# FUNCTIONALITY OF DOMAINS OF *XIST* AND THEIR ROLE IN ESTABLISHING CHROMOSOME INACTIVATION

by

Thomas Dixon-M<sup>c</sup>Dougall

H.B.S., The University of Toronto, 2013

## A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

## THE REQUIREMENTS FOR THE DEGREE OF

## DOCTOR OF PHILOSOPHY

in

## THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

## THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2020

© Thomas Dixon-M<sup>e</sup>Dougall, 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

| Functionality of domains | s of XIST and | d their role in | establishing | chromosome | inactivation |
|--------------------------|---------------|-----------------|--------------|------------|--------------|
| 2                        |               |                 | 0            |            |              |

| submitted by     | Thomas Dixon-McDougall             | in partial fulfillment of the requirements for |
|------------------|------------------------------------|--|
|                  |                                    |  |
| the degree of    | Doctor of Philosophy               |  |
| in               | Medical Genetics                   |  |
| Examining Co     | mmittee:                           |  |
| Dr. Carolyn B    | rown, Professor, Medical Genetics  | , UBC  |
| Supervisor       |                                    |  |
| Dr. Peter Stirli | ng, Associate Professor, Medical ( | Genetics, UBC                                  |
| Supervisory C    | ommittee Member                    |  |
| Dr. William T    | homas Gibson, Professor, Medical   | Genetics, UBC                                  |
| University Exa   | uminer                             |  |
| Dr. Thibault M   | layor, Professor, Biochemistry and | l Molecular Biology, UBC                       |
| University Exa   | uminer                             |  |
| Additional Sup   | pervisory Committee Members:       |  |
| Dr. Matthew L    | orincz, Professor, Medical Geneti  | cs, UBC  |
| Supervisory C    | ommittee Member                    |  |

Dr. Dixie Mager, Professor, Medical Genetics, UBC Supervisory Committee Member

### Abstract

X-chromosome inactivation establishes dosage compensation between the sexes of eutherian mammals through the long non-coding RNA gene, *XIST*. During development one of the female X chromosomes up-regulates *XIST*, which coats that chromosome and causes the large-scale silencing of genes. The X chromosome being inactivated by *XIST* is repositioned within the nucleus, condensed and enriched with heterochromatin associated factors. Understanding the mechanisms of how XIST functions has provided insights into how non-coding RNAs regulate cellular biology, the process of X-chromosome inactivation and novel tools for regulating the epigenome.

The function of *XIST* was investigated using an inducible *XIST* cDNA construct integrated into the autosome of a male fibrosarcoma cell line, HT1080. The chromatin domain surrounding the *XIST loci* had a significant effect on its activity, and of all the autosomal loci the 8p integration site functioned most effectively and was thus used for further studies. A series of isogenic inducible partial *XIST* constructs were created by modifying the Full length *XIST* construct in 8p to study the importance of individual regions of *XIST*. The functions dependent on each region of *XIST* were identified and the relationships between these identified processes were then examined through the use of chemical inhibitors.

XIST silencing of genes was demonstrated to depend upon two distinct regions at the extreme ends of the transcript, but the internal sequences spanning these regions were dispensable. Silencing occurred without obvious dependence on chromatin modifications such as those established by the two polycomb group complexes, that in turn relied on distinct regions of *XIST* suggesting entirely independent mechanisms. Both polycomb complexes were crucial, along with additional elements, for the recruitment of additional heterochromatin factors. This study in

iii

human differentiated cells yielded important insights beyond those seen in mouse differentiating cells. The results of this thesis revealed the regions of *XIST* that were both crucial and dispensable for its activity, and offer novel insights into the mechanisms that lead to chromosome inactivation.

## Lay Summary

In mammals the number of X chromosomes differs between males and females, so *XIST* inactivates one of the two X chromosomes in females so that all mammals only have a single functioning X chromosome. *XIST* produces a long RNA that coats its chromosome of origin and recruits factors that cause it to become inactivated. This thesis expanded the current knowledge of how *XIST* functions. I designed and characterized a series of partially deleted *XIST* genes that when compared with the full length *XIST* revealed the function of each region. Studying these partial *XIST* genes revealed that the entire length of *XIST* was crucial for the numerous pathways it initiates. I then determined how disrupting certain pathways affected how chromosome inactivation proceeded. This work provided new insights into how XIST RNA functions as well as how it might be best implemented as a therapeutic or tool for controlling chromosome activity.

## Preface

I performed all of the work presented in this thesis with a single exception. Bradley Balaton from the Brown Lab assisted in writing the R script and merging the .csv tables used when measuring the distribution of chromatin features at the *XIST* RNA cloud as described in the methods section 2.13.2.

### Part of chapter 3 of this thesis was previously published in:

Kelsey, Angela D., Christine Yang, Danny Leung, Jakub Minks, **Thomas Dixon-McDougall**, Sarah E.L. Baldry, Aaron B. Bogutz, Louis Lefebvre, and Carolyn J. Brown. (2015) "Impact of Flanking Chromosomal Sequences on Localization and Silencing by the Human Non-Coding RNA *XIST*." Genome Biology 16 (1). https://doi.org/10.1186/s13059-015-0774-2.

• Part of the data published within this paper is contained within chapter 3. I contributed to the analysis of heterochromatin mark recruitment and Cot-1 depletion by the *XIST* integration cell lines within this publication. Only those experiments performed by the candidate directly are included in this chapter of the thesis.

# **Table of Contents**

| Abstract     | iii   |
|--------------|---|
| Lay Summa    | aryv  |
| Preface      | vi  |
| Table of Co  | ontents vii   |
| List of Tabl | les xiv   |
| List of Figu | res xvi   |
| List of Abb  | reviations xix  |
| List of gene | e names and genetic factors xxi   |
| Acknowled    | gements xxiii   |
| Chapter 1:   | Introduction1   |
| 1.1 Tł       | nesis overview  |
| 1.2 Th       | ne mammalian sex chromosomes and dosage compensation                            |
| 1.3 Th       | ne initiation of XCI7   |
| 1.4 XI       | ST as the master regulator of eutherian XCI                                     |
| 1.4.1        | The localization of <i>XIST</i> to chromatin                                    |
| 1.4.2        | The pathways and factors linked to XIST-mediated transcriptional silencing      |
| 1.4.3        | Mechanisms governing histone deacetylation of the XIST RNA coated chromatin. 18 |
| 1.4.4        | Role of the polycomb group complexes in XIST-mediated chromatin remodeling 19   |
| 1.4.5        | Methylation of H4K20 and H3K9 during XCI  |
| 1.4.6        | The late chromatin modifications associated with the Xi                         |
| 1.4.7        | Nuclear positioning, structure and folding of the Xi                            |
| 1.5 Tł       | ne context-dependent factors contributing to XCI                                |

| 1.6 The      | model systems used to study XCI in mice and humans    | . 29 |
|--------------|---|------|
| 1.7 The      | HT1080 model system used to study XIST activity       | . 30 |
| 1.8 CRIS     | SPR, theory and application for modifying <i>XIST</i> | . 35 |
| 1.9 Thes     | is objective  | . 38 |
| Chapter 2: M | ethodologies  | 39   |
| 2.1 Cell     | culture and materials                                 | . 39 |
| 2.1.1 Tr     | ansferring adherent cells between containers          | . 40 |
| 2.2 Proc     | uring RNA for analysis                                | . 42 |
| 2.2.1 RI     | NA extraction and purification                        | . 42 |
| 2.2.2 D      | NAse treatment and reverse transcription of RNA       | . 43 |
| 2.3 Meas     | suring transcript levels by RT-qPCR                   | . 44 |
| 2.4 PCR      | to amplify DNA sequences                              | . 47 |
| 2.4.1 PC     | CR product purification                               | . 49 |
| 2.5 Pyro     | sequencing  | . 50 |
| 2.6 Prote    | ein Extraction and western Blotting                   | . 52 |
| 2.6.1 Pr     | otein Extraction from cells                           | . 52 |
| 2.6.2 SI     | DS-PAGE gel   | . 52 |
| 2.6.3 W      | estern blotting                                       | . 54 |
| 2.7 Gene     | erating gRNA plasmids                                 | . 58 |
| 2.7.1 Se     | electing gRNA targeting oligonucleotide sequences     | . 58 |
| 2.7.2 Cl     | loning targeting oligonucleotides into pSPgRNA vector | . 59 |
| 2.7.3 Tr     | ransforming gRNA plasmid into DH5α Competent Cells    | 61   |
| 2.8 Plasi    | mid purification                                      | . 64 |
|              |   | viii |

| 2.9     | DNA             | A electrophoresis through agarose gel  | 64  |
|---------|-----------------|--|-----|
| 2.10    | DNA             | A extraction from mammalian cells  | 66  |
| 2.1     | 0.1             | DNAzol DNA purification from Life Technologies method                            | 66  |
| 2.1     | 0.2             | Mouse Homogenizing Buffer (MHB) method of DNA purification                       | 66  |
| 2.11    | Trar            | nsfection and selection of deletion construct containing HT1080 cell lines       | 67  |
| 2.12    | Fluc            | prescently labelling cells (IF-FISH)   | 74  |
| 2.12    | 2.1             | Fixing cells on coverslips   | 74  |
| 2.12    | 2.2             | Creating fluorescent RNA probes through nick translation                         | 74  |
| 2.12    | 2.3             | Immunofluorescence and Fluorescent in situ Hybridization (IF-FISH)               | 75  |
| 2.13    | Ana             | lysis of IF-FISH images by qualitative and quantitative methods                  | 78  |
| 2.1     | 3.1             | Qualitative assessment of IF-FISH images   | 78  |
| 2.1     | 3.2             | Quantifying the distribution of Xi associated factors at the XIST RNA cloud      |     |
| rela    | tive to         | o the nuclear background   | 80  |
| Chapter | r <b>3: V</b> a | alidation of <i>XIST</i> Inducible Model   | .85 |
| 3.1     | Intro           | oduction   | 85  |
| 3.2     | Resi            | ılts   | 88  |
| 3.2     | .1 E            | nrichment of H4K20me1 and H3K27me3 at induced XIST RNA clouds                    | 88  |
| 3.2     | .2 C            | lonal variability of H3K27me3 and H4K20me1                                       | 93  |
| 3.2     | .3 T            | he extent of XIST functional capacity when expressed from 8p                     | 95  |
| 3.2     | .4 E            | stablishing the dynamics of XIST activity in the 8p XIST integration cell line 1 | 02  |
| 3       | 3.2.4.1         | Expression of XIST transgene RNA rapidly reached an upper plateau 1              | 02  |
| 3       | 3.2.4.2         | Dynamics of XIST-mediated distal gene silencing 1                                | 05  |
| 3       | 8.2.4.3         | Dynamics of XIST-mediated chromatin reorganization lag behind silencing 1        | 09  |
|         |                 |  | ix  |

| 3.3    | Discus    | sion  | . 115 |
|--------|-----------|---|-------|
| Chapte | r 4: Deve | elopment and validation of <i>XIST</i> deletion constructs      | 120   |
| 4.1    | Introdu   | iction  | . 120 |
| 4.2    | Results   | 5   | . 125 |
| 4.2    | .1 The    | generation of XIST deletion constructs using CRISPR Cas9        | . 125 |
| 2      | 4.2.1.1   | Δ Α   | . 131 |
| 2      | 4.2.1.2   | $\Delta$ FBh  | . 131 |
| 2      | 4.2.1.3   | $\Delta Bh$   | . 132 |
| 2      | 4.2.1.4   | Δ BC  | . 133 |
| 2      | 4.2.1.5   | Δ 3'PflMI   | . 133 |
| 2      | 4.2.1.6   | Δ D   | . 134 |
| 2      | 4.2.1.7   | Δ 3D5E  | . 134 |
| 2      | 4.2.1.8   | Δ Ε   | . 135 |
| 2      | 4.2.1.9   | Δ 3'  | . 136 |
| 4.2    | .2 Test   | ing the inducibility of the XIST Deletion constructs            | . 136 |
| 4.2    | .3 Visu   | alizing XIST RNA localization in the deletion cell lines        | . 146 |
| 4.3    | Discus    | sion  | . 149 |
| Chapte | r 5: Test | ing of Deletion Constructs                                      | 154   |
| 5.1    | Introdu   | iction  | . 154 |
| 5.2    | Results   | 5   | . 159 |
| 5.2    | .1 The    | regions of XIST responsible for the silencing of distal genes   | . 159 |
| 5.2    | .2 H3K    | 27me3 enrichment by XIST was facilitated by Repeats F, Bh and E | . 164 |

| 5.2.3   | UbH2A enrichment by XIST was facilitated by numerous distinct and essential       |
|---------|---|
| doma    | ains of <i>XIST</i>   |
| 5.2.4   | Two distinct regions of XIST were both essential for H4K20me1 enrichment 175      |
| 5.2.5   | MacroH2A enrichment requires large segments along the length of XIST 180          |
| 5.2.6   | SMCHD1 enrichment depends upon a unique series of segments of XIST 185            |
| 5.2.7   | CIZ1 enrichment depends upon Repeat E and a novel non-repeat region 191           |
| 5.2.8   | XIST-mediated histone deacetylation associated with regions critical for PRC2 but |
| not P   | PRC1 recruitment  |
| 5.2.9   | Perinucleolar localization of <i>XIST</i> is highly stable and resilient          |
| 5.3     | Discussion  |
| Chapter | 6: Chemical Inhibition of Key <i>XIST</i> Binding Factors210                      |
| 6.1     | Introduction  |
| 6.1.1   | General HDAC inhibition with Trichostatin A (TSA)                                 |
| 6.1.2   | HDAC3 specific inhibition with RGFP966  |
| 6.1.3   | PRC1 inhibition with PRT4165  |
| 6.1.4   | PRC2 inhibition with GSK343   |
| 6.2     | Results   |
| 6.2.1   | Validating Chemical Inhibitors in HT1080 inducible XIST model 217                 |
| 6.2     | 2.1.1 Survival and stress during chemical inhibition treatments                   |
| 6.2     | 2.1.2 Western blotting to determine effectiveness of chemical inhibitors          |
| 6.2.2   | Chemical inhibitor effects on <i>XIST</i> induced transcript levels               |
| 6.2.3   | Chemical inhibitors effects on XIST-mediated distal gene silencing                |
| 6.2.4   | HDAC3 inhibition specifically linked to XIST-mediated H3K27me3 enrichment 243     |
|         | xi  |

| 6.2.    | 5 PRC2 activity necessary for MacroH2A enrichment and histone hypoacetylation     | 247  |
|---------|---|------|
| 6.2.    | 6 PRC1 activity was necessary for XIST enrichment of SMCHD1 but not PRC2          | 252  |
| 6.3     | Discussion  | 256  |
| Chapter | · 7: Conclusion   | .261 |
| 7.1     | XIST homologues have retained some functional compatibility but were affected by  | 7    |
| their c | ontext  | 261  |
| 7.2     | XIST localization and regulation of chromatin architecture were highly stable and |      |
| likely  | facilitated by several partially redundant processes                              | 263  |
| 7.3     | Silencing mediated by XIST occurs through a unique pathway preceding chromatin    |      |
| remod   | leling  | 265  |
| 7.4     | Insights into chromatin remodeling by XIST  | 267  |
| 7.5     | Final conclusions about the mechanisms governing XIST activity                    | 270  |
| Referen | ces   | .275 |
| Appendi | ices  | .322 |
| Appen   | ndix A Tables of primers and antibodies used throughout this thesis               | 322  |
| A.1     | Table of primer combinations used to identify CRISPR deletions in XIST            | 322  |
| A.2     | Additional primers used throughout this thesis                                    | 323  |
| A.3     | Table of primary antibodies used for IF and western blotting                      | 325  |
| A.4     | Table of secondary antibodies used for IF and western blotting                    | 326  |
| Appen   | ndix B List of sequences across the cut sites of CRISPR deletion constructs       | 327  |
| B.1     | Δ A #3 Sequencing Results   | 327  |
| B.2     | Δ A #12 Sequencing Results  | 327  |
| B.3     | Δ FBh #21   | 328  |
|         |   | xii  |

|   | B.4     | Δ FBh #22   | 28 |
|---|---------|---|----|
|   | B.5     | Δ Bh #5   | 29 |
|   | B.6     | Δ Bh #7   | 29 |
|   | B.7     | Δ Bh #11  | 30 |
|   | B.8     | Δ BC #2   | 30 |
|   | B.9     | Δ BC #8   | 31 |
|   | B.10    | Δ BC #17  | 32 |
|   | B.11    | Δ 3'PflMI #3  | 32 |
|   | B.12    | Δ 3'PflMI #6  | 33 |
|   | B.13    | Δ D #3  | 33 |
|   | B.14    | Δ D #10   | 34 |
|   | B.15    | Δ 3D5E #13  | 34 |
|   | B.16    | Δ 3D5E #14  | 35 |
|   | B.17    | Δ 3D5E #15  | 35 |
|   | B.18    | Δ E #6  | 35 |
|   | B.19    | Δ E #10   | 36 |
|   | B.20    | Δ 3' #1   | 36 |
|   | B.21    | Δ 3' #7   | 37 |
| A | ppendix | x C Cross species compatibility of human XIST and mouse Xist                | 38 |
|   | C.1     | Cot-1 holes mediated by human and mouse XIST in their native and reciprocal |    |
|   | species | s   | 38 |

