S, 10 U/µl) in total volume 200 µl at 37°C for 1 hr, followed by heat inactivation at 80°C for 20 min. Then the digestion mix was added with 10 µl Nb.BsrDI (New England Biolabs, R0648S, 10 U/µl) and incubated at 65°C for 1 hr. The nicked DNA was purified with Qiaquick columns and eluted with 50 µl EB buffer.

Single-strand padlock probes were purified by 6% TBE-Urea gel selection. Nicked DNA was mixed with 2 x TBE-urea sample buffer in 1:1 ratio. After denatured the mixture at 75°C for 8-10 min, the mixture was quickly placed on ice for at least 2 min. denatured DNA was loaded to the

gel and after gel running, 70 bp band was cut using a scalpel. The gel slide was chopped into small pieces and transferred to a new tube. 400-500 μ l 1 x TAE buffer was added to the gel slides followed by vigorously shaking on vortex at 37°C for 60 min. The solution was then collected using Nanosep column (PALL, ODM02C34) and centrifuged at 15,000 rpm for 3 min. To purify padlock probes, ethanol precipitation was applied. The DNA concentration was estimated by running 6% TBE-urea gel.

3.12 Padlock probe SNP capture:

The padlock probe SNP capture reaction was performed based on previously published paper (Shoemaker, Deng et al. 2010).

200 ng single-stranded cDNA (1.2 μ g for non-coding regions SNP capture) was mixed with 2 pmol padlock probes (4 pmol for non-coding regions SNP capture), 3.4 μ l 5 x Phusion HF buffer (New England biolabs, M0530S) and nuclease free water to make the total volume as 17 μ l. The mixture was heated at 95°C for 30 sec and cool down to 60°C at 0.02 °C/sec. At the same time, a TLN mix was made by mixing 0.2 μ l Ampligase (Epicentre, A3210K, 5 U/ μ l), 0.5 μ l Phusion HF taq (New England Biolabs, M0530S, 2 U/ μ l), 0.6 μ l 5x Phusion HF buffer, 0.2 μ l dNTP (1 mM each) and 1.5 μ l nuclease free water together (for non-coding regions SNP capture, the 1.5 μ l nuclease free water was replaced by 1.36 μ l 7.35 mM NAD⁺ (Sigma- Aldrich, N0632) and 0.14 μ l nuclease free water;

NAD⁺ was used to enhance the specifically binding of ampligase and DNA). After the SNP capture reaction was cooled down to 60°C, TLN mix was added to the reaction mixture and to prevent evaporating, 40 µl mineral oil (Sigma-Aldrich, M5904) was added into the tube. The reaction was performed at 60°C overnight. The reaction mixture was heated at 94°C for 2 min to denature the enzymes. 2 µl of exonuclease I /exonuclease III mix (New England Biolabs, M0293S, M0206S, 1:1 ratio) was added to remove linear DNA.

The circularized DNA was amplified by real-time PCR to synthesize amplicons with Solexa sequencing adaptors. The primers used for amplification are listed as follows:

CA-2-RA.Miseq:

5' –

AATGATACGGCGACCACCGAGATCTATCGGCTACACGCCTATCGGGAAGCTG

AAG-3' ;

CA-2-FA. Indx7Sol:

5' –

CAAGCAGAAGACGGCATAACGAGATGATCTGCGGTCTGCCATCCGACGGTAGT

GT-3' ;

CA-2-FA. Indx45Sol:

5' –

CAAGCAGAAGACGGCATAACGAGATCGTAGTCGGTCTGCCATCCGACGGTAGT

GT-3' ;

CA-2-FA.Indx76Sol :

5' -

CAAGCAGAAGACGGCATAACGAGATAATAGGCGGTCTGCCATCCGACGGTAGT

GT-3' ;

CA-2-FA.Indx91Sol :

5' -

CAAGCAGAAGACGGCATAACGAGATACATCGCGGTCTGCCATCCGACGGTAGT

GT-3' ;

CA-2-FA.Indx92Sol :

5' -

CAAGCAGAAGACGGCATAACGAGATTCAAGTCGGTCTGCCATCCGACGGTAGT

GT-3' ;

CA-2-FA.Indx93Sol :

5' -

CAAGCAGAAGACGGCATAACGAGATATTGGCCGGTCTGCCATCCGACGGTAGT

GT-3' ;

The real-time PCR program is shown as follows:

98°C 30 sec -> (98°C 10 sec -> 52°C 20 sec -> 72°C 15 sec) x 8 cycles ->
(98°C 10 sec -> 72°C 20 sec) x 22 cycles -> 72°C 3 min -> 15°C hold

The amplicons were then purified using AMPure beads (Beckman Coulter, A63880). The ratio of PCR product to AMPure beads was 5:9.

Primers which were used for Solexa sequencing were listed below:

Read1.Miseq: 5'-ATCGGCTACACGCCTATCGGGAAGCTGAAG-3' ;

IndexRead: 5'-ACACTACCGTCGGATGGCAGACCG-3' ;

Read2 (Zhang): 5'-CGGTCTGCCATCCGACGGTAGTGT-3' ;

3.13 KAPA Quantification

The quantification of deep sequencing samples is done by following the protocol of the Library Quantification Kit – Illumina (KAPA Biosystems, KK4844). The dilution buffer was made by mixing 500 µl 1 M Tris-HCl, pH8, 25 µl Tween-20 and 49,475 µl nuclease free water together. The qPCR/Primer mix was then prepared by adding 1 ml Illumina Primerpremix (10x) and 5ml KAPA STBR FAST qPCR MasterMix to make total volume as 6 ml. Dilution of samples was done in triplicates. The dilution factor was 50 for the first dilution and it was achieved by adding 98 µl dilution buffer to 2 µl stock sample. The second, third and fourth dilutions were achieved by mixing 10 µl previously diluted DNA with 90 µl dilution buffer. This dilution step finally made the dilution factor as 50000. To do quantitative PCR, reactions for 6 standards were set up by mixing 6 µl of the qPCR/Primer mix and 4 µl of KAPA DNA standard as one reaction. Each standard was set up in triplicate. Totally 18 reactions were set up for standard running. The reactions for 50000 x diluted samples were also set up by mixing 4 µl

diluted samples with 6 µl qPCR/Primer mix. The program for this quantitative PCR is shown as following:

95°C 5 min -> (95°C 30 sec -> 60°C 45 sec) x 35 -> 15°C hold

3.14 RNA Pol III inhibition

RNA polymerase III inhibitor ML-60218 (Merch Millipore, Cat# 557403) (CAS5777-84-91-9; (Wu, Pan et al. 2003) was used at the concentration of 30 µM in all experiments. For in vitro differentiation of 3F1 cells, the cells were treated with the inhibitor in ES cell differentiation medium for 2 days.

3.15 Isolation of bromouridine (BRU) labelled nascent RNA and reverse transcription

2loxT.S2 cells were treated with Pol III inhibitor for 1 hr and treated with the Pol III inhibitor together with 2 mM bromouridine (Sigma–Aldrich, 850187) for another 4 hrs. The BRU labelled RNA was isolated by following the instruction from (Paulsen, Veloso et al. 2014) with minor revise. To prepare the conjugated beads, 50 µl of anti-mouse IgG magnetic Dynabeads (Invitrogen, 110.33) per sample was added to fresh 1.5 ml microfuge tubes. Beads were captured with a magnetic stand (Invitrogen, 12321D) and the storage buffer was aspirated. 200 µl 0.1% Bovine Serum Albumin in DEPC-PBS (BSA, Roche, 03116999001) was added to each tube.

The tubes were flicked to resuspend beads and the beads were captured on the magnetic stand and supernatant was aspirated. This washing step was repeated two more times. After the final wash, each bead pellet was resuspended in 200 μ l 0.1% BSA in DEPC-PBS. 1 μ l RNaseOUT (Invitrogen, 10777-019) and 4 μ l (2 μ g) anti-BrdU antibody (BD Pharmingen, 555627) were then added to each tube. The mixture was incubated with gentle rotation for 1.5 hrs at room temperature. After incubation, the conjugated beads were washed 3 times with 200 μ l 0.1% BSA in DEPC-PBS as detailed above. After the final wash, beads were resuspended in 200 μ l 0.1% BSA in DEPC-PBS and 0.5 μ l RNaseOUT was added. Total RNA of each sample was isolated by following steps in 3.4. The isolated RNA was dissolved in 200 μ l DEPC-water, heated in an 80°C heat block for 10 min, and immediately put on ice for at least 2 min. 180 μ l of the RNA was added to the prepared beads. The tubes were flicked to ensure sample was mixed well. The samples were incubated with gentle rotation for 1.5 hrs at room temperature. After incubation, the beads were washed with 200 μ l 0.1% BSA in DEPC-PBS for 5 min on the rotator. This washing step was repeated two more times and the final wash solution was completely removed. 40 μ l DEPC-water was added to resuspend the bead pellet and the beads were incubated for 10 min in a 95°C heat block to elute BRU labelled RNA.

Reverse transcription of BRU labelled RNA was achieved by using SuperScript II (Invitrogen, 18064-014). For each sample, 8 μ l 5 x first – strand buffer, 1 μ l random primer (3 μ g/ μ l, Invitrogen, 48190-011) and 1

μl gene specific primer mix (GSP, 10 μM each) were mixed with 23.7 μl BRU labelled RNA. The mixture was incubated in a PCR machine at 85°C for 10 min and was immediately put on ice for at least 1 min. 4 μl 100 mM DTT, 0.8 μl 25 mM dNTP , 0.5 μl RNaseOUT and 1 μl Superscript II were added to the pre-mix. The samples were incubated using the following program:

25°C 10 min -> 42°C 50 min -> 70°C 15 min -> 4°C hold

The synthesized cDNA was directly used in a quantitative PCR experiment. The quantitative PCR was carried out on the CFX Connect real-time PCR system (Bio-Rad) using the SsoAdvanced Universal STBY Green Supermix (Bio-Rad, #172-5271).

Primers used in this section:

SNP35888319P1 : 5' -

TAACATAAATAAGTATCTCTACATCACTTACAATACCTGAGA-3'

SNP35888319P2 : 5' -AATATTGCTGCCTCAGCTGGTAGCT-3'

SNP35888319P3 : 5' -ACATGTCCTACAGACAGTGAATGCAT-3'

SNP35888319P4 : 5' -CCCATGGGCTTGCTCTTCCCC-3'

SNP35888319P5 : 5' -TCTCAAACATGCCCATGGGCTT-3'

tRNA^{LeuF} : 5' -CGCCAGACTCAAGCTATGGC-3'

tRNA^{Leu}R: 5' -TGTCAGAAGTGGGATTCGAACC-3'

tRNA^{Tyr}F: 5' -CCTTCGATAGCTCAGTTGGTAGAG-3'

tRNA^{Tyr}R: 5' -GGATTCGAACCAGCGACCTAAGGATATC-3'

SNP147087190F: 5' -AGCAGCAGGCAGAAGCCAGGACT-3'

SNP147087190R: 5' -AGTCAGCGGTGGAAACACAGTT-3'

SNP8441359F: 5' -CTATGGGACCCACAGCAGGC-3'

SNP8441359R: 5' -ACAGCCCTGTTGAGAGGCCCTCTCC-3'

SNP12807658F: 5' -AAACCCTCCCCCTTCAAGCCC-3'

SNP12807658R: 5' -AATGGATATTGTCACTTTGCGTGAATTGTGC-3'

SNP98673711F: 5' -ATACTGTGTGCAGAAAAGCCTGCTGA-3'

SNP98673711R: 5' -AGGGGAGCAGCCTTCAGGCCATACA-3'

GSPNC545_sqz: 5' -CTACTTTGTGATCCGTTGGAAGGCAGGAT-3'

SNP12344545F: 5' -GTAGCTCAGGCGAGAAGCACCTGC-3'

SNP12344545R: 5' -GGATGGCACACACAGCCTTTCAGC-3'

tRNA^{Leu}R, tRNA^{Tyr}R, SNP147087190F, SNP35888319P4, SNP8441359R, SNP12807658R, SNP98673711F and GSPNC545_sqz were used as gene-specific primersGSP in reverse transcriptionRT-PCR.

To concentrate cDNA, 3~5 reactions were combined together to proceed ethanol precipitation. cDNA pellet was dissolved in ~15 µl H₂O. 1~2 µl cDNA was taken to measure concentration using Quibit. 900 ng cDNA and 70 ng padlock probe were used for one SNP capture reaction to detect non-coding transcriptions.

3.16 Immunostaining

Slides were treated in 1xPBS with 0.5% Triton for 30 min, followed by washing with 1xPBS with 0.2% Tween-20 twice at 60 rpm for 2 min each time. Primary antibody was mixed with same volume 0.5% BSA (1:1) and the mixture was diluted 200 times in 1 x PBS with 0.2% Tween-20 (1:200). Diluted antibody was added onto the slide and incubated in a humid box at 4°C overnight. After incubation, the slide was washed in 1 x PBS with 0.2% Tween-20 at room temperature, 60 rpm three times, for 5 min each time. Secondary antibody was mixed with same volume 0.5% BSA (1:1) and the mixture was diluted in 1 x PBS with 0.2% Tween-20 (1:500). The antibody was added onto the slide and incubated in a humid and dark environment at room temperature for 2 hrs. The slide was then washed in 1 x PBS with

0.2% Tween-20 at room temperature, 60 rpm three times, for 5 min each time.

3.17 Chromosome conformation capture carbon copy (5C)

The 5C libraries were generated following the protocol described in the previous publication (Dostie, Zhan et al. 2007). The 5C primer was designed using a series of in-house written perl scripts. Totally 190 primers were designed to target an X chromosome region (ChrX: 70,831,000-71,821,000), including 100 forward primers and 90 reverse primers.

Primers used to amplify 5C ligation products are listed below:

5C-F: 5'-AATGATACGGCGACCACCGAGATCTACACGCTACACGCCATT
AACCCCTCACTAAAGGGA-3'

5C-R-index1: 5'-CAAGCAGAAGACGGCATACGAGATGATCTGCGGTC
TGCCATCCGCCCTATAGTGAGTCGTATTA-3'

5C-R-index2: 5'-CAAGCAGAAGACGGCATACGAGATTGGTCACGGTC
TGCCATCCGCCCTATAGTGAGTCGTATTA-3'

5C-R-index3: 5'-CAAGCAGAAGACGGCATACGAGATCACTGTGCGGTC
TGCCATCCGCCCTATAGTGAGTCGTATTA-3'

5C-R-index4: 5'-CAAGCAGAAGACGGCATACGAGATATTGGCCGGTC
TGCCATCCGCCCTATAGTGAGTCGTATTA-3'

3.18 shRNA knockdown

To knockdown RNA Pol III, a RNA interference system (OligoEngine, pSUPER RNAi System) was adopted. Two DNA oligos containing shRNA sequence targeting mouse general transcription factor IIIC, polypeptide 4 (Gtf3c4, gene ID: 269252) were designed. The two oligos were dissolved with ddH₂O to a concentration of 3 µg/µl. 1 µl of each oligo was combined with 48 µl annealing buffer (100 mM NaCl and 50 mM HEPES, pH7.4). The mixture was heated up to 90°C for 4 min, then at 70°C for 10 min, followed by slowly cooling down to room temperature within 20 min to finish annealing.

To prepare a backbone, vector pSUPER-hygro-sv40 was digested by BglII (New England Biolabs, R0144S) and HindIII (New England Biolabs, R0104S). After double enzyme digestion, a 5.1kb backbone was purified by gel extraction. The annealed oligos and the backbone were ligated by T4 DNA ligase (New England Biolabs, M0202S). 0.5 µl BglII was used to remove unmodified vectors. 2 µl of the ligation solution was used for transformation, followed by plasmid extraction. The yield was linearized by HindIII and the final DNA product was dissolved with ddH₂O to a concentration of 0.5-1 µg/µl. The DNA was then transfected into HR7 cells and 48 hrs post transfection, hygromycin B selection was started on the cells. The selection was lasted for 10 days. The established cell line is called HR7.shRNA.

DNA oligos containing desired shRNA sequence:

84573F: 5' – GATCCCCGCCAGCTCTTTAATATGTAATTCAAG
AGATTACATATTAAAGAGCTGGGCTTTTAA–3'

84573R: 5' –AGCTTAAAAAGCCAGCTCTTTAATATGTAATCTCT
TGAATTACATATTAAAGAGCTGGGCGGG–3'

3.19 Allele specific RT-PCR and southern blotting

Allele-specific RT-PCR was carried out following the protocol described in the previous publication (Huynh and Lee 2003).

Two primers (Sqz893StuIF and Sqz893StuIR) were designed and used to amplify a 90bp DNA fragment cross SNP12321893 of *Med14* (geneID: 26896) from GSP reverse transcribed cDNA. The PCR was carried out using ampliTaq® 360 DNA polymerase (ThermoFisher Scientific, 4398818) and the program is shown as follows:

95°C 2 min -> (95°C 30 sec -> 60°C 20 sec -> 72°C 20 sec) x 35 cycles ->
72°C 5 min -> 15°C hold

After the 35 cycles PCR, one additional cycle was performed to remove heteroduplex. The PCR product was desalted through illustra MicroSpin G-25 Columns (GE Healthcare, 27-5325-01) and the desalted

amplicons were digested by restriction enzyme *Stu*I (New England Biolabs, R0187S) at 37°C overnight.

Southern blotting was carried out according to the method described in (Southern 1975). Restriction enzyme digested product was loaded on a 4% agarose gel and gel running was performed at 100 Voltage, 90 min. The gel was denatured in transfer buffer (1 M NaCl, 400 mM NaOH) for 40 min. After denaturing, southern blot transfer tower was set up and DNA was transferred to an amersham hybond-N membrane (GE Healthcare, RPN303N). The membrane was then crosslinked by UV at 1,200 dosage. For hybridization, P³² labeled DNA probes were prepared according to the components below:

ddH ₂ O	9 µl
Oligo probe (10 µM)	2µl
10 x T4 polynucleotide kinase (PNK) buffer (M0201S)	2µl
T4 PNK enzyme (M0201S)	1 µl
ATP, [γ-32P]- 6000Ci/mmol 10mCi/ml , 250 µCi (PerkinElmer, BLU002Z250UC)	6 µl

The mixture was incubated at 37 °C for 1 hr, followed by illustra MicroSpin G-50 Columns (GE Healthcare, 27-5330-01) purification. Hybridization was carried out in a rolling hybridization bottle with P³²

labeled DNA probes and 20 ml church buffer (500 mM NaPhosphate, pH 7.5, 6.8% SDS, 1% BSA, 2 mM EDTA) at 50 °C overnight.

After hybridization, the membrane was washed by 2 x SSC, 0.1% SDS for 2 x 10 min and 0.1 X SSC, 0.1% SDS for 10 min. The membrane was then wrapped and was ready to be exposed to phosphor screen.

Primers used in amplification:

Sqz893StuIF: 5' -AGAATGGCCTTCAGTTCCTGAGGC-3'

Sqz893StuIR: 5' -GAAGATTGACCACTTATCAATAGA-3'

Southern blotting oligo probe:

Sqz893-sprobe: 5' -CTGGTGGGCTCTTGCGTGTACACTG-3'

3.20 Data analysis

To analyze the data of gene reactivation pattern, a simple method and a more sophisticated statistical method were developed. Since the analysis results from the two methods largely agree to each other. We choose to use the simple method for our data analysis. The simple method is described in the result section and the sophisticated statistical method, developed by Dr. Lin Bing Qing is described here:

First, for each gene, $r = \log_2 \left(\frac{r_e}{r_c} \right)$ was calculated where r_e is the ratio of the number of reads for the 129 allele to that for the *Castaneus* allele in the experiment sample and r_c is the ratio of the number of reads for the 129 allele to that for the *Castaneus* allele in the control sample. Genes were assumed as either “non-reactivated” with prior probability π_0 or “reactivated” with prior probability $\pi_1 = 1 - \pi_0$. Moreover, The r values of non-reactivated genes and of reactivated genes were assumed to have probability density functions $f_0(r)$ and $f_1(r)$, respectively.

Then, a two-component mixture model(Efron, Tibshirani et al. 2001) was built to depict r values of all the (reactivated or non-reactivated) genes,

$$f(r) = \pi_0 f_0(r) + \pi_1 f_1(r).$$

According to Bayes’ Theorem, the posterior probability that a gene is reactivated is given by

$$p_1(r) = \Pr(\text{reactivated} | r) = \pi_1 \frac{f_1(r)}{f(r)}$$

In order to compute this posterior probability, the prior probability π_0 , the null density $f_0(r)$ and the mixture density $f(r)$ were estimated.

The plotted distribution of r values from our experimental data is skewed with heavy tails. Therefore $f_0(r)$ and $f(r)$ were assumed to be both skew t-distribution density functions with unknown parameters. Under this distribution assumption, the function `mst.mle` in R package “sn”(Azzalini and Capitanio 2003) was directly applied to estimate $f(r)$ from all the observed r values.

A re-sampling method to estimate $f_0(r)$ was proposed. Re-sampling was performed by randomly selecting a set of genes without replacement, with the size being half the number of r values. When re-sampling, the chance of choosing the genes by their absolute r values was weighted. For the selected set of genes, the function `mst.mle` was applied to estimate the parameters of the skew t-distribution function. This procedure was repeated 100 times. The final estimation of the parameters for $f_0(r)$ is taken to be the median of the resulting 100 estimators.

To make sure a positive probability value $p_1(r)$ to be derived, $\pi_0 \leq \min_r \{f(r)/f_0(r)\}$ was required. Here, π_0 was simply estimated to be $\min_r \{f(r)/f_0(r)\}$. Let $\hat{\pi}_0$, $\hat{f}_0(r)$ and $\hat{f}(r)$ denote the estimates of π_0 , $f_0(r)$ and $f(r)$, respectively. Then the estimated posterior probability is

$$\hat{p}_1(r) = 1 - \hat{\pi}_0 \frac{\hat{f}_0(r)}{\hat{f}(r)}.$$

Finally, a gene was called as “reactivated” if its $\hat{p}_1(r)$ value is bigger than 0.8. Otherwise, it was called “non-reactivated”.

4 Results

4.1 Generation of mutant fibroblast cell lines with allele-specific deletion of *Xist* on the Xi

Mutant mice carrying a genetically engineered *Xist* allele for conditional gene knockout was generated in previous studies (Csankovszki, Panning et al. 1999). Two *loxP* sites were inserted into the *Xist* gene body: one at 5 kb upstream of the *Xist* promoter and the other at the intron 3 of *Xist*. The structure of the modified *Xist* gene is shown in Fig. 4A. This mutant X allele carries a genetic background of the mouse strain “129”. By hybridizing the mutant mouse with a wildtype Castaneous mouse, female mouse embryonic fibroblast (MEF) cell lines $X^{2lox-129}X^{WT-Cast}$ were isolated in a previous research (Zhang, Huynh et al. 2007). One of the two X chromosomes of this cell line is the wild type X chromosome carrying the genetic background of the mouse strain “Castaneous”. The other is the mutant X chromosome carrying the genetic background of the mouse strain “129”. (Note: to simplify the terms used in communication, we name the $X^{2lox-129}X^{WT-Cast}$ genotype cell as the “2loxT” cell)

In order to conditionally delete *Xist* gene of $X^{2lox-129}$, a plasmid, called “pSALK-Cre”, carrying the expression cassette of Cre-recombinase was transiently transfected into the 2loxT fibroblast cells. Cre-recombinase mediates the excision of the region between the two *loxP* sites. Since the transfection efficiency is not 100%, after transient transfection of Cre, the cell population is made of two possible genotypes: the 2loxT cells (cells failed to be modified) and $X^{1lox}X^{WT-Cast}$ (the “1lox” cells). Selection of the 1lox clones was achieved by limiting dilution. In this study, 842 candidate clones were isolated; and the genotype of each clone was analyzed by colony PCR (Fig.5B). Three primers were used to distinguish the genotypes of 1lox and 2loxT cell lines (Fig.5A&C). Six 1lox cell lines (HR1, HR5, HR6, HR7, HR8, and HR9) were detected in the initial PCR genotyping (Fig.6A&B). A final round of genotyping PCR was carried out to confirm their 1lox genotype (Fig.6C). Considering random XCI in fibroblast cells, we carried out *Xist* RNA FISH to distinguish the two possible phenotypes for a 1lox clone: $Xi^{1lox}Xa^{WT-Cast}$ and $Xa^{1lox}Xi^{WT-Cast}$ and select for the 1lox cell line, in which *Xist* is allele-specifically deleted from the Xi (Fig.7). All the six 1lox

cell lines established were confirmed as $Xi^{1lox}Xa^{WT-Cast}$. This result is consistent with the *Xce* (X-controlling element) strength of the two X alleles (Courtier, Heard et al. 1995). *Xce* is an X-linked DNA element playing a role at the “Choice” step of the random XCI. The X chromosome carrying a stronger *Xce* is more likely to remain active after XCI. It is known that, in an $X^{129}X^{Cast}$ cell, although XCI is random, the X^{Cast} allele has a much higher chance (~80-90%) to remain active. Therefore, it is not a surprising result that all the six established 1lox cell lines were confirmed as $Xi^{1lox}Xa^{WT-Cast}$. These cell lines are suitable for the subsequent studies.

Another aim of this study is to investigate whether the gene reactivation pattern of 1lox cells could be changed after long term cell culture. To obtain the long term cultured 1lox cells, HR5 and HR6 were cultured for 25 cell population doubling. The two established cell populations are named as HR5.D25 and HR6.D25.

To obtain a control cell line, six 2loxT cell lines were isolated by limiting dilution and they are named as 2loxT.S1 to 2loxT.S6, respectively. The *Xist* gene expression of these cell lines was allelotyped by allele-

specific reverse transcription and all these six 2loxT cell lines were confirmed as $\text{Xi}^{129}\text{Xa}^{\text{Cast}}$, as their *Xist* RNA is expressed from “129” allele (Fig.8). Cell line 2loxT.S2 was selected as the control cell line for the subsequent study.

Figure 5 Experimental designs of generating female fibroblast cell lines in which *Xist* is conditionally and allele-specifically deleted from the Xi.

A. The genetically engineered *Xist* allele generated by previous studies for conditional gene knockout (Csankovszki, Panning et al. 1999).

Zp370, Zp371&Zp389 are PCR primers used for genotyping. **B.**

Transient transfection of Cre results in three types of cells. **C.**

Predicted PCR genotyping results of “1lox” and “2loxT” clones.

Figure 6 PCR genotyping results of the established $X^{WT}X^{\Delta Xist}$ cell lines.

A. A representative gel image of the initial colony PCR results on limiting dilution clones using PCR primer pair Zp370 & Zp389. A potential 1lox clone was found in *lane3* (marked in the red box). **B.** Genotype of the potential 1lox clones was further confirmed using primer pair Zp370 & Zp371. **C.** Genotype confirmation of all the 6 established 1lox cell lines (HR1, 5-9) after they were expanded from the 96-well plate. Negative control: 2loxT cells. Positive control: III.3, a 1lox cell line with a confirmed $Xi^{1lox}Xa^{WT-Cast}$ genotype established by previous studies (Zhang, Huynh et al. 2007).

Figure 7 $Xa^{WT}Xi^{\Delta Xist}$ cell lines identified by *Xist* RNA FISH.

Cell nuclei were stained with Dapi (blue). The *Xist* RNA was detected by Cy3 labeled probes (red). 2loxT: $X^{2lox-129}X^{WT-Cast}$ cells. III.3, a 1lox cell line with a confirmed $Xi^{1lox}Xa^{WT-Cast}$ genotype established by previous studies (Zhang, Huynh et al. 2007). HR1, 5-9: 1lox cell lines established in this study. Arrow: an *Xist* RNA cloud. Asterisk: a tetraploid cell carries two *Xist* clouds per nucleus.

Figure 8 Allele-specific RT-PCR on Xist RNA of the 6 clones generated by limiting dilution of 2loxT cells (2loxTS1-S6). The genomic DNA of castaneous genetic background was used as the control.

4.2 The design of individual padlock probes

The basic structure of a padlock probe is shown in Fig.9. The padlock probe recognizes its cDNA target through the binding of its extension arm and ligation arm with the template cDNA. After binding with its cDNA template, the oligo forms a “padlock”-like structure. The 3’ end of the extension arm triggers DNA elongation by DNA polymerases (nucleotides are added from 5’ to 3’ on the 3’ end of the probe). After the DNA chain elongation reaches the 5’ end of the ligation arm of the padlock probe, the gap will be sealed by DNA ligase, and the padlock probe is “locked up” to form a circular DNA structure. The circular DNA structure can be further amplified using primers complementary to the “common linker” region of each padlock probe. After PCR amplification, the DNA sample was subjected to Solexa sequencing.

In this study, the length of the extension arm and the ligation arm of each padlock probe is ~18-20 nt (nucleotides). Each padlock probe was designed to target one SNP. As shown in Fig.9, each padlock probe will capture not only the SNP but also a 16 nt region and an 8 nt region

flanking the target SNP. We chose this design because although the DNA ligase does not require a specific DNA sequence for its “binding site”, it does have a small footprint about 12 nt on either side of its ligation site. For the different SNPs captured by one padlock probe, if the position of SNP is located within the footprint of the DNA ligase, it may slightly vary the working efficiency of the DNA ligase. This is a minor concern in theory. However, this concern can be eliminated by reserving a fixed distance between the target SNP and the 5'-end of the ligation arm. Because of this reason, each padlock probe in our probe library was designed to capture a ~25 nt region surrounding each target SNP.

Figure 9 The principle of SNP capture using Padlock probes.

4.3 The design of the padlock probe library

Two Padlock probe libraries were designed to target SNPs along the X chromosome. One targets X-linked gene coding regions, and the other targets non-coding regions, which include intergenic regions and introns.

To design padlock probe library that targets gene coding regions, the following five criteria were taken into consideration. (1) The target SNPs should locate along mouse X chromosome. (2) The SNP should be between mouse strain 129 and castaneous. (3) The target SNPs should locate inside gene coding regions. (4) The target SNP should not be located too close to the boundary of an exon, so that enough space is left between the SNP and its adjacent exon boundary for designing the two arms of the padlock probe. We name this as the “boundary effect”. (5) The space between a target SNP and its adjacent SNPs should not be too narrow, so that no SNPs interferes the binding of the extension arm and the ligation arm of the padlock probe to its template. We name this as the “neighbor effect”.

The mouse SNP data file (20110602-final-snps.vcf.gz) was obtained from Wellcome Trust Sanger institute database (ftp://ftp-mouse.sanger.ac.uk/current_snps/). This data file is in VCF format carrying the SNPs from 17 different mouse strains including castaneous and three substrains of “129”. SNPs from all the three substrains of 129 are included in the padlock probe library design because which substrain of 129 mouse that was adopted to generate the $X^{2lox-129}X^{WT-Cast}$ fibroblast cell line is unclear. In total, the data file downloaded from Sanger database contains 2,505,518 X-linked SNPs. Among them, 502,558 SNPs are confirmed with high genotype quality calls (ATG (above threshold genotype) score = 0 or 1) (Keane, Goodstadt et al. 2011, Yalcin, Wong et al. 2011) and are capable of distinguishing between strain 129 and castaneous. These SNPs were selected (Fig.10A). The mouse genome annotation was obtained from UCSC genome annotation database (<http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/>), which provides the position of each gene and its exons. 4857 SNPs located inside X-linked gene coding regions are selected after considering the shoulder and the neighbor

effects. To design padlock probes, the assembled X chromosome sequence (mm_ref_MGSCv37_chrX.mfa.gz) downloaded from NCBI database (ftp://ftp.ncbi.nih.gov/genomes/M_musculus/CHR_X/) was used as the reference genome. To further improve the quality of each probe design, we utilized a perl program package called “ppDesigner” (Diep, Plongthongkum et al.). This program helps to fine tune the length of each capture arm to provide a unified control on the T_m of each designed probe. It also utilizes another algorithm, called UNAFold (Markham and Zuker 2008), to eliminate over-complicated secondary structure of the DNA template of each padlock probe. 2,969 padlock probes are included in our final library design targeting gene coding regions (Fig.10A). These padlock probes are able to cover 1,110 X-linked genes among the 2,021 genes along the mouse X chromosome (Fig.10B).

To design padlock probe library targeting the non-coding regions of X chromosome, we started with those qualified SNPs that can distinguish between 129 and castaneous mouse strains. Among these SNPs, those located inside coding-regions were removed. ppDesigner was then used to

design qualified padlock probes targeting the rest SNPs. After the design, the list was narrowed down to 163,383. Two concerns were covered to further narrow down our list. First, the SNPs are not equally distributed. Different areas contain different numbers of SNPs. For example, between Position 4963172 to 4963356 (184 bp), there are five SNPs, but between Position 4979770 to 4984380 (4610 bp), there is no SNPs. If all the padlock probes generated by ppDesigner were used to detect SNPs, it may cause the bias that most reads are from SNPs enriched areas while transcripts from areas containing few SNPs may be underestimated. Secondly, given the vastness of non-coding regions, the SNPs we can target are numerous, but to profile the expression pattern of non-coding regions, only some representative SNPs are required. In that case, we selected SNPs with the frequency as two SNPs for every 5 kb. One targets the sense strand and the other targets the antisense strand. Finally, 6,361 padlock probes targeting sense strand and 6,495 padlock probes targeting antisense strand were selected to synthesize a padlock probe library (total number: 12,856) (Fig.10B).

A	SNP in X chromosome	2505518
	After ATG selection (ATG is 0 or 1)	
	129s1	2482676
	129p2	2480011
	129s5	2461909
	Cast	2367496
	SNP between cast and all 129 strains	502558
	SNP inside gene coding region	7032
	After considering boundary	5218
	After considering neighbor	4857
	After ppDesigner	2969

B

Figure 10 The established padlock probe library.

A. Summary of the padlock probe library design. **B.** The distribution of the padlock probes along the entire mouse X chromosome.

4.4 Padlock probe library synthesis and amplification

When a padlock probe was designed by ppDesigner, two 15bp common adapters were also added to the 5' and the 3' end of the padlock probe. To amplify the library, a pair of common primers was used. The common adapters contain two nickase digestion sites; thus, after amplification, the adapters can be removed by double nickase digestion (Fig.11).

Oligonucleotides of both coding and non-coding padlock probe libraries were synthesized by LC Sciences. Production PCR to obtain a sufficient amount of padlock probes was carried out as previously described (Zhang, Li et al. 2009). To avoid over amplification of the padlock probes, the production PCR was carried out using real-time PCR and was stopped at the linear range of the amplification curve. Thus, before the production PCR, a test PCR run was carried out to determine the linear range of the amplification curve (Fig.12). Interestingly, unlike the normal amplification curve which eventually reaches the plateau phase, a drooping amplification curve was observed (Fig.12). This observation is consistent with the heterogeneity of the PCR template (the

oligo library). During the early cycles of a PCR reaction, since the concentration of PCR primer is overwhelming, an annealing event almost always occurs between a template oligo and a PCR primer. During the later cycles of the PCR reaction, as the PCR primers are consumed and the PCR products accumulate, an annealing event becomes more and more likely to occur between two single-stranded PCR products. Because of the heterogeneity of the PCR products, such an annealing event produces a heteroduplex DNA fragment, in which the two ends (primer binding sites) are complimentary and the long middle portion is not. Since the fluorescent dye used in the real-time PCR reaction can only fluorescence after being inserted into complementary base pairs. The formation of heteroduplex DNA makes the amplification curve drop.

Based on this test PCR run, the cycle number chosen for the production PCR is 12 (Fig.13). The amplicons were next digested by nickases Nt. AwaI and Nb. BsrDI. After digestion, the single-stranded padlock probe was shorten to 70 nt. The digestion products were separated by denaturing PAGE gel containing 6% urea and the 70nt band was isolated by gel purification (Fig.14).

Figure 11 Schematic diagram of padlock probe library preparation process.

The composition of the padlock probe is shown: adapters, ligation arm, common linker and extension arm.

Figure 12 Amplification curves obtained from the heterogeneous padlock probe oligo library in a test PCR run.

Drooping amplification curves are consistent with the heterogeneity of a mixed-template.

Figure 13 Amplification curves obtained from the heterogeneous padlock probe oligo library in a production PCR run.

Reaction was stopped at the linear phase of amplification to avoid over amplification.

Figure 14 Gel analysis of single-stranded padlock probes after double nikase digetions.

The 70-nt long padlock probes and the 120-nt long complementary strands were electrophoretically separated in a denaturing gel.

4.5 SNP capture and sequencing samples preparation

cDNA samples of X-linked genes and non-coding transcripts were reverse transcribed using oligo(dT) and random primer mix respectively. Concerning the proportion of mRNA in the total RNA is only 1-3% and a large amount of cDNA was required for the SNP capture reactions, we started with 10 µg total RNA to reverse transcribe mRNA. To prepare cDNA for non-coding transcription targeting, the started total RNA was ~20 µg.

During the SNP capture reaction, each padlock probe attaches to the flanking regions of one SNP; the extension arm was elongated and ligated to the ligation arm to “lock” the padlock-like structure. The product from this reaction is a circular single-stranded DNA molecule, which bears the SNP of interest (Fig.15). After the SNP capture was finished, exonucleases (exonuclease I /exonuclease III) were used to remove linear DNA fragments such as the large amount of cDNA and non-functional padlock probes. To do deep sequencing, a series of index primers and a reverse primer were used to amplify short linear DNA fragments from the SNP capture products. Each index primer contains three parts: an adapter, a

barcode and a region complementary to the common linker of the circular SNP capture product. Two DNA polymerases (sso7d fusion polymerase (Biorad, 1725270) and KAPA SYBR FAST DNA polymerase (KAPABIOSYSTEMS, KK4600) were used to amplify the padlock capture products (Fig.16). KAPA SYBR FAST DNA polymerase is capable of strand displacement. When the DNA template is circular, the enzyme is able to finish one full circle of DNA chain elongation and then “pushes off” the PCR product from the circular template and continues the DNA chain elongation. The strand displacement activity is not a well-documented enzyme activity for the commercially available DNA polymerases. In our study, we were fortunate to find two enzymes, one with the strand displacement activity (KAPA), and one without (sso7d). Since the padlock capture products are the only circular DNA molecules in the padlock capture reaction, KAPA enzyme was used in optimizing the padlock capture experiments, as the PCR products are a series of DNA fragments corresponding to PCR extension around the circular SNP capture product 1, 2, 3 and 4 times (Fig.16), which can be used as a feature to confirm that padlock capture is working. The expected DNA fragment size of PCR extension around the circular template a single time is 175-179bp.

According to the established protocol (Diep, Plongthongkum et al. 2012), this DNA fragment is gel purified before being subjected to sequencing. In this study, we identified that sso7d is an Enzyme without strand displacement activity. Therefore, once the experimental conditions were optimized, padlock capture was carried out using sso7d and the PCR products were subjected directly for DNA sequencing (Fig.17). The padlock capture products for non-coding transcription contained unspecific bands which are shorter than the expected size of 175-179bp (Fig.17B). This was probably resulted from the complexity of the non-coding padlock probe library. To isolate the ~175 bp padlock capture products, gel extraction was done.

The deep sequencing equipment and run mode we adopted is the HiSeq 2500, rapid run mode. This mode is able to output 300 million reads for one run, ensuring that there are sufficient reads generated for analysis. On average, each sample can be covered with about 13-25 million sequencing reads.

Figure 15 Schematic diagram of SNP capture process, followed by amplification using solexa sequencing primers.

The expected PCR product size is between 174 to 179 bp, which is subjected to the consequent deep sequencing. Red: the ligation arm. Blue: the extension arm. Green: the common linker. Yellow: the captured region. Black: the binding of cDNA and capturing arms.

Figure 16 Amplification of a padlock probe SNP capture products for allelotyping X-linked genes.

Two different DNA polymerases were used to amplify the product. The ~175bp band produced by sso7d enzyme was subjected to the deep sequencing. The PCR products of KAPA Sybr DNA polymerase, show that the PCR template is a circular DNA molecule. Bands around 265 bp and 355 bp are corresponding to the PCR extension around the circle 2 and 3 times. The smallest fragment in the KAPA lane is a non-specific PCR product, as its size is smaller than the expected PCR product (~175bp).

Figure 17 Gel analysis of padlock probe SNP capture amplicons using sso7d fusion polymerase.

A. samples in the gene reactivation study. *Lane1:* 10bp ladder. *Lane2:* -RT, which is the amplification result of a SNP capture product using -RT as cDNA template. The band in the -RT lane is a non-specific PCR product, as its size is smaller than the expected PCR product (~175bp).

B. samples in non-coding transcription study.

4.6 Reactivation pattern study

The sequencing reads were aligned back to the assembled X chromosome (mm_ref_MGSCv37_chrX.mfa) using Bowtie2. We adopted the default setting of Bowtie2 but not allowed gaps in the alignment. In the study of gene reactivation, around 80% reads can be mapped back to the X chromosome. Reads that have more than one mapping positions were discarded and the rest were head counted using a SNP index list as reference. This index list contains the targeted SNP positions, SNPs of “129” allele as well as “Castaneous” allele. Ideally, one read can be either mapped to the “129” allele or mapped to “Castaneous” allele, contributing to the read counts of 129 or Castaneous, respectively. There is one other possibility, which is that a read neither belongs to the “129” read counts nor belongs to the “Castaneous” read counts. These reads took a very small proportion (~0.1%). These reads were regarded as sequencing errors. We used these reads to calculate sequencing error rate for each SNP. The sequencing error was deducted from the allelotyping data of the SNP.

In the study on the gene reactivation pattern, on average, each SNP was allelotyped by ~15,000 reads. There are a small fraction of SNPs covered by very few sequencing reads (for example, less than 10 reads). Including these SNPs into the data analysis certainly cannot provide meaningful information. Therefore, a detection threshold (10) was set to remove these SNPs from the data analysis. To study genes, read counts of the SNPs belonging to one gene were added together. The sum of the read count was considered as the read counts of the gene. We did this to all the cell samples and generated a gene read counts list. Genes that were not detected in any of the cell samples were removed from the list. 11 genes escaped XCI are identified in the previous study (Carrel and Willard 2005). These genes were removed from the list (Fig.18). 5 genes in the control cell sample (2loxT.S2) showing a larger head counts number of the “129” allele than the “Castaneous” allele were also removed from the list. The rest of the genes were used to analyze the reactivation pattern.

Gene	Synonym
Ddx3x	Utx
Kdm6a	
6720401G13Rik	
Eif2s3x	
Xist	
2010000I03Rik	Jpx/Enox
B230206F22Rik	Ftx
5530601H04Rik	
2610029G23Rik	
Kdm5c	Jarid1c
Mid1	

Figure 18 Eleven X-linked genes which escape XCI.

To prevent dividing by zero, a pseudo-count (10 reads) was set to replace all the read counts that were smaller than 10. To analyze the data of gene reactivation pattern, we developed a simple method (Minajigi, Froberg et al. 2015) and a more sophisticated statistical method. Since the analysis results for the two methods largely agree to each other, we chose to present the analysis result of the simple method. The sophisticated statistical method was described in the Material and Method section.

A ratio of the read counts of 129 allele to the read counts of Castaneous allele ($r = \text{Reads}^{129} / \text{Reads}^{\text{Cast}}$) was calculated to quantify the allele-specific expression pattern and another ratio ($R = r_e / r_c$, r_e : $\text{Reads}^{129} / \text{Reads}^{\text{Cast}}$ of a 1lox cell sample, r_c : $\text{Reads}^{129} / \text{Reads}^{\text{Cast}}$ of 2loxT.S2) was calculated to quantify gene reactivation status of the 1lox cell. If the value of R was bigger than 3, the gene in that 1lox cell sample would be considered as a potential reactivated gene. To rule out a small fraction of extreme data points, we also required $r > 1\%$ for a reactivated gene (a significant amount of transcription has to be detected from the 129 X allele).

On average, 32-57% of the targeted SNPs and 22-33% of the X-linked genes were allelotyped in each sample and overall 38 genes out of 309 profiled genes were determined as the reactivated genes in the six 1lox cell samples (Fig.19). Different cell lines have different reactivated genes and on average, about 4% of the X-linked genes were reactivated in each 1lox cell sample (Fig.20). Notably, 25 genes were reactivated in more than one 1lox cell sample and these genes were called “the frequently reactivated genes” in this study (Fig.21).

No.	Gene ID (UCSC Known Gene)	No.	Gene ID (UCSC Known Gene)
1	uc009tqi.2	20	uc009stl.1
2	uc009uss.2	21	uc009tbx.2
3	uc009tit.1	22	uc009tqc.1
4	uc009tcb.1	23	uc009tcc.1
5	uc012hkf.1	24	uc009sty.1
6	uc009sqy.1	25	uc009sma.1
7	uc009tzw.2	26	uc009smt.1
8	uc012hma.1	27	uc009upz.1
9	uc009uav.1	28	uc009urs.1
10	uc009tpe.2	29	uc009uxz.1
11	uc009uvm.1	30	uc009ujg.1
12	uc009ssq.1	31	uc009stp.2
13	uc009sqz.1	32	uc009tsv.1
14	uc009sra.1	33	uc009too.1
15	uc009ssf.1	34	uc009ufr.2
16	uc009txs.2	35	uc009utb.2
17	uc009tug.1	36	uc009uct.1
18	uc009tuf.1	37	uc009tqd.1
19	uc009uar.2	38	uc009ue.1

Figure 19 Reactivated genes detected in 1lox cell lines.

Padlock Allelotyping of Coding Region SNPs						
Cell Lines	Genotype	Cell Type	Mapped Reads	Allelotyped SNPs	Profiled Genes	Reactivated Genes
2loxT.S2	Xi ¹²⁹ Xa ^{Cast}	Fibroblast	25,158,923	667	NA	NA
HR1	Xi ^{129-1lox} Xa ^{Cast}	Fibroblast	13,543,143	338	207	9
HR5	Xi ^{129-1lox} Xa ^{Cast}	Fibroblast	15,114,424	453	234	9
HR6	Xi ^{129-1lox} Xa ^{Cast}	Fibroblast	17,523,524	404	214	5
HR7	Xi ^{129-1lox} Xa ^{Cast}	Fibroblast	18,631,601	1051	298	21
HR8	Xi ^{129-1lox} Xa ^{Cast}	Fibroblast	23,947,887	947	297	23
HR9	Xi ^{129-1lox} Xa ^{Cast}	Fibroblast	19,718,479	827	292	10
HR5.D25	HR5 after long term cell culture	Fibroblast	18,754,821	556	256	9
HR6.D25	HR6 after long term cell culture	Fibroblast	15,232,743	507	250	10

Figure 20 The summary of allelotyping read counts of the gene reactivation study.

Gene_ID	HR1	HR5	HR6	HR7	HR8	HR9	HR5D25	HR6D25	Frequency
uc009tqi.2	0	1	1	1	1	1	1	1	7
uc009tit.1	1	0	0	1	1	1	0	1	5
uc009uss.2	0	1	0	1	1	0	1	1	5
uc009tcb.1	0	1	1	0	1	0	1	1	5
uc009tpe.2	1	0	1	0	0	0	1	1	4
uc009uav.1	0	0	0	1	1	1	0	1	4
uc012hkf.1	1	0	0	1	0	0	1	1	4
uc009sqy.1	1	0	0	1	1	0	1	0	4
uc012hma.1	0	1	0	0	1	1	1	0	4
uc009tzw.2	0	1	1	0	1	1	0	0	4
uc009sra.1	0	0	0	1	1	1	0	0	3
uc009ssf.1	1	0	0	1	0	1	0	0	3
uc009txs.2	0	0	0	1	1	1	0	0	3
uc009tug.1	1	0	0	0	1	0	1	0	3
uc009sqz.1	0	0	0	1	1	1	0	0	3
uc009uvm.1	0	1	0	1	1	0	0	0	3
uc009ssq.1	0	1	0	1	1	0	0	0	3
uc009sty.1	0	0	0	1	1	0	0	0	2
uc009tbx.2	0	0	0	0	1	0	0	1	2
uc009uar.2	0	0	0	1	1	0	0	0	2
uc009stl.1	1	0	0	1	0	0	0	0	2
uc009tcc.1	1	1	0	0	0	0	0	0	2
uc009sma.1	0	0	0	0	1	0	1	0	2
uc009tuf.1	0	0	0	1	0	0	0	1	2
uc009tqc.1	0	1	0	1	0	0	0	0	2

Figure 21 The 25 frequently reactivated genes.

1: Gene was reactivated in the corresponding cell line. Frequency: the number of cell lines in which the gene was reactivated.

The distribution of the reactivated genes was aligned with genomic features, such as repetitive DNA elements and *Xist* RNA coating pattern along the X chromosome (Fig.22). The *Xist* RNA coating distribution was replotted from the data generated by previous studies (Engreitz, Pandya-Jones et al. 2013, Simon, Pinter et al. 2013). The data of repetitive DNA elements were adopted from the RepeatMasker track file (chrX-rmsk.txt), which was downloaded from the UCSC website (<http://hgdownload.soe.ucsc.edu/downloads.html>).

The correlation between different genomic features and the gene reactivation pattern along $Xi^{129-1lox}$ is measured by Spearman's rank correlation coefficient. To calculate the Spearman correlation coefficient, we divided the mouse X chromosome into non-overlapping windows at 8 x 106 bp resolution. The alignment reveals that most reactivated genes locate in the gene-rich areas and the reactivation pattern is positively corresponded to *Xist* RNA coating (Spearman's correlation = 0.59) and SINE density, for example mouse SINE family B2 (Spearman's correlation = 0.67), Alu (Spearman's correlation = 0.64), ID (Spearman's correlation =

0.63) and B4 (Spearman's correlation = 0.61) (Fig.23). The four mouse SINE families are described in the previous publications (Okada 1991, Mouse Genome Sequencing, Waterston et al. 2002). Conversely, a small value of the Spearman coefficient between L1 and the reactivated genes was observed (-0.13). As it was mentioned in the introduction section, L1 was predicted to serve as "way station" DNA element facilitating the *Xist* RNA spreading. However, in our study, the distribution of L1 does not have a significant correlation with the reactivated genes.

Figure 22 The alignment of reactivated genes with padlock probes targeting genes and other genomic features along mouse X chromosome.

Ch X: the X chromosome ideogram, indicating the loci of *Hprt* and *Xist*. *Xist* RNA: the distribution of *Xist* RNA coating density in mouse lung fibroblast cells, replotted from data generated by (Engreitz, Pandya-Jones et al. 2013). L1, Alu and B4: the distribution of repetitive DNA elements (LINE1/Alu/B4). HR1-HR9: the distribution of reactivated genes, each bar represents one gene. Red bars: frequently reactivated genes. Padlock probes: the distribution of padlock probes targeting the gene coding regions. Genes: UCSC genes.

No.	Genomic Features	Spearman coefficient	No.	Genomic Features	Spearman coefficient
1	B2	0.67	22	Y-chromosome	0.25
2	Alu	0.64	23	MIR	0.24
3	ID	0.63	24	Genie	0.23
4	B4	0.61	25	CR1	0.22
5	MaLR	0.60	26	MER2_type	0.20
6	Xist RNA (CHART-seq)	0.59	27	LTR	0.16
7	L2	0.53	28	TcMar	0.15
8	Tc2	0.53	29	ERV1	0.15
9	tRNA	0.52	30	Gypsy	0.14
10	snRNA	0.50	31	AcHobo	0.12
11	Xist RNA (RAP)	0.48	32	Tip100	0.10
12	MuDR	0.44	33	Tigger	0.08
13	ERVL	0.41	34	srpRNA	0.07
14	Mariner	0.39	35	Satellite	0.02
15	RTE	0.38	36	TcMar	0.01
16	MER1_type	0.35	37	Charlie	-0.10
17	rRNA	0.35	38	L1	-0.13
18	hAT	0.33	39	ERV1	-0.22
19	scRNA	0.33	40	ERVK	-0.30
20	Simple_repeat	0.32	41	Low_complexity	-0.31
21	ERV	0.31	42	Helitron	-0.32

Figure 23 The spearman coefficient values between reactivated genes and different genomic features.

The list is ranked by the coefficient value in descending order.

4.7 SINEs are enriched along the upstream regions of the reactivated genes in comparison to the genes remain silenced

The reactivation pattern we found suggests that the gene reactivation along $Xi^{\Delta Xist}$ is correlated to the genomic features such as SINE density and Xist coating density. However, this correlation could be a consequence of high gene density in gene-rich areas. Obviously, gene reactivation is more likely to occur along the gene-rich areas and these areas are known to be enriched for SINE density (Batzner and Deininger 2002) and Xist coating density ((Batzner and Deininger 2002, Engreitz, Pandya-Jones et al. 2013, Simon, Pinter et al. 2013)

To rule out this possibility, the copy number of SINEs in the upstream 20kb and downstream 20kb regions of the profiled genes was analyzed. The SINE copy number along the upstream and downstream regions of these genes were counted in the frequency of every 3kb. To present the density of SINE within the gene body, the average SINE copy number of a 3 kb intronic region of a gene was counted.

Strikingly, SINEs, e.g. B4 and Alu were found significantly enriched along the upstream regions of the reactivated genes compared to the genes remained silenced (Fig.24 & Fig.25, the yellow-highlighted region). Collectively, the gene reactivation along $Xi^{\Delta Xist}$ is rare but not sporadic. A gene with SINE enrichment along its upstream region is prone to be reactivated.

Figure 24 Distribution of SINE elements along the X-linked genes.

The element numbers in the upstream and downstream of each X-linked gene were counted for every 3kb with a step size of 1 kb. To represent the distribution of SINE in the gene region, an average copy number of a 3kb intronic region of each gene was calculated. The gene body is represented as a single point along the X-axis at position "0". Position 10k indicates the position 10k downstream of the gene's transcription stop site. Position -10k indicates the position 10k upstream of the gene's transcription start site. *Data* are shown as mean $\pm$ S.E.M. Box plots of SINE copy numbers of the 10 kb DNA region located 15-5 kb upstream of X-linked genes (yellow-highlighted region) is shown. P-value was calculated by Wilcoxon Test.

Figure 25 Distribution of sense/antisense SINE elements along the X-linked genes.

The statistic method was described in Fig.24. The gene body is represented as a single point along the X-axis at position "0". Position 10k indicates the position 10k downstream of the gene's transcription stop site. Position -10k indicates the position 10k upstream of the gene's transcription start site. Data are shown as mean $\pm$ S.E.M. Box plots of SINE copy numbers of the 10 kb DNA region located 15-5 kb upstream of X-linked genes (yellow-highlighted region) is shown. P-value was calculated by Wilcoxon Test.

4.8 Gene reactivation pattern in long term cultured cell lines

The purpose to study the gene reactivation pattern in long term cell culture is to find out whether the epigenetic lesion caused by *Xist* deletion can be enlarged during long term cell culture. However, limited effect of long term culture was observed. The gene reactivation level remained low in the two long term cultured cell populations. Moreover, long term cell culture could either reactivate a silenced gene or silence a reactivated gene (Fig.26). 11 genes were reactivated after long term cell culture and 6 genes were re-silenced. The result suggests that the epigenetic status of $Xi^{\Delta Xist}$ is not duplicated with 100% fidelity during each cell cycle.

Selection effect could be one cause to explain the re-silencing phenomenon. It is possible that during high generation growing, cells with some specific genes reactivated are more likely to be dysfunctional. Thus, during long term cell culture, this selection effect might gradually eliminate cells carrying these reactivation events, although no obvious cell death was observed. In summary, long term cell culture has very limited effect on both gene reactivation and re-silencing.

Figure 26 The distribution of gene reactivation of HR5, HR6, HR5.D25 and HR6.D25.

Ch X: the X chromosome ideogram, indicating the loci of *Hprt* and *Xist*. HR5: the reactivated X-linked genes in HR5 cell line. HR5.D25: the reactivated X-linked genes in HR5.D25 cell population. HR6: the reactivated X-linked genes in HR5 cell line. HR6.D25: the reactivated X-linked genes in HR6.D25 cell population. Black bars indicate the genes are reactivated in 1lox cells and are not re-silenced after long term cell culture. Blue bars indicate the genes were silenced after long term cell culture. Red bars indicate the genes were reactivated after long term cell culture.

4.9 Non-coding transcription study on XCI

In the X chromosome non-coding transcription study, two mouse female ES cell lines named as Ls2, which was established by my lab members and EL16.7 (Lee, Davidow et al. 1999) were also involved. Both Ls2 and EL16.7 carried one X chromosome with the genetic background of 129 and the other X chromosome with the genetic background of Castaneous ($Xa^{129}Xa^{Cast}$). In total, three cell types were included in the non-coding regions expression study ($Xa^{129}Xa^{Cast}$, $Xi^{129}Xa^{Cast}$ and $Xi^{129-1lox}Xa^{Cast}$). Each cell type was represented by two cell lines (Fig.27).

After deep sequencing, the deep sequencing reads were mapped back to the whole mouse genome as a clearance step. Reads that can be multiply aligned to the genome were removed, ensuring the transcripts for subsequent analysis are specifically transcribed from the X chromosome. Using Bowtie2, on average, about 30% of the total reads are uniquely aligned to the X chromosome. Among 12,856 designed padlock probes, 396 in total detected transcripts from the non-coding regions. On average, ~250 non-coding SNPs were allelotyped in each cell line and each SNP was allelotyped by ~5,000 reads (Fig.27).

Padlock Allelotyping of Non-coding Region SNPs				
Cell Lines	Genotype	Cell Type	Mapped Reads	Allelotyped SNPs
EL16.7	Xa ¹²⁹ Xa ^{Cast}	ES Cell	823897	227
LS2	Xa ¹²⁹ Xa ^{Cast}	ES Cell	1705408	278
HR6	Xi ^{129-1lox} Xa ^{Cast}	Fibroblast	1140780	227
HR7	Xi ^{129-1lox} Xa ^{Cast}	Fibroblast	1430764	262
2loxT.S1	Xi ¹²⁹ Xa ^{Cast}	Fibroblast	1216252	233
2loxT.S2	Xi ¹²⁹ Xa ^{Cast}	Fibroblast	1277267	283

Figure 27 The summary of allelotyping read counts of the non-coding transcription study.

In the comparison to gene coding regions expression, non-coding transcription is rare. To analyze the data, we define two terms, “allelotype” and “expression pattern”. A SNP’s allelotype is the SNP’s allele-specific expression status in a given cell line, which includes “129-specific”, “cast-specific”, “biallelic”, “undetermined” or “undetected” (read counts of both alleles < detection threshold) (Fig.28). If a SNP’s allelotype is detectable in all three cell types, the SNP’s “expression pattern” can be determined. The expression pattern of a SNP could be either *Xist*-dependent or *Xist*-independent (Fig.29). If a transcription is biallelic in ES cells ($Xa^{129}Xa^{Cast}$) and become cast-specific in fibroblasts ($Xi^{129}Xa^{Cast}$), the transcription is silenced by *Xist* (*Xist*-dependent). All the other possible expression patterns were considered as *Xist*-independent. In our study, we did not detect any non-coding transcription with the expression pattern similar to *Xist*. Therefore, if the expression pattern of a SNP is *Xist*-dependent, the non-coding transcription is silenced by the *Xist* RNA, and it either remains silenced ($r_e/r_c < 3$) or is reactivated ($r_e/r_c \geq 3$) along $Xi^{\Delta Xist}$.

Strikingly, completely different from X-linked genes, which are exclusively cast-specific in the control fibroblasts (2loxT.S2) ($Xi^{129}Xa^{Cast}$), the allelotype of non-coding transcription showed a high degree of allelic bias and independence from *Xist*-mediated transcriptional silencing (Fig.30).

When we further analyzed the expression pattern of each SNPs, since two different cell lines were included for each cell type, we analyzed the data by classifying them into four different combinations (“Combo”) and each combo includes three cell types ($Xa^{129}Xa^{Cast}$, $Xi^{129}Xa^{Cast}$ and $Xi^{129-1lox}Xa^{Cast}$) (Fig.29). Most of the SNPs (85) showed *Xist*-independent expression pattern (Fig.31 & Fig.33) and these SNPs were highly consistent among the four combinations of data analysis (Fig.31). Only 12 SNPs were *Xist*-dependent (Fig.32). Notably, among the 12 identified *Xist*-dependent SNPs, 11 locate within introns and along the coding strands of the resident genes, strongly suggesting these detectable transcripts were unspliced nascent transcripts of X-linked genes (Fig.32). This result further confirmed the sensitivity and consistency of the experiments.

Calculation & Threshold	Allelotype
$r = \text{Reads}^{129} / \text{Reads}^{\text{Cast}}$	
$r > 3$	129-specific
$r < 0.3$	Cast-specific
$0.5 < r < 2$	Biallelic
$0.3 < r < 0.5$ & $2 < r < 3$	Undetermined
Calculation & Threshold	Expression Pattern
$R = \frac{r \text{Xi}^{129-1\text{lox}} \text{Xa}^{\text{Cast}}}{r \text{Xi}^{129} \text{Xa}^{\text{Cast}}}$	
$R \leq 3$	Silenced
$R > 3$	Reactivated
Detection Thresholds = 10	
Pseudo-count = 10	

Figure 28 The calculations and thresholds used to define allelotype and expression pattern.

The allelotype is determined by the ratio of the 129 read counts to the Castaneous read counts. The expression pattern is determined by dividing of r of 1lox cell to r of 2loxT cell. The ratio is represented as R.

	$Xa^{129}Xa^{Cast}$	$Xi^{129}Xa^{Cast}$	$Xi^{129-1lox}Xa^{Cast}$	Expression Pattern
Tested Cell Lines	EL16.7	2loxT.S1	HR6	NA
	LS2	2loxT.S2	HR7	NA
Allelotypes and Expression Patterns	Biallelic	Cast-specific	$R \leq 3$	Silenced <i>Xist</i>-dependent
	Biallelic	Cast-specific	$R > 3$	Reactivated
	All the other possible expression patterns as far as the SNP is allelotyped in all three kinds of cells			<i>Xist</i>-independent
Combo 1	LS2	2loxT.S1	HR6	NA
Combo 2	EL16.7	2loxT.S1	HR6	NA
Combo 3	LS2	2loxT.S2	HR6	NA
Combo 4	EL16.7	2loxT.S2	HR6	NA

Figure 29 The definitions of allelotype and expression pattern, which are used to describe the allele-specific expression profile of X-linked non-coding SNPs.

The examined cell lines contain three genotypes, $Xa^{129}Xa^{Cast}$, $Xi^{129}Xa^{Cast}$ and $Xi^{129-1lox}Xa^{Cast}$. Each genotype was represented by two cell samples. The *Xist*-dependent expression patterns are determined by the expressions of all three genotypes. The data from different cell lines were combined in four different ways to test the consistency of the experiments.

Figure 30 The non-coding SNPs and the X-linked genes allele-specific expression profile of “129” allele to “Castaneous” allele in $\text{Xi}^{129}\text{Xa}^{\text{Cast}}$ cells.

Expression ratio of 129/Cast was taken for 281 non-coding transcripts and 371 coding transcripts of 2loxT cells.

Figure 31 A venn diagram showing the expression patterns between the four combos of data sets described in Fig.29.

The Venny Diagram was generated using (Oliveros, J.C. (2007) VENNY. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.)

SNP Position	Expression Pattern	Intronic?	Coding-Strand?
12278498	Silenced	Yes	Yes
20286300	Silenced	Yes	Yes
39515436	Reactivated	Yes	Yes
49017955	Silenced	Yes	Yes
71198141	Silenced	Yes	Yes
91346283	Silenced	Yes	Yes
98401546	Silenced/Reactivated	Yes	Yes
99573768	Reactivated	Yes	Yes
148710921	Silenced	No	NA
156023819	Silenced	Yes	Yes
156141156	Silenced	Yes	Yes
161425885	Silenced	Yes	Yes

Figure 32 Non-coding SNPs with *Xist*-dependent expression pattern from the data analysis of four different combinations of data sets.

Whether the transcripts cross these SNP positions are intronic and whether they are of the same orientation as the coding strand are shown.

No.	SNP Position	Strand	No.	SNP Position	Strand	No.	SNP Position	Strand
1	5,277,117	+	30	74,771,919	-	58	130,809,258	+
2	7,227,208	-	31	75,813,496	+	59	131,227,063	-
3	7,768,502	+	32	75,870,843	-	60	131,304,049	-
4	8,239,774	+	33	77,780,358	+	61	133,464,368	+
5	8,845,449	-	34	85,650,688	-	62	133,539,347	+
6	12,344,545	-	35	90,818,736	+	63	137,356,878	+
7	12,807,658	+	36	91,423,174	+	64	146,819,569	+
8	12,879,544	-	37	92,018,272	+	65	146,877,146	+
9	13,373,553	+	38	93,117,207	+	66	146,974,711	+
10	17,833,105	-	39	97,175,901	-	67	147,017,388	+
11	18,519,844	-	40	98,292,961	+	68	147,029,398	-
12	19,816,795	-	41	98,395,512	+	69	147,087,190	-
13	20,235,745	-	42	98,673,711	-	70	147,809,926	-
14	20,435,354	-	43	99,496,460	+	71	148,004,349	-
15	20,634,135	+	44	99,557,757	+	72	148,468,717	-
16	34,158,892	+	45	99,595,155	+	73	148,545,040	+
17	34,169,014	+	46	100,496,027	+	74	148,756,047	+
18	35,535,031	-	47	100,502,644	+	75	149,833,356	-
19	35,888,319	+	48	100,580,718	-	76	151,938,596	+
20	39,428,600	-	49	101,701,054	-	77	156,023,004	+
21	39,505,390	-	50	101,817,822	+	78	157,237,893	+
22	47,911,112	+	51	101,857,200	+	79	157,309,699	+
23	49,502,374	+	52	102,677,289	-	80	159,220,820	+
24	51,187,025	+	53	118,449,334	+	81	160,380,742	+
25	64,979,833	-	54	126,884,955	-	82	165,517,285	-
26	70,691,121	+	55	126,903,279	-	83	165,806,562	-
27	71,552,769	-	56	130,179,355	+	84	166,173,763	-
28	72,632,962	-	57	130,452,728	-	85	166,446,176	+
29	72,966,295	+						

Figure 33 The 85 non-coding region SNPs with *Xist*-independent expression pattern.

The mouse chromosome X DNA sequence (mm_ref_MGSCv37_chrX.mfa.gz) was used as the reference genome. The “Strand” column indicates the orientation of the detected transcription. A “+” sign indicates that the detected RNA transcript carries the same sequence as the reference genome. A “-” sign indicates that the detected RNA transcript carries the complementary sequence of the reference genome.

To avoid subjectivity of threshold choice in this analysis, the data were re-analyzed using a different set of thresholds (Fig.34). The two data analysis results were highly consistent (Fig.35). To rule out the possibility that RNA editing (Danecek, Nellaker et al. 2012) interferes data interpretation, we compared the positions of all the allelotyped non-coding SNPs with the RNA editing sites collected from 15 mouse strains (Danecek, Nellaker et al. 2012). No match was found. To rule out the possibility of DNA contamination in the RNA samples, we included –RT controls in all our experiments (data not shown). Moreover, DNA contamination is a random and accidental event. The fact that a same set of Xist-independent non-coding SNPs was consistently detected from different RNA samples cannot be explained by DNA contamination.

Collectively, the allelic expression of transcriptions along non-coding regions of the X chromosome can be studied by padlock probe method. The transcriptions show a high degree of allelic bias and are largely independent of Xist-mediated transcriptional silencing.

The Thresholds Used in This Figure	
Calculation & Threshold	Allelotype
$r = \text{Reads}^{129} / \text{Reads}^{\text{Cast}}$	
$r > 4$	129-specific
$r < 0.25$	Cast-specific
$0.4 < r < 2.5$	Biallelic
Calculation & Threshold	Expression Pattern
$R = \frac{\sum_i X_i^{129-110X} X_a^{\text{Cast}}}{\sum_i X_i^{129} X_a^{\text{Cast}}}$	
$R \leq 3$	Silenced
$R > 3$	Reactivated
Detection Thresholds = 15	
Pseudo-count = 5	

Figure 34 A different threshold used in analyzing transcription from non-coding regions.

Definitions of “allelotype” and “expression pattern” used to describe the allele-specific expression profile of an X-linked non-coding SNP. Cell lines of three different genotypes were used in the study. Two cell lines were selected for each genotype. The data from different cell lines were formed in different combinations to test the consistency of the experiments.

Figure 35 Data analysis result using the threshold shown in Fig.34.

A. A Venn diagram showing the non-coding SNPs with *Xist*-independent expression patterns from the data analysis of four different combinations of data sets. **B.** Non-coding SNPs with *Xist*-dependent expression pattern from the data analysis of four different combinations of data sets.

4.10 RNA Polymerase III transcription shows chromatin restructuring function

The alignment between the *Xist*-independent SNPs and Alu density again reveals a significant positive correlation (Spearman's correlation = 0.72) (Fig.36). We next tested whether some of these *Xist*-independent transcripts were directly from SINEs. By reverse transcription, the range of the transcription across SNP35888319 was narrowed down to a ~400bp region (Fig.37A), which is a typical SINE transcript length. The region contains a putative SINE element (Fig.37A). SINEs are retrotransposons transcribed by Pol III (Batzner and Deininger 2002). Using BRU labeling method and quantitative PCR, nascent transcription across this SNP position can be inhibited by a Pol III specific inhibitor (ML-60218) (Fig.37B). In addition to SNP35888319, the transcriptions across five other non-coding SNPs were also repressed by the Pol III inhibitor (Fig.37B). Moreover, padlock probe library targeting X-linked non-coding regions was used to capture nascent BRU RNA before and after Pol III inhibitor treatment. The Pol III inhibitor showed a clear inhibitory effect on the non-coding transcription along the entire X chromosome effect (Fig.37C). Putting together, the results confirm the non-coding transcripts we detected are Pol III products.

Figure 36 The distribution of the detectable *Xist*-independent non-coding regions transcripts, aligned with *Xist* RNA coating and Alu density.

Ch X: the X chromosome ideogram, indicating the loci of *Hprt* and *Xist*. Alu: the distribution of Alu along the X chromosome. Padlock probes: the distribution of padlock probes targeting the non-coding regions. Genes: UCSC genes.

Figure 37 The non-coding transcriptions are Pol III products.

A. RT-PCR to check the range of the non-coding transcription across SNP35888319. **B.** Quantitative RT-PCR to check the expression level of five selected non-coding transcriptions from 2loxT.S2 cells treated with the Pol III inhibitor. Two tRNA precursors (Leu and Tyr) were used as positive controls and the result collected from biologically triplicated experiments were normalized to an internal control (*Gapdh*). Data are shown as mean $\pm$ S.E.M of biological triplicate. The statistical analysis used is the Student's *t*-test. One asterisk indicates *P*-values smaller than 0.05. **C.** SNP capture result of X-linked non-coding transcripts from 2loxT.S2 cells treated with the Pol III inhibitor. Arrowhead indicates the reduced captured product.

Pol III transcripts are short, for example, tRNAs and SINEs. They may come from individual nucleosomes or regions, which are possibly located at special positions “impervious” to or independent of the influence of the neighboring higher order chromatin structure (“ihoc”, independent of higher order chromatin structure) (Fig.38). *Xist*-mediated transcriptional silencing depends on a compact higher order chromatin structure. Therefore, the *Xist*-independent expression of ihoc Pol III transcripts can be explained (Fig.38). Furthermore, we assume that ihoc nucleosomes located near a gene may influence the packaging of the gene into its higher order chromatin structure, which may explain the positive correlation between detectable reactivated genes distribution and SINE density as well as the enrichment of SINE in the upstream of the reactivated genes. To test this hypothesis, we carried out allele-specific RT PCR targeting on a gene *Med14* (uc009sqz.1), which was reactivated in HR7 cells. By inhibiting Pol III transcriptions located in the upstream region of this reactivated gene (SNP12344545) (Fig.37B), this gene was re-silenced. The re-silencing effect can be observed in the cells treated with Pol III inhibitor overnight or transfected with shRNA against TFIIIC (transcription factor for polymerase III C, a key transcription factor in the Pol III transcription machinery) (Fig.39).

Figure 38 A model of *Xist*-independent Pol III transcription from “ihoc” nucleosomes/regions regulating the epigenetic architecture of the Xi.

Figure 39 Allele-specific RT-PCR on Med14 (uc009sqz.1), a reactivated gene along the Xi^{129-Δ*Xist*} allele in HR7 cells.

129/Cast: density ratio of the 129 allele to the Cast allele within lanes.

Immunostaining of Pol II and Pol III combined with RNA FISH was also performed (Fig.40). Two mouse monoclonal antibodies (Santa Cruz, sc-47701 for Pol II and Santa Cruz, sc-47701 for Pol III) against Pol II and Pol III were used in the immunostaining, which was followed by previously described *Xist* RNA FISH(Zhang, Huynh et al. 2007). The result shows that, in contrast to Pol II, which is deficient in the Xi chromosome territory covered by the *Xist* RNA cloud, Pol III was not excluded (Fig.40), indicating on-going Pol III transcription along the Xi. This result supports our hypothesis.

Figure 40 Co-localization of RNA polymerases and *Xist* RNA using immuno-RNA FISH.

Immunofluorescence was performed using mouse monoclonal antibodies against RNA polymerase II (Pol II) and RNA polymerase III (Pol III) with a secondary antibody conjugated with FITC. Immunostaining was followed by RNA FISH with Cy3-labeled probe to detect the *Xist* RNA. Images shown are deconvolved images from a single z-section. Blue, DAPI; Green, Pol II and Pol III; Red, *Xist* RNA.

The Pol III transcription activities may not merely be a consequence of the ihoc positions. They may also involve regulating the chromatin architecture. To study the chromatin architecture remodeling function of Pol III transcriptions, chromatin immunoprecipitation (ChIP) was performed using antibody towards H3K27me3 in 2loxT.S2 cells with and without Pol III inhibitor treatment.

Another cell sample was generated using 2 mM Hydroxyurea (HU) to synchronize the cells at G1/S boundary, followed by 4 hrs Pol III inhibitor treatment (Fig.42A). Histone modifications play a role in maintenance of heterochromatin and mutually, heterochromatin regions passively maintain the histone modifications outside S phase (Probst, Dunleavy et al. 2009). To encourage the histone modification change to occur along those heterochromatin regions, it is better to treat the cells with Pol III inhibitor at S phase. Therefore the cells were synchronized and released at G1/S boundary before Pol III inhibition.

5 regions of X-linked gene *Hprt* and 4 regions of *Mecp2* were analyzed with the internal control (*GAPDH*). The result suggested Pol III inhibitor

may lead to a more compact chromatin structure (Fig.41, Fig.42B&Fig.38), which in consequence would either facilitate the *Xist* RNA to deposit the *Xist*-associated histone modification (H3K27me3) or increase the efficiency of antibody “pull down” during chromatin IP.

Figure 41 Chromatin IP to check the enrichment of H3K27me3 along the Xi in 2loxT.S2 cells with and without the Pol III inhibitor treatment.

Two X-linked genes, *Hprt* and *Mecp2*, were selected. *Gapdh* was used as the control. Data are shown as mean $\pm$ S.E.M of biological triplicate. The statistical analysis used is the Student's *t*-test. One asterisk indicates *P*-values smaller than 0.05.

Figure 42 Enrichment of H3K27me3 along the Xi in 2loxT.S1 cells arrested at G1/S boundary before Pol III inhibition.

A. Cells were arrested at the G1/S boundary by hydroxyurea treatment before Pol III inhibition. Cells treated with Pol III inhibitor were harvested at S phase. **B.** Chromatin IP results. Two X-linked genes, *Hprt* and *Mecp2*, were examined. *Gapdh* was used as the control. Data are shown as mean $\pm$ S.E.M of biological triplicate. The statistical analysis used is the Student's *t*-test. One asterisk indicates *P*-values smaller than 0.05.

A genome-wide ChIP-Seq experiment using antibody towards H3K27me3 was also performed to study chromatin architecture change caused by Pol III inhibitor treatment. The cell samples were synchronized by HU treatment. To analyze the data, Model-based Analysis of ChIP-Seq (MACS) (Zhang, Liu et al. 2008), a ChIP-Seq analyzing software, was used to call peaks. Peaks called by MACS were further analyzed by another software Homer (Heinz, Benner et al. 2010). By the `annotatePeaks` function of Homer, a heat map data matrix file of these peaks was generated (command options: `-size 6000 -hist 25 -ghist`). The heat map data matrix was clustered by Cluster3.0 (de Hoon, Imoto et al. 2004) using “Correlation (centered)” as the similarity metric and was visualized by Java Tree View (Saldanha 2004). Although most peaks look similar between the control sample and the Pol III inhibitor-treated sample, differences can be detected in some regions (Fig.43). Along the X chromosome, 1970 enrichment peaks of H3K27me3 appeared after Pol III inhibition and 1146 peaks disappeared (Fig.43), indicating a disorganized and more compact chromatin structure of the X chromosome after Pol III inhibition. A multi-

experiment overlap tracks were also generated using Homer, which aligned the peaks from the control sample and the Pol III inhibitor-treated sample (Fig.44). The result revealed the epigenetic landscape change of H3K27me3 enrichment after Pol III inhibition.

Figure 43 Heatmaps of ChIP-Seq data showing H3K27me3 enrichment along the X chromosome.

Peaks specific to the control sample and the Pol III inhibition sample were isolated, sorted according to $-10 \cdot \log_{10}(\text{p-value})$ of the peaks and presented in separated views.

Figure 44 The alignment of local-viewed ChIP-Seq peaks distribution and SINE density.

Orange: peaks specifically shown up in the control sample; purple: peaks specifically shown up in the Pol III inhibitor-treated sample; dark purple: peaks in common.

To detect chromatin structural change upon Pol III inhibition, a 5C (Carbon-Copy Chromosome Conformation Capture) assay was carried out on a ~ 1mb region of the X chromosome (Fig. 45B). Two cell types were involved in this study, 2loxT.S2 cells and 3F1 cells, which are mouse female ES cells ($Xa^{129}Xa^{Cast}$). 2loxT.S2 cells were treated with Pol III inhibitor overnight and 3F1 cells were differentiated in vitro for four days with the first two days treated with Pol III inhibitor (Fig. 45A). It is known that during the first few days of ES cell differentiation, de novo XCI is initiated and on-going in a significant percentage of the differentiating cells (Chaumeil, Le Baccon et al. 2006). Our 5C result clearly revealed more extensive and chromatin interactions in the fibroblast cells than the differentiating ES cells (Fig. 45D), which is consistent with the notion that fibroblast cells carry a fully inactivated X chromosome. Meanwhile, Pol III inhibition caused detectable change in chromatin looping in both the fibroblast cells and the differentiating ES cells, while more significant effects were observed on the differentiating ES cells (Fig. 45C-E). Collectively, Pol III inhibition affected chromatin looping, which confirmed Pol III transcription's role in regulating chromatin structure.

Figure 45 Pol III inhibition causes chromatin conformation change.

A. The experimental design of 5C. **B.** The ~1 mb region along the mouse X chromosome selected for 5C analysis (ChrX: 70,831,000-71,821,000). **C - E.** Chromatin interaction profiles of one anchor point (71176362) close to the transcription start site of a X-linked gene (*Renbp*, GenelD: 19703). Arrows: the differences of long distance chromatin looping between the two samples in each plot.

4.11 Pol III inhibition impairs XCI during in vitro differentiation

To study the function of Pol III transcription on XCI, the mutant mouse female ES cell line 3F1 ($X^{129}X^{Cast}$) was used. The *Tsix* gene on its X^{129} is mutated. *Tsix* gene is transcribed into non-coding RNA, which is antisense to *Xist*. During the random XCI occurrence, female cells undergo the “Choice” step. One of the two X chromosomes is determined as the future active X chromosome, on which *Tsix* is upregulated to prevent *Xist* RNA from covering (Lee, Davidow et al. 1999). In 3F1 cells, mutant *Tsix* causes non-random inactivation of the X^{129} allele, which enables us to monitor XCI occurrence by padlock probe SNP capture method.

We carried out in vitro differentiation of 3F1 for 4 days with or without Pol III inhibitor treatment. In this study, Pol III inhibitor was used to treat the cells during the first 2 days of differentiation (Fig.46A). X-linked genes of the differentiating cells (day4, control and day4, Pol III inhibition) and the (day0, control) 3F1 cells were allelotyped. The result demonstrated that from the d3F1.control sample, gene silencing was detected on 117 genes out of the allelotyped 171 genes (68%, $\log R_c < 0$)

(Fig.46B) and among the 117 being-silenced genes, the XCI process of 94 genes (80%) was impaired or delays upon Pol III inhibition ($\log R_i > 0$) (Fig.46B&C). The distribution of the 94 genes showed positive correlation with SINE density (Spearman's correlation = 0.73) (Fig.46C). Further analysis suggested SINE elements are enriched along the upstream regions of the 94 affected genes comparing to the 23 unaffected genes (Fig.47). Since chromatin structure re-organization is involved in *Xist*-mediated gene silencing, inhibition of Pol III changes the chromatin structure, thereby resulting in the aberrancy of XCI. This observation also confirmed that Pol III transcription plays a role in regulation of chromatin architecture.

Figure 46 Pol III inhibition impairs XCI in vitro.

A. The experimental design to study Pol III function on XCI. 3F1 cells were in vitro differentiated for 4 days, with or without Pol III inhibition for the first two days. **B.** The effect of Pol III inhibition on the XCI of 3F1 cells in vitro. Each bar represents one gene. X-linked genes were sorted according to their corresponding values on the y-axis. 117 genes showed signs of XCI in the control sample ($\text{Log Rc} < 0$), out of which 94 genes were affected by Pol III inhibition ($\text{Log Ri} > 0$). **C.** The alignment of Alu density and the X-linked genes affected by Pol III inhibition. Each bar represents one gene. Blue bar: genes with negative log value. Red bar: genes with positive log value. The Log Rc and Log Ri values of the X-linked genes are plotted along the X chromosome at the corresponding position of each gene.

Figure 47 Distribution of SINE elements along the X-linked genes.

The 20 kb upstream and the 20 kb downstream regions of each X-linked gene were scanned by a sliding window of 3 kb with a step size of 1 kb. The average copy number of a repetitive element within a 3 kb intronic region of a gene is used to represent the density of the repetitive element within the gene body. Data are shown as mean $\pm$ S.E.M. Box plots show the of copy number of SINE elements within 20 kb upstream of X-linked genes. 0: gene body.

4.12 Labeling the Xi in live cells using miniSOG and fluorescent protein

Live-cell labeling of the Xi has been achieved in previous studies (Csankovszki, Panning et al. 1999, Perche, Vourc'h et al. 2000, Chadwick, Valley et al. 2001) by fusing macroH2A (mH2A), or its histone domain (H2AD), with GFP. In this study, we aim to label the Xi by fusing macroH2A with miniSOG. The miniSOG-labeled Xi will be subjected to the subsequent EM analysis. The isoform mH2A1.2 (NM_001159513.1), which enriches in Xi, was cloned from cDNA of a mouse cell line with C57Bl6 genetic background. The coding cassette of miniSOG was obtained directly from Dr. Roger Tsien's laboratory ([http:// www.tsienlab.ucsd.edu/](http://www.tsienlab.ucsd.edu/)).

In our initial attempt, we have followed the same strategy used in previous studies of fluorescence labeling of the Xi in live cells. miniSOG protein was fused to the C terminal of mH2A or H2AD. Two plasmid constructs, p-mH2A-SOG and p-H2AD-SOG were established by gateway cloning. These two plasmid constructs were transfected into a transformed female mouse fibroblast cell line, called 129T. However, after transfection, under the fluorescent microscope, we did not observe the Xi

labeled by the green fluorescence of miniSOG (Fig.48). To trouble shoot, we utilized the plasmid construct, p-miniSOG-mCherry-C1, in which miniSOG is fused with a bright fluorescent protein mCherry. When pminiSOG-mCherry-C1 was transfected into 129T cells, we observed the bright red fluorescence emitted by mCherry. However, within a cell which was brightly labeled by mCherry, no green fluorescence was observed (Fig.49), which reveals that the green fluorescence of miniSOG might be too weak to be observed. To verify, we transfected the cells with ptdminiSOG-N1, a plasmid that expresses tandem dimer miniSOG (tdminiSOG). Although the green fluorescence of tdminiSOG was visible, the signal was weak and we observed severe photo-bleaching effect when we were examining the signals through microscope eye-pieces. Taken together, these data show that miniSOG, an artificial protein engineered mainly as a live-cell label for EM analysis, is too weak for the fluorescent labeling purpose.

In order to locally polymerize DAB, miniSOG labeled target has to be identified under a fluorescent microscope using a fluorescent label. In the

original study, miniSOG was engineered to produce green fluorescence to serve this purpose. However, in our study, we found the green fluorescence of miniSOG is too weak to fulfill this purpose. To solve this problem, several new plasmid constructs (p-H2AD-mc, p-H2AD-SOG-mc, p-H2AD-tdSOG, p-mH2A-mc-SOG, p-mH2A-SOG-mc and p-H2AD-mc-SOG) were generated (Fig.50&Fig.51). The idea for these designs is to use mCherry to replace miniSOG as the fluorescent label. Cells Transfected with p-H2AD-mc plasmid construct showed specifically labeled Xi (Fig.50). This result confirms the previous studies that the H2A domain of mH2A can be used as a Xi marker in live cells when the peptide is fused with fluorescent proteins at its C terminal. We cannot find cells with fluorescently labeled Xi when cells were transfected with p-H2AD-SOG-mc and p-H2AD-tdSOG. For p-H2AD-SOG-mc labeling, fluorescent signal can be observed but no enrichment signal shows up to determine the Xi territory. This is probably because of the steric influence of the miniSOG fusing protein and therefore affects the labeling. For p-H2AD-tdSOG labeling, no signal can be observed, probably because the fluorescent

signal emitted by tminiSOG is too weak to observe. Fortunately, the other three plasmid constructs (p-mH2A-mc-SOG, p-H2AD-mc-SOG and p-mH2A-SOG-mc) can give us good labeling of the Barr body(Fig.51). Moreover, in some p-mH2A-SOG-mc transfected cells, even the signal of miniSOG was successfully captured though it suffers severe photo-bleaching (Fig.51).

Considering the transfection efficiency of 129T cells is too low (~10%), in the following study, we also used 293T cells, which are human female transformed fibroblast cells with high transfection efficiency (~80%). Because mH2A1.2 is highly conserved between species (Pehrson, Changolkar et al. 2014), fusion protein macroH2A-miniSOG-mCherry, which contains the mouse mH2A, is also capable of labeling human Xi. In the p-mH2A-SOG-mc transfected 293T cells, Xi with fluorescent labeling can be clearly observed (Fig.53). Transiently transfected 129T cells and 293T cells were next used for photo-oxidation, which is the reaction that miniSOG catalyze the localized polymerization of DAB under the illumination of blue light. Cells were transfected with p-mH2A-SOG-mc

when cell population reached ~70%. Photo-oxidation was performed 2 days post-transfection. After photo-oxidation, the area with miniSOG labeling showed as brownish view, which is a sign of photo-oxidation (Fig.52&Fig.53). Successful photo-oxidation of miniSOG labeled Xi was observed in a transiently transfected 293T cell (Fig.53).

Figure 48 The labeling of Xi with mH2A-SOG and H2AD-SOG.

A. the designs of p-mH2A-SOG and p-H2AD-SOG plasmid constructs. **B.** The transient transfection result of 129T mouse fibroblast cells with these two plasmid constructs.

Figure 49 The transient transfection result of 129T mouse fibroblast cells with two plasmids, pminiSOG-mCherry-C1 and ptdminiSOG-N1.

Photobleaching of tminiSOG fluorescence is shown in the 3 sec postponed image.

Figure 50 The labeling of Xi with H2AD-mc, H2AD-SOG-mc and H2AD-tdSOG.

A. the designs of p-H2AD-mc, p-H2AD-SOG-mc and p-H2AD-tdSOG plasmid constructs. **B.** The transient transfection result of 129T mouse fibroblast cells with these three plasmid constructs. Arrowhead: a macrochromatin body.

Figure 51 The labeling of Xi with mH2A-mc-SOG, H2AD-mc-SOG and mH2A-SOG-mc.

A. the designs of p-mH2A-mc-SOG, p-H2AD-mc-SOG and p-mH2A-SOG-mc plasmid constructs. **B.** The transient transfection result of 129T mouse fibroblast cells with these three plasmid constructs. Arrowhead: a macrochromatin body.

Figure 52 A photo-oxidation effect of a mH2A-miniSOG-mCherry labeled cell which is a transient transfected 129T mouse fibroblast cell with p-mH2A-SOG-mc.

miniSOG: the miniSOG labeling in a Cy2 channel. Before: a white field imaging before photo-oxidation. After: a white field imaging after photo-oxidation.

Figure 53 A photo-oxidation effect of a mH2A-miniSOG-mCherry labeled cell which is a transient transfected tetraploid 293T human fibroblast cell with p-mH2A-SOG-mc.

mCherry: the mCherry labeling in a fluorescent channel. 4 minutes: a white field imaging when photo-oxidation was carried out for 4 mins. 6 minutes: a white field imaging when photo-oxidation was carried out for 6 mins. Arrows: Barr bodies.

5 Conclusion and Discussion

In this study, using the padlock probes SNP capture method (Zhang, Li et al. 2009), we are able to target thousands of genes in an allele-specific fashion. Gene reactivation pattern was profiled in six mouse fibroblast cell lines of which *Xist* was conditionally deleted. The results reveal that among detected genes, ~4% gene reactivation on average occurred, suggesting a low level of gene reactivation, which is consistent with previously published studies (Brown and Willard 1994, Csankovszki, Nagy et al. 2001). The reactivated genes varied among cell lines, but the overall pattern showed positive correlation with *Xist* RNA coating, gene density and SINEs density. No gene was consistently reactivated in all the mutant cell lines. Instead, there were 25 genes (frequently reactivated genes) reactivated in at least 2 mutant cell populations. Further analysis reveals that SINE are significantly enriched in the upstream regions of the reactivated genes comparing to the genes remain silenced, which suggests a gene with its upstream region enriched with SINE is more likely to be reactivated. Long term cell culture study further confirmed the stubborn

gene silencing status in *Xist* deleted cells, as long term cell culture cannot cause a larger percentage of genes reactivated or re-silenced.

The expression pattern of X-linked non-coding transcripts were also studied using Padlock capture. Surprisingly, the expression pattern of non-coding transcription is largely independent of *Xist*-mediated gene silencing. Again, the distribution of non-coding transcription is positively correlated to SINE density. Immunostaining-RNA FISH reveals the co-localization of Pol III and Xi territory. The transcripts can be inhibited by a Pol III specific inhibitor. Moreover, one transcript from the non-coding regions was identified as a putative SINE, which can be narrowed down to a typical SINE length. These evidences suggest the transcripts we detected are Pol III products and some of them are SINEs. Function of these transcriptions was also explored. Inhibition of these transcriptions caused the re-silencing of downstream reactivated gene as well as the chromatin architecture change. The results show that Pol III transcription along the non-coding regions is involved in regulating chromatin structure.

Currently, more and more evidence suggest the involvement of Pol III

transcriptions, especially SINE, in the regulation of gene expression and chromatin architecture (Dixon, Selvaraj et al. 2012, Pascali and Teichmann 2013) Ferrigno, Virolle et al. 2001, Allen, Von Kaenel et al. 2004. Lunyak, Prefontaine et al. 2007. Roman, Gonzalez-Rico et al. 2011). In yeast, tRNA genes are impervious to heterochromatin formation (Haldar and Kamakaka 2006). B2 SINE is characterized to have an active Pol II promoter at one end away from the Pol III promoter. The binding of this Pol II promoter with transcription factor USF (upstream stimulatory factor) promotes the expression from the core Pol II promoter (Ferrigno, Virolle et al. 2001). A B2 SINE repeat is also found to function as a domain boundary for the growth hormone (GH) gene to restrict interference of silent chromatin domains (Lunyak, Prefontaine et al. 2007). Another study revealed that a B1 SINE (B1-X35S) is a potential insulator (Roman, Gonzalez-Rico et al. 2011). Enrichment of B1-X35S is found in 1,398 gene promoters, suggesting the insulator activity of B1-X35S may affect thousands of genes expression. The study also finds that B1-X35S insulation is consistent with the enrichment of heterochromatic marks

H3K9me3 and H3K27me3 in downstream region and is associated with the binding of CTCF (Roman, Gonzalez-Rico et al. 2011). Enrichments of SINE, tRNA, CTCF and housekeeping genes are observed in the boundaries of TADs, suggesting these factors may play a role in defining the TADs construction (Dixon, Selvaraj et al. 2012). Positive correlation of SINE density and *Xist* RNA coating is also identified in a *Xist* RNA coating pattern study (Simon, Pinter et al. 2013).

Although a number of Pol III target genes were characterized, it is still unclear that how many Pol III binding loci and how many Pol III transcriptions exist in the genome. Chromatin immunoprecipitation (ChIP) followed by DNA microarrays or massively parallel sequencing has been widely used to identify novel Pol III binding sites in a genome wide scale (White 2011). These studies discover that except those known Pol III target genes, there are a number of Pol III-associated loci distributing in human genome with functions still unclear (Barski, Chepelev et al. 2010, Canella, Praz et al. 2010, Moqtaderi, Wang et al. 2010, Oler, Alla et al. 2010, Raha, Wang et al. 2010).

In our study, hundreds of Pol III transcripts from Xi were identified. These transcription activities are independent of the higher order chromatin structure and some of them are assumed to occur at iHOC positions. The activities promote nearby gene reactivation after Xi deletion. Inhibition of the Pol III transcriptions results in disordered chromatin structure, suggesting that Pol III transcription not only involves in nearby gene expression but also functions as chromatin architecture regulator.

Fusing macroH2A with miniSOG can successfully label the inactive X chromosome. We have generated several plasmid constructs which encode mH2A-miniSOG or H2AD-miniSOG fused protein to label the inactive X chromosomes of mouse fibroblast 129T cells. Through transient transfection, some of these plasmid constructs can give us clear labeling of Xi under the fluorescent microscope, such as p-mH2A-SOG-mc, p-H2AD-SOG-mc, p-mH2A-mc-SOG and p-H2AD-mc-SOG. Inactive X chromosomes labeled with the mH2A-miniSOG-mCherry fusion protein in a transiently transfected 293T cell can be clearly observed after photo-oxidation,

indicating the mH2A-miniSOG-mCherry fusing protein can function robustly.

We next plan to prepare resin embedded sample and observe the labeled Xi under EM. However, although transient transfection can give us a successful labeling of Xi, the high expression level of the fused protein (mH2A-miniSOG-mCherry) may lead to heavy background and poor resolution. To solve the issue, establishment of a stable cell line with endogenously modified mH2A to express mH2A-miniSOG-mCherry fused protein is necessary. On the other hand, we plan to label transcriptionally active chromosomes using H2A-miniSOG-mCherry fused protein, thereby solving the euchromatin structure.

A previous study uses EM to observe the interphase Xi. The morphology of Xi suggests a loosely packaged heterochromatin structure, of which the intensity is between euchromatin and heterochromatin. As the structure of heterochromatin is more condensed than euchromatin, heterochromatin is proposed to have a smaller volume. Another study reports that Xi and Xa have the same volume in the resolution of

fluorescence microscopy (Eils, Dietzel et al. 1996). We expect that, using miniSOG labeling, high resolution structures of Xi and transcriptionally active chromosome regions can be solved, which may help us to better understand the difference between the structure of Xi and euchromatin.

6 Reference

- Agrelo, R., A. Souabni, M. Novatchkova, C. Haslinger, M. Leeb, V. Komnenovic, H. Kishimoto, L. Gresh, T. Kohwi-Shigematsu, L. Kenner and A. Wutz (2009). "SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells." Dev Cell **16**(4): 507-516.
- Alvarez, J. D., D. H. Yasui, H. Niida, T. Joh, D. Y. Loh and T. Kohwi-Shigematsu (2000). "The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development." Genes Dev **14**(5): 521-535.
- Azzalini, A. and A. Capitanio (2003). "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution" Journal of the Royal Statistical Society Series B-Methodological **65**: 367-389.
- Bailey, J. A., L. Carrel, A. Chakravarti and E. E. Eichler (2000). "Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis." Proc Natl Acad Sci U S A **97**(12): 6634-6639.
- Barr, M. L. and E. G. Bertram (1949). "A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis." Nature **163**(4148): 676.
- Barski, A., I. Chepelev, D. Liko, S. Cuddapah, A. B. Fleming, J. Birch, K. Cui, R. J. White and K. Zhao (2010). "Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes." Nat Struct Mol Biol **17**(5): 629-634.
- Batzner, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." Nat Rev Genet **3**(5): 370-379.
- Becker, K. G., G. D. Swergold, K. Ozato and R. E. Thayer (1993). "Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element." Hum Mol Genet **2**(10): 1697-1702.
- Berendes, H. D. and G. F. Meyer (1968). "A specific chromosome element, the telomere of Drosophila polytene chromosomes." Chromosoma **25**(2): 184-197.
- Bernstein, B. E., E. L. Humphrey, R. L. Erlich, R. Schneider, P. Bouman, J. S. Liu, T. Kouzarides and S. L. Schreiber (2002). "Methylation of histone H3 Lys 4 in coding regions of active genes." Proc Natl Acad Sci U S A **99**(13): 8695-8700.
- Brockdorff, N., A. Ashworth, G. F. Kay, P. Cooper, S. Smith, V. M. McCabe, D. P. Norris, G. D. Penny, D. Patel and S. Rastan (1991). "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome." Nature **351**(6324): 329-331.

Brockdorff, N., A. Ashworth, G. F. Kay, V. M. McCabe, D. P. Norris, P. J. Cooper, S. Swift and S. Rastan (1992). "The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus." Cell **71**(3): 515-526.

Brown, C. J. and H. F. Willard (1994). "The human X-inactivation centre is not required for maintenance of X-chromosome inactivation." Nature **368**(6467): 154-156.

Brownell, J. E. and C. D. Allis (1996). "Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation." Curr Opin Genet Dev **6**(2): 176-184.

Brownell, J. E., J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth and C. D. Allis (1996). "Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation." Cell **84**(6): 843-851.

Cai, S., H. J. Han and T. Kohwi-Shigematsu (2003). "Tissue-specific nuclear architecture and gene expression regulated by SATB1." Nat Genet **34**(1): 42-51.

Canella, D., V. Praz, J. H. Reina, P. Cousin and N. Hernandez (2010). "Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells." Genome Res **20**(6): 710-721.

Cantrell, M. A., B. C. Carstens and H. A. Wichman (2009). "X chromosome inactivation and Xist evolution in a rodent lacking LINE-1 activity." PLoS One **4**(7): e6252.

Carrel, L., C. Park, S. Tyekucheva, J. Dunn, F. Chiaromonte and K. D. Makova (2006). "Genomic environment predicts expression patterns on the human inactive X chromosome." PLoS Genet **2**(9): e151.

Carrel, L. and H. F. Willard (2005). "X-inactivation profile reveals extensive variability in X-linked gene expression in females." Nature **434**(7031): 400-404.

Chadwick, B. P., C. M. Valley and H. F. Willard (2001). "Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome." Nucleic Acids Res **29**(13): 2699-2705.

Chan, S. R. and E. H. Blackburn (2004). "Telomeres and telomerase." Philos Trans R Soc Lond B Biol Sci **359**(1441): 109-121.

Chaumeil, J., P. Le Baccon, A. Wutz and E. Heard (2006). "A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced." Genes Dev **20**(16): 2223-2237.

Chow, J. C., C. Ciaudo, M. J. Fazzari, N. Mise, N. Servant, J. L. Glass, M. Attreed, P. Avner, A. Wutz, E. Barillot, J. M. Greally, O. Voinnet and E. Heard (2010). "LINE-1

activity in facultative heterochromatin formation during X chromosome inactivation." *Cell* **141**(6): 956-969.

Chow, J. C., Z. Yen, S. M. Ziesche and C. J. Brown (2005). "Silencing of the mammalian X chromosome." *Annu Rev Genomics Hum Genet* **6**: 69-92.

Civiero, L., R. Vancraenenbroeck, E. Belluzzi, A. Beilina, E. Lobbetael, L. Reyniers, F. Gao, I. Micetic, M. De Maeyer, L. Bubacco, V. Baekelandt, M. R. Cookson, E. Greggio and J. M. Taymans "Biochemical characterization of highly purified leucine-rich repeat kinases 1 and 2 demonstrates formation of homodimers." *PLoS One* **7**(8): e43472.

Costanzi, C. and J. R. Pehrson (1998). "Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals." *Nature* **393**(6685): 599-601.

Courtier, B., E. Heard and P. Avner (1995). "Xce haplotypes show modified methylation in a region of the active X chromosome lying 3' to Xist." *Proc Natl Acad Sci U S A* **92**(8): 3531-3535.

Csankovszki, G., A. Nagy and R. Jaenisch (2001). "Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation." *J Cell Biol* **153**(4): 773-784.

Csankovszki, G., B. Panning, B. Bates, J. R. Pehrson and R. Jaenisch (1999). "Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation." *Nat Genet* **22**(4): 323-324.

Danecek, P., C. Nellaker, R. E. McIntyre, J. E. Buendia-Buendia, S. Bumpstead, C. P. Ponting, J. Flint, R. Durbin, T. M. Keane and D. J. Adams (2012). "High levels of RNA-editing site conservation amongst 15 laboratory mouse strains." *Genome Biol* **13**(4): 26.

de Hoon, M. J., S. Imoto, J. Nolan and S. Miyano (2004). "Open source clustering software." *Bioinformatics* **20**(9): 1453-1454.

Denisenko, O., M. Shnyreva, H. Suzuki and K. Bomsztyk (1998). "Point mutations in the WD40 domain of Eed block its interaction with Ezh2." *Mol Cell Biol* **18**(10): 5634-5642.

Diep, D., N. Plongthongkum, A. Gore, H. L. Fung, R. Shoemaker and K. Zhang (2012). "Library-free methylation sequencing with bisulfite padlock probes." *Nat Methods* **9**(3): 270-272.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu and B. Ren (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions." *Nature* **485**(7398): 376-380.

Dostie, J., Y. Zhan and J. Dekker (2007). "Chromosome conformation capture carbon copy technology." Curr Protoc Mol Biol **Chapter 21**: Unit 21 14.

Efron, B., R. Tibshirani, J. D. Storey and V. Tusher (2001). "Empirical Bayes analysis of a microarray experiment." Journal of the American Statistical Association **96**(456): 1151-1160.

Eils, R., S. Dietzel, E. Bertin, E. Schrock, M. R. Speicher, T. Ried, M. Robert-Nicoud, C. Cremer and T. Cremer (1996). "Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure." J Cell Biol **135**(6 Pt 1): 1427-1440.

Engreitz, J. M., A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath and M. Guttman (2013). "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome." Science **341**(6147): 1237973.

Ferrigno, O., T. Virolle, Z. Djabari, J. P. Ortonne, R. J. White and D. Aberdam (2001). "Transposable B2 SINE elements can provide mobile RNA polymerase II promoters." Nat Genet **28**(1): 77-81.

Hajkova, P., S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter and M. A. Surani (2002). "Epigenetic reprogramming in mouse primordial germ cells." Mech Dev **117**(1-2): 15-23.

Haldar, D. and R. T. Kamakaka (2006). "tRNA genes as chromatin barriers." Nat Struct Mol Biol **13**(3): 192-193.

Hall, L. L., D. M. Carone, A. V. Gomez, H. J. Kolpa, M. Byron, N. Mehta, F. O. Fackelmayer and J. B. Lawrence (2014). "Stable COT-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes." Cell **156**(5): 907-919.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.

Helbig, R. and F. O. Fackelmayer (2003). "Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain." Chromosoma **112**(4): 173-182.

Huynh, K. D. and J. T. Lee (2003). "Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos." Nature **426**(6968): 857-862.

Jeon, Y. and J. T. Lee (2011). "YY1 tethers Xist RNA to the inactive X nucleation center." Cell **146**(1): 119-133.

Katayama, S., Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, H. Nishida, C. C. Yap, M. Suzuki, J. Kawai, H. Suzuki, P. Carninci, Y. Hayashizaki, C. Wells, M. Frith, T. Ravasi, K. C. Pang, J. Hallinan, J. Mattick, D. A. Hume, L. Lipovich, S. Batalov, P. G. Engstrom, Y. Mizuno, M. A. Faghihi, A. Sandelin, A. M. Chalk, S. Mottagui-Tabar, Z. Liang, B. Lenhard, C. Wahlestedt, R. G. E. R. Group, G. Genome Science and F. Consortium (2005). "Antisense transcription in the mammalian transcriptome." *Science* **309**(5740): 1564-1566.

Kay, G. F., G. D. Penny, D. Patel, A. Ashworth, N. Brockdorff and S. Rastan (1993). "Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation." *Cell* **72**(2): 171-182.

Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N. A. Furlotte, E. Eskin, C. Nellaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T. G. Belgard, P. L. Oliver, R. E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. van der Weyden, C. A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I. J. Jackson, A. Czechanski, J. A. Guerra-Assuncao, L. R. Donahue, L. G. Reinholdt, B. A. Payseur, C. P. Ponting, E. Birney, J. Flint and D. J. Adams (2011). "Mouse genomic variation and its effect on phenotypes and gene regulation." *Nature* **477**(7364): 289-294.

Kurose, K., K. Hata, M. Hattori and Y. Sakaki (1995). "RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system." *Nucleic Acids Res* **23**(18): 3704-3709.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chisoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P.

Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglu, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen and C. International Human Genome Sequencing (2001). "Initial sequencing and analysis of the human genome." *Nature* **409**(6822): 860-921.

Lee, J. T., L. S. Davidow and D. Warshawsky (1999). "Tsix, a gene antisense to Xist at the X-inactivation centre." *Nat Genet* **21**(4): 400-404.

Lee, J. T. and N. Lu (1999). "Targeted mutagenesis of Tsix leads to nonrandom X inactivation." *Cell* **99**(1): 47-57.

Lunyak, V. V., G. G. Prefontaine, E. Nunez, T. Cramer, B. G. Ju, K. A. Ohgi, K. Hutt, R. Roy, A. Garcia-Diaz, X. Zhu, Y. Yung, L. Montolieu, C. K. Glass and M. G. Rosenfeld (2007). "Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis." *Science* **317**(5835): 248-251.

Lyon, M. F. (1998). "X-chromosome inactivation: a repeat hypothesis." Cytogenet Cell Genet **80**(1-4): 133-137.

Markham, N. R. and M. Zuker (2008). "UNAFold: software for nucleic acid folding and hybridization." Methods Mol Biol **453**: 3-31.

Mazo, A., J. W. Hodgson, S. Petruk, Y. Sedkov and H. W. Brock (2007). "Transcriptional interference: an unexpected layer of complexity in gene regulation." J Cell Sci **120**(Pt 16): 2755-2761.

Minajigi, A., J. E. Froberg, C. Wei, H. Sunwoo, B. Kesner, D. Colognori, D. Lessing, B. Payer, M. Boukhali, W. Haas and J. T. Lee (2015). "Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation." Science **349**(6245).

Minakami, R., K. Kurose, K. Etoh, Y. Furuhashi, M. Hattori and Y. Sakaki (1992). "Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to the element." Nucleic Acids Res **20**(12): 3139-3145.

Monk, M. and M. I. Harper (1979). "Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos." Nature **281**(5729): 311-313.

Moqtaderi, Z., J. Wang, D. Raha, R. J. White, M. Snyder, Z. Weng and K. Struhl (2010). "Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells." Nat Struct Mol Biol **17**(5): 635-640.

Mouse Genome Sequencing, C., R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyas, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers,

J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody and E. S. Lander (2002). "Initial sequencing and comparative analysis of the mouse genome." *Nature* **420**(6915): 520-562.

Muthurajan, U. M., S. J. McBryant, X. Lu, J. C. Hansen and K. Luger (2011). "The linker region of macroH2A promotes self-association of nucleosomal arrays." *J Biol Chem* **286**(27): 23852-23864.

Nechanitzky, R., A. Davila, F. Savarese, S. Fietze and R. Grosschedl "Satb1 and Satb2 are dispensable for X chromosome inactivation in mice." *Dev Cell* **23**(4): 866-871.

Okada, N. (1991). "SINEs: Short interspersed repeated elements of the eukaryotic genome." *Trends Ecol Evol* **6**(11): 358-361.

Oler, A. J., R. K. Alla, D. N. Roberts, A. Wong, P. C. Hollenhorst, K. J. Chandler, P. A. Cassidy, C. A. Nelson, C. H. Hagedorn, B. J. Graves and B. R. Cairns (2010). "Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors." *Nat Struct Mol Biol* **17**(5): 620-628.

Ostertag, E. M. and H. H. Kazazian, Jr. (2001). "Biology of mammalian L1 retrotransposons." *Annu Rev Genet* **35**: 501-538.

Pascali, C. and M. Teichmann (2013). "RNA polymerase III transcription - regulated by chromatin structure and regulator of nuclear chromatin organization." *Subcell Biochem* **61**: 261-287.

Pasque, V., A. Gillich, N. Garrett and J. B. Gurdon (2011). "Histone variant macroH2A confers resistance to nuclear reprogramming." *EMBO J* **30**(12): 2373-2387.

Paulsen, M. T., A. Veloso, J. Prasad, K. Bedi, E. A. Ljungman, B. Magnuson, T. E. Wilson and M. Ljungman (2014). "Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA." Methods **67**(1): 45-54.

Pazin, M. J. and J. T. Kadonaga (1997). "What's up and down with histone deacetylation and transcription?" Cell **89**(3): 325-328.

Pehrson, J. R., L. N. Changolkar, C. Costanzi and N. A. Leu (2014). "Mice without macroH2A histone variants." Mol Cell Biol **34**(24): 4523-4533.

Pehrson, J. R. and R. N. Fuji (1998). "Evolutionary conservation of histone macroH2A subtypes and domains." Nucleic Acids Res **26**(12): 2837-2842.

Penny, G. D., G. F. Kay, S. A. Sheardown, S. Rastan and N. Brockdorff (1996). "Requirement for Xist in X chromosome inactivation." Nature **379**(6561): 131-137.

Perche, P. Y., C. Vourc'h, L. Konecny, C. Souchier, M. Robert-Nicoud, S. Dimitrov and S. Khochbin (2000). "Higher concentrations of histone macroH2A in the Barr body are correlated with higher nucleosome density." Curr Biol **10**(23): 1531-1534.

Pereira, L. V. and L. R. Vasques (2000). "X-chromosome inactivation: lessons from transgenic mice." Gene **255**(2): 363-371.

Plath, K., J. Fang, S. K. Mlynarczyk-Evans, R. Cao, K. A. Worringer, H. Wang, C. C. de la Cruz, A. P. Otte, B. Panning and Y. Zhang (2003). "Role of histone H3 lysine 27 methylation in X inactivation." Science **300**(5616): 131-135.

Probst, A. V., E. Dunleavy and G. Almouzni (2009). "Epigenetic inheritance during the cell cycle." Nat Rev Mol Cell Biol **10**(3): 192-206.

Pullirsch, D., R. Hartel, H. Kishimoto, M. Leeb, G. Steiner and A. Wutz (2010). "The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation." Development **137**(6): 935-943.

Raha, D., Z. Wang, Z. Moqtaderi, L. Wu, G. Zhong, M. Gerstein, K. Struhl and M. Snyder (2010). "Close association of RNA polymerase II and many transcription factors with Pol III genes." Proc Natl Acad Sci U S A **107**(8): 3639-3644.

Rasmussen, T. P., T. Huang, M. A. Mastrangelo, J. Loring, B. Panning and R. Jaenisch (1999). "Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing." Nucleic Acids Res **27**(18): 3685-3689.

Rego, A., P. B. Sinclair, W. Tao, I. Kireev and A. S. Belmont (2008). "The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure." J Cell Sci **121**(Pt 7): 1119-1127.

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal and H. Y. Chang (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell **129**(7): 1311-1323.

Robertson, K. D., E. Uzvolgyi, G. Liang, C. Talmadge, J. Sumegi, F. A. Gonzales and P. A. Jones (1999). "The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors." Nucleic Acids Res **27**(11): 2291-2298.

Rogers, J. H. (1985). "The origin and evolution of retroposons." Int Rev Cytol **93**: 187-279.

Roman, A. C., F. J. Gonzalez-Rico, E. Molto, H. Hernando, A. Neto, C. Vicente-Garcia, E. Ballestar, J. L. Gomez-Skarmeta, J. Vavrova-Anderson, R. J. White, L. Montoliu and P. M. Fernandez-Salguero (2011). "Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch." Genome Res **21**(3): 422-432.

Saldanha, A. J. (2004). "Java Treeview--extensible visualization of microarray data." Bioinformatics **20**(17): 3246-3248.

Sanchez-Elsner, T., D. Gou, E. Kremmer and F. Sauer (2006). "Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax." Science **311**(5764): 1118-1123.

Savarese, F., K. Flahndorfer, R. Jaenisch, M. Busslinger and A. Wutz (2006). "Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation." Mol Cell Biol **26**(19): 7167-7177.

Schneider, R., A. J. Bannister, F. A. Myers, A. W. Thorne, C. Crane-Robinson and T. Kouzarides (2004). "Histone H3 lysine 4 methylation patterns in higher eukaryotic genes." Nat Cell Biol **6**(1): 73-77.

Sheardown, S. A., S. M. Duthie, C. M. Johnston, A. E. Newall, E. J. Formstone, R. M. Arkell, T. B. Nesterova, G. C. Alghisi, S. Rastan and N. Brockdorff (1997). "Stabilization of Xist RNA mediates initiation of X chromosome inactivation." Cell **91**(1): 99-107.

Shoemaker, R., J. Deng, W. Wang and K. Zhang (2010). "Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome." Genome Res **20**(7): 883-889.

Shu, X., V. Lev-Ram, T. J. Deerinck, Y. Qi, E. B. Ramko, M. W. Davidson, Y. Jin, M. H. Ellisman and R. Y. Tsien "A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms." PLoS Biol **9**(4): e1001041.

Silva, J., W. Mak, I. Zvetkova, R. Appanah, T. B. Nesterova, Z. Webster, A. H. Peters, T. Jenuwein, A. P. Otte and N. Brockdorff (2003). "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.

Simon, M. D., S. F. Pinter, R. Fang, K. Sarma, M. Rutenberg-Schoenberg, S. K. Bowman, B. A. Kesner, V. K. Maier, R. E. Kingston and J. T. Lee (2013). "High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation." *Nature* **504**(7480): 465-469.

Simon, M. D., C. I. Wang, P. V. Kharchenko, J. A. West, B. A. Chapman, A. A. Alekseyenko, M. L. Borowsky, M. I. Kuroda and R. E. Kingston (2011). "The genomic binding sites of a noncoding RNA." *Proc Natl Acad Sci U S A* **108**(51): 20497-20502.

Singer-Sam, J., V. Chapman, J. M. LeBon and A. D. Riggs (1992). "Parental imprinting studied by allele-specific primer extension after PCR: paternal X chromosome-linked genes are transcribed prior to preferential paternal X chromosome inactivation." *Proc Natl Acad Sci U S A* **89**(21): 10469-10473.

Southern, E. M. (1975). "Detection of specific sequences among DNA fragments separated by gel electrophoresis." *J Mol Biol* **98**(3): 503-517.

Strahl, B. D., R. Ohba, R. G. Cook and C. D. Allis (1999). "Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena." *Proc Natl Acad Sci U S A* **96**(26): 14967-14972.

Tanasijevic, B. and T. P. Rasmussen (2011). "X chromosome inactivation and differentiation occur readily in ES cells doubly-deficient for macroH2A1 and macroH2A2." *PLoS One* **6**(6): e21512.

Tang, S., M. Mishra, D. P. Frazier, M. L. Moore, K. Inoue, R. Deora, G. Sui and P. Dubey (2012). "Positive and negative regulation of prostate stem cell antigen expression by Yin Yang 1 in prostate epithelial cell lines." *PLoS One* **7**(4): e35570.

Taunton, J., C. A. Hassig and S. L. Schreiber (1996). "A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p." *Science* **272**(5260): 408-411.

Tse, C., T. Sera, A. P. Wolffe and J. C. Hansen (1998). "Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III." *Mol Cell Biol* **18**(8): 4629-4638.

van Dijk, K., K. E. Marley, B. R. Jeong, J. Xu, J. Hesson, R. L. Cerny, J. H. Waterborg and H. Cerutti (2005). "Monomethyl histone H3 lysine 4 as an epigenetic mark for silenced euchromatin in Chlamydomonas." *Plant Cell* **17**(9): 2439-2453.

Washietl, S., J. S. Pedersen, J. O. Korbel, C. Stocsits, A. R. Gruber, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Reiche, A. Tanzer, C. Ucla, C. Wyss, S. E. Antonarakis, F. Denoeud, J. Lagarde, J. Drenkow, P. Kapranov, T. R. Gingeras, R. Guigo, M. Snyder, M. B. Gerstein, A. Reymond, I. L. Hofacker and P. F. Stadler (2007). "Structured RNAs in the ENCODE selected regions of the human genome." Genome Res **17**(6): 852-864.

Weiner, A. M., P. L. Deininger and A. Efstratiadis (1986). "Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information." Annu Rev Biochem **55**: 631-661.

West, J. D., W. I. Frels, V. M. Chapman and V. E. Papaioannou (1977). "Preferential expression of the maternally derived X chromosome in the mouse yolk sac." Cell **12**(4): 873-882.

White, C. L., R. K. Suto and K. Luger (2001). "Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions." EMBO J **20**(18): 5207-5218.

White, R. J. (2011). "Transcription by RNA polymerase III: more complex than we thought." Nat Rev Genet **12**(7): 459-463.

Wolffe, A. P. (1997). "Transcriptional control. Sinful repression." Nature **387**(6628): 16-17.

Wolffe, A. P. and D. Pruss (1996). "Targeting chromatin disruption: Transcription regulators that acetylate histones." Cell **84**(6): 817-819.

Wu, L., J. Pan, V. Thoroddsen, D. R. Wysong, R. K. Blackman, C. E. Bulawa, A. E. Gould, T. D. Ocain, L. R. Dick, P. Errada, P. K. Dorris, T. Parkinson, T. Wood, D. Kornitzer, Z. Weissman, I. M. Willis and K. McGovern (2003). "Novel small-molecule inhibitors of RNA polymerase III." Eukaryot Cell **2**(2): 256-264.

Yalcin, B., K. Wong, A. Agam, M. Goodson, T. M. Keane, X. Gan, C. Nellaker, L. Goodstadt, J. Nicod, A. Bhomra, P. Hernandez-Pliego, H. Whitley, J. Cleak, R. Dutton, D. Janowitz, R. Mott, D. J. Adams and J. Flint (2011). "Sequence-based characterization of structural variation in the mouse genome." Nature **477**(7364): 326-329.

Zhang, K., J. B. Li, Y. Gao, D. Egli, B. Xie, J. Deng, Z. Li, J. H. Lee, J. Aach, E. M. Leproust, K. Eggan and G. M. Church (2009). "Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human." Nat Methods **6**(8): 613-618.

Zhang, L.-F., K. D. Huynh and J. T. Lee (2007). "Perinucleolar Targeting of the Inactive X during S Phase: Evidence for a Role in the Maintenance of Silencing." Cell **129**(4): 693-706.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.

Zhao, J., B. K. Sun, J. A. Erwin, J. J. Song and J. T. Lee (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." Science **322**(5902): 750-756.

Zheng, D., A. Frankish, R. Baertsch, P. Kapranov, A. Reymond, S. W. Choo, Y. Lu, F. Denoeud, S. E. Antonarakis, M. Snyder, Y. Ruan, C. L. Wei, T. R. Gingeras, R. Guigo, J. Harrow and M. B. Gerstein (2007). "Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution." Genome Res **17**(6): 839-851.