Understanding transcriptional regulation of the long noncoding RNA XIST

A Thesis

Submitted in partial fulfillment of the requirements

Of the degree of

Doctor of Philosophy

By

Rini Shah

20103047

Under the guidance of
Prof. Sanjeev Galande

at

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

December 2016

Dedicated to my late grandmother…

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources that have not been properly cited or from whom proper permission has not been taken when needed.

Rini Shah

20103047

Date: 26th December, 2016

CERTIFICATE

Certified that the work incorporated in the thesis entitled, “Understanding transcriptional regulation of the long noncoding RNA XIST”submitted by Rini Shah was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or Institution.

Prof. Sanjeev Galande

Supervisor

Date: 26th December, 2016

Acknowledgements

This thesis is a result of years of work done during my PhD, which has been a remarkable learning experience. Throughout this time I have been extremely fortunate to have very many people around who have contributed in different ways to make this journey a beautiful and memorable one. It gives me immense pleasure in extending a note of sincere gratitude to all of them.

First and foremost, I wish to thank my thesis supervisor, Prof. Sanjeev Galande for his constant encouragement, support and guidance all throughout my PhD. It has truly been a privilege working alongside him. He has been an excellent mentor, always open to new ideas and challenges and a big science enthuse. Thanks to him that I have had a wonderful and rigorous training which has instilled in me confidence to think, plan and execute the experiments independently. Working with him has been an ever learning experience which has helped me grow as a researcher and prepared me for the next stage of my career.

Acknowledgements are due to my Research Advisory Committee members, Dr. Sorab Dalal from ACTREC, Navi Mumbai and Dr. Girish Ratnaparakhi from IISER Pune. I thank them for their valuable time and insights on my work all these years which has definitely helped shaped this thesis.

I am deeply thankful to Prof. K.N.Ganesh, director, Prof. L.S.Shashidhara, head, biology division, Prof. V.S.Rao, former registrar and Col. G. Raja Sekhar, current registrar at IISER Pune for providing me with an opportunity to join IISER Pune as a graduate student and for their extended financial support to complete my PhD.

I would also like to express my gratitude to all the staff from academic section, IT section,accounts section, library and stores and purchase section of IISER for all their help and cooperation.

I wish to thank the funding source, Council of Scientific and Industrial Research for providing me with the fellowship for pursuing PhD.

My sincere gratitude to IISER Pune for providing state-of-the-art infrastructure and an exciting scientific environment to do research. It has been a pleasure to grow with the institute of national repute. I wish to thank the entire biology department for creating a healthy research environment and for all their suggestions during my annual talks and poster presentations. Special thanks to the biomanager Dr. Mrinalini Virkar as well as Shabnam Patil, Kalpesh Thakare and Piyush Gadekar for all their hard work in maintaining instruments and taking care of the consumable orders to ensure smooth running of the labs. I would also like to thank Mr. Mahesh Rote for all his help in the official matters. I take this opportunity to thank all the security and the housekeeping staff for their round the clock hard work.

I consider myself extremely privileged to have worked with a talented and enthusiastic bunch of people in the SG lab throughout my PhD. A big thankyou is due to Ranveer for being a wonderful senior and a teacher to me. He patiently taught me all the basic molecular biology techniques and guided me at the start of my journey in the lab. His cheerful mannerisms and ever helping attitude all throughout have been commendable. I would also like to duly acknowledge Praveena for her selfless help, support and contributions as an amazing friend. She has been a constant source of inspiration and taught me a great deal about life. Thankyou is not sufficient to pay my dues to her. Special thanks is due to the fellow X-men, Ashwin for

expanding my understanding of the field of X inactivation, for being an excellent critique and a fan of the work and for all his help in reviewing my thesis, reports and proposals. I am grateful for all your invaluable contributions to my project. I would like to express my gratitude to Manjunath and Kamal for sharing the same scientific zeal. All the discussions with you have truly helped my growth as a scientist. Thank you for being there for me when I needed the most. I would also like to thank Sunita for teaching me differentiation experiments. I have learnt a lot from you, especially the meticulous planning and execution of the experiments. I thank Rahul and Rafeeq for all their help in the lab, be it troubleshooting a failed experiment or planning an ambitious one. I feel extremely glad to have worked alongside the likes of such hard working guys. I wish to thank a great friend Rahul Sharma. Discussions and arguments with you from time to time have helped boost my morale and improve my scientific foresight. I thank Satyajeet and Mouli for all the interesting discussions related to my work and science in general. Your experiences have taught me a lot. I would also like to extend my gratitude to Dr. Krishanpal Karmodiya for his critical evaluation of my work and for teaching me basics of NGS analysis. I thank Manu, Ankitha Shetty, Saurabh, Suyog and Ankita Sharma for maintaining a fun-filled atmosphere in the lab. I wish to thank all the other present and past members of SG lab, Indumathi, Akhila, Ameya, Rutika, Ayush, Sneha, Priyanka, Sonam, Farhat, Vijay, Ram, Sameet, Dipeshwari, Pavana, Ishwar, Rachel, Suyash, Mithila, Trupti, Rasika, Shashikant, Amruth, Poornima, Meghna, Neha Sirwani, Neha Oli, Prachi, Georgina, Divya, Priyasha, Nethra, Tejas, Geeta and Aditi for providing an excellent work environment. I would like to sincerely thank our former lab manager, Vipradas Sir. He was extremely

organized, always helpful and very supportive. I also thank SG lab attendant Santosh for all the maintenance work in the lab. I also would like to thank present and past KK lab members, Mukul, Abhishek, Pratima and Madhavi.

I have been fortunate enough to be blessed with many dear friends and would like to take this opportunity to thank them for their kindness. Special thanks to my dear friends, Mithila, Kanika and Payal for their unconditional support throughout my PhD, especially during the times when I was at my lowest. Thank you guys for always being there, cheering me up and making my stay in Pune a truly awesome experience. I also thank Vimal, Shreeharsha, Abhishek , JP, Aniruddha, Vallari and Prema for all the fun moments spent discussing, debating and learning together. I would also like to thank my long-term MSc friends Shalmali, Sahir, Shweta, Sarita, Samay, Pooja, Arpit and DD for constantly motivating me to excel. I thank my dear friend Rashmi for her motivational speeches from time to time.

Finally, I could not have made it to where I am now without the infinite love, encouragement and support from my family. No amount of thank you is enough to express my deepest gratitude to my mother, father and brother for always believing in me and being a constant source of motivation. It is you all who give me the courage and strength to fight all the difficulties and keep moving ahead. I am forever indebted to you for everything.

- Rini

List of Contents

Contents

Synopsis 1

Chapter 1: General Introduction 13

1.1 Evolution of dosage compensation 13

1.2 Evolution of mammalian XY chromosomes 15

1.3 Evolution of X chromosome inactivation 16

1.4 X chromosome inactivation (XCI) 19

1.4.1 Meiotic sex chromosome inactivation (MSCI) 20

1.4.2 Imprinted XCI 20

1.4.3 Random XCI (rXCI) 21

1.4.3.1 Initiation phase of rXCI 23

1.4.3.2 Establishment phase of rXCI 26

1.4.3.4 Maintenance phase of rXCI 28

1.5 Regulation of Xist 29

1.5.1 Regulation of Xist during the initiation phase of rXCI 29

1.5.1.1 Regulation of Xist by XIC region 29

1.5.1.2 Regulation of Xist by pluripotency factors 33

1.6 Comparison of mouse Xist and human XIST 36

1.7 Rationale of the present study 37

1.8 References 38

Chapter 2: Regulation of XIST in the context of initiation phase of XCI 49

2.1 Introduction 49

2.1.1 SP1, YY1 and CTCF 51

2.1.2 OCT4, SOX2 and NANOG 53

2.1.3 SATB1 and SATB2 53

2.2 Rationale of the present study 55

2.3 Summary of the work 55

2.4 Materials and Methods 56

2.4.1 Cloning of XIST promoter elements into pGL3 Basic vector 56

2.4.2 Cloning of XIST exon1 regions into pGL3 Basic, XIST1.1 and XIST1 57

2.4.3 Cell Culture 57

2.4.4 RA-mediated differentiation of EC cells 58

2.4.5 Transfection of DNA and siRNA 58

2.4.6 Luciferase Reporter Assay 59
2.4.7 RNA extraction and cDNA synthesis 60
2.4.8 Chromatin immunoprecipitation 60
2.4.9 Protein extraction and immunoblotting 61
2.4.10 Quantitative real-time PCR 62
2.4.11 Antibodies, siRNAs and other reagents 63
2.5 Results 64
2.5.1 Human embryonic carcinoma (EC) cells responds to differentiation cues 64
2.5.2 Expression of XIST is upregulated upon RA-mediated differentiation of EC cells 66
2.5.3 Identification of promoter elements for XIST 67
2.5.4 SP1, YY1 and CTCF are the potential factors regulating transcription from XIST promoter 69
2.5.5 Pluripotency factors bind XIST Exon1 region 72
2.5.6 Pluripotency factors negatively regulate XIST 75
2.5.7 Depletion of OCT4, SOX2 and NANOG levels in EC cells does not lead to upregulation of XIST 78
2.5.8 Pluripotency factors and SATB proteins cross-regulate each other 80
2.5.9 SATB1 and SATB2 positively regulate XIST expression in EC cells 83
2.5.10 Chromatin organizer proteins SATB1, SATB2 and CTCF bind to two distinct on XIST Exon1 86
2.5.11 XIST Exon1 site with SATB1, SATB2 and CTCF binding might be involved in positive regulation of XIST 90
2.6 Discussion 92
2.6.1 Identification of XIST promoter 93
2.6.2 Identification of transcription factors governing transcription from XIST promoter 94
2.6.3 Role of pluripotency factors in regulating XIST expression in EC cells 96
2.6.4 Role of chromatin organizer proteins – SATB1, SATB2 and CTCF in regulating XIST expression in EC cells 97
2.7 References 103
Chapter 3: Regulation of XIST in the context of maintenance phase of X inactivation 110
3.1 Introduction 110
3.2 Rationale of the present study 112
3.3 Summary of the work 113
3.4 Materials and Methods 113
3.4.1 Cell Culture 113
3.4.2 Transfection of DNA and siRNA 114
3.4.3 Luciferase Reporter Assay 115

3.4.4 RNA extraction and cDNA synthesis 115
3.4.5 Protein extraction and immunoblotting 116
3.4.6 Quantitative real-time PCR 116
3.4.7 Antibodies, siRNAs and other reagents 117
3.5 Results 117
3.5.1 Assaying for the promoter activities in HEK293T cells 117
3.5.2 Effect of increasing the levels of SATB proteins on XIST minimal promoter activity and endogenous XIST levels 118
3.5.3 YY1 is the primary transcription factor regulating XIST promoter activity determined via biochemical assays 121
3.5.4 YY1 is the primary transcription factor regulating XIST expression in HEK 293T cells 125
3.5.5 SP1 and CTCF are also potential transcription factors regulating XIST expression in addition to YY1 128
3.5.6 CTCF binding affinities to the XIST genomic locus may be an important determinant of XIST expression 131
3.6 Discussion 134
3.7 References 138
Chapter 4: Conclusions and future directions 141
Appendix 1 143
Details of primers used for quantitative real time PCR of cDNA 143
Appendix 2 144
Details of primers used for cloning 144
Appendix 3 145
Details of primers used for ChIP – PCR 145
Publications 146

List of Figures

Chapter 1

Figure 1.1 Timeline for the evolution of XCI across several mammalian species 19
Figure 1.2 X inactivation cycle during female mouse development 22
Figure 1.3 Molecular events leading to XCI 23
Figure 1.4 Xist RNA interacting proteins 28
Figure 1.5 Tsix mediated repression of *Xist* 30
Figure 1.6 Jpx mediated upregulation of Xist expression 31
Figure 1.7 Rnf12 relieves Tsix mediated repression by targeting Rex1 for degradation 32
Figure 1.8 Regulation of *Xist* by pluripotency factors 35
Figure 1.9 Comparison between mouse and human *Xist*/*XIST* 36

Chapter 2

Figure 2.5.1 Human EC cells respond to RA-mediated differentiation cues 65
Figure 2.5.2 Expression of XIST increases during RA-mediated differentiation of EC cells 66
Figure 2.5.3 Identification of *XIST* promoter region by luciferase reporter assay 68
Figure 2.5.4 SP1, YY1 and CTCF are the potential transcription regulators of XIST 72
Figure 2.5.5 Pluripotency factors bind a site on *XIST* Exon1 75
Figure 2.5.6 Pluripotency factors negatively regulate *XIST* 77
Figure 2.5.7 Depletion of OCT4, SOX2, NANOG in EC cells does not cause upregulation of XIST 79
Figure 2.5.8 Pluripotency factor and SATB proteins cross-regulate each other in EC cells 83
Figure 2.5.9 SATB1 and SATB2 positively regulate XIST expression in EC cells 85
Figure 2.5. 10 SATB1, SATB2 and CTCF occupy two distinct sites on *XIST* Exon1 89
Figure 2.5.11 SATB1, SATB2 and CTCF binding to *XIST* exon1 (2.5 Kb downstream of TSS) might be positively regulating XIST transcription 92

Chapter 3

Figure 3.5.1 XIST promoter activity in HEK293T cells as measured via luciferase reporter assay 118
Figure 3.5.2 Effect of increasing the levels of SATB proteins on XIST minimal promoter activity and endogenous XIST levels 120

Figure 3.5.3 YY1 is the primary transcription factor regulating XIST promoter as determined biochemically 124
Figure 3.5.4 YY1 is the primary transcription factor regulating XIST expression in HEK 293T cells 128
Figure 3.5.5 SP1 and CTCF are also potential transcription factors regulating XIST expression in addition to YY1 131
Figure 3.5.6. CTCF binding affinities to the XIST genomic locus may be an important determinant of XIST expression 133

List of Abbreviations

AS	Antisense
Atrx	Alpha thalassemia/mental retardation syndrome X-linked
BAC	Bacterial artifical chromosome
BCA	Bicinchoninic acid
bp	Basepair
BSA	Bovine serum albumin
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP-sequencing
CSBS	Consensus SATB1 binding sequence
Ct	Threshold cycle
CTCF/Ctcf	CCCTCC binding factor
DCC	Dosage compensation complex
DMEM	Dulbecco's Modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dnmt3a	DNA methyltransferae 3a
Dnmt3b	DNA methyltransferae 3b
DTT	Dithiothretol
E	Embyronic day
EC	Embryonic carcinoma
EDTA	Ethylene di-amine tetra-acetic acid
EGTA	Ethylene Glycol Tetra-Acetic acid
EMSA	Electrophoretic mobility shift assay
ENCODE	Encyclopedia of DNA elements
ESC/ESCs	Embryonic stem cell
Ezh2	Enhancer of zeste
FBS	Fetal Bovine Serum
FISH	Fluorescent *in situ* hybridization
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
H2AK119ub	Histone H2A Lysine 119 Monoubiquitination
H3K27ac	Histone H3 Lysine 27 Acetylation
H3K27me3	Histone H3 Lysine 27 Trimethylation
H3K4m3	Histone H3 Lysine 4 Trimethylation
H3K4me	Histone H3 Lysine 4 Monomethylation
H3K9ac	Histone H3 Lysine 9 Acetylation
H3K9me3	Histone H3 Lysine 9 Trimethylation
H4K20me	Histone H4 Lysine 20 Monomethylation
Hdac3	Histone deacetylase 3
HEK 293T	Human embryonic kidney 293T
HEPES	(4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid)

hESC/hESCs	Human embryonic stem cell
HMG	High mobility group
hnRNP C	Heterogenous ribonuclear protein C
hnRNP K	Heterogenous ribonuclear protein K
hnRNP M	Heterogenous ribonuclear protein M
hnRNP U	Heterogenous ribonuclear protein U
Hprt	Hypoxanthine guanine phosphoribosyl transferase
HRP	Horse Raddish Peroxidase
IgG	Immunoglobulin G
iXCI	Imprinted XCI
Kb	Kilobase
KDa	KiloDalton
Klf4	Kruppel-like factor 4
LINE 1	Long interspersed nuclear elements 1
LNA	Locked nucleic acid
Lnc	Long noncoding
Lnx3	Ligand of numb-protein X 3
Mb	Million base
MEF	Mouse embryonic fibroblast
mESC/mESCs	Mouse embryonic stem cell
MSCI	Meiotic sex chromosome inactivation
MSL	Male specific lethal
MSUC	Meiotic silencing of unsynapsed chromatin
MYA	Million years ago
NaCl	Sodium chloride
$NaHCO_3$	Sodium bicarbonate
Nc	Noncoding
NEB	New England Biolabs
NT2D1	NTERA-2 clone D1
Oct4/OCT4	Octamer binding transcription factor 4
PAM	Parental antagonism model
PAR	Pseudoautosomal region
PAX6	Paired Box6
PBS	Phosphate buffered saline
PcG	Polycomb group
PCI	Phenol-chloroform-isoamylalcohol
PCR	Polymerase chain reaction
PFB	Pluripotency factor binding
Pgk1	Polyglycerate kinase 1
PMSC	Postmeiotic sex chromatin
PMSF	Phenylmethyl sulfonly fluoride
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PVDF	Polyvinylidene Difluoride

qRT	Quantitative real time
RA	All-trans-retinoic acid
Rbm15	RNA binding motif protein 15
Rex1/REX1	Ring-exported protein 1
RIPA Buffer	Radio immunoprecipitation assay buffer
RNA	Ribonucleic acid
RNA PolII	RNA Polymerase II
RNAi	RNA interference
Rnf12/RNF12	RING-H2 finger protein12
rRNA	Ribosomal RNA
RT	Reverse transcriptase
rXCI	Random XCI
Satb1/SATB1	Special AT-rich binding protein 1
Satb2/SATB2	Special AT-rich binding protein 2
SHARP	SMRT and HDAC associated repressor protein
siRNA	Small interfering RNA
Sox2/SOX2	SRY box containing gene 2
SOX3	SRY box containing gene 3
Sp1/SP1	Specificity protein 1
Sry/SRY	Sex determining region Y
STORM	Stochastic Optical Reconstruction Microscopy
TAF	TBP associated factor
TBP	TATA- binding protein
TSS	Transcription start site
TST	Tris buffered salin - Tween 20
UCSC	University of California SantaCruz
Wtap	Wilms tumour 1-associating protein
Xa	Active X
XCI	X chromosome inactivation
Xi	Inaative X
XIC	X inactivation centre
Xist/XIST	Xi-specific transcript
XistAR	Xist Activating RNA
Xm	Maternal X
Xp	Paternal X
Xpr	X pairing region
Yy1/YY1	Yin-Yang1

Synopsis

Abstract

XIST, Xi-specific transcript is a long non-coding RNA (lncRNA) made exclusively from the inactive X chromosome (Xi) in mammals. It plays an indispensable role in regulating X chromosome inactivation (XCI), a mechanism of dosage compensation in mammalian females. XIST physically coats the chosen X chromosome in *cis* initiating dramatic epigenetic reprogramming resulting in heterochromatinization of the entire chromosome rendering it stably inactive in a mitotically heritable manner. Since XIST is the pioneering molecule initiating the process of chromosome-wide silencing, regulating its timely activation is essential not only for the proper execution of XCI but also for the survival and the development of embryo. Although we understand a great deal about the process of XCI as well as regulation of *Xist* in mouse system, the mechanisms regulating the human *XIST* still remains elusive. Our study aims to address this aspect in the context of initiation as well as maintenance phases of XCI.

The work presented in this thesis demonstrates for the first time that both the promoter region and the gene body are essential for regulating the promoter of *XIST*. We have also identified and uncovered the role(s) of certain common as well as distinct factors in governing transcriptional activity from the promoter of *XIST*. More specifically, we demonstrate that the pluripotency factors (OCT4, SOX2, NANOG) bind to exon1 and act as repressors of *XIST* and hence are important determinants shaping temporal upregulation of *XIST* during the initiation phase. Our work also highlights the involvement of the chromatin organizer proteins – SATB1, SATB2 and CTCF and two previously reported transcription factors – SP1 and YY1 in controlling the expression of *XIST*. We discover YY1 to be the primary regulator of *XIST*, whereas SP1 and CTCF to be the secondary ones. Nevertheless, our work provides novel evidence towards the unique role played by CTCF in the process of *XIST* regulation. The findings from our study suggest that the pattern of CTCF enrichment on the several sites located in the *XIST* gene body as well as the identified promoter element can actually prove to be an essential dictating factor influencing the transcriptional outcome from its promoter. In summary, the current study identifies roles of multiple transcription

factors and explores their complex interplay in an attempt to provide the mechanistic basis of XIST regulation.

Introduction

Various mechanisms of dosage compensation have evolved time and againbecause any fluctuation(s) causing an imbalance in the gene expression can prove to be detrimental to the survival of the organism. The phenomenon of XCI is one of such dosage compensation mechanisms which evolved to resolve the gene dosage disparity between males and females of mammalian species. As the name suggests, it is manifested by stable silencing of one of the two X chromosomes in the mammalian females (Lyon, 1961). Studies over the past two decades have been instrumental in expanding our understanding of the subject of XCI at the molecular level. The foremost player initiating the process of silencing is the 17 Kb long non-coding RNA (lncRNA) (in the case of mouse), called Xist. Xist is upregulated from the X chromosome which randomly chosen to be inactivated during differentiation of embryonic stem cells (ESCs) or at the time of embryonic implantation. The accumulated Xist forms a repressive nuclear compartment, devoid of RNA PolII and transcription factors, into which X-linked genes relocate resulting in their silencing (Chaumeil et al., 2006; reviewed in Nora and Heard, 2010). The sequence of events occurring either simultaneous to or following the spreading of Xist RNA along the X chromosome is the recruitment of repressive chromatin modifiers such as Polycomb group (PcG) protein complexes - PRC1 and PRC2, leading to heterochromatinization of the entire chromosome (Chow and Brown, 2003; Li et al., 2012; Marks et al., 2009; Plath et al., 2003; Wakefield et al., 1997). Subsequently, the Xi shifts to late replication (Morishima et al., 1962; Takagi et al., 1982), incorporates a histone variant macroH2A (Mermoud et al., 1999) and the promoters of X-linked genes undergo CpG methylation (Cotton et al., 2015; Sharp et al., 2011), resulting in stable silencing of most of the genes on Xi.

Xist is kept repressed in the undifferentiated ESCs and in preimplantation embryos. Modulating its levels is essential to ensure timely manifestation of XCI during development. Hence, expression of *Xist* is tightly regulated by multiple players such as the pluripotency proteins (Oct4, Sox2, Nanog, Klf4, c-Myc, Rex1) as well as by various factors encoded from

the X inactivation centre (XIC) (lncRNAs Tsix, Jpx, Ftx and E3 ubiquitin ligase Rnf12). A brief summary of their roles is as follows:

(i) Tsix – It is one of the key *cis* regulators of *Xist* in mouse(Lee et al., 1999). *Tsix* codes for 40 Kb lncRNA, is arranged antisense to *Xist* in the genome and overlaps the whole of *Xist* gene. It has been clearly demonstrated that transcription from the *Tsix* promoter results in the enrichment of repressive histone modifications as well as DNA methylation on the promoter of *Xist*, thereby preventing its upregulation (Navarro et al., 2005; Ohhata et al., 2008; Sado et al., 2005).

(ii) Jpx – It escapes XCI and has recently been shown to be involved in positively regulating Xist in differentiating mESCs by causing eviction and titration of CTCF from the promoter of *Xist*(Sun et al., 2013).

(iii) Ftx - It also escapes silencing. However, there is no definite understanding about its role in regulating Xist expression aside from a single report wherein Xist promoter was shown to be hypermethylated leading to reduced Xist expression in male mESCs harbouring Ftx deletion (Chureau et al., 2011).

(iii) Rnf12 – It is an E3 ubiquitin ligase and has earned the reputation of X-encoded dose dependent activator (Barakat et al., 2014; Hosler et al., 1989; Patrat et al., 2009; Shin et al., 2010). It works through targeting Rex1, the positive regulator of *Tsix*, to degradation (Gontan et al., 2012).

(iv) Pluripotency factors – Both XCI and cellular differentiation and hence development are tightly coupled (Schulz et al., 2014). Therefore, it is essential to temporally control Xist levels. The pluripotency factors – Oct4, Sox2, Nanog, whose levels go down during differentiation, have been shown to link the two developmental events. Work from several labs has discovered pluripotency factors to be the repressors of *Xist* and activators of *Tsix* in stem cells. Specifically, Oct4, Sox2 and Nanog have been demonstrated to bind intron1 site of *Xist* gene, thereby keeping Xist repressed in the undifferentiated ESCs (Navarro et al., 2008; Nesterova et al., 2011). In addition to this, they also exert positive influence on Tsix expression either by directly binding to the 5' region of *Tsix* (Klf4, Rex1 and c-Myc) or indirectly by acting on Xite (Oct4, Sox2, Klf4), the latter being the enhancer of Xist (Donohoe et al., 2009; Navarro et al., 2010).

On the basis of findings described, it can be convincingly stated that Xist expression is robustly regulated by host of factors, acting synergistically or independently so as to ensure timely and proper execution of developmentally significant process of XCI. For the past two decades, mESCs have clearly been the system of choice to understand the molecular events leading to XCI including regulation of *Xist* gene. Our understanding with respect to the factors regulating *XIST* gene as well as the dynamics of XCI in other eutherian mammals is limited thus far. It is important to decipher the pathways of XCI in multiple systems so as to be able to address the conservation of the process which in turn will help in answering the fundamental question about the evolution of XCI. For the scope of the current work, we will be dealing with human *XIST* gene regulation.

Rationale of the present study

Although *Xist* gene from mouse and human were discovered around the same time and are also functionally conserved (Brown et al., 1991; Borsani et al., 1991), the understanding with regards to regulation of human *XIST* has been rather limiting. Unlike the ease of manipulating mESCs, studying the process by making use of human ESCs (hESCs) has been challenging, not only owing to associated ethical concerns but also due to variability in the system.Despite significant conservation of X-linked genes, chromosomal synteny as well as the process of XCI guided by XIST lncRNA between several eutherian mammals, XCI is initiated in diverse ways in different species (studied in Okamoto et al., 2011). One of the possible explanations could be evolutionary alterations in the developmental programmes across species necessitating robustness in the regulation of XCI to accommodate these changes (Okamoto et al., 2011). Additionally,lncRNA called Tsixthat isknown to be a critical negative regulator of *Xist* in mouse is not functionally conserved across species. We now know that the TSIX RNA discovered in human cells may not function similar to its murine counterpart since it is truncated, not homologous to mouse *Tsix*, does not extend into the promoter of *Xist* and it is made from the inactive X along with XIST in human foetal cells (Migeon et al., 2001; Migeon et al.,2002). Therefore, it can be stated that *XIST* in humans might be regulated differently and therefore addressing the mechanism(s) ofits regulation is of utmost importance.

A few studies have addressed the mechanisms of *XIST* promoter regulation. A pioneering study was contributed by Huntington Willard's group as early as 1997 (Hendrich et al., 1997). The authors of the paper attempted to identify and characterize the promoter and discovered that only the first 100 bases from the TSS of *XIST* gene shared significant homologies between mouse, rabbit, horse and human. Moreover, their work also led to identification of potential transcription factors such as SP1, YY1 and TBP, possibly involved in regulating the promoter of *XIST*. But the question whether these factors can actually bind the promoter in the cells and regulate *XIST* transcription remained unanswered until recently. Two independent studies which were published while we were pursuing our study, uncovered the role of YY1 as the key transcription factor regulating *XIST* transcription from two different promoters (Chapman et al., 2014; Makhlouf et al., 2014). As per studies carried out using mouse system, there are multiple pathways and molecules involved in controlling the expression of Xist in robust manner. Therefore, the quest for understanding the mechanism(s) controlling human XIST expression is far from complete. There are many important questions that are still left unaddressed. For instance, our knowledge with respect to *XIST* promoter behaviour during the initiation phase of XCI is still lacking. It is known that the levels of XIST are upregulated during a certain time window of development. But whether this is the consequence of YY1 acting on *XIST* promoter during a specific period or a concerted outcome of other additional players involved still remains elusive.Therefore, with the objective of unravelling the mechanism(s) of transcriptional regulation of human *XIST*, we undertook the current study with the following aims:

(1) To understand the transcriptional regulation of human *XIST* in the context of initiation phase of XCI.

(2) To understand the transcriptional regulation of human *XIST* in the context of maintenance phase of XCI.

Summary of the work done

1. To understand the transcriptional regulation of human *XIST* in the context of initiation phase of XCI

In the second chapter of the thesis, we present our findings with respect to regulation of *XIST* in human embryonic carcinoma (EC) cells. This cell line has been used as a proxy model for understanding the features of human ES cells as it expresses all the pluripotency factors and differentiates into neuronal lineage upon administering all-trans-retinoic acid (RA) (Andrews et al., 1980; Andrews et al., 1984). Also, undifferentiated EC cells express XIST in low amounts as assessed by RNA-FISH (Chow et al., 2003). Therefore, this provides a good system to probe the dynamic pattern of *XIST* promoter activity and carefully tease the roles of several other factors in tightly regulating the transcription from *XIST* promoter in the context of initiation phase of XCI.

Firstly, we demonstrate that EC cells respond to differentiation cues mediated by subjecting the cells to all-*trans*-retinoic acid (RA) treatment by assessing the expression of pluripotency and differentiation markers. Following this, we show for the first time that the expression of XIST is upregulated upon administration of RA to these cells. Next, we made use of ENCODE (Encyclopedia of DNA elements) data sets publicly available on UCSC (University of California, SantaCruz) genome browser and made the reporter constructs to better characterize the promoter element(s) of *XIST*. In addition to previously identified binding sites for SP1, YY1 and TBP observed close to TSS, we also observed the ChIP-sequencing peak for CTCF, both upstream as well as downstream of the TSS, in various ENCODE cell lines. We observed that the activity of various promoter constructs harbouring binding sites for the said factors remained comparable to the 100 bp region found to be partly conserved across several species as per the first study (Hendrich et al., 1997).

To determine the behaviour of *XIST* promoter constructs in response to the differentiation cues which provides the initiation context, we tested their activities in differentiating EC cells. Surprisingly, it was observed that while the expression of endogenous XIST increases, the activity of *XIST* promoter elements with binding sites for SP1, YY1 and CTCF decrease during the course of differentiation and so does the expression of these transcription regulators. These contradictory observations led us to speculate that region(s) other than

the promoter elements could be important determinants of *XIST* transcription. And indeed by performing ChIP we show that the pluripotency proteins OCT4, SOX2 and NANOG display a unique differentiation driven temporal pattern of binding to a distinct site on *XIST* exon1 which negatively correlates with the expression of XIST. Additionally, we show that forced expression of OCT4 or SOX2 leads to downregulation of XIST in a cell line wherein Xi is stably established and maintained (HEK 293T). We further determined XIST expression upon depletion of OCT4/SOX2/NANOG in EC cells. However, contrary to our expectation, we did not observe any upregulation in the levels of XIST. This suggested that yet unidentified factors might be involved in temporally modulating the levels of XIST lncRNA.

Based on the literature survey and preliminary data from our lab, we decided to test if the nuclear matrix proteins – SATB1 and SATB2 could regulate *XIST*. To address this, we monitored the expression of OCT4, SOX2, NANOG as well as XIST upon perturbing the levels of SATB1 or SATB2 in EC cells. The results obtained seem to suggest that SATB proteins positively regulate *XIST* and acts as repressors for pluripotency factors. Moreover, we also show that the chromatin organizer proteins – SATB1, SATB2 and CTCF bind to two distinct sites on *XIST* exon1. Interestingly, they also show similar binding kinetics as exhibited by the pluripotency proteins. Our work provides evidence that the site occupied by OCT4, SOX2, NANOG, SATB1 and SATB2 (~ 5.5 Kb downstream of TSS) might be involved in repressing *XIST* and the site bound by CTCF, SATB1 and SATB2 (~ 2.5 Kb downstream of TSS) might act as either an activator or repressor, possibly dependent upon its interaction with the upstream elements.

Collectively, the results discussed in the current chapter provide compelling evidence towards coordinated action by a multitude of factors leading to timely activation of transcription from *XIST* promoter.

2. To understand the transcriptional regulation of human *XIST* in the context of maintenance phase of XCI

It has long been thought that Xist is required only during the establishment phase of XCI and is actually dispensable during the maintenance phase in the differentiated cells despite its continuous synthesis (Brown and Willard, 1994; Csankovszki et al., 1999). The reason being that the alterations brought about by the accumulation of repressive epigenetic

modifications on the underlying chromatin architecture of Xi as a consequence of Xist RNA coating is an important step causing a shift from reversible, Xist-dependent silencing to irreversible, Xist-independent one (Wutz and Jaenisch, 2000). This widely held belief was countered by another study demonstrating the significance of Xist RNA in maintaining the perinucleolar positioning of Xi in MEF, which in turn is essential for maintenance of heterochromatin marks and stable silencing (Zhang et al., 2007).However, our knowledge about the role and regulation of *Xist*/*XIST* during the maintenance phase is lacking. Evidently much of the field's focus has been drawn towards understanding the events leading to the establishment of inactive X using mESCs and probing the questions pertaining to stable maintenance of Xi in mouse as well as human system have gained rather limited attention.

There have been a couple of recent studies addressing the aspect of *XIST* transcriptional regulation in the maintenance phase using fibroblast and mouse-human hybrid cell lines (Chapman et al., 2014; Makhlouf et al., 2014). Both these studies attribute the role of transcription factor regulating *XIST* to YY1. These findings were published while we were pursuing the work presented in the third chapter of the thesis. In our study, we set out to uncover the roles of SP1, YY1, CTCF as well as SATB1 and SATB2 towards governing the transcription from *XIST* promoter using HEK 293T cells which is a female cell line and harbours an established Xi. The results of our experiment revealed that the promoter activityexhibited by various reporter constructs remains more or less constant in the EC and HEK 293T cells. Interestingly, we observed that SATB proteins actually behaved as the repressors of *XIST* in HEK 293T cells as opposed to their roles as activators in EC cells. However, whether they exert their regulatory influence directly or indirectly remains to be determined.

Further we attempted to address the roles of SP1, YY1 and CTCF in the process by perturbing their levels and testing the promoter activities biochemically as well as scoring for endogenous XIST expression. Our results also support the published findings that YY1 is the primary factor governing transcription from XIST promoter (P1). Interestingly, upon depletion of SP1/YY1/CTCF for prolonged duration (72 hours) by performing RNAi-mediated knockdown, it was observed that even SP1 and CTCF can act as potential transcription factors for XIST promoter. Besides these findings, our work also discovers a unique mode of regulation brought about by CTCF. We demonstrate that the combination of high-affinity

and low-affinity CTCF binding sites, present on the *XIST* genomic locus (promoter as well as gene body) can have an important impact on the differential outcomes from the *XIST* promoter.

Collectively, the results discussed in this chapter suggest that the regulation of *XIST* in the maintenance phase is not as simple as considered so far and instead production of XIST in the somatic cells might be an outcome of concerted action of multiple factors acting either together or independently.

Conclusions

Based on the work presented in the thesis, we derive following conclusions:

(1) The promoter of XIST seems to extend beyond the previously considered minimal promoter region (100bp from the TSS) (Hendrich et al., 1997).

(2) The results obtained in our study provide compelling evidence towards the involvement of *XIST* exon1 region in controlling transcription from its promoter. More specifically, we demonstrate for the first time that the pluripotency factors (OCT4, SOX2 and NANOG) and the chromatin organizer proteins (CTCF, SATB1 and SATB2) exhibit a unique pattern of binding to two distinct sites on the XIST exon1 region. We believe that this can possibly have an impact on the transcriptional activity from the XIST promoter.

(3) Findings from our work attribute the role of *XIST* repressors to the pluripotency factors. This is in accordance with the results obtained using mouse system (Navarro et al., 2008).

(4) The three zinc-finger transcription regulators – SP1, YY1 and CTCF are important factors governing XIST transcription in the initiation as well as maintenance phases of XCI. Interestingly, CTCF seems to play a unique role in this process. It displays strong as well as weak mode of binding to multiple sites at the promoter as well as within the gene body of *XIST*. Our work highlights the significance of this observation and suggests that a combinatorial usage of these multiple CTCF binding motifs may presumably be involved in a looping mediated interaction between *XIST* exon1 and its promoter, thereby influencing firing of the promoter.

References

Andrews, P.W., Bronson, D.L., Benham, F., Strickland, S., and Knowles, B.B. (1980). A comparative study of eight cell lines derived from human testicular teratocarcinoma. Int. J. Cancer *26*, 269–280.

Andrews,P.W., I.Damjanov, D.Simon, G.Banting, C.Carlin, N.C.Dracopoli, and J.Fogh. 1984. Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2: differentiation *in vivo* and *in vitro*. *Lab. Invest.* 50, 147-162

Barakat, T.S., Loos, F., Van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J.A., and Gribnau, J. (2014). The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. Mol. Cell *53*, 965–978.

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

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.

Chapman, A.G., Cotton, A.M., Kelsey, A.D., and Brown, C.J. (2014). Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. BMC Genet. *15*, 89.

Chaumeil, J., Baccon, P. Le, Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear. Genes Dev. *20*, 2223–2237.

Chow, J.C., and Brown, C.J. (2003). Forming facultative heterochromatin: Silencing of an X chromosome in mammalian females. Cell. Mol. Life Sci. *60*, 2586–2603.

Chow, J.C., Hall, L.L., Clemson, C.M., Lawrence, J.B., and Brown, C.J. (2003). Characterization of expression at the human XIST locus in somatic, embryonal carcinoma, and transgenic cell lines. Genomics *82*, 309–322.

Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., and Rougeulle, C. (2011). Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. Hum. Mol. Genet. *20*, 705–718.

Cotton, A.M., Price, E.M., Jones, M.J., Balaton, B.P., Kobor, M.S., and Brown, C.J. (2015). Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum. Mol. Genet. *24*, 1528–1539.

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

Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature *460*, 128–132.

Giuseppe Borsani, Rossana Tonlorenzi, M. Christine Simmler, Luisa Dandolo, Danielle Arnaud, Valeria Capra, Markus Grompe, Antonio Pizzuti, Donna Muzny, Charles Lawrence, H.F.W., Avner, P., and Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *354*, 56–58.

Gontan, C., Achame, E.M., Demmers, J., Barakat, T.S., Rentmeester, E., van IJcken, W., Grootegoed, J.A., and Gribnau, J. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature *485*, 386–390.

Hendrich, B.D., Plenge, R.M., and Willard, H.F. (1997). Identification and characterization of the human XIST gene promoter: Implications for models of X chromosome inactivation. Nucleic Acids Res. *25*, 2661–2671.

Hosler, B. a, LaRosa, G.J., Grippo, J.F., and Gudas, L.J. (1989). Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. Mol. Cell. Biol. *9*, 5623–5629.

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

Li, Y., Tan, T., Zong, L., He, D., Tao, W., and Liang, Q. (2012). Study of methylation of histone H3 lysine 9 and H3 lysine 27 during X chromosome inactivation in three types of cells. Chromosom. Res. *20*, 769–778.

Lyon, M.F. (1961). Gene Action in the X-chromosome of the Mouse (Mus musculus L.). Nature *4773*, 372–373.

Makhlouf, M., Ouimette, J.-F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014). A prominent and conserved role for YY1 in Xist transcriptional activation. Nat. Commun. *5*, 1–12.

Marks, H., Chow, J.C., Denissov, S., Françoijs, K.-J., Brockdorff, N., Heard, E., and Stunnenberg, H.G. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res. *19*, 1361–1373.

Mermoud, J.E., Costanzi, C., Pehrson, J.R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J. Cell Biol. *147*, 1399–1408.

Migeon, B.R., Chowdhury, A.K., Dunston, J.A., and Mcintosh, I. (2001). Identification of TSIX , Encoding an RNA Antisense to Human XIST , Reveals Differences from its Murine Counterpart : Implications for X Inactivation. Am. J. Hum. Genet. *69*, 951–960.

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

Morishima, A., Grumbach, M.M., and Taylor, J.H. (1962). Asynchronous duplication of human chromosomes and the origin of sex chromatin. Proc. Natl. Acad. Sci. U. S. A. *48*, 756–763.

Navarro, P., Pichard, S., Ciaudo, C., Avner, P., and Rougeulle, C. (2005). Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: Implications for X-chromosome inactivation. Genes Dev. *19*, 1474–1484.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693–1695.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457–460.

Nesterova, T.B., Senner, C.E., Schneider, J., Alcayna-Stevens, T., Tattermusch, A., Hemberger, M., and Brockdorff, N. (2011). Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. Epigenetics Chromatin *4*, 17.

Nora, E.P., and Heard, E. (2010). Chromatin structure and nuclear organization dynamics during X-chromosome inactivation. Cold Spring Harb. Symp. Quant. Biol. *75*, 333–344.

Ohhata, T., Hoki, Y., Sasaki, H., and Sado, T. (2008). Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. Development *135*, 227–235.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

Patrat, C., Okamoto, I., Diabangouaya, P., Vialon, V., Le Baccon, P., Chow, J., and Heard, E. (2009). Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. Proc. Natl. Acad. Sci. U. S. A. *106*, 5198–5203.

Plath, K. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation. Science (80-.). *300*, 131–135.

Sado, T., Hoki, Y., and Sasaki, H. (2005). Tsix silences Xist through modification of chromatin structure. Dev. Cell *9*, 159–165.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell *14*, 203–216.

Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., and Antonarakis, S.E. (2011). DNA methylation profiles of human active and inactive X chromosomes. Genome Res. *21*, 1592–1600.

Shin, J., Bossenz, M., Chung, Y., Ma, H., Byron, M., Taniguchi-Ishigaki, N., Zhu, X., Jiao, B., Hall, L.L., Green, M.R., et al. (2010). Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. Nature *467*, 977–981.

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

Takagi, N., Sugawara, O., and Sasaki, M. (1982). Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. Chromosoma *85*, 275–286.

Wakefield, M.J., Keohane, A.M., Turner, B.M., and Graves, J. a (1997). Histone underacetylation is an ancient component of mammalian X chromosome inactivation. Proc. Natl. Acad. Sci. U. S. A. *94*, 9665–9668.

Wutz, A., and Jaenisch, R. (2000). A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation. Mol. Cell *5*, 695–705.

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

Chapter 1: General Introduction

Regulating gene dosage is crucial for organismal development as well as for its ability to respond to various environmental cues. Any gross perturbation(s) leading to the imbalance in gene expression can have deleterious effects on the survival of the organism. Hence, there are multiple mechanisms in place to ascertain and tightly control the correct amount of gene dosage. Broadly, this can be accomplished by either of the following two ways:

(i) by equally modulating the expression from both alleles. For example, dosage compensation of X chromosome in hermaphrodites of *Caenorhabditis elegans*(Burgoyne, 1993; Meyer and Casson, 1986; Thornhill and Burgoyne, 1993; Tsunoda et al., 1985) and (ii) by differentially regulating both alleles. In the second mode of regulation, expression from one of the alleles is either lowered or silenced completely. The key for following such a strategy of expression optimization must be efficient coordination between the two alleles. The best understood examples of this method of gene regulation are mammalian X chromosome inactivation (XCI) and autosomal imprinting. For the scope of this thesis, I will elaborate upon dosage compensation and XCI.

1.1 Evolution of dosage compensation

Dosage compensation is the phenomenon of equalising the gene expression from sex chromosomes between males and females in sexually dimorphic species. The necessity of dosage compensation on chromosome-wide scale arose with the evolution of chromosome based sex determination. Across the animal kingdom, sex determination depends either on the environmental factors or the genetic constitution (reviewed in Payer and Lee, 2008). The environment based sex determination is common in few fish and reptile species, wherein the environmental cues – particularly egg incubation temperature post fertilization decides the sex of the offspring (reviewed in Crews, 2003). There is no need for dosage compensation in these species as males and females have identical chromosomes and lack any specific sex chromosomes. This mode of sex determination is beneficial in that the offspring of better adapted sex is preferentially produced, maximizing the reproductive fitness in a given environmental condition. However, the major disadvantage of this system

is that sudden environmental changes such as global warming can be detrimental to the existence of the organism.

It has been suggested that the sex chromosomes evolved *de novo* in the ancestors of today's reptiles, birds, mammals upon acquisition of sex-determining gene(s) on one of the two autosomes that superseded the effect of temperature and hence, transiting from the environment guided sex determination to the predetermined one, based on chromosomal constitution (Muller, 1914) (reviewed in Bachtrog, 2006). Such sex determination system comprises of a homogametic sex (one with two identical sex chromosomes) and a heterogametic sex (one with two different sex chromosomes). There are two distinct systems depending on the inheritance of the sex chromosomes. In the first scheme followed by birds and few reptiles, males represent the homogametic sex, with two Z chromosomes and females are heterogametic, harbouring Z and W chromosomes. The second system is a well-characterized XY-based sex determination observed in most mammals. Here, the males are heterogametic with XY and females are homogametic with XX. Elaborating on the XY system, both the sex chromosomes shared significant homology earlier in the evolutionary time-scale. However, they seem to have diverged presumably due to chromosomal rearrangements and accumulation of maleness specific genes on the Y chromosome over time. This led to suppression of recombination outside of a small stretch called the pseudoautosomal region (PAR), driving independent evolution of both the sex chromosomes. As a consequence, degeneration of gene-content on the Y chromosome ensued (reviewed in Graves, 2006), creating the dosage inequality problem owing to the following reasons :

(i) The equivalents of these genes are present in two copies in the homogametic sex.

(ii) The autosomal genes are also present in two copies.

Hence, there is a wide difference between the ratio of the sex chromosomal and autosomal gene outputs in the males and females of the same species. To circumvent this issue, various dosage compensation mechanisms have evolved multiple times in different species (reviewed in Payer and Lee, 2008). Exemplifying this, it has been demonstrated that in the invertebrates such as *Drosophila melanogaster* and *Caenorhabditis elegans*, sex is determined by measuring X to autosome ratio. However, both the species have devised completely different strategies of compensating for gene dosage imbalance. In flies for instance, males (XY) upregulate the gene expression from their single X chromosome

twofold as compared to females (XX) (Mukherjee and Beermann, 1965).This transcriptional enhancement is regulated by roX RNA containing protein complex called Male Specific Lethal (MSL) which binds to and spreads along the X chromosome in males modifying the underlying chromatin status (reviewed in Lucchesi and Kuroda, 2015). In contrast, the hermaphrodites in nematodes (XX) exhibit reduction in expression from both their X chromosomes by half to match the expression levels in males (XO) (Meyer and Casson, 1986; reviewed in Strome et al., 2014). This downregulation is manifested by the action of Dosage Compensation Complex (DCC) which binds and spreads *in cis* along the X chromosomes in hermaphrodites causing alterations in histone modifications (Ercan et al., 2007; McDonel et al., 2006; reiewed in Strome et al., 2014). The dosage compensation mechanism selected in mammals is transcriptional silencing of one of the two X chromosomes in females rendering it functionally inactive (reviewed in Lee, 2011). To understand the evolution of XCI, it is imperative to first look at the evolution of mammalian sex chromosomes.

1.2 Evolution of mammalian XY chromosomes

Mammals are categorised into three groups depending on their divergence on the evolutionary timescale – (1) The prototherians (e.g., platypus) are the closest relatives to birds and reptiles and emerged around 165 million years ago (mya). (2) The metatherians (e.g., kangaroo, opossum) diverged from other mammals 150 mya and (3) The eutherians (e.g., rodents, humans) are the placental mammals which emerged around 100 mya (reviewed in Waters et al., 2007). All three groups have XY based sex determination system, wherein males are the heterogametic sex. To trace the ancestry of XY pair in mammals, many researchers have compared and mapped the genes on X and Y chromosomes between species belonging to these three categories. Platypus, an egg-laying mammal provides an ideal system to determine the evolution of therian XY system. It harbours an overwhelming 5 sets of sex chromosomes, with males of $X_1Y_1X_2Y_2X_3Y_3X_4Y_4X_5Y_5$ karyotype and females of $X_1X_1X_2X_2X_3X_3X_4X_4X_5X_5$ karyotype. These 10 sex chromosomes align in a chain during meiosis to ensure faithful segregation. Interestingly, X_1Y_1 on one end of the chain share the maximum homology with each other and X_5Y_5 on the other end show complete divergence. This indicates that X_1Y_1 might be evolutionary the youngest and X_5Y_5, the oldest. Moreover, X_5 chromosome contain *DMRT1* gene, which is the sex-determining gene present on the Z chromosome in birds suggesting that bird-like ZW system might be ancestral to the XY

system in mammals. However, extensive FISH based mapping and BAC sequencing of stretches of platypus sex chromosomes carried out by Jennifer Graves' lab yielded no evidence of homology with their therian counterparts (Veyrunes et al., 2008). Therefore, it can be speculated that therian XY pair evolved *de novo* after the monotreme lineage branched off and can be rightly called neoX and neoY chromosomes. This hypothesis is validated by another independent study assessing the mobility of retrogenes from X chromosome to autosomes (Potrzebowski et al., 2008).

Therian species acquired a key maleness determining gene – *SRY* (sex determining region Y), which encodes a high mobility group (HMG) – box transcription factor from their Y chromosome (Sinclair et al., 1990). *SRY* seems to be a truncated version of *SOX3* gene, which is autosomal in non-mammalian vertebrates and monotremes but X-linked in therian mammals (Wallis et al., 2007; Waters et al., 2007). Hence, the mutation of an ancestral proto-sex chromosome gene – *SOX3* into *SRY* could have paved the way for testis determining switch. This can be considered as an important and decisive event in shaping up the divergence between X and Y chromosomes, ensuing the degeneration of Y chromosome. The human X chromosome (165 Mb) contains about 1000 genes with the functions ranging from housekeeping in nature to brain development (Ross et al.,2005). The X chromosome is highly conserved across various placental species. The comparison of X chromosome sequences from human and mouse has revealed remarkable synteny between them. In contrast, the Y chromosome is much smaller (~60 Mb) and codes for only 45 unique proteins, with a majority involved in male-specific specialized functions (Skaletsky et al., 2003). If the dosage difference is left uncorrected, it will result in natural aneuploidy which is not tolerated. Hence, the need for dosage compensation arose which is achieved by XCI in the mammalian females.

1.3 Evolution of X chromosome inactivation

Evolution of XCI across several mammalian species is illustrated in Figure 1.1. However, the evolutionary basis of XCI is highly debated and still remains an open question. Many theories have been put forth to explain this phenomenon. Susumu Ohno formulated a theory suggesting stepwise evolution of XCI: (1) two-fold expression from the X chromosome in both the sexes, solving X:AA (X chromosome:Autosome) imbalance problem in males, (2) inactivation of one of the two Xs in females to restore the balance (Ohno, 1967). For this

theory to be true, the expression ratio of X chromosome to autosomes should be close to 1 in both males and females. Indeed the global X:AA expression ratio as assessed by microarray based gene expression profiling was discovered to be close to 1 in both the sexes of several mammalian species including human, macaque, mouse and rat (Nguyen and Disteche, 2005). However, a recent transcriptome-wide study using RNA-sequencing analysis yielded contradictory results and X:AA ratios were found to be close to 0.5 and 0.3 respectively in human and mouse male and female tissues (Xiong et al., 2010). Moreover, the expression ratio of orthologs present on the X in mammals and on autosomes (presumably an ancestor to proto-X chromosome) in chicken also turned out to be close to 0.5, hence refuting Ohno's hypothesis (He et al., 2011; Lin et al., 2012). These results were challenged by another recent report which reanalysed the same transcriptome data by taking into consideration the tissue-restricted expression patterns of many X-linked genes and found evidence for upregulation of X-linked genes in comparison to the autosomal genes, hence supporting Ohno's hypothesis (Deng et al., 2011). Interestingly, RNA PolII occupancy as well as active epigenetic modifications were also found to be enriched at highly expressed genes on X chromosome compared to autosomes, providing a mechanistic basis to the observed upregulation (Deng et al., 2011; Yildirim et al., 2011). Another possibility is that not all the genes are compensated to the same extent. The genes coding for proteins involved in complexes seem to be dosage sensitive since attaining proper stoichiometry is crucial to the functioning of protein complex (Hall and Wayne, 2012; Pessia et al., 2012). However, such haploinsufficiency is known only for a few X-borne genes. Moreover, few studies reported increase in the expression of X-linked genes (Lin et al., 2012; Pessia et al., 2012), while another study demonstrated decrease in the expression of autosomal genes (Julien et al., 2012). Hence it is also possible that such a mechanism could have evolved on a gene-by-gene basis. Nonetheless, genes escaping Xist-mediated inactivation offer one of the most striking evidences of such haploinsufficiency based mechanism. Studies employing the approach of allele-specific ChIP-sequencing for certain active and repressive histone modifications from both mouse and human systems have convincingly demonstrated a distinct epigenetic profile on the escape genes versus silenced genes which correlates strongly with the allele-specific RNA-sequencing profile and quantitative real time PCR (Carrel and Willard, 2005; Yang et al., 2012; Berletch and Ma et al., 2015; reviewed in Peeters et al., 2014). A few of these escapees are pseudoautosomal

genes and hence present on both X and Y chromosomes and therefore not dosage sensitive. While the number of identified escapee genes is around 3% of the total X-linked genes in mouse, the figure is as high as 15% in humans. The expression from these genes is the primary cause of phenotypic abnormalities observed in the humans with X chromosome aneuploidies such as Turner's syndrome and Klinefelter's syndrome.

Based on the findings discussed, if XCI initially evolved to counteract the hyperactivation of X-linked genes in females as proposed by Ohno, it seems perplexing as to why does XCI affect the majority of X-linked genes, when only a minority of them show hyperactivation. An alternative theory – parental antagonism model (PAM), proposed by Haig attempts to address this issue (Engelstädter and Haig, 2008; Haig, 2006). According to this model, XCI evolved not as a means of dosage compensation but instead as a method to silence growth inhibiting genes on the X chromosomes during the embryonic development. This model is devised on an assumption that growth inhibiting genes are enriched on the X chromosome. There are some evidences which seem to be consistent with this notion. For example, it has been demonstrated that XO or XY embryos develop faster than XX embryos until XCI is established (Burgoyne, 1993; reviewed in Schulz and Heard, 2013; Thornhill and Burgoyne, 1993; Tsunoda et al., 1985). A recent study demonstrated that the presence of two X chromosomes interferes with the developmental signaling networks and prevents exiting from the pluripotent state until one of the two Xs is inactivated, hence providing a mechanistic basis to the previous observations (Schulz et al., 2014). Another salient feature of PAM is that it predicts that XCI should have evolved in the groups where parental conflicts over maternal resource allocation to embryonic growth are stronger, which is indeed the case in therian mammals. However, this model still remains a hypothetical one since it has not been subjected to rigorous testing.

Another possibility is the coupling of evolution of XCI to sex determination. It has been observed that the overexpression of *Sox3*, gene ancestral to *Sry*, induces the formation of testis (Sutton et al., 2011). Hence, it may have been crucial to balance the amount of *Sox3*/*Sry* gene pair to ensure proper development of females and males respectively. Therefore, it is postulated that initial function of XCI would have been to maintain these dosage differences. Supporting this theory is the fact that *Sox3* is among the X-inactivated genes in the females (Splinter, et al., 2011). Intriguingly, *Rnf12* and *Lnx3*, genes crucial for the evolution of XCI, were possibly physically close to *Sox3* genes on the proto-sex

chromosomes (~180 Mya), as per the analysis carried out in birds (reviewed in Graves, 2008). Although the implication of this theory seems fascinating, it still remains speculative.

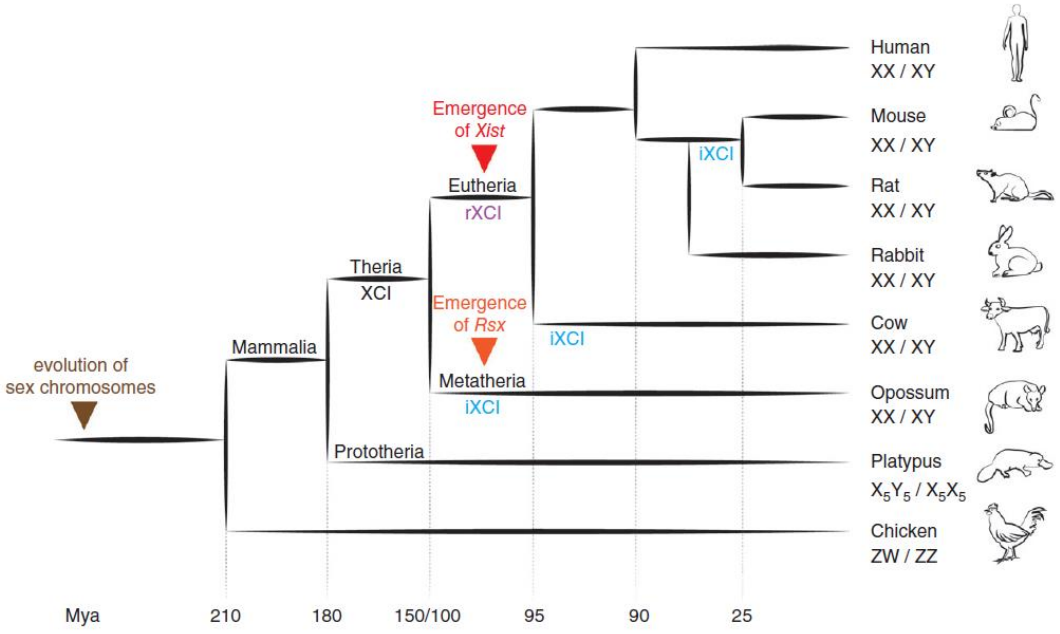


Figure 1.1 Timeline for the evolution of XCI across several mammalian species.

The X-axis here represents the timeline as Million years ago (Mya) and the Y-axis indicates several mammalian species. iXCI = imprinted XCI and rXCI = random XCI. Imprinted XCI occurs in metatherians mammals mediated by lncRNA Rsx. Evolution of random XCI correlates with the emergence of Xist lncRNA in eutherian mammals. The mechanisms of iXCI as well as rXCI are discussed in the later sections of this chapter. The illustration is adapted from Furlan and Rougeulle, 2016.

1.4 X chromosome inactivation (XCI)

The mammalian X chromosome gathered unprecedented attention in the field of genetics for more than a decade before the discovery of XCI. The puzzle in the field was to frame a hypothesis that could satisfactorily explain why the excess expression of most X-linked genes from the two X chromosome in females as opposed to a single X in males was not harmful or how was this compensated for. As early as 1949, the presence of sex chromatin in the female cats had already been demonstrated (Barr and Bertram, 1949). A key observation which proved to be instrumental in the discovery of XCI came from the work carried out in Susumu Ohno's lab. Ohno's work on mice and rats identified that one X chromosome appeared more condensed and heterokypnotic in contrast with the other, which showed similar staining pattern as the autosomes (Ohno, 1959; Ohno and Hauschka, 1960). Based on these prior findings and the phenotypic observations made by her on various X chromosome mutants such as *tortoiseshell*, *mottled*, *brindled* and *tabby* (Lyon,

1960), Mary Lyon came up with a unified hypothesis of XCI as the dosage compensation mechanism explaining all these evidences in her seminal paper published in 1961 (Lyon, 1961). Since the discovery, the quest for understanding the mechanism of XCI gained significant attention worldwide. XCI can be classified into three types – Meiotic sex chromosome inactivation, imprinted XCI, random XCI.

1.4.1 Meiotic sex chromosome inactivation (MSCI)

MSCI occurs in the male germline of many species (reviewed in Cloutier and Turner, 2010; Kelly and Aramayo, 2007). It has been shown in mice that during the sister chromatid pairing at the pachytene stage of meiosis, unpaired regions on both autosomes and sex chromosomes are transcriptionally silenced by a process called Meiotic silencing of unsynapsed chromatin (MSUC) (Turner et al., 2005). The function of MSUC might be to serve as a genomic defence mechanism against the spreading of transposons or retroviruses as well as to avoid the chromosomal abnormalities in the mature gametes. As discussed before, X and Y chromosomes have diverged significantly and can no longer form a chromatin synapse except for a short stretch (PAR). During MSCI in mice, the unpaired regions are recognized by DNA double-strand break repair pathway, resulting in the establishment of repressive histone marks and hence causing transcriptional silencing (reviewed in Cloutier and Turner, 2010; Kelly and Aramayo, 2007). Silenced sex chromosomes form XY-body or sex chromatin and this state is maintained as postmeiotic sex chromatin (PMSC) throughout spermatogenesis except for the genes required during the process, which are reactivated by an unknown mechanism (Namekawa et al., 2006). MSCI might be the ancestral form of XCI since silencing mechanism as part of genome defence strategy was already present. Moreover, when a silent paternal X (Xp) is inherited in the female embryo, the X-linked gene dosage parity will be automatically resolved in the males and females. Such an imprinted mechanism of XCI is predominantly used in the marsupials where Xp is always inactivated. Indeed, recent studies have documented evidence for MSCI and PMSC in opossum (Namekawa et al. 2007; Showell and Conlon, 2007).

1.4.2 Imprinted XCI

In imprinted XCI, the choice of X chromosome to be inactivated is non-random. The paternal X chromosome inactivated during spermatogenesis (MSCI) is inherited as such and maintained in all the cells except epiblast cells throughout preimplantation development.

After implantation, the imprinted Xp undergoes global transcriptional silencing in the extraembryonic tissue (Huynh and Lee, 2003). Evidence for such parent-of-origin effect came from the studies looking at the development of XO mouse embryos with X chromosomes of different parental origins. While XpO embryos were developmentally retarded during early post-implantation stages, XmO embryos were indistinguishable from XX controls. These observations suggested that Xp fails to provide the appropriate dosage of X-linked genes (Burgoyne et al., 1983; Jamieson et al., 1998; Thornhill and Burgoyne, 1993). Such method of XCI has been adapted by marsupial females (Cooper DW, VandeBerg JL, Sharman GB, 1971; Sharman GB, 1971) as well as extraembryonic tissues of mouse (Huynh and Lee, 2003; IOkamoto et al., 2004; Takagi and Sasaki, 1975; Mak et al., 2004). Understanding of the regulation of imprinted XCI in eutherian mammals at the molecular level has emerged from the studies carried out on mice oocytes and preimplantation embryos. Multiple evidences suggest that the imprinting of genomic region of long non-coding (lnc) RNA – *Xist*, the primary player in XCI is a key to imprinted XCI. There are evidences for both, establishment of repressive imprint on the maternal X in the form of repressive histone modification (H3K9me3) and chromatin condensation of *Xist* genomic region during oogenesis (Fukuda et al., 2014, Fukuda et al.,2015), as well as that of permissive imprint of the paternal X during MSCI (Sun et al., 2015).

1.4.3 Random XCI (rXCI)

The random XCI (rXCI) happens in the epiblast cells which give rise to embryo proper. Prior to rXCI, the imprinted silencing of Xp is reversed and it is activated during the generation of epiblast cells. Upon implantation there is a second wave of XCI which is stochastic in nature and both paternal as well as maternal X chromosomes have an equal probability of being silenced (Kay et al., 1993; Williams et al., 2011) (illustrated in figure 1.2). In this mode of XCI, the cells follow n-1 rule, wherein only one X chromosome is kept active per diploid set of autosomes. It has been documented that XO females do not exhibit XCI and one of the Xs in XXY males (Klinefelter's Syndrome) is inactivated (Grumbach et al., 1963; Monkhorst et al., 2009). Random XCI has been demonstrated to occur in the entire soma of mouse as well as both extraembryonic and embryonic lineages in humans. As a result, the somatic tissues in female are a mosaic of cells with either the paternally derived or the maternally derived X inactive. Emergence of random XCI correlates with the evolution

of gene encoding for lncRNA – Xist. Important model systems to understand the molecular events underlying XCI have been early mouse embryos as well as mESCs derived from the inner cell mass of blastocyst. Female mESCs (XX) have been extensively used for this purpose since they recapitulate the rXCI upon induction of differentiation *in vitro*. Based on the work done using these systems over the past two decades, the process of rXCI can be categorized into initiation, establishment and maintenance phases (shown as an illustration in Figure 1.3).

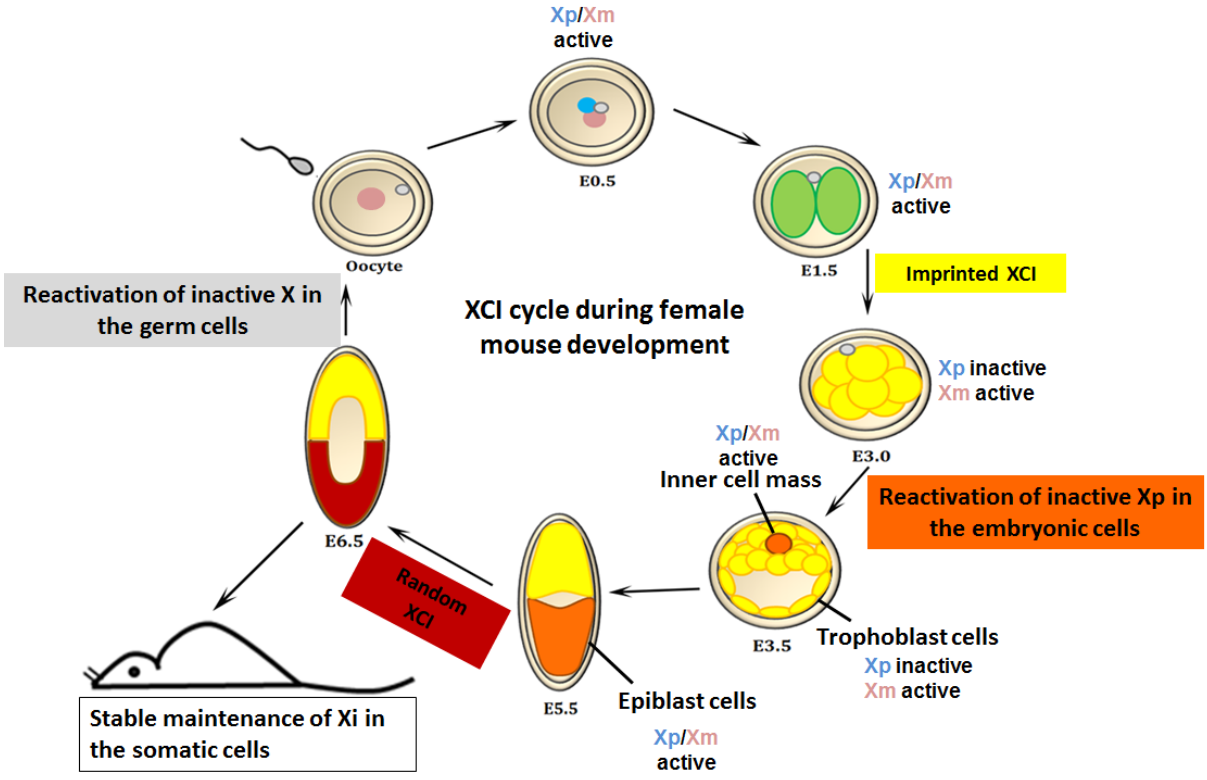


Figure 1.2 X inactivation cycle during female mouse development.

The figure illustrates the cycle of X inactivation during mouse embryonic development. Paternal and the maternal X chromosomes (Xp and Xm) are active in the zygote and 2-cell stage. Imprinted XCI (iXCI) occurs during the morula stage (yellow). The inactive X is reactivated in the cells of the inner cell mass (brown) and iXCI is maintained in the trophoblast and primitive endoderm cells. Random XCI is initiated in the epiblast cells at E5.5 (red). Once inactivated, Xi is stably maintained in the somatic cells of the animal. Xi is reactivated in the primordial germ cells. The illustration is modified from (Wutz, 2011).

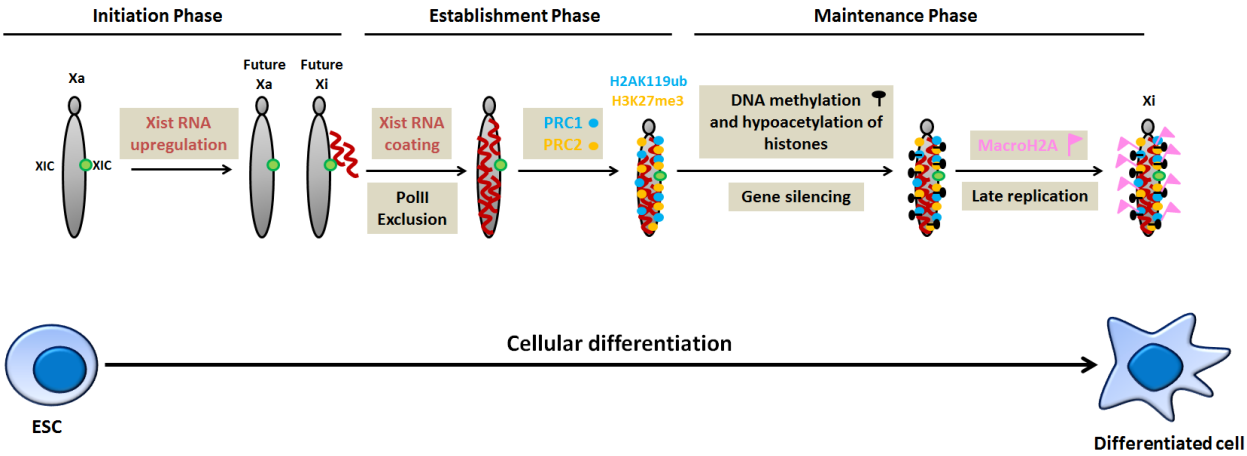


Figure 1.3 Molecular events leading to XCI.

In ESC, both the X chromosomes are active. Upon induction of differentiation, Xist RNA is upregulated from one of the Xs (red) (Initiation phase). In the establishment phase of XCI, Xist RNA coats the X chromosome in cis causing RNA Polymerase II exclusion. This is followed by the recruitment of Polycomb repressive complexes (PRC1 and PRC2) which brings about H2AK119ub and H3K27me3 modifications respectively, leading to heterochromatinization of the entire chromosome. Subsequently, the promoters of X-linked genes undergo DNA methylation and hypoacetylation of histones causing global gene silencing. Finally, macroH2A is recruited on Xi and the inactive X shifts to late replication. DNA methylation and epigenetic modifications ensure stable maintenance of Xi (Maintenance phase).

1.4.3.1 Initiation phase of rXCI

The initiation of XCI in mammalian females is developmentally regulated. While both X chromosomes are active in the early zygote (Epstein et al., 1978), rXCI is initiated during the implantation stage, concomittant with the differentiation of embryonic stem cells. Initiation of XCI is controlled by a 4.5 Mb locus in mouse called the X-inactivation centre (XIC)(Lee et al., 1996). Several studies over the past two decades have documented that XIC locus codes for many lncRNAs, a majority of which are involved in the initiation of XCI (reviewed in Maclary et al., 2013). The two most important lncRNAs are Xist (Xi-specific transcript) and Tsix. *Xist*, evolved from a protein coding gene *Lnx3*, is 17Kb long in mice and 19Kb long in humans and it is exclusively made from the inactive X (Xi) (Brown et al., 1991; Borsani et al., 1991; Kinzler et al., 2006). *Tsix*, 40Kb lncRNA in mice and absent in humans, is synthesized specifically from the active X chromosome (Xa) (Lee et al., 1999). Briefly, the function of Xist is to physically coat the X chromosome in cis, followed by the downstream chromatin modifying events resulting in the transcriptional silencing of the chosen X (Figure 3). And the function of Tsix is to prevent one of the Xs from meeting the same fate. Since, all Xs in excess of one are inactivated per diploid set, it suggests that the cells can sense and count the number of X chromosomes before choosing n-1 for inactivation. Therefore, it is

important to understand the events occurring prior to initiation of XCI and Xist coating. Multiple models have been put forth to explain the process of counting as discussed below:
(i) The one-factor model proposes that XCI is the default pathway and that it is regulated by autosomally encoded blocking factor made in quantities sufficient only to bind and repress single XIC (Rastan, 1983).

(ii) The two-factor model takes into consideration an X-encoded competence factor in addition to the blocking factor coded by autosome. The model predicts that the amount of blocking factor is enough only to titrate the competence factor on a single XIC. The other XIC can get occupied by the remaining competence factor leading to the onset of XCI (Gartler and Riggs, 1983). Though this model was proposed before the discovery of mechanism of action Xist and Tsix, it helped explain the experimental observations from Tsix deletion (Lee and Lu, 1999).

(iii) The stochastic model suggests that the autosomal factors promote *Xist* repression, for example by inducing Tsix, whereas the X chromosome makes activators of *Xist*. The balance of these factors is ought to change as the Xist spreads on one of the X chromosomes inducing silencing, thereby reducing the levels of activator. This is a comprehensive model since it not only predicts the probability of single X staying active in the cells but also the feedback and checkpoint mechanisms to ensure only a single X remains active (Monkhorst et al., 2008).

These models are neither mutually exclusive nor sufficient to reconcile for all the available data. Nevertheless, they do provide substantial framework to integrate the experimental observations thus far and determine the course of future experiments. Evidences from multiple lines of investigation have led to the identification of proposed autosome encoded repressors (pluripotency factors) as well as X-encoded activators of *Xist* (Rnf12, Jpx, Ftx). A complex picture emerges from these studies with respect to regulation of Xist and hence XCI (discussed later in the section describing Regulation of *Xist* promoter).

After the counting step all but one X upregulates the expression of Xist. A number of experimental evidences stated below conclusively demonstrated that Xist RNA is both necessary and sufficient for chromosome wide silencing:

(1) Xist RNA is made only from the Xi (Brockdorff et al., 1991; Brown et al., 1991).

(2) Xist expression is upregulated in the preimplantation embryos and precedes the onset of inactivation (Kay et al., 1993, Kay et al., 1994).

(3) Targeted deletion of *Xist* from one of the X chromosomes in mESCs (XX) leads to non-random XCI of the X bearing the Xist allele, suggesting the absolute requirement of Xist coating in cis for successful XCI (Penny et al., 1996)

(4) Xist RNA specifically localizes to inactive X at interphase (Clemson et al., 1996; Duthie et al., 1999)

(5) Integration of *Xist*/XIC containing transgenes into the autosomes lead to Xist coating and spreading in cis, altering the underlying chromatin architecture causing transcriptional inactivation (Lee et al., 1996; Lee and Jaenisch, 1997; Herzing et al., 1997;Heard et al., 1999;Wutz and Jaenisch, 2000; Hall et al., 2002;Chow et al., 2007; Tang et al., 2010).

The critical regulator of *Xist* in mice is the antisense lncRNA Tsix. Following studies investigated the role of Tsix in regulating the process of rXCI:

(1) Tsix is exclusively made from the active X chromosome (Lee et al., 1999).
(2) Upon targeted deletion of *Tsix* from one of the X chromosomes in mESCs (XX), the X chromosome with the deletion always undergoes XCI (Lee and Lu, 1999; Nesterova et al., 2003).
(3) Transcription from the *Tsix* promoter brings about repressive chromatin alterations over *Xist* promoter, thereby preventing Xist upregulation (Navarro et al., 2005; Ohhata et al., 2008; Sado et al., 2005).

This tug of war between different pathways and the molecules thereof paves the way for stochastic upregulation of Xist from one of the X chromosomes (future Xi) and Tsix from the other X (future Xa).

1.4.3.2 Establishment phase of rXCI

The accumulated Xist forms a repressive nuclear compartment which is devoid of RNA PolII and transcription factors, into which X-linked genes relocate concomitant with their silencing (Chaumeil et al., 2006; reviewed in Nora and Heard, 2010). The sequence of events occurring either simultaneous to or following the spreading of Xist RNA along the X chromosome is the recruitment of repressive chromatin modifiers such as Polycomb group (PcG) protein complexes such as PRC1 and PRC2. As a result, chromatin marks associated with gene activation such as H3K9ac, H3K4me3 are depleted and heterochromatin marks such as H3K27me3, H4K20me are enriched (reviewed in Chow and Brown, 2003; Plath et al., 2002; Li et al., 2012; Marks et al., 2009; Wakefield et al., 1997). Subsequently, the Xi shifts to late replication (Morishima et al., 1962; Takagi et al., 1982), incorporates a histone variant macroH2A (Mermoud et al., 1999) and the promoters of X-linked genes undergo CpG methylation (Cotton et al., 2015; Sharp et al., 2011), resulting in stable silencing of most of the genes on Xi. It is also known that some of the genes such as *Jarid1c* can escape inactivation (reviewed in Berletch et al., 2010; Yang et al., 2010).

Despite gaining mechanistic understanding of the process, the kinetics of Xist spreading over 160 Mbp mouse X chromosome has remained an intriguing question in the field. The fact that X:Autosome translocations show only partial silencing has led to the speculation that the X chromosome bear certain permissive features aiding in Xist spreading process (Cotton et al., 2014; Duthie et al., 1999; Hall et al., 2002b; Keohane et al., 1999; Popova et al., 2006). To solve this puzzle, LINE1 repeat hypothesis was postulated wherein the repetitive elements, highly enriched on the X chromosome compared to autosomes, were speculated to enhance the spreading of inactivation (Lyon, 2000). Indeed it was demonstrated that LINE1 elements facilitate the heterochromatinization of X chromosome undergoing silencing (Chow et al., 2010). At this point, it seems rather intuitive that the chromatin organization might actually play a pivotal role in bringing about this chromosome-wide change. One of the very first studies addressing this aspect provided evidence for the role of Xist itself in shaping the architecture of Xi which is different from Xa (Splinter et al., 2011). Interestingly, a couple of recent studies have unequivocally demonstrated that Xist RNA actually spreads along the X chromosome by a proximity based transfer mechanism, not dictated by the linear distance from *Xist* locus but instead aided by the three dimensional

chromatin architecture of Xi (Engreitz et al., 2013; Simon et al., 2013). However, both the studies observed an anti-correlation between Xist spreading and LINE1 repeat sequences and thus questioning their hypothesized role as booster elements. Nevertheless, it can be speculated that LINE1 elements may be involved in shaping the higher-order topological folding of the chromosome assisting in Xist spreading. Moreover, the authors also demonstrated that the hotspots for Xist RNA-binding on the X chromosome correlate remarkably well with the enrichment of Ezh2 and the H3K27me3 modification it causes (Engreitz et al., 2013; Simon et al., 2013), concurrent with the earlier evidences of coordinated recruitment and spreading of Xist and PRC2 (Maenner et al., 2010; Pinter et al., 2012; Sanulli et al., 2015; Sarma et al., 2014; Sunwoo et al., 2015; Zhao et al., 2008). Although the described experimental claims are well supported, the kinetics of XCI as measured by allele-specific RNA sequencing of differentiating mESCs offer somewhat contradictory results. The aforementioned study particularly observes that the dynamics of silencing over X chromosome follows a linear pattern, with the genes closest to XIC undergoing inactivation earlier than the distally located ones (Marks et al., 2015).

It has long been thought that the action of Xist may be facilitated by its protein partners. The modular structure of Xist RNA can act as docking sites for proteins. For example, the nuclear matrix associated heterogenous ribonuclear protein U (hnRNPU) was one of the first proteins to be shown to bind A-repeat of Xist RNA and determine its chromosomal localization (Hasegawa et al., 2010). Recent technological innovations has made it possible to systematically explore the protein interactome of Xist RNA in order to better understand the chromatin architecture guided silencing process (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015; Moindrot et al., 2015; Monfort et al., 2015). In addition to the previously characterized interactor such as hnRNPU, these studies also discovered many novel interactors such as histone modifiers - SHARP/SPEN and PRC2, chromatin organizers – CTCF, cohesin, nuclear membrane protein – Lamin B receptor as well as RNA-binding proteins – hnRNPK, hnRNPC, hnRNPM, to name a few. Various Xist RNA-binding proteins identified in several studies are depicted in the illustration below (Figure 1.4) along with their functions.

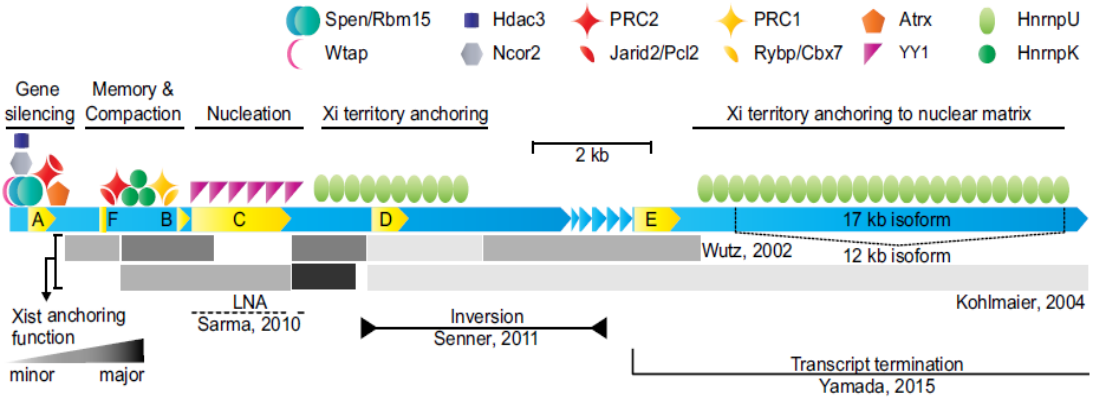


Figure1.4 Xist RNA interacting proteins.

The schematic summarises the proteins identified to be the interactor of Xist lncRNA repeat elements (A-F) along with their functions. The illustration is adapted from Pinter, 2016.

1.4.3.4 Maintenance phase of rXCI

Once established, the inactive X is stably maintained in the somatic cells in mitotically heritable manner. An intriguing question is to understand the mechanism by which the fates of two epigenetically diverse chromosomes – Xa and Xi are irreversibly maintained in the nucleus. Some of the initial studies suggested that the function of the XIC and Xist RNA is crucial only during the establishment of XCI and despite its continued synthesis, it is dispensable during the maintenance phase (Brown and Willard, 1994; Csankovszki et al., 1999). However, careful examination of Xi position in the nucleus indicated that its perinucleolar localization is affected upon loss of Xist RNA, further leading to loss of repressive histone marks and reactivation of certain X-linked genes upon continued passaging of the cells (Zhang et al., 2007). Whether the observed alterations in the chromatin state of Xi is the consequence of loss of Xist RNA or due to its dissociation from the nucleus still remains an open question. Nonetheless, a speculation in this regard can be made based on the recent findings providing compelling mechanistic evidence for the role of Xist RNA in repelling cohesin proteins on Xi to evade acquisition of transcriptionally-permissive state (Minajigi et al., 2015). Both these results put together give a perception for the active involvement of Xist RNA in maintaining the inactive state of Xi during the maintenance phase. Interestingly, using super-resolution microscopy (Stochastic Optical Reconstruction Microscopy), it has been elegantly demonstrated that the stoichiometry of Xist RNA molecules to Xi in the maintenance phase is much lower than previously thought. As per this study, calculating the number of Xist RNA as well as PRC2 molecules at a single

cell level, there are only 50-100 binding stations for Xist RNA and PRC2 (Buzin et al., 1994; Sun et al., 2006; Sunwoo et al., 2015). Therefore, it can be concluded that Xist RNA does not coat the entire Xi and the epigenetic modifications serve to retain the robustness of silencing in the maintenance phase.

1.5 Regulation of Xist

As discussed in the previous sections, Xist lncRNA is the master regulator for the process of XCI. Hence, keeping its levels under check is of utmost significance. Clearly, Xist is required both during the initiation as well as maintenance phases of XCI. Regulation of *Xist* is determined by certain common as well as variable factors during both these phases.

1.5.1 Regulation of *Xist* during the initiation phase of rXCI

Monoallelic upregulation of *Xist* and hence the initiation of rXCI coincides with the differentiation of embryonic stem cells into specific cell types. During the embryonic development the levels of Xist are dynamically regulated. Being a molecule of central significance for the process of XCI, there are multiple mechanisms in place to robustly regulate its expression.

1.5.1.1 Regulation of *Xist* by XIC region

The XIC encompasses several lncRNAs and other elements that are known to affect Xist expression and activity as discussed below.

1) Regulation by Tsix

The critical *cis* regulator of *Xist* in mouse is its antisense lncRNA, Tsix. Tsix is 40 Kb long and its gene-body extends into and spans the whole of *Xist* gene. It is evident from the studies employing heterozygous targeted deletions or truncations of Tsix that the choice of XCI becomes skewed towards the mutant allele bearing X chromosome in such scenarios (Lee and Lu, 1999; Nesterova et al., 2003). Mechanistically, Tsix transcription rather than the transcript is involved in altering the chromatin state of *Xist* promoter, thereby causing the repression (Shibata and Lee, 2004; Navarro et al.,2005; Sado et al., 2005; Ohhata et al., 2008). Also, Tsix transcript plays an active role in recruiting DNA methyltransferase to *Xist* promoter thereby repressing it during differentiation process (Navarro et al., 2006; Sun et al., 2006). However, it has also been reported that embryos mutant for Dnmt3a and Dnmt3b do not show any dysregulation in XCI (Sado et al., 2004). But disruption of Dnmt3a and

Dnmt3b leads to aberrant upregulation of Xist in differentiated male mESCs, suggesting that methylation of *Xist* promoter may be required for *Xist* repression in the maintenance phase (Beard et al., 1995). The mechanism of action of Tsix on *Xist* promoter is illustrated in a schematic (Figure 1.5).

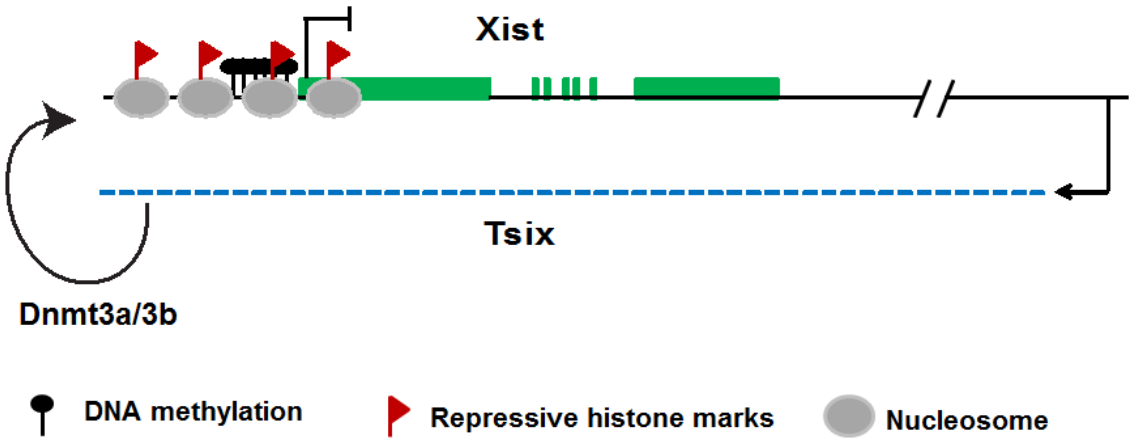


Figure 1.5 Tsix mediated repression of *Xist*.

Transcription of Tsix across *Xist* gene locus causes repressive histone modifications on the *Xist* promoter. Tsix also helps in the recruitment of de novo DNA methylatransferases Dnmt3a and Dnmt3b to the *Xist* promoter promoting DNA methylation, hence repressing *Xist* expression. Shown in green are the exons of *Xist* and blue dashed lines represent Tsix transcript.

2) Regulation by Jpx

Jpx maps to 10 Kb upstream of *Xist* promoter and transcribed in the opposite orientation to *Xist* (Johnston et al., 2002). The observation that it is upregulated in differentiating female mESCs and it actually escapes inactivation suggested its putative role in XCI. Indeed, female mESCs with heterozygous deletion of *Jpx* exhibit increased cell death and fewer cells show Xist coating upon differentiation (Tian et al., 2011). It has also been demonstrated that Jpx acts as a dose-dependent activator of *Xist* by driving overexpression of Jpx with varying promoter strengths in WT female mESCs (Sun et al., 2013). Both these studies indicate that Jpx might be the positive regulator of *Xist*. Mechanistic evidence for the same emerges from the latter report describing the role of Jpx in activating *Xist* upon differentiation of female mESCs by causing eviction and titration of CTCF from *Xist* promoter, thereby releasing the repression mediated by CTCF (Sun et al., 2013) (illustrated in Figure 1.6). Basically, Jpx in female cells is sufficient to directly bind CTCF at the P2 promoter of *Xist* causing its removal from the site which in effect leads to transcription from the promoter. However, similar pathway is prevented in males since the dose of Jpx from a single X chromosome is not

sufficient for the action. Based on their experimental evidences, Sun et al., 2013 have ascribed the role of blocking factor to CTCF and that of competence factor to Jpx.

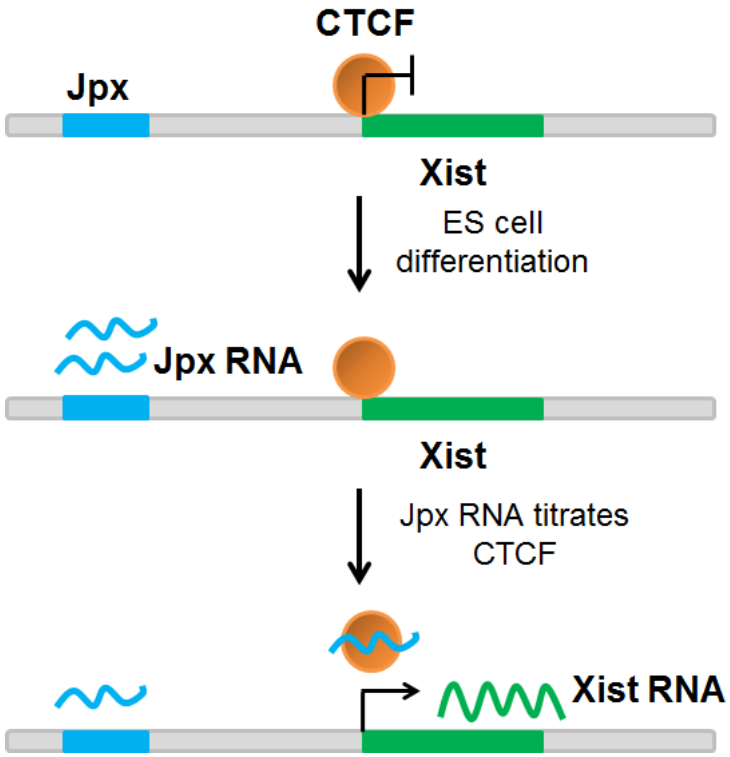


Figure 1.6 Jpx mediated upregulation of Xist expression.

In ESC, CTCF occupies *Xist* promoter and prevents its expression. Upon induction of differentiation, expression of Jpx is upregulated which directly interacts with CTCF and removes it from the *Xist* promoter thereby causing upregulation of Xist lncRNA. The illustration is modified from Sun et al., 2013.

3) Regulation by Ftx

Ftx lncRNA gene lies approximately 150 Kb upstream of *Xist* and is transcribed in the same orientation as *Xist* (Chureau et al., 2002). Similar to Jpx, Ftx is also upregulated during differentiation of mESCs and escapes Xist-mediated silencing (Chureau et al., 2011). There is no defect in terms of Xist expression and activity in the female embryos bearing paternally deleted Ftx allele, asserting that it is dispensable for imprinted XCI (Soma et al., 2014). The role of Ftx in regulating Xist expression remains largely unclear aside from a single report wherein *Xist* promoter was shown to be hypermethylated leading to reduced Xist expression in male mESCs harbouring Ftx deletion (Chureau et al., 2011).

4) Regulation by Rnf12

Rnf12 is a protein coding gene, located 500 Kb upstream of *Xist*. It was identified as E3 ubiquitin ligase, whose ectopic expression leads to inactivation of the single X chromosome in male mESCs and both the chromosomes in the female mESCs (Jonkers et al., 2009). Further work from the same group has provided mechanistic basis for Rnf12 action, which is through targeting Rex1 for degradation upon differentiation of mESCs (Gontan et al., 2012). Rex1 is a pluripotency factor and follows the expression kinetics similar to Oct4, Sox2 and Nanog. It positively regulates *Tsix* in undifferentiated mESCs and hence causes repression of *Xist* (Navarro et al., 2010). However, upon differentiation, levels of Rnf12 increase which in turn ubiquitinates Rex1 and targets it for degradation, relieving Rex1/Tsix mediated repression of Xist expression (Gontan et al., 2012) (shown in Figure 1.7). Furthermore, in a study employing an elegant strategy of XX-XY diploid heterokaryons, it was shown that the physical contact between the X chromosomes is not necessary for XCI to initiate as considered thus far (Bacher et al., 2006; Ogawa and Lee, 2003; Xu et al., 2007). Instead, a trans-acting factor serves as an indicator of X chromosome numbers and dosage. Rnf12 was identified to be the trans-acting factor which along with certain *cis*-acting elements such as Jpx, Ftx, Xpr can uplift Tsix mediated repression of *Xist* in X:X pairing independent manner (Barakat et al., 2014). In addition to this, another important characteristic which attributes the function of X-encoded dose-dependent activator to Rnf12 is its early silencing mediated by Xist, shown in the case of imprinted XCI (Hosler et al., 1989; Patrat et al., 2009; Shin et al., 2010). It is also noteworthy that Rnf12 itself may be negatively regulated by the pluripotency factors, which suggests that the pluripotency factors can influence the expression kinetics of Xist indirectly (Barakat et al., 2011; Kim et al., 2008).

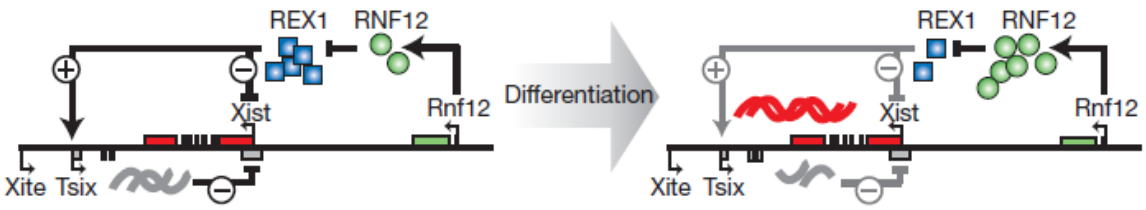


Figure 1.7 Rnf12 relieves Tsix mediated repression by targeting Rex1 for degradation.

In ESC, levels of Rex1 are high which positively regulates Tsix and hence represses *Xist*. Upon differentiation, expression of Rnf12 increases which cause degradation of Rex1, thereby releasing Tsix mediated repression on *Xist*. The illustration is adapted fromGontan et al., 2012.

5) Regulation by 5' end of Xist – A-repeat and Xist AR

5' end of *Xist* is highly conserved and functional end of the Xist lncRNA. It harbours A-repeat comprising of conserved direct and inverted repeats, which leads to the formation of secondary stem-loop structure. Study by Wutz et al., 2002 provided the foremost evidence highlighting the role of A-repeat in initiating Xist mediated silencing. The authors demonstrated that Dox-inducible Xist cDNA lacking the A-repeat accumulates on the X chromosome in male ES cells upon induction but is defective in initiating chromosomal silencing (Wutz et al., 2002). Subsequent studies have unequivocally demonstrated that the A-repeat actually acts as the docking site for a number of histone modifiers contributing to the silencing function of the element (See figure 1.4). Another study sought to determine the function of A-repeat in regulating XCI during embryogenesis in mouse (Hoki and Kimura et al., 2009). Towards this, the authors generated $Xist^{\Delta A}$ mouse and demonstrated that upon paternal inheritance of the defective allele, X chromosome fails to undergo iXCI in the extraembryonic tissues leading to selective loss of female embryos. Interestingly, this study also demonstrated that even though the basal transcription of Xist is not affected upon deletion of A-repeat in undifferentiated male ES cells, synthesis of Xist transcript is abrogated during embryogenesis in the mutant embryos. The discrepancies between the two studies described above can be resolved by a recent discovery of Xist Activating RNA (XistAR) (Sarkar, Gayen and Kumar et al., 2015). This study identifies and characterizes the function of novel lncRNA XistAR, which is encoded within exon 1 of Xist, transcribed in the antisense orientation to *Xist* only from Xi, and extends to its 5' end. The authors propose that the act of transcription of XistAR in the antisense orientation may serve to induce active chromatin changes on *Xist* locus, thereby enhancing its transcription. On the basis of these findings, it can be speculated that the abrogation of *Xist* transcription observed upon deletion of A-repeat in previous report (Hoki and Kimura et al., 2009) can possibly be an outcome of lack of XistAR transcript.

1.5.1.2 Regulation of *Xist* by pluripotency factors

Xist is induced during preimplantation stage in the embryo and early during differentiation of female mESCs. Both XCI and cellular differentiation and hence development are tightly coupled (Schulz et al., 2014). Therefore, it is essential to temporally control Xist levels. It is an established fact that there are two waves of XCI during mouse

embryonic development and the dynamics of inactivation/reactivation correlate with the levels of pluripotency factors. Moreover, despite being a critical *cis* regulator of *Xist*, heterozygous deletion of *Tsix* does not lead to aberrant upregulation of Xist in undifferentiated mESCs (Morey et al., 2001). These indicate that there are additional factors involved in modulating Xist levels at the onset of XCI. The pluripotency factors – Oct4, Sox2, and Nanog that are downregulated during differentiation, have been shown to link the two developmental events. In a study by Navarro et al., 2008, it has been demonstrated that loss of Nanog or Oct4 leads to upregulation of Xist in undifferentiated male and female mESCs. The authors also show that these factors directly bind intron1 region of *Xist* and hence, may be involved in repressing *Xist* in undifferentiated ES cells directly (Navarro et al., 2008). Evidence supporting this claim came from another study employing the strategy of deleting intron1 from *Xist* genomic locus and observing an increase in the expression of Xist (Nesterova et al., 2011). In addition to this it has also been shown that *Tsix* is also under the control of pluripotency factors as depicted in the Figure 1.8. They exert positive influence on Tsix expression either by directly binding to the 5' region of *Tsix* or indirectly by acting on Xite, the latter being the enhancer of *Xist* (Donohoe et al., 2009; Navarro et al., 2010). A complex picture of coupling between the onset of XCI and the state of the cells emerges from both the studies described, which is summarized in the schematic shown below (Figure 1.8).

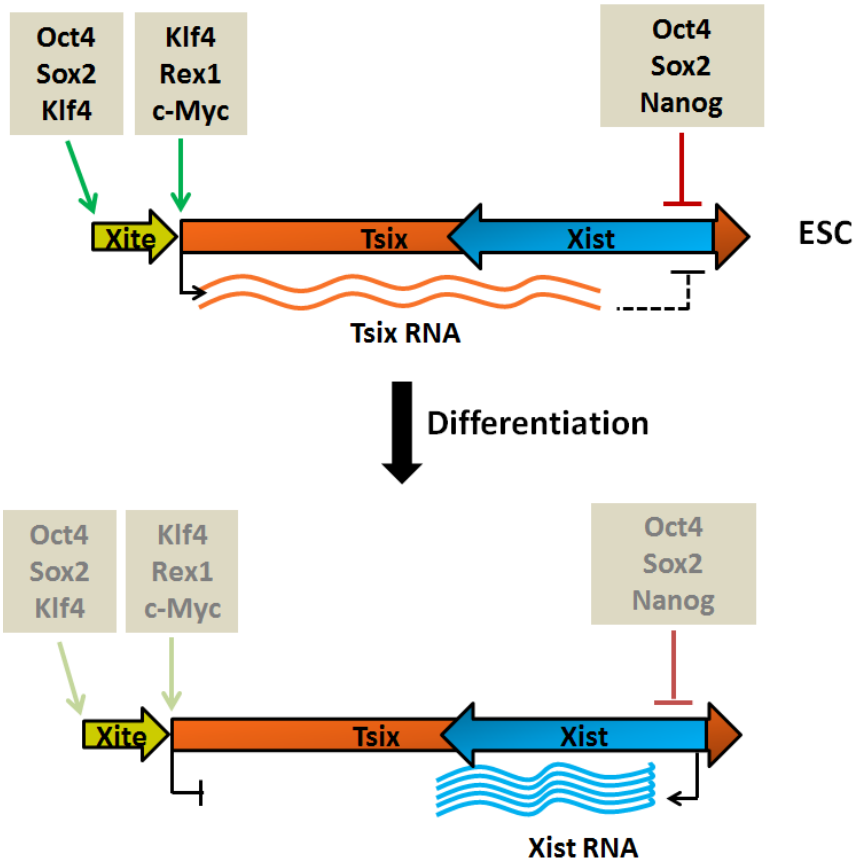


Figure 1.8 Regulation of *Xist* by pluripotency factors.

Pluripotency factors bind to intron 1 of *Xist* and repress it. They also positively regulate Tsix by binding to its 5' region or its enhancer, Xite in ESC. Upon differentiation, the expression of these factors decreases and hence Xist is upregulated.

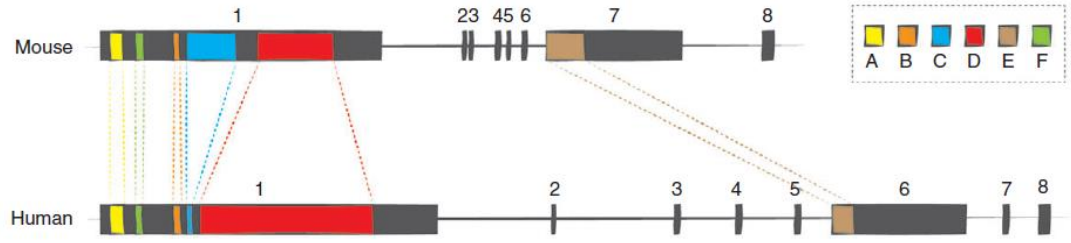
One of the major caveats is that perturbing the levels of pluripotency factors can cause multiple alterations and affect known *Xist* regulators such as Rnf12 (Navarro et al., 2011) and hence an indirect effect on Xist/Tsix levels cannot be ruled out. To circumvent this issue, two independent studies generated a mESC line as well as a mouse model with *Xist* intron1 region deleted. Contrary to the previous reports, these two studies unequivocally demonstrated that intron1 region of *Xist* with the binding sites for Oct4, Sox2 and Nanog is actually dispensable for *Xist* regulation (Barakat et al., 2011; Minkovsky et al., 2013).

On the basis of above mentioned findings, it can be convincingly stated that Xist expression is robustly regulated by host of factors acting synergistically or independently so as to ensure timely and proper execution of developmentally significant process of XCI. For the past two decades, mESCs have been the system of choice to understand the molecular events leading to XCI including regulation of *Xist* gene. Our understanding with respect to

the factors regulating *XIST* gene as well as the dynamics of XCI in other eutherian mammals is limited thus far owing to technical difficulties and ethical concerns. It is important to decipher the pathways of XCI in multiple systems so as to be able to address the conservation of the process which in turn will help in answering the fundamental question about the evolution of XCI. For the scope of the current work, we will be dealing with human *XIST* gene regulation.

1.6 Comparison of mouse *Xist* and human *XIST*

Human XIST mature transcript is 19 Kb long. Ectopic insertion of human *XIST* transgene in the murine and human cells induces XCI in these cells (Hall et al., 2002a; Heard et al., 1999; Migeon et al., 1999). This signifies the functional conservation between the properties of mouse and human *Xist*/*XIST*. However, at the sequence level, they display only 49% conservation, with the maximum homology observed in the 1st exon (refer Figure 1.9). XIST transcript across mammalian species harbors six repeat elements, A-F, which show the highest degree of sequence similarity and quite possibly functional conservation too (Brockdorff et al., 1992; Hendrich et al., 1993; Nesterova et al., 2001).



Repeat	Content and structure	Function		Transcribed	Conservation
		Silencing	*Xist* localization on Xi		
A	A-rich;2 stem loops, 4 repeats each	✓	✗	✓	✓
B	C-rich	✗	✗	✗	✓
C	C-rich; 14 tandem repeats	✗	✓	✗	✓
D	Complex, truncated	✗	Partial	✗	Poor
E	T-rich; 5 tandem repeats	✓	✓	✗	✓
F		✗	?	✗	✓

Figure 1.9 Comparison between mouse and human *Xist*/*XIST*.

Maximum conservation at the level of sequence as well as function is observed between repeat elements present in the first exon. The various repeat elements are colour coded as shown on the right side. Numbers indicate various exons. The table below illustrates the comparison between the repeat elements present in mouse and human *Xist*/*XIST*. The schematic is adapted from Makhlouf and Rougeulle, 2011.

Apart from this, even the developmental window during which XIST is expressed is distinct from its mouse counterpart. Not only is XIST induced in both the male as well as the female embryos from the maternal as well as paternal X chromosomes (in female embryos) as early as blastocyst stage, it also coats the X chromosomes. However, the coating may not be efficient as the chromosome-wide silencing is not initiated as inferred from H3K27me3 staining and RNA-FISH for X-linked genes (Okamoto et al., 2011). Based on their results, the authors suggest that in humans, induction of XIST expression and random choice of X chromosome to be inactivated is not coupled. Also, it is remarkable that although the process of XCI is conserved across eutherians and is dependent on XIST RNA, there are diverse ways in which the process is initiated in different species. They speculate that these discrepancies arose because of poor conservation of *TSIX* lncRNA across species. We now know that the TSIX RNA discovered in human may not function similar to its murine counterpart since it is truncated, not homologous to mouse *Tsix* , does not extend into the promoter of *Xist* and it is made from the inactive X along with XIST in human fetal cells (Migeon et al., 2001; Migeon et al.,2002).

The spatiotemporal kinetics of the cascade of events during the establishment of XCI in human is still elusive. Unlike the ease of manipulating mESCs, studying the process by making use of human ESCs (hESCs) has been challenging, not only owing to associated ethical concerns but also due to variability in the system. Few research groups have attempted to isolate and characterize female hESCs with respect to their XCI dynamics by staining for XIST RNA and H3K27me3 (Dvash et al., 2010; Hoffman et al., 2005; Silva et al., 2008; Tomoda et al., 2012). On this basis, the stem cell lines were classified into 3 distinct classes:

(1) Class I – Able to recapitulate XCI in culture upon differentiation like mESCs system.
(2) Class II – Spontaneously differentiates and induces XCI.
(3) Class III – Already undergone XCI but loses XIST expression in culture.

1.7 Rationale of the present study

The mechanisms regulating mouse *Xist* gene have been very well understood as discussed elaborately in the section 1.5. However, studies on mechanisms regulating human *XIST* have been limited thus far. As mentioned before, ethical concerns and technical

difficulties with hESCs have been the rate limiting factors. Evidently, XIST is the first and the foremost player of XCI. Therefore, deciphering its regulation is of utmost importance which will eventually pave the way towards addressing the most fundamental questions related to evolutionary conservation of the process at large. In the current study we aim to understand the molecular mechanism of transcriptional regulation of human *XIST* by characterizing its promoter elements in the context of initiation as well as maintenance phases of XCI. Since very little knowledge exists about the transcription factors involved in directly controlling XIST expression in both the scenarios, we sought to identify and understand the interplay between different factors potentially governing XIST expression. More specifically, we attempt to probe the connections between the pluripotency factors and XIST expression in the perspective of initiation phase (Chapter 2), as observed in mESCs. XIST continues to be synthesized during the maintenance phase but the transcription factors governing its robust expression still remains elusive. The current study makes use of inadequate previous knowledge in this regard as well as attempts to identify and characterize novel factors involved in maintaining XIST expression levels (Chapter 3). Overall, we present our findings towards the characterization of the transcriptional regulation of human *XIST* by a multitude of factors and unravel the complex regulatory interplay involved.

1.8 References

Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat. Cell Biol. *8*, 293–299.

Bachtrog, D. (2006). A dynamic view of sex chromosome evolution. Curr. Opin. Genet. Dev. *16*, 578–585.

Barakat, T.S., Gunhanlar, N., Pardo, C.G., Achame, E.M., Ghazvini, M., Boers, R., Kenter, A., Rentmeester, E., Grootegoed, J.A., and Gribnau, J. (2011). RNF12 activates Xist and is essential for X chromosome inactivation. PLoS Genet. *7*, 1–12.

Barakat, T.S., Loos, F., Van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J.A., and Gribnau, J. (2014). The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. Mol. Cell *53*, 965–978.

Barr, M.L., and Bertram, E.G. (1949). A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature *163*, 676.

Beard, C., Li, E., and Jaenisch, R. (1995). Loss of methylation activates Xist in somatic but not in embryonic cells. Genes Dev. *9*, 2325–2334.

Berletch, J.B., Yang, F., and Disteche, C.M. (2010). Escape from X inactivation in mice and humans. Genome Biol. *11*, 213.

Berletch, J.B., Ma, W., Yang, F., Shendure, J., Noble, W.S., Disteche, C.M., Deng, X. (2015). Escape from X inactivation varies in mouse tissues. PLoS Genet. *11*, 1-26.

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature *351*, 329–331.

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

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

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.

Burgoyne, P.S. (1993). A Y-chromosomal effet on blastocyst cell number in mice. Development *117*, 341–345.

Burgoyne, P.S., Tam, P.P., and Evans, E.P. (1983). Retarded development of XO conceptuses during early pregnancy in the mouse. J. Reprod. Fertil. *68*, 387–393.

Buzin, C.H., Mann, J.R., and Singer-Sam, J. (1994). Quantitative RT-PCR assays show Xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. Development *120*, 3529–3536.

Carrel, L and Willard, H. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400-403.

Chaumeil, J., Baccon, P. Le, Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear. Genes Dev. *20*, 2223–2237.

Chow, J.C., and Brown, C.J. (2003). Forming facultative heterochromatin: Silencing of an X chromosome in mammalian females. Cell. Mol. Life Sci. *60*, 2586–2603.

Chow, J.C., Hall, L.L., Baldry, S.E.L., Thorogood, N.P., Lawrence, J.B., and Brown, C.J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. Proc. Natl. Acad. Sci. U. S. A. *104*, 10104–10109.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., et al. (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956–969.

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

Chureau, C., Prissette, M., Bourdet, A., Cattolico, L., Jones, L., Avner, P., and Duret, L. (2002). Comparative sequence analysis of the X-Inactivation center region in mouse , human and bovine. Genome Res. *12*, 894–908.

Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., and Rougeulle, C. (2011). Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. Hum. Mol. Genet. *20*, 705–718.

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

Cooper DW, VandeBerg JL, Sharman GB, P.W. (1971). Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. Nat New Bio *230*, 155–157.

Cotton, A.M., Chen, C.Y., Lam, L.L., Wasserman, W.W., Kobor, M.S., and Brown, C.J. (2014). Spread of X-chromosome inactivation into autosomal sequences: Role for DNA elements, chromatin features and chromosomal domains. Hum. Mol. Genet. *23*, 1211–1223.

Cotton, A.M., Price, E.M., Jones, M.J., Balaton, B.P., Kobor, M.S., and Brown, C.J. (2015). Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum. Mol. Genet. *24*, 1528–1539.

Crews, D. (2003). Sex determination: Where environment and genetics meet. Evol. Dev. *5*, 50–55.

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

Deng, X., Hiatt, J.B., Nguyen, D.K., Ercan, S., Sturgill, D., Hillier, L.W., Schlesinger, F., Davis, C.A., Reinke, V.J., Gingeras, T.R., et al. (2011). Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. Nat. Genet. *43*, 1179–1185.

Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature *460*, 128–132.

Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M., and Brockdorff, N. (1999). Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. Hum. Mol. Genet. *8*, 195–204.

Dvash, T., Lavon, N., and Fan, G. (2010). Variations of X chromosome inactivation occur in early passages of female human embryonic stem cells. PLoS One *5*, 1–9.

Engelstädter, J., and Haig, D. (2008). Sexual antagonism and the evolution of X chromosome inactivation. Evolution (N. Y). *62*, 2097–2104.

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

Epstein, C., Smith, S., Travis, T., and Ucker, G. (1978). Both X chromosomes function before visible X chromosome inactivation in female mouse embryos. Nature *274*, 500–503.

Ercan, S., Giresi, P.G., Whittle, C.M., Zhang, X., Green, R.D., and Lieb, J.D. (2007). X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation. Nat. Genet. *39*, 403–408.

Erik Splinter, Elzo de Wit, Elphège P. Nora, et al. The inactive X chromosome adopts a unique three-dimensional (Genes Dev.)-2011-Splinter-1371-83.pdf.

Fukuda, A., Tomikawa, J., Miura, T., Hata, K., Nakabayashi, K., Eggan, K., Akutsu, H., and Umezawa, A. (2014). The role of maternal-specific H3K9me3 modification in establishing imprinted X-chromosome inactivation and embryogenesis in mice. Nat. Commun. *5*, 5464.

Fukuda, A., Mitani, A., Miyashita, T., Umezawa, A., and Akutsu, H. (2015). Chromatin condensation of Xist genomic loci during oogenesis in mice. Development *142*, 4049–4055.

Furlan, G., and Rougeulle, C. (2016). Function and evolution of the long noncoding RNA circuitry orchestrating X-chromosome inactivation in mammals. Wiley Interdiscip. Rev. RNA *7*, 705–722.

G.B.Sharman (1971). Late DNA replication in the paternally derived X chromosome of female kangaroos. Nat. Phys. Sci. *231*, 231–232.

Gartler, S.M., and Riggs, A.D. (1983). Mammalian X-chromosome inactivation. Ann. Rev. Genet *17*, 155–190.

Giuseppe Borsani, Rossana Tonlorenzi, M. Christine Simmler, Luisa Dandolo, Danielle Arnaud, Valeria Capra, Markus Grompe, Antonio Pizzuti, Donna Muzny, Charles Lawrence, H.F.W., Avner, P., and Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *354*, 56–58.

Gontan, C., Achame, E.M., Demmers, J., Barakat, T.S., Rentmeester, E., van IJcken, W., Grootegoed, J.A., and Gribnau, J. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature *485*, 386–390.

Graves, J.A.M. (2006). Sex chromosome specialization and degeneration in mammals. Cell *124*, 901–914.

Grumbach, M.M., Morishima, A., and Taylor, J.H. (1963). Human Sex Chromosome Abnormalities in Relation To Dna Replication and Heterochromatinization. Proc. Natl. Acad. Sci. U. S. A. *49*, 581–589.

Haig, D. (2006). Self-Imposed Silence : Parental Antagonism and the Evolution of X-Chromosome Inactivation. Evolution (N. Y). *60*, 440–447.

Hall, D.W., and Wayne, M.L. (2012). Ohno's "peril of hemizygosity" revisited: Gene loss, dosage compensation, and mutation. Genome Biol. Evol. *5*, 1–15.

Hall, L.L., Byron, M., Sakai, K., Carrel, L., Willard, H.F., and Lawrence, J.B. (2002a). An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc. Natl. Acad. Sci. U. S. A. *99*, 8677–8682.

Hall, L.L., Clemson, C.M., Byron, M., Wydner, K., and Lawrence, J.B. (2002b). Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. Hum. Mol. Genet. *11*, 3157–3165.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of xist RNA. Dev. Cell *19*, 469–476.

He, X., Chen, X., Xiong, Y., Chen, Z., Wang, X., Shi, S., Wang, X., and Zhang, J. (2011). He et al. reply. Nat. Genet. *43*, 1171–1172.

Heard, E., Mongelard, F., Arnaud, D., Chureau, C., Vourc 'h, C., and Avner, P. (1999). Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. Proc Natl Acad Sci U S A *96*, 6841–6846.

Hendrich, B.D., Browp, C.J., and Willard, H.F. (1993). Evolutionary conservation of possible functional domains of the human and murine XIST genes. Hum. Mol. Genet. *2*, 663–672.

Herzing, L.B., Romer, J.T., Horn, J.M., and Ashworth, A. (1997). Xist has properties of the X-chromosome inactivation centre. Nature *386*, 272–275.

Hoffman, L.M., Hall, L., Batten, J.L., Young, H., Pardasani, D., Baetge, E.E., Lawrence, J., and Carpenter, M.K. (2005). X-inactivation status varies in human embryonic stem cell lines. Stem Cells *23*, 1468–1478.

Hoki, Y., Kimura, N., Kanbayashi, M., Amakawa, Y., Ohhata, T., Sasaki, H.and Sado, T. (2009).A proximal conserved repeat in the *Xist* gene is essential as a genomic element for X-inactivation in mouse. Development *136*, 139-146.

Hosler, B. a, LaRosa, G.J., Grippo, J.F., and Gudas, L.J. (1989). Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. Mol. Cell. Biol. *9*, 5623–5629.

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

I Okamoto, A P Otte, C D Allis, D Reinberg, E.H. (2004). Epigenetic Dynamics of Imprinted X Inactication During Early Mouse Development. Science*303*, 644–650.

Jacey L. Hornecker, Paul B. Samollow, Edward S. Robinson, J.L.V. and J.R.M. (2007). Decoding development in Xenopus tropicalis. Genesis 696–708.

Jamieson, R. V, Tan, S.S., and Tam, P.P. (1998). Retarded postimplantation development of X0 mouse embryos: impact of the parental origin of the monosomic X chromosome. Dev. Biol. *201*, 13–25.

Johnston, C.M., Newall, A.E., Brockdorff, N., and Nesterova, T.B. (2002). Enox, a novel gene that maps 10 kb upstream of Xist and partially escapes X inactivation. Genomics *80*, 236–244.

Jonkers, I., Barakat, T.S., Achame, E.M., Monkhorst, K., Kenter, A., Rentmeester, E., Grosveld, F., Grootegoed, J.A., and Gribnau, J. (2009). RNF12 is an X-encoded dose-dependent activator of X chromosome inactivation. Cell *139*, 999–1011.

Julien, P., Brawand, D., Soumillon, M., Necsulea, A., Liechti, A., Schütz, F., Daish, T., Grützner, F., and Kaessmann, H. (2012). Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. PLoS Biol. *10*, 1–20.

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

Kay, G.F., Barton, S.C., Surani, M.A., and Rastan, S. (1994). Imprinting and X chromosome counting mechanisms determine Xist expression in early mouse development. Cell *77*, 639–650.

Kelly, W.G., and Aramayo, R. (2007). Meiotic silencing and the epigenetics of sex. Chromosom. Res. *15*, 633–651.

Keohane, A.M., Barlow, A.L., Waters, J., Bourn, D., and Turner, B.M. (1999). H4 acetylation, XIST RNA and replication timing are coincident and define X;autosome boundaries in two abnormal X chromosomes. Hum. Mol. Genet. *8*, 377–383.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049–1061.

J.C.Lucchesi. and M.I.Kuroda (2015). Dosage compensation in Drosophila. Cold Spring Harb. Perspect. Biol. *5*, 1–20.

Laurent Duret, Corinne Chureau, Sylvie Samain, Jean Weissenbach, P.A. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science *312*, 1653–1655.

Lee, J.T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nat. Rev. Mol. Cell Biol. *12*, 815–826.

Lee, J.T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275–279.

Lee, J.T., and Lu, N. (1999). Targeted Mutagenesis of Tsix Leads to Nonrandom X Inactivation. Cell *99*, 47–57.

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

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

Li, Y., Tan, T., Zong, L., He, D., Tao, W., and Liang, Q. (2012). Study of methylation of histone H3 lysine 9 and H3 lysine 27 during X chromosome inactivation in three types of cells. Chromosom. Res. *20*, 769–778.

Lin, F.Q., Xing, K., Zhang, J.Z., and He, X.L. (2012). Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. Proc. Natl. Acad. Sci. U. S. A. *109*, 11752–11757.

Lyon, M.F. (1954). A Further Mutation Of The Mottled Type In the House Mouse. J. Hered. 116–121.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). Nature *4773*, 372–373.

Lyon, M.F. (2000). LINE-1 elements and X chromosome inactivation: A function for "junk" DNA? Proc. Natl. Acad. Sci. *97*, 6248–6249.

Maclary, E., Hinten, M., Harris, C., and Kalantry, S. (2013). Long nonoding RNAs in the X-inactivation center. Chromosom. Res. *21*, 601–614.

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianférani, S., Van Dorsselaer, A., Clerc, P., Avner, P., et al. (2010). 2-D structure of the a region of Xist RNA and its implication for PRC2 association. PLoS Biol. *8*, 1–16.

Makhlouf, M., and Rougeulle, C. (2011). Linking X chromosome inactivation to pluripotency: Necessity or fate? Trends Mol. Med. *17*, 329–336.

Marks, H., Chow, J.C., Denissov, S., Françoijs, K.-J., Brockdorff, N., Heard, E., and Stunnenberg, H.G. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res. *19*, 1361–1373.

Marks, H., Kerstens, H.H.D., Barakat, T.S., Splinter, E., Dirks, R.A.M., van Mierlo, G., Joshi, O., Wang, S.-Y., Babak, T., Albers, C.A., et al. (2015). Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. Genome Biol. *16*, 1–20.

Marshall Graves, J. a (2008). Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. Annu. Rev. Genet. *42*, 565–586.

McDonel, P., Jans, J., Peterson, B.K., and Meyer, B.J. (2006). Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. Nature *444*, 614–618.

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Mermoud, J.E., Costanzi, C., Pehrson, J.R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J. Cell Biol. *147*, 1399–1408.

Meyer, B.J., and Casson, L.P. (1986). *Caenorhabditis elegans* compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. Cell *47*, 871–881.

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

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

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

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., et al. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science (80-.). *349*, 1–19.

Minkovsky, A., Barakat, T.S., Sellami, N., Chin, M.H., Gunhanlar, N., Gribnau, J., and Plath, K. (2013). The pluripotency factor-bound intron 1 of xist is dispensable for X chromosome inactivation and reactivation In Vitro and In Vivo. Cell Rep. *3*, 905–918.

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A pooled shrna screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. Cell Rep. *12*, 562–572.

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. Cell Rep. *12*, 554–561.

Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F., and Gribnau, J. (2008). X Inactivation Counting and Choice Is a Stochastic Process: Evidence for Involvement of an X-Linked Activator. Cell *132*, 410–421.

Monkhorst, K., de Hoon, B., Jonkers, I., Achame, E.M., Monkhorst, W., Hoogerbrugge, J., Rentmeester, E., Westerhoff, H. V., Grosveld, F., Grootegoed, J.A., et al. (2009). The probability to initiate X chromosome inactivation is determined by the X to autosomal ratio and X chromosome specific allelic properties. PLoS One *4*, 1–14.

Morey, C., Arnaud, D., Avner, P., and Clerc, P. (2001). Tsix-mediated repression of Xist accumulation is not sufficient for normal random X inactivation. Hum. Mol. Genet. *10*, 1403–1411.

Morishima, A., Grumbach, M.M., and Taylor, J.H. (1962). Asynchronous duplication of human chromosomes and the origin of sex chromatin. Proc. Natl. Acad. Sci. U. S. A. *48*, 756–763.

Mukherjee, A.S., and Beermann, W. (1965). Synthesis of ribonucleic acid by the X-chromosomes of Drosophila melanogaster and the problem of dosage compensation. Nature *207*, 785–786.

Muller, H.J. (1939). A Gene For The Fourth Chromosome Of Drosophila. 325–336.

Namekawa, S.H., Park, P.J., Zhang, L.F., Shima, J.E., McCarrey, J.R., Griswold, M.D., and Lee, J.T. (2006). Postmeiotic sex chromatin in the male germline of mice. Curr. Biol. *16*, 660–667.

Namekawa, S.H., Vandeberg, J.L., Mccarrey, J.R., and Lee, J.T. (2007). Sex chromosome silencing in the marsupial male germ line. Proc Natl Acad Sci U S A *104*, 9730–9735.

Navarro, P., Pichard, S., Ciaudo, C., Avner, P., and Rougeulle, C. (2005). Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: Implications for X-chromosome inactivation. Genes Dev. *19*, 1474–1484.

Navarro, P., Page, D.R., Avner, P., and Rougeulle, C. (2006). Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. Genes Dev. *20*, 2787–2792.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693–1695.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457–460.

Navarro, P., Moffat, M., Mullin, N.P., and Chambers, I. (2011). The X-inactivation trans-activator Rnf12 is negatively regulated by pluripotency factors in embryonic stem cells. Hum. Genet. *130*, 255–264.

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., and Zakian, S.M. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res. *11*, 833–849.

Nesterova, T.B., Johnston, C.M., Appanah, R., Newall, A.E.T., Godwin, J., Alexiou, M., and Brockdorff, N. (2003). Skewing X chromosome choice by modulating sense transcription across the Xist locus. Genes Dev. *17*, 2177–2190.

Nesterova, T.B., Senner, C.E., Schneider, J., Alcayna-Stevens, T., Tattermusch, A., Hemberger, M., and Brockdorff, N. (2011). Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. Epigenetics Chromatin *4*, 17.

Nguyen, D.K., and Disteche, C.M. (2005). Dosage compensation of the X chromosome in mammals. Nat Genet. *38*, 47–53.

Nora, E.P., and Heard, E. (2010). Chromatin structure and nuclear organization dynamics during X-chromosome inactivation. Cold Spring Harb. Symp. Quant. Biol. *75*, 333–344.

Ogawa, Y., and Lee, J.T. (2003). Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol. Cell *11*, 731–743.

Ohhata, T., Hoki, Y., Sasaki, H., and Sado, T. (2008). Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. Development *135*, 227–235.

Ohno, S., and Hauschka, T.S. (1960). Allocycly of the X-chromosome in tumors and normal tissues. Cancer Res. *20*, 541–545.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

Patrat, C., Okamoto, I., Diabangouaya, P., Vialon, V., Le Baccon, P., Chow, J., and Heard, E. (2009). Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. Proc. Natl. Acad. Sci. U. S. A. *106*, 5198–5203.

Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annu. Rev. Genet. *42*, 733–772.

Peeters, S.B., Cotton, A.M., Brown, C.J. (2014). Variable escape from X chromosome inactivation : Identifying factors that tip the scales towards expression. Bioessays *36*, 746-756.

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

Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A., and Marais, G.A.B. (2012). Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. Proc. Natl. Acad. Sci. U. S. A. *109*, 5346–5351.

Pinter, S.F. (2016). A Tale of Two Cities: How Xist and its partners localize to and silence the bicompartmental X. Semin. Cell Dev. Biol.*56*, 19-34.

Pinter, S.F., Sadreyev, R.I., Yildirim, E., Jeon, Y., Ohsumi, T.K., Borowsky, M., and Lee, J.T. (2012). Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. Genome Res. *22*, 1864–1876.

Plath, K. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation. Science *300*, 131–135.

Popova, B.C., Tada, T., Takagi, N., Brockdorff, N., and Nesterova, T.B. (2006). Attenuated spread of X-inactivation in an X;autosome translocation. Proc. Natl. Acad. Sci. U. S. A. *103*, 7706–7711.

Potrzebowski, L., Vinckenbosch, N., Marques, A.C., Chalmel, F., Jégou, B., and Kaessmann, H. (2008). Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biol. *6*, 709–716.

Rastan, S. (1983). Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre. J Embryol Exp Morphol *78*, 1–22.

Ross, M.T., Grafham, D. V, Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., et al. (2005). The DNA sequence of the human X chromosome. Nature *434*, 325–337.

S. Ohno, W. D. Kaplan, and R.K. (1959). Formationof the sex chromatin by a single x-chromosome in LIVER *Rattus norvegicus*. Exp. Cell Res. *18*, 415–418.

Sado, T., Okano, M., Li, E., and Sasaki, H. (2004). De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. Development *131*, 975–982.

Sado, T., Hoki, Y., and Sasaki, H. (2005). Tsix silences Xist through modification of chromatin structure. Dev. Cell *9*, 159–165.

Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T., Offer, J., et al. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. Mol. Cell *57*, 769–783.

Sarkar, K.M., Gayen, S., Kumar, S., Maclary, E., Buttigieg, E., Hinten, M., Kumari, A., Harris, C., Sado, T.and Kalantry, S. (2015). An Xist-activating antisense RNA required for X-chromosome inactivation. Nat. Comm. *6*, 1-13.

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

Schulz, E.G., and Heard, E. (2013). Role and control of X chromosome dosage in mammalian development. Curr. Opin. Genet. Dev. *23*, 109–115.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell *14*, 203–216.

Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., and Antonarakis, S.E. (2011). DNA methylation profiles of human active and inactive X chromosomes. Genome Res. *21*, 1592–1600.

Shibata, S., and Lee, J.T. (2004). Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice. Curr. Biol. *14*, 1747–1754.

Shin, J., Bossenz, M., Chung, Y., Ma, H., Byron, M., Taniguchi-Ishigaki, N., Zhu, X., Jiao, B., Hall, L.L., Green, M.R., et al. (2010). Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. Nature *467*, 977–981.

Silva, S.S.S., Rowntree, R.K.K., Mekhoubad, S., and Lee, J.T.T. (2008). X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. Proc Natl Acad Sci U S A *105*, 4820–4825.

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

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, a M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature *346*, 240–244.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.

Soma, M., Fujihara, Y., Okabe, M., Ishino, F., and Kobayashi, S. (2014). Ftx is dispensable for imprinted X-chromosome inactivation in preimplantation mouse embryos. Sci. Rep. *4*, 5181.

Splinter, E., de Wit, E., Nora, E.P., Klous, P., van de Werken, H.J.G., Zhu, Y., Kaaij, L.J.T., van Ijcken, W., Gribnau, J., Heard, E., et al. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. *25*, 1371–1383.

Strome, S., Kelly, W.G., Ercan, S., and Lieb, J.D. (2014). Regulation of the X chromosomes in *Caenorhabditis elegans*. Cold Spring Harb. Perspect. Biol. *6*, 1–21.

Sun, B.K., Deaton, A.M., and Lee, J.T. (2006). A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol. Cell *21*, 617–628.

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

Sun, S., Payer, B., Namekawa, S., An, J.Y., Press, W., Catalan-Dibene, J., Sunwoo, H., and Lee, J.T. (2015). Xist imprinting is promoted by the hemizygous (unpaired) state in the male germ line. Proc. Natl. Acad. Sci. U. S. A. *112*, 14415–14422.

Sunwoo, H., Wu, J.Y., and Lee, J.T. (2015). The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. Proc. Natl. Acad. Sci. U. S. A. *112*, E4216-25.

Sutton, E., Hughes, J., White, S., Sekido, R., Tan, J., Arboleda, V., Rogers, N., Knower, K., Rowley, L., Eyre, H., et al. (2011). Identification of SOX3 as an XX male sex reversal gene in mice and humans. J. Clin. Invest. *121*, 328–341.

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature *256*, 640–642.

Takagi, N., Sugawara, O., and Sasaki, M. (1982). Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. Chromosoma *85*, 275–286.

Tang, Y.A., Huntley, D., Montana, G., Cerase, A., Nesterova, T.B., and Brockdorff, N. (2010). Efficiency of Xist-mediated silencing on autosomes is linked to chromosomal domain organisation. Epigenetics Chromatin *3*, 1–11.

Thornhill, A.R., and Burgoyne, P.S. (1993). A paternally imprinted X chromosome retards the development of the early mouse embryo. Development *118*, 171–174.

Tian, D., Sun, S., and Lee, J.T. (2011). The long noncoding RNA, Jpx, is a molecular switch for X-chromosome inactivation. Cell *143*, 390–403.

Tomoda, K., Takahashi, K., Leung, K., Okada, A., Narita, M., Yamada, N.A., Eilertson, K.E., Tsang, P., Baba, S., White, M.P., et al. (2012). Derivation conditions impact X-inactivation status in female human induced pluripotent stem cells. Cell Stem Cell *11*, 91–99.

Tsunoda, Y., Tokunaga, T., and Sugie, T. (1985). Altered sex ratio of live young after transfer of fast- and slow-developing mouse embryos. Gamete Res. *12*, 301–304.

Turner, J.M.A. (2007). Meiotic sex chromosome inactivation. Curr. Biol. *134*, 1823–1831.

Turner, J.M., Mahadevaiah, S.K., Fernandez-Capetillo, O., Nussenzweig, A., Xu, X., Deng, C.X., and Burgoyne, P.S. (2005). Silencing of unsynapsed meiotic chromosomes in the mouse. Nat Genet *37*, 41–47.

Veyrunes, F., Waters, P.D., Miethke, P., Murchison, E.P., Kheradpour, P., Sachidanandam, R., Park, J., Semyonov, J., Chang, C.L., Whittington, C.M., et al. (2008). Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes of platypus and reptile venom genes Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Res. *18*, 965–973.

Wakefield, M.J., Keohane, A.M., Turner, B.M., and Graves, J. a (1997). Histone underacetylation is an ancient component of mammalian X chromosome inactivation. Proc. Natl. Acad. Sci. U. S. A. *94*, 9665–9668.

Wallis, M.C., Waters, P.D., Delbridge, M.L., Kirby, P.J., Pask, A.J., Gr??tzner, F., Rens, W., Ferguson-Smith, M.A., and Graves, J.A.M. (2007). Sex determination in platypus and echidna: Autosomal location of SOX3 confirms the absence of SRY from monotremes. Chromosom. Res. *15*, 949–959.

Waters, P.D., Wallis, M.C., and Graves, J.A.M. (2007). Mammalian sex-Origin and evolution of the Y chromosome and SRY. Semin. Cell Dev. Biol. *18*, 389–400.

Williams, L.H., Kalantry, S., Starmer, J., and Magnuson, T. (2011). Transcription precedes loss of Xist coating and depletion of H3K27me3 during X-chromosome reprogramming in the mouse inner cell mass. Development *138*, 2049–2057.

Winifred Mak, Tatyana B. Nesterova, Mariana de Napoles, Ruth Appanah, Shinya Yamanaka, Arie P. Otte, N.B. (2004). Reactivation of the Paternal X Chromosome in Early Mouse Embryos. Science (80-.). *303*, 666–669.

Wutz, A. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat. Rev. Genet. *12*, 542–553.

Wutz, A., Rasmussen, P.T. and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet. *30*, 167-174

Wutz, A., and Jaenisch, R. (2000). A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation. Mol. Cell *5*, 695–705.

Xiong, Y., Chen, X., Chen, Z., Wang, X., Shi, S., Wang, X., Zhang, J., and He, X. (2010). RNA sequencing shows no dosage compensation of the active X-chromosome. Nat. Genet. *42*, 1043–1047.

Xu, N., Donohoe, M.E., Silva, S.S., and Lee, J.T. (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. Nat Genet *39*, 1390–1396.

Yang, F., Babak, T., Shendure, J., and Disteche, C.M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. *20*, 614–622.

Yildirim, E., Sadreyev, R.I., Pinter, S.F., and Lee, J.T. (2011). X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. Nat. Struct. Mol. Biol. *19*, 56–61.

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

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

Chapter 2: Regulation of XIST in the context of initiation phase of XCI

2.1 Introduction

Over the past two decades, a large body of evidence has significantly improved our understanding of the regulation of *Xist* as well as XCI in mouse system. Although *Xist*/*XIST* were discovered from mouse and human around the same time (Brown et al., 1991; Borsani et al., 1991), understanding with regards to the regulation of *XIST* and the process of XCI in human has gained attention rather recently owing to the limitations of the system discussed in the section 1.6 of Chapter 1. Nevertheless, these limited studies have been instrumental in drawing the field's attention to the fact that the process of XCI, despite being fairly similar, can be manifested and regulated differently in diverse species.

The first study attempted at unravelling the regulation of human *XIST* gene identified and characterized the promoter of *XIST* (Hendrich et al., 1997). The authors in the study also compared sequence upstream of *XIST* TSS from mouse, human, horse and rabbit and discovered that significant homology exist only within the first 100bp upstream of TSS. Furthermore, they also identified the binding sites for three transcription factors – SP1, YY1 and TBP by carrying out saturated site-directed mutagenesis and validated the binding *in vitro* by performing electrophoretic mobility shift assay (EMSA). The question whether these factors can actually bind the promoter in the cells and regulate *XIST* transcription remained unanswered until recently. Two independent studies uncovered the role of YY1 as the key transcription factor regulating *XIST* expression by binding to its promoters (Chapman et al., 2014; Makhlouf et al., 2014). In silico analysis of *XIST* promoter region for seven eutherian species in the study by Makhlouf et al., 2014 led to the identification of conserved YY1 binding sites arranged as clusters close to *XIST* TSS as shown in the Figure 2.1. The authors also identified binding sites for chromatin organizer proteins such as CTCF and RAD21 in close vicinity to YY1 clusters.

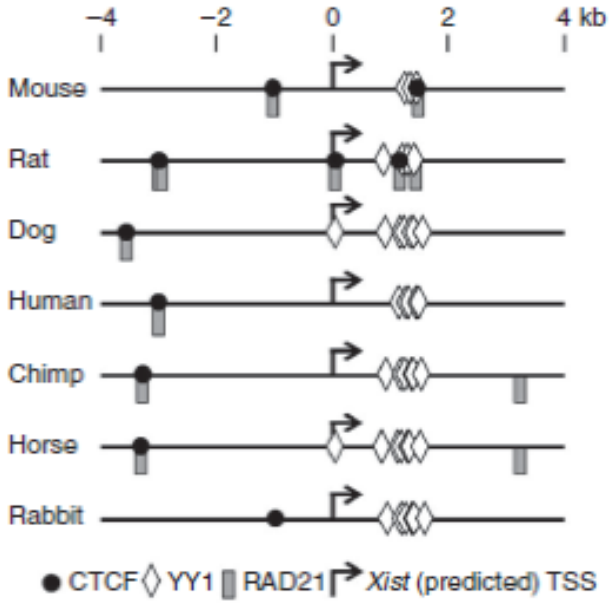


Figure 2.1 Comparison of XIST genomic sequence across seven eutherian species.

The schematic map of Xist/XIST 5' region depicting the binding sites for CTCF (black closed circle), YY1 (open diamond) and RAD21 (grey rectangle) in seven eutherian species. The image shown here is adapted from Makhlouf et al., 2014.

In this report, it was convincingly demonstrated that YY1 occupies the promoter and regulates *Xist*/*XIST* specifically in the female lines of mESCs, XIST expressing hESC, mouse embryonic fibroblast (MEF) and human fibroblast. They argued against the roles of CTCF and RAD21 in regulating asymmetric expression of Xist/XIST since they were found to be equally enriched on *XIST* genomic locus in both male and female lines of mESCs, MEF and human fibroblast as well as *XIST+* and *XIST*- hESCs. Furthermore, it was also demonstrated that in wildtype female mESCs, Yy1 and Rex1 compete for binding on *Xist* promoter in undifferentiated state and as the levels of Rex1 decrease due to Rnf12 mediated degradation during the initiation of XCI, Yy1 occupancy on the unmethylated *Xist* promoter increases which then results in upregulation of Xist (Makhlouf et al., 2014). Whether such a mechanism exists in human cells remains to be tested.

Second study identified an alternative promoter for human *XIST* (P2) within a CpG-island that is differentially methylated between male and female lines and located 1.4 Kb downstream of the first promoter (P1) by performing DNase hypersensitivity mapping using male and female lymphoblast cells (Chapman et al., 2014). This study also discovered YY1 to be an essential factor regulating transcription from the unmethylated P2 promoter. Interestingly, Yy1 has also been demonstrated to interact with Xist lncRNA in mESCs which is

thought to be crucial for tethering Xist RNA to the X chromosome from which it is synthesized (future Xi) to prevent its diffusion (Jeon and Lee, 2011).

The major conclusion from these studies is that YY1 is an essential factor governing *XIST* transcription from both P1 and P2 promoters in hESCs as well as human fibroblast lines. A vast number of studies using mouse system have identified and deciphered the roles played by a multitude of factors in modulating the expression of *Xist* in temporal manner (discussed in section 1.5 of Chapter 1). Our understanding with respect to temporal regulation of human *XIST* is far from complete. Therefore, it is important to address this aspect of human *XIST* regulation during the initiation phase of XCI. The present study attempts to probe the roles of already identified transcription factors – SP1 and YY1 along with pluripotency factors – OCT4, SOX2, NANOG and chromatin organizer proteins – CTCF, SATB1 and SATB2. Below is a brief description about transcription regulators.

2.1.1 SP1, YY1 and CTCF

SP1 is an acronym for Specificity Protein 1. It was identified as early as 1980s (Kadonaga et al., 1987) and subsequently shown to recognize and specifically bind to GC-rich sites via three Cys_2his_2 zinc-finger domains (Dynan and Tjian, 1983). It also harbours two transactivation domains known as Domain A and Domain B (Courey and Tjian, 1988; Pascal and Tjian, 1991), by virtue of which it can interact with general transcription factors such as TATA-binding protein TBP or TBP associated factor TAF (Chiang and Roeder, 1995; Emili et al., 1994; Hoey et al., 1993; Tanese et al., 1996). It is ubiquitously expressed in a variety of tissues (Saffer et al., 1991) and is present in the species ranging from *Caenorhabditis elegans* to humans (Brown et al., 1996; Ernst A. Wimmer, Herbert Jackie, 1993; Kadonaga et al., 1987). SP1 is an important transcription factor and regulates genes involved in a variety of biological functions including embryonic development, cell growth and proliferation (reviewed in Suske, 1999).

Yin-Yang1 or YY1 was identified and characterized simultaneously independently by two research groups (Park and Atchison, 1991; Shi et al., 1991). It harbors four Cys_2his_2 zinc-finger domains at its C-terminal which confers upon it sequence specific DNA binding ability (Hyde-deruyscher et al., 1995; Yant et al., 1995) as well as the repression function (Austen et al., 1997; Galvin and Shi, 1997; Hariharan et al., 1991; Lewis et al., 1995; Montalvo et al.,

1995). The activation domain is present at the N-terminus of YY1 which consists of 11 consecutive acidic amino acids in a stretch. YY1 can act both as transcriptional activator or repressor depending on its interaction with other proteins and hence the name Yin-Yang (Chang et al., 1989; Gordon et al., 2006; Shi et al., 1991, 1997). It is ubiquitously expressed in multiple tissues as well as a variety of cell lines (reviewed in Shi et al., 1997) and evolutionary conserved between Xenopus, mouse and humans (Pisaneschi et al., 1994). YY1 is involved in transcriptionally regulating genes involved in a variety of processes including cell growth, proliferation and differentiation (Donohoe et al., 1999; Palko et al., 2004; Petkova et al., 2001).

CCCTCC binding factor – CTCF is an evolutionary conserved protein in higher eukaryotes. The maximum homology is observed in the eleven zinc-finger central DNA binding domain (reviewed in Ohlsson et al., 2001). It is ubiquitously expressed in a manner similar to the housekeeping gene in most of the metazoan tissues. One peculiar property of CTCF is that it can bind to a wide variety of DNA sequences as well as interact with numerous coregulatory proteins by combinatorial use of the eleven zinc-finger domains (Filippova et al., 1996). Since its discovery (Klenova et al., 1993), it has been actively worked upon which has led to the appreciation of its role as a context dependent transcriptional activator or repressor or insulator acting globally (reviewed in Phillips and Corces, 2009). Accumulating evidences have deciphered its role in governing higher order chromatin architecture and hence CTCF is now regarded as a global chromatin organizer (reviewed in Phillips and Corces, 2009). It has been demonstrated to be indispensable for various biological processes including embryogenesis, genomic imprinting (Fedoriw et al., 2004; Kurukuti et al., 2006; Pant et al., 2003; Szabo et al., 2004; Szabó et al., 2000), X chromosome inactivation in mouse (Bacher et al., 2006; Chao et al., 2002; Filippova et al., 2005; Navarro et al., 2006; Spencer et al., 2011; Xu et al., 2007).

As mentioned earlier, YY1 has already been shown to be the activator of human XIST transcription (Chapman et al., 2014; Makhlouf et al., 2014). SP1 has been demonstrated to bind XIST promoter by EMSA (Hendrich et al., 1997) and CTCF too has binding sites on the XIST genomic locus (Makhlouf et al., 2014). However, the significance of these observations apart from the role of YY1 remains to be tested, which we attempt to address in our study.

2.1.2 OCT4, SOX2 and NANOG

OCT4, SOX2 and NANOG are the pluripotency factors expressed in embryonic stem cells. They form a core regulatory network to maintain the self-renewal pluripotent state (Boiani and Schöler, 2005; Boyer et al., 2005; Loh et al., 2006). OCT4 is a POU-domain containing conserved transcription factor, whose precise levels are essential for maintaining the stemness of ES cells (Nichols et al., 1998). For example, it has been demonstrated that reduction in OCT4 levels leads to differentiation of ES cells to trophoectoderm lineage, whereas induction of OCT4 expression causes ES cells to differentiate into primitive endoderm and mesoderm (Niwa, 2001; Yeom et al., 1996). SOX2 and NANOG belong to HMG-domain containing family of proteins. SOX2 co-occupies many of the regulatory regions of multiple genes with OCT4 and aid in maintaining the self-renewal state of stem cells (Ambrosetti et al., 2000; Avilion et al., 2003; Masui et al., 2007). Basically, the function of these three proteins is to restrict the expression of lineage-specific genes and promoter the expression of pluripotency related genes (Boyer et al., 2005; Chambers and Smith, 2004; Chambers et al., 2007; Loh et al., 2006).

The contribution of the pluripotency factors in different facets of mouse XCI has been reported previously. Citing few studies, they have been demonstrated to be involved in controlling monoallelic upregulation of Xist (Donohoe et al., 2009; Navarro et al., 2008, Navarro et al., 2010). However, there is a controversy over these reports since two independent research groups argued against their roles in repressing Xist transcription (Minkovsky et al., 2013; Nesterova et al., 2011) (discussed elaborately in the 1st chapter). Since human *XIST* is presumably regulated differently from mouse *Xist*, the pluripotency factors may or may not contribute towards *XIST* transcriptional regulation. In the present study, we have attempted to understand their roles in an unbiased manner.

2.1.3 SATB1 and SATB2

SATB1 and SATB2 were identified to be matrix attachment region (MAR) binding proteins, characterized by their abilities to bind AT-rich sequences and hence the name Special AT-rich binding proteins (Dickinson, 1992; Szemes et al., 2006). SATB1 harbours DNA binding domains at their C-termini comprising two Cut-repeat domains (supports binding to DNA) and a homeodomain (provides sequence specificity) and a Ubiquitin like domain at their N-termini, which not only mediates homo- as well as hetero-dimerization between

SATB1 and SATB2 (Dickinson et al., 1997; Purbey et al., 2008; Wang et al., 2014), but also provides the protein interaction interface (Notani et al., 2010, 2011; Pavan Kumar et al., 2006; Purbey et al., 2009). SATB1 and SATB2 share 61% homology at the protein levels and their domain architectures are more or less conserved. Earliest studies using SATB1 null mice identified its role in the development of T-lymphocytes in thymus (Alvarez et al., 2000). Subsequent studies helped in gaining significant insights into their role as global chromatin organizer in T cells (Cai et al., 2006; Galande et al., 2007; Kumar et al., 2007; Yasui et al., 2002). Functional role for SATB2 is majorly implicated in the neuronal development and osteogenesis (Balamotis et al., 2012; Britanova et al., 2006; Dobreva et al., 2006; FitzPatrick et al., 2003; Leone et al., 2015; Schütz et al., 2015). Several studies have also discovered SATB proteins to be involved in tumorigenesis (Han et al., 2008; Mir et al., 2015).

In addition, SATB proteins have been demonstrated to regulate Nanog in mESCs and influence the functional properties of stem cells (Savarese et al., 2009). The two members of the SATB family of proteins are also implicated in the initiation context for Xist-mediated silencing (Agrelo et al., 2009). The latter report however has been a subject of debate since another study employing $Satb1^{-/-}Satb2^{-/-}$ mESCs demonstrated them to be dispensable for XCI (Nechanitzky et al., 2012). It is worth mentioning that the authors of this paper arrived at this conclusion based on the formation of Xist cloud and H3K27me3 accumulation and assaying for the expression of only two X-linked genes (*Pgk1* and *Hprt*), which in our opinion is an inefficient measure for X inactivation. Also, they provide no mechanistic insights to substantiate their claims. On the contrary, Agrelo et al., 2009 have scored for the efficiency of initiating silencing in male MEFs by scoring for their survival by using inducible Xist in the background of Satb1 or Satb2 over-expression. This is a better and robust method of determining Xist mediated silencing since it scores for a global effect. Moreover, it is already known that Satb1 or Satb2-null mice are not embryonic lethal. However, no double knockout mouse has been reported to survive yet. Indeed in the Nechanitzky et al., 2012 study the number of double knockout embryos obtained is extremely small. Despite the controversies over the roles of Satb proteins in the process of XCI in mouse, we have made an unbiased attempt to uncover their roles in regulating human *XIST*.

2.2 Rationale of the present study

As stated earlier, there have been only a few studies towards understanding the regulation of human *XIST* promoter. Our knowledge with respect to *XIST* promoter behaviour during the initiation phase of XCI is still lacking. It is known that the levels of XIST are upregulated during a certain time window of development. But whether this is the consequence of YY1 acting on *XIST* promoter during a specific period or a concerted outcome of other additional players involved still remains an open question. The present study aims to address this aspect of *XIST* regulation by making use of embryonic carcinoma (EC) cell line. Like embryonic stem cells, EC cells express the pluripotency factors – OCT4, SOX2 and NANOG. This cell line was derived from a patient of testicular cancer and has been used as a proxy model for understanding the features of human ES cells as it differentiates into neuronal lineage upon administering all-trans-retinoic acid (RA) (Andrews et al., 1980; Andrews et al., 1984; Lee and Andrews, 1986; Skotheim et al., 2005). Moreover, as per ATCC (CRL-1973), these cells exhibit 1.6% polyploidy and the chromosomes other than 1, 10, 11 and 13, including X chromosome are present in 2 to 3 copies per cell. Also, undifferentiated EC cells express low amounts of XIST as assessed by RNA-FISH (Chow et al., 2003). Therefore, this provides a good system to probe the dynamic pattern of *XIST* promoter activity and carefully tease the roles of several other factors in tightly regulating the transcription from *XIST* promoter in the context of initiation phase of XCI. We do realize that being a cell line originating from a cancer patient, EC cells have its own limitations compared to female ES cells in terms of drawing conclusions implicating the physiological scenarios. Nevertheless, there is no ideal system currently available in the field to address the question of regulation of *XIST* and XCI in humans. Therefore, we have carefully devised experiments and derived conclusions keeping these limitations in mind.

2.3 Summary of the work

In this chapter of the thesis, we present results pertaining to the regulation of human *XIST* promoter in the context of initiation phase of XCI using EC cells as the model system. After establishing that they undergo differentiation and upregulate XIST lncRNA upon treatment with RA, we biochemically assessed the activity of *XIST* promoter elements. It was however observed that while the expression of endogenous XIST increases, the activity of *XIST* promoter elements with binding sites for SP1, YY1 and CTCF decrease during the course

of differentiation and so does the expression of these transcription regulators. These contradictory observations led us to speculate that region(s) other than the promoter elements could be important determinants of *XIST* transcription. And indeed by performing ChIP we show that the pluripotency proteins OCT4, SOX2 and NANOG display a unique differentiation driven temporal pattern of binding to a distinct site on XIST exon1 which negatively correlates with the expression of XIST. Additionally we also show that forced expression of OCT4 or SOX2 leads to downregulation of XIST in a cell line where Xi is stably established and maintained. We further determined XIST expression upon depletion of OCT4/SOX2/NANOG in EC cells. However, contrary to our expectation, we did not observe any upregulation in the levels of XIST in EC cells. This suggested that yet unidentified factors might be involved in temporally modulating the levels of XIST lncRNA. Based on the literature survey and preliminary data from our lab, we decided to test if the nuclear matrix proteins – SATB1 and SATB2 could regulate *XIST*. To address this, we monitored the expression of OCT4, SOX2, NANOG as well as XIST upon perturbing the levels of SATB1 or SATB2 in EC cells. The results obtained seem to suggest that SATB proteins positively regulate *XIST* and act as repressors for pluripotency factors. Moreover, we also show that the chromatin organizer proteins – SATB1, SATB2 and CTCF bind to distinct sites on *XIST* exon1. Interestingly, they also show similar binding kinetics as exhibited by the pluripotency proteins. The data presented here provides evidence that the site(s) occupied by OCT4, SOX2, NANOG, SATB1 and SATB2 might be involved in repressing *XIST* and the site bound by CTCF, SATB1 and SATB2 might act as either an activator or repressor, possibly dependent upon its interaction with the upstream elements. Collectively, the results discussed in this chapter provide compelling evidence towards coordinated action by a multitude of factors leading to timely activation of transcription from *XIST* promoter.

2.4 Materials and Methods

2.4.1 Cloning of *XIST* promoter elements into pGL3 Basic vector

In order to characterize *XIST* promoter, genomic regions ranging from 101 bp to 4458 bp upstream of *XIST* Transcription start-site (TSS) including a part of 5' region of the gene were cloned into a promoter- less reporter vector – pGL3 Basic procured from Promega (Wisconsin, Madison, USA). XIST genomic regions were PCR amplified from the genomic DNA

extracted from EC cells using specific primer pairs (see Appendix 2). The amplicons were subjected to phenol-chloroform-isoamylalcohol (25:24:1) (Sigma-Aldrich, St. Louis, Missouri, USA) extraction and precipitated by adding 1/10th volume of 3M Sodium acetate (pH 5.6) and 2.5 volumes of absolute ethanol to the isolated aqueous layer and allowed to precipitate at -20°C for 2h. DNA was purified by spinning the contents down in a centrifuge at 4°C, 12000rpm for 10 min after incubation period and the pellet was washed with 70% ethanol, air-dried and dissolved in sterile H_2O. The vector (pGL3 Basic) and the purified insert were digested with specific set of restriction enzymes (supplied by NEB, Ipswich, Massachusetts, USA). The linearized DNA products were separated using agarose gel electrophoresis, the bands were excised and DNA was purified from the agarose gel pieces by Gel extraction kit provided by Qiagen (Hilden, Germany). The purified vector and insert were ligated subsequently using T4 DNA ligase (supplied by NEB, Ipswich, Massachusetts, USA) at 16° C for 16 hours. The ligated mixtures were transformed into competent *Escherichia coli*DH5αcells and the colonies were screened for presence of the insert of interest by performing DNA extraction by alkaline lysis method followed by restriction enzyme digestion. The positive clones were confirmed by sequencing before using for further experiments.

2.4.2 Cloning of *XIST* exon1 regions into pGL3 Basic, XIST1.1 and XIST1

In order to test the significance of *XIST* exon1 region (~ 5.5 Kb downstream of XIST TSS) exhibiting binding to pluripotency factors and SATB proteins and the region, ~ 2.5 Kb downstream of XIST TSS with binding sites for CTCF, SATB1 and SATB2, both these sites were individually cloned into pGL3 Basic as well as into XIST1.1 and XIST1 constructs, between the promoter elements and luciferase gene as depicted in the schematic in the results section (Figure 2.5.6(A) and 2.5.11(A)). The regions were cloned in both the sense as well as anti-sense orientation by making use of restriction enzymes (NEB, Ipswich, Massachusetts, USA) followed by ligation using T4 DNA ligase (NEB, Ipswich, Massachusetts, USA) as described in the section 2.4.1. The positive clones were confirmed by sequencing before using for further experiments.

2.4.3 Cell Culture

Human embryonic carcinoma (EC) cell line NT2D1 (NTERA2-clone D1) was obtained as a kind gift from Dr. Peter Andrews, University of Sheffield, UK. They were grown in

Dulbecco's Modified Eagle's Medium with sodium pyruvate, high glucose (DMEM, Sigma-Aldrich, St. Louis, Missouri, USA) supplemented with 10% foetal bovine serum (Invitrogen, Carlsbad, California, USA), 2mM L-glutamine (Invitrogen, Carlsbad, California, USA) and penicillin-streptomycin (Invitrogen, Carlsbad, California, USA) and maintained at 37°C under 5% CO_2 atmosphere. EC cells were passaged upon reaching 70% confluency by gentle scraping and trypsin was not used for passaging. Human embryonic kidney cells (HEK293T) were grown in DMEM (Invitrogen, Carlsbad, California, USA) without sodium pyruvate, high glucose, supplemented with 10% foetal bovine serum (Invitrogen, Carlsbad, California, USA) and penicillin-streptomycin (Invitrogen, Carlsbad, California, USA) and maintained at 37°C under 5% CO_2 atmosphere. HEK293T cells were maintained passaged upon reaching 70-80% confluency using 0.05% trypsin (Invitrogen, Carlsbad, California, USA).

2.4.4 RA-mediated differentiation of EC cells

All-trans-retinoic acid (RA) (procured from Sigma-Aldrich, St. Louis, Missouri, USA) used for inducing differentiation of EC cells was reconstituted at a concentration of 5mg/ml in DMSO (Sigma-Aldrich, St. Louis, Missouri, USA) and stored as 20µl aliquots at -80°C. For differentiation experiments, EC cells were harvested using 0.05% trypsin, resuspended in fresh medium and seeded at a density of 0.15 X 10^6 cells in a 6-well plate or 1 X 10^6 cells in a 100mm tissue culture dish (Corning, Corning, NewYork, USA). Cells were allowed to grow for 24 hour following which few cells were harvested for RNA and protein extractions as 0 day control. RA was added to the remaining wells/plates at a concentration of 13.7µM for the remaining 6 days. Each day cells were either replenished with fresh medium and RA or harvested as day1/2/3/4/5/6 samples. Harvesting of the cells was done by scraping the cells off the plates using cell scrapers (Corning, Corning, NewYork, USA) in 1X chilled PBS, followed by spinning the cells down at 2000rpm, 4°C for 5 minutes. The cell pellets were either homogenized and resuspended in TriZol reagent (Invitrogen, Carlsbad, California, USA) or lysed using RIPA buffer for RNA and protein extractions respectively as described in sections 2.4.6 and 2.4.8.

2.4.5 Transfection of DNA and siRNA

EC cells or HEK293T cells were transfected with the DNA constructs using Lipofectamine 2000 (Invitrogen, Carlsbad, California, USA) as per manufacturer's guidelines. Briefly the cells were seeded and grown in the medium conditions described in section2.4.3.

Cells were transfected with equimolar amounts XIST promoter DNA constructs (0.3µM) in a 12-well plate for the reporter assay shown in Figure 2.5.3(B). For all the other reporter assays, EC cells were transfected with 1µg of Firefly luciferase construct along with 100ng of Renilla luciferase construct in a 12-well plate. For overexpression of OCT4 and SOX2, HEK293T cells were transfected with 3µg each of vector control, OCT4 or SOX2 DNA plasmids in a 6-well plate (Figure 2.5.6(C)). For overexpression of SATB1/2, EC cells were transfected with 8µg each of FLAG or FLAG-SATB1 or FLAG-SATB2 in a 100mm dish (Figure 2.5.8 (A) and Figure 2.5.9(A)) or 0.5µg each of FLAG or FLAG-SATB1 or FLAG-SATB2 along with 1µg of XIST1.1 and 100ng of Renilla luciferase for the reporter assay (Figure 2.5.9(C)). For transfecting EC cells with siRNAs, RNAiMax transfection reagent (Invitrogen, Carlsbad, California, USA) was used according to the protocol provided by the manufacturer. For RNAi-mediated knockdown of OCT4, SOX2 or NANOG in EC cells, 10 picomoles (pm) of each of siLuciferase, siOCT4, siSOX2 or siNANOG were used (Figure 2.5.7(B)). For siRNA mediated knockdown of SATB1 or SATB2 in EC cells, 150pm of siSATB1 and 100pm of siSATB2 were used (Figure 2.5.8(B)). All the transfections were carried out in OptiMEM medium (Invitrogen, Carlsbad, California, USA). 6 hours post transfection, culture dishes were replaced with the fresh complete DMEM medium. Cells were harvested for RNA isolation or protein extraction or reporter assays 48 hour post transfection or as indicated in the individual experimental schemes. For the results shown in the Figure2.5.4(C) and Figure2.5.9(D), reverse transfection were performed. Briefly, EC cells were transfected with the equal amounts of desired DNA constructs during the time of seeding. The transfection was done in DMEM + 10% serum + 2mM L-glutamine containing medium. Twelve hours post transfection, fresh medium containing antibiotic and 13.7 µM or equivalent DMSO was added to the culture dish. The cells were harvested at the indicated time points for the reporter assays.

2.4.6 Luciferase Reporter Assay

Luciferase reporter assays were performed using the Dual luciferase assay kit from Promega (Madison, Wisconsin, USA). Cells were transfected with Firefly luciferase and Renilla luciferase DNA constructs as described in section 2.4.5. After harvesting, the cells were lysed using 1X Passive lysis buffer (Promega, Madison, Wisconsin, USA) as per manufacturer's instructions. The lysates and substrates were mixed in the optical bottom 96-

well plate (ThermoFisher Scientific, Waltham, Massachusetts, United States) according to the guidelines provided by Promega. The reporter activities were measured using luminometry mode on the Varioskan machine (ThermoFisher Scientific, Waltham, Massachusetts, United States). In all the assays, Renilla luciferase activity measurement serves as an internal control. The fold change was calculated with respect to either the vector control or 0 day control as and when mentioned.

2.4.7 RNA extraction and cDNA synthesis

To isolate RNA, cell pellets were resuspended and homogenized in Trizol reagent (Invitrogen, Carlsbad, California, USA). 0.2 volumes of chloroform was added and mixed gently. The aqueous and the organic phases were allowed to get separated by standing the tubes for 5 to 10 minutes. The tubes were spun at 14000 rcf at 4°C for 10 minutes. The aqueous layer was collected in the fresh tube and 0.8 volume of isopropanol was added for precipitating RNA. The tubes were incubated at room temperature for 15 to 20 minutes and then spun at 14000 rcf at 4°C for 10 minutes, followed by washing the nucleic acid pellet with 75% ethan ol, air drying the pellet, dissolving it nuclease-free water and incubating at 55°C for 10 minutes. After assessing the quality and quantity of RNA using Nanodrop (ThermoFisher Scientific, Waltham, Massachusetts, United States), 2 µg of RNA was used for DNase treatment (Promega, Madison, Wisconsin, USA) as per the protocol provided by the manufacturer. This was followed by cDNA synthesis using reverse trascriptase kits either from Promega (ImpromII) (Madison, Wisconsin, USA) or Applied Biosystems (High capacity cDNA synthesis kit)(Foster City, California, USA). Also, -RT control was set up to verify the efficiency of DNase treatment. The synthesized cDNA was used to set up quantitative real-time PCR (qRT-PCR).

2.4.8 Chromatin immunoprecipitation

ChIP was performed essentially as per Jayani et al., 2010. Briefly, cells were crosslinked by adding formaldehyde to final concentration of 1% directly in the cell culture plate and keeping the plate on a rocker (Tarsons, Kolkata, West Bengal, India) for 10min at room temperature. The crosslinking was quenched by adding glycine to a final concentration of 125mM. The cells were washed twice with ice-cold 1X phosphate-buffered saline. The cells were collected by scraping and centrifuged at 2000rpm, 4°C for 5min. Cell pellets were washed sequentially in wash buffer 1 (0.25% Triton X-100, 10 mM EDTA, 0.5mM EGTA, 10

mM HEPES [pH 7.5], 1X Protease inhibitor cocktail [Roche]) and wash buffer 2 (0.2 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM HEPES [pH 7.5], 1X Protease inhibitor cocktail [Roche, Basel, Switzerland]). Cell pellets were resuspended in the lysis buffer (150 mM NaCl, 25 mM Tris-HCl [pH 7.5], 5 mM EDTA [pH 8.0], 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate, 1 mM PMSF, 10 mM sodium butyrate, 1X Protease inhibitor cocktail [Roche, Basel, Switzerland]). This was followed by sonication cycles (20-25 pulses of 10s on/off) using Bioruptor (Diagenode, Belgium) to obtain the chromatin fragment of the size 200 to 400bp. After getting desired chromatin fragment size, lysed cells were spun at 13000 rpm in a microcentrifuge at 4°C, 15 min. The supernatant is the input soluble cross-linked chromatin which can be diluted appropriately with the lysis buffer for further steps. 10-30 µg of fragmented chromatin was precleared by adding 20 µl/ml of protein A/G-plus bead cocktail (50% slurry, 100 mg of salmon sperm DNA/ml, 500 µg of BSA) and kept on the end to end rocker at 4^{o}C for 1-4 h. Post incubation, the precleared samples were centrifuged at 1000 X g, 4^{o}C, 5min. The precleared supernatant was divided into equal aliquots and incubated with 2µg of specific as well as isotype control antibodies, incubated overnight on the end to end rocker at 4^{o}C to immunoprecipitate DNA-protein complex. Post incubation, 20 µl of protein A/G-plus bead cocktail was added and kept on the end to end rocker at 4^{o}C for 2-4 h. Following this, the beads were harvested by centrifuging them at 1000g, 4^{o}C, 5min. The beads were washed twice with RIPA buffer (10mM Tris [pH 8.0], 1mM EDTA [pH 8.0], 0.5mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140mM NaCl) and twice with TE (10mM Tris, pH 8.0 and 1mM EDTA, pH 8.0). The chromatin antibody complexes were eluted by adding 2% SDS, 0.1 M $NaHCO_3$, and 10 mM DTT to the pellet. This was followed by reverse cross-linking by adding 0.05 volumes of 4M NaCl to the eluted complex and incubating at 65°C for 4 hours. Released protein was degraded by adding 0.025 volume of 0.5 M EDTA, 0.05 volume of 1 M Tris-HCl (pH 6.5), proteinase K (100 µg/ml) and incubated for 1 h at 45^{o}C. DNA was recovered by phenol:chloroform:isoamylalcohol extraction and purification (as described in the section 2.4.1). The recovered DNA was used to set up quantitative PCR as described in the section 2.4.10.

2.4.9 Protein extraction and immunoblotting

Cell pellets were resuspended in RIPA buffer (10mM Tris (pH 8.0), 1mM EDTA (pH 8.0), 0.5mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140mM NaCl)

containing 5% glycerol and 1X protease inhibitors (procured from Roche, Basel, Switzerland) and lysed by repeated freeze-thaw cycles (2 to 3). The lysates were centrifuged at 14000 rpm, 4°C, 30 minutes to get rid of the cellular debris. The supernatant was collected in the fresh microfuge tube. The concentrations of protein were determined by performing BCA assay (purchased from ThermoFisher Scientific, Waltham, Massachusetts, United States). Equal amounts of protein lysates were boiled in 1X Laemlli buffer (0.5 M Tris-HCl pH 6.8, 28% glycerol, 9% SDS, 5% 2-mercaptoethanol, 0.01% bromophenol blue) for 10-15 minutes and subjected to electrophoresis on a polyacrylamide gel. The separated proteins were transferred onto PVDF membrane (Millipore, Billerica, Massachusetts, USA) using phosphate based transfer buffer (10mM sodium phosphate monobasic, 10mM sodium phosphate dibasic) at 4°C, 400mA, 3 hours. After the completion of transfer, membranes were blocked in 5% skimmed milk, incubated overnight at 4°C with the primary antibodies prepared either in 5% milk or 5% BSA. The membranes were washed thrice with the buffer containing 20 mM Tris buffer pH 7.4, 500 mM NaCl and 0.1% tween 20 (TST) the next day and incubated with the appropriate secondary antibodies conjugated with horseradish peroxidase for an hour at room temperature. Following this, the membranes were again washed thrice with TST buffer. The blots were developed using Immobilon Western Chemiluminescent HRP Substrate (Millipore, Billerica, Massachusetts, USA) and detected using ImageQuant LAS 4000 (GE Healthcare, Piscataway, NJ, USA) according to the manufacturer's instructions.

2.4.10 Quantitative real-time PCR

For transcript quantitation, cDNA prepared (as described in the section 2.4.7) was diluted 5 times with nuclease free water and used as template for PCR along with specific set of primer pairs (see Appendices). SYBR Green chemistry (Roche and Applied Biosystems) was used and the reaction was set up on the Applied Biosystems ViiA7 thermal cycler. Changes in threshold cycles were calculated by subtracting the Ct values of the gene of interest from that of housekeeping control ($Ct_{target\ genes} - Ct_{\beta\text{-actin/18s rRNA}}$). ΔCt values of specific target genes from the experimental samples were then subtracted from their respective control samples to generate ΔΔCt values. The fold changes were calculated using the formula : $2^{(-(\Delta\Delta Ct\ value))}$.

For quantification after ChIP, DNA recovered post ChIP is diluted 5 times with nuclease free water. Diluted ChIP products were used as template for the PCR with specific

set of primers. For quantification of enrichment, the efficiency of chromatin immunoprecipitation of particular genomic locus can be calculated from qPCR data and reported as a percentage of starting material: % (ChIP/ Total input) which is calculated according to following formula:

% (ChIP/ Total input)= 2^[(Ct(x%input) – log(x%)/log2) - Ct(ChIP)]x 100%

Here, 2 is the amplification efficiency (AE); Ct (ChIP) and Ct (x%input) are threshold values obtained from exponential phase of qPCR for the IPed DNA sample and input sample respectively; the compensatory factor (logx%/log2) is used to take into account the dilution 1:x of the input. The recoveryis the % (ChIP/ Total input). Relative occupancy can be calculated as a ratio of specific signal over background:

Occupancy= % input (specific loci) / % input (background loci)

Relative occupancy is then used as a measure of the protein association with a specific locus.

2.4.11 Antibodies, siRNAs and other reagents

SATB1 (3650S) and SP1 (5931S) antibodies for immunoblotting were procured from Cell Signaling Technology (Danvers, USA), SATB2 (ab51502) and YY1 (ab12132) antibodies for western blotting were purchased from Abcam (Cambridge, UK), SATB1 (ab92307) and SATB2 (ab34735) for ChIP were procured from Abcam (Cambridge, UK), SOX2 (AF2018) and NANOG (AF1997) antibodies for immunoblotting and ChIP were from R & D(Menomonie, USA), OCT4 antibody (sc-9081) for western blotting and ChIP and goat-HRP secondary antibody were purchased from Santacruz Biotechnologies (Dallas, Texas, USA), CTCF antibody (07-729) procured from Millipore/Upstate (Billerica, Massachusetts, USA) was used for ChIP and CTCF antibody (sc-21298) purchased from Santacruz Biotechnologies (Dallas, Texas, USA) was used for immunoblotting, Normal rabbit IgG (12-370) and Normal mouse IgG (12-371) for ChIP were purchased from Millipore (Billerica, Massachusetts, USA), Normal goat IgG (sc-2028) used is from Santacruz Biotechnologies (Dallas, Texas, USA), siRNA targeting SATB1 (sc-36460), SATB2 (sc-76456) were procured from SantaCruz Biotechnologies (Dallas, Texas, USA), β-ACTIN (VMA00048) and γ-TUBULIN primary antibodies, mouse-HRP and rabbit-HRP secondary antidbodies were purchased from Biorad Laboratories (Hercules, California, USA), Dual luciferase assay kit (E1960) was purchased from Promega, Lipofectamine 2000 and

RNAiMax transfection reagents were procured from Invitrogen, all the restriction enzymes and T4 DNA ligase was procured from NEB, RNase-free DNase and ImpromII reverse transcriptase enzyme was purchased from Promega (Madison, Wisconsin, USA), High specificity cDNA synthesis kit is from Applied Biosystems (ABI) (Foster City, California, USA), SYBR Green for qPCR was procured from ABI (Foster City, California, USA) and Roche (Basel, Switzerland), Protein A/G Plus Ultralink resin (Cat. No. 53135) was obtainedfrom ThermoFisher Scientific (Waltham, Massachusetts, United States).

2.5 Results

2.5.1 Human embryonic carcinoma (EC) cells responds to differentiation cues

In order to validate if EC cells can be used as a system to understand the regulation of *XIST* in the context of initiation phase of XCI, we induced the differentiation of these cells upon treating them with RA for 6 days. The experimental paradigm and the associated gross morphological changes of the cells are depicted in the Figure 2.5.1(A). These cells grow in cluster when in undifferentiated state and exhibit flattened and elongated appearance with neuron-like projections upon administering RA as shown in the Figure 2.5.1 (A). As expected, the levels of pluripotency factors decrease, both at the transcript as well as protein levels as assessed by quantitative real-time PCR (qRT-PCR) and immunoblotting (Figure 2.5.1(B-E)). Also, the transcript levels of differentiation marker – PAX6, a key transcription factor for specifying neuronal lineage, increase upon differentiation (Figure 2.5.1 (D)). These results suggest that the differentiation regime tested was successful.

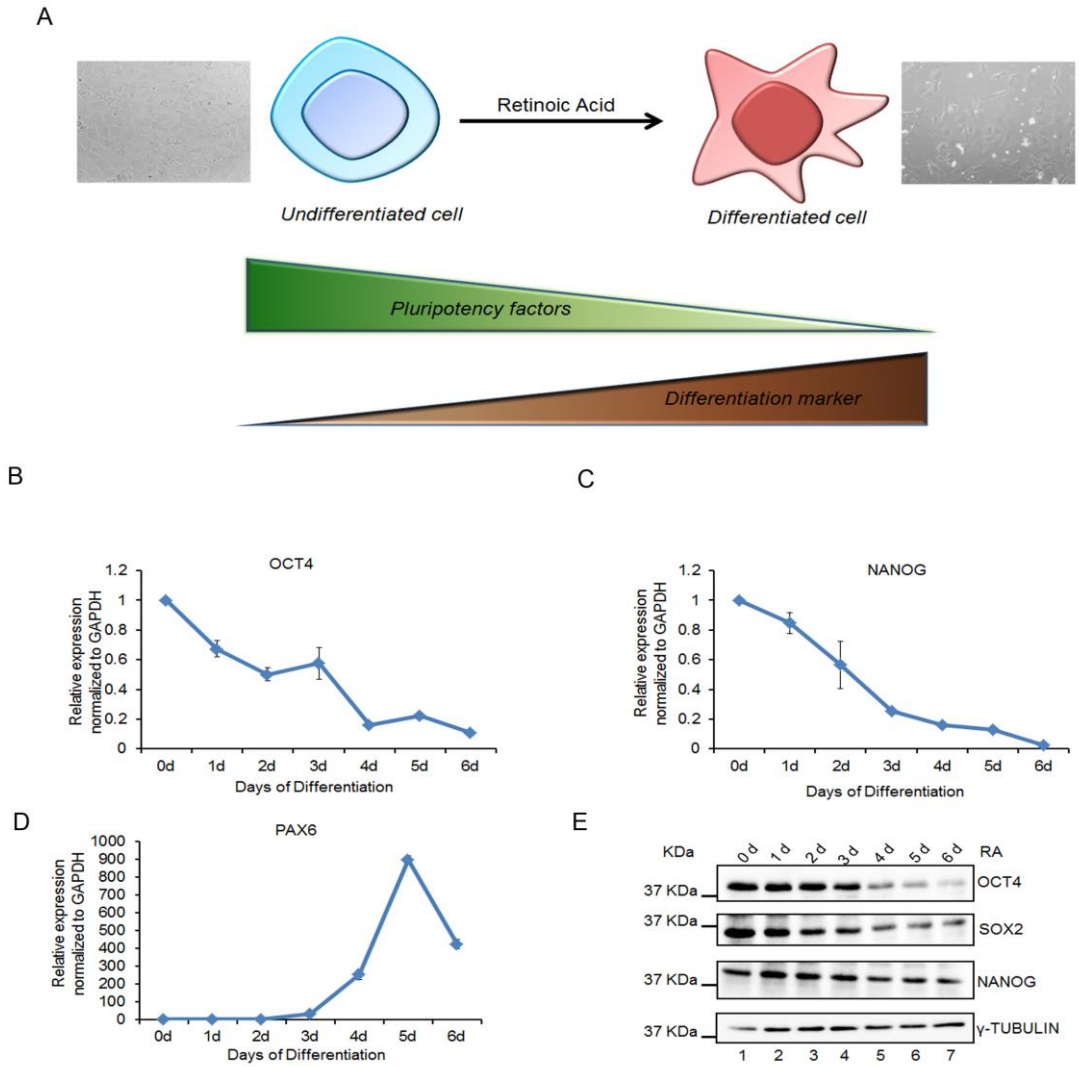


Figure 2.5.1 Human EC cells respond to RA-mediated differentiation cues.

EC cells were subjected to RA-mediated differentiation as described in 'Materials and Methods'section (2.4.4). Cells were harvested at indicated days and frozen immediately. At the end of differentiation series, RNA and protein extraction and cDNA synthesis was performed as described in 'Materials and Methods' section (2.4.7).

(A) Experimental paradigm for RA-mediated differentiation of EC cells. DIC image showing gross morphological changes upon administering RA are taken from Sunita Singh, PhD thesis, 2013.

(B-D) qRT-PCR for OCT4, NANOG and PAX6 using gene specific primers (see Appendix1) indicate that levels of pluripotency factors OCT4 and NANOG decrease progressively while the expression of differentiation marker PAX6 increases during differentiation. Fold change was calculated using ΔΔCt method. GAPDH was used as an endogenous control for normalization.

(E) Immunoblot for OCT4, SOX2 and NANOG indicate that their levels decrease during differentiation as compared to the loading control γ-TUBULIN. Treatment with RA is indicated on the top and the antibodies used for western blotting are indicated on the side of the immunoblot image.

2.5.2 Expression of XIST is upregulated upon RA-mediated differentiation of EC cells

After assessing the RA-mediated differentiation kinetics of EC cells, we next sought to determine the expression of XIST in these cells during differentiation. As mentioned earlier, EC cells express low amounts of XIST as detected by RNA-FISH (Chow et al., 2003). Moreover, this cell line has more than one copy of X chromosome (as per ATCC CRL-1973). It is a well-established fact that cells have built-in mechanism to count the number of X chromosomes they harbour and inactivates all in excess of one. For instance, it has been documented that XO females do not exhibit XCI and one of the Xs in XXY males (Klinefelter's Syndrome) is inactivated (Grumbach et al., 1963; Monkhorst et al., 2009). Moreover, in mouse system, expression of Xist is upregulated upon induction of differentiation (Kay et al., 1993, 1994). Therefore, we wished to determine the status of XIST expression in response to differentiation cues. Towards this, EC cells were treated with RA for 6 days and expression of XIST RNA was measured by qRT-PCR for the indicated days. It was observed that the XIST RNA is upregulated upon RA-induced differentiation of EC cells (Figure 2.5.2). Hence, EC cells provide a good model system to understand the regulation of *XIST* in the context of initiation phase of XCI (as mentioned previously Chow et.al 2003).

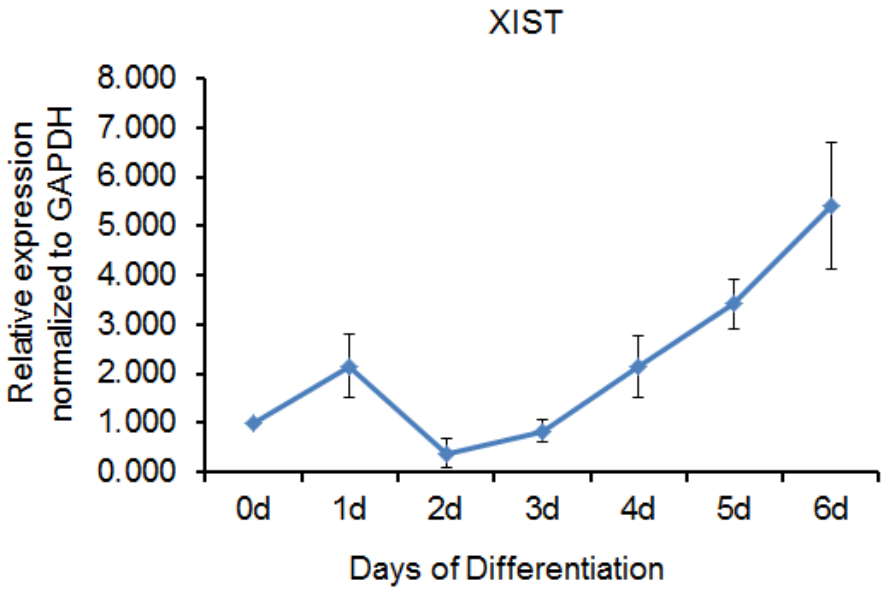


Figure 2.5.2 Expression of XIST increases during RA-mediated differentiation of EC cells.

EC cells were subjected to RA-mediated differentiation as described in 'Materials and Methods' section (2.4.4). Cells were harvested at indicated days and frozen immediately. At the end of differentiation series, RNA was extracted and cDNA was synthesized as described in 'Materials and Methods' section (2.4.7). Quantitative RT-

PCR for XIST was performed using gene-specific primers (see Appendix 1). Fold change was calculated using ΔΔCt method and GAPDH was used as an endogenous control for normalization. X-axis indicates the days of differentiation and Y-axis indicate the fold change in expression of XIST normalized to GAPDH.

2.5.3 Identification of promoter elements for *XIST*

To understand the regulation of human *XIST* gene, it is important to identify its promoter elements. Towards this, we cloned genomic regions ranging from 101 bp to 4458 bp upstream of *XIST* Transcription start-site (TSS) including a part of 5' region of the gene into a promoter less reporter vector – pGL3 Basic. The schematic of the cloned regions and the image for the positive clones is shown in the Figure 2.5.3(A). These regions were selected on the basis of the ChIP-sequencing peaks observed for some of the identified transcription factors such as SP1 and YY1 from the ENCODE data set as discussed and depicted in section 2.5.4. These DNA constructs were then transfected into EC cells and assayed for the potential promoter elements by measuring the firefly luciferase activity. Similar to the first study on characterizing *XIST* promoter elements (Hendrich et al., 1997), it was observed that 101 bp region including a few bases from the 5' region of *XIST* gene and bases from the region immediately upstream of the TSS was sufficient to drive the transcription of luciferase gene. Hence this can be called the 'minimal promoter region'. Also the promoter activities for the all other fragments with increasing distance from the TSS (except for the fragment XIST1.2, bar 3 in Figure 2.5.3(B)) remained constant when compared to the vector control (Figure 2.5.3(B)). As a control, we also cloned one of the fragments – XIST1 which includes genomic region 1100 bp in the anti-sense orientation. Since *XIST* promoter is a unidirectional promoter, the fragment cloned in anti-sense orientation fails to transcribe the reporter gene and hence does not exhibit any measurable reporter activity (Bar 7, Figure 2.5.3(B)).

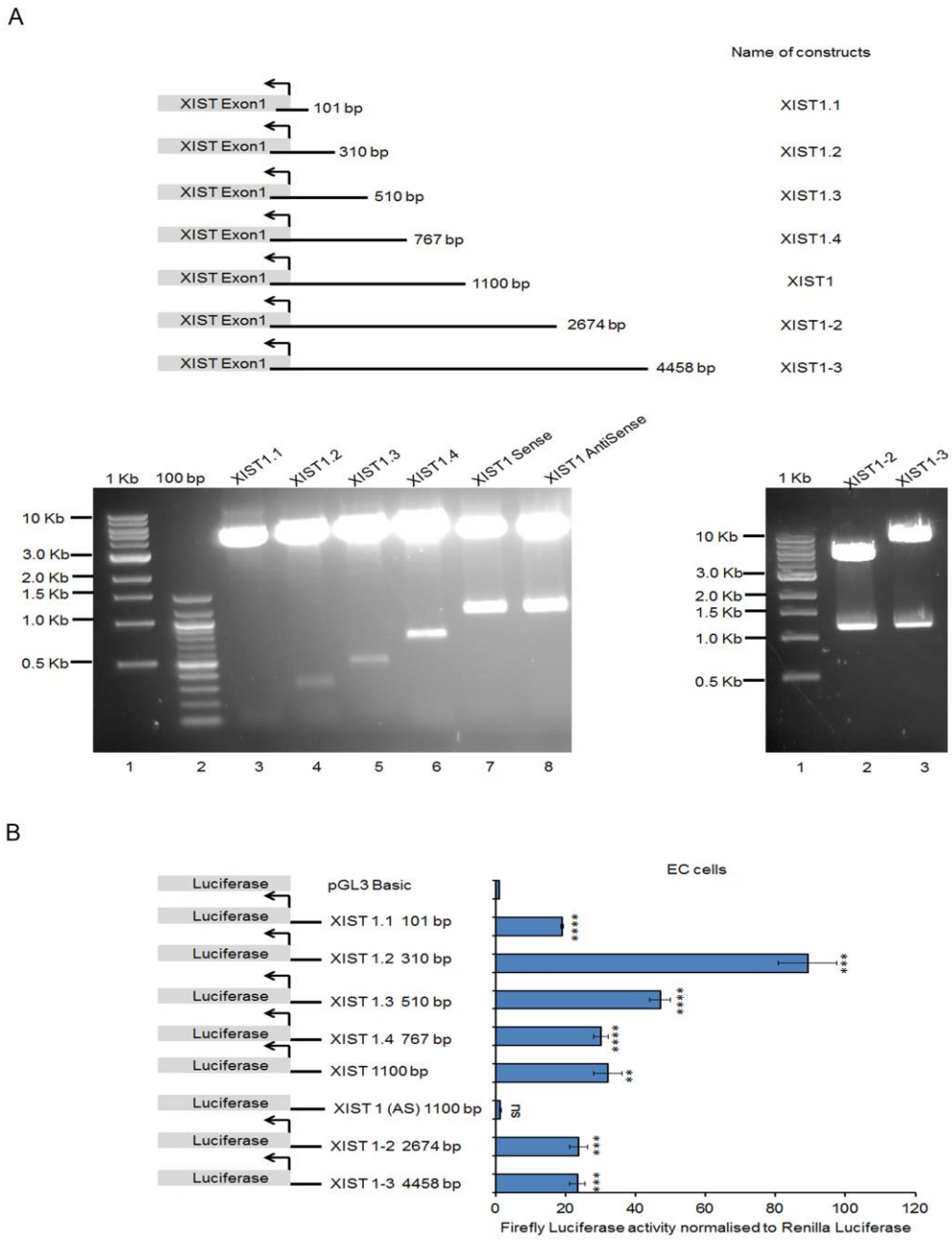


Figure 2.5.3Identification of *XIST* promoter region by luciferase reporter assay.

Genomic region upstream of *XIST* TSS including a part of 5' region of gene was cloned into pGL3 Basic as described in the 'Materials and Methods' section (2.4.1) (see Appendix 2 for primer details). Equimolar amounts of the confirmed clones along with Renilla Luciferase were transfected into EC cells. 48 hour post transfection, cell were harvested and the reporter assay was performed as described in the section 2.4.6.

(A) Shown on the left is the schematic of genomic regions of *XIST* promoter fragments cloned into promoter less pGL3 Basic vector. Depicted on the right are the names of these DNA constructs. Below the schematic is the agarose gel image showing the promoter clones.

(B) Luciferase reporter activities for XIST promoter constructs transfected into EC cells compared to pGL3 Basic vector control. All the promoter constructs tested except for XIST1 (AS) generate reporter transcription. Firefly

luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis represents the normalized fold change in the firefly luciferase activity and Y-axis indicates DNA constructs transfected into EC cells. Each bar represents values from three independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, *** P value < 0.001, ** P value < 0.01, * P value < 0.05, ns = non-significant).

2.5.4 SP1, YY1 and CTCF are the potential factors regulating transcription from *XIST* promoter

The first study attempting to uncover the regulation of human *XIST* gene identified SP1, YY1 and TBP as its potential transcriptional regulators by performing site directed mutagenesis and *in vitro* binding assays (Hendrich et al., 1997). However, the *in vitro* techniques although useful have their own limitations and may not be a true picture of *in vivo* scenario. In the post human genome project and ENCODE consortium era, the binding sites and motifs for many more transcription factors have been unraveled. We made use of publicly available ENCODE (Encyclopedia of DNA elements) data from the UCSC (University of California, SantaCruz) genome browser and verified if any of these identified factors also bind *XIST* promoter region in the cell system. Shown in the Figure 2.5.4 (A-a) is the UCSC genome track snapshot for *XIST* genomic region including its promoter from a male human ES cell line H1 and a female lymphoblastoid line GM12878. In addition to finding binding sites for SP1 and YY1, we also observed occupancy of CTCF as per ChIP-sequencing profiles from GM12878. However, in case of the GM12878 cell lines it is unknown whether the binding of said factors is on the active or inactive X since it is not allele-specific ChIP-sequencing.

As can be seen from the image, SP1 and YY1 show occupancies at the *XIST* TSS which was known previously as well as 1 Kb downstream of the TSS into the 1st exon. Also, CTCF shows binding approximately 2.5 Kb upstream of TSS along with YY1 as well as 2.5 Kb downstream of TSS. The binding sites for these 3 factors also correlate with the DNase protected region as determined by sequencing following DNase treatment. Overlaying of promoter/enhancer specific active histone marks such as H3K4me3 and H3K27ac indicates enrichment of these marks at the *XIST* TSS as well as 1 Kb region downstream of TSS, both coinciding with the occupancy of SP1 and YY1. The chromatin status as determined by ChromHMM (ENCODE) on the basis of histone modification and RNA sequencing profiles found in H1 and GM12878 cell lines were completely distinct. This correlates well with the state of promoter in these cells (Figure 2.5.4 (A-a)). The observations from the ENCODE data

set described are depicted in a simplified manner below the UCSC track snapshot (Figure 2.5.4 (A-b))

Based on these, we sought to address the role of SP1, YY1 and CTCF in regulating *XIST* promoter in EC cells. Since levels of XIST increase upon differentiation of these cells (Figure 2.5.2.), we first determined the expression of SP1, YY1 and CTCF under differentiation conditions. As can be observed, SP1, YY1 and CTCF are downregulated during the course of differentiation (Figure 2.5.4(B)). Next we tested the activity of *XIST* promoter constructs containing binding sites for SP1 or YY1 or CTCF biochemically. Towards this we transiently transfected them into EC cells and subjecting the cells to RA-induced differentiation for 3 days following which the reporter activity was measured. The scheme of the experiment is shown in Figure 2.5.4(C). In accordance with decrease in the levels of the transcription factors under investigation in our study, a significant reduction in the promoter activities of the three reporter constructs was also observed, declining maximally at day 3 of differentiation (Figure 2.5.4(D-F)). These observations put together suggest that SP1, YY1 and CTCF could be the potential factors regulating transcription from *XIST* promoter.

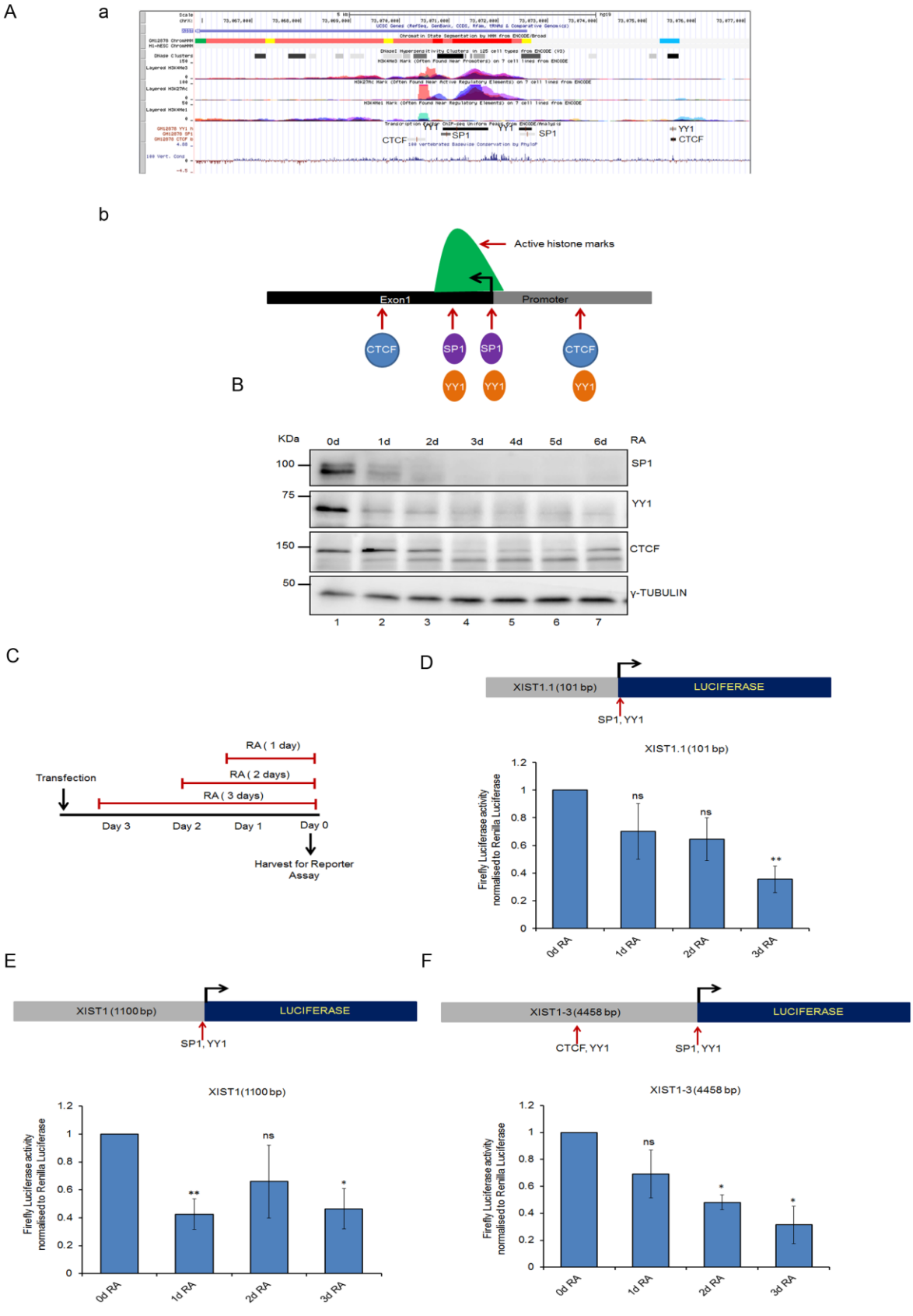

A
a
CTCF
SP1
YY1
b
Active histone marks
Exon1
Promoter
CTCF
SP1
SP1
CTCF
YY1
YY1
YY1
B
KDa
0d
1d
2d
3d
4d
5d
6d
RA
100
SP1
75
YY1
150
CTCF
50
γ-TUBULIN
1
2
3
4
5
6
7
C
RA (1 day)
RA (2 days)
RA (3 days)
Transfection
Day 3
Day 2
Day 1
Day 0
Harvest for Reporter Assay
D
XIST1.1 (101 bp)
LUCIFERASE
SP1, YY1
XIST1.1 (101 bp)
Firefly Luciferase activity normalised to Renilla Luciferase
1.2
1
0.8
0.6
0.4
0.2
0
ns
ns
**
0d RA
1d RA
2d RA
3d RA
E
XIST1 (1100 bp)
LUCIFERASE
SP1, YY1
XIST1(1100 bp)
Firefly Luciferase activity normalised to Renilla Luciferase
1.2
1
0.8
0.6
0.4
0.2
0
**
ns
*
0d RA
1d RA
2d RA
3d RA
F
XIST1-3 (4458 bp)
LUCIFERASE
CTCF, YY1
SP1, YY1
XIST1-3 (4458 bp)
Firefly Luciferase activity normalised to Renilla Luciferase
1.2
1
0.8
0.6
0.4
0.2
0
ns
*
*
0d RA
1d RA
2d RA
3d RA

Figure 2.5.4 SP1, YY1 and CTCF are the potential transcription regulators of *XIST*.

(A)(a) UCSC genome browser track showing *XIST* genomic region (chrX:73,067,000-73,077,000) overlayed with promoter/enhancer specific histone modifications – H3K4me3 and H3K27ac, DNase protected regions, chromatin segmentation state for male hESC line H1 and female lymphoblastoid cell line GM12878. Also shown is the ChIP-seq profile for SP1, YY1 and CTCF from GM12878 cell line. All the information displayed is generated by ENCODE consortium. (b) Simplified schematic depicting binding sites for SP1, YY1 and CTCF as well as histone modification profile as determined from the ENCODE data set.

(B) Expression profile of SP1, YY1 and CTCF upon RA-mediated differentiation of EC cells by immunoblotting. γ-TUBULIN serves as a loading control. Treatment with RA is indicated on the top and the antibodies used for western blotting are indicated on the side of the immunoblot image.

(C) Experimental scheme for measuring reporter activities in differentiation conditions for XIST1.1, XIST1 and XIST1-3 constructs by transiently transfecting them into EC cells as mentioned in the Section 2.4.5.

(D-F) Reporter activity for all the three constructs decline significantly at Day 3 of differentiation relative to Day 0, in accordance with decreased levels of SP1, YY1 and CTCF. Firefly luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis represent days of differentiation and Y-axis indicate normalized fold change values for the promoter constructs. Each bar represents values from three independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (**P value < 0.01, *P value < 0.05, ns=non-significant).

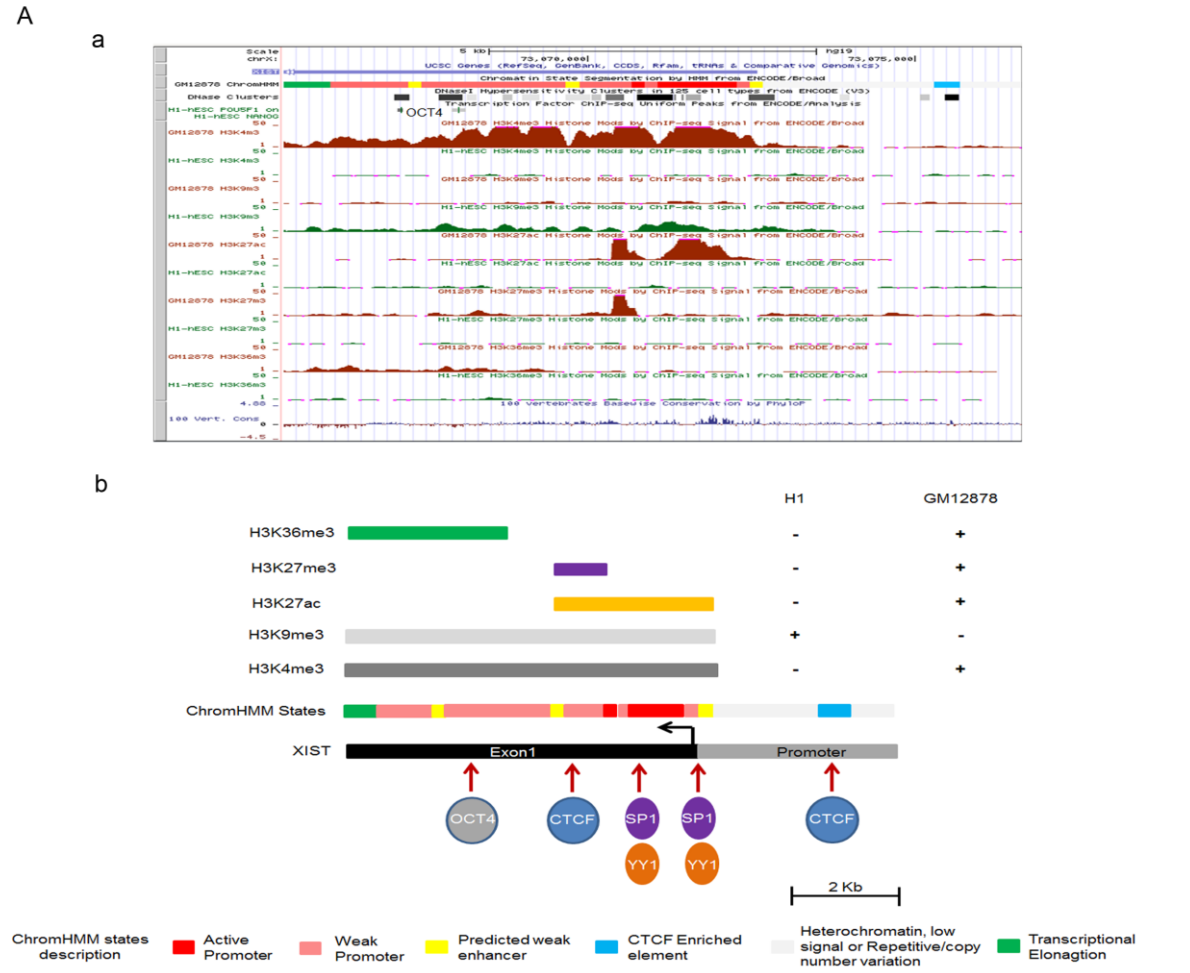
2.5.5 Pluripotency factors bind *XIST* Exon1 region

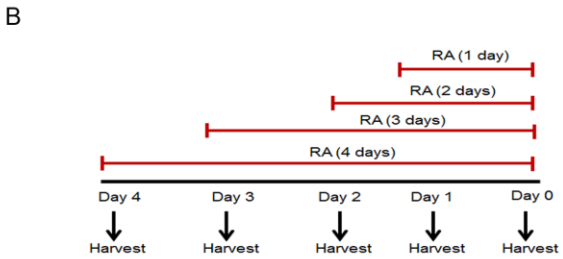
The two contradictory observations of XIST levels increasing upon differentiation (Figure 2.5.2) and the levels of SP1, YY1, CTCF as well as the promoter activities assayed biochemically decreasing (Figure 2.5.4(C) & (D)) indicated that there could be additional factors involved in temporally modulating the levels of XIST in EC cells. Since the levels of pluripotency factors go down upon differentiation of EC cells and the finding, although debatable, that they are important regulators of mouse Xist (Donohoe et al., 2009; Minkovsky et al., 2013; Navarro et al., 2008, 2010; Nesterova et al., 2011) prompted us to investigate if they are involved in regulating human *XIST*. We first determined if the pluripotency factors were known to bind *XIST* genomic region as per ChIP-sequencing profiles of these proteins from male hESC line H1 in the publicly available ENCODE data set. OCT4 exhibited binding to *XIST* Exon1 region, ~ 5.0 Kb and ~5.5 Kb downstream of TSS (Figure 2.5.5(A-a,b)). Moreover, the occupancy of OCT4 correlates with DNase protected region as well as a putative weak promoter element as per ChromHMM chromatin state segmentation. In addition, enrichment of promoter specific active histone mark – H3K4me3 and active elongation mark H3K36me3 were found to be widespread spanning well into the 1st exon of *XIST* in a female cell line GM12878 which expresses XIST as opposed to male hESC H1 which does not express XIST. Also, repressive histone modification H3K9me3 was enriched on *XIST* genomic region in H1 cells and absent on GM12878 (Figure 2.5.5(A-a,b)).

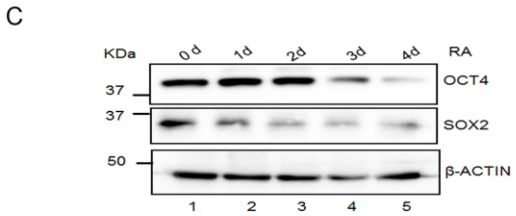
This suggests that the epigenetic status of *XIST* genomic locus in both these cell lines may be essential for governing XIST expression. However, what factors contribute towards these distinct states in a temporal fashion still remains elusive. This epigenetic profile is summarized in a simplified manner in Figure 2.5.5(A-b).

We validated the occupancy of OCT4 and also tested if SOX2 and NANOG occupied the same site on *XIST* Exon1 (~ 5.5 Kb downstream of TSS) by performing ChIP in differentiating EC cells. We chose to test their occupancies in differentiating EC cells to be able to probe whether their binding kinetics alter during the process of differentiation which could have temporal influence on the induction of XIST levels. Towards this, seeded EC cells were induced for differentiation by treating them with RA for 4 days. The experimental scheme followed is shown in the Figure 2.5.5(B). Each day cells were harvested for protein and cross-linked for ChIP. Successful differentiation was verified by checking the expression of OCT4 and SOX2 by immunoblotting (Figure 2.5.5(C)). ChIP was performed using OCT4, SOX2 and NANOG specific antibodies as described in 'Materials and Methods' section 2.4.8. Quantitative RT-PCR was performed using the primers flanking OCT4 binding region (as per ENCODE) (see Appendix 3) with the DNA recovered post ChIP. It was observed that all the three pluripotency factors showed occupancy to the region tested.

A distinct profile of binding was observed, with lowest occupancies observed at day 0, 1 and peaking at day 2 of differentiation and again declining by day 3 and 4 (Figure 2.5.5 (D)). Interestingly, similar to mESCs (Navarro et al., 2008, 2010), the binding kinetics of the pluripotency factors showed inverse correlation with XIST expression in our system as well (compare Figure 2.5.2 with Figure 2.5.5(C)). This suggests that pluripotency factors could act as the negative regulators of XIST expression in EC cells.


A
a
OCT4
b
H1
GM12878
H3K36me3
H3K27me3
H3K27ac
H3K9me3
H3K4me3
ChromHMM States
XIST
Exon1
Promoter
OCT4
CTCF
SP1
SP1
CTCF
YY1
YY1
2 Kb
ChromHMM states description
Active Promoter
Weak Promoter
Predicted weak enhancer
CTCF Enriched element
Heterochromatin, low signal or Repetitive/copy number variation
Transcriptional Elonagtion


B
RA (1 day)
RA (2 days)
RA (3 days)
RA (4 days)
Day 4
Day 3
Day 2
Day 1
Day 0
Harvest
Harvest
Harvest
Harvest
Harvest


C
KDa
0 d
1 d
2 d
3 d
4 d
RA
37
37
50
OCT4
SOX2
β-ACTIN
1
2
3
4
5

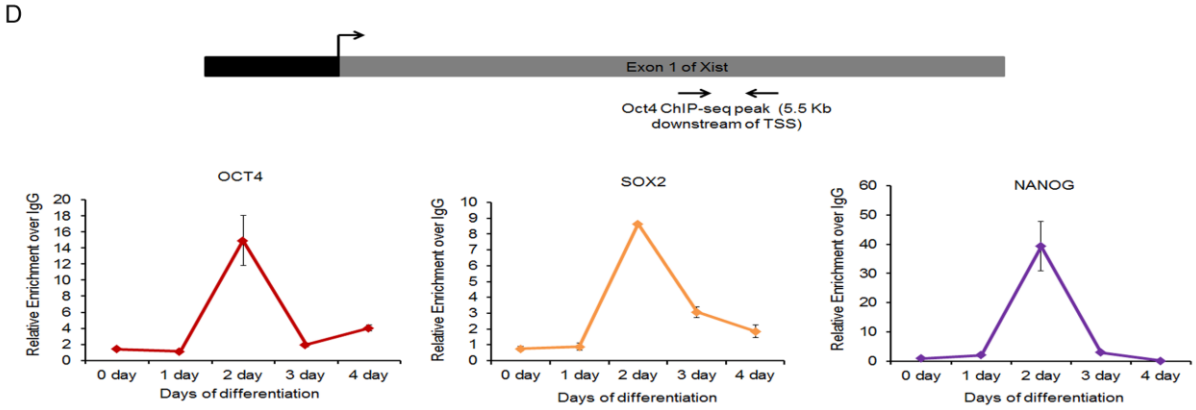

D
Exon 1 of Xist
Oct4 ChIP-seq peak (5.5 Kb downstream of TSS)
OCT4
SOX2
NANOG
Relative Enrichment over IgG
0 day
1 day
2 day
3 day
4 day
Days of differentiation

Figure 2.5.5 Pluripotency factors bind a site on *XIST* Exon1.

(A) (a) UCSC genome browser track showing *XIST* genomic region (chrX:73,067,000-73,077,000) overlayed with active (H3K4me3, H3K27ac, H3K36me3) and repressive (H3K9me3, H3K27me3) histone modifications, DNase protected regions, chromatin segmentation state for male hESC line H1 and female lymphoblastoid cell line GM12878. Also shown is the ChIP-seq profile for OCT4 from H1 cells. All the information displayed is generated by ENCODE consortium. (b) Schematic depicting binding sites for SP1, YY1, CTCF, OCT4 as well as histone modification profile as determined from the ENCODE data set from H1 and GM12878 cell lines.

(B) Experimental scheme for RA-induced differentiation of EC cells for 4 days.

(C) Immunoblotting for OCT4 and SOX2 under these conditions confirm successful differentiation as their levels decrease compared to the loading control β-ACTIN. Treatment with RA is indicated on the top and the antibodies used for western blotting are indicated on the side of the immunoblot image

(D) qRT-PCR for the region indicated in *XIST* genomic region schematic for the DNA recovered post ChIP using antibodies specific for OCT4, SOX2 and NANOG at Day 0, 1, 2, 3 and 4 of differentiation. Maximum occupancies were observed at Day 2 of differentiation with OCT4, SOX2 and NANOG showing 16-fold, 8-fold and 40-fold relative enrichment over IgG control. Ct values for ChIP were normalized to that for IgG control and plotted using the formula described in 'Material and Methods' section 2.4.10. X-axis represents Days of differentiation and Y-axis represents normalized fold change for ChIP.

2.5.6 Pluripotency factors negatively regulate XIST

From the previous experiment (2.5.5) it seemed likely that OCT4, SOX2 and NANOG may be acting as repressors of *XIST*. In order to validate this, we cloned the region to which the pluripotency factors exhibited binding into two of the promoter constructs – XIST1.1 (101 bp region) and XIST 1 (1100 bp region), between the promoter element and luciferase gene to resemble *XIST* genomic context as shown in the schematic (Figure 2.5.6(A)). This region is labelled as Pluripotency Factor Binding (PFB) site. Next we transfected XIST1.1 and XIST1 constructs with and without the pluripotency factor binding site into EC cells along with Renilla luciferase as an internal control. The cells were harvested for testing the promoter activity biochemically 48 hour post transfection. It was observed that the presence of the region to which OCT4, SOX2 and NANOG could potentially bind led to a significant decrease in the promoter activities of the both the DNA constructs (Figure 2.5.6(A)). However, it can be appreciated that the decrease observed for 1.1 Kb long promoter fragment is much less compared to that observed for the minimal promoter region (101 bp). Importantly, the promoter activities for both these fragments are more or less similar (Figure 2.5.3 (B)).

Interestingly, as per ENCODE CTCF ChIP-seq data from multiple cell lines, multiple low affinity peaks for CTCF binding were observed in 1.1 Kb region and not in the 101 bp promoter regions (Figure 2.5.11 (C)). Besides, it was also observed that CTCF occupies this

PFB site in EC cells (shown in Figure 2.5.10 (D)). Taken together, it can be speculated that when the pluripotency factor binding site is present alongside low affinity CTCF binding sites, there is a partial rescue in the promoter activities. Further experimentation is needed to conclusively state the significance of this observation. However, the above experiment is just a biochemical assay and has its own limitations. To conclusively argue that the pluripotency factors are the potential repressors of *XIST* transcription, one needs to score for the levels of XIST upon perturbing the levels of pluripotency factors. Therefore, to probe further into their roles, we chose to make use of a female cell line which does not express any of these stem cell factors but being a cell line from a female source, does express XIST. Towards this we transfected HEK293T cells with OCT4 and SOX2 overexpression DNA constructs. HEK293T is a female cell line which expresses significant amounts of XIST from the maintenance phase Xi. If OCT4 and SOX2 are negative regulators of *XIST*, then levels of XIST should decrease upon their forced expression in these cells. And indeed it was observed that SOX2 overexpression led to decrease in XIST levels as determined by qRT-PCR (Bar 3, blue colour, figure 2.5.6(C)). We also determined the levels of unspliced version of XIST, referred to as premature XIST in the graph by qRT-PCR in order to confirm that the decline in XIST levels was due to transcriptional repression and not due to decreased stability of mature XIST lncRNA (Bar 2, 3, red colour, Figure 2.5.6(C)). We also tested whether overexpression of OCT4 and SOX2 affected levels of potential activators of *XIST* – SP1, YY1 and CTCF. It was observed that while SP1 levels show a mild reduction, YY1 levels are increased marginally and CTCF expression remains unperturbed as compared to the vector control in the experiment (Figure 2.5.6(B)). These results suggest that the pluripotency factors play the role of repressors, fine-tuning XIST expression and possibly providing the necessary temporal control for the initiation of XCI.

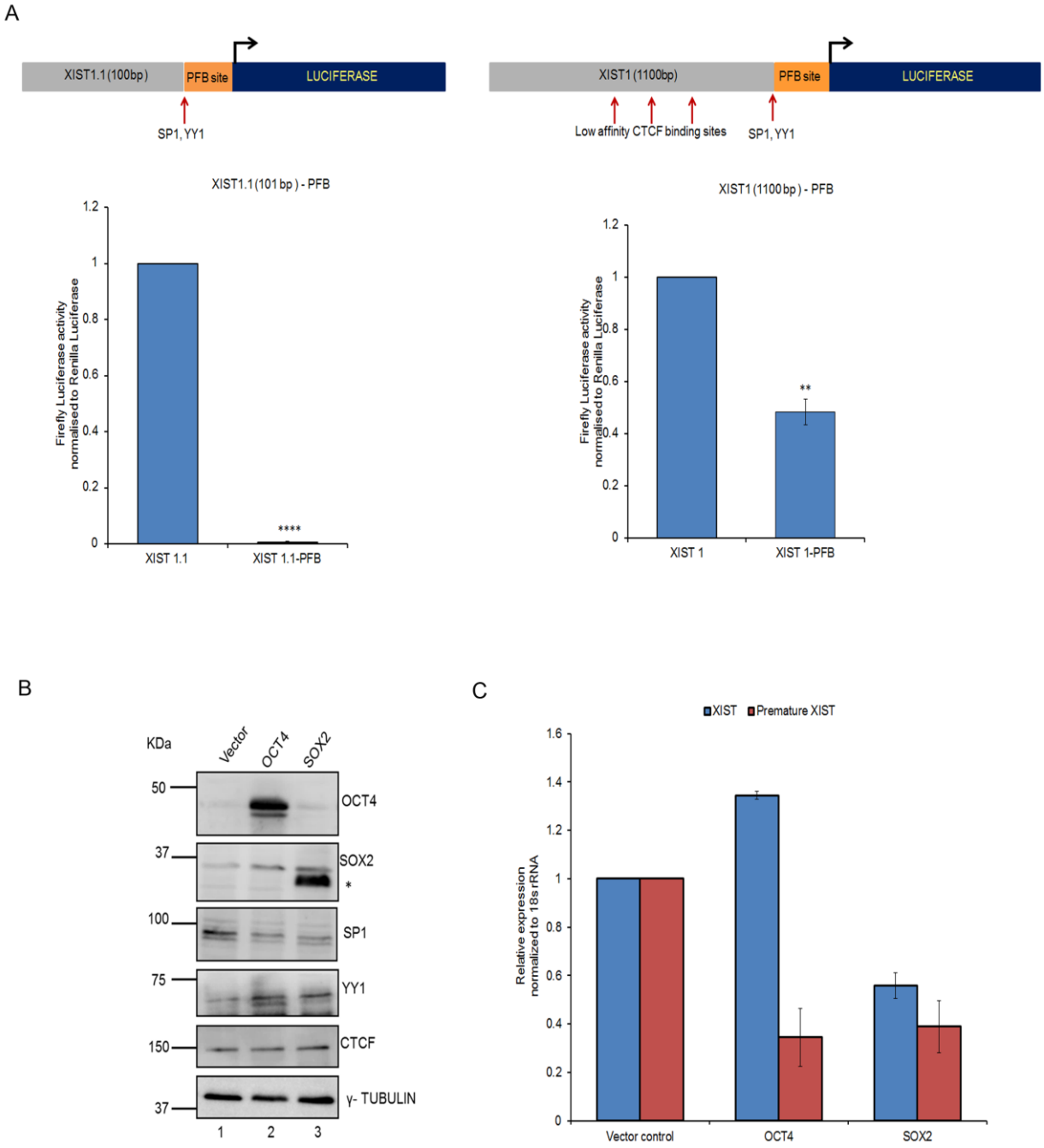


Figure 2.5.6 Pluripotency factors negatively regulate *XIST*.

(A) Schematic showing insertion of pluripotency factor binding (PFB) site between XIST1.1/XIST1 promoter element and luciferase gene. Also shown are the low affinity CTCF binding sites as per CTCF ChIP-seq peaks from ENCODE data set. Luciferase reporter activities of XIST1.1-PFB and XIST1-PFB DNA constructs measured 48 hour post transfection in EC cells. Firefly luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis indicates the DNA constructs transfected into EC

cells and Y-axis indicates the normalized fold change observed. Each bar represents values from three independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over XIST1.1 or XIST1 control as per Student's T-test (****P value < 0.0001, ** P value < 0.01).

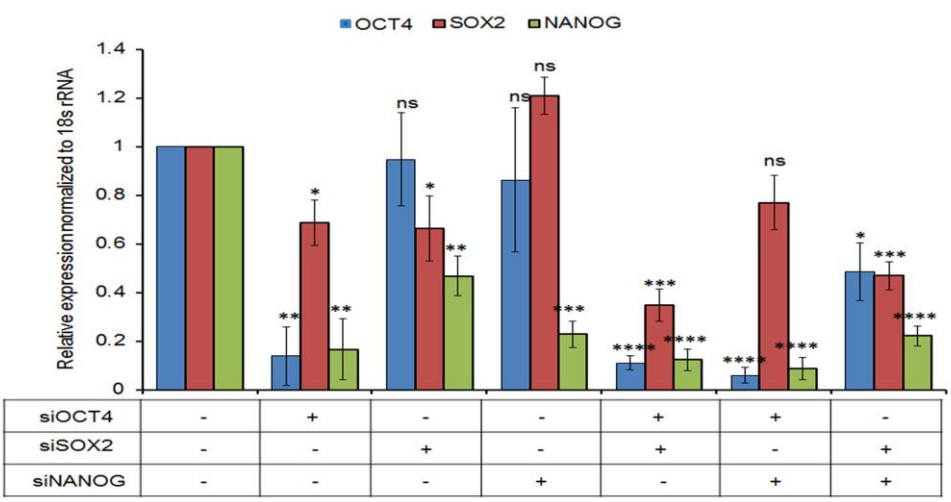
(B) Validation of OCT4 and SOX2 overexpression in HEK293T cells by immunoblotting using specific antibodies. Immunoblotting for SP1, YY1 and CTCF using specific antibodies. Levels of SP1 show mild reduction, YY1 shows a slight upregulation, whereas CTCF expression remains undisturbed upon forced expression of OCT4 and SOX2 in HEK293T cells. Western blot for γ-TUBULIN serves as a loading control. The DNA constructs transfected into HEK293T cells are indicated on the top and the antibodies used for probing the proteins are indicated on the side of the immunoblot image.

(C) qRT-PCR for mature (C) as well as premature XIST using gene specific primers (see Appendix 1) indicate that their expression declines by 0.5-fold relative to 18s rRNA upon overexpression of SOX2. X-axis indicated the DNA constructs transfected into HEK293T cells and Y-axis represents the normalized fold change.

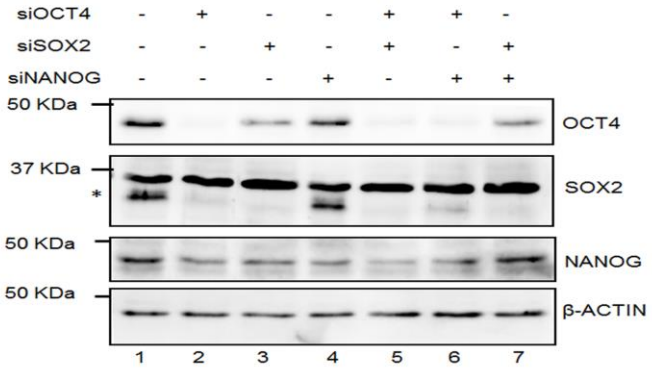
2.5.7 Depletion of OCT4, SOX2 and NANOG levels in EC cells does not lead to upregulation of XIST

Results from the previous experiments led us to speculate that the pluripotency factors may be playing an essential role in keeping *XIST* repressed in undifferentiated EC cells and quite possibly in hESCs. Therefore we next asked if the levels of XIST can be induced upon depleting the levels of OCT4, SOX2 and NANOG in EC cells. To address this, we transfected EC cells with siRNA targeting either OCT4, SOX2 or NANOG. The morphology of the cells resembled that of differentiating cells as determined microscopically (Figure 2.5.7 (C)). Cells were harvested for RNA and protein 48 hour post-transfection. The efficiencies of knockdowns were validated by qRT-PCR (Figure 2.5.7(A)) and immunoblotting for OCT4, SOX2 and NANOG (Figure 2.5.7(B)). Depletion of these proteins led to differentiation of EC cells as assessed by measuring the transcript levels of a differentiation marker – PAX6, which is known to be upregulated upon RA-induced differentiation (Figure 2.5.7(D)). However, Ct values for mature as well as premature XIST remained undetermined as in undifferentiated cells. Our results indicate that decreasing the levels of OCT4 or SOX2 or NANOG, either alone or in combination, surprisingly may not be sufficient to upregulate XIST. Nonetheless, it is possible that differentiation induced by perturbing the levels of pluripotency factors may in turn have affected the activities of yet unidentified regulator(s) of *XIST*.

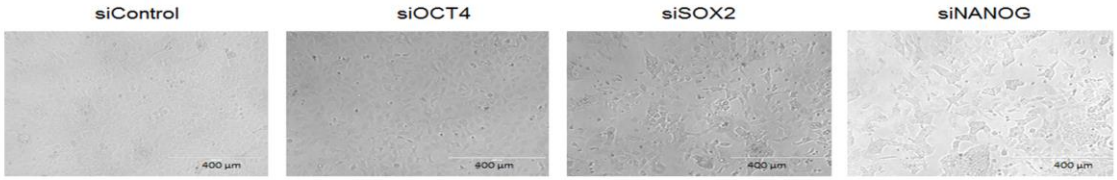
A


OCT4
SOX2
NANOG
Relative expression normalized to 18s rRNA
1.4
1.2
1
0.8
0.6
0.4
0.2
0
ns
*
**
***
****
siOCT4 - + - - + + -
siSOX2 - - + - + - +
siNANOG - - - + - + +

B


siOCT4 - + - - + + -
siSOX2 - - + - + - +
siNANOG - - - + - + +
50 KDa
OCT4
37 KDa
*
SOX2
50 KDa
NANOG
50 KDa
β-ACTIN
1 2 3 4 5 6 7

C


siControl
siOCT4
siSOX2
siNANOG
400 µm

D

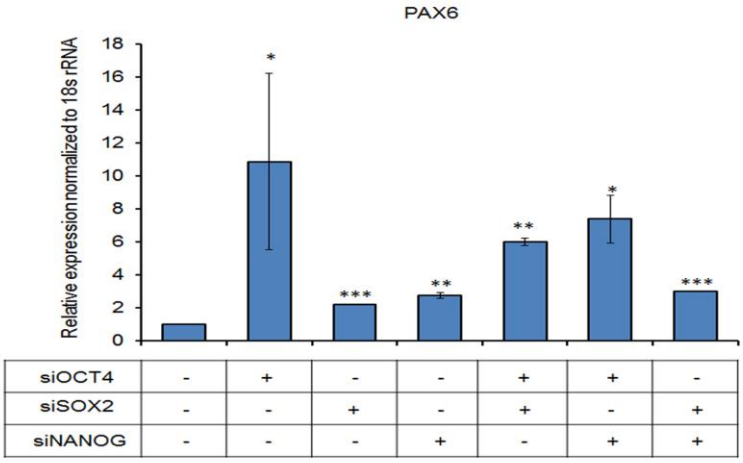

PAX6
Relative expression normalized to 18s rRNA
18
16
14
12
10
8
6
4
2
0
*
**
***
siOCT4 - + - - + + -
siSOX2 - - + - + - +
siNANOG - - - + - + +

Figure 2.5.7 Depletion of OCT4, SOX2, NANOG in EC cells does not cause upregulation of XIST.

EC cells were transfected with siRNA targeting OCT4, SOX2 or NANOG. 48 hour post transfection, cells were harvested and subjected to RNA and protein extraction as described in Section 2.4.7 and 2.4.9.Each bar represents values from three independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, *** P value < 0.001, ** P value < 0.01, * P value< 0.05, ns = non-significant).

(A) qRT-PCR for OCT4 (blue bar), SOX2 (red bar) and NANOG (green bar) using specific primers (see Appendix 1) indicate successful knockdowns as compared to siLuciferase control. Relative expression was calculated and normalized with respect to 18s rRNA levels as described in Section 2.4.10. X-axis indicates siRNA transfected into EC cells and Y-axis represents normalized fold change.

(B) Immunoblotting using antibodies specific for OCT4, SOX2 and NANOG confirms knockdowns at the protein level. β-ACTIN serves as loading control. siRNAs transfected into EC cells are indicated on the top and the antibodies used for probing the proteins are indicated on the side of the immunoblot image.

(C) Morphology of EC cells upon siRNA mediated knockdown of OCT4, SOX2, NANOG as compared to Control siRNA.

(D) qRT-PCR using PAX6 specific primers (see Appendix 1) indicate increase in its levels upon depletion of pluripotency factors as compared to siLuciferase control. Relative expression was calculated and normalized with respect to 18s rRNA levels. X-axis indicates siRNA transfected into EC cells and Y-axis represents normalized fold change.

2.5.8 Pluripotency factors and SATB proteins cross-regulate each other

The results discussed thus far suggest that regulation of *XIST* is under the control of multiple factors, similar to mouse *Xist*. To gain a comprehensive picture of *XIST* regulation, we set out to uncover additional factors that may be potentially involved in controlling *XIST* transcription.

It was already known that levels of Satb1 and Satb2 are important determinants of mESC functions such as self-renewal and differentiation (Savarese et al., 2009). Moreover, they have also been speculated to be involved in providing the initiation context for Xist-mediated silencing in mESCs and thymic lymphoma cells (Agrelo et al., 2009). Therefore, we sought to determine whether SATB1 and SATB2 are also involved in regulating the levels of pluripotency factors in EC cells before attempting to address their roles in regulating XIST levels. Towards this, we perturbed the levels of SATB1 or SATB2 in EC cells and scored for the expression of OCT4, SOX2 and NANOG by qRT-PCR. Overexpression and knockdown of SATB1/2 were validated by immunoblotting (Figure 2.5.8(A) & (B) respectively). Both these experiments yielded opposite results in terms of the expression levels of pluripotency factors. It was observed that while the expression of OCT4, SOX2 and NANOG decreased upon overexpression of SATB1 or SATB2 (Figure 2.5.8(A)), their levels were upregulated

upon depletion of SATB1 or SATB2 (Figure 2.5.8(B)). This suggests that the pluripotency factors are negatively regulated by SATB proteins.

Next, we asked if the expression of SATB1 and SATB2 are under the control of pluripotency factors in EC cells. To determine this, protein levels of SATB1 and SATB2 were assessed by immunoblotting upon individual or combinatorial depletion of OCT4, SOX2, NANOG in EC cells. Knockdowns of these factors were validated by qRT-PCR as well as immunoblotting as shown in Figure 2.5.7(A) & (B) respectively. It was observed that the levels of SATB1 increased marginally upon knockdown of NANOG and OCT4/NANOG together. And SATB2 decreased upon depletion of any of the pluripotency factors (Figure 2.5.8(C)).

Since the levels of the pluripotency factors decrease upon RA-induced differentiation of EC cells (Figure 2.5.1(B)), we also assessed the levels of SATB1 and SATB2 at transcript as well as protein levels upon differentiation of EC cells for 6 days. As evident from the Figure 2.5.8(D), levels of SATB1 steadily increase upon differentiation, while SATB2 levels initially follow a somewhat zig-zag pattern of expression. Based on these observations it can be stated that SATB1 is negatively regulated by the stem cell factors while SATB2 seems to be positively regulated. Collectively these results suggest that pluripotency factors and SATB proteins cross-regulate each other in EC cells. Therefore, it can be speculated that SATB proteins might play an important role in influencing the functions of EC cells similar to mESCs system.

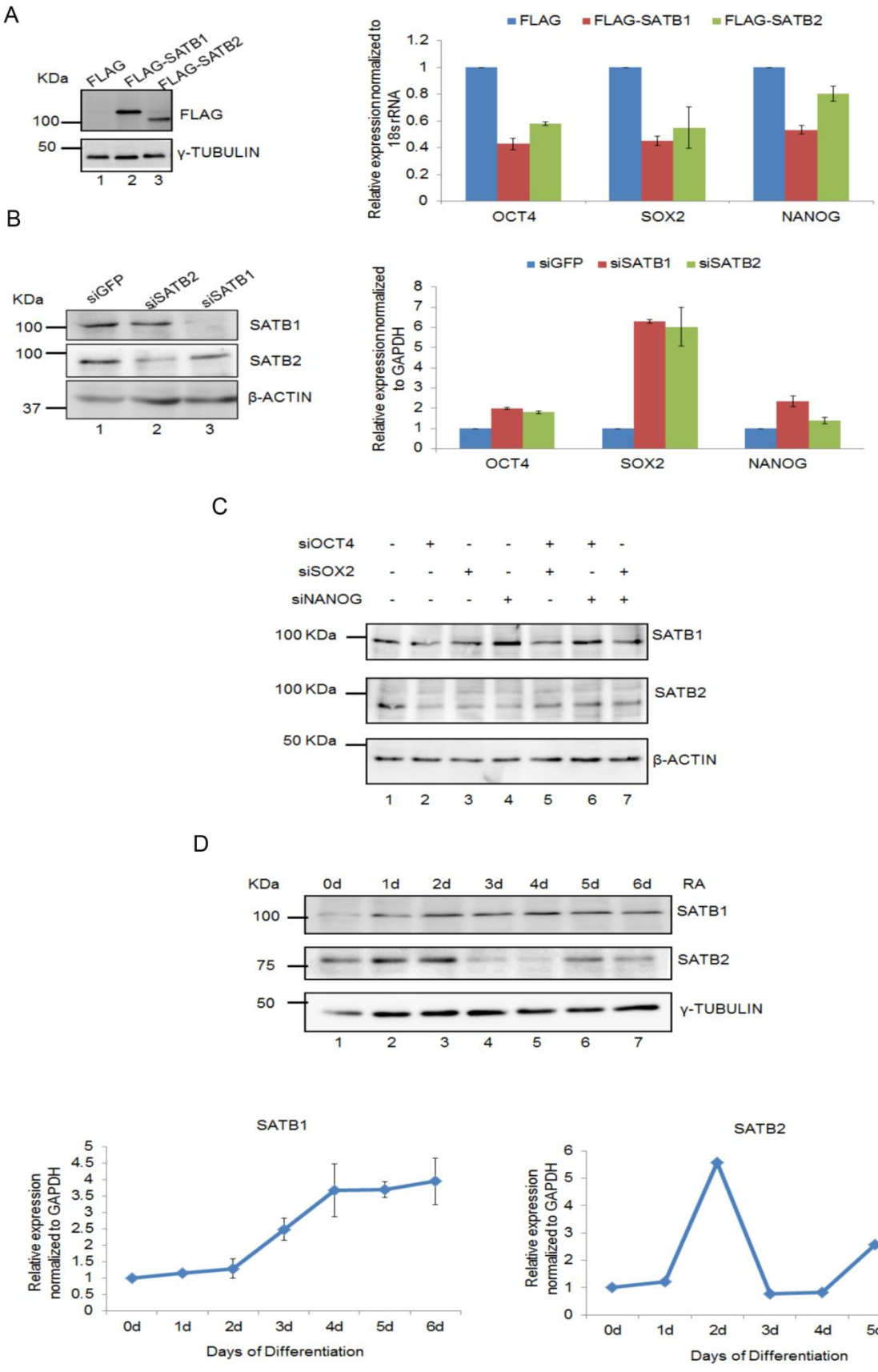

A
KDa
FLAG
FLAG-SATB1
FLAG-SATB2
100
FLAG
50
γ-TUBULIN
1 2 3
FLAG
FLAG-SATB1
FLAG-SATB2
Relative expression normalized to 18s rRNA
1.2
1
0.8
0.6
0.4
0.2
0
OCT4
SOX2
NANOG
B
KDa
siGFP
siSATB2
siSATB1
100
SATB1
100
SATB2
37
β-ACTIN
1 2 3
siGFP
siSATB1
siSATB2
Relative expression normalized to GAPDH
8
7
6
5
4
3
2
1
0
OCT4
SOX2
NANOG
C
siOCT4 - + - - + + -
siSOX2 - - + - + - +
siNANOG - - - + - + +
100 KDa
SATB1
100 KDa
SATB2
50 KDa
β-ACTIN
1 2 3 4 5 6 7
D
KDa
0d
1d
2d
3d
4d
5d
6d
RA
100
SATB1
75
SATB2
50
γ-TUBULIN
1 2 3 4 5 6 7
SATB1
Relative expression normalized to GAPDH
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0d 1d 2d 3d 4d 5d 6d
Days of Differentiation
SATB2
Relative expression normalized to GAPDH
6
5
4
3
2
1
0
0d 1d 2d 3d 4d 5d 6d
Days of Differentiation

Figure 2.5.8 Pluripotency factor and SATB proteins cross-regulate each other in EC cells.

(A) Overexpression of SATB1 and SATB2 by transfecting EC cells with FLAG-SATB1 or FLAG-SATB2 DNA constructs. Overexpression is validated by immunoblotting using FLAG antibody. γ-TUBULIN serves as a loading control. The DNA constructs transfected into EC cells are indicated on the top and the antibodies used for probing the proteins are indicated on the side of the immunoblot image.

qRT-PCR for OCT4, SOX2 and NANOG using gene-specific primers (see Appendix 1) upon transfecting EC cells with FLAG (Blue bar), FLAG-SATB1 (Red bar) or FLAG-SATB2 (Green bar). Levels of OCT4, SOX2 and NANOG decrease upon overexpression of SATB1 or SATB2 as compared to FLAG vector control. Ct values were normalized with respect to 18s rRNA. X-axis indicates the genes and Y-axis represents the normalized fold change.

(B) Knockdown of SATB1 and SATB2 by transfecting EC cells with siRNA targeting SATB1 or SATB2. Knockdowns were validated by immunoblotting using specific antibodies. β-ACTIN serves as a loading control. siRNAs transfected into EC cells are indicated on the top and the antibodies used for probing the proteins are indicated on the side of the immunoblot image.

qRT-PCR for OCT4, SOX2 and NANOG using gene-specific primers (see Appendix 1) upon transfecting EC cells with siGFP (Blue bar), siSATB1 (Red bar) or siSATB2 (Green bar). Levels of OCT4, SOX2 and NANOG increase upon knockdown of SATB1 or SATB2 as compared to siGFP control. Ct values were normalized with respect to GAPDH. X-axis indicates the genes and Y-axis represents the normalized fold change.

(C) Protein levels of SATB1 and SATB2 upon knockdown of OCT4/SOX2/NANOG in EC cells. Levels of SATB1 marginally increase upon depletion of NANOG and OCT4/NANOG and that of SATB2 decrease upon knockdown of any of OCT4/SOX2/NANOG as compared to siLuciferase control. β-ACTIN serves as a loading control. β-ACTIN blot shown here is the same as for Figure 2.5.7(B) since the western blotting images shown in Figure 2.5.7(B) and Figure 2.5.8(C) are part of the same experiment and performed simultaneously. siRNAs transfected into EC cells are indicated on the top and the antibodies used for probing the proteins are indicated on the side of the immunoblot image.

(D) Immunoblotting and qRT-PCR for SATB1 and SATB2 upon differentiation of EC cells for 6 days. Levels of SATB1 increase steadily upon differentiation. Expression of SATB2 increases at Day 2 of differentiation and then declines at Days 3 and 4 following which it again increases at Days 5 and 6.

2.5.9 SATB1 and SATB2 positively regulate XIST expression in EC cells

Following the leads from the studies carried out in mESCs and the results obtained in the earlier set of experiments described in 2.5.8, we next asked if SATB1 and SATB2 can regulate XIST expression. To address this, we monitored the endogenous levels of XIST by qRT-PCR as well as biochemically assessed the promoter activity of XIST1.1 DNA construct (101 bp promoter element) upon overexpression of SATB1 or SATB2 in EC cells. It was observed that levels of XIST increased 15-fold and 4-fold upon overexpression of SATB1 or SATB2 respectively (Figure 2.5.9(B)). It is known that the levels of OCT4, SOX2 and NANOG decrease upon overexpression of either SATB1 or SATB2 (Figure 2.5.8(A)). In addition, we also observed that the levels of SP1, YY1 and CTCF remained unaltered upon increasing the levels of SATB1/2 in EC cells (Figure 2.5.9(A)). With regards to the promoter activity, only a marginal increase was observed upon increasing the levels of SATB1 whereas SATB2 overexpression had no significant effect (Figure 2.5.9(C)). Since the promoter activity of XIST1.1 declines

progressively upon RA-induced differentiation for 3 days (Figure 2.5.4(D)), we also sought to determine if increasing SATB1 expression was able to rescue this effect. The experimental scheme is depicted in Figure 2.5.9 (D)) and it was observed that mere overexpression of SATB1 was not sufficient to rescue the reporter activity (Figure 2.5.9 (E)). On the basis of results shown in Figure 2.5.9 (C) and (E), we speculate that the promoter element tested (XIST1.1) does not harbor the sequence features that could be influenced by increasing SATB levels. Two possibilities, essentially not mutually exclusive, arise from the results described in this section:

(i) The positive regulation of endogenous XIST by SATB1/2 could be an indirect effect observed due to decrease in the levels of negative regulators of XIST, possibly the pluripotency factors and at the same time no alteration in the levels of characterized and putative positive regulators of *XIST* – YY1 and SP1 respectively.

(ii) SATB1 or SATB2 could also be playing a direct role in regulating *XIST* by binding to *XIST* genomic loci and modulating its epigenetic status.

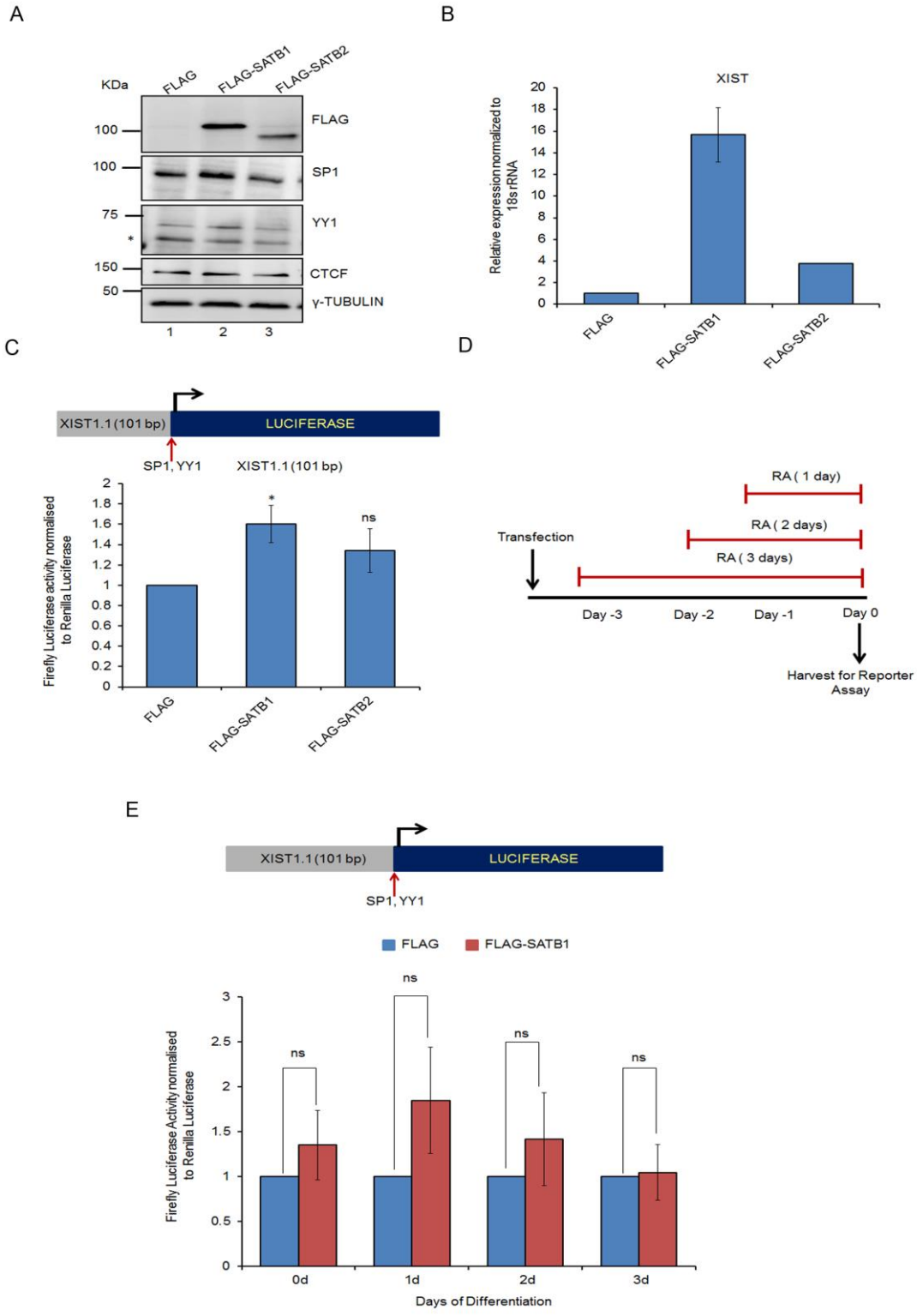


Figure 2.5.9 SATB1 and SATB2 positively regulate *XIST* expression in EC cells.

(A) Equal amounts of 3XFLAG CMV10, FLAG-SATB1 or FLAG-SATB2 were transfected into EC cells. Cells were harvested 48 hour post transfection and subjected to RNA and protein extraction.Immunoblotting using FLAG-antibody to validate SATB1/2 overexpression. Western blotting for SP1, YY1 and CTCF indicates that their levels do not alter significantly upon increased expression of SATB1 or SATB2 in EC cells for 48 hour. γ-TUBULIN

serves as the loading control. The DNA constructs transfected into EC cells are indicated on the top and the antibodies used for probing the proteins are indicated on the side of the immunoblot image. FLAG and γ-TUBULIN blot shown here is the same as for Figure 2.5.8(A) since the western blotting images shown in Figure 2.5.8(A) and Figure 2.5.9(B) are part of the same experiment and performed simultaneously.

(B) RNA was used for cDNA synthesis and qRT-PCR was done using XIST gene specific primers (see Appendix 1). Expression of endogenous XIST increase 15-fold and 4-fold respectively upon overexpression of SATB1 or SATB2 as compared to vector control. Ct values for XIST were normalized with respect to 18s rRNA. X-axis indicates the DNA constructs transfected into EC cells and Y-axis represents normalized fold-change.

(C) Schematic representing XIST1.1 promoter element with binding sites for SP1 and YY1 at the TSS. Firefly luciferase activities for XIST1.1 promoter fragment measured 48 hour post transfection with SATB1 or SATB2 overexpression DNA constructs in EC cells indicate 1.5-fold upregulation in the reporter activity upon SATB1 overexpression as compared to the vector control. No significant change in the activity was observed upon overexpression of SATB2. Firefly luciferase activities represented here are normalized with respect to Renilla luciferase activity which serves as an internal control. X-axis indicates the DNA constructs transfected into EC cells and Y-axis indicates the normalized fold change observed. Each bar represents values from three independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over FLAG vector control as per Student's T-test (* P value < 0.05, ns = non-significant).

(D) Experimental scheme for measuring XIST1.1 reporter activity upon overexpression of FALG-SATB1 in differentiating EC cells.

(E) Briefly, cells were transfected with XIST1.1, Renilla luciferase and FLAG or FLAG-SATB1 DNA constructs. 8h-post transfection, differentiation was induced by adding RA either for 1, 2 or 3 days. At the end of 3 days cells were harvested for the reporter assay. Luciferase activity from XIST1.1 promoter construct does not show any significant change upon overexpression of SATB1 compared to that of vector control. Firefly luciferase activities represented here are normalized with respect to Renilla luciferase activity which serves as an internal control. Each value is further normalized with respect to the corresponding value obtained for the vector control. X-axis indicates the days of differentiation and Y-axis indicates the normalized fold change observed. Each bar represents values from three independent experiments. Error bar represent the ±S.E.M. No significant difference (ns) was observed as per Student's T-test.

2.5.10 Chromatin organizer proteins SATB1, SATB2 and CTCF bind to two distinct on *XIST* Exon1

To assess if SATB proteins could directly bind *XIST* promoter or gene body, we made use of a freely available online tool FIMO (MEME) (Grant et al., 2011) to determine if *XIST* locus harboured a consensus SATB1 binding sequence (CSBS). The consensus for SATB1 has been identified from the previous studies carried out in Galande laboratory (Purbey et al., 2008). We used CSBS and *XIST* genomic region as queries in FIMO and it predicted several motifs for SATB1 in the region as shown in the UCSC genome browser track (Figure 2.5.10(A-a)) and summarised in the Table containing the significance values for each predicted site (Figure 2.5.10(A-b)). Also shown are OCT4 ChIP-seq peak from H1 ESCs as well as CTCF ChIP-seq peaks from H1 ESCs and GM12878 line. The same is summarised in a simplified schematic (Figure 2.5.10 (B)). Most of the predicted motifs were in the *XIST* exon1 region. From several identified motifs, we decided to validate the occupancy of SATB1 or SATB2 on two distinct sites which exhibit binding by OCT4 and CTCF respectively (Figure 2.5.10(A,B)).

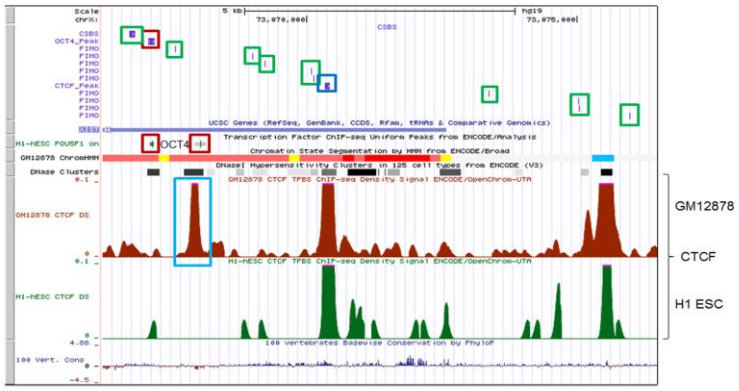
One of these sites (~5.5 Kb downstream of *XIST* TSS) was previously tested for OCT4, SOX2 and NANOG occupancies as shown in the Figure. 2.5.5(D). We performed ChIP analysis for SATB1 and SATB2 in differentiating EC cells (as carried out for 2.5.5) and determined their binding on this site. Like pluripotency factors, we found that SATB proteins also showed similar kinetics of binding to this particular site, with highest occupancies observed at Day 2 of differentiation (Figure 2.5.10(C)).

We also validated the binding of CTCF (as observed by ChIP-seq for CTCF from H1 cells in ENCODE data), SATB1 and SATB2 (based on CSBS predictions by FIMO) on *XIST* exon1 region, ~ 2.5 Kb downstream of *XIST* TSS by performing ChIP in differentiating EC cells. It was observed that the binding patterns exhibited by all the three proteins followed a similar profile (Figure 2.5.10(C), (D)). While CTCF and SATB2 showed a similar kinetics of binding, with maximum enrichment at Day 2 of differentiation, SATB1 displayed higher binding at day1 of differentiation. Interestingly, it was also observed that the pluripotency factor binding site on *XIST* exon1 (~ 5.5 Kb downstream of *XIST* TSS) was occupied by CTCF in GM12878 cells but not in H1 ESC (as per ENCODE) (shown by blue box in UCSC browser track – Figure 2.5.10 (A-a)). Moreover, this region is predicted to be a weak promoter in GM12878 as per ENCODE chromatin segmentation states (Figure 2.5.10 (B)). Therefore, we speculated that in the absence of pluripotency factors, CTCF can bind to this site and possibly positively regulate XIST expression. We tested this by performing ChIP for CTCF in differentiating EC cells and determining its occupancy on this particular site. The dynamics of enrichment obtained resembled to that observed for pluripotency proteins and SATB proteins, with maximum occupancies at day 2 of differentiation. Occupancy of CTCF decreases at day 3 but shows a slight increment (~1.5-fold) again at day 4. Upon careful comparison of the results obtained in Figure 2.5.5 (D) and Figure 2.5.10 (D), it was observed that a similar pattern was observed for OCT4 and SOX2, whereas NANOG exhibited the lowest occupancy at day 4. Interestingly, the sites occupied by CTCF on *XIST* exon1 as per ENCODE as well as in our study are the regions close to the predicted weak enhancer elements (yellow box in the ChromHMM state) according to chromatin segmentation states. And it is evident from Figure 2.5.2 that XIST is upregulated from day 3 of differentiation onwards. All these results put together provide possible clues towards the dynamic and complex regulatory

mechanism governed by multiple factors in controlling transcription from *XIST* promoter in EC cells in a temporal manner.

A

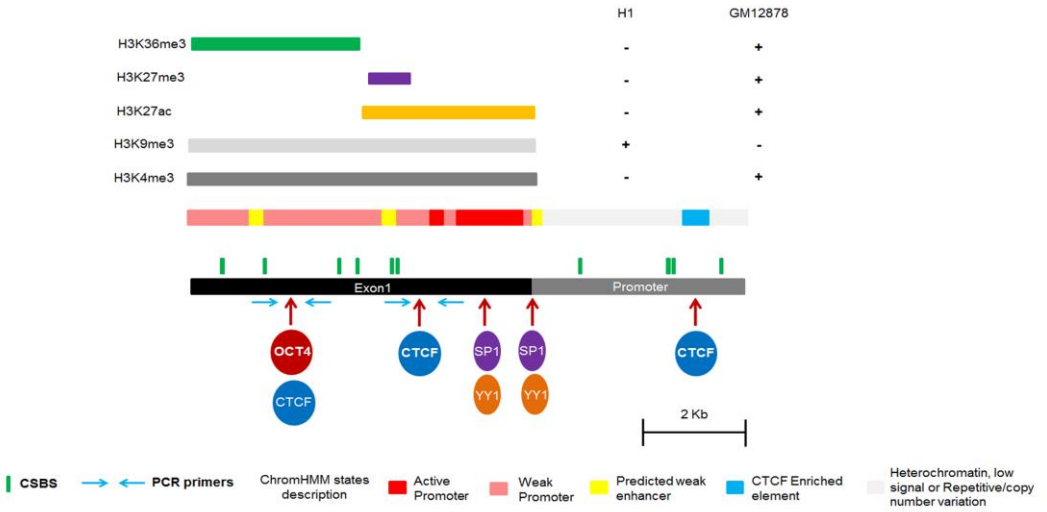
a



b

Motif	Sequence Name	Strand	Start	End	p-value	q-value	Matched Sequence
1	hg19_dna	+	8909	8920	2.05e-05	0.202	TATTTGTAATAT
1	hg19_dna	+	3114	3125	0.000341	0.85	TATTGGCAATTA
1	hg19_dna	+	4021	4032	0.000341	0.85	AATTTGTAATGA
1	hg19_dna	+	1461	1472	0.000521	0.85	AATTAGAAATAC
1	hg19_dna	+	2900	2911	0.000521	0.85	AATTAGCAATTA
1	hg19_dna	+	3971	3982	0.000521	0.85	CATTAGAAATGA
1	hg19_dna	+	8920	8931	0.000604	0.85	TATTTTTAATCA
1	hg19_dna	+	9858	9869	0.000715	0.88	TAGTTCTAATAA
1	hg19_dna	+	7254	7265	0.000853	0.934	TGTTTGTAATTT

B



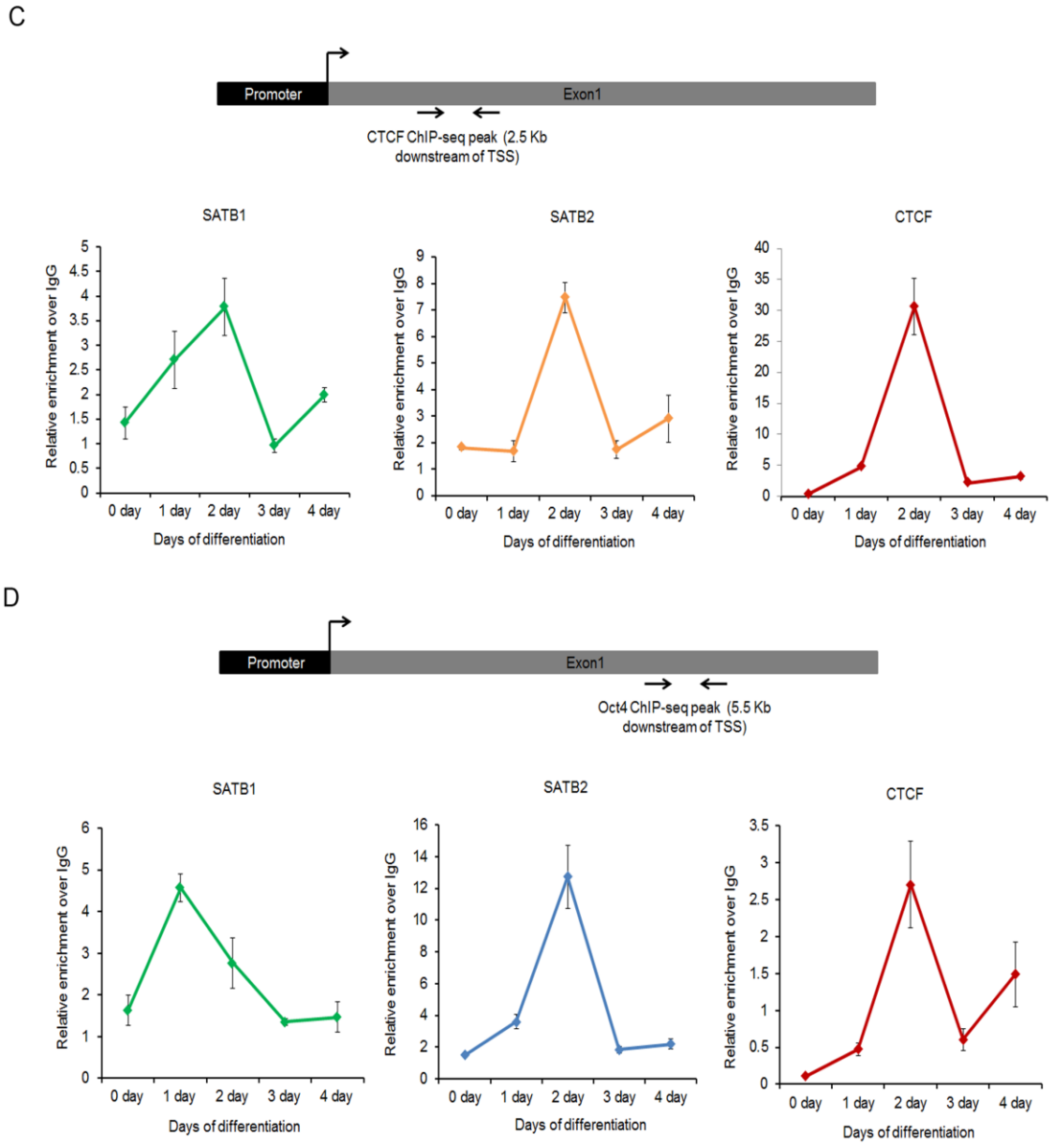


Figure 2.5.10 SATB1, SATB2 and CTCF occupy two distinct sites on *XIST* Exon1.

(A) (a) UCSC genome browser track for *XIST* genomic locus (chrX:73,067,000-73,078,000) displaying predicted CSBS (green rectangles), OCT4 ChIP-seq peak from H1 ES cells (red rectangle) and CTCF ChIP-seq peak from H1 ES cells (blue rectangle). Also shown are DNase hypersensitivity clusters from 125 ENCODE cell lines and the ChromHMM state from GM12878 cell line. OCT4 ChIP-seq peak from H1 ESC and CTCF ChIP-seq peaks from H1 ESC and GM12878 cell line are also displayed. (b) Table depicting the location of CSBS and the significance score as predicted by FIMO.

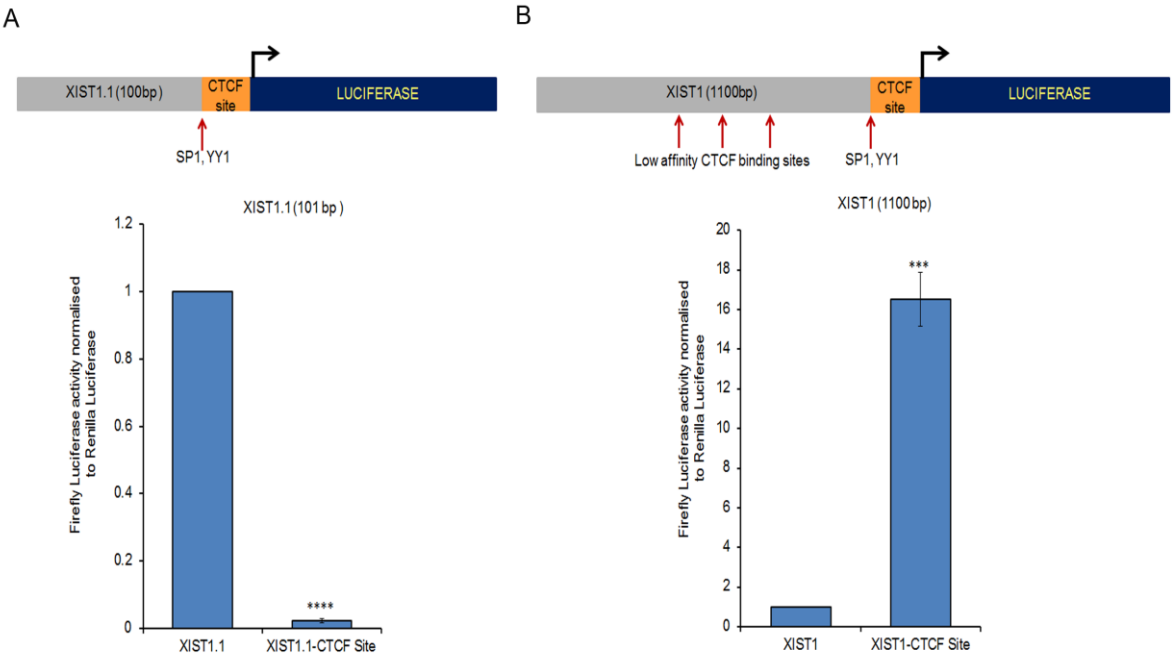
(B) Simplified illustration showing CSBS (green), OCT4 and CTCF binding sites, ChromHMM state from GM12878, comparison of histone modification profiles between H1 and GM12878 and the primers used for PCR post ChIP experiment (blue arrows) (see Appendix 3 for primer sequence).

(C) qRT-PCR for the region indicated in *XIST* genomic region schematic for the DNA recovered post ChIP using antibodies specific for SATB1 and SATB2 at Day 0, 1, 2, 3 and 4 of differentiation. Ct values for ChIP were normalized to that for IgG control and plotted using the formula described in the "Material and Methods Section". X-axis represents Days of differentiation and Y-axis represents normalized fold change for ChIP.

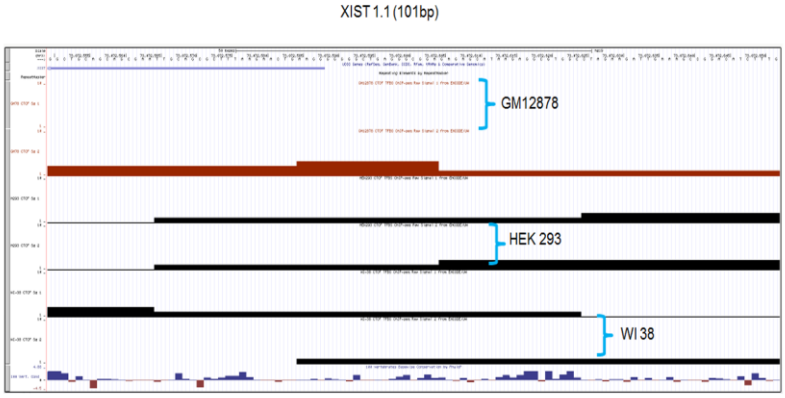
(D) qRT-PCR for the region indicated in *XIST* genomic region schematic for the DNA recovered post ChIP using antibodies specific for CTCF at Day 0, 1, 2, 3 and 4 of differentiation. Ct values for ChIP were normalized to that for IgG control and plotted using the formula described in the 'Material and Methods' section 2.4.10. X-axis represents Days of differentiation and Y-axis represents normalized fold change for ChIP.

2.5.11 *XIST* Exon1 site with SATB1, SATB2 and CTCF binding might be involved in positive regulation of *XIST*

The significance of the site on *XIST* Exon1 on which the binding of pluripotency factors and SATB proteins was demonstrated has been tested earlier (Figure 2.5.6(B)). It was observed that this site which is located ~ 5.5 Kb downstream of *XIST* TSS may be the negative regulator of *XIST*. We next wanted to test whether the other site on *XIST* Exon1, located closer to the TSS (~2.5 Kb downstream of *XIST* TSS) and identified to be bound by SATB1, SATB2 and CTCF (Figure 2.5.10(C)) has any regulatory potential. Towards this, we cloned the region which exhibited binding to these factors into two of the promoter constructs – XIST1.1 (101 bp region) and XIST 1 (1100 bp region), between the promoter element and luciferase gene as shown in the schematic (Figure 2.5.11(A)) to recapitulate the genomic context. This region is labelled as CTCF site. Next we transfected XIST1.1 (101 bp) and XIST1 (1100 bp) constructs with and without the CTCF binding region into EC cells along with Renilla luciferase as an internal control. The cells were harvested for testing the promoter activity biochemically 48 hour post transfection. It was observed that the presence of the region to which CTCF, SATB1 and SATB2 could potentially bind led to a significant decrease in the promoter activity when fused with XIST1.1 DNA construct (Figure 2.5.11(A)). However, the same region increased the promoter activity manifold when put together with XIST1 DNA construct (Figure 2.5.11(B)). Upon careful examination of XIST1 region, we observed less prominent but multiple CTCF ChIP-seq peaks from many ENCODE female cell lines (GM12878, HEK 293 and primary fibroblast WI38) indicated by green arrow (Figure 2.5.11(C)). Therefore, it can be speculated that the presence of these additional binding sites along with the prominent CTCF binding site on *XIST* Exon1 can have a positive impact on *XIST* regulation. However, further experiments such as mutations or deletions of few or all of these site are needed to attribute functional characteristics to these different modes of binding.


A
XIST1.1(100bp)
CTCF site
LUCIFERASE
SP1, YY1
XIST1.1(101 bp)
Firefly Luciferase activity normalised to Renilla Luciferase
1.2
1
0.8
0.6
0.4
0.2
0
****
XIST1.1
XIST1.1-CTCF Site
B
XIST1(1100bp)
CTCF site
LUCIFERASE
Low affinity CTCF binding sites
SP1, YY1
XIST1(1100bp)
Firefly Luciferase activity normalised to Renilla Luciferase
20
18
16
14
12
10
8
6
4
2
0
***
XIST1
XIST1-CTCF Site

C


XIST1.1(101bp)
GM12878
HEK 293
WI38

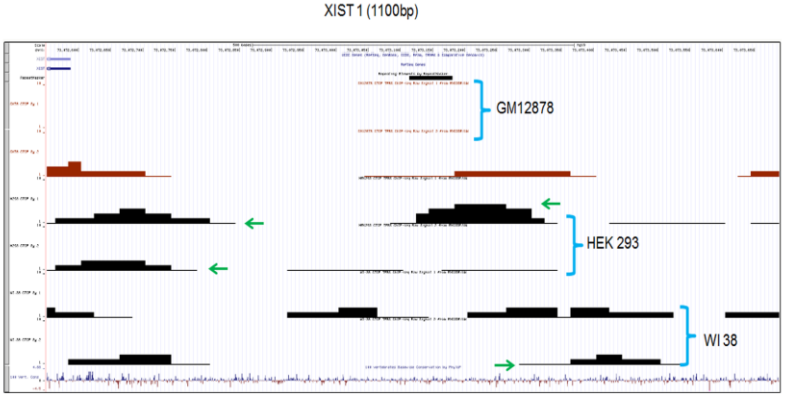

XIST1(1100bp)
GM12878
HEK 293
WI38

Figure 2.5.11 SATB1, SATB2 and CTCF binding to *XIST* exon1 (2.5 Kb downstream of TSS) might be positively regulating *XIST* transcription.

(A) Schematic showing the CTCF binding site present ~ -2.5 Kb downstream of *XIST* TSS into the exon1 cloned between XIST1.1 promoter (101 bp region) and luciferase gene. Luciferase reporter activities of XIST1.1-CTCF DNA construct measured 48 hour post transfection in EC cells. The fold change is calculated over XIST1.1 construct without the CTCF binding site. Firefly luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis indicates the DNA constructs transfected into EC cells and Y-axis indicates the normalized fold change observed. Each bar represents values from three independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over XIST1.1 control as per Student's T-test (****P value < 0.0001).

(B) Schematic showing the CTCF binding site present ~ -2.5 Kb downstream of *XIST* TSS into the exon 1 cloned between XIST1 promoter (1100 bp region) and luciferase gene. Also shown are the low affinity binding sites of CTCF as per ENCODE ChIP-seq peaks for CTCF. Luciferase reporter activities of XIST1-CTCF DNA construct measured 48 hour post transfection in EC cells. The fold change is calculated over XIST1 construct without the CTCF binding site. Firefly luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis indicates the DNA constructs transfected into EC cells and Y-axis indicates the normalized fold change observed. Each bar represents values from three independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over XIST1.1 or XIST1 control as per Student's T-test (***P value < 0.001).

(C) UCSC genome browser tracks for XIST1.1 (101 bp) and XIST1 (1100 bp) promoter regions. XIST1.1 shows no CTCF ChIP-seq peaks, whereas XIST1 displays multiple low peaks in GM12878, HEK 293 andprimary fibroblast line WI38. These peaks are indicated by green arrows.

2.6 Discussion

In this study we sought to determine the regulation of *XIST* expression in the context of initiation phase of XCI. Research over the past two decades has been instrumental in elaborating our understanding of the subject of XCI. Extensive work carried out in mouse models has generated a vast knowledge not only pertaining to the molecular details of the process at large but also unraveled certain fundamental features of chromosomal organization and epigenetic regulation. Based on these studies it can be convincingly stated that there are multiple mechanisms in place to tightly control the expression of pioneering molecule of XCI – Xist (reviewed in Augui et al., 2011; Deuve and Avner, 2011; Goodrich et al., 2016). Since Xist is central to the process of XCI, it is absolutely essential to timely modulate its levels so as to ensure faithful execution of the process and hence the developmental programs. Despite significant conservation of X-linked genes, chromosomal synteny as well as the process of XCI guided by XIST lncRNA between several eutherian mammals, XCI is initiated in diverse ways in different species (Okamoto et al., 2011). One of the possible explanations could be evolutionary alterations in the developmental programs across species necessitating robustness in the regulation of XCI to accommodate these changes (Okamoto et al., 2011). In addition to this, lncRNA called Tsix that is known to be a

critical negative regulator of *Xist* in mouse is not functionally conserved across species. TSIX RNA in humans is distinct from its murine counterpart in sequence. It is also truncated (overlaps only upto 5-8 exons of *XIST*) and hence does not extend into *XIST* promoter to be able to alter its epigenetic status. Besides, it is transcribed even from the inactive X in human foetal cells (Migeon et al., 2001, 2002). Therefore, it can be said that *XIST* in humans may be regulated differently and so addressing this question is of utmost importance in order to improve our understanding of the phenomenon of XCI in human.

2.6.1 Identification of *XIST* promoter

There have been only a few studies looking into this aspect. The pioneering work was carried out by Hendrich et al. in 1997, where they identified and made an attempt to characterize human *XIST* promoter elements by employing biochemical tools such as luciferase reporter assays, saturated site directed mutagenesis and EMSA to decipher the potential transcription factors. However, all their experiments were performed either in HEK 293 cells (female) or HT1080 fibrosarcoma cells (male), which does not provide the context for the initiation phase of XCI. It is fairly well known from the several studies carried out in mouse system that temporal regulation of *Xist* during the initiation phase is very crucial and is brought about by multiple players. Therefore, there arises a need to address this aspect of *XIST* regulation in the context of initiation phase of XCI in human.

Our study attempts to provide clues with regards to the aforementioned question. To comprehend this, we chose human EC cells as our model system which has been used as a proxy for understanding the biology of human ES cells in the past) (Andrews et al., 1980; Andrews et al., 1984; Lee and Andrews, 1986; Skotheim et al., 2005). We demonstrate that these cells respond to RA-mediated differentiation paradigm as reported earlier (Andrews et al., 1980; Andrews et al., 1984) and show for the first time that the low expression of XIST (Chow et al., 2003) is upregulated over the course of differentiation. Similar to the first report on human *XIST* promoter characterization (Hendrich et al., 1997), we also observed that the region conserved across several eutherian species, encompassing 101 bp including a fraction of *XIST* 5' and the region upstream of the TSS is sufficient to drive the transcription and can be called the minimal promoter of *XIST*. Based on these findings, it can be concluded that the identified minimal promoter element (101 bp) behaves in a similar manner in the cells that provide the context for initiation (EC) (our study) as well as in a cell line that

provides the context for the maintenance phase (HEK293) (Hendrich et al., 1997) (Chapter 3 of current study) as assessed by biochemical assays.

2.6.2 Identification of transcription factors governing transcription from *XIST* promoter

The immediate question that follows is the identity of transcription factors regulating *XIST* promoter. The study by Hendrich et al. (1997) identified SP1, YY1 and TBP to be the potential transcription factors by performing saturated site-directed mutagenesis and EMSA. In the post human genome and ENCODE era, there is a bulk of information available on the binding sites and motifs for many more transcription factors as well as genome wide epigenetic signatures for 125 different human cell lines including male hESC (H1). In order to verify whether the factors identified in the mentioned study actually bind *XIST* promoter, we made use of publicly available ChIP-sequencing data from ENCODE. We observed that in addition to the binding sites identified in the previous study, SP1 and YY1 also showed occupancy ~1.4 Kb downstream of *XIST* TSS, in the exon1. In addition to this, we also observed peaks for CTCF in various cell lines (H1, GM12878, HEK293, BJ, WI38) at ~ 2.5 Kb upstream as well as downstream of *XIST* TSS. Out of these three factors, SP1 and YY1 had already been demonstrated to bind the minimal promoter of *XIST* (Hendrich et al., 1997). In addition to its other roles in the process of XCI in mouse, CTCF was recently shown to act as a repressor of *Xist* promoter in mESCs (Sun et al., 2013). However, we still do not completely understand if it is involved in repressing human *XIST* expression in a similar manner.

In the system of our choice – EC cells, the expression of XIST increases linearly upon subjecting the cells to the differentiation cues by administering RA. In order to understand whether the three zinc-finger transcription regulators (SP1, YY1 and CTCF) have any role(s) in governing the expression from *XIST* promoter, we analyzed their expression patterns upon RA-mediated differentiation of EC cells. Surprisingly, we observed that all three of them are actually downregulated during differentiation, which anti-correlates with XIST expression profile. Two possibilities, though not mutually exclusive, stem from these findings:

(1) Firstly, it is possible that SP1/ YY1/CTCF can exhibit temporal enrichment on the *XIST* promoter during differentiation despite a sharp decline in their expression levels. This is an

important question which can be addressed by performing ChIP for these factors in differentiation dependent manner.

(2) Secondly, there may be additional players involved in the process which can influence the transcriptional outcome from the promoter under investigation.

In order to reconcile our observations, we further assessed the activities of XIST1.1 (contains binding site for SP1 and YY1), XIST1 (contains binding site for SP1, YY1 and weak affinity CTCF sites) and XIST1-3 (contains binding site for SP1, YY1 and weak as well as strong affinity CTCF sites) promoter constructs in response to differentiation cues and observed a significant decline in the manner similar to that observed for SP1, YY1 and CTCF expressions. Collectively, these results not only indicated that *XIST* promoter might be regulated by these three zinc-finger containing transcription factors – SP1, YY1 and CTCF but also provided clues towards the involvement of additional players.

While we were pursuing our current study, YY1 was identified to be the key factor controlling transcription from *XIST* promoters by two independent research groups (Chapman et al., 2014; Makhlouf et al., 2014). One of the studies demonstrated YY1 as the activator of*Xist*/*XIST* promoter in mouse and human ES cells as well as fibroblast cells (Makhlouf et al., 2014). The latter study identified a second promoter of *XIST*, albeit weaker to the first one, located 1.4 Kb downstream of the TSS and also regulated by YY1 (Chapman et al., 2014). The authors also provided the first ever evidence of antisense transcript initiating within the first exon of *XIST* close to TSS and extending upto 5' end of *XIST*. However, unlike the reports in mouse system (Sarkar, Gayen and Kumar et al., 2015) this antisense transcript is expressed at levels 0.001% of sense XIST transcript. Therefore, it's role in regulating the levels of XIST is improbable. The former study also performed comparative in silico analysis of *XIST* promoter sequences from seven species of eutherian mammals and observed the presence of conserved YY1 clusters located close to the TSS. Interestingly, the authors also highlighted that the occurrence of conserved CTCF motif present downstream of TSS and in close vicinity to the YY1 cluster was observed only in the case of rodents (mouse and rat) and absent from other mammals tested in their study. But the motif located upstream of the *XIST* TSS is more or less conserved with an exception of one species (rabbit) (Makhlouf et al., 2014). However, CTCF ChIP-sequencing profiles generated using multiple

cell lines by the ENCODE consortium seems to suggest otherwise atleast in the case of humans. Upon examination of these data sets, we observed multiple CTCF peaks even in the exon1 of *XIST* gene. The significance of these observations is stated in the later parts of the discussion section. The findings so far from our experiments indicated that the elements other than the promoter could be playing a role in regulating *XIST* in EC cells.

2.6.3 Role of pluripotency factors in regulating *XIST* expression in EC cells

In order to gain significant insights into regulation of *XIST*, we next asked if the pluripotency proteins could be a part of this regulation since increment in XIST levels in EC cells correlates with the downregulation of OCT4, SOX2 and NANOG. Besides, there are also a number of evidences linking Xist upregulation to downregulation of pluripotency factors in mouse system (reviewed in Deuve and Avner, 2011; Donohoe et al., 2009; Navarro et al., 2008; Navarro et al., 2010). Upon making use of OCT4, SOX2 and NANOG ChIP-sequencing data sets from male hESC H1 generated by ENCODE, we observed that OCT4 displayed binding on the exon 1 region, ~ 5.5 Kb downstream of *XIST* TSS. Moreover, this region was identified as weak promoter as per ChromHMM profile (ENCODE). We also observed that it showed enrichment of repressive histone marks such as H3K9me3 in H1 hESC whereas the same site was enriched for active histone mark H3K4me3 in a female lymphoblastoid cells (GM12878). Since XIST is not made in H1 cells, we thought that OCT4 binding may be playing a role in inhibiting *XIST* transcription by modulating the epigenetic status of the gene locus. Firstly, we validated OCT4 occupancy on this site and also demonstrated that even SOX2 and NANOG bind to this region. Interestingly, all these proteins display a peculiar pattern of enrichment, with the occupancies peaking at the 2nd day of RA-induced differentiation of EC cells and immediate decline observed thereafter. This correlates with our finding that XIST expression is the lowest at the 2nd day of differentiation. It would be interesting to address the functional and mechanistic significance of this distinct binding profile. Whether it is a consequence of specific post-translational modifications on the factors tested or a result of co-recruitment or an outcome of competition for the binding site remain open questions. Moreover, it would also be interesting to determine the implication of this binding profile in terms of governing the establishment of promoter specific histone modifications and hence regulating *XIST* expression.

To determine the significance of the results from our ChIP experiments, we introduced this pluripotency factor binding (PFB) site between *XIST* promoter (XIST1.1 or XIST1) and the reporter gene (firefly luciferase) to closely match the genomic context and assessed the promoter activities biochemically in EC cells. The results from this experiment suggested that PFB site might actually be acting as a negative regulator of transcription from the *XIST* promoter. However, it can be appreciated that despite having similar promoter firing activity, the decrease in the promoter activity observed in the case of XIST1 (1100 bp) construct upon introduction of PFB site was less dramatic compared to XIST1.1 (101) element. The significance of the stated observation is not completely understood but we do attempt to make a speculation about the same towards the end.

Furthermore, in order to better understand the roles of pluripotency factors as possible repressors of *XIST*, we chose to use a cell line which does not express these factors but makes significant amounts of *XIST* and harbours maintenance phase Xi. Towards this we overexpressed either OCT4 or SOX2 in HEK 293T cells and observed that the levels of spliced as well as unspliced XIST are significantly downregulated upon overexpression of SOX2 in these cells. Importantly, the expression of SP1 and CTCF remained unaffected whereas there was a marginal increase in YY1 levels. These results put together suggest that the pluripotency factors could potentially be involved in repressing XIST during the initiation phase of XCI. However, this speculation was not supported well by another experiment involving depletion of OCT4/SOX2/NANOG in EC cells and scoring for the expression of XIST. If these factors are indeed the negative regulators of XIST, then their depletion should have led to upregulation of XIST, which was not found to be the case in our study. Nevertheless, it is possible that perturbing the levels of pluripotency factors may in turn have affected certain yet unidentified regulators of *XIST*. Nonetheless, it can be realized that like regulation of any other gene, the balance of activators and repressors is a crucial determinant of the fate of *XIST* transcription.

2.6.4 Role of chromatin organizer proteins – SATB1, SATB2 and CTCF in regulating *XIST* expression in EC cells

In our quest to decipher the complex regulation of *XIST*, we set out to uncover additional players involved. It was already demonstrated that the nuclear matrix binding proteins Satb1 and Satb2 have an impact on the self-renewal and differentiation properties of mESCs

(Savarese et al., 2009). Also, they were suggested to be important for providing the initiation context for XCI in mESC (Agrelo et al., 2009). The latter finding however is debatable since another study came up with contradictory results with regards to initiation of XCI when using $Satb1^{-/-}Satb2^{-/-}$ mESCs and argued that Satb proteins are dispensable for XCI (Nechanitzky et al., 2012). This is not the first case of discrepancy observed between *in vivo* and *ex vivo* results in the field of XCI. For instance, there have been reports proving otherwise for the involvement of Ftx, pluripotency factors, DNA methyltransferases, Xist and Tsix themselves in the process of XCI (Kalantry et al., 2009; Sado et al., 2004; Shin et al., 2014; Soma et al., 2014). Therefore, it suffices to say that the process of XCI is executed in a very robust manner since it is extremely essential for the survival of the embryo. Hence, we went ahead with addressing the roles of SATB proteins in regulating *XIST* as well as pluripotency factors in EC cells.

Our results suggest that SATB proteins positively regulate *XIST* in these cells. Moreover, we also observed that there is cross-regulation between SATB1/2 and the pluripotency proteins. Elaborating on this, our results show that OCT4, SOX2 and NANOG are negatively regulated by SATB1/2. SATB1 itself seems to be repressed by pluripotency proteins (OCT4 and NANOG) but SATB2 appears to be regulated in an opposite manner by the stem cell factors as determined by measuring their expression either upon siRNA mediated depletion of OCT4/SOX2/NANOG or RA-induced differentiation. Interestingly, we observed that although overexpression of SATB1 led to upregulation of XIST, it was not sufficient to rescue the decline in the *XIST* minimal promoter activity during differentiation. On the basis of these results, we speculate that the promoter element tested (XIST1.1) does not harbor the sequence features that could be influenced by increasing SATB levels. Two possibilities, essentially not mutually exclusive, arise from the findings described:

(i) The positive regulation of endogenous XIST by SATB1/2 could be an indirect effect observed due to decrease in the levels of negative regulators of *XIST* (possibly OCT4, SOX2 and NANOG) and at the same time no alteration in the levels of positive regulators of *XIST* (YY1 and SP1). However, whether perturbing the levels of SATB1/2 in EC cells has any effect on the occupancies of any of the said positive and negative regulators on *XIST* genomic locus remains to be tested.

(ii) SATB1 or SATB2 could also be playing a direct role in regulating *XIST* by binding to *XIST* genomic locus and modulating its epigenetic status.

In order to explore the 2nd possibility, we checked for the occurrence of consensus SATB1 binding sequence (CSBS), characterized earlier by Galande group (Purbey et al., 2008) on the *XIST* genomic locus. Towards this, we made use of an *in silico* tool – FIMO that scans the given sequence for the presence of a motif provided and obtained multiple CSBS over the *XIST* locus (including promoter and exon1) (Bailey et al., 2009). Two of these high-confidence predicted binding sites overlapped with OCT4 ChIP-sequencing peak (H1 cells) and CTCF ChIP-sequencing peak (multiple cell lines), located ~ 5.5 Kb downstream and ~ 2.5 Kb downstream of *XIST* TSS on the exon 1 respectively. Therefore, we verified if SATB1 and SATB2 could indeed bind to these two sites and observed that they exhibited temporal kinetics of enrichment, with highest occupancies observed at the start of differentiation (Day1 or Day2), followed by a sharp depression upon receiving prolonged differentiation cues (Day3 and Day4). We also observed a similar pattern of enrichment profile for another well characterized chromatin organizer protein – CTCF, at ~ 2.5 Kb and ~ 5.5 downstream of *XIST* TSS into the 1st exon. It is noteworthy that while occupancy by CTCF at the site closer to the TSS (~ 2.5 Kb downstream) was observed in both H1 as well as GM12878 cell lines (as per ENCODE), the 2nd site (~ 5.5 Kb downstream) was found to be occupied by it only in GM12878. Interestingly, this particular site exhibited higher enrichment for the pluripotency factors (OCT4, SOX2, NANOG) compared to CTCF. This raises the possibility of competition between the proteins highlighted in our study to bind to this site. In addition to this, both these sites are predicted to be weak enhancer element by ChromHMM (ENCODE). Overall the observations hint towards the prospect of differential occupancy by CTCF which may be instrumental in the distinctive outcome from the *XIST* promoter in the two cell lines. It is important to mention that none of the ChIP-sequencing information stated or the ChIP performed by us is allele-specific. Hence, it cannot be said with certainty whether the proteins tested in our study actually bind to the *XIST* genomic locus on (future) Xa or (future) Xi.

The significance of the region shown to be occupied by OCT4, SOX2, NANOG, CTCF, SATB1 and SATB2 (~ 5.5 Kb downstream of TSS) was previously described by incorporating it in between the promoter element and the reporter gene. In order to gain insights into the

role played by the second binding site (2.5 Kb downstream of *XIST* TSS) which was shown to be occupied by the three chromatin organizer proteins – SATB1, SATB2 and CTCF and also displayed active histone modifications (as per ENCODE), we introduced this site between the *XIST* promoter element and the reporter gene on the plasmid. Interestingly, we observed that while this particular site led to a significant reduction in the activity of the minimal promoter (XIST1.1), the same site along with 1.1 Kb long promoter fragment (XIST1) caused a significant increase in the promoter activity. It is noteworthy that there are no inherent differences in the activities of just the minimal promoter (101 bp) and a longer promoter fragment (1.1 Kb) otherwise. Upon closer examination of the CTCF ChIP-sequencing peaks from the ENCODE consortium, we observed multiple small peaks in the promoter as well as exon 1 of *XIST*. We call these sites as the low affinity sites and the site verified by ChIP as the high affinity site, solely on the basis of the heights of the peaks obtained by ChIP-sequencing. The minimal promoter tested does not harbour any of these low affinity binding sites but the extended promoter fragment (1.1Kb) contains a number of such sites. Therefore, we think that a strong CTCF binding site when present alongside these low affinity sites changes the transcriptional outcome from negative to positive.

On the basis of this preliminary result, we hypothesize that binding of CTCF both upstream as well as downstream of *XIST* TSS plays an important role in governing the transcription from *XIST* promoter. Results obtained in our study with respect to the significance of the two CTCF binding sites (~5.5 Kb and ~2.5 Kb downstream of TSS) extend support to this hypothesis. Interestingly, the region on *XIST* exon1 (~+2.5 Kb from TSS) as well as upstream of the TSS (~ -2.5 Kb from TSS) with strong CTCF binding site also shows occupancy of RAD21 (RAD21 ChIP-seq from ENCODE and supplementary information from Makhlouf et al., 2014). Both CTCF and RAD21 have been long known to be involved in chromatin looping interactions (Splinter et al., 2006). Also, the studies carried out in our lab and elsewhere have convincingly demonstrated SATB1 to be a mediator of long range chromatin associations and the determinant of the 3D chromatin loopscape (Cai et al., 2006; Galande et al., 2007; Kumar et al., 2007). Therefore, it can be speculated that these factors may possibly be involved in a looping interaction between the promoter and gene body of *XIST*, aiding in the enrichment of specific histone marks and hence enabling the transcriptional control from the promoter. This raises questions over previous study which

ruled out the contribution by CTCF in regulating *XIST* promoter based on their ChIP experiment which showed more or less equal enrichment of CTCF in male and female cell lines or XIST+ and XIST- hESC (Makhlouf et al., 2014). We think that although CTCF occupancies on the *XIST* genomic locus are comparable between male and female cells, the possibility of differential looping mediated interaction cannot be ruled out solely on the basis of this finding. However, our hypothesis is merely a conjecture as of now and needs to be verified by making use of chromosome conformation capture (3C) technique supported by either site-directed mutagenesis or deletion using CRISPR-Cas9 to disrupt CTCF binding to any or all of these sites.

In conclusion, the results presented in this chapter provide compelling evidence towards the involvement of multiple factors in regulating *XIST* in the context of initiation phase of XCI. Based on the results presented, it can be appreciated that the identified minimal promoter of *XIST* does not actually have the complete regulatory potential. And interestingly, the regions present in the gene body of *XIST* (exon1) can be important determinants of transcription from the promoter. Based on our work, we propose a model of complex and dynamic regulatory interplay between the potential activators and repressors of *XIST* gene as summarized in the Figure 2.6.

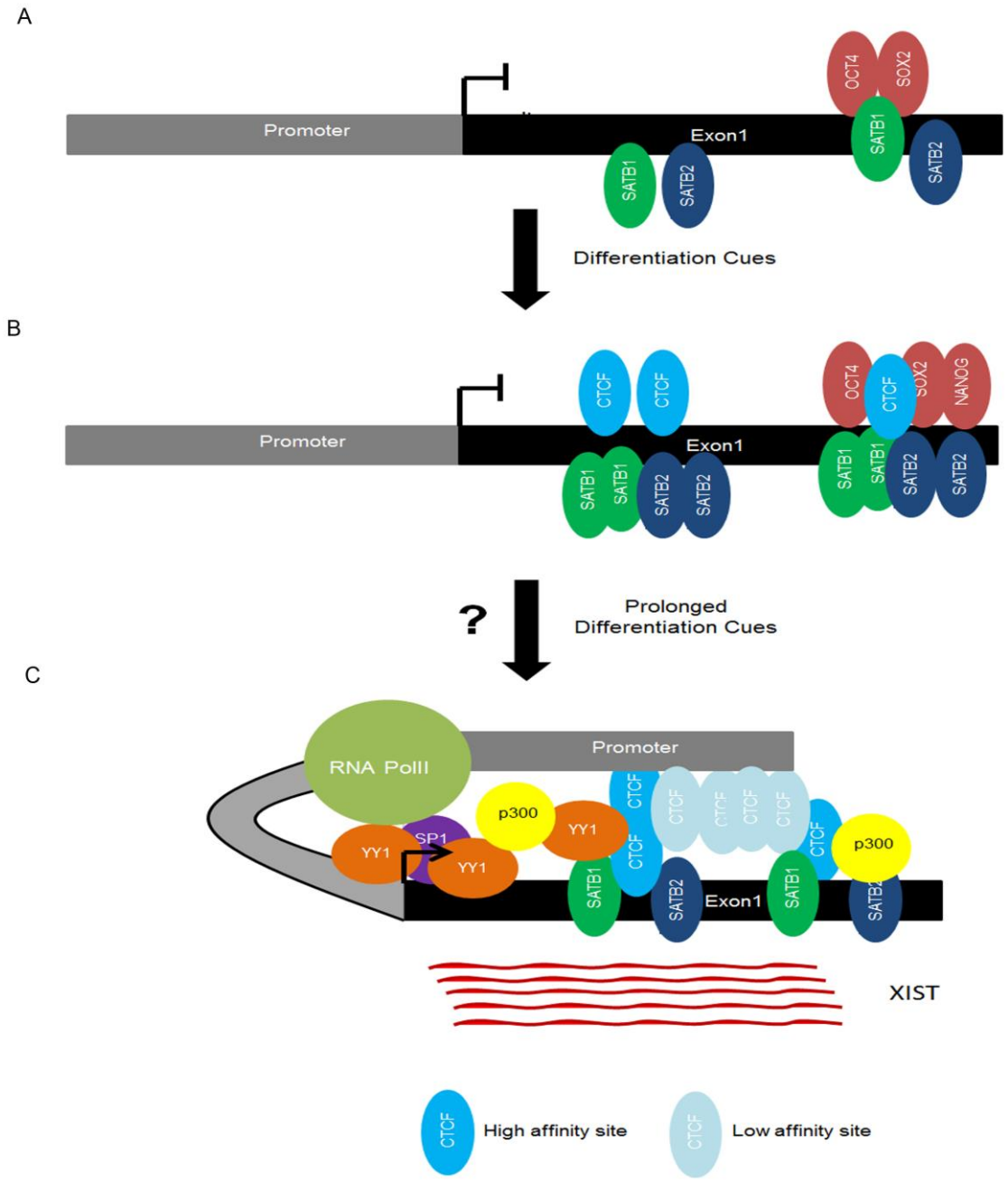


Figure 2.6.Proposed model for the regulation of human *XIST* promoter in the context of initiation phase of XCI.

The promoter is kept repressed in the undifferentiated state presumably by the pluripotency factors bound to *XIST* exon1 (A). Upon receiving differentiation cues, more factors such as SATB1, SATB2 and CTCF occupy multiple sites on *XIST* exon1 and the promoter is maintained in the repressed state (B). Prolonged differentiation cues results in the decrease in the levels of pluripotency factors (OCT4, SOX2 and NANOG), thereby causing a reduction in their enrichment on the exon1 site. This in turn possibly paves the way for the looping mediated interaction between CTCF enriched on the promoter as well as exon1 of *XIST* gene, recruitment of active histone modifiers (p300), transcription factors (SP1 and YY1) and RNA Polymerase II, resulting in the activation of transcription from the *XIST* promoter (C).

2.7 References

Agrelo, R., Souabni, A., Novatchkova, M., Haslinger, C., Leeb, M., Komnenovic, V., Kishimoto, H., Gresh, L., Kohwi-Shigematsu, T., Kenner, L., et al. (2009). SATB1 Defines the Developmental Context for Gene Silencing by Xist in Lymphoma and Embryonic Cells. Dev. Cell *16*, 507–516.

Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y., and Kohwi-shigematsu, T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Genes Dev. *14*, 521–535.

Ambrosetti, D.C., Schöler, H.R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the Fibroblast growth factor-4 enhancer. J. Biol. Chem. *275*, 23387–23397.

Andrews, P.W., Bronson, D.L., Benham, F., Strickland, S., and Knowles, B.B. (1980). A comparative study of eight cell lines derived from human testicular teratocarcinoma. Int. J. Cancer *26*, 269–280.

Andrews,P.W., I.Damjanov, D.Simon, G.Banting, C.Carlin, N.C.Dracopoli, and J.Fogh. 1984. Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2: differentiation *in vivo* and *in vitro*. *Lab. Invest.* 50, 147-162.

Augui, S., Nora, E.P., and Heard, E. (2011).Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. *12*, 429–442.

Austen, M., Lu, B., and Lu, J.M. (1997).Characterization of the Transcriptional Regulator YY1. Biochemistry *272*, 1709–1717.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development on SOX2 function. Genes Dev. *17*, 126–140.

Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat. Cell Biol. *8*, 293–299.

Balamotis, M. a., Tamberg, N., Woo, Y.J., Li, J., Davy, B., Kohwi-Shigematsu, T., and Kohwi, Y. (2012). Satb1 Ablation Alters Temporal Expression of Immediate Early Genes and Reduces Dendritic Spine Density during Postnatal Brain Development. Mol. Cell. Biol. *32*, 333–347.

Boiani, M., and Schöler, H.R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. Nat. Rev. Mol. Cell. Biol. *6*, 872–884.

Boyer, L. a L. a., Lee, T.I.T.I., Cole, M.F.M.F., Johnstone, S.E.S.E., Stuart, S., Zucker, J.P.J.P., Guenther, M.G.M.G., Kumar, R.M.R.M., Murray, H.L.H.L., Jenner, R.G.R.G., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell *122*, 947–956.

Britanova, O., Depew, M.J., Schwark, M., Thomas, B.L., Miletich, I., Sharpe, P., and Tarabykin, V. (2006).Satb2 haploinsufficiency phenocopies 2q32-q33 deletions, whereas loss suggests a fundamental role in the coordination of jaw development. Am. J. Hum. Genet.*79*, 668–678.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.

Brown, D.D., Wang, Z., Furlow, J.D., Kanamori, A., Schwartzman, R. a, Remo, B.F., and Pinder, A. (1996).The thyroid hormone-induced tail resorption program during Xenopus laevis metamorphosis. Proc. Natl. Acad. Sci. U. S. A. *93*, 1924–1929.

Cai, S., Lee, C.C., and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat. Genet. *38*, 1278–1288.

Chambers, I., and Smith, A. (2004).Self-renewal of teratocarcinoma and embryonic stem cells.Oncogene *23*, 7150–7160.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chang, L.S., Shi, Y., and Shenk, T. (1989). Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. J Virol *63*, 3479–3488.

Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S., and Lee, J.T. (2002).CTCF, a candidate trans-acting factor for X-inactivation choice. Science *295*, 345–347.

Chapman, A.G., Cotton, A.M., Kelsey, A.D., and Brown, C.J. (2014). Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. BMC Genet.*15*, 89.

Chiang, C.M., and Roeder, R.G. (1995).Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. Science *267*, 531–536.

Chow, J.C., Hall, L.L., Clemson, C.M., Lawrence, J.B., and Brown, C.J. (2003).Characterization of expression at the human XIST locus in somatic, embryonal carcinoma, and transgenic cell lines. Genomics *82*, 309–322.

Courey, A.J., and Tjian, R. (1988). Analysis of Sp1 in vivo reveals mutiple transcriptional domains, including a novel glutamine-rich activation motif. Cell *55*, 887–898.

Deuve, J.L., and Avner, P. (2011).The Coupling of X-Chromosome Inactivation to Pluripotency.Annu. Rev. Cell Dev. Biol. *27*, 611–629.

Dickinson, L. (1992). A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. Cell *70*, 631–645.

Dickinson, L.A., Dickinson, C.D., and Kohwi-Shigematsu, T. (1997). An atypical homeodomain in SATB1 promotes specific recognition of the key structural element in a matrix attachment region. J. Biol. Chem. *272*, 11463–11470.

Dobreva, G., Chahrour, M., Dautzenberg, M., Chirivella, L., Kanzler, B., Fariñas, I., Karsenty, G., and Grosschedl, R. (2006). SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. Cell *125*, 971–986.

Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E., and Shi, Y. (1999).Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. Mol. Cell. Biol. *19*, 7237–7244.

Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature *460*, 128–132.

Dynan, W.S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*, 79–87.

Emili, A., Greenblatt, J., and Ingles, C.J. (1994).Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. Mol. Cell. Biol. *14*, 1582–1593.

Ernst A. Wimmer, Herbert Jackie, C.P.& S.M.C. (1993). A Dorsophila homolouge of human SP1 is a head-specific segmentation gene. Nature *366*, 690–694.

Fedoriw, A.M., Stein, P., Svoboda, P., Schultz, R.M., and Bartolomei, M.S. (2004).Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science*303*, 9–12.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V. V (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol. Cell. Biol. *16*, 2802–2813.

Filippova, G.N., Cheng, M.K., Moore, J.M., Truong, J.P., Hu, Y.J., Nguyen, D.K., Tsuchiya, K.D., and Disteche, C.M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. Dev. Cell *8*, 31–42.

FitzPatrick, D.R., Carr, I.M., McLaren, L., Leek, J.P., Wightman, P., Williamson, K., Gautier, P., McGill, N., Hayward, C., Firth, H., et al. (2003). Identification of SATB2 as the cleft palate gene on 2q32-q33. Hum. Mol. Genet. *12*, 2491–2501.

Galande, S., Purbey, P.K., Notani, D., and Kumar, P.P. (2007). The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. Curr.Opin.Genet. Dev. *17*, 408–414.

Galvin, K.M., and Shi, Y. (1997).Multiple mechanisms of transcriptional repression by YY1. Mol. Cell. Biol. *17*, 3723–3732.

Giuseppe Borsani, Rossana Tonlorenzi, M. Christine Simmler, Luisa Dandolo, Danielle Arnaud, Valeria Capra, Markus Grompe, Antonio Pizzuti, Donna Muzny, Charles Lawrence, H.F.W., Avner, P., and Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *354*, 56–58.

Goodrich, L., Panning, B., and Leung, K.N. (2016). Activators and repressors: A balancing act for X-inactivation. Semin. Cell Dev. Biol. *56*, 3–8.

Gordon, S., Akopyan, G., Garban, H., and Bonavida, B. (2006). Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. Oncogene *25*, 1125–1142.

Grant, C., Bailey, T. and Noble, W. (2011).FIMO : Scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018.

Grumbach, M.M., Morishima, A., and Taylor, J.H. (1963). Human Sex Chromosome Abnormalities in Relation To Dna Replication and Heterochromatinization. Proc. Natl. Acad. Sci. U. S. A. *49*, 581–589.

Han, H.-J., Russo, J., Kohwi, Y., and Kohwi-Shigematsu, T. (2008). SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. Nature *452*, 187–193.

Hariharan, N., Kelley, D.E., and Perry, R.P. (1991). Delta, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. Proc. Natl. Acad. Sci. U. S. A. *88*, 9799–9803.

Hendrich, B.D., Plenge, R.M., and Willard, H.F. (1997). Identification and characterization of the human XIST gene promoter: Implications for models of X chromosome inactivation. Nucleic Acids Res. *25*, 2661–2671.

Hoey, T., Weinzierl, R.O.J., Gill, G., Chen, J.L., Dynlacht, B.D., and Tjian, R. (1993). Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. Cell *72*, 247–260.

Hyde-deruyscher, R.P., Jennings, E., and Shenk, T. (1995).DNA binding sites for the transcriptional activator/repressor YY1. Nucleic Acids Res. *23*, 4457–4465.

Jeon, Y., and Lee, J.T. (2011). YY1 Tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987).Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell *51*, 1079–1090.

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

Kay, G.F., Barton, S.C., Surani, M.A., and Rastan, S. (1994). Imprinting and X-chromosome counting mechanisms determine Xist expression in early mouse development. Cell *77*, 639–650.

Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkov, V. V (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol. Cell. Biol. *13*, 7612–7624.

Kumar, P.P., Bischof, O., Purbey, P.K., Notani, D., Urlaub, H., Dejean, A., and Galande, S. (2007). Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. Nat. Cell Biol. *9*, 45–56.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc. Natl. Acad. Sci. U. S. A. *103*, 10684–10689.

Lee, V.M., and Andrews, P.W. (1986). Differentiation of NTERA-2 clonal human embryonal carcinoma cells into neurons involves the induction of all three neurofilament proteins. J. Neurosci. *6*, 514–521.

Leone, D.P., Heavner, W.E., Ferenczi, E.A., Dobreva, G., Huguenard, J.R., Grosschedl, R., and McConnell, S.K. (2015). Satb2 regulates the differentiation of both callosal and subcerebral projection neurons in the developing cerebral cortex. Cereb.Cortex *25*, 3406–3419.

Lewis, B. a, Tullis, G., Seto, E., Horikoshi, N., Weinmann, R., and Shenk, T. (1995). Adenovirus E1A proteins interact with the cellular YY1 transcription factor. J. Virol. *69*, 1628–1636.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

Makhlouf, M., Ouimette, J.-F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014).A prominent and conserved role for YY1 in Xist transcriptional activation. Nat. Commun. *5*, 1–12.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625-U26.

Migeon, B.R., Chowdhury, A.K., Dunston, J.A., and Mcintosh, I. (2001). Identification of TSIX , Encoding an RNA Antisense to Human XIST , Reveals Differences from its Murine Counterpart : Implications for X Inactivation. Am. J. Hum. Genet.*69*, 951–960.

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

Minkovsky, A., Barakat, T.S., Sellami, N., Chin, M.H., Gunhanlar, N., Gribnau, J., and Plath, K. (2013). The pluripotency factor-bound intron 1 of xist is dispensable for X chromosome inactivation and reactivation In Vitro and In Vivo. Cell Rep. *3*, 905–918.

Mir, R., Pradhan, S.J., Patil, P., Mulherkar, R., and Galande, S. (2015). Wnt/β-catenin signaling regulated SATB1 promotes colorectal cancer tumorigenesis and progression. Oncogene *35*, 1–13.

Monkhorst, K., de Hoon, B., Jonkers, I., Achame, E.M., Monkhorst, W., Hoogerbrugge, J., Rentmeester, E., Westerhoff, H. V., Grosveld, F., Grootegoed, J.A., et al. (2009). The probability to initiate X chromosome inactivation is determined by the X to autosomal ratio and X chromosome specific allelic properties. PLoS One *4*, 1–14.

Montalvo, E.A., Cottam, M., Hill, S., and Wang, Y.J. (1995). YY1 binds to and regulates cis-acting negative elements in the Epstein-Barr virus BZLF1 promoter. J. Virol. *69*, 4158–4165.

Navarro, P., Page, D.R., Avner, P., and Rougeulle, C. (2006). Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. Genes Dev. *20*, 2787–2792.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008).Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693–1695.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010).Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457–460.

Nechanitzky, R., Dávila, A., Savarese, F., Fietze, S., and Grosschedl, R. (2012). Satb1 and Satb2 are dispensable for X Chromosomeinactivation in mice. Dev. Cell *23*, 866–871.

Nesterova, T.B., Senner, C.E., Schneider, J., Alcayna-Stevens, T., Tattermusch, A., Hemberger, M., and Brockdorff, N. (2011). Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. Epigenetics Chromatin *4*, 17.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998).Formation of pluripotent stem cells in the mammalian embryo dependes on the POU transcription factor Oct4. Cell *95*, 379–391.

Niwa, H. (2001). Molecular mechanism to maintain stem cell renewal of ES cells.Cell Struct.Funct.*26*, 137–148.

Notani, D., Gottimukkala, K.P., Jayani, R.S., Limaye, A.S., Damle, M. V., Mehta, S., Purbey, P.K., Joseph, J., and Galande, S. (2010). Global regulator SATB1 recruits Beta-catenin and regulates TH2 differentiation in Wnt-dependent manner. PLoS Biol. *8*.

Notani, D., Ramanujam, P.L., Kumar, P.P., Gottimukkala, K.P., Kumar-Sinha, C., and Galande, S. (2011). N-terminal PDZ-like domain of chromatin organizer SATB1 contributes towards its function as transcription regulator. J. Biosci. *36*, 461–469.

Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet.*17*, 520–527.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

Palko, L., Bass, H.W., Beyrouthy, M.J., and Hurt, M.M. (2004). The Yin Yang-1 (YY1) protein undergoes a DNA-replication-associated switch in localization from the cytoplasm to the nucleus at the onset of S phase. J. Cell Sci. *117*, 465–476.

Pant, V., Mariano, P., Kanduri, C., Mattsson, A., Lobanenkov, V., Heuchel, R., and Ohlsson, R. (2003). The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCFand the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. Genes Dev. *17*, 586–590.

Park, K., and Atchison, M.L. (1991). Isolation of a candidate repressor/activator, NF-E1 (YY1, d), that binds to the immunoglobulin k 3' enhancer and the immunoglobulin heavy-chain mE1 site. Pnas *88*, 9804–9808.

Pascal, E., and Tjian, R. (1991). Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. Genes Dev. *5*, 1646–1656.

Pavan Kumar, P., Purbey, P.K., Sinha, C.K., Notani, D., Limaye, A., Jayani, R.S., and Galande, S. (2006). Phosphorylation of SATB1, a Global Gene Regulator, Acts as a Molecular Switch Regulating Its Transcriptional Activity In Vivo. Mol. Cell *22*, 231–243.

Petkova, V., Romanowski, M.J., Sulijoadikusumo, I., Rohne, D., Kang, P., Shenk, T., and Usheva, A. (2001).Interaction between YY1 and the Retinoblastoma Protein.Biochemistry *276*, 7932–7936.

Phillips, J.E., and Corces, V.G. (2009). CTCF: Master Weaver of the Genome. Cell *137*, 1194–1211.

Pisaneschi, G., Ceccotti, S., Falchetti, M.L., Fiumicino, S., Carnevali, F., and Beccari, E. (1994). Characterization of FIII/YY1, a Xenopus laevis conserved zinc-finger protein binding to the first exon of L1 and L14 ribosomal protein genes. Biochem.Biophys. Res. Commun. *205*, 1236–1242.

Purbey, P.K., Singh, S., Kumar, P.P., Mehta, S., Ganesh, K.N., Mitra, D., and Galande, S. (2008). PDZ domain-mediated dimerization and homeodomain-directed specificity are required for high-affinity DNA binding by SATB1. Nucleic Acids Res. *36*, 2107–2122.

Purbey, P.K., Singh, S., Notani, D., Kumar, P.P., Limaye, A.S., and Galande, S. (2009). Acetylation-dependent interaction of SATB1 and CtBP1 mediates transcriptional repression by SATB1. Mol. Cell. Biol. *29*, 1321–1337.

Saffer, J.D., Jackson, S.P., and Annarella, M.B. (1991).Developmental expression of Sp1 in the mouse. Mol. Cell. Biol. *11*, 2189–2199.

Sarkar, K.M., Gayen, S., Kumar, S., Maclary, E., Buttigieg, E., Hinten, M., Kumari, A., Harris, C., Sado, T.and Kalantry, S. (2015). An Xist-activating antisense RNA required for X-chromosome inactivation. Nat. Comm. *6*, 1-13.

Savarese, F., Dávila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C.F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A.G., et al. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes Dev. *23*, 2625–2638.

Schütz, B., Schäfer, M.K.H., Gördes, M., Eiden, L.E., and Weihe, E. (2015).Satb2-Independent Acquisition of the Cholinergic Sudomotor Phenotype in Rodents.Cell. Mol. Neurobiol. *35*, 205–216.

Shi, Y., Seto, E., Chang, L.S., and Shenk, T. (1991).Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell *67*, 377–388.

Shi, Y., Lee, J.S., and Galvin, K.M. (1997). Everything you have ever wanted to know about Yin Yang 1... Biochim. Biophys. Acta - Rev. Cancer *1332*, F49–F66.

Skotheim, R.I., Lind, G.E., Monni, O., Nesland, J.M., Abeler, V.M., Fosså, S.D., Duale, N., Brunborg, G., Kallioniemi, O., Andrews, P.W., et al. (2005). Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal development. Cancer Res. *65*, 5588–5598.

Spencer, R.J., del Rosario, B.C., Pinter, S.F., Lessing, D., Sadreyev, R.I., and Lee, J.T. (2011). A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-Chromosome inactivation. Genetics *189*, 441–454.

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

Suske, G. (1999). The Sp-family of transcription factors. Gene *238*, 291–300.

Szabo, P.E., Tang, S.E., Silva, F.J., Tsark, W.M.K., and Mann, J.R. (2004).Role of CTCF Binding Sites in the Igf2 / H19 Imprinting Control Region †. Mol. Cell. Biol. *24*, 4791–4800.

Szabó, P.E., Tang, S.H.E., Rentsendorj, A., Pfeifer, G.P., and Mann, J.R. (2000). Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr. Biol. *10*, 607–610.

Szemes, M., Gyorgy, A., Paweletz, C., Dobi, A., and Agoston, D. V. (2006). Isolation and characterization of SATB2, a novel AT-rich DNA binding protein expressed in development- and cell-specific manner in the rat brain. Neurochem. Res. *31*, 237–246.

Tanese, N., Saluja, D., Vassallo, M.F., Chen, J.L., and Admon, A. (1996). Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100. Proc. Natl. Acad. Sci. U. S. A. *93*, 13611–13616.

Wang, Z., Yang, X., Guo, S., Yang, Y., Su, X.C., Shen, Y., and Long, J. (2014). Crystal structure of the ubiquitin-like domain-cut repeat-like tandem of special AT-rich sequence binding protein 1 (SATB1) reveals a coordinating DNA-binding mechanism. J. Biol. Chem. *289*, 27376–27385.

Xu, N., Donohoe, M.E., Silva, S.S., and Lee, J.T. (2007).Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. Nat Genet *39*, 1390–1396.

Yant, S.R., Zhu, W., Millinoff, D., Slightom, J.L., Goodman, M., and Gumucio, D.L. (1995). High affinity YY1 binding motifs: Identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human ?? globin cluster. Nucleic Acids Res. *23*, 4353–4362.

Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. Nature *419*, 641–645.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, a, Ohbo, K., Gross, M., Hübner, K., Schöler, H.R., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*, 881–894.

Chapter 3: Regulation of XIST in the context of maintenance phase of X inactivation

3.1 Introduction

As elaborated in the Chapter 1, upregulation of Xist marks the initiation event of XCI. Work from few earlier studies have suggested that Xist RNA is required only during the initiation and establishment phases of XCI and is dispensable during the maintenance phase despite its continuous synthesis (Brown and Willard, 1994; Csankovszki et al., 1999). The proposed reason for this switch is the alterations of the underlying chromatin structure of the Xi due to accumulation of repressive epigenetic modifications as a consequence of Xist RNA coating. This causes a shift from reversible, Xist-dependent silencing to irreversible, Xist-independent one (Wutz and Jaenisch, 2000). This widely held belief was countered by another study demonstrating the significance of Xist RNA in maintaining the perinucleolar positioning of Xi in MEF, which in turn is essential for maintenance of heterochromatin marks and stable silencing (Zhang et al., 2007). Furthermore, a recent study from the same group also highlighted the role of Xist RNA in repelling cohesin molecules to prevent acquisition of transcriptionally permissive state on the Xi in the maintenance phase (Minajigi et al., 2015). These findings provide clear evidence that in addition to the repressive epigenetic marks on the Xi, Xist RNA is also required for stable maintenance of silencing in the somatic cells. However, our knowledge about the role and regulation of *Xist*/*XIST* during the maintenance phase is lacking.

Evidently much of the field's focus has been drawn towards understanding the events leading to the establishment of inactive X using mESCs (discussed in Chapters 1 and 2) and probing the questions pertaining to stable maintenance of Xi have gained rather limited attention. Using super-resolution microscopy (Stochastic Optical Reconstruction Microscopy), we now know that the stoichiometry of Xist RNA molecules to Xi in the maintenance phase is much lower than previously thought. It was found that there are only 50-100 binding stations for Xist RNA and PRC2 (Buzin et al., 1994; Sun et al., 2006; Sunwoo et al., 2015). Therefore, it can be concluded that Xist RNA though indispensable for maintenance of heterochromatin status of the Xi, does not actually coat the entire Xi and the

epigenetic modifications serve to retain the robustness of silencing in the maintenance phase. Also, it is an established fact that Xist RNA is synthesized and remains tethered to Xi throughout mitosis and that the amount of Xist bound to DNA doubles from G1 to G2 phases indicating that the binding capacity is proportional to the amount of chromosomal DNA (Jonkers et al., 2008; Ng et al., 2011). This is also indicative of the fact that Xist binding on the Xi is persistent and is maintained, possibly to prevent a transcriptionally active status being achieved by the Xi.

It is imperative to mention that all the above mentioned studies have been carried out using differentiating mESCs or mouse fibroblast cells and the findings cannot be directly extrapolated to human *XIST* with certainty. This is majorly owing to the challenges arising due to inherent variability of human ES cell line system in terms of initiating and establishing XCI. In addition, the expression kinetics of human XIST and mouse Xist also display remarkable differences during the initiation of rXCI (discussed in Chapter 1 and 2) (Dvash et al., 2010; Hoffman et al., 2005; Silva et al., 2008; Tomoda et al., 2012). Having said this, there have been a few studies attempting to characterize the promoter of *XIST* and identify the transcription factors essential for governing and maintaining robust levels of XIST during the maintenance phase.

The first study towards this uncovered the binding sites for three transcription factors - SP1, YY1 and TBP on the promoter region by carrying out saturated site directed mutagenesis and validated the same by *in vitro* binding assay (Hendrich et al., 1997). Although EMSA serves as a good tool to discover the binding site(s) of a particular protein on DNA, it has its own associated limitations because the *in vitro* and *in vivo* scenarios can be totally different. Therefore addressing the role(s) of these factors in the cell system is necessary to be able to conclusively demonstrate the significance of the observations made in the original paper. We have presented our findings with regards to *XIST* regulation in the context of initiation phase of XCI in the second chapter of the thesis. But an important question about whether these factors can actually bind the promoter and play any role in controlling *XIST* transcription during the maintenance phase remained unanswered until recently. Two independent studies demonstrated that YY1 is an essential transcription factor involved in regulating *Xist*/*XIST* transcription in both ES cells as well as differentiated cells (Chapman et al., 2014; Makhlouf et al., 2014). The results in Makhlouf et al., 2014

highlighted the role of YY1 in regulating P1 promoter, while those published by Chapman et al., 2014 showed its significance in regulating a second promoter P2. The former study identified conserved YY1 and CTCF binding sites on the *XIST* locus from seven different mammals. However, they ruled out CTCF's role in asymmetrically regulating human/mouse *XIST*/*Xist* since it did not show any differential enrichment when compared between their respective male and female fibroblast cells. However, as discussed in Chapter 2, we have discovered that CTCF shows a distinct pattern of occupancy on the *XIST* exon1 region as determined by ChIP in human EC cells. Moreover, by performing luciferase reporter assay, we observe that this high affinity CTCF binding site present on *XIST* exon1, when fused to the minimal promoter element leads to decline in the measured reporter activity as opposed to the dramatic increase observed when the same site is present along with a number of low affinity CTCF binding sites located on the extended promoter region. Interestingly, YY1 and CTCF seem to occupy the same site, ~2.5 Kb upstream of *XIST* TSS, as per ChIP-sequencing profile from multiple cell lines (depicted in results section, Figure 3.5.3(A)). Moreover, YY1 and CTCF are known to physically interact with each other in mouse as well as human cell lines (Donohoe et al., 2007). Therefore, we sought to determine and assign specific roles to the three ubiquitously expressed zinc-finger transcription factors –SP1, YY1 and CTCF in the present study. Findings discussed in Chapter 2 highlighted the significance of chromatin organizer proteins – SATB1 and SATB2 towards transcriptional regulation of *XIST*. Therefore, we also chose to probe their roles in terms of regulation of *XIST* in the maintenance phase so as to arrive at the consensus of factors regulating *XIST* in the initiation and maintenance phases.

3.2 Rationale of the present study

As discussed in the section 3.1, there have been only two studies addressing transcriptional regulation of *XIST* in the maintenance phase using fibroblast and mouse-human hybrid cell lines (Chapman et al., 2014; Makhlouf et al., 2014). Both these studies attributed the role of transcription factor regulating *XIST* to YY1. These findings were published while we were pursuing the work presented here. In this part of the thesis, we set out to determine the roles of factors in addition to YY1 such as SP1, CTCF, SATB1 and SATB2 based on the results discussed in Chapter 2. To address the question of understanding transcriptional regulation of *XIST* we chose HEK 293T cell line. This is a partially triploid

female cell line (ATCC CRL 3216) harbouring at least one inactive X and expresses XIST robustly. Importantly, it is a widely used cell line to study a number of basic biological processes due to its ease of culturing, transfections and other such biochemical assays. Therefore, this provides a good system to look into the aspect of *XIST* regulation during the maintenance phase of XCI.

3.3 Summary of the work

In this chapter, we present our results with respect to the regulation of *XIST* promoter in the context of maintenance phase of XCI. Firstly, we biochemically assessed the activities of several *XIST* promoter constructs by measuring the amount of firefly luciferase produced. Next, we tested the effect of perturbing the levels of SATB1 and/or SATB2 on the promoter activities measured biochemically as well as scored for the expression of endogenous mature and premature XIST transcripts. Contrary to the findings described in Chapter 2, we observed that overexpression of either SATB1 and/or SATB2 actually led to downregulation of XIST transcript levels. Further we attempted to address the roles of SP1, YY1 and CTCF in the process by perturbing their levels and testing the promoter activities biochemically as well as scoring for endogenous XIST expression. Our results also support the published findings that YY1 is the primary factor governing transcription from *XIST* promoter (P1). Interestingly, upon depletion of SP1/YY1/CTCF for prolonged duration (72 hours) by performing RNAi-mediated knockdown, it was observed that even SP1 and CTCF can act as potential transcription factors for *XIST* promoter. Collectively, the results discussed in this chapter suggest that the regulation of *XIST* in the maintenance phase is not as simple as thought previously and instead production of XIST in the somatic cells may be an outcome of concerted action of multiple factors acting either together or independently.

3.4 Materials and Methods

3.4.1 Cell Culture

Human embryonic kidney cells (HEK293T) were grown in DMEM (Invitrogen, Carlsbad, California, USA) without sodium pyruvate, high glucose, supplemented with 10% fetal bovine serum (Invitrogen, Carlsbad, California, USA) and penicillin-streptomycin (Invitrogen, Carlsbad, California, USA) and maintained at 37°C under 5% CO_2 atmosphere. HEK293T cells

were passaged upon reaching 70-80% confluency using 0.05% trypsin (Invitrogen, Carlsbad, California, USA).

3.4.2 Transfection of DNA and siRNA

HEK293T cells were transfected with the DNA constructs using Lipofectamine 2000 (Invitrogen, Carlsbad, California, USA) as per manufacturer's guidelines. Briefly the cells were seeded and grown in the medium conditions described in the section 3.4.1. Cells were transfected with equimolar amounts XIST promoter DNA constructs (0.3µM) in a 12-well plate for the reporter assay shown in Figure 3.5.1. For all the other reporter assays, EC cells were transfected with 1µg of Firefly luciferase construct along with 100ng of Renilla luciferase construct in a 12-well plate. For overexpression of SATB1/2, cells were transfected with 8µg each of FLAG or FLAG-SATB1 or FLAG-SATB2 in a 100mm dish (Figure 3.5.2 (B)) or 0.5µg each of FLAG or FLAG-SATB1 or FLAG-SATB2 along with 1µg of XIST1.1 and 100ng of Renilla luciferase for the reporter assay (Figure 3.5.2(A)). For transfecting HEK 293T cells with siRNAs, RNAiMax transfection reagent (Invitrogen, Carlsbad, California,USA) was used according to the protocol provided by the manufacturer. For RNAi mediated knockdown of SP1, YY1 and CTCF for the reporter assays, cells were transfected with 5 picomoles of control siRNA, SP1 siRNA or YY1 siRNA and 10 picomoles of CTCF siRNA (Figure 3.5.3 (B)). For siRNA mediated knockdown of SP1, YY1 or CTCF in HEK 293T cells to determine endogenous XIST levels, cells were transfected with 12.5 picomoles of control siRNA, SP1 siRNA, YY1 siRNA and 25 picomoles of CTCF siRNA (Figure 3.5.4 (B), 3.5.5 (B)). All the transfections were carried out in OptiMEM medium (Invitrogen, Carlsbad, California,USA). 6 hours post transfection, culture dishes were replaced with the fresh complete DMEM medium. Cells were harvested for RNA isolation or protein extraction or reporter assays 48 hour post transfection or as indicated in the individual experimental schemes. For the results shown in the Figure 3.5.5 reverse transfection were performed. Briefly, HEK 293T cells were transfected with the equal amounts of desired siRNAs during the time of seeding. The transfection was done in DMEM + 10% serum containing medium. 12 hours post transfection, fresh medium containing antibiotic was added to the culture dish. The cells were harvested 72 hours post transfection for RNA and protein.

3.4.3 Luciferase Reporter Assay

Luciferase reporter assays were performed using the Dual luciferase assay kit from Promega (Madison, Wisconsin, USA). Cells were transfected with Firefly luciferase and Renilla luciferase DNA constructs as described in the section 3.4.2. After harvesting, the cells were lysed using 1X Passive lysis buffer (Promega, Madison, Wisconsin, USA) as per manufacturer's instructions. The lysates and substrates were mixed in the optical bottom 96-well plate (ThermoFisher Scientific, Waltham, Massachusetts, United States) according to the guidelines provided by Promega (Madison, Wisconsin, USA). The reporter activities were measured using luminometry mode on the Varioskan machine. In all the assays, Renilla luciferase activity measurement serves as an internal control. The fold change was calculated with respect to the vector control or siRNA control as and when mentioned.

3.4.4 RNA extraction and cDNA synthesis

To isolate RNA, cell pellets were resuspended and homogenized in Trizol reagent (Invitrogen, Carlsbad, California, USA). 0.2 volumes of chloroform was added and mixed gently. The aqueous and the organic phases were allowed to get separated by standing the tubes for 5 to 10 minutes. The tubes were spun at 14000 rcf at 4°C for 10 minutes. The aqueous layer was collected in the fresh tube and 0.8 volume of isopropanol was added for precipitating RNA. The tubes were incubated at room temperature for 15 to 20 minutes and then spun at 14000 rcf at 4°C for 10 minutes, followed by washing the nucleic acid pellet with 75% ethanol, air drying the pellet, dissolving it nuclease-free water and incubating at 55°C for 10 minutes. After assessing the quality and quantity of RNA using Nanodrop (ThermoFisher Scientific, Waltham, Massachusetts, United States), 2µg of RNA was used for DNase treatment (Promega, Madison, Wisconsin, USA) as per the protocol provided by the manufacturer. This was followed by cDNA synthesis using Reverse trascriptase kits either from Promega (ImpromII) (Madison, Wisconsin, USA) or Applied Biosystems (Foster City, California, USA) (High capacity cDNA synthesis kit). Also, -RT control was set up to verify the efficiency of DNase treatment. The synthesized cDNA was used to set up quantitative real-time PCR (qRT-PCR).

3.4.5 Protein extraction and immunoblotting

Cell pellets were resuspended in RIPA buffer (10mM Tris (pH 8.0), 1mM EDTA (pH 8.0), 0.5mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140mM NaCl) containing 5% glycerol and 1X protease inhibitors (procured from Roche, Basel, Switzerland) and lysed by repeated freeze-thaw cycles (2 to 3). The lysates were centrifuged at 14000 rpm, 4°C, 30 minutes to get rid of the cellular debris. The supernatant was collected in the fresh microfuge tube. The concentrations of protein were determined by performing BCA assay (purchased from Thermoscientific, Waltham, Massachusetts, USA). Equal amounts of protein lysates were boiled in 1X Lamelli buffer (0.5 M Tris-HCl pH 6.8, 28% glycerol, 9% SDS, 5% 2-mercaptoethanol, 0.01% bromophenol blue) for 10-15 minutes and subjected to electrophoresis on a polyacrylamide gel. The separated proteins were transferred onto PVDF membrane using phosphate based transfer buffer (10mM sodium phosphate monobasic, 10mM sodium phosphate dibasic) at 4°C, 400mA, 3 hours. After the completion of transfer, membranes were blocked in 5% skimmed milk, incubated overnight at 4°C with the primary antibodies prepared either in 5% milk or 5% BSA. Next day, the membranes were washed thrice with the buffer containing 20 mM Tris buffer pH 7.4, 500 mM NaCl and 0.1% tween 20 (TST) and incubated with the appropriate secondary antibodies conjugated with horseradish peroxidase for an hour at room temperature. Following this, the membranes were again washed thrice with TST buffer. The blots were developed using Immobilon Western Chemiluminescent HRP Substrate (Millipore, Billerica, MA, USA) and detected using ImageQuant LAS 4000 (GE Healthcare, Piscataway, NJ, USA) according to the manufacturer's instructions.

3.4.6 Quantitative real-time PCR

For transcript quantitation, cDNA prepared (as described in the section 3.4.4) was diluted 5 times with nuclease free water and used as template for PCR along with specific set of primer pairs. SYBR Green chemistry (Roche, Basel, Switzerland and ABI, Foster City, California, USA) was used and the reaction was set up on the ABI Viia7 cycler (Foster City, California, USA). Changes in threshold cycles were calculated by subtracting the Ct values of the gene of interest from that of housekeeping control ($Ct_{target\ genes} - Ct_{\beta\text{-actin/18s rRNA}}$). ΔCt values of specific target genes from the experimental samples were then subtracted from

their respective control samples to generate ΔΔCt values. The fold changes were calculated using the formula : $2^{(-(\Delta\Delta Ct\ value))}$.

3.4.7 Antibodies, siRNAs and other reagents

SATB1 (3650S) and SP1 (5931S) antibodies for immunoblotting were procured from Cell Signaling Technology (Danvers, Massachusetts, USA), SATB2 (ab51502) and YY1 (ab12132) antibodies for western blotting were purchased from Abcam (Cambridge, UK), CTCF antibody (sc-21298) purchased from Santacruz Biotechnologies (Dallas, Texas, USA) was used for immunoblotting. β-ACTIN (VMA00048) and γ-TUBULIN primary antibodies, mouse-HRP and rabbit-HRP secondary antidbodies were purchased from Biorad Laboratories (Hercules, California, USA). siRNA targeting SP1, YY1 and Luciferase were synthesized by Sigma (St. Louis, Missouri, USA), siRNA targeting CTCF (6265) was procured from Cell Signaling Technology (Danvers, Massachusetts, USA), Non targeting siRNA (D-001810-10-05) was obtained from Dharmacon (Thermoscientific, Waltham, Massachusetts, USA). Dual luciferase assay kit (E1960) was purchased from Promega (Madison, Wisconsin, USA), Lipofectamine 2000 and RNAiMax transfection reagents were procured from Invitrogen (Carlsbad, California, USA), RNase-free DNase and ImpromII reverse transcriptase enzyme was purchased from Promega (Madison, Wisconsin, USA), High specificity cDNA synthesis kit is from Applied Biosystems (ABI) (Foster City, California, USA), SYBR Green for qPCR was procured from ABI (Foster City, California, USA) and Roche (Basel, Switzerland).

3.5 Results

3.5.1 Assaying for the promoter activities in HEK293T cells

Human embryonic kidney cells – HEK293T are female in origin, harbour X chromosome(s) which have already undergone inactivation and are stably maintained as Xi and hence express XIST robustly. In order to use these cells for characterizing the promoter and understand *XIST* regulation, we first tested the activities of the luciferase reporter constructs described in the chapter 1 by transient transfection. Similar to the results described in the previous chapter, we observed that all the promoter constructs (101 bp to 4458 bp) led to an increase in the transcription from the luciferase reporter gene over vector control (Figure 3.5.1). It can be observed that just the 101 bp region is sufficient to drive the transcription of the reporter gene and hence can be considered the 'minimal promoter

element'. The DNA construct encompassing 1100 bp region, cloned in the antisense orientation to luciferase gene, fails to show the reporter activity and serves as an important negative control for the experiment (Bar7, Figure 3.5.1).

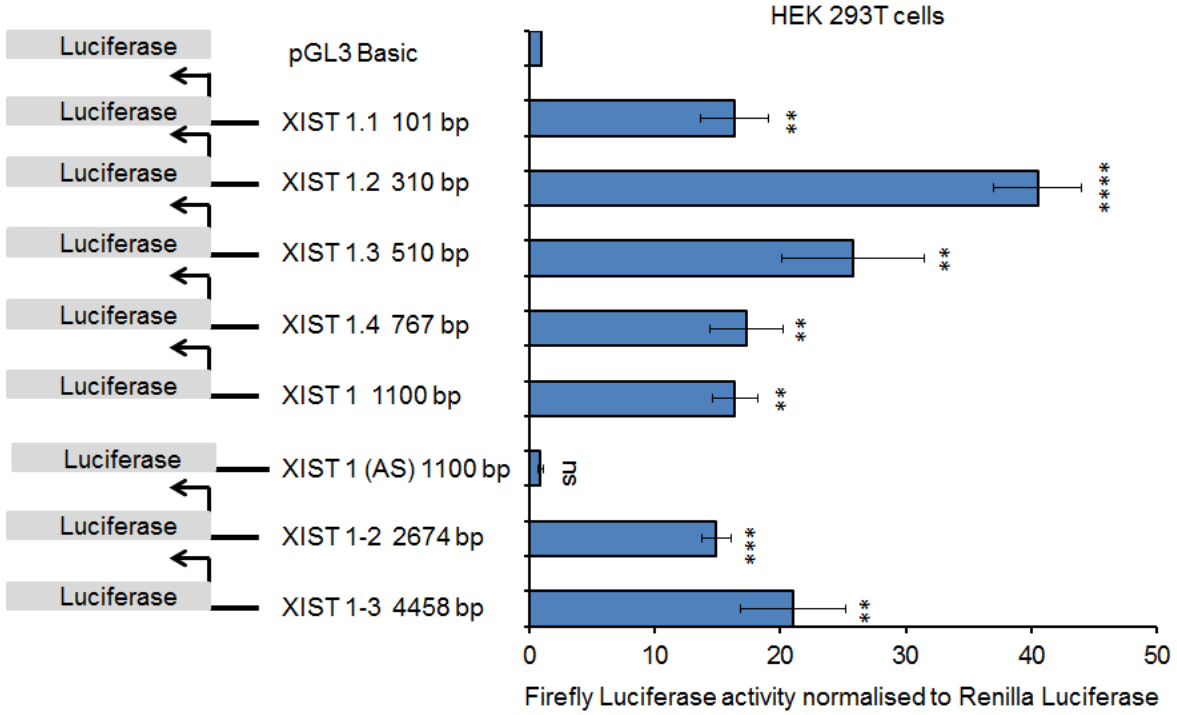


Figure 3.5.1 *XIST* promoter activity in HEK293T cells as measured via luciferase reporter assay.

Genomic region upstream of *XIST* TSS including a part of 5' region of gene was cloned into pGL3 Basic as described in the 'Materials and Methods' section (2.4.1). Equimolar amounts of the confirmed clones along with Renilla Luciferase were transfected into EC cells. 48 hour post transfection, cell were harvested and the reporter assay was performed as described in the Section 3.4.3. Shown on the left is the schematic of genomic regions of *XIST* promoter fragments cloned into promoter less pGL3 Basic vector. Luciferase reporter activities for *XIST* promoter constructs transfected into HEK293T cells compared to pGL3 Basic vector control. 101 bp region is sufficient to act as the minimal promoter element for *XIST*. Firefly luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis represents the normalized fold change in the firefly luciferase activity and Y-axis indicates DNA constructs transfected into cells. Each bar represents values from three independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, *** P value < 0.001, ** P value < 0.01, ns = non-significant).

3.5.2 Effect of increasing the levels of SATB proteins on *XIST* minimal promoter activity and endogenous XIST levels

The results discussed in Chapter 2 suggested that the chromatin organizer proteins – SATB1 and SATB2 positively regulate *XIST* expression in the context of initiation phase of XCI. Therefore, we tested their roles in regulating *XIST* in HEK 293T cells as well which provide the

maintenance phase context for XCI. To probe their roles, we transfected HEK 293T cells with the minimal promoter construct XIST1.1 along with either SATB1 and/or SATB2 overexpression constructs. 48 hours post transfections, the cells were harvested for measuring the firefly luciferase activity biochemically as a read out of the promoter activity. This experiment offered results distinct to that described in Chapter 2. The results from the current experiment suggested that the promoter activity remained unaffected upon forced overexpression of either SATB1 or SATB2 alone. Increasing the levels of both the proteins however led to a significant, though marginal upregulation in the reporter activity assayed (Figure 3.5.2 (A)).

In order to understand this further, we next forcefully increased the expression of SATB1 and/or SATB2 by transfecting HEK 293T cells with overexpression constructs and scored for the levels of premature as well as mature XIST transcripts so as to be able to understand their effect on endogenous XIST transcription. Again it was observed that the results obtained were in contrast to that discussed in the previous chapter. Moreover, they also did not match our reporter assay results shown in Figure 3.5.2 (A). It was observed that overexpression of SATB1 and/or SATB2 actually led to significant downregulation of mature as well as premature XIST transcripts We also scored for the expression of known (YY1) and putative transcription regulators (SP1 and CTCF) of *XIST* promoter and observed that the expression of SP1 decreases upon increasing the levels of SATB1 and/or SATB2. Also, YY1 levels decrease only when SATB2 is overexpressed and expression of CTCF remains unaltered. These observations put together suggest that the reduction in XIST levels obtained upon increasing the levels of SATB1 or SATB2 can either be an (i) indirect effect caused by reduction in the levels of SP1 or YY1 or (i) a direct effect of SATB proteins binding to *XIST* genomic locus and influencing the transcriptional outcome from the promoter as suggested in the previous chapter.

It has been established that SATB1/2 can form homo as well as heterodimers and act in a context dependent manner either as an activator or a repressor (Pavan Kumar et al., 2006; Purbey et al., 2008, 2009). Also the disparity observed between the measured XIST1.1 promoter activity (Figure 3.5.2 (A)) and the endogenous XIST levels (Figure 3.5.2 (C), (D)) upon overexpression of SATB1/2 suggests that SATB proteins may actually be binding to the regions aside from the minimal promoter element (as predicted and demonstrated by ChIP

for two such regions on exon1 in Chapter 2) and leading to opposite outcomes. However, whether SATB1 and SATB2 actually bind to the predicted sites on *XIST* exon1 remain to be tested.

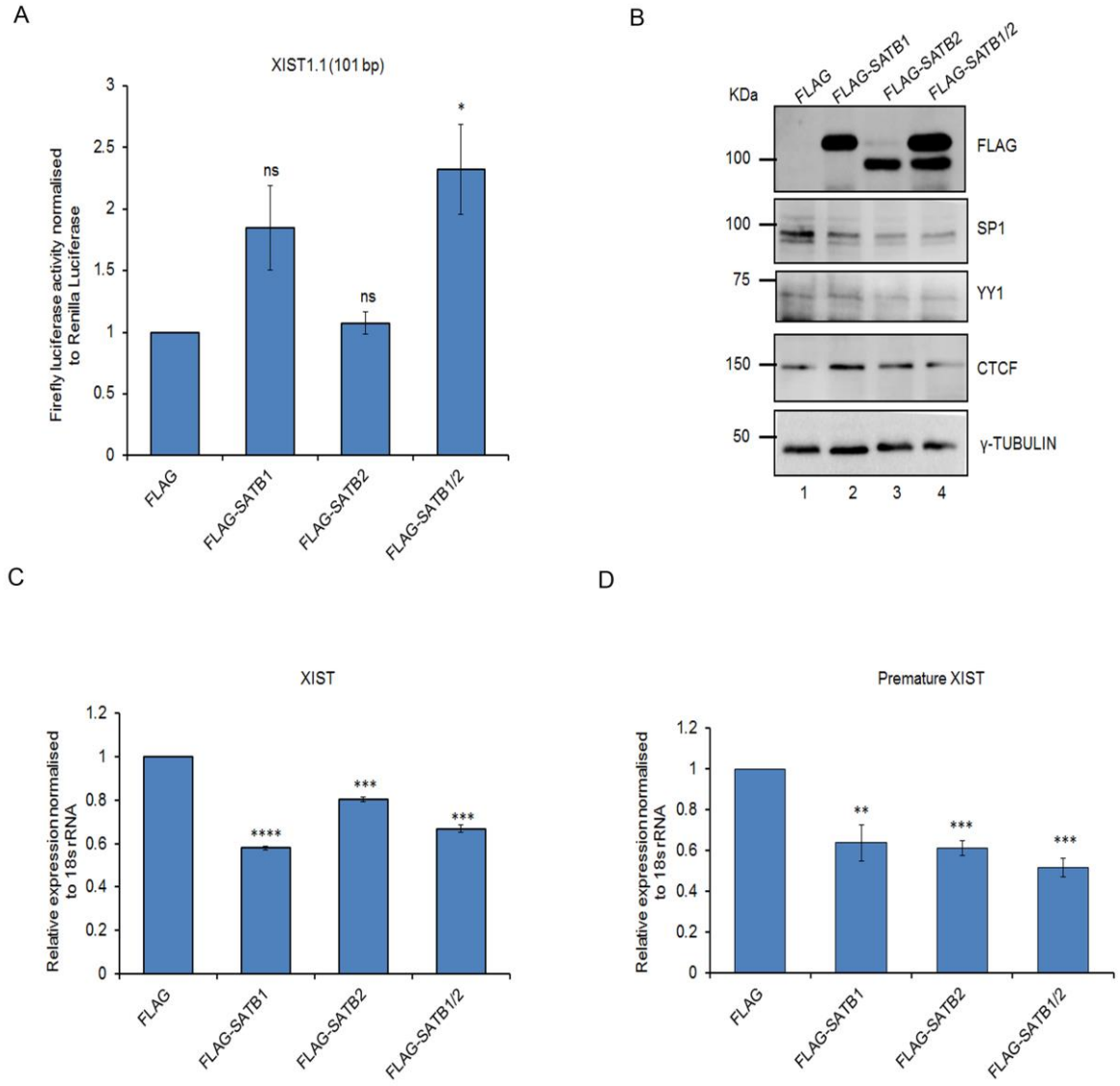


Figure 3.5.2 Effect of increasing the levels of SATB proteins on *XIST* minimal promoter activity and endogenous XIST levels.

(A) Firefly luciferase activity of XIST1.1 promoter construct measured upon overexpression of SATB1 and/or SATB2 in HEK 293T cells. Overexpression of both SATB1 and SATB2 leads to significant increase in the reporter activity as compared to FLAG vector control (compare Bar 4 with Bar1). Firefly luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis represent DNA transfected into the cells and Y-axis indicate normalized fold change values for the promoter constructs. Each bar represents values from three independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (* P value < 0.05, ns = non-significant).

(B) Immunoblotting to confirm overexpression of SATB1/2 as well as probing the expression of SP1, YY1, CTCF. γ-TUBULIN serves as a loading control. DNA transfected into HEK 293T cells is indicated on the top and the antibody used for immunoblotting is indicated on the side of western blot image.

(C-D) qRT-PCR for mature XIST (C) and premature XIST transcripts (D) using specific primers (see Appendix 1). Overexpression of either SATB1 and/or SATB2 leads to significant decrease in their levels as compared to the vector control (compare Bars 2, 3, 4 with 1 in (C) and (D)). X-axis represents DNA transfected and Y-axis indicatefold change normalized to 18s rRNA levels. Each bar represents values from three independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, *** P value < 0.001, ** P value < 0.01).

3.5.3 YY1 is the primary transcription factor regulating *XIST* promoter activity determined via biochemical assays

As described and discussed in the Chapter 1, SP1, YY1 and CTCF could be the potential transcription factors governing XIST expression. These three zinc-finger domain containing proteins are more or less ubiquitously expressed in various cell lines of different tissue origin including the cell line of our choice, HEK293T. Moreover, they were also found to be enriched on the promoter based on their ChIP-seq profiles from the ENCODE consortium (Figure 3.5.3 (A)), although it cannot be said with certainty whether they bind to the Xa or Xi since this is not allele-specific ChIP-seq.

Nevertheless, we sought to determine the promoter activities of 101 bp (XIST1.1), 1100 bp (XIST1) and 4458 bp region (XIST1-3) upon depletion of SP1/YY1/CTCF in HEK293T cells in order to (i) identify the key transcription factor(s) regulating XIST promoter in this cell line and (ii) determine if these factors act in a concerted manner. XIST1.1 and XIST1 constructs have the binding sites for SP1 and YY1, while XIST1-3 DNA construct harbours binding site for CTCF as well (Figure 3.5.3 (C), (D), (E)). Knockdown of each of these factors was validated by immunoblotting (Figure 3.5.3 (B)). The results from this experiment suggested that YY1 is the key factor responsible for transcriptional activation of *XIST* promoter since decreasing its levels by siRNA mediated knockdown, either singly or in combination with SP1 and/or CTCF resulted in the similar levels of decrease in the reporter activities from XIST1.1 and XIST1 promoter fragments as compared to the control siRNA (Compare Bars 3, 5, 7, 8 with Bar 1 in Figure 3.5.3 (C) and (D)). However, a different pattern was observed for XIST1-3 promoter construct which also harbours binding site for CTCF in addition to SP1 and YY1. Here, single knockdown of YY1 failed to cause a decline in the reporter activity (compare Bar 3 with Bar 1 in Figure 3.5.3 (E)), but instead its knockdown in combination with SP1 and/or CTCF caused reduction in the measured promoter activity

(Compare Bar 5, 7, 8 with Bar 1 in Figure 3.5.3 (E)). Interestingly, for XIST1-3 promoter fragment (4458 bp), even the single knockdown of CTCF led to a marginal decrease in the reporter activity which was not found to be the case for XIST1.1 (101 bp) and XIST1 (1100 bp) promoter clones (compare Bar 4 with Bar 1 in Figure 3.5.3 (E)). Importantly, there are no inherent differences in the reporter activities measured for the three promoter constructs employed in the current experiment (as shown in Figure 3.5.1). Also we do not observe enrichment of any specific histone modifications at or near the CTCF binding site in the ENCODE data set and it has predicted as an insulator element as per ChromHMM (Figure 3.5.3 (A)). The observed results suggest that the promoter element of *XIST* may not be limited to just the first 100 bp or so close to its TSS which is shown to be bound by SP1 and YY1. The promoter element tested in our study may actually be serving as a more complete promoter of *XIST* than the previously identified minimal promoter element.

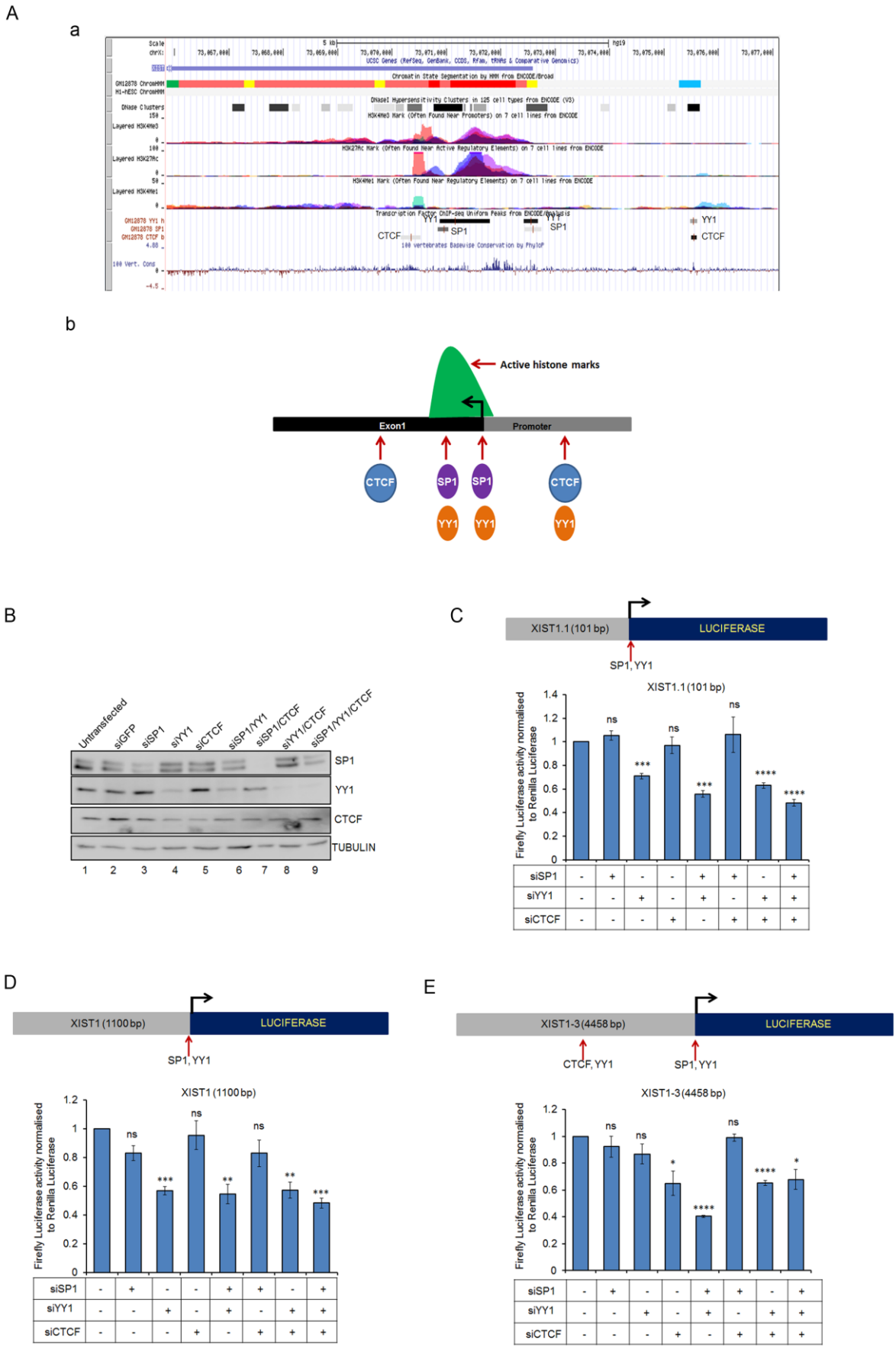

A
a
b
Active histone marks
Exon1
Promoter
CTCF
SP1
SP1
CTCF
YY1
YY1
YY1
B
Untransfected
siGFP
siSP1
siYY1
siCTCF
siSP1/YY1
siSP1/CTCF
siYY1/CTCF
siSP1/YY1/CTCF
SP1
YY1
CTCF
TUBULIN
1 2 3 4 5 6 7 8 9
C
XIST1.1 (101 bp)
LUCIFERASE
SP1, YY1
XIST1.1 (101 bp)
Firefly Luciferase activity normalised to Renilla Luciferase
siSP1
siYY1
siCTCF
D
XIST1 (1100 bp)
LUCIFERASE
SP1, YY1
XIST1 (1100 bp)
Firefly Luciferase activity normalised to Renilla Luciferase
siSP1
siYY1
siCTCF
E
XIST1-3 (4458 bp)
LUCIFERASE
CTCF, YY1
SP1, YY1
XIST1-3 (4458 bp)
Firefly Luciferase activity normalised to Renilla Luciferase
siSP1
siYY1
siCTCF

Figure 3.5.3 YY1 is the primary transcription factor regulating *XIST* promoter as determined biochemically.

(A) (a) UCSC genome browser track showing *XIST* genomic region (chrX:73,067,000-73,077,000) superimposed with promoter/enhancer specific histone modifications – H3K4me3 and H3K27ac, DNase protected regions, chromatin segmentation state for female lymphoblastoid cell line GM12878. Also shown is the ChIP-seq profile for SP1, YY1 and CTCF from GM12878 cell line. All the information displayed is generated by ENCODE consortium. (b) Simplified schematic depicting binding sites for SP1, YY1 and CTCF as well as histone modification profile as determined from the ENCODE data set.

(B) Immunoblotting for SP1, YY1, CTCF to validate siRNA mediated knockdown in HEK 293T cells. Western blotting for γ-TUBULIN and GAPDH serves as loading control. siRNA transfected is indicated on the top of the panel and the antibodies used for immunoblotting are indicated on the side of each panel.

(C-E) Firefly luciferase reporter activity for XIST1.1 (C), XIST1 (D) and XIST1-3 (E) promoter constructs upon knockdown of SP1/YY1/CTCF. For XIST1.1 and XIST1, only knockdown of YY1 causes significant reduction in the reporter activity as compared to control siRNA (compare Bars 3, 5, 7, 8 with Bar1 in (C) and (D)). A different pattern is observed for XIST1-3 construct (described in the text). The schematic of the reporter constructs along with the binding sites for each of these factors is depicted on the top of each graph. Firefly luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis represent siRNA transfected into the cells and Y-axis indicate normalized fold change values for the promoter constructs. Each bar represents values from three independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over control siRNA as per Student's T-test (****P value < 0.0001, *** P value < 0.001, ** P value < 0.01, * P value < 0.05, ns = non-significant).

The results discussed in 3.5.3 provided clues towards the role of CTCF in regulating the extended promoter of *XIST* gene as determined by measuring the promoter activity biochemically. In order to draw significance inference from these preliminary results, it is necessary to score for the levels of endogenous XIST lncRNA upon depletion of SP1, YY1 and CTCF. Few important facts to bear in mind before we discuss the findings from this experiment are the following :

(i) XIST is continuously synthesized throughout the cell cycle in all female derived tissues and cell lines (Jonkers et al., 2008; Ng et al., 2011)

(ii) It is pretty stable and the half-life of mature Xist transcripts in the absence of transcription is 4-6 hours (Ng et al., 2011).

(iii) The doubling time of the cell line of our choice - HEK 293T cells is ~ 20 hours.

Therefore, we decided to perturb the levels of the said factors by RNAi-mediated knockdown in HEK 293T cells. We chose 48 hour and 72 hour knockdown as two different time points for this experiment. The rationale for doing this is to identify and segregate the primary and secondary factors involved in regulating *XIST*. The published studies have highlighted the role of YY1 in governing transcription from the *XIST* promoter, which possibly could be the potential primary factor. However, the results from our study indicate

involvement of CTCF as well in the process. The results from the experiment are described in 3.5.4 and 3.5.5.

3.5.4 YY1 is the primary transcription factor regulating XIST expression in HEK 293T cells

Based on the results obtained in 3.5.3, we next sought to determine the effect of perturbing the levels of SP1/YY1/CTCF on the endogenous levels of XIST in HEK 293T cells. Towards this, the cells were transfected with siRNA targeting either SP1/YY/CTCF. Experimental scheme is depicted in the Figure 3.5.4 (A). 48 hour post transfection, the cells were harvested for RNA and protein. Successful knockdowns of SP1/YY1/CTCF were validated by immunoblotting and qPCR (Figure 3.5.4 (B)). Levels of endogenous XIST as well as premature XIST (the unspliced version of XIST) were determined by qRT-PCR as shown in Figure 3.5.4 (C) and (D). It was observed that depletion of YY1 led to a significant decrease in the levels of XIST as well as premature XIST. Interestingly, siRNA mediated knockdown of SP1 and/or CTCF actually led to a significant increment in the levels of premature XIST, which however returned to the levels close to control siRNA upon triple knockdown of SP1, YY1 and CTCF. It is noteworthy that YY1 levels increase upon knockdown of CTCF as per Figure 3.5.3 (B) and Figure 3.5.4 (B). Therefore, it is possible that the upregulation in premature XIST transcript levels observed may be an indirect effect of CTCF knockdown.

We also determined the levels of another lncRNA – JPX, which lies close to *XIST* gene on XIC and is known to escape XCI in mouse. Also, it has been demonstrated to be involved in positively regulating mouse *Xist* by relieving the repression mediated by CTCF on *Xist* promoter (Sun et al., 2013) (discussed in section 1.5.1.1 of Chapter 1). In the current experiment, we observed that the levels of JPX do not seem to change substantially except for upon knockdown of YY1 and CTCF together (Figure 3.5.4 (E)). Upon careful examination it was observed that XIST, premature XIST and JPX however followed a similar pattern of expression upon treating the cells with siSP1/CTCF, siYY1/CTCF and siSP1/YY1/CTCF (Bars 6, 7, 8 in Figure 3.5.4 (C), (D) and (E)). Besides this, it is notable that the findings from this experiment are not in accordance with that obtained by the reporter assays (3.5.3 (C-E)) except for the effect observed upon depletion of YY1. From the current results as well as those discussed in Chapter 2, it is apparent that transcriptional outcome from the *XIST* promoter is not solely under the control of region(s) upstream of TSS of the gene. We think

that the upstream promoter region collaborates with certain regions on the exon1 which confers the regulatory potential and assists in controlling the transcription fate of *XIST*. The evidence for the same were presented in the previous chapter and we provide more results towards the end of this chapter to support our hypothesis. Overall, we can conclude from this experiment that YY1 may be acting as a transcriptional activator while SP1 and CTCF may be playing a role of transcriptional repressors either directly or indirectly. And more importantly, it is the balance of these factors that might be crucial in deciding the fate of transcription from *XIST* promoter.

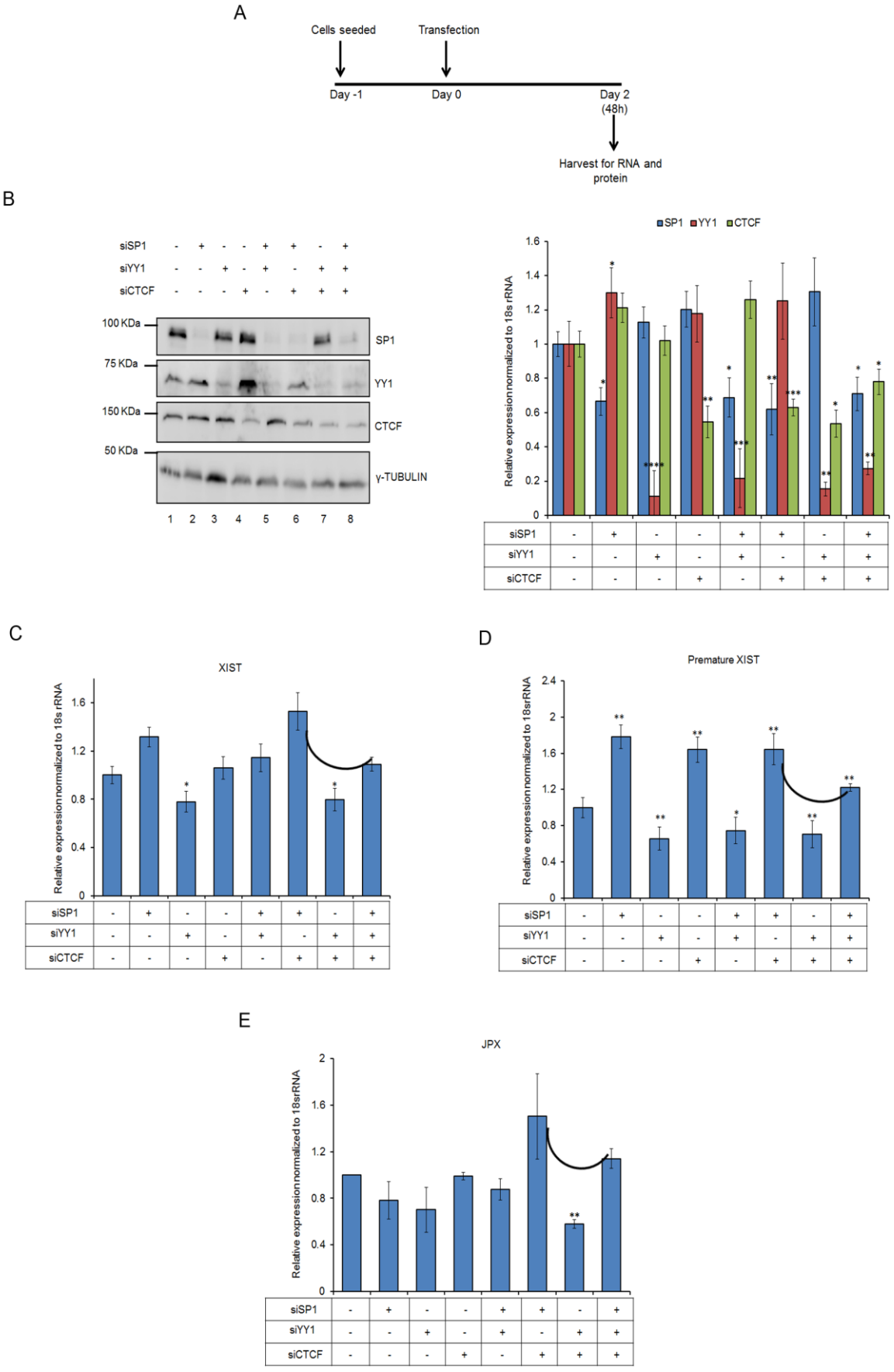

A
Cells seeded
Transfection
Day -1
Day 0
Day 2 (48h)
Harvest for RNA and protein
B
siSP1
siYY1
siCTCF
100 KDa
75 KDa
150 KDa
50 KDa
SP1
YY1
CTCF
γ-TUBULIN
1 2 3 4 5 6 7 8
SP1 YY1 CTCF
Relative expression normalized to 18s rRNA
C
XIST
Relative expression normalized to 18s rRNA
D
Premature XIST
Relative expression normalized to 18srRNA
E
JPX
Relative expression normalized to 18srRNA

Figure 3.5.4 YY1 is the primary transcription factor regulating *XIST* expression in HEK 293T cells.

(A) Experimental scheme followed for performing siRNA mediated knockdown of SP1/YY1/CTCF in HEK 293T cells for 48 hours.

(B) Validation of knockdown by immunoblotting and qRT-PCR. For immunoblotting, γ-TUBULIN serves as loading control. siRNA transfected is indicated on the top and the antibodies used are indicated on the side of the western blot image. qRT-PCR was performed using SP1, YY1, CTCF specific primers (see Appendix 1). X-axis represent siRNA and Y-axis indicate fold change normalized to 18s rRNA levels. Each bar represents values from three independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over control siRNA as per Student's T-test (****P value < 0.0001, *** P value < 0.001, ** P value < 0.01, * P value < 0.05).

(C-E) qRT-PCR of mature XIST (C), premature XIST (D) and JPX (E) using specific primers (see Appendix 1). XIST and premature XIST levels significantly decrease upon knockdown of YY1 (compare Bars 3, 5, 7 with Bar1 in (D)), whereas depletion of SP1 and CTCF lead to marginal increase in premature XIST levels (compare Bars 2, 4, 6 with Bar1 in (D)). Expression of JPX does not vary significantly except for upon dual knockdown of YY1 and CTCF (compare Bar 7 with Bar 1 in (E)). X-axis represent siRNA and Y-axis indicate fold change normalized to 18s rRNA levels. Each bar represents values from three independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over control siRNA as per Student's T-test (****P value < 0.0001, *** P value < 0.001, ** P value < 0.01, * P value < 0.05).

3.5.5 SP1 and CTCF are also potential transcription factors regulating *XIST* expression in addition to YY1

As can be observed from the results discussed in section 3.5.4, the depletion of SP1 or YY1 or CTCF for 48 hour (~2.4 cell cycles) did not exhibit a remarkable effect on the levels of mature XIST. Therefore, we next performed the similar experiment of siRNA mediated knockdown of SP1/YY1/CTCF to determine their effect on XIST levels, except that this time reverse transfection was carried out and the cells were harvested for RNA and protein after 72 hourours of transfection (~3.6 cell cycle). The experimental scheme is outlined in Figure 3.5.5 (A). Successful knockdowns of SP1, YY1 and CTCF were validated by immunoblotting and qRT-PCR (Figure 3.5.5 (B)). Levels of endogenous XIST, premature XIST as well as JPX were determined by qRT-PCR. As can be seen from the Figure 3.5.5 (C), (D), (E), all the three transcripts tested showed a significant decline in their expression levels. Interestingly, CTCF levels show significant reduction upon depletion of SP1 and/or YY1 as well. This set of results indicate that SP1 and CTCF can also act as a potential transcription activators for XIST expression, either directly by binding to *XIST* regulatory region or indirectly by influencing JPX expression. However, we did not observe any additive effect on the levels of XIST (both spliced and unspliced) upon perturbing the levels of all the three proteins in question. But the levels of JPX seem to decline in an additive manner upon single, double and triple knockdowns of SP1, YY1 and CTCF. Also, it can be appreciated that while the levels of knockdown of SP1 and YY1 are comparable to that observed in Figure 3.5.4(B), expression of

CTCF shows a greater reduction here. Altogether, based on the results obtained in 3.5.4 and 3.5.5, it can be stated that YY1 seems to be the primary transcription activator regulating *XIST* promoter. The effect observed upon disturbing the levels of SP1 and CTCF could be secondary in nature and further experimentation is needed to carefully look into their roles in the process.

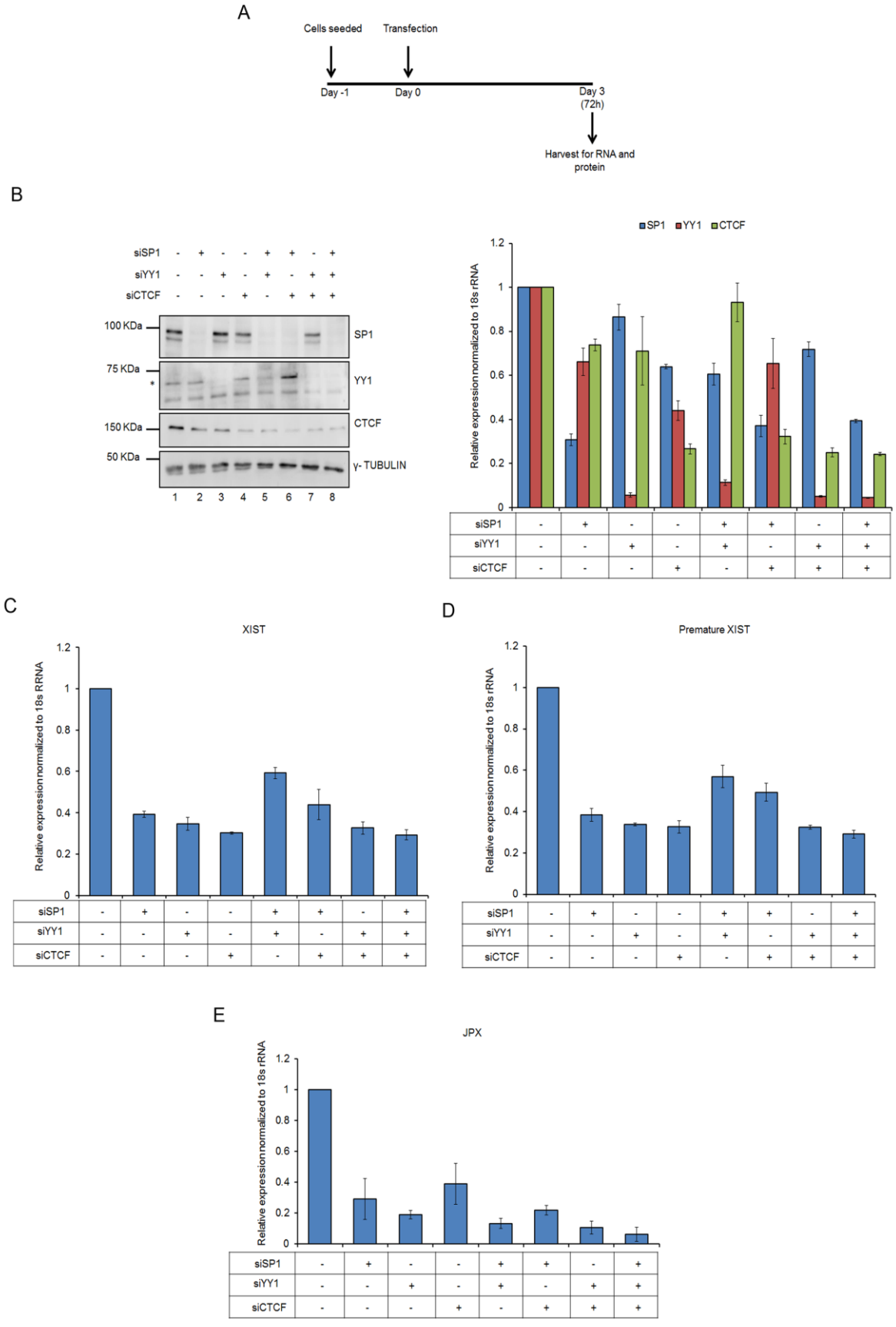

A
Cells seeded
Transfection
Day -1
Day 0
Day 3 (72h)
Harvest for RNA and protein
B
siSP1
siYY1
siCTCF
100 KDa
75 KDa
150 KDa
50 KDa
SP1
YY1
CTCF
γ- TUBULIN
1 2 3 4 5 6 7 8
SP1 YY1 CTCF
Relative expression normalized to 18s rRNA
C
XIST
Relative expression normalized to 18s RRNA
D
Premature XIST
Relative expression normalized to 18s rRNA
E
JPX
Relative expression normalized to 18s rRNA

Figure 3.5.5 SP1 and CTCF are also potential transcription factors regulating XIST expression in addition to YY1.

(A) Experimental scheme followed for performing siRNA mediated knockdown of SP1/YY1/CTCF in HEK 293T cells for 72 hours.

(B) Validation of knockdown by immunoblotting and qRT-PCR. For immunoblotting, γ-TUBULIN serves as a loading control. siRNA transfected is indicated on the top and the antibodies used are indicated on the side of the western blot image. qRT-PCR was performed using SP1, YY1, CTCF specific primers (see Appendix 1). X-axis represent siRNA and Y-axis indicate fold change normalized to 18s rRNA levels.

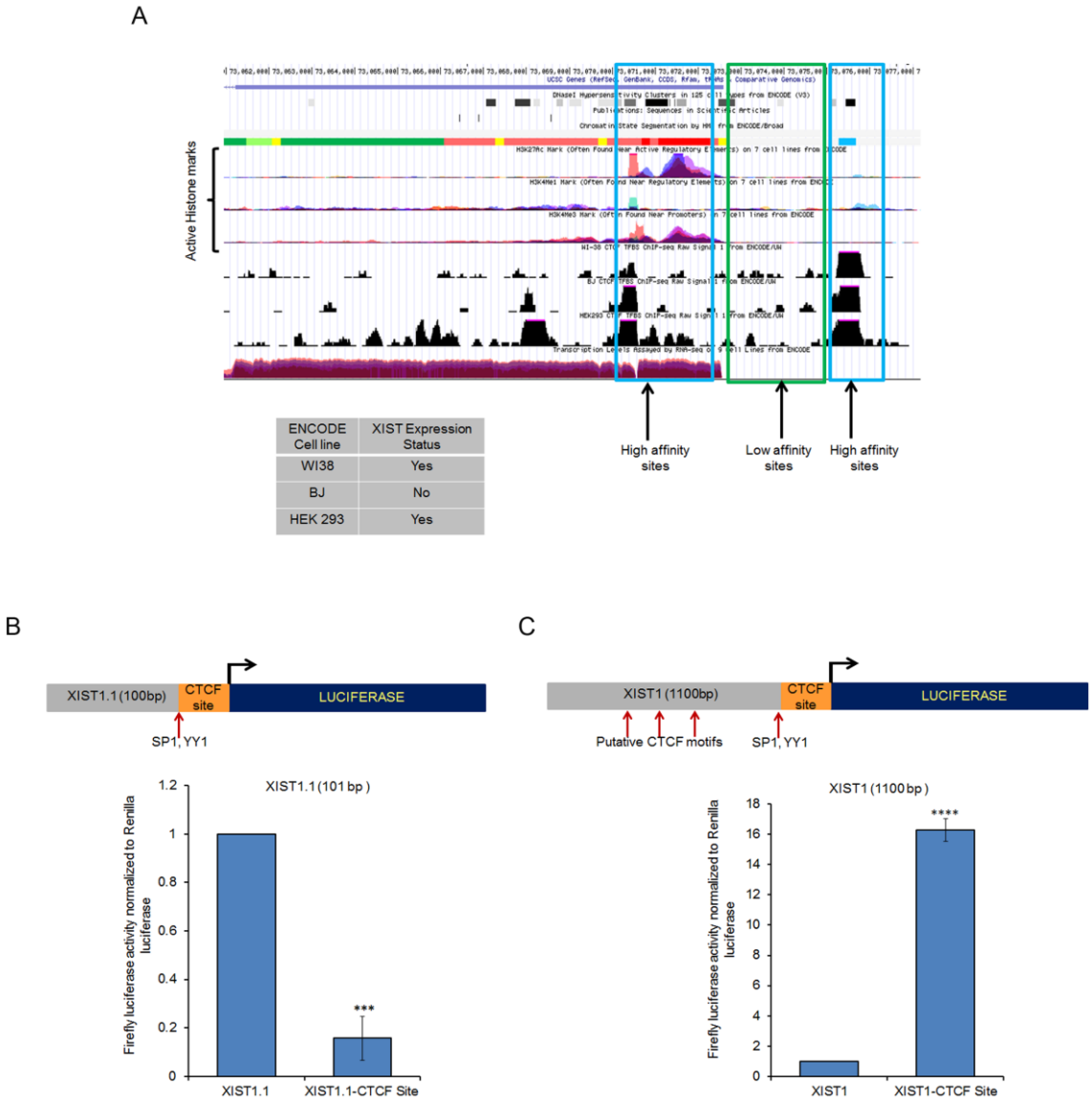
(C-E) qRT-PCR of mature XIST (C), premature XIST (D) and JPX (E) using specific primers (see Appendix 1). Levels of all the three transcripts decrease significantly upon depletion of SP1/YY1/CTCF as compared to control siRNA transfected cells. X-axis represent siRNA and Y-axis indicate fold change normalized to 18s rRNA levels.

3.5.6 CTCF binding affinities to the *XIST* genomic locus may be an important determinant of XIST expression

In order to understand the function of CTCF in terms of its role in regulating *XIST* promoter, we made use of publicly available ENCODE CTCF ChIP-seq data set from HEK293, primary male fibroblast cell line BJ and primary female fibroblast cell line WI38. As shown in the UCSC genome browser track in Figure 3.5.6 (A), there are multiple CTCF binding sites upstream of TSS as well as downstream into the 1st exon of *XIST*. For simplicity, we have categorized these sites into low affinity and high affinity sites solely on the basis of the height of the ChIP-seq peaks observed. One major caveat however is that it cannot be determined whether CTCF peaks observed in HEK293 and WI 38 cells are the reads from active or inactive X. Based on the results discussed in 3.5.4 and 3.5.5, we hypothesized that the combinatorial binding of CTCF to these sites is essential in deciding the transcriptional outcome from *XIST* promoter. Elaborating on this, it is plausible that upon siRNA mediated knockdown of CTCF in HEK 293T cells for 48 hours, CTCF may have lost its occupancy from certain such low affinity sites resulting in increased levels of premature XIST. However, upon depletion of CTCF for 72 hours, even the high affinity sites may lose CTCF binding, thereby causing a reduction in the levels of XIST.

In order to test this hypothesis, we cloned the region exhibiting high affinity CTCF binding site, located ~ 2.5 Kb downstream of *XIST* TSS between the minimal promoter region (XIST1.1) and luciferase as well as between XIST1 promoter element and luciferase gene so as to retain the *XIST* genomic context as closely as possible. This region is labelled as CTCF site. Next we transfected XIST1.1 and XIST1 DNA constructs, with and without CTCF site into HEK 293T cells along with Renilla luciferase vector for 48 hour. Upon biochemically

measuring the amount firefly luciferase gene product, it was observed that while the fusion of CTCF site to XIST1.1 minimal promoter element led to a significant decrease (Figure 3.5.6 (B)), the same region increased the promoter activity manifold when put together with XIST1 DNA construct (Figure 3.5.6 (C)). It is noteworthy that the promoter activities of XIST1.1 and XIST1 DNA construct are more or less identical as shown in Figure 3.5.1. These results provide preliminary support to our hypothesis that the combination of high affinity and low affinity binding mode of CTCF to the sites present on *XIST* genomic locus, more precisely on the 1st exon and putative promoter region, can be a potential determinant of *XIST* promoter activity aside from the role of YY1.



ENCODE Cell line	XIST Expression Status
WI38	Yes
BJ	No
HEK 293	Yes

Figure 3.5.6. CTCF binding affinities to the *XIST* genomic locus may be an important determinant of XIST expression.

(A) UCSC genome browser snapshot depicting CTCF ChIP-sequencing peaks from diploid female fibroblast WI38, diploid male fibroblast BJ and immortalized cells HEK 293 on *XIST* genomic locus. The status of XIST expression in these cell lines is indicated in the table below. Blue and green rectangles indicate high and low affinity CTCF binding sites respectively. Also shown are active histone modification – H3K27ac, H3K4me, H3K4me3 and ChromHMM profiles generated by ENCODE.

(B-C) Firefly luciferase activities for XIST1.1 and XIST1 promoter constructs with or without the CTCF site cloned between the promoter element and luciferase gene. The schematic of the reporter constructs along with the binding sites for SP1, YY1, CTCF is depicted on the top of each graph. Firefly luciferase activities represented here are normalized to Renilla luciferase activity which serves as an internal control. X-axis represent the constructs transfected into HEK 293T cells and Y-axis indicate normalized fold change values for each construct. Each bar represents values from three independent experiments. Error bar represents ±S.E.M. Asterisks

represent the significance over XIST1.1 or XIST1 control as per Student's T-test (****P value < 0.0001 and *** P value < 0.001).

3.6 Discussion

Through this study we sought to determine transcriptional regulation of human *XIST* gene in the context of maintenance phase of XCI. For a long time it was widely thought that XIST RNA is required only during the initiation and establishment phases of XCI and dispensable for maintaining the inactive X in the somatic cells (Brown and Willard, 1994; Csankovszki et al., 1999). As discussed in Chapter 1, coating by Xist RNA ensues a dramatic epigenetic reprogramming of the chosen Xi brought about by the coordinated action of the histone and DNA modifying machineries. Accumulation of such repressive alterations of Xi chromatin have shown to be stably maintained and inherited during subsequent cell divisions thereby maintaining the inactive X in the somatic cells (Barakat et al., 2010; Blewitt et al., 2008; Casas-Delucchi et al., 2011; Chan et al., 2011; Hernández-Muñoz et al., 2005; Jeppesen and Turner, 1993; Li et al., 2012; Marks et al., 2009; Mermoud et al., 1999). Notably, work from Jeannie Lee's group has elegantly demonstrated that Xist RNA is required for targeting Xi to the perinucleolar position in the maintenance phase (Zhang et al., 2007). In addition to this, the authors also determined that the retention of heterochromatin marks on Xi is affected upon loss of Xist RNA thereby leading to reactivation of X-linked genes. This study convincingly argues that Xist RNA after all is actually necessary for maintaining stable silencing of Xi in the somatic cells. Whether the observed alterations in the chromatin state of Xi is the consequence of loss of Xist RNA or due to its dissociation from the nucleus still remains an open question. In an attempt to identify the protein interactome of Xist RNA, the same group established that Xist RNA actually repels cohesin and thereby possibly evades acquisition of transcriptionally-permissive state (Minajigi et al., 2015). Both these finding put together provide a compelling mechanistic basis to the role of Xist RNA in preferentially targeting Xi to the perinucleolar position in the nucleus so as to possibly avoid cohesin molecules and hence retain stable silencing of Xi. However, our understanding about the regulation of *Xist*, especially human *XIST* gene, in the maintenance phase nucleus is still lacking.

As evident from the literature discussed throughout the thesis, most of our understanding about the process of XCI as well as *Xist* regulation has been from the studies

carried out using mouse or mESCs as a model system. Through the work described herewith, we have attempted to characterize human *XIST* gene promoter in terms of its transcriptional regulation during the initiation as well as maintenance phases of XCI.

In the current chapter, we have presented results with regards to *XIST* promoter regulation in the context of maintenance phase of XCI using HEK 293T cells as a model cell line system. The first ever study attempting to identify and characterize human XIST promoter was published way back in 1997. The authors in this study not only identified the minimal promoter region of *XIST* but also discovered binding sites for SP1, YY1 and TBP transcription factors by *in vitro* biochemical assays (Hendrich et al., 1997). Since the promoter of *XIST* is not very well conserved across species except for the first 100 bp or so and the fact that upregulation of Xist/XIST follows distinct pattern in a variety of species (discussed in Chapter 2), we also tested the activity of sequences farther away from the TSS, ranging from 101 bp, to 4458 bp upstream to identify the promoter elements regulating *XIST*. Similar to the first report on human *XIST* promoter characterization (Hendrich et al., 1997), we also observed that the region conserved across several eutherian species, encompassing 101 bp including a fraction of XIST 5' and the region upstream of XIST TSS is sufficient to drive the transcription and can be called the minimal promoter of *XIST*. Moreover, the promoter activity remains more or less constant for all the upstream elements tested.

Although it has been a common knowledge for a long time that SP1, YY1 and TBP bind to *XIST* minimal promoter region and mutating their binding sites lead to decrease in the promoter activities measured biochemically (Hendrich et al., 1997), work to decipher their roles in the cellular context was not followed up. Moreover, these three factors as well as CTCF also showed enrichment on *XIST* promoter region in a variety of cell lines as per their ChIP-sequencing profiles from ENCODE consortium. Therefore, we sought to address the roles of three zinc-finger proteins – SP1, YY1 and CTCF in the cell system by assaying for the promoter activities biochemically as well as determining the expression of endogenous XIST upon perturbing their levels. Based on these experiments, it was observed that YY1 is the key transcription factor regulating *XIST* transcription. Interestingly, even though depletion of SP1 and/or CTCF failed to show any appreciable effect as per the luciferase reporter assays of the XIST1.1 (101 bp) and XIST1 (1100 bp) promoter constructs, both of which harbour

binding sites for SP1 and YY1, we did observe a distinct pattern for XIST1-3 (4458 bp) promoter construct which possesses a conserved CTCF binding motif and also displays enrichment of CTCF in a variety of cell lines as per ENCODE data. The results from this experiment seemed to indicate that these factors might be acting independent of each other since we did not observe any additive effect on the measured promoter activities upon perturbing the levels of all the 3 factors by RNAi mediated knockdown.

To gain further insights into their roles, we next decided to score for the expression of endogenous XIST upon depletion of SP1/YY1/CTCF in HEK 293T cells. throughout the cell cycle (Jonkers et al., 2008; Ng et al., 2011) It was observed that when transient siRNA mediated knockdown was performed for 48 hours, only reduction in YY1 levels showed decrement in premature as well as mature XIST transcript levels as determined by qRT-PCR, whereas knockdown of SP1 and/or CTCF caused an increment. However, when the same experiment of siRNA mediated knockdown was carried out for 72 hours, a profile distinct from the previous experiment was observed. The results from this experiment showed that XIST levels significantly decrease upon prolonged depletion of any of the three factors. In mouse system, it has been shown that YY1 not only regulates *Xist* transcription but also actually binds to Repeat C of Xist lncRNA and helps in tethering it to the chromosome from which it is synthesized (Jeon and Lee, 2011; Makhlouf et al., 2014). Therefore, we also scored for the expression of premature XIST transcript levels to be able to probe for the effect on XIST transcription and observed that it also followed the trend similar to that of the mature transcript. Additionally, we also assessed the levels of another X-linked lncRNA, JPX, which in mouse has been shown to be an important positive regulator of Xist (Sun et al., 2013; Tian et al., 2010). Our results demonstrate that the pattern of expression of JPX upon perturbing the levels of SP1/YY1/CTCF is similar to that observed for XIST. Interestingly, JPX is actually co-regulated by SP1, YY1 and CTCF because its levels progressively decline upon single, double and triple knockdown of these factors. It would be interesting to determine if JPX, like in mouse, acts by titrating out CTCF molecules and helping in positively regulating *XIST* promoter. It is also known that SP1, YY1 and CTCF interact with each other. However, atleast in our study, we can rule out the possibility of their co-recruitment on the *XIST* promoter since we did not observe any additive effect upon depleting any two or even all three of them together. An important conclusion that can be made from the results described is that

SP1, YY1 and CTCF play parallel roles in regulating *XIST* transcription, with YY1 being the primary factor.

In order to expand our understanding on the regulation of *XIST*, we looked into the publicly available ChIP-sequencing profiles of CTCF generated by ENCODE programme using male as well as female cell lines (BJ, WI38, HEK 293) and observed that CTCF showed enrichment over a wide region of *XIST* genomic locus, both upstream as well as downstream of *XIST* TSS. We categorised these sites into high-affinity and low-affinity binding sites solely on the basis of the height of ChIP-seq peaks. An important limitation to be considered is that it is not possible to determine whether CTCF occupies the active X (XIST repressed) or inactive X (XIST expressed) on the basis of these ChIP-seq data since it is not allele-specific sequencing. However, upon closer examination it can be observed that there are multiple such low-affinity sites bound by CTCF exclusively in female cells (WI38 and HEK 293) and absent from male fibroblast cell line BJ. The strong binding sites are present ~ 2.5 Kb upstream as well as downstream of TSS. In Chapter 2, we have shown that CTCF displays manifold enrichment over one of these high affinity sites located ~ 2.5 Kb into the exon 1 of XIST. Interestingly, this site also shows enrichment for active histone modifications such as H3K4me3 and H3K27ac. In order to understand the regulatory role of CTCF binding to this region, we made reporter constructs with this site incorporated between XIST1.1 or XIST1 promoter element and the firefly luciferase gene. Interestingly, we observed that while this particular site led to a significant reduction in the minimal promoter activity, the same site along with 1.1 Kb long promoter fragment caused a significant increment in the promoter activity. It is noteworthy that there are no inherent differences in the activities of just the minimal promoter (101 bp) and a longer promoter fragment (1.1 Kb) otherwise.

We made a similar observation in the previous chapter using EC cells. This finding can be explained by looking into the differences between XIST1.1 and XIST1 constructs. As discussed in the previous chapter as well, XIST1.1 promoter element does not exhibit any CTCF occupancy, while XIST1 has multiple low-affinity CTCF binding sites. Therefore, we think that a strong CTCF binding site when present alongside these low affinity sites changes the transcriptional outcome. The discrepancy in the result with regards to endogenous XIST levels upon depletion of CTCF for 48 hour and 72 hour can also be explained by same logic. Elaborating on this, we think that when siRNA mediated knockdown of CTCF is allowed for

48 hours, CTCF loses its occupancy from the low-affinity sites but when the knockdown is performed for 72 hours, then CTCF might lose its occupancy even from such high-affinity sites, thereby causing a reduction in the expression of XIST as opposed to marginal increase observed in the former case. On the basis of these findings, we hypothesize that binding of CTCF both upstream as well as downstream of *XIST* TSS plays an important role in governing the transcription from *XIST* promoter. Interestingly, the region on XIST exon1 (~+2.5 Kb from TSS) as well as upstream of the TSS (~ -2.5 Kb from TSS) with strong CTCF binding site also shows occupancy of RAD21 (RAD21 ChIP-seq from ENCODE and supplementary information from Makhlouf et al., 2014). Both CTCF and RAD21 have been long known to be involved in chromatin looping interactions (Splinter et al., 2006). Therefore, it can be speculated that these factors may possibly be involved in a looping interaction between the promoter and gene body of *XIST*, aiding in the enrichment of specific histone marks and hence enabling the transcriptional control from the promoter.

While we were pursuing this question, a couple of recent reports uncovered the role of transcription factor YY1 in positively regulating transcription from *XIST* promoters (P1 and P2) (Chapman et al., 2014; Makhlouf et al., 2014). Both these studies however ruled out the role of CTCF as a regulator of *XIST* since they did not observe any differential enrichment of CTCF between male and female fibroblast lines or XIST+ and XIST- hESCs. This disparity in our and published results can be solved by performing chromosome conformation capture assay in male and female cells to delineate if these multiple CTCF sites are involved in looping mediated interaction and differential regulation.

Collectively, the results presented in this chapter provide evidence for the involvement of parallel pathways controlled by SP1, YY1 and CTCF in controlling transcription from *XIST* promoter in the context of maintenance phase of XCI.

3.7 References

Barakat, T.S., Jonkers, I., Monkhorst, K., and Gribnau, J. (2010).X-changing information on X inactivation. Exp. Cell Res. *316*, 679–687.

Blewitt, M.E., Gendrel, A.-V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat. Genet. *40*, 663–669.

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

Buzin, C.H., Mann, J.R., and Singer-Sam, J. (1994). Quantitative RT-PCR assays show Xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. Development *120*, 3529–3536.

Casas-Delucchi, C.S., Brero, A., Rahn, H.-P., Solovei, I., Wutz, A., Cremer, T., Leonhardt, H., and Cardoso, M.C. (2011). Histone acetylation controls the inactive X chromosome replication dynamics. Nat. Commun. *2*, 222.

Chan, K.M., Zhang, H., Malureanu, L., van Deursen, J., and Zhang, Z. (2011). Diverse factors are involved in maintaining X chromosome inactivation. Proc. Natl. Acad. Sci. *108*, 16699–16704.

Chapman, A.G., Cotton, A.M., Kelsey, A.D., and Brown, C.J. (2014). Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. BMC Genet.*15*, 89.

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

Donohoe, M.E., Zhang, L.F., Xu, N., Shi, Y., and Lee, J.T. (2007). Identification of a Ctcf Cofactor, Yy1, for the X Chromosome Binary Switch. Mol. Cell *25*, 43–56.

Hendrich, B.D., Plenge, R.M., and Willard, H.F. (1997). Identification and characterization of the human XIST gene promoter: Implications for models of X chromosome inactivation. Nucleic Acids Res. *25*, 2661–2671.

Hernández-Muñoz, I., Lund, A.H., van der Stoop, P., Boutsma, E., Muijrers, I., Verhoeven, E., Nusinow, D.A., Panning, B., Marahrens, Y., and van Lohuizen, M. (2005). Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. Proc. Natl. Acad. Sci. U. S. A. *102*, 7635–7640.

Jeon, Y., and Lee, J.T. (2011). YY1 Tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

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

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA Is Confined to the Nuclear Territory of the Silenced X Chromosome throughout the Cell Cycle. Mol. Cell. Biol. *28*, 5583–5594.

Li, Y., Tan, T., Zong, L., He, D., Tao, W., and Liang, Q. (2012).Study of methylation of histone H3 lysine 9 and H3 lysine 27 during X chromosome inactivation in three types of cells.Chromosom. Res. *20*, 769–778.

Makhlouf, M., Ouimette, J.-F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014).A prominent and conserved role for YY1 in Xist transcriptional activation. Nat. Commun. *5*, 1–12.

Marks, H., Chow, J.C., Denissov, S., Françoijs, K.-J., Brockdorff, N., Heard, E., and Stunnenberg, H.G. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res. *19*, 1361–1373.

Mermoud, J.E., Costanzi, C., Pehrson, J.R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J. Cell Biol. *147*, 1399–1408.

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., et al. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science*349*, 1–19.

Ng, K., Daigle, N., Bancaud, A., Ohhata, T., Humphreys, P., Walker, R., Ellenberg, J., and Wutz, A. (2011).A system for imaging the regulatory noncoding Xist RNA in living mouse embryonic stem cells.Mol. Biol. Cell *22*, 2634–2645.

Kumar, P., Purbey, P.K., Sinha, C.K., Notani, D., Limaye, A., Jayani, R.S., and Galande, S. (2006). Phosphorylation of SATB1, a Global Gene Regulator, Acts as a Molecular Switch Regulating Its Transcriptional Activity In Vivo. Mol. Cell *22*, 231–243.

Purbey, P.K., Singh, S., Kumar, P.P., Mehta, S., Ganesh, K.N., Mitra, D., and Galande, S. (2008). PDZ domain-mediated dimerization and homeodomain-directed specificity are required for high-affinity DNA binding by SATB1. Nucleic Acids Res. *36*, 2107–2122.

Purbey, P.K., Singh, S., Notani, D., Kumar, P.P., Limaye, A.S., and Galande, S. (2009). Acetylation-dependent interaction of SATB1 and CtBP1 mediates transcriptional repression by SATB1. Mol. Cell. Biol. *29*, 1321–1337.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and De Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the B-globin locus. Genes Dev. *20*, 2349–2354.

Sun, B.K., Deaton, A.M., and Lee, J.T. (2006). A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol. Cell *21*, 617–628.

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

Sunwoo, H., Wu, J.Y., and Lee, J.T. (2015). The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. Proc. Natl. Acad. Sci. U. S. A. *112*, E4216-25.

Tian, D., Sun, S., and Lee, J.T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell *143*, 390–403.

Wutz, A., and Jaenisch, R. (2000). A Shift from reversible to irreversible x inactivation is triggered during ES cell differentiation. Mol. Cell *5*, 695–705.

Zhang, L.F., Huynh, K.D., and Lee, J.T. (2007). Perinucleolar targeting of the inactive x during S phase: Evidence for a role in the maintenance of silencing. Cell *129*, 693–706.

Chapter 4: Conclusions and future directions

Based on the work presented in the thesis, we can arrive at the conclusions discussed below. These findings not only address some of the key questions of the field but also lead to new lines of investigations.

Our work demonstrates that the promoter of *XIST* may not be limited to the first 100 bp upstream of TSS as previously reported. In fact the results obtained in our study suggest it to extend upto ~ 4.5 Kb upstream of XIST TSS. The extended promoter region does not show conservation across several eutherian species except for the conserved CTCF binding motif. Based on our findings, we hypothesize an important role for this particular site in influencing *XIST* promoter behaviour.

The results obtained in our study provide compelling evidence towards the involvement of *XIST* exon1 region in controlling transcription from its promoter. More specifically, we demonstrate for the first time that the pluripotency factors (OCT4, SOX2 and NANOG) and the chromatin organizer proteins (CTCF, SATB1 and SATB2) exhibit a unique pattern of enrichment over two distinct sites on the XIST exon1 region (~2.5 Kb and ~ 5.5 Kb downstream of TSS) in the EC cells which provide the context for initiation phase of XCI. We think that this can possibly have an impact on the transcriptional activity from XIST promoter. Additionally, we also discover the pluripotency factors – OCT4 and SOX2 can act as repressors of *XIST*. Whether these transcriptional regulators can govern the deposition of active/repressive histone marks across *XIST* promoter is an interesting question to be addressed. This can be answered either by scoring for the promoter specific histone marks, both active (H3K4m3, H3K27ac) and inactive (H3K9me3) upon either perturbing the levels of the said factors or more cleanly by deleting their binding sites using CRISPR-Cas9 tool.

Our work also shows that the three zinc-finger transcription regulators – SP1, YY1 and CTCF are important factors governing XIST transcription in the initiation as well as maintenance phases of XCI. We identify YY1 to be the primary regulator of *XIST* promoter. Interestingly, CTCF seems to be playing a unique role in the aspect. It displays strong and weak modes of binding to multiple sites on the promoter as well as the gene body of *XIST*. Our work highlights the significance of this observation and suggests that a combinatorial

usage of these multiple CTCF binding motifs can actually be involved in a looping mediated interaction between *XIST* exon1 and its promoter, thereby influencing the promoter firing. It would be interesting to determine the existence of such looping interactions mediated by CTCF during the initiation as well as maintenance phases of XCI by employing chromosome conformation capture based techniques in female and male cell lines. Furthermore, it would be important to understand the influence of these loops (if any) on the establishment of specific epigenetic features across XIST genomic locus including promoter region.

Collectively, our study convincingly demonstrates the involvement of a myriad of factors in regulating transcription from *XIST* promoter during the initiation and maintenance phases of XCI. It would be interesting to decipher the complex regulatory cascade mediated by these various factors in ensuring timely activation and persistence of XIST expression.

Appendix 1

Details of primers used for quantitative real time PCR of cDNA

Description	Sequence (5' to 3')
hSATB1 Forward	ACCAGTGGGTACGCGATGA
hSATB1 Reverse	TGTTAAAAGCCACACGTGCAA
hSATB2 Forward	CCAGAGCACATTAGCCAAAGA
hSATB2 Reverse	TGTGCTATTTACAATGGATGAAATC
hOCT4 Forward	AGCAAAACCCGGAGGAGT
hOCT4 Reverse	CCACATCGGCCTGTGTATATC
hNANOG Forward	CCTGAACCTCAGCTACAAACAG
hNANOG Reverse	GCTATTCTTCGGCCAGTTGT
hPAX6 Forward	GGCACACACACATTAACACACTT
hPAX6 Reverse	GGTGTGTGAGAGCAATTCTCAG
hSP1 Forward	GCCCCAGGTGATCATGGA
hSP1 Reverse	CTGGGCTGTTTTCTCCTTCCT
hYY1 Forward	TGGAGAGAACTCACCTCCTGA
hYY1 Reverse	TCTTTAATTTTTCTTGGCTTCATTC
hCTCF Forward	GGAAGGTGATGCAGTCGAAG
hCTCF Reverse	GTATCGTCCACAGCAGCCTC
hXIST Forward	GATGTCAAAAGATCGGCCCA
hXIST Reverse	CAAGAGGAGCCTAAGGAGAC
Human Premature XIST Forward	TGCTTTAGCATCAAAGCCCT
Human Premature XIST Reverse	GCCTTAGATTCCCAGTTCCA
hJPX Forward	CTCACCCTCTGCCTCCTGAC
hJPX Reverse	TGACACCAGAACACCATCATG
hGAPDH Forward	CTGCACCACCAACTGCTTAG
hGAPDH Reverse	GTCTTCTGGGTGGCAGTGAT
h18s rRNA Forward	CGCCGCTAGAGGTGAAATTCT
h18s rRNA Reverse	CGAACCTCCGACTTTCGTTCT

Appendix 2

Details of primers used for cloning

Description	Sequence (5' to 3')	RE site	Vector
hXIST1.1 Forward	ACGAGATCTCTGCAGCAGCGAATTGCAG	BglII	pGL3 Basic
hXIST1.1 Reverse	GCTGAGCTCCAAAGATGTCCGGCTTTC	SacI	pGL3 Basic
hXIST1.2 Forward	ACGAGATCTCTGCAGCAGCGAATTGCAG	BglII	pGL3 Basic
hXIST1.2 Reverse	GCTGAGCTCGCAGTTTATGGAGGATTTTAGC	SacI	pGL3 Basic
hXIST1.3 Forward	ACGAGATCTCTGCAGCAGCGAATTGCAG	BglII	pGL3 Basic
hXIST1.3 Reverse	GCTGAGCTCGGAATGGGAAGTCCCTTGAAG	SacI	pGL3 Basic
hXIST1.4 Forward	ACGAGATCTCTGCAGCAGCGAATTGCAG	BglII	pGL3 Basic
hXIST1.4 Reverse	GCTGAGCTCGCCATTCTATGAAATGTCTTTC	SacI	pGL3 Basic
hXIST1 Forward	AGCAAGCTTCTGCAGCAGCGAATTGCAG	HindIII	pGL3 Basic
hXIST1 Reverse	GCTGAGCTCGCCAGTGGGAGGGTAATGTA	SacI	pGL3 Basic
hXIST1 Antisense Forward	GCTGAGCTCCTGCAGCAGCGAATTGCAG	SacI	pGL3 Basic
hXIST1 Antisense Reverse	AGCAAGCTTGCCAGTGGGAGGGTAATGTA	HindIII	pGL3 Basic
hXIST1-2 Forward	ACGAGATCTCTGCAGCAGCGAATTGCAG	BglII	pGL3 Basic
hXIST1-2 Reverse	GCTGAGCTCTGGAGCCAAGCAGTAGTGAA	SacI	pGL3 Basic
hXIST1-3 Forward	ACGAGATCTCTGCAGCAGCGAATTGCAG	BglII	pGL3 Basic
hXIST1-3 Reverse	GCTGAGCTCTTCCTCCCTCTCCCTAGTGTTT	SacI	pGL3 Basic
pGL3-PFB site Sense Forward	GCAAGGAAGCGGGATTCTA	Blunt	pGL3 Basic
pGL3-PFB site Sense Reverse	CAAAGCAGCAGGAGTGCTAA	Blunt	pGL3 Basic
XIST1.1-PFB site Sense Forward	GCAAGGAAGCGGGATTCTA	Blunt	pGL3 Basic
XIST1.1-PFB site Sense Reverse	CAAAGCAGCAGGAGTGCTAA	Blunt	pGL3 Basic
XIST1-PFB site Sense Forward	GCAAGGAAGCGGGATTCTA	Blunt	pGL3 Basic
XIST1-PFB site Sense Reverse	CAAAGCAGCAGGAGTGCTAA	Blunt	pGL3 Basic
pGL3-CTCF site Sense Forward	CAGCATTAGCCAAGGGGTAG	Blunt	pGL3 Basic
pGL3-CTCF site Sense Reverse	GGACCAGAATGGATCACAGA	Blunt	pGL3 Basic
XIST1.1-CTCF site Sense Forward	CAGCATTAGCCAAGGGGTAG	Blunt	pGL3 Basic
XIST1.1-CTCF site Sense Reverse	GGACCAGAATGGATCACAGA	Blunt	pGL3 Basic
XIST1-CTCF site Sense Forward	CAGCATTAGCCAAGGGGTAG	Blunt	pGL3 Basic
XIST1-CTCF site Sense Reverse	GGACCAGAATGGATCACAGA	Blunt	pGL3 Basic
hSATB2 Forward	GCGAATTCGATGGAGCGGCGGAGCGAGAGCCCG	EcoRI	3X FLAG CMV10
hSATB2 Reverse	GCGTCTAGATTATCTCTGGTCAATTTCGGCAGG	XbaI	3X FLAG CMV10

Appendix 3

Details of primers used for ChIP – PCR

Description	Sequence (5' to 3')
Primer for OCT4 Peak on XIST Exon1 Forward	GCAAGGAAGCGGGATTCTA
Primer for OCT4 Peak on XIST Exon1 Reverse	CAAAGCAGCAGGAGTGCTAA
Primer for CTCF Peak on XIST Exon1 Forward	CAGCATTAGCCAAGGGGTAG
Primer for CTCF Peak on XIST Exon1 Reverse	GGACCAGAATGGATCACAGA

Publications

Manuscript under preparation

Shah, R., Kelkar, A. and Galande, S. Transcriptional regulation of human XIST lncRNA is governed by the dynamic interplay between pluripotency factors and chromatin organizer proteins – CTCF, SATB1 and SATB2.

Poster Presentation

Presented a poster on the **'Roles of SATB1 and SATB2 proteins in mammalian X-chromosome inactivation'**atthe18thTranscription Assembly Meeting 2015 held at IISER Pune. (12th March – 14th March, 2015).