# List of Tables

| Table 1.1 Characteristic features of the inactive X chromosome    6                                      |
|--|
| Table 1.2 XIST Integration constructs in HT1080 cell line    33  |
| Table 1.3 Distal genes that undergo allele-specific silencing by <i>XIST</i> induction from 8p(23.1). 35 |
| Table 2.1 Surface area of cell growth containers along with the approximate volumes of growth            |
| media and Trypsin used for culturing cells   |
| Table 2.2 M-MLV reaction conditions for a single RNA reaction  |
| Table 2.3 Recipe for 20ul RT-qPCR using Fermentas products    45   |
| Table 2.4 Recipe for 20ul RT-qPCR using Promega GoTaq G2 products    45                                  |
| Table 2.5 Recipe for 25ul PCR    48  |
| Table 2.6 List of antibodies used for Western Blotting and their concentrations                          |
| Table 2.7 Phosphorylation of targeting gRNA oligonucleotides   |
| Table 2.8 Reaction conditions for ligating target sequence into digested oligonucleotides 61             |
| Table 3.1 Summary of XIST induced silencing dynamics    108  |
| Table 3.2 Dynamics of Xi associated chromatin featured recruited by 8p inducible XIST 114                |
| Table 4.1 List of gRNAs    128   |
| Table 4.2 List of $\Delta$ cell lines  |
| Table 4.3 Relative expression of XIST cell lines and constructs    142                                   |
| Table 5.1 Summary of strength of silencing of distal genes by induced <i>XIST</i> deletion constructs.   |
|  |
| Table 5.2 Average H3K27me3 enrichment and XIST RNA cloud    168  |
| Table 5.3 Average ubH2A enrichment and XIST RNA cloud    174   |
| Table 5.4 Average H4K20me1 enrichment and XIST RNA cloud    179  |
| xiv  |

| Table 5.5 Average MacroH2A enrichment and XIST RNA cloud                                     |
|--|
| Table 5.6 Average SMCHD1 enrichment and XIST RNA cloud                                       |
| Table 5.7 Average H3K27ac enrichment and XIST RNA cloud                                      |
| Table 6.1 Inhibitors and concentrations effective for HT1080s    224                         |
| Table 6.2 Expression of induced Full XIST is affected by GSK343                              |
| Table 6.3 Silencing strength of <i>XIST</i> in HT1080 cells treated with chemical inhibitors |
| Table 6.4 Comparison of how allelic contributions were affected by HDAC inhibition           |
| Table 6.5 Summary of effects of HDAC inhibition on chromatin remodeling by XIST 246          |
| Table 6.6 Summary of effects of PRC2 inhibition on chromatin remodeling by XIST 251          |
| Table 6.7 Summary of effects of PRC1 inhibition on chromatin remodeling by XIST 255          |
| Table A.1 Primary antibody list  |
| Table A.2 Secondary antibody list  |

# List of Figures

| Figure 1.1 Comparison of mouse and human XIST transcripts 12                                      |
|---|
| Figure 1.2 The human Xi is surrounded by XIST RNA, enriched for heterochromatin and               |
| compacted chromatin 15  |
| Figure 2.1 Summary of method used to generate targeted gRNA expressing plasmids                   |
| Figure 2.2 Transfection protocol for generating XIST deletion cell lines using targeted gRNA and  |
| Cas9 expressing plasmids  |
| Figure 2.3 Categorizing cells as either enriched or not enriched for factors at XIST RNA cloud 79 |
| Figure 2.4 Quantification of the relative distribution of factors at the XIST RNA cloud           |
| Figure 3.1 Enrichment of H3K27me3 and H4K20me1 vary depending on site of XIST                     |
| integration   |
| Figure 3.2 Clonal variability minimally affected XIST-mediated recruitment of H3K27me3 and        |
| H4K20me194  |
| Figure 3.3 Summary of published results of <i>XIST</i> integrations into HT1080 cells             |
| Figure 3.4 Qualitative assessment of 8p inducible <i>XIST</i> chromatin reorganization            |
| Figure 3.5 Induced XIST from 8p enriched for CIZ1 without increased chromatin density 100         |
| Figure 3.6 XIST transcript levels rapidly reach an upper level then plateau 104                   |
| Figure 3.7 XIST induced silencing of distal genes occurs primarily within 2 days 107              |
| Figure 3.8 XIST-mediated chromatin remodeling reaches its upper limit following 5 days of         |
| induction   |
| Figure 4.1 The structure of the inducible XIST constructs integrated into the 8p FRT site in the  |
| HT1080 cell line 2-3-0.5a   |
| Figure 4.2 Map of the CRISPR generated <i>XIST</i> deletion constructs                            |

| Figure 4.3 Induced XIST transcript levels in $\Delta$ cell lines relative to Full XIST                                       |
|--|
| Figure 4.4 Transcript levels of $\Delta$ constructs relative to Full <i>XIST</i>   |
| Figure 4.5 Transcript levels of $\Delta$ <i>XIST</i> constructs and cell lines of interest                                   |
| Figure 4.6 ΔA #3 transcript levels relative to Full <i>XIST</i>  |
| Figure 4.7 $\triangle$ 3D5E and $\triangle \triangle$ XIST forms a less unified RNA domain and $\triangle$ XB RNA completely |
| failed to localize   |
| Figure 5.1 XIST silencing of distal genes depends upon its 5' and 3' regions of XIST 161                                     |
| Figure 5.2 H3K27me3 enrichment by XIST was disrupted primarily by two regions of XIST  |
| centered around repeat F and Repeat E 167  |
| Figure 5.3 XIST enrichment of ubH2A was facilitated by four distinct interdependent domains of                               |
| <i>XIST</i>  |
| Figure 5.4 XIST-mediated enrichment of H4K20me1 depends upon the non-repeat 3'PflMI  |
| region as well as Repeat E   |
| Figure 5.5 MarcoH2A enrichment by XIST RNA depends on Repeat F, Repeat E and a broad   |
| internal region of XIST  |
| Figure 5.6 SMCHD1 enrichment by XIST depends upon three distinct domains of XIST   |
| Figure 5.7 CIZ1 enrichment by XIST RNA depended on both Repeat E and a small non-repeat                                      |
| region of exon 1 193   |
| Figure 5.8 XIST-mediated histone deacetylation weakly associated with PRC2 but not ubH2A                                     |
| essential domains  |
| Figure 5.9 XIST RNA localization to the perinucleolar compartment is highly stable, not                                      |
| depending upon any single region of <i>XIST</i>  |
| Figure 5.10 Summary of the functional importance of the regions of <i>XIST</i>   |
| xvii   |

| Figure 6.1 Survival of cells over the course of HDAC inhibition treatments                    |
|---|
| Figure 6.2 Treatment plan for HT1080 cells with chemical inhibitors                           |
| Figure 6.3 Chemical inhibitors produced expected effects on global histone mark levels in     |
| HT1080 cells  |
| Figure 6.4 Induced XIST transcript levels effected by PRC2 inhibition                         |
| Figure 6.5 PRC2 activity associated with XIST induced distal gene repression                  |
| Figure 6.6 Stronger XIST induced silencing observed following HDAC(3) inhibition treatment in |
| three of four genes   |
| Figure 6.7 HDAC3 inhibition affected H3K27me3 enrichment by XIST                              |
| Figure 6.8 PRC2 activity necessary for XIST regulation of H3K27me3, MacroH2A and H3K27ac      |
|   |
| Figure 6.9 PRC1 inhibition affects SMCHD1 and ubH2A enrichment by XIST RNA 254                |
| Figure 7.1 Summary of the identified functional domains pathways of <i>XIST</i>               |

## List of Abbreviations

| 5ddox         | Cells were treated with doxycycline at a concentration of $1\mu g/\mu l$ for five days |  |  |  |
|---------------|--|--|--|--|
| BLAST         | Basic local alignment search tool  |  |  |  |
| BSA           | Bovine Serum Albumin   |  |  |  |
| CRISPR        | Clustered regulatory interspaced short palindromic repeats                             |  |  |  |
| crRNA         | CRISPR RNA   |  |  |  |
| СТ            | Cycle threshold  |  |  |  |
| ddH20         | Double distilled H2O   |  |  |  |
| DEPC          | Diethyl Pyrocarbonate  |  |  |  |
| DMEM          | Dulbecco's modified Eagle medium   |  |  |  |
| DOX           | doxycycline  |  |  |  |
| EMEM          | Eagle's Minimum Essential Medium   |  |  |  |
| ESC           | Embryonic stem cell  |  |  |  |
| FCS           | fetal calf serum   |  |  |  |
| FISH          | fluorescent in situ hybridization  |  |  |  |
| FRET          | fluorescence recovery after photobleaching   |  |  |  |
| gRNA          | Guide RNA  |  |  |  |
| H3K27ac       | Acetylation of lysine 27 of histone H3   |  |  |  |
| H3K27me3      | Trimethylation of lysine 27 of histone H3  |  |  |  |
| H3K9me3       | Trimethylation of lysine 9 of histone H3   |  |  |  |
| H4K20me1      | Monomethylation of lysine 20 of histone H4   |  |  |  |
| H4K8ac        | Acetylation of lysine 8 of histone H4  |  |  |  |
| HDAC          | Histone deacetylase  |  |  |  |
| IF            | Immunofluorescence   |  |  |  |
| kb            | Kilobases  |  |  |  |
| KD            | Knock down (incomplete disruption of function)   |  |  |  |
| kDa           | Kilodalton (1.66x10-15 µg)   |  |  |  |
| KO            | Knock out (complete loss of function)  |  |  |  |
| lncRNA        | long non coding RNA  |  |  |  |
| MHB           | Mouse homogenizing buffer  |  |  |  |
| mRNA          | Messenger RNA  |  |  |  |
| NMR           | Nuclear magnetic resonance   |  |  |  |
| nt            | Nucleotide(s)  |  |  |  |
| PAM           | Proto-spacer adjacent motif  |  |  |  |
| PBS           | Phosphate-buffered saline  |  |  |  |
| PBST          | PBS supplemented with 0.1% Tween-20  |  |  |  |
| PBT           | PBS based blocking buffer supplemented with 1% v/v bovine serum albumin and            |  |  |  |
| 0.1% v/v Twee | en-20  |  |  |  |
| PCR           | Polymerase chain reaction  |  |  |  |
| PRC1          | Polycomb repressive complex 1  |  |  |  |
| PRC2          | Polycomb repressive complex 2  |  |  |  |
| qRT-PCR       | Real-Time Quantitative Reverse Transcription PCR                                       |  |  |  |
| rcf           | Relative Centrifugal Force (multiples of force of gravity)                             |  |  |  |
| RI            | RiboLock RNAse Inhibitor   |  |  |  |

| rpm      | Rotations per minute  |  |  |  |
|----------|---|--|--|--|
| RQ       | Relative Quantification   |  |  |  |
| RT-qPCR  | Reverse transcribed quantitative polymerase chain reaction            |  |  |  |
| s.d.     | Standard deviation  |  |  |  |
| (s)gRNA  | Single guide RNA  |  |  |  |
| SHAPE    | Selective 2' hydroxyl acylation analyzed by primer extension          |  |  |  |
| spCas9   | Streptococcus pyogenes derived Cas9 (double strand cleaving, NGG PAM) |  |  |  |
| SSC      | 0.15M sodium chloride and 0.15 sodium citrate solution                |  |  |  |
| TAD      | Topologically associated domain                                       |  |  |  |
| TAE      | Tris-acetate-EDTA   |  |  |  |
| TBS      | Tris buffered saline  |  |  |  |
| TBST     | Tris buffered saline with 0.1% v/v Tween-20                           |  |  |  |
| TetR     | Tet repressor   |  |  |  |
| tracrRNA | Trans-activating RNA  |  |  |  |
| Tris     | tris(hydroxymethyl)aminomethane                                       |  |  |  |
| TSA      | Trichostatin A  |  |  |  |
| UbH2A    | Ubiquitination of lysine 119 of histone H2A                           |  |  |  |
| Xa       | Active X chromosome   |  |  |  |
| XCI      | X-chromosome inactivation   |  |  |  |
| Xi       | Inactive X chromosome   |  |  |  |
| XIC/Xic  | Human / mouse X-inactivation centre                                   |  |  |  |

## List of gene names and genetic factors

The genes described throughout this thesis were listed below, along with their full names. In most instances the factors were described in the context of either the human specific or eutherian homologs of that gene. In the few instances where homologues between species were actively distinguished the species name was included in brackets.

| ACTB      | Actin Beta  |  |  |  |
|-----------|---|--|--|--|
| ASH2L     | ASH2 like histone lysine methyltransferase complex subunit    |  |  |  |
| ATRX      | Alpha thalassemia/mental retardation syndrome X-linked        |  |  |  |
| Cas9      | CRISPR-associated endonuclease Cas9/Csn1                      |  |  |  |
| CBX       | Chromobox   |  |  |  |
| CDX4      | Caudal type homeobox 4  |  |  |  |
| CIZ1      | Cip1-interacting zinc finger protein 1                        |  |  |  |
| CTCF      | CCCTC-binding factor  |  |  |  |
| DNMT3B    | DNA methyltransferase 3 beta                                  |  |  |  |
| EED       | embryonic ectoderm development                                |  |  |  |
| EZH2      | Enhancer of zeste 2 polycomb repressive complex 2 subunit     |  |  |  |
| FIRRE     | Functional intergenic repeating RNA element                   |  |  |  |
| G9a/EHMT2 | Euchromatic Histone Lysine Methyltransferase 2                |  |  |  |
| H2A       | Histone H2A   |  |  |  |
| H3        | Histone H3  |  |  |  |
| H4        | Histone H4  |  |  |  |
| HDAC3     | Histone deacetylase 3   |  |  |  |
| HnRNPK    | Heterogeneous nuclear ribonucleoprotein K                     |  |  |  |
| HNRNPU/SA | F-A Heterogeneous nuclear ribonucleoprotein U                 |  |  |  |
| HP1       | Heterochromatin protein 1                                     |  |  |  |
| HPH (hyg) | hygromycin B phosphotransferase (HPH)                         |  |  |  |
| JARID2    | Jumonji And AT-Rich Interaction Domain Containing 2           |  |  |  |
| KLF4      | Kruppel Like Factor 4   |  |  |  |
| KMT5A     | Lysine methyltransferase 5A (alternate name PR-SET7 or SETD8) |  |  |  |
| L3MBTL1   | Lethal(3)malignant brain tumor-like protein 1                 |  |  |  |
| LBR       | Lamin B Receptor  |  |  |  |
| LINE-1    | Long interspersed nuclear elements 1                          |  |  |  |
| MacroH2A  | Core histone macro-H2A.1                                      |  |  |  |
| NANOG     | Homeobox protein NANOG  |  |  |  |
| OCT4      | Octamer-binding transcription factor 4                        |  |  |  |
| PCGF(3/5) | Polycomb group factor homologs 3 and 5                        |  |  |  |
| PTBP1     | Polypyrimidine tract-binding protein 1                        |  |  |  |
| PuroR     | Puromycin resistance gene                                     |  |  |  |
| RBAP46/48 | Retinoblastoma-Binding Protein P46 and P48 subunits           |  |  |  |
| RBM15     | RNA Binding Motif Protein 15                                  |  |  |  |

RING1A/B Really Interesting New Gene 1 / Ring Finger Protein 1 paralogs A and B

RNF12 Ring Finger Protein 12

RSX (Metatherians) RNA on silent X

SHARP/SPENRNA-on-the-silent X

SMCHD1 Structural maintenance of chromosome flexible hinge domain-containing protein

1

SMRT Silencing Mediator For Retinoid And Thyroid Hormone Receptors

SUV4-20H1/2 Suppressor Of Variegation 4-20 Homolog 1/2

SUZ12 Suppressor Of Zeste 12

TetR Tet repressor

Tsix Mus musculus) antisense Xist (

XACT (Homo sapiens) X Active Specific Transcript

XIST (Homo sapiens/eutherian) X-Inactive Specific Transcript

Xist (Mus musculus) X-Inactive Specific Transcript

## Acknowledgements

This thesis could not have been completed without the help and support of many important people. Thank you so much to my supervisor Dr. Carolyn Brown for all the time, energy and support and guidance she provided though out this process. Thank you also to my committee members including Dr. Matthew Lorincz, Dr. Dixie Mager and Dr Peter Stirling who all provided incredible guidance and advice though out this process.

The advice and support received from members of the Brown Lab and LSI were instrumental to completing the work described in this thesis. Thank you to the members of the Brown lab, in particular Christine Yang, Dr. Samantha Peeters, Bradley Balaton, Angela Kelsey and Sarah Baldry. Thank you also to various members of the Molecular Epigenetics research group, in particular Riley Horvath and Aaron Bogutz.

Throughout this whole process my family Elizabeth Riddell-Dixon, Allan K. McDougall and Lisa Philips have been more supportive and wonderful than I could ever sufficiently express. Their constant support and encouragement were unbelievably important to this work. Finally, I would like to thank coffee, which has been a constant source of warmth, support and encouragement throughout every day of this long process.

### **Chapter 1: Introduction**

### 1.1 Thesis overview

X-chromosome inactivation (XCI) is the complex process that solves the otherwise lethal problem of the different copy number of over a thousand genes between the roughly two halves of the population in all species of eutherian mammals. This process is unusual as it is facilitated by a single gene, a long non-coding RNA (lncRNA) named *XIST* (X-Inactive Specific Transcript). Early in the development of all eutherian females, up-regulation of *XIST* leads to the inactivation of one of the X chromosomes, though the mechanisms governing this early stage of XCI vary between species [1]. *XIST* RNA spreads specifically across its X chromosome of origin and silences the majority of genes along its length while reorganizing the chromosome itself. The inactivated X chromosome (Xi) then remains stably inert in all the somatic cells of an individual for the remainder of their life.

Because *XIST* is such a potent epigenetic regulator, it has been and remains of great interest to researchers for the insights it provides into the role of RNA as functional elements, how dosage compensation is established and as a therapeutic tool. However, over the years despite incredible research being done, the mechanisms governing *XIST* activity were complicated by its large size and the complexity of the various pathways it initiates. The research performed thus far has relied almost exclusively on mouse models, and therefore there is a comparative paucity of knowledge about human *XIST* and XCI.

The focus of this research was to further develop a system that will allow for human XIST to be studied, to create an initial map of the functions of the various regions of human XIST and to provide novel insights into how human XIST functions. The model used throughout this thesis had an inducible XIST cDNA transgene integrated into chromosome 8 of a male cell line that allowed the activity of XIST to be characterized free of other XCI contributing factors. A series of isogenic cell lines with inducible XIST constructs containing serial deletions were generated using CRISPR to allow for each region of XIST to be examined individually. These partial XIST constructs were compared to the full-length progenitor to create a map showing the functions of each region of XIST. Finally, chemical inhibitors for key pathways downstream of XIST were used to create the first outline in a human model of the interdependence and independence of XCI processes. This work expands the current understanding of XCI as a whole as well as provides insights into the functional domains of XIST that contribute to silencing distal genes and modifying its surrounding chromatin along with insights into how these various pathways relate to each other. Silencing by the 5' and 3' most regions of XIST was observed to precede modifications to the surrounding chromatin. Several independent pathways facilitating the conversion of the XIST RNA bound chromatin to heterochromatin were identified, as well as novel insights into the hierarchy of how these heterochromatin marks were established. This work further examined how XIST activity has functionally diverged and adapted between mouse and humans as well as offering novel findings that may help reconcile current conflicting views within the field. Overall XIST was demonstrated to be extremely resilient as it retained most of its functional capacity even when large internal regions were excised. The results of this work will provide a launching point for further studies of the mechanisms of XCI and lncRNA activity.

### 1.2 The mammalian sex chromosomes and dosage compensation

In 1949 a dense region of DNA was observed in the nuclei of female cat neuronal cells, and this dense region of chromatin at the nucleolar periphery was originally labelled the nucleolar satellite but came to be referred to as the Barr body, after its discoverer [2]. In the following years it was demonstrated that this body was female specific and not unique to neuronal cells or cats, though it would ultimately be shown that the relative number of X chromosomes rather than sex was responsible for the observed Barr body [2]–[4]. In 1959 it was clearly demonstrated that the dense region of chromatin was a single X chromosome [5], [6]. Within a few years Dr Lyon proposed that one of the X chromosomes in mammalian females was being randomly inactivated early in an organism's development, and this inactivated X chromosome (Xi) then remained transcriptionally inert heterochromatin in somatic tissues throughout the individual's life [7], [8].

Across all mammalian species the mechanism used to compensate for the imbalance of gene products between the sexes involves inactivation of one allele for each of the X-linked genes. Both eutherian and metatherian (marsupial) mammals have a single pair of sex chromosomes, with males being heterogametic (X and Y chromosomes) while females being homogametic with two X chromosomes, one of which becomes inactivated through the process of XCI [9]. Prototherians developed their sex chromosomes independently of therian mammals, possessing 5 X chromosomes and 4-5 Y chromosomes, the difference resulting from the fusion of two Y chromosomes in the platypus, and were found to regulate each gene individually rather than on a chromosome-wide scale [9]–[13]. The random choice of which X chromosome was inactivated, proposed in the Lyon hypothesis, was unique to, and universal among, eutherian mammals. This process of random XCI was believed to have developed from imprinting based inactivation of the X chromosome which developed shortly after the therian mechanism of sex determination arose ~181 million years ago [11], [14], [15].

Over evolutionary time the therian sex-determining chromosome (Y) underwent significant degradation, with only 27 distinct proteins produced from its length, compared to the ~1000 produced from the X chromosome [16]. In marsupials, imprinted XCI has remained as the system used to establish dosage compensation between the sexes [17]. Analysis of multiple marsupial species indicated that the imprinted XCI leaves a greater proportion of X-linked genes transcriptionally active relative to eutherians [18]. Random XCI allowed for both X chromosomes to be transcriptionally active in half of cells, resulting in eutherian females having a mosaic expression profile [15]. Among eutherian mammals only rodents (mice and rats) still utilize systems for both imprinted and random XCI in embryonic lineages, with the former state being established prior to implantation then subsequently being replaced with random inactivation [19]. Imprinted XCI was proposed to occur in the extra-embryonic tissues of bovines, though subsequent work has opposed this [20], [21]. In mice random inactivation was demonstrated to be more stable and extensive than imprinted XCI, with more chromatin modifications and DNA methylation pathways unique to the process of random inactivation [22]–[24]. The inactivation of genes along the Xi during XCI is not universal, as ~15% of human and ~3% of mouse X-linked genes consistently avoid inactivation through unknown mechanisms, and remain active on the Xi across tissues [25]–[27].

The processes involved in random XCI were observed to vary across species of eutherian mammals, but gene silencing has generally been found to occur as the epiblast forms (E6.5 in

mice) followed by subsequent steps ensuring the silencing will be maintained [28]–[33]. Following XCI, the Xi is immediately identifiable within the nuclei of the cell by microscopy for its clear depletion of transcriptional activity as well as euchromatic features and its enrichment of heterochromatin-associated features and DNA methylation. The very earliest studies revealed that the compact chromatin of the Xi was typically associated with the nuclear and nucleolar peripheries. The characteristic features associated with the Xi that can be clearly visualized in the nuclei of female cells are listed in table 1.1.

| Xi feature                           | Factor responsible           | Reference(s)    |
|--------------------------------------|------------------------------|-----------------|
|                                      | (where known/relevant)       |                 |
| Histone hypoacetylation              | Histone deacetylases (HDACs) | [34], [35]      |
| (e.g. absence of H4K8ac and H3K27ac) |                              |                 |
| H3K27me3 enrichment                  | PRC2                         | [36], [37]      |
| H2A119Kub (ubH2A)<br>enrichment      | PRC1                         | [38]            |
| H4K20me1 enrichment                  | PRSET7/SETD8                 | [39]            |
| SMCHD1 enrichment                    |                              | [40]            |
| MacroH2A enrichment                  |                              | [41]            |
| H3K9me2/3 enrichment                 | SETDB1                       | [42]–[45]       |
| HP1 enrichment                       |                              | [46]            |
| Promoter DNA                         | DNMT3B                       | [47]–[49]       |
| hypermethylation                     |                              |                 |
| RNA polymerase II exclusion          |                              | [50], [51]      |
| Chromatin compaction                 |                              | [2], [52], [53] |
| Nucleolar association                | CTCF, FIRRE                  | [54], [55]      |
| Nuclear periphery association        | LBR                          | [56]            |
| Transcriptional inactivity           | SPEN                         | [57]–[59]       |

 Table 1.1 Characteristic features of the inactive X chromosome

### **1.3** The initiation of XCI

Some of the first mechanistic insights into XCI initiation came with the identification of an essential locus on the chromosome, located at Xq13.2 in humans, and named the X inactivation center (XIC) [60], [61]. The XIC is necessary for a cell to initially 'count' the number of X chromosomes and to then select one (in a diploid cell) of those chromosomes for subsequent inactivation though unknown mechanisms currently being investigated [62], [63]. The process of XCI itself is initiated by the expression of a long non-coding RNA (lncRNA) within the center of the XIC on the future Xi. This lncRNA was discovered in humans, and was termed the Xinactivation specific transcript (XIST) [64]. The XIST RNA specifically coats its X chromosome of origin and is ultimately responsible for establishing all of the characteristics of the Xi described in table 1.1 [65], [66]. XIST has been identified as the critical element for XCI in all eutherian mammals where it is expressed as a 15-24kb RNA transcript depending upon its species of origin [67], [68]. Marsupials independently evolved a completely distinct lncRNA, RSX, to regulate their imprinted XCI. No sequence homology exists between RSX and XIST, and it is generally accepted that they have entirely different evolutionary origins [69]. It was demonstrated, however, that RSX was capable of silencing genes on a transgenic chromosome when expressed in mouse cells, suggesting that the mechanisms employed by these lncRNAs were still functionally compatible across therians despite the mechanisms of XCI evolving independently [69], [70]. The convergent evolution of lncRNAs as master regulators of the X chromosome was part of the growing understanding that lncRNAs are prevalent and essential to many aspects of cell biology (reviewed in [71]). Therefore, an understanding of XIST function will prove illuminating to the broader field of lncRNA biology.

#### 1.4 XIST as the master regulator of eutherian XCI

The discovery of human *XIST* RNA coating the Xi in human females was quickly substantiated in mice, and rapidly followed up with functional studies establishing that *XIST/Xist* truly was a non-coding RNA and was essential for the silencing of X-linked genes [65], [66], [72], [73]. The human *XIST* RNA transcript was initially discovered as a 17kb isoform, though subsequent work identified a 19.3kb transcript, making it much larger than the vast majority of lncRNAs identified in mammals, which on average are less than 2kb [65], [74]–[76]. Multiple mouse *Xist* isoforms have also been discovered, with the first discovered transcript being 15kb and the longest version being 17kb [77], [78]. The spliced, polyadenylated and capped *XIST* RNA transcripts spread from the site of expression along its chromosome of origin to ultimately coat it in RNA [65], [78], [79]. As *XIST* spreads, it remodels the structural organization, histone code and the proteins/complexes bound to its chromosomes of origin. A representation of human and mouse transcripts is shown at the end of this section.

One of the most common challenges researchers have faced when studying *XIST* was predicting which regions of *XIST* were likely functionally significant, as sequence conservation across the locus is generally low across species. Chureau *et al* first calculated a 66% sequence conservation between human and mouse *XIST* transcripts, and a 62% sequence conservation between mouse and bovine *XIST* transcripts [80]. Using the Basic Local Alignment Search Tool (BLAST) to align the mouse *Xist* 17.9kb transcript (NCBI reference NR-001463) and human mRNA sequences *XIST* 19.3kb transcript (NCBI reference NR\_001564.2) revealed a 67% sequence conservation. A series of tandem repeat regions found in all *XIST* homologues were discovered

and their presence across species though a great deal of expansion and contraction of the repeats themselves has occurred between homologues [65], [68], [78], [81]. It was thought that the conservation of these repeats indicated that they had a significant role in the activity of XIST [82]. These tandem repeats were initially named A-E in order 5' to 3' along XIST, with the later discovered 6<sup>th</sup> repeat, Repeat F, discovered 3' of Repeat A [65], [78], [81]. Additional homolog specific tandem repeat sequences were discovered in bovine XIST (Repeat G) and dog XIST (Repeat H) which are of unknown function and as far as the author is aware have yet to be investigated [67], [68]. The 5' most repeat sequence, Repeat A, was also found to be the most conserved between humans and mice and is the most extensively studied region of XIST. Human Repeat A consists of a total of 429 nucleotides (nt) in length and is predicted to produce a series of stem loop structures [65], [67]. Numerous studies using SHAPE (Selective 2' Hydroxyl Acylation analyzed by Primer Extension), NMR (Nuclear Magnetic Resonance) fingerprinting and computational modeling of Repeat A in both mouse and humans have predicted a range of potential stem loop secondary structure conformations [82]-[87]. Later studies probing for direct RNA binding demonstrated a strong presence of inter-repeat binding between the individual repeat segments of Repeat A and one of the key conclusions was that secondary structure might be inherently dynamic [88]–[90]. This proposed dynamic range of conformations for Repeat A was subsequently supported by statistical modeling of Xist Repeat A structure that revealed a wide range of conformations likely to occur and additionally that there was no evidence of secondary structure conservation at Repeat A between 13 XIST homologues analyzed [91]. Numerous Repeat A binding factors have recently been identified in mice, some of which (SPEN and RBM15) were proposed to bind Xist semi-competitively at U-rich motifs between sections of double stranded RNA [87], [89]. The conflicting models predicting the secondary structure of

human *XIST* Repeat A by NMR may indicate that it too exists in a series of dynamic conformations, however it has yet to be directly determined which factors bind to human Repeat A [85], [86].

The remainder of the XIST repeats remain far less extensively studied than Repeat A, though in recent years an increasing number of studies have begun to be performed probing the structure and function of these other repeats. Repeat F was identified ~1.4kb into the sequence of human XIST, and consists of a more weakly conserved 101 nucleotides [81]. Nothing has been determined of the secondary structure of Repeat F or whether it interacts with any other regions of the XIST transcript, though one study proposed long range interactions with the 3' most region of exon 1 [87], [89]. In the primate lineages the Repeat B of XIST split into two distinct repeats, with the 5' most region called Repeat Bh (human B) while the 3' Repeat B was thought to have remained in the ancestral position [65], [67], [81]. Repeat Bh, identified as a 93nt long sequence and located ~1.9kb into the human XIST sequence has never been studied functionally [65], [67], [81]. Repeat B was identified ~800bp 3' of repeat Bh in human XIST and consists of 118nt, although using the tandem repeat finder indicated that Repeat B may only consist of ~43nt of true repeat sequence [65], [67]. Human and mouse Repeat C differed strikingly in the number of their repeats, as mouse Xist Repeat C consisted of 14 repeats sequences that were over 1.4kb, making it by far the largest Repeat C in any XIST homologue; while Human Repeat C does not even contain a complete repeat, being only 45nt in length [67]. The human XIST Repeat D contained both the largest monomer of all the repeats (289bp) and was by far the largest of the 7 human repeat regions, at 2.2kb [65], [67]. It has been debated whether the 1.5kb region upstream of this core Repeat D sequence was also part of the repeat given its much weaker conservation,

though recent evidence has suggested that long range interactions may exist between the core region of Repeat D and that upstream region [65], [67], [89]. Repeat D was observed to be typically more extensively degraded in rodent *Xist* homologues than other mammals with only 128nt of a single complete repeat monomer of mouse *Xist* still remaining entirely intact and 10 degenerate monomers discovered subsequently [67], [78]. The final repeat of *XIST*, Repeat E, was also the only one not located in exon 1 but instead located in exon 6, 12kb into the *XIST* RNA transcript. Whether Repeat E is actually a repeat sequence in human *XIST* or whether it simply reflects an AT rich and loosely structured region of *XIST* extensively bound by numerous factors is controversial [65], [67], [88]. The debate has led to inconsistency in where the edges of the Repeat are delineated and for the purposes of this work the encompassing 1.4kb is referred to as a repeat sequence [65], [67]. Evidence of protein binding within Repeat E as well as evidence that Repeat E participates in long range binding with non-repeat regions at the 3' most extreme end of *XIST* has been reported [89].

Since the discovery of the repeat sequences and early studies of Repeat A the prevailing view for nearly 20 years of research into the function of *XIST* has been that the conserved Repeats A-F are the essential functional elements of *XIST* in mammals [82], [92]. Conservation has not always been a clear indication of function when studying *Xist*, however, as exon 4 was shown to be a short but highly conserved sequence between *XIST* homologues and yet its removal from *Xist* revealed it to be dispensable for XCI [93], [94]. The research examining the regions of the *XIST* homologues critical for various aspects of XCI have therefore been almost exclusively focused upon the repeats.



### Figure 1.1 Comparison of mouse and human XIST transcripts

Summary of the human *XIST* and mouse *Xist* transcripts structure, with the repeat regions highlighted by colour to illustrate the expansion and contractions that have occurred over time. The gaps between exons are not to scale and are merely shown stylistically as gaps between the sections of the transcript to demonstrate their position. The regions of human *XIST* predicted to bind to specific factors or be crucial for certain processes are listed above the construct, while a sample of the relevant binding factors associated with mouse *Xist* are shown below. Solid black lines indicate the regions critical for a process, while dotted lines indicate regions of questionable significance.

#### **1.4.1** The localization of *XIST* to chromatin

Once an X chromosome was selected to be inactivated in a female cell, mammalian XIST RNA coated the chromosome rapidly. A model of spreading was proposed where Xist RNA first associates with sites in close 3D spatial proximity to the *Xist* locus, and then spreads outwards along the DNA itself [95], [96]. This model was supported when Xist RNA induced outside of its natural locus on the X chromosome bound to chromatin in close spatial proximity to its transgenic locus [95]. It has not yet been determined in any species what mediates the spread of XIST RNA across the X chromosome, though it was determined that the interaction of Xist with chromatin was likely mediated by protein intermediates rather than through direct RNA-DNA hybridization [79]. The localization of the XIST RNA in both mice and humans results in the associated X chromosome forming large chromatin domains, with the majority of the topologically associated domains (TADs) of the active X chromosome reformed into two large megadomains [55], [97]. When visualized using microscopy the distribution of XIST RNA appeared as a localized mass surrounding the inactivated chromatin, which has become referred to conventionally as the XIST RNA cloud (figure 1.2) [98]. XIST RNA in both mice and humans remains associated with the X chromosome during interphase, but were shown to be lost during mitosis and reaccumulated after replication, though it has been reported that fluorescently labelled foci of mouse *Xist* were still visible during mitosis [79], [98], [99].

Two proteins strongly associated with localizing *Xist* RNA to chromatin are HNRNPU and CIZ1. HNRNPU has been observed to be enriched on the Xi in mouse cells [100], [101], and it was subsequently demonstrated that HNRNPU knockdown critically disrupted the ability of mouse *Xist* to localize into an RNA cloud [102]. The model of HNRNPU directly binding to
mouse *Xist* RNA was subsequently validated in the various 2015 publications screening for potential binding factors [57], [59], [103]. In humans hnRNPU also binds *XIST* RNA, however unlike mice it is not essential for *XIST* RNA to effectively localize to the chromatin fraction in most somatic cell lines, suggesting additional factors also contribute to the process such as the closely related hnRNPUL1 [104], [105]. HNRNPU was observed to be crucial for mouse *Xist* localization and spreading across the X chromosome during XCI, though additional factors were proposed to also be involved [103], [106]. HNRNPU was observed binding broadly across both mouse *Xist* and human *XIST* at both the region surrounding Repeat D and the 3' most non-repeat region of the transcript of both homologues [105].

Another localization factor, CIZ1, was first observed to be associated with *Xist* in 2015 in one of the large screens for *Xist* interacting factors and its association was subsequently validated in 2017 [59], [107], [108]. Knockdown experiments have demonstrated that CIZ1 is necessary for maintaining proper *Xist* localization and silencing of genes along the mouse Xi, although it is important to note that significant XCI was still able to occur, given that female CIZ1 KO mice are able to survive to term [107]. There was evidence that CIZ1 acts as one of many *Xist* localization factors during the establishment of XCI, but in later time points takes on a more essential role in ensuring that *Xist* remained bound to the Xi [107], [108]. CIZ1 binding to mouse *Xist* RNA was demonstrated to depend specifically on Repeat E, but was clearly found not to require any region of exon 1 (Repeats A-D and F) [107]. Human *XIST* RNA becomes enriched for CIZ1 at Repeat E, but no research was done to determine the mechanisms responsible for that enrichment [109].



# Figure 1.2 The human Xi is surrounded by *XIST* RNA, enriched for heterochromatin and compacted chromatin

A female (IMR90) cell was fluorescently labelled by IF-FISH to illustrate the relative enrichment of the heterochromatin mark H3K27me3 (red) and greater density of DAPI labelled DNA (blue) at the XIST RNA cloud (green) on the Barr body (Xi). The first panel shows the merged imaged of these marks within the nuclei and the scale bar indicates 10µm. Arrows are used to indicate the position of the XIST RNA cloud across panels.

#### 1.4.2 The pathways and factors linked to *XIST*-mediated transcriptional silencing

Most of the  $\sim 1000$  X-linked genes on the Xi were found to be consistently and stably silenced across individuals and cell lines as a result of XIST spreading across the X chromosome. The loss of transcriptional activity was observed using fluorescent microscopy using fluorescent in situ hybridization (FISH) and immunofluorescence (IF) as the depletion of Cot-1 RNA and transcriptional machinery from the Barr Body [50], [110]. Repeat A was first demonstrated to be essential for silencing using transgenic inducible Xist constructs in mouse embryonic stem cells (ESC) [82]. Since then the Repeat A has become widely accepted as essential for silencing in multiple studies of mouse Xist and has even been demonstrated to be critical for human XISTmediated silencing of a proximal reporter gene [103], [111], [112]. A huge breakthrough in understanding how Repeat A mediated silencing came in 2015 when independent studies identified SPEN as a mouse Xist interacting factor. SPEN (also referred to as SHARP) contains multiple RNA binding motifs and was subsequently shown to interact with several other factors that included histone deacetylases (HDACs) and repressors of gene activity [57]–[59], [103]. SPEN association with mouse Xist Repeat A was essential for the initial silencing of genes during XCI in mouse ESCs [113]. It was proposed that SPEN recruited by Xist may activate prebound HDAC3 via the SMRT intermediate as a potential mechanism for establishing gene silencing, though HDAC3-independent mechanisms of silencing at a minority of genes are also believed to exist [103], [114]. In the SPEN-SMRT-HDAC3 model system, it was proposed that Xist-mediated activation of HDAC3 along the X chromosome led to the deacetylation of histones along the Xi. Additional factors and pathways contribute to silencing a further subset of genes through their interaction with Xist, including the polycomb group complexes (PRC) [59], [115].

Conflicting reports have emerged about the role of RBM15 methylation of N6-adenosine (m6A) of the *XIST* transcripts in mice and humans. Current evidence seems to suggest that both mouse and human *XIST* RNA can undergo methylation in specific regions, including the region adjacent to Repeat A and at the 3' most region of *XIST* [58], [116]. Subsequent studies of mouse *Xist* have found that disruption of methylation along the transcripts did not disrupt *Xist*-mediated silencing [115], [117]. Critically, from the perspective of understanding the activity of human *XIST*, none of the mechanisms proposed to be crucial for *XIST*-mediated gene silencing have been tested outside of mouse cells.

An important aspect of the ability of *XIST* to silence genes is its ability to spread to silence large swaths of a chromosome. Work by Dr Minks of the Brown lab demonstrated that just expressing Repeat A of human *XIST* was sufficient to repress an immediately proximal reporter gene [111]. However, it was not known what was necessary for *XIST* to be able to silence distal genes and it seemed probable that more extensive pathways were necessary to silence those genes. The evidence of this increased complexity was supported indirectly by the observation that even in its natural context, a subset of the ~1000 protein coding genes identified on the Xi avoid silencing, remaining stably expressed from both X chromosomes. The proportion of genes silenced by XIST differs noticeably between mice and humans, as ~15% or more of human X-linked genes escape inactivation consistently, while only ~3% of mouse X-linked genes escape inactivation [118], [119]. Additional X-linked genes escape silencing in a tissue-specific or individual-specific manner [120], [121]. The factors that mediate *XIST* spreading across the X to silence genes were conventionally referred to as either waystations or boundary elements depending on whether they promote or prevent *XIST*-mediated silencing of a locus respectively [122]. Recent

evidence from mouse studies suggested that gene repression was mediated by *Xist* recruiting SPEN specifically to the sites of enhancers and promoters [113]. It was therefore clear that to understand how *XIST* established gene silencing required identifying both the mechanisms underlying targeting of transcriptionally active genes as well as how it silences them.

#### 1.4.3 Mechanisms governing histone deacetylation of the XIST RNA coated chromatin

Deacetylation of histones on the Xi has received increased attention recently in mouse XCI as it was proposed to be a causative factor in the earliest stages of establishing gene silencing [114]. Hypoacetylation of H2A, H3 and H4 is a common feature of the Xi in mice and humans [34], [35], [123], [124]. In both mice and human cells, deacetylation of the future Xi occurs very early in the process of XCI and is established by *XIST/Xist* expression and spreading [110], [125]. Numerous HDAC proteins exist, but recent work identified HDAC3 specifically to be the sole essential HDAC necessary for histone deacetylation mediated indirectly by mouse *Xist* Repeat A binding to SPEN [103], [114]. Once *Xist* has established hypoacetylation of the Xi that state of deacetylation was found to be stably maintained in differentiated cells even without the continued presence of *Xist* RNA [126]. The current model of histone deacetylation during XCI suggests that *Xist* activates HDAC3 already bound to the X chromosome via the SPEN-SMRT intermediates and that it does not directly recruit HDAC3 [114]. An additional role for histone deacetylation contributing to the establishment of heterochromatin territories along the Xi was proposed, though the strength of the effect on PRC associated marks was relatively weak [114].

#### 1.4.4 Role of the polycomb group complexes in *XIST*-mediated chromatin remodeling

The polycomb group complexes PRC1 and PRC2 were the most studied and divisive XIST associated factors, and their interaction and interdependence with each other was extensively studied [127]–[133]. PRC2 is responsible for trimethylating lysine 27 on the histone tale of H3 (H3K27me3) which is commonly associated with transcriptional inactivity [134]. PRC2 consists of four core component proteins that include catalytically active EZH2 as well as SUZ12, EED and RbAp46/48 and cofactors such as JARID2 that are believed to be involved in regulating the activity of EZH2 as well as regulating where it binds in the genome [135]–[137]. PRC1 consists of a variable array of core factors that ubiquitinate lysine 119 of histone H2A (H2AK119ub or ubH2A) via the catalytically active component RING1B [138]. The subtypes of PRC1 are generally subdivided into the canonical groups that contained the CBX core component or the non-canonical RYBP subunit. Both versions of PRC1 contain one paralog of PCGF, labelled 1 through 6 [138]. The non-canonical PRC1 containing specifically the PCGF3 or PCGF5 core components are believed to bind indirectly to *Xist* and be the earliest PRC1 types recruited by Xist during XCI, though both non-canonical and canonical PRC1 have been shown to be ultimately recruited to the Xi [139], [140].

H3K27me3 and ubH2A were some of the earliest heterochromatin associated marks observed to accumulate on the Xi during XCI, and their establishment being directly dependent upon *Xist* RNA was well established [36]–[38]. How PRC1 and PRC2 are bound and/or activated by *Xist* has been extensively studied, with seemingly conflicting results over the years, thus what follows is a brief overview of the key relevant findings.

The importance of PRC2 for maintaining silencing of genes on the imprinted Xi was first demonstrated in mouse Eed knockout trophectoderm cells, though it was less crucial in maintaining the randomly inactivated Xi in the embryo proper [141]–[143]. Some studies suggested that PRC2 bound directly to mouse Repeat A, though even at the time those studies were contentious, as previous studies had found enrichment of H3K27me3 at the Xi in cells with Repeat A absent from the Xist transcript [37], [86]. Later work suggested that PRC2 recruitment to the Xi by mouse Xist was mediated by the cofactor Jarid2 binding to a region extending from Repeat F to B, with every region 3' dispensable for PRC2 recruitment and enrichment of the chromatin with H3K27me3 [131]. Subsequent studies of Xist deletion constructs led the same research group to suggest that PRC2 was not interacting with Xist at all, but was being recruited by PRC1 [129]. PRC1 was shown to begin to enrich the mouse Xi with ubH2A slightly before PRC2 began to enrich the chromatin with H3K27me3 [114]. The first proposed mechanism of how PRC1 was recruited by mouse Xist came with the observation that hnRNPK acted as an intermediate between Repeats B and C of Xist RNA [59] and the PCGF3 and PCGFP5 containing types of PRC1 [127]. Disruption of the PCGF3/5 containing PRC1, knockdown of hnRNPK and removal of Repeat B & C containing regions of Xist all resulted in loss of both ubH2A and H3K27me3 enrichment during mouse XCI [59], [129], [140]. As a result, one popular current hypothesis is that hnRNPK binding to Xist Repeats B and C directly recruits PRC1 to ubiquitinate H2A on the future Xi and that these marks then recruit PRC2 to trimethylated H3 [127]. There was some evidence that PRC2 binding to ubH2A was facilitated by the JARID2 cofactor, though this remains uncertain [133]. Some reports have suggested that PRC1 and PRC2 create feedforward mechanisms to spread their marks on the Xi, with H3K27me3 recruiting additional (non PCGF3/5 containing) PRC1 to further enrich and maintain

ubH2A enrichment [139], [144]. Currently it remains unclear how this model can be reconciled with the observation that *Xist* preferentially binds to regions enriched specifically with H3K27me3/PRC2 at the very onset of XCI [95], [114]. The pathway of *Xist* recruiting PRC1 which then recruits PRC2 has thus far only been observed in mouse models, so an outstanding question remains, namely whether the same pathway is also applicable to XCI in humans.

#### 1.4.5 Methylation of H4K20 and H3K9 during XCI

While the PRC associated marks were the most studied Xi associated heterochromatin marks, numerous other heterochromatin modifications have been identified as enriched along the Xi, the functional significance of which remains mostly unknown. The mono-methylation of lysine 20 of histone H4 (H4K20me1) becomes enriched on both the human and mouse Xi, yet little is known about how it is established or what role it plays in XCI [39], [145]. H4K20me1 is established by KMT5A (also referred to as PR-SET7 or SETD8), a histone methyltransferase that is presumed to be recruited to the future Xi but has yet to be investigated directly [146]. Two forms of H4K20 methylation, H4K20me1 and H4K20me3 (established by SUV4-20H1/2), were found to be maintained on the Xi in visibly distinct domains [145]. H4K20me3 is generally associated with gene repression as well as chromatin compaction and it must be determined why a significant amount of H4K20me1 remains enriched on the Xi without undergoing subsequent dimethylation [147], [148]. One of the few studies that examined H4K20me1 enrichment on the Xi observed that disruption of PRC2 (EED KO) in mouse cell lines attenuated the proportion of Xi with clear enrichment of H4K20me1( from 83% to 50%) [128]. H4K20me1 enrichment may contribute to chromatin compaction through indirect chromatin binding factors (e.g. L3MBTL1),

but these observations have yet to be confirmed in the context of XCI [149]. No analysis in any species has been performed to determine whether KMT5A binds *XIST* via intermediates or is recruited through downstream processes (like those described for PRC2 recruitment by PRC1 in mice). Outside of the field of XCI, H4K20me1 has been associated with both gene activation and gene repression in a context-dependent manner and was regularly associated with DNA repair indicating that its role on the Xi may similarly be context dependent [146].

One of the other heterochromatin marks enriched early during XCI of both mice and humans was H3K9me3 [45], [150], [151]. The histone methyltransferase SETDB1 was proposed to be responsible for establishing H3K9me3 enrichment during XCI but loss of SETDB1 in humans somatic cells did not result in an accompanying loss of H3K9me3 enrichment, suggesting that other factors are capable of maintaining that mark [152], [153]. H3K9me3 enrichment during mouse XCI was found to be essential for the initial silencing of ~150 genes along the Xi, but its role maintaining silencing was minimal [44]. It was shown that the enrichment of H3K9me3 was associated with the DNA hypermethylation of the Xi, believed to be one of the key mechanisms responsible for maintaining the inactivated status of the Xi [44], [154]. Loss of H3K9me3 has also been associated with proper chromatin compaction and folding of the Xi potentially through the recruitment of factors such as HP1a [46], [153]. H3K9me3 seemed to be enriched independently of the Polycomb established marks, as these two groups of marks were enriched in distinct regions of the Xi and disruption of either has not been associated with major effects on the other mark [42], [44]. At time of writing it is not known how SETDB1 becomes recruited/activated or what processes control the extent of H3K9me3 enrichment during XCI. Enrichment of H3K9me2 along the Xi has also been suggested, though studies have been

hampered due to issues of antibody cross-reactivity, and is thought to recruit CYDL to the Xi promoting the further spread of heterochromatin marks [44], [155].

### 1.4.6 The late chromatin modifications associated with the Xi

After the establishment of the early chromatin modifications described so far during XCI, the histone variant MacroH2A and the SMC family protein SMCHD1 are recruited to the Xi [40], [156]. It is believed that their role is to maintain the inactivation status given their late recruitment, and both were recently proposed to interact directly with Xist [157]. MacroH2A was discovered first of the two marks, when it was observed to be enriched on the mouse Xi after the initial propagation of XCI, first leading to its association with maintenance of XCI in the late 1990s [41], [156]. MacroH2A referred to the variant histories of H2A containing a 30kDa globular domain at the C-terminal proposed to interfere with transcription and recruit HDACs [158], [159]. The distribution of MacroH2A in humans was observed to colocalize with heterochromatin including H3K27me3 both on autosomes and on the Xi, while negatively correlating with active histone marks across the genome [145], [160]. It was discovered that the continued enrichment of MacroH2A on the Xi depended upon the presence of Xist through as yet unknown mechanisms [126]. MacroH2A was demonstrated to be dispensable for mouse viability during development and loss of both variants of MacroH2A had no observable effect on XCI [161], [162]. The proposed mechanisms governing how *Xist* facilitates the recruitment of MacroH2A have become increasingly unclear with recent evidence that ubiquitination of MacroH2A is also essential for its recruitment to the Xi [163], [164]. It was proposed that PRC1 activity represented an important first step in MacroH2A becoming incorporated into the Xi

though a recent paper under review suggested that this ubiquitination may instead be mediated by BRCA1 [165]. The way that MacroH2A becomes recruited to the Xi either directly or indirectly by *XIST* has not been determined in any mammalian species nor is it known what purpose MacroH2A plays in maintaining XCI.

The second late Xi associated heterochromatin mark that was of interest to this thesis was SMCHD1. Like MacroH2A, SMCHD1 becomes enriched on the Xi late in the process of XCI though knockout of SMCHD1 in mice produced clear defects in the maintenance of XCI at ~10% of inactivated genes and was associated with the chromatin compaction characteristic of the Xi [40], [46]. SMCHD1 contributed to the accumulation of DNMT3B-mediated DNA methylation of the mouse Xi, though SMCHD1 independent DNA methylation pathways also exist, raising the question of whether its role in maintaining silencing was a result of establishing DNA methylation [48]. SMCHD1 may contribute to the compaction of the Xi, allowing for Xist to spread into a small subset of euchromatin TADs during XCI and ultimately incorporate them into the large megadomains associated with the Xi (described in greater detail in section 1.5) [166]. Contrary to the reports of direct interaction between Xist and SMCHD1 [157], a separate study has proposed that SMCHD1 enrichment to the mouse Xi depended entirely upon the enrichment of ubH2A by PRC1 [167]. This latter model concluded that SMCHD1 enrichment being dependent on Repeat B and C of mouse Xist through the mediation of HNRNPK and PRC1 [167]. At time of writing the conflicting views of whether SMCHD1 binds mouse Xist directly or due to ubH2A enrichment remain unreconciled, with little known about how SMCHD1 might be recruited in humans.

#### 1.4.7 Nuclear positioning, structure and folding of the Xi

The unique position of the inactive X chromosome within the nucleus was an early characteristic ascribed as a feature of the Barr body [2]. The Xi associates preferentially with both the nuclear and nucleolar periphery, both of which are regions typically associated with heterochromatin [168]–[170]. It is still being investigated what role the position of the Xi plays in establishing/maintaining XCI, though disruption of perinucleolar localization and Xist expression from the Xi was shown to affect maintenance of the Xi [171]. The primary factor believed to be involved in the localization of the Xi to the nuclear periphery is LBR, a transmembrane protein of the nuclear membrane that was found to associate with mouse *Xist* potentially by binding both Repeat F and E [56], [103], [172]. Initially the association with LBR was proposed to be important for proper silencing of a subset of genes on the Xi and for *Xist* to properly coat the chromosome [56]. However the importance of LBR was questioned subsequently as it was observed to have little effect on maintaining silencing in ESCs or the viability of female mice [115], [173].

A multifactorial model was proposed for how the Xi becomes associated with the nucleolar periphery or perinucleolar compartment. The perinucleolar localization of the Xi was suggested to be dependent upon the continued expression of *Xist* from the Xi [171]. Recent work has revealed that an additional lncRNA, *Firre*, is critical for the localization of the Xi to the perinucleolar compartment as well as for the continued enrichment of H3K27me3 in mouse differentiated cells [174], [175]. *Firre* was reported to be transcribed from the active X chromosome and act *in trans* on the Xi to ensure its association with nucleolar membrane [174]. *Firre* transcripts were found to interact with the *Dxz4* locus of the X chromosome which was

found to be critical for perinucleolar association [174]. The *Dxz4* locus has previously been identified as the boundary sequence between the two megadomains of the Xi, established during XCI in both humans and mice [55], [97], [176]. CTCF was proposed to act as a tether between the nucleolar membrane and the Xi, principally by binding the *Dxz4* locus in an orientation dependent manner [177], [178]. *Firre* RNA was shown to be critical for CTCF accumulation at the edge of the Xi adjacent to the nucleolar membrane but the mechanisms governing this interaction are not yet known [174], [175]. *Firre* knockouts were found to affect heterochromatin mark retention on the Xi but did not show signs of gene reactivation on the Xi, however *Firre* has also been found to affect autosomal gene expression indicating that it has multiple roles within a cell [179], [180]. *Xist* expressed from mouse autosomes can frequently result in perinucleolar association, suggesting that *Xist* itself is ultimately responsible for nucleolar association with *Firre* RNA potentially recruited *in trans* through as yet unknown mechanisms [171]. How *XIST* RNA itself directed the reorganization of its chromosome to the nucleolar periphery has yet to be identified in any species.

#### **1.5** The context-dependent factors contributing to XCI

Extensive research over the last several decades has identified numerous factors and loci along the X chromosome that drive the expression of *XIST* and contribute to its function. The concept of a critical region of the X chromosome being essential for the chromosome to undergo XCI was proposed shortly after the Lyon hypothesis, and this region was termed the X-inactivation center (*XIC*) [181]. The *XIC* was discovered across eutherian species (located at human Xq13 and mouse X D) and was defined to include a region that consisted of numerous genes crucial for regulating the onset and initiation of XCI, and generally defined as just under 1 Mb in size in humans and mice [182]–[184]. The *XIC* comprises the *XIST* gene itself as well as a range of regulatory factors generally thought to act *in cis*. Regulatory lncRNAs were identified within the *XIC* which ranged in function from promoting *XIST* up-regulation, such as *JPX* and *FTX* [185], [186], to preventing *Xist* expression, best exemplified by the mouse-specific *Xist* antisense/repressor lncRNA, *Tsix* [80]. The *XIC* also contained genes for regulatory proteins critical for XCI and conserved between humans and mice, including *RNF12* and *CDX4* [80], [187]. One recent study using mouse ESCs suggested that the establishment of TADs within the *XIC* itself was a key step in the up-regulation of *Xist* and chromosome inactivation [188]. Additional studies have supported this proposed model, that specialized sequences and the 3D organization of the *XIC* actively regulated *Xist* up-regulation and repression by determining which factors are able to interact with its promoter [185]. Understanding of the *XIC* has continued to develop over the years, with a novel conserved lncRNA gene *LINX* recently identified as a context dependent *cis* regulator of genes within its TAD on *XIC* [189].

Outside of the *XIC* itself, several additional X-linked factors were found to affect the expression or activity of *XIST* including the previously discussed lncRNA *FIRRE* and the human specific lncRNA *XACT*, proposed to act *in trans* as a *XIST* antagonist specifically in humans pluripotent cells [175], [190], [191]. It was proposed that *XACT* expression in human ESCs from both X chromosomes during pluripotency might prevent *XIST*-mediated inactivation in humans during the early stages of development [190]. In addition to these lncRNAs, loci along the X chromosome itself were proposed to contribute with *XIST* during inactivation. The importance of the *Dxz4* loci as a hinge mediating the formation of the Xi megadomains, associating with the

perinucleolar compartment and potentially contributing to heterochromatin enrichment along the Xi across mammals has been previously described [97], [177].

In addition to specific X-linked factors or loci discussed, the broader sequence of the X chromosome was proposed to have adapted to contribute to ensure efficient and rapid XCI, with long interspersed nuclear element-1 (LINE-1) being the primary examples of this chromosomewide adaptation [192], [193]. The X chromosome was shown to generally be enriched with twofold as many LINE-1 elements relative to the average autosome and these LINEs were proposed to act as waystations for XIST RNA to spread across the X chromosome in multiple species [122], [194]–[197]. Correlation between LINE density and efficiency of Xist/XIST-mediated spreading as well as inactivation, have been demonstrated through studies of instances of X: autosome translocations [121], [192], [198]. Conflicting reports have emerged about the role of LINE-1 elements during XCI. Xist RNA has proportionally lower occupancy at LINE-1 elements on the Xi which led to the hypothesis that LINE-1 elements promoted silencing at regions that Xist does not coat [199]. The most recent study used computational predictive modeling to conclude that both mouse Xist and human XIST RNA might bind to LINE-1 elements through the formation of DNA:RNA triplexes, though no functional testing was performed to support this yet [200]. LINE-1 elements may have acted in a context dependent manner during XCI, with Xist recruited to active LINEs and silent lines promoting the spread of heterochromatin [197]. CTCF binding along roughly 2000 sites on the X chromosome was proposed to be critical element separating the remaining euchromatin and heterochromatin regions of the Xi [201], [202]. The current knowledge of waystation and boundary elements does not allow for genes subject to inactivation and escaping to be accurately predicted however indicating that additional factors

still remain to be identified contributing to the X chromosomes unique ability to undergo inactivation [176].

## 1.6 The model systems used to study XCI in mice and humans

The studies of XCI that have provided the insights described throughout this chapter were primarily performed in mouse ESCs which have formed the basis of the current paradigm of how XCI occurs and how Xist functions. Mouse models have offered numerous insights into XCI through both in vitro and in vivo studies, and the indispensable nature of Xist for mammalian viability was first demonstrated through mouse knockout studies [66]. Later studies of in vivo mouse knockouts established the mouse specific two-rounds of XCI, imprinted inactivation and subsequent random inactivation in the embryo proper, were both indispensable for viability [73], [203]. Xist has traditionally been shown to be indispensable for both types of mouse XCI, however additional Xist independent mechanisms for establishing dosage compensation in the epiblast may exist and act collaboratively with the Xist-initiated pathways [204]. Studies of human XCI were traditionally limited to *in vitro* models but it was established that humans, like most eutherian mammals, only undergo one wave of XCI during development. It was proposed that prior to XCI, a general 'dampening' of gene expression from both X chromosomes occurred as a temporary solution to dosage inequality [205]. The processes of mouse XCI and differentiation were shown to be interdependent, with studies suggesting that XCI depended upon and promoted differentiation [206]. This meant that studies of Xist activity in mice can only be effectively performed in ESCs as otherwise *Xist* did not induce gene silencing. The connection between the differentiation on *Xist* activity was proposed to be a result of its

interactions with differentially regulated factors such as the SATB1/2 proteins, expressed during this narrow window for inactivation [207]. The developmental window for XCI was associated with known pluripotency factors OCT4, KLF4 and NANOG all found to influence mouse *Xist* and Tsix [208], [209]. There was even evidence that the role of *Xist* binding factors shifted with differentiation, as CIZ1 was found to be essential for *Xist* localization in somatic cells but dispensable in stem cells [105].

Human *XIST* was shown to be less dependent upon the state of differentiation of its cells and was shown to effectively silence genes along the X chromosome in differentiated somatic cell lines [110], [112]. This has allowed for *XIST* to be studied in inducible somatic cells lines which were the predominant model for studying *XIST* in humans [111], [112], [210]. The ability of *XIST* to function in somatic cells and autosomally in humans has led it to be studied as a powerful therapeutic tool for individuals living with Down syndrome or large scale chromosome duplications [211], [212]. It has not yet been possible to consistently recapitulate the process of random XCI through cell reprogramming to an embryonic like state, as treatments were shown to permanently silence *XIST* expression while also being unable to return the Xi to its pre-inactivated condition, potentially due to the retention of Xi associated chromatin marks [213]–[217].

#### 1.7 The HT1080 model system used to study XIST activity

The purpose of the research performed in this thesis was to gain a novel understanding of how *XIST* itself functions without the confounding effects of the factors that affected *XIST* activity or

independently promote XCI (described in sections 1.5). Previous members of the Brown lab created a model system for studying *XIST* as an inducible transgene integrated into the autosome of a male somatic cell line (HT1080) [112]. It was demonstrated that the induced *XIST* transgene, which was the sole source of *XIST* RNA in the cells, produced a localized RNA cloud, silenced a proximal reporter gene *in cis* and enriched the local chromatin with Xi associated heterochromatin features [111], [112].

A series of HT1080 cell lines with inducible *XIST* cDNA constructs integrated into unique locations throughout the genome was generated by first transfecting with the Tet repressor (TetR) expressing element pcDNA6/TR, which integrated in two unique HT1080 cell lines. The TetR expressing HT1080 cell lines (2-3 and 2-12) were subsequently transfected with an FRT integration site containing plasmid, pFRT/lacZeo, which randomly integrated into the genome. Single cell clones with single FRT integrations were identified by Southern blotting. TetR was continuously expressed by a CMV promoter ensuring that the subsequently integrated *XIST* cDNA construct, controlled by a Tet Response Element controlled construct through a Tet-Off system, remained transcriptionally inactive [218].

The 'Full *XIST*' cDNA construct was cloned by successive integrations of the *XIST* cDNA sequence into the pcDNA5/FRT/TO construct. The sequence of *XIST* exon 1 (Genebank accession number M97168) extending from nucleotides 105 to 10747 was amplified from genomic DNA and cloned into the EcoRV site of the pcDNA5/FRT/TO vector. The sequence corresponding to exons 2-8 of the short *XIST* isoform was obtained from cDNA amplified and integrated 15bp downstream of the 3' region of exon 1 using the NotI and XhoI restriction enzymes. The 3'cDNA sequence of *XIST* was based on the short cDNA isoform of *XIST* that has

2.8kb at the 3' end of *XIST* spliced out [65], [76]. A short ~100 bp sequence of the pcDNA5 vector sequence was expressed at the 5' end of the *XIST* cDNA construct downstream as a result of the cloning protocol leaving that sequence spanning the distance from the CMV promoter to *XIST* integration site.

The pcDNA5/FRT/TO plasmid containing the *XIST* cDNA construct was integrated into the TetR expressing HT1080 cells with single FRT sites through transient co-transfection with the Flp-In T-Rex containing pOG44 plasmid. Clones were selected for Hygromycin resistance through the expression of the *Hyg* gene (HPH) by a CMV promoter upon the successful integration of the pcDNA construct. The location of the 8 autosomal *XIST* integration constructs were identified by inverse polymerase chain reaction (PCR) performed by members of the Brown lab [112]. An Xq *XIST* integration system was additionally generated from the HT1080 cell line, F55, that contained an X-linked FRT site received from Yan and Boyd who created and characterized the cell line (position in UCSC Human Genome Browser 2004 assembly 113,493,021) [219]. The initial random integration of the FRT sites resulted in three of the eight constructs being located in the intronic regions of genes (1p, 8p, 12q) with the rest being located kilobases from the nearest known gene. The surrounding chromatin environment also varied, with 3 constructs located in G-dark bands (3q, 7p, 15q) while the rest were in G-light bands. The sites of *XIST* integration are listed in table 1.2.

| Name (location) | Cell Line    | Nearest gene to integration | G-banding |
|-----------------|--------------|-----------------------------|-----------|
| 1p              | 2-3-1.0d     | MACF1 (intronic)            | Light     |
| 3q              | 2-12-0.5#3   | CLDN1 (10kb away)           | Dark      |
| 4q              | 2-12-0.5#8   | DCHS2 (55kb away)           | Light     |
| 7p              | 2-3-1.0 #5   | BBS9 (20kb away)            | Dark      |
| 7q              | 2-3-0.5+3 #1 | MTERF (215kb away)          | Light     |
| 8p              | 2-3-0.5a     | AGPAT5 (intronic)           | Light     |
| 12q             | 2-12-1.0 #14 | FAM222A (intronic)          | Light     |
| 15q             | 2-3-1.0 #12  | LINC00052 (89kb away)       | Dark      |
| Xq              | F55-6B1      | HTR2C (150kb away)          | Dark      |

 Table 1.2 XIST Integration constructs in HT1080 cell line

*XIST* transcription was induced in these HT1080 cells by treating the cells with doxycycline (Dox) which bound and released the TETR homodimers blocking the CMV promoter [218], [220]. Initially,  $1\mu$ g/ml Dox was determined to be sufficient to induce *XIST* RNA expression in the HT1080 3q cell line (2-12-0.5#3) and produce a localized and unified *XIST* RNA cloud and this concentration of Doxycycline was found to be sufficient in all the other cell lines [112]. All

9 of the induced *XIST* integration sites could produce a clear *XIST* RNA cloud that was depleted for transcriptional activity as evidenced by an overlapping Cot-1 RNA hole.

Members of the Brown lab observed that 4 days of *XIST* cDNA induction with doxycyline in the 3q cells resulted in the 80% reduction of fluorescence of a proximal EGFP reporter gene [112]. Numerous time points of *XIST* induction extending up to 31 days were examined, though EGFP fluorescence was found to reach a lower plateau day 7 of induction with its fluorescence ~10% of pre *XIST* induction levels. This indicated that the majority of the effects of *XIST*-mediated silencing of the proximal reporter had occurred after ~4 days of dox induction. Subsequent work published by Dr. Jakub Minks adopted the five-day induction period when examining the EGFP reporter gene and additionally observed a strong silencing effect of a ~80% decrease in the transcript levels by qPCR [111]. The five day period of dox induction (5ddox) subsequently became standard operation procedure for other members of the Brown lab when examining the inducible *XIST* constructs in the HT1080 cell lines. The induced *XIST* RNA in 3q became enriched with some chromatin marks associated with the Xi, such as H4K20me1, while becoming depleted for marks associated with transcriptional activity such as acetylation of H4 [112].

Previous work by Jennifer Chow had shown that by 4 days of induction a reporter gene proximal to induced *XIST* underwent significant silencing [112]. Work by Christine Yang identified four genes on 8p with coding SNPs that underwent silencing after five days of *XIST* induction in the 8p *XIST* inducible HT1080 cell line (2-3-0.5a). These genes, in order of proximity to the *XIST* cDNA constructs, were *CTSB*, *DLC1*, *SLC25A37* and *STC1* (table 1.3).

| Gene            | Transcript start (hg38) | Location (G-band) | Distance from XIST |
|-----------------|-------------------------|-------------------|--------------------|
|                 |                         |                   |                    |
| CTSB            | 11,842,524              | 8p23.1            | 5.11 Mb            |
|                 |                         |                   |                    |
| DLC1            | 13,083,361              | 8p22              | 6.35 Mb            |
|                 |                         |                   |                    |
| <i>SLC25A37</i> | 23,528,956              | 8p21.2            | 16.80 Mb           |
|                 |                         |                   |                    |
| STC1            | 23,841,929              | 8p21.2            | 17.11 Mb           |
|                 |                         |                   |                    |

Table 1.3 Distal genes that undergo allele-specific silencing by XIST induction from8p(23.1)

## **1.8** CRISPR, theory and application for modifying *XIST*

There was a need to ensure maximal genetic and epigenetic similarity between the novel deletion constructs to be generated during this research and the Full length *XIST* construct. Previous work had relied on the Flp-In recombinase to create different inducible *XIST* constructs [112] However this research project began shortly after the advent of CRISPR technology as a novel alternative that allowed for novel deletion constructs to be generated directly by modifying Full *XIST* and was adopted for these reasons.

CRISPR (clustered regularly interspaced short palindromic repeats) is the adaptive immune defense of bacteria and archaea that developed to remove pathogenic sequences from the hosts genome and provide pathogen resistance [221]–[223]. Its potency as a tool for gene editing was realized when it was demonstrated to be functional with eukaryote cells, and could be easily

directed to produce double strand breaks of targeted sequences in the genome [224]-[226]. Of the three types of CRISPR system, the Type II includes the most commonly used system, Cas9. The Cas9 nuclease is derived from Streptococcus pyogenes (SpCas9) and cleaves a targeted DNA sequence when bound to two RNAs, the CRISPR RNA (crRNA) and the trans activator RNA (tracrRNA) [227], [228]. It was quickly determined, however, that these two RNAs could be fused together into a single guide RNA (sgRNA or gRNA) that comprised a targeting sequence complementary to the sequence of DNA to be cleaved and a Cas9 interacting domain [228]. In order for a DNA sequence to be targeted by Cas9 it was determined that a 23-17nt complementary RNA sequence was needed that was adjacent to a short nucleotide sequence bound directly by the Cas9. This Cas9 binding sequence was termed the proto-spacer adjacent motif (PAM) which differ based on the type of Cas9 enzyme being used. The commonly used Streptococcus pyogenes Cas9 enzyme used in this research recognized the sequence NGG, with a much weaker affinity for NAG, though numerous other PAM recognizing enzymes were subsequently developed, including the highly permissive xCas9 (NG) [229]–[231]. Cas9 unwinds DNA at the PAM sequence and checks for complementarity between its gRNA and the DNA sequence. If complementarity occurred and formed an R-loop then the HNH endonuclease domain of Cas9 cleaves the DNA bound to the gRNA while the RuvC-like endonuclease domain cuts the non-complemented DNA strand. These two cleavage events produce a blunt end cut of the double stranded DNA sequence ~3bp from the PAM sequence [228], [232]. Numerous research groups have demonstrated that the Cas9 directed system for mediating double stranded breaks was highly efficient with most complementary target sequences being successfully cut resulting in a high efficiency when generating homozygous modifications within the genome [224], [233]–[236]. A major initial concern of the Cas9 system was the fear of off-target

cleavage of sections of the genome caused by sequence complementation to the gRNA sequence [237], [238]. However this concern has decreased subsequently with numerous algorithms and online tools such as GUIDE-seq testing preemptively for potential off target effects [239]–[241]. Specificity of a target sequence was found to be dramatically increased by shortening the targeting sequence of a gRNA to lengths of 17-19nt as well as other modifications to the gRNA such as using locked nucleic acids [237], [242]–[244].

#### 1.9 Thesis objective

The overarching goal of this thesis was to create a map of the regions of XIST responsible for its various functions, determine what connections if any exist between the various silencing pathways initiated by XIST and to create a model system that will facilitate further research of said pathways. In chapter 3 the activity of XIST expressed from different chromosomes in the male HT1080 cell line was tested to identify the best model system available for further studies and to identify how the integration site from which XIST was expressed affected its reorganization of the surrounding chromatin. In chapter 4 a series of isogenic partial XIST constructs were generated to allow for the functional significance of individual regions of XIST to be determined and to identify regions of XIST that were crucial for its stability and ability to localize to a single unified region of the nucleus. In chapter 5, to identify the function of each region of XIST and identify potential interdependencies between the XCI associated pathways, the partial XIST constructs were tested relative to full-length controls. The objective was to identify the key elements of XIST crucial for gene silencing, chromatin remodeling, and perinucleolar localization of XIST. The final objective was to identify how key factors in the XCI affected each other and other aspects of XIST function and to create an outline of the steps in the pathways initiated by XIST. In chapter 6 chemical inhibitors were used to disrupt histone deacetylation and polycomb group complexes and the resulting effects on gene silencing and chromatin remodeling were studied. The results obtained in this thesis significantly advances understanding of how XIST functions in humans, provides both a map and validated model system to direct future research and presents a model of the relationship between key pathways of human XCI.

# **Chapter 2: Methodologies**

The methods listed in this chapter were written as step by step protocols for the various experiments performed in this work. Each section describes a specific protocol, and it was intended that a researcher following the instructions in this protocol would be able to replicate any of the work or research performed with the need for referring to additional sources minimized. The figures that complement these protocols were included in their relevant sections rather than at the end of the chapter to make each section as intuitive and easy to follow as possible.

## 2.1 Cell culture and materials

The human immortalized male fibrosarcoma cell line, HT1080, and its derivatives were grown at 37°C in a humidified incubator maintaining 5% CO<sub>2</sub> in Dulbecco's modified Eagle medium (DMEM) purchased from Gibco. The DMEM media was supplemented to 10% v/v Fetal Calf Serum (FCS), 1X Penicillin-Streptomycin, 1X non-essential amino acids and 2mM L-Glutamine all from Gibco. Induction of the doxycycline inducible *XIST* constructs within the HT1080 cell lines was performed by further supplementing the DMEM media with 1µg/ml doxycycline.

The female immortalized fibroblast cell line IMR-90 was cultured in Eagle's Minimum Essential Medium purchased from Gibco supplemented with 10% v/v FCS, 1X Penicillin-Streptomycin, 1X non-essential amino acids and 2mM L-Glutamine all purchased from Gibco. The female immortalized cell line, HTERT-RPE1, was cultured in DMEM F-12 from Gibco supplemented with 10% v/v FCS from Gibco and 10µg/ml hygromycin B from Thermo Fisher. Both female cell lines were cultured at 37°C in an incubator maintaining a 5% CO<sub>2</sub> concentration. Regular testing to ensure that all cell lines were free of mycoplasma contamination was performed using the MycoFluor Mycoplasma Kit from Thermo Fisher.

#### 2.1.1 Transferring adherent cells between containers

As all the cells worked with were adherent and generally grew in monolayers, they would frequently need to be diluted or 'split' to a lower concentration in their flasks. Splitting was performed when the cells reached > 80% confluency by adding ~15 $\mu$ l of 0.25% v/v Trypsin-EDTA (Invitrogen) per cm<sup>2</sup> of the container's surface. These volumes were often rounded to the nearest 10 $\mu$ l or 100 $\mu$ l depending on the pipette being used. The Trypsin was gently rocked back and forth to ensure it completely covered the surface area of the container and was then allowed to sit for 2-4 minutes. Once the cells began to slough off, a volume of phosphate buffered saline (PBS), twice that of Trypsin, was added to the flask and was gently pipetted up and down to ensure an even mixture of the cells in suspension and transferred into a fresh container. An appropriate volume of media was then rapidly added to the fresh container.

If the cells were being harvested for some other purpose other than simply maintaining the cell lines then the cell suspension was transferred into a 1.5ml microcentrifuge tube and was centrifuged at 1000 relative centrifugal force (rcf) for 5 minutes. This caused the cells to form a pellet at the bottom of the tube, allowing the supernatant to be aspirated off. This cell pellet was either stored at -70°C for subsequent analysis or was used for subsequent analysis immediately. Typically, the number of cells in the tube were estimated based on the initial confluency of the cells growing in the container. Table 2.1 summarizes the volumes used throughout this project as well as the approximate number of HT1080 cells that will fill each type of container if allowed to reach confluency.

| Container       | Surface area              | Number of HT1080 cells     | Volume of | Volume Trypsin |
|-----------------|---------------------------|----------------------------|-----------|----------------|
|                 |                           | when confluent             | media     | to split cells |
| T25             | 25 cm <sup>2</sup>        | ~2.5 x 10 <sup>6</sup>     | 6 ml      | ~400 µ1        |
| 6-well plate    | $\sim 10 \text{ cm}^2$    | $\sim 1.0 \text{ x } 10^6$ | 3 ml      | ~200 µl        |
| 24-well plate   | $\sim 2 \text{ cm}^2$     | $\sim 2.0 \text{ x } 10^5$ | 1 ml      | ~30 µl         |
| 96-well plate   | $\sim 0.3 \text{ cm}^2$   | $\sim 3.0 \text{ x } 10^4$ | 300 µl    | ~5-10 µl       |
| 100mm petri     | $\sim$ 57 cm <sup>2</sup> | $\sim 5.7 \text{ x } 10^6$ | 12ml      | ~900 µ1        |
| dish            |                           |                            |           |                |
| 60mm petri dish | $\sim 22 \text{ cm}^2$    | $\sim 2.2 \text{ x } 10^6$ | 5ml       | ~300 µl        |

| Table 2.1 Surface area of cell growth containers along with the approximate volu | mes of |
|--|--------|
| growth media and Trypsin used for culturing cells.                               |        |

A summary table of the dimensions and working volumes of the various tissue culture containers used throughout this work. The name of each container is listed in the first column followed by the surface area available for adherent cells to grow on and the approximate number of HT1080 cells (the main type used in this research) that could live on that surface area. The working volume of media that cells were grown in was listed in the fourth column and the volume of 0.25% Trypsin that was found to be most effective for splitting confluent populations of cells was listed in the final column

# 2.2 Procuring RNA for analysis

#### 2.2.1 RNA extraction and purification

The purification of RNA was performed using Trizol from Invitrogen following the harvesting of cell pellets. Trizol was added to the cell pellets in 1.5ml centrifuge tubes and caused the cells to rapidly disintegrate into the Trizol. Within a few minutes until the mixture appeared homogeneous. If the RNA was not to be immediately purified these Trizol mixtures were stored at -70°C.

To extract RNA from the cell pellets, the cell-Trizol mixtures sat at room temperature for 5 minutes before 200µl of chloroform was added and the tubes were shaken vigorously for 15 seconds by hand until the mixture was a clear opaque pink. The mixture was allowed to sit for 2-3 minutes at room temperature before being centrifuged at 12,000 rcf for 30 minutes at 4°C. A clear colourless aqueous phase containing the RNA and consisting of slightly less than half the total volume of the mixture was visible after being centrifuged. The aqueous phase was transferred to a fresh tube by gently pipetting the mixture and avoiding any of the white sediment found in the interphase.

To precipitate the RNA, 500 µl of isopropanol was added to the roughly 500 µl of RNA containing liquid, gently mixed by inversion, and then allowed to sit for 10 minutes at room temperature. The mixture was then centrifuged at 12,000 rcf for 10 minutes at 4°C. This caused the RNA to precipitate out of solution and form a white translucent pellet at the bottom of the tube. The liquid phase was carefully poured out to avoid disturbing the pellet and 1ml of 75% ethanol was pipetted onto the pellet. The 75% ethanol was obtained by diluting pure alcohol with

DEPC treated double distilled water (ddH2O). The pellet was then disturbed from the bottom of the tube by gently vortexing or shaking by hand. The mixture was then centrifuged at 8,000 rcf for 5 minutes at 4°C to pellet the washed RNA. The ethanol mixture was poured off gently and the final drops were pipetted away, leaving only the RNA pellet. To avoid the pellet completely drying out, 30-50  $\mu$ l of ddH2O was immediately added to the pellet once the last amount of 75% ethanol was removed. The RNA mixture was then heated to 55°C for 10 minutes to fully resuspend the pellet.The RNA was either used immediately, stored at -20°C if it was to be used within a couple weeks, or else it was stored at -70°C.

#### 2.2.2 DNAse treatment and reverse transcription of RNA

Prior to reverse transcription the RNA obtained from the Trizol purification described in 2.2.1 was treated with DNAse1 from Roche according to their protocol to digest any DNA which may contaminate the RNA. To digest the DNA, the RNA mixture was treated with the appropriate amount of 10x DNAse1 buffer, 10 units (1.25 µl) of DNAse1 and 1.25µl of 40 U/µl RiboLock RNase Inhibitor from Thermo Fisher. The samples were then incubated at 35°C for 20 minutes, allowing for the degradation of any DNA. The DNAse1 enzyme was then heat inactivated by incubating the RNA mixtures at 75°C for 10 minutes.

The RNA was then reverse transcribed using the M-MLV reverse transcriptase from Invitrogen. The recipe used for the reverse transcription reaction is listed below in Table 2.2. The reverse transcription mixtures were allowed to sit at room temperature for 5 minutes then were incubated at 42°C for 2hours before the M-MLV enzyme was heat inactivated at 95°C for 5 minutes.

| Reagent                  | Volume for single reaction |
|--------------------------|----------------------------|
| 5x First Strand Buffer   | 4 µl                       |
| 0.1 mM DTT               | 2 µl                       |
| 1.25 μM dNTPs            | 2 µl                       |
| 50 µM Random hexamers    | 1 μl                       |
| Ribolock RNAse inhibitor | 0.5 μl                     |
| M-MLV                    | 1 μl                       |
| RNA                      | 9.5µl                      |

Table 2.2 M-MLV reaction conditions for a single RNA reaction

# 2.3 Measuring transcript levels by RT-qPCR

Two different sets of qPCR reagents were used throughout the course of this research, Hot Start Fermentas Taq and Promega Go Taq G2. The change in reagents was due to the discontinuation of Fermentas Taq. In both cases the fluorescent nucleic acid dye EVA green (20x from Biotium) was used to measure amplification of cDNA sequences. The recipes used to prepare the reaction mixtures were listed in table 2.3 and 2.4. The primer sequences used for the various qPCRs described throughout this work were included in the appendix primer table A.2.

| Reagent                                  | Volume for single reaction (µl) |
|--|---------------------------------|
| 25mM dNTPs                               | 0.16                            |
| 50mM MgCl <sub>2</sub>                   | 2                               |
| 10x Buffer                               | 2                               |
| 25nmol/ul sense and antisense primer mix | 0.2                             |
| cDNA template                            | 1.5                             |
| HS Fermentas Taq                         | 0.16                            |
| EVA green 20x from Biotium               | 1                               |
| ddH <sub>2</sub> O                       | 12.98                           |

 Table 2.3 Recipe for 20ul RT-qPCR using Fermentas products

| Reagent                                  | Volume for single reaction (µl) |
|--|---------------------------------|
| 25mM dNTPs                               | 0.16                            |
| 25mM MgCl <sub>2</sub>                   | 1.2                             |
| 5x Buffer                                | 4                               |
| 25nmol/ul sense and antisense primer mix | 0.2                             |
| cDNA template                            | 1.5                             |
| Promega HS GoTaq G2                      | 0.1                             |
| EVA green 20x                            | 1                               |
| ddH <sub>2</sub> O                       | 11.84                           |

 Table 2.4 Recipe for 20ul RT-qPCR using Promega GoTaq G2 products

The reaction mixtures were prepared on ice and were run in either MicroAmp Optical 96-well plates or 8-cap strips from Applied Biosystems. Hot Start Fermentas Taq reactions were heated to 95°C for 3 minutes, then completed 40 repeating cycles of 95°C for 20 seconds, 60°C for 30 seconds and 72°C for 1 minute. The Promega reactions were treated the same way, only at 96°C instead of 95°C. The cycle threshold (CT) was calculated using the software included with the various qPCR machines. The primary interest of this research was to compare the relative quantification (RQ) of a gene of interest (i.e. *XIST*) in different cell types or conditions. The RQ would be calculated using an endogenous control gene (e.g. *PGK1*). The difference in CTs between the gene of interest and control would be used to determine the  $\Delta$ CT value for a sample, which in turn would be used to determine the RQ according to the formula below.

$$RO = 2^{-\Delta\Delta CT}$$

$$\Delta \Delta CT = \Delta CT_{(test \ sample)} - \Delta CT_{(reference \ sample)}$$

$$\Delta CT = CT_{(gene \ of \ interest)} - CT_{(endogenous \ control \ gene)}$$

In certain cases, multiple endogenous control genes were used to calculate the RQ to avoid potential confounding effects caused by variation of a single endogenous control. When the two endogenous controls were used their CT values were combined into a geometric mean based on their respective primer efficiencies. Primer efficiency was calculated by graphing the Ct values of serially diluted cDNA on a log scale in Microsoft excel [245]. The calculation used to calculate the relative gene expression using multiple controls as well as the efficiency calculation shown below [246], [247].

 $Quantity of XIST RNA (RQ) = \frac{(Efficiency of XIST primers)^{\Delta Ct of XIST}}{Geo. Mean(Efficiency of control primers^{\Delta Ct of controls})}$ 

 $Primer \ Efficiency = 10^{\left(\frac{-1}{slope \ of \ Ct \ values \ on \ log \ scale}\right)} - 1$ 

#### 2.4 PCR to amplify DNA sequences

The amplification of DNA sequences of interest by PCR used hot start Taq DNA polymerase from Invitrogen. PCR mixtures were prepared according to the recipe listed in table 2.5. PCRs were initially heated to 95°C for five minutes, then completed 35 cycles alternating between 95°C for 30 seconds, 55-62°C for 30 seconds then 72°C for 1 minute per kilobase of sequence to be amplified (minimum time of 1 minute). The only exception to this procedure came when preparing PCR products for pyrosequencing. The PCR products for pyrosequencing were required to be biotinylated. Only the *STC1* and *CTSB* primers were directly biotinylated, in the cases of the *DLC1* and *SLC25A37* pyrosequencing primers, the 'reverse' primers in each mixture were fused to a M13 universal primer (Appendix primer table A.2). 15 PCR cycles were carried out using the M13 fusion primers and the resulting PCR product was used as a template for a subsequent reaction using biotinylated M13 universal primers for 20 cycles. Following the final amplification cycle of all PCRs, the products were incubated for 5 additional minutes at 72°C then either stored at -20°C or used immediately for subsequent purposes. In instances where amplifying the DNA template proved difficult the PCRs were supplemented with 20% Betaine (from Sigma) by volume, adding 5ul of Betaine and 14.5  $\mu$ l of water rather than the full 19.5  $\mu$ l of water. Betaine acts to equilibrate Tm differentials between the two types of base pairings, which in practice results in improved target synthesis and specificity across difficult sequences while preventing non-specific products from being produced [248].

| Reagent                                  | Volume for single reaction (µl) |
|--|---------------------------------|
| 25mM dNTPs                               | 0.2                             |
| 50mM MgCl <sub>2</sub>                   | 0.75                            |
| 10x Buffer                               | 2.5                             |
| 25nmol/ul sense and antisense primer mix | 1                               |
| DNA template                             | 1                               |
| Taq DNA Polymerase                       | 0.125                           |
| ddH2O                                    | 19.5                            |
|  |                                 |

Table 2.5 Recipe for 25ul PCR

#### 2.4.1 PCR product purification

The purification of PCR products was carried out using the Qiaquick PCR purification kit (commercially available from Qiagen). In every purification 20µl of PCR product was combined with 100µl of the column binding buffer, Buffer PB, and mixed gently by pipetting. The mixed 120µl were then placed on top of the membrane of the assembled Qiaquick column and collection tube provided in the kit and centrifuged at 17,900 rcf for 1 minute. The column and collection tube were removed from the centrifuge and the flow through disposed. The column was filled with 750µl of wash buffer, Buffer PE, and centrifuged at 17,900 rcf for 1 minute. The column and collection tube were removed from the centrifuge and the flow through discarded. The column and tube were again spun at 17,900 rcf for 1 minute to ensure that as much of Buffer PE had been removed as is possible. The column was removed from the collection tube and placed in an appropriately labelled 1.5ml microcentrifuge tube. Typically, 50µl of ddH<sub>2</sub>O was added to each column, though this could be adjusted to increase or decrease the final concentration of the DNA as the column only held 10µg of DNA. The ddH<sub>2</sub>O was added to the column and allowed to sit for 1 minute before being centrifuged for 2 minutes at 17,900 rcf. The column was then discarded and the DNA solution could be either stored at -20°C or used for subsequent applications immediately.
## 2.5 Pyrosequencing

Pyrosequencing was performed on biotinylated PCR products obtained by the amplification of specific sequences of cDNA. All the PCRs of pyrosequencing primer pairs were performed at a melting temperature of 58.3°C and extension times of 1 minute. 12µl of PCR product was used in each pyrosequencing reaction in a PyroMark MD machine from Qiagen using CDT tips according to the recommended protocol. The PCR products combined with 38µl of Pyromark's binding buffer, 35µl of ddH<sub>2</sub>O and 2µl of Streptavidin Sepharose High Performance beads from GE Healthcare. The mixture was then agitated at 1400 rpm to allow the DNA to bind to the beads. While the PCR/bead mixture was shaking, 0.144 µl of 25nmol/µl SNP primer was combined with 11.856µl of Pyromark Annealing buffer in the well of a Pyromark optical plate. The containers of the Pyromark workstation were filled with the following: 1 with 100ml of 70% ethanol, 2 with ~100ml denaturing solution, 3 with ~120ml of wash buffer and 5 with ~100ml of  $ddH_2O$ . The workstations comb was rinsed in  $ddH_2O$  to remove any residual dust or contaminants, then the PCR product containing mixture was sucked up through the comb of the vacuum apparatus. The DNA and beads remained held against the outside of the semi permeable membrane of the comb by the negative pressure generated by the vacuum apparatus. As soon as the liquid was sucked up, the comb was gently placed in each of the numbered 1 - 3 containers of the workstation, allowing the PCR product coated beads to be treated with the various liquids. The comb of the vacuum apparatus was then carefully lowered directly over the optical plate, without touching the primer mixture. The vacuum was then disconnected using the switch at the front of the workstation and the teeth of the comb were lowered directly into the wells of the optical plate, swirling gently to ensure that the beads were released into the primer containing

annealing mixture. The combs were then carefully removed, and the optical plate was heated to 80°C for 2 minutes. The plate was then loaded into the PyroMark MD machine and the CDT tips were filled with their respective primers, enzyme and substrate components from Qiagen. The Pyromark software was used to run the PyroMark MD machine and calculate the allelic ratio present within the PCR product.

Throughout this work the relative strength of induced gene silencing was measured by comparing the change in allelic contribution to total gene expression across the test condition relative to the change induced by the control condition. The formula below was used to calculate the normalized strength of silencing between a test condition (e.g.  $\Delta$  A) and a control condition (e.g. Full *XIST*) in cells prior to and following the treatment.

$$Silencing Strength = \frac{(treatment induced change in allelic contribution)_{test}}{(treatment induced change in allelic contribution)_{control}}$$

Change in allelic ratio =  $allelic \ contribution_{post-treatment} - allelic \ contribution_{initial}$ 

This simple formula is useful as a value of 1 indicates that the test condition did not differ from the control, while a value of 0 indicates that the treatment did not affect the relative contribution of an allele to the total gene expression. Statistical significance was calculated using the Mann Whitney U-test comparing multiple biological replicates performed in technical duplicates [249]. When multiple testing was being performed the significance was adjusted based on the Bonferroni correction [adjusted p = p/(number of tests)] [250].

## 2.6 Protein Extraction and western Blotting

### 2.6.1 Protein Extraction from cells

Whole cell protein was extracted from cells by treating cell pellets with Radio Immunoprecipitation Assay buffer (RIPA buffer). The cell pellets were harvested by Trypsin treatment as described in section 2.1.1 and the remaining liquid on top of the pellet was aspirated off. RIPA buffer (1M NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1%SDS, 50mM pH 7.4 Tris in ddH<sub>2</sub>O) prepared according to the Cold Spring Harbor [251] protocol was supplemented with 1µl of Protease Inhibitor Cocktail from Roche per 100µl of RIPA buffer. RIPA buffer was added at a concentration of 500µl per ~2x10<sup>6</sup> cells (the approximate number of HT1080 cells in a nearly confluent T25 flask). Samples were gently agitated by shaking for 30 minutes at 4°C to allow for the breakdown of cells, then were centrifuged at max speed (>14,000 rcf) at 4°C for 20 minutes. The supernatant containing the protein component was transferred to a fresh microcentrifuge tube to be loaded immediately into an SDS-PAGE gel or stored at -20°C.

#### 2.6.2 SDS-PAGE gel

In preparation for running protein samples on the SDS-PAGE gel a 1.5mm glass spacer plate and cover plate were sandwiched together in a casting clamp and placed upon a gel casting stand (all from Bio Rad). The volume within the plates was filled with water and allowed to sit for several minutes to ensure that there was no leakage. As the water sat in the plates, 10ml of 12%

acrylamide lower running gel consisting of 4.2ml of 29:1 acrylamide/bis-acylamide (Bio Rad), 2.5ml of 4x lower gel buffer (1.5M Tris and 0.4%SDS in distilled water, pH of 8) and 3.3ml of distilled water mixed thoroughly then 40µl of 10% ammonium persulfate (prepared freshly each time) as well as 10µl of TEMED from Fisher Scientific was added. The water was poured from the watertight gel plates and this gel solution was pipetted in, filling the volume to within a cm of the top. The final cm of the plates was filled with 100% isopropanol. The lower gel solidified for a minimum of 30 minutes at which point the researcher consumed a coffee. 3ml of upper gel mixture consisting of 0.75ml 29:1 acrylamide/bis-acylamide, 0.45ml of 4x upper gel buffer (0.5M Tris and 0.4%SDS in distilled water, pH of 6.8), 1.8ml of distilled water, 10µl of of 10% ammonium persulfate and 5µl of TEMED (Fisher Scientific) was prepared. The isopropanol covering the lower gel was poured off completely, exposing the top of the lower gel. The upper gel mixture was poured on top, filling the remaining centimeter in the gel plates. A 1.5mm gel comb that accompanied the spacer plate was then inserted into the top of the plates. This upper gel mixture solidified over the course of 30 minutes.

The solid gel and its glass case were removed from the casting stand and clamp and were fitted into the electrode assembly (Bio Rad) with the short plates facing inwards; the comb was then removed. If only one gel was being run then a mock spacer plate and short plate combination was inserted on the other side of the electrode assembly with short plate on the inside. This created a watertight container that was fitted into the Bio Rad buffer tank and filled to overflowing with Running buffer. Running buffer was produced 1 liter at a time by diluting 100ml of 10x Running buffer (0.25M Tris, 1.92M Glycine and 1% w/v SDS in distilled water) in 900ml distilled water. The tank itself was then filled with the remaining Running buffer. Once

the gel was prepared for electrophoresis, 15µl of each protein extract was mixed with 2X SDS gel-loading buffer (4% SDS, 0.2% bromophenol blue, 20% glycerol and 200mM dithiothreitol) according to the Cold Spring Harbor protocol and the resulting mixture was heated to 95°C for 2-3 minutes [252]. The protein mixture was loaded into each well of the gel, and 5µl BenchMark Pre-stained Protein Ladder from Thermo Fisher was loaded into one of the wells abutting the samples. The gel was run at 180 volts until the bromophenol blue dye in the various samples had nearly reached the bottom of the gel but had not run off. The bottom band of the Pre-stained Protein Ladder corresponded to ~10kDa and so it was ensured that there was always a significant gap between this band of the ladder and the bottom of the gel to ensure that the histones had not been run off the gel.

## 2.6.3 Western blotting

As the gel in section 2.7.2 was running, fresh Transfer buffer was prepared and cooled either in a freezer or on ice. Transfer buffer was prepared by combining 100ml of 10x Transfer buffer (0.25M Tris and 1.9M glycine in distilled water and stored at 4°C), 200ml of methanol and 700ml of distilled water. The gel was removed from the tank, electrode assembly and spacer plates. The gel was immersed for 5 minutes in cold Transfer buffer and the transfer system was prepared. The Bio Rad transfer holder cassette was arranged with the black frame (which would face the negative electrode) on one side and the clear frame (which would face the positive electrode) on the other. A foam pad from Bio Rad was placed on top of the black frame and upon that was placed a piece of the filter paper supplied in the Nitrocellulose/Filter Paper Sandwich from Thermo Fisher. The now soaked gel was placed carefully upon that piece of filter paper so

as to be completely supported by it and not touching any part of the sponge beneath. Carefully and using metal tweezers, the 0.2µm nitrocellulose was first briefly soaked in Transfer buffer then placed on top of the gel. Pressure was applied with a smooth flat implement to remove any gaps that might exist between the nitrocellulose and gel and a small amount of transfer buffer was poured on top to ensure the nitrocellulose was completely soaked. The other piece of filter paper was placed on top of the nitrocellulose to cover it completely. An additional piece of spongy foam was then placed on top of the filter paper. The clear frame was then folded over and fastened to the black frame and the cassette was fitted into the Bio Rad Core Assembly Module for protein transfer with the black of the cassette facing the black side of the module. The Core Assembly Module was inserted into the Bio Rad protein transfer tank and a cold ice pack was placed in the transfer tank as well. The tank was then filled with Transfer buffer and was hooked up to a Bio Rad PowerPac HC power supply. Protein transfer was carried out at 90 volts for one hour at 4°C with constant gentle agitation of the Transfer buffer using a magnetic stir bar to ensure dispersion of heat.

After the one hour the protein bound nitrocellulose membrane was transferred into a container filled with 40ml of blocking buffer consisting of Tris buffered saline (TBS) with 0.1% v/v Tween-20 and 3% Bovine Serum Albumin (BSA: 1.2 grams in 40ml TBS). The membrane was gently rocked back and forth for one hour in the blocking buffer to prevent the non-specific absorption of antibodies into the nitrocellulose. After blocking, the nitrocellulose was placed inside a 50ml falcon tube containing 20ml of primary antibody solution (3% w/v BSA, 0.1% v/v Tween-20 in TBS plus the relevant antibodies, diluted as listed in table 2.6). The membrane was incubated in this solution overnight at 4°C on a tube rotator to allow even coating of the entire

surface of the membrane with antibodies. The next day the membrane was retrieved from its falcon tube and was washed four times for 5 minutes each in room temperature TBST (0.1% v/v Tween-20 in TBS). During each wash the nitrocellulose membrane was placed on a rocker to help wash off any unbound proteins. The nitrocellulose membrane was then placed in a container containing 20ml of secondary antibody solution (3% w/v BSA, 0.1% v/v Tween-20 in TBS plus either goat anti-mouse or goat anti-rabbit fluorescently labelled secondary antibodies listed in table 2.6). The membrane was incubated in the secondary antibody solution for one hour at room temperature on a gently rocking rocker. Following this secondary incubation, the membrane was washed twice with TBST then once in TBS for 5 minutes each at room temperature. This was done to remove any unbound secondary antibodies. The protein and antibody bound membrane was then imaged at the relevant wavelengths using an LI-COR Odyssey machine from BioAgilytix and the software package, Image Studio.

| Primary western blotting antibodies               |               |           |           |              |          |  |  |  |  |
|---|---------------|-----------|-----------|--------------|----------|--|--|--|--|
| Target  | Company       | Catalog # | Lot #     | Host species | Dilution |  |  |  |  |
| H3K27me3  | Diagenode     | C15410069 | A1824D    | rabbit       | 1:500    |  |  |  |  |
| H4K8ac  | Abcam         | Ab45166   | EP1002Y   | rabbit       | 1:2500   |  |  |  |  |
| H2AK119ub   | Sigma-Aldrich | 05-678    | 2591849   | mouse        | 1:2000   |  |  |  |  |
| H4  | Upstate       | 25296     | 07-108    | rabbit       | 1:2000   |  |  |  |  |
| β-actin   | Invitrogen    | MA5-15739 | TC263471  | mouse        | 1:5000   |  |  |  |  |
| Secondary fluorescent western blotting antibodies |               |           |           |              |          |  |  |  |  |
| Target  | Company       | Catalog # | Lot #     | Host species | Dilution |  |  |  |  |
| Rabbit  | LI-COR        | 925-68071 | B81113-01 | goat         | 1:10,000 |  |  |  |  |
| antibodies  |               |           |           |              |          |  |  |  |  |
| Mouse   | LI-COR        | 925-32210 | 926-32201 | goat         | 1:10,000 |  |  |  |  |
| antibodies  |               |           |           |              |          |  |  |  |  |

## Table 2.6 List of antibodies used for Western Blotting and their concentrations

The table lists the targets of western blotting that were examined in chapter 6 as well as the product information for each antibody and its species of origin. Antibodies were subdivided between those that were used for primary and secondary labelling. All of the targets of the primary antibodies had been shown to work in human models. The dilutions listed in the final column indicate the concentration of the antibody in solution during the Western blotting reaction and was determined through trial with multiple concentrations. Only antibodies that proved effective for Western blotting were shown here, and a complete list of the antibodies used throughout this work is in the appendices (table A.3 and A.4).

# 2.7 Generating gRNA plasmids

The methods described in section 2.7 are summarized in graphical format in figure 2.1, which illustrates the key steps involved in creating a targeted gRNA expressing plasmid described in greater detail throughout sections 2.7.1-2.7.3.

# 2.7.1 Selecting gRNA targeting oligonucleotide sequences

To generate guide RNAs (gRNAs) that would allow for the specific modifications desired throughout this project the gRNA containing plasmid pSPgRNA from Charles Gersbach was ordered (Addgene plasmid # 47108 ; http://n2t.net/addgene:47108 ;

RRID:Addgene\_47108)[253]. This plasmid uses a U6 promoter directly upstream of two BbsI restriction sites that allowed for the oriented integration of a gRNA sequence by attaching the sequence CACC to the 5' end sense oligonucleotide and AAAC to the 5' end of the anti-sense gRNA oligonucleotide. The gRNA target sequences were generated using the E-CRISP online design tool (e-crisp.org) produced by the German Cancer Research Center. The human genome *Homo sapiens* GRCh38 was selected for reference and the FASTA sequence for the *XIST* cDNA region of interest was pasted into the tool. The 'strict' default setting was selected and then several advanced options adjusted. The minimum guide length was set to 19 and maximum to 20. The 5' preceding base requirement was set to 'G' as this is essential for transcript initiation by the U6 promoter of the pSPgRNA. Off-target analysis was set to use Bowtie2 with high sensitivity for off-targets. Additional off target effects on the puromycin resistance gene were selected against, as the Cas9 plasmid contained a PuroR gene. The maximum number of results

was increased to over 50 to provide a range of options. The gRNAs were selected based on their location within the region of interest as well as their specificity (S-score)[241] and efficiency (E-score)[254], in that order of importance. The gRNA sequences were then checked by BLASTn to check for potential off-target effects not identified in the program. If no serious concerns were uncovered then sense and antisense targeting oligonucleotide sequences were ordered from IDT.

# 2.7.2 Cloning targeting oligonucleotides into pSPgRNA vector

The targeting oligonucleotides were resuspended in double distilled water (ddH<sub>2</sub>O) to a concentration of 100µM. The oligos were phosphorylated by incubating with T4 polynucleotide kinase according to the recipe in table 2.7 at 37°C for 30 minutes. The T4 kinase was inactivated at 95°C for 5 minutes, then the temperature was ramped down to 25°C at 5°C/min to ensure proper hybridization. This mixture was then diluted 200-fold to a final oligonucleotide concentration of 50nM.

| Reagent                                  | Volume for single reaction |
|--|----------------------------|
| Sense targeting oligonucleotide          | 1 μ1                       |
| Anti-sense targeting oligonucleotide     | 1 μ1                       |
| T4 ligation buffer, 10x with ATP (1mM)   | 1 μ1                       |
| (Thermo Fisher)                          |                            |
| T4 polynucleotide kinase (Thermo Fisher) | 1 μ1                       |
| ddH2O                                    | 6 μl                       |

 Table 2.7 Phosphorylation of targeting gRNA oligonucleotides

The pSPgRNA plasmid was digested using the BbsI restriction enzyme from New England Biolabs. 1ug of plasmid was digested by 10 units of BbsI enzyme in 1x NEBuffer 2.1 (diluted down from 10x NEBuffer 2.1) for 3 hours at 37°C and then heat inactivated at 65°C for 20 minutes. The plasmid was then dephosphorylated as an additional step to ensure that it would only be able to re-ligate back together with the phosphorylated targeting oligonucleotides. To ensure that the plasmid was dephosphorylated 1x Antarctic phosphatase buffer was added to the digested pSPgRNA mixture along with 10 units of Antarctic phosphatase enzyme from New England BioLabs. This reaction mixture was incubated at 37°C for 30 minutes then was inactivated by heating the reaction mixture to 80°C for 2 minutes. The digested and dephosphorylated plasmid pSPgRNA plasmid was then column purified using the QIAquick PCR Purification Kit from Qiagen and resuspended in ddH<sub>2</sub>O.

1.5 ng of targeting oligonucleotide and 100ng of pSPgRNA linear plasmid were incorporated into the ligation mixture, to give a 2:1 molar ratio of targeting sequence to vector. The

concentration of pSPgRNA was determined using a Ultraspec 2000 from Pharmacia Biotech measuring the 260nm and 280nm absorbances according to the manufacturer's directions. Each *XIST* targeting gRNA was generated using the following reaction setup, and the reaction proceeded either for 2 hours at room temperature or overnight at 4°C.

| Reagent                                  | Volume for single reaction     |
|--|--------------------------------|
| Digested and dephosphorylated pSPgRNA    | volume needed for 100ng of DNA |
| Targeting oligonucleotides (50nm)        | 2 µl (each)                    |
| 10x T4 DNA Ligase Buffer (Thermo Fisher) | 2 µl                           |
| T4 DNA ligase (Thermo Fisher)            | 1 μl                           |
| ddH2O                                    | Fill volume up to 20ul         |

 Table 2.8 Reaction conditions for ligating target sequence into digested oligonucleotides.

After ligation, the reaction mixture was then treated with T5 Exonuclease (New England Biolabs) to clean up linear DNA after ligation reactions. 10X NEBuffer 4 was added to the reaction conditions to a concentration of 1X (8 µl to the 20µl ligation reaction) along with 1µl of T5 Exonuclease. The exonuclease reaction incubated at 37°C for 30 minutes according to the New England Biolabs protocol before being used immediately for transformations.

# 2.7.3 Transforming gRNA plasmid into DH5α Competent Cells

Subcloning Efficiency<sup>™</sup> DH5α Competent Cells were ordered from Invitrogen and all the gRNA plasmids generated through this procedure were transformed into these cells. Transformations

were carried out using the manufacturers protocol. These cells were stored at -70°C and during transformations were thawed on ice and 50µl was transferred into a 1.5ml microcentrifuge tube. To these cells was added 1.5µl of the gRNA solution created in 2.7.2 (~7.5ng of gRNA plasmid). The cells were then gently mixed and allowed to sit on ice for 30 minutes before undergoing heat shock by placing the tube in a 42°C water bath for 20 seconds. The cells were then returned to their ice filled box for a further 2 minutes. 750µl of 37°C SOC media made according to the Cold Spring Harbor protocol [255] and was added to the DH5 $\alpha$  cells. The cells were incubated at 37°C for 1 hour on a shaker rotating at 225rpm. LB agar plates were prepared according to the Cold Spring Harbor protocol [256] and supplemented with Ampicillin to a final concentration of 100ug/ml. Volumes of 50µl, 150µl and 300µl of gRNA transformed DH5 $\alpha$  cells were each added to the ampicillin LB agar plates. The DH5 $\alpha$  cell mixtures were spread evenly across the surface of their LB agar plate. The plates were incubated overnight at 37°C and the next day colonies were picked to be tested for the successful integration of the gRNA target sequence into the pSPgRNA plasmid.



# Figure 2.1 Summary of method used to generate targeted gRNA expressing plasmids.

A graphical overview of the process used to insert a targeting DNA sequence into the pSPgRNA plasmid. Two BBsI endonuclease target sequences allowed for a target sequence (blue) to be integrated upstream of the remainder of the gRNA repeat and tracrRNA sequence (green). The gRNA expression was driven by a U6 promoter (yellow). The target sequence of the gRNA was ligated into the BBsI digested gRNA sequence to create a plasmid that expressed the custom gRNA.

#### 2.8 Plasmid purification

All plasmids were extracted from their host bacterial cells using the QIAprep Spin Miniprep Kit according to the Qiagen's directions. Bacterial cultures were grown overnight in 2ml of LB media with ampicillin at a concentration of 100µg/ml. These cultures were pelleted by spinning down at 10,000 rcf for 3 minutes at room temperature before being resuspended in 250µl of lysis buffer (Buffer P1) and adding 250µl of Buffer P2. The reaction was mixed gently until clear and 350µl of Buffer N3 was added; the ~850µl of reaction volume was then mixed again. The reactions were centrifuged at 18,000 rcf for 10 minutes. The remaining supernatant was then applied to the QIAprep spin columns and centrifuged for 1 minute at 18,000 rcf and the flow through was discarded. 500µl of wash buffer (Buffer PB) was then added to the column which was centrifuged for 1 minute at 18,000 rcf. Then 750µl of the other wash buffer Buffer PE was added to the column before being centrifuged twice for 1 minute each at 18,000 rcf, with the flow through discarded after each wash. The column was then placed in a fresh 1.5ml microcentrifuge tube and was filled with 50ul of ddH<sub>2</sub>O. The water was allowed to sit in the tube at room temperature for 1 minute before being centrifuged for 2 minutes at 18,000 rcf. The column was then discarded and the plasmid containing water in the the microcentrifuge tube was either stored at -20°C or used immediately for subsequent applications.

# 2.9 DNA electrophoresis through agarose gel

Whether checking PCR product or digested plasmid sizes the DNA was run on an 1% w/v agarose gel in Tris-acetate-EDTA (TAE) buffer. TAE buffer was produced by diluting a stock

solution of 40x TAE (1.6M Tris, 0.52M sodium acetate and 80mM EDTA with the pH adjusted to 8.0 using acetic acid in distilled water) 1:39 in distilled water. The 1x TAE was used as the solvent for UltraPure agarose powder from Invitrogen. The agarose was dissolved into TAE by microwaving the agarose and TAE in an Erlenmeyer flask with a volume at least 4x greater than the volume of TAE until it started to boil gently. The contents were then swirled gently to ensure that the agarose had completely melted into the TAE. SYBR Safe Stain from Invitrogen was added to the mixture to a concentration of 1/10,000 and swirled gently to ensure equal distribution of the stain. Before the gel began to solidify, it was poured into a Bio Rad gel casting tray with the complementary comb inserted to be nearly touching the bottom of the casting tray. The gel solidified for at least 30 minutes before the comb was removed and the gel was placed in Bio Rad gel electrophoresis system filled with TAE at room temperature. The edge of the gel with the loading wells was placed next to the cathode (black electrode) so that the DNA would flow through the length of the gel towards the anode (red electrode). DNA samples were mixed with DNA Gel Loading Dye at a ratio of 5:1. A DNA ladder of appropriate size (usually GeneRuler 100bp Plus DNA Ladder from Thermo Fisher) was loaded next to the DNA samples. The volumes loaded into an agarose gel varied based on the application but in general 5µl of DNA sample would be loaded into each well of a gel. The samples were run at ~80 volts using a Bio Rad Power Pac 300 power supply. Once the leading edge of the loading bromophenol blue dye had travelled a significant distance along the gel the power supply was turned off and the gel transferred into an Alpha Innotech AlphaImager 2200 gel imaging machine. The UV light of the imaging machine caused the SYBR safe stained DNA to fluoresce, which was captured using the built-in camera system. The machine was controlled and the photos were analyzed using the AlphaImager HP software suite for Windows.

#### 2.10 DNA extraction from mammalian cells

#### 2.10.1 DNAzol DNA purification from Life Technologies method

To purify DNA using DNAzol (Life Technologies) cells were pelleted as described in section 2.1.1. Room temperature DNAzol was then added to the cell pellet, at a concentration of  $100\mu$ l per  $10^6$  cells according to the manufacturer's instructions. The tube was inverted several times to allow for the complete lysis of the cell pellet, which rapidly dissolved into the DNAzol, forming a clear and colourless solution. The DNA was precipitated by adding a volume of pure ethanol half that of the volume of DNAzol (ex 50µl of ethanol was used to precipitate DNA from  $100\mu$ l of DNAzol). The tube was mixed by inversion, then sat for 3 minutes at room temperature. The tube was then centrifuged at 4,000 rcf for 2 minutes at room temperature to pellet the precipitated DNA. The DNAzol and ethanol was then removed and the DNA pellet was washed twice with 1ml of 75% ethanol in ddH<sub>2</sub>O and gently agitated each time until the pellet released from the bottom of the tube and floated freely within the ethanol solution. After the final wash the ethanol was removed and the DNA pellet was resuspended in a volume of 20-100µl of ddH<sub>2</sub>O, depending on the application and size of the pellet.

# 2.10.2 Mouse Homogenizing Buffer (MHB) method of DNA purification

This method of DNA purification is based on a protocol shared with the Brown lab by Andrea Korecki of the Simpson lab in the Department of Medical Genetics at the University of British Columbia. This protocol was designed to digest mouse ear notches but was highly efficient at producing usable DNA samples from cell pellets (section 2.1.1). Immediately before using Mouse Homogenizing Buffer (MHB: 50 mM KCL, 10 mM Tris- HCL at a pH 8.3, 2 mM MgCl2, 0.1 mg/ml gelatin, 0.45% IGEPAL CA-630 from Sigma-Aldrich and 0.45% Tween 20 from Sigma-Aldrich all suspended in ddH20) Proteinase K (10mg/ml Sigma Aldrich) was added to the buffer to a concentration of 120µg/ml. After extensive troubleshooting, volumes of 300µl to 500µl of MHB supplemented with Proteinase K were effective for converting a pellet of roughly 10<sup>6</sup> cells into a solution ideal for subsequent PCRs. After the MHB Proteinase K solution was added to a cell pellet, the solution was incubated overnight at 55°C. The next day the Proteinase K was inactivated by heating the solution to 95°C for 5 minutes. The DNA solution was then either stored at -20°C or was immediately used for subsequent purposes.

# 2.11 Transfection and selection of deletion construct containing HT1080 cell lines

The CRISPR system used throughout this research involved a standard SpCas9 expressing plasmid (pSpCas9(BB)-2A-Puro) and sgRNA expressing plasmids (pSpgRNA) that could be transfected into cells together. The Cas9 was codon optimized for human models and expression was driven by a CMV enhancer and chicken-β-actin promoter. A SV40 nuclear localization signal was fused to the Cas9 to ensure its transport into the nucleus [225]. The pSPgRNA produced a *Streptococcus pyogenes* specific guide RNA under the control of a U6 promoter designed to be transcribed in mammalian cells [253]. Digesting the plasmid with BbsI resulted in non-matching 4 nt sticky ends on the linearized plasmid allowing for an orientable integration of

a target sequence into the gRNA. The final size of the gRNA plasmid following the integration of a target sequence was 3.2kb and the Cas9 plasmid was 9.2kb.

All of the *XIST* deletion constructs that were generated in this thesis were a result of transfecting gRNA and Cas9 plasmids into the male fibrosarcoma cell line HT1080 cell line 2-3-0.5a 1c.1, which carries a doxycycline inducible Full *XIST* cDNA construct integrated into chromosome 8p. The following methodology was arrived at through extensive troubleshooting of conditions to precisely excise a specific region of the *XIST* transgene using transient selection. To excise specific regions, three plasmids were transfected into the HT1080 cells, two gRNA expressing plasmids targeting the regions of *XIST* to be cut as well as a Cas9 expressing plasmid that provided transient puromycin resistance. The purified ~3.4kb gRNA plasmids were described in section 2.7 and 2.8. The ~9.2kb Cas9 plamid, pSpCas9(BB)-2A-Puro (PX459), was obtained from Feng Zhang (Addgene plasmid # 48139 ; http://n2t.net/addgene:48139 ; RRID:Addgene 48139)[225]. Figure 2.2 at the end of this section provides a graphical outline of

the steps involved in generating clonal potential deletion cell lines by transfection and selection.

The three plasmids were transfected into the HT1080 cells using Lipofectamine 3000 reagents (ThermoFisher). The transfection protocol that was adapted from the manufacturers protocol. The concentration and purity of the purified gRNA and Cas9 plasmids were determined by measuring the 260nm and 280nm absorbance in an Ultraspec 2000 from Pharmacia Biotech. As the protocol for lipofectamine suggested that the DNA should be concentrated to greater than  $0.5\mu g/\mu l$  the purified plasmids were typically concentrated in their ddH<sub>2</sub>O solutions in a SpeedVac system from Savant Integrated with the vacuum itself produced by a motor from A Vac Industries.

The Cas9 expressing plasmid that was to be used contained a puromycin resistance gene and was the only one of the three plasmids that could be selected for. At the time the molar ratios recommended based on the literature suggested 1:2 to 1:6 molar ratios of Cas9 plasmids to individual gRNA plasmid [237], [257], [258]. It was eventually determined that a molar ratio of 1:4.5:4.5 (mass ratio of 1:1.5:1.5) for the Cas9 and two gRNA plasmids was optimal. A 1µg plasmid solution was created prior to each transfection reaction which consisted of 0.375µg of each gRNA plasmid and 0.25µg of the Cas9 plasmid (figure 2.2 A).

The day before transfection, the HT1080 cells to be modified were seeded into the wells of a 24well plate to be  $\sim$ 40-50% confluent by the next day. When performing the transfection protocol, Lipofectamine 3000 was first diluted into 25µl volumes of Opti-MEM media (Gibco) and the resulting solutions were shaken vigorously for several seconds then allowed to sit at room temperature while the plasmid solutions were prepared. For each grouping of plasmids to be transfected two different Lipofectamine 3000 dilutions were tested, either consisting of 0.75µl or 1.5µl of the reagent (figure 2.2 B). The plasmid solutions were made by adding a total of 1µg of plasmids to 50µl of Opti-MEM which was additionally supplemented with 2µl of the P3000 reagent, included in the Lipofectamine reagent kit. Equal volumes of the plasmid solution (~27µl) were then added to the ~26µl Lipofectamine 3000 solutions. The plasmid-Lipofectamine mixtures were incubated for 15-20 minutes at room temperature before being gently pipetted into the appropriate well of the 24-well plate. The media was gently swirled to ensure an even distribution of the DNA-Lipofectamine particles. The transfected cells were grown for at least 16 hours before their media was removed and replaced with prewarmed media containing puromycin from Sigma at a concentration of  $1\mu g/ml$ . Selection lasted for 72 hours (+/- 1 hour)

with the puromycin containing media being changed out on the second day. Immediately following selection, the cells were washed twice with PBS and split with Trypsin as described in section 2.1.1, and 20-40 cells were transferred each to two 100mm petri dishes. This number was determined by counting the cells in a  $<10\mu$ l volume on the surface of the petri dish prior to the addition of media. The remaining cells had their DNA extracted using MHB as described in section 2.11.2. 12ml of media was added to the droplet of cells in each 100mm petri dish and was gently shaken to disperse the cells evenly across the surface of the plate. Over the next 10-14 days the cells were regularly observed as they grew to form colonies larger than 120 cells (figure 2.2 C).

Colonies of sufficient size were picked by removing the media and washing the petri dishes gently with PBS twice. Vacuum grease was applied to the base of PYREX Cloning Cylinders (Corning) that were then placed around individual colonies. Into each cloning cylinder was applied  $30\mu$ l of 0.25% v/v Trypsin that dissociated the cells from the plate surface over the course of 2-4 minutes. 70µl of DMEM HT1080 media was added to each of the flasks and was gently pipetted up and down to help dissociate the cells. The cells from each colony were then transferred into wells of a 24 well plate filled with 1ml of media. The cells in each well were allowed to grow until they reach near confluency before being split. Roughly a quarter to half of the cells being split were moved to a fresh well with the remainder having their DNA extracted with MHB as described in section 2.10.2. The extracted DNA was tested for the desired modification to *XIST* by PCR as described in section 2.4 using primers spanning the region of interest (primers used to test  $\Delta$  were listed in appendices table A.1). The PCR products from the different colonies were run on agarose gels as described in section 2.10. If a cell line produced

only the correct sized PCR product then the product was column purified as described in section 2.4.1. The concentration of the purified PCR product was determined by measuring the 260nm and 280nm absorbance using the Ultraspec 2000 spectrometer from Pharmacia Biotech. Sanger sequencing was performed initially by the Integrated DNA Technologies (IDT) operating out of the BRC Sequencing Core at the University of British Columbia and subsequently by the company Genewiz. When preparing the samples for both companies the PCR products were diluted to the appropriate concentrations in  $ddH_2O$  and one of the  $\Delta$  primers (Appendix table A.1) used to create that PCR product was sent along to be used to initiate the sequencing reaction. It was important to determine whether different deletion cell lines for each type of deletion construct were from different progenitor cells and a result of unique CRISPR events. As the PCR targeting within the first exon of XIST could not distinguish between the transgene and endogenous gene, only cell lines with homozygous and uniform PCR products were sent for sequencing. The different cell lines for each deletion construct could only be confirmed to be unique if they were from different transfection treatments or if they carried unique sequences at the cut site. If multiple deletion cell lines could not be proven to be unique then one was randomly selected to be kept and the others were discarded, ensuring that every remaining deletion cell line for each type of construct was unique.





# gRNA and Cas9 expressing plasmids

A graphical summary of the steps involved in generating colonies of potential deletion constructs. A) The concentrations of plasmids expressing the two gRNAs (green and blue), Cas9 (orange) as well as P3000 was added to one volume of Opti-MEM. The lipofectamine 3000 reagent was added in 0.75 and 1.5 $\mu$ l volumes to two separate volumes of Opti-MEM B) The plasmid mixture was evenly distributed into the two lipofectamine 3000 reagent containing solutions. C) The plasmid and lipofectamine containing solutions were added to HT1080 cells in a 24-well plate. The next day the cells were treated with DMEM media supplemented with 1 $\mu$ g/ml puromycin for 3 days. The surviving cells were transferred to 100ml petri dishes at low numbers in DMEM free of puromycin and grew for ~2 weeks to form colonies that could be picked and tested.

#### 2.12 Fluorescently labelling cells (IF-FISH)

#### 2.12.1 Fixing cells on coverslips

Adherent cells were cultured under the same conditions on glass 22mm x 22mm square coverslips one to two days before they were to be permeabilized and fixed. At the start of the permeabilization and fixing procedure, coverslips were transferred into Coplin jars filled with cold CSK buffer (100mM NaCl, 300mM sucrose, 3mM MgCl<sub>2</sub>, 10mM PIPES) and making sure to note which direction the cell coated side of the coverslip was facing. The coverslips were then transferred into cold CSK buffer containing 0.5ml of Triton-X in 10ml of CSK buffer for 8 minutes. The permeabilized coverslips were then transferred into 10ml of 4% paraformaldehyde in PBS for 8 minutes at room temperature to fix them. The coverslips were washed for 3-4 minutes in 70% ethanol then stored in jars of 70% ethanol at 4°C.

#### 2.12.2 Creating fluorescent RNA probes through nick translation

The fluorescent probes to be used for RNA fluorescent *in situ* hybridization (FISH) were created from purified plasmids containing multi kilobase *XIST* sequences. Three probes were used in descending order of frequency: HBC1a, 3'*XIST* and G1a. HBC1a targets the 1.6kb of *XIST* spanning the A repeats to the very start of the Bh repeats. 3'*XIST* targets the latter 3.6kb *XIST* cDNA short isoform sequence. G1a is a 10kb plasmid that targets both the intronic and exon regions of *XIST* extending from intron 4 onwards. The Abbott Molecular nick translation protocol was used according to the manufacturer's (Abbott Molecular) instructions to nick translate these probes and label the resulting 50nt to 200nt fragments with either Green 496 or

Red 598dUTPs from the company Enzo. 1-2 $\mu$ g of plasmid was used in each nick translation reaction and the resulting probes were resuspended in 50 $\mu$ l of DEPC treated ddH<sub>2</sub>O and stored at -20°C in the dark as the fluorescently labelled probes were light sensitive and prone to photobleaching.

#### 2.12.3 Immunofluorescence and Fluorescent *in situ* Hybridization (IF-FISH)

Immunofluorescence (IF) and RNA FISH were performed jointly in most experiments. Permeabilized and fixed cells on coverslips were incubated face down on 100ul droplets of PBT consisting of 1% v/v Bovine Serum Albumin and 0.1% v/v Tween-20 in PBS supplemented with 0.4U/µl RiboLock RNAse Inhibitor (RI) from Thermo Fisher for twenty minutes. The coverslips were incubated at room in 100ul PBT supplemented with 0.4U/µl RI and 1ul of primary antibody for 4-6 hours at room temperature or overnight at 4°C. The primary antibodies (all used at 1:100 dilution) used successfully throughout this work were listed in the appendices, table A.3. To prevent the coverslip and antibody solution from drying, each coverslip was individually sealed between two layers of Parafilm, creating a semi-air tight Parafilm pocket.

Just prior to retrieving the coverslips from their Parafilm pockets, hybridization mixtures were created from 5µl of the nick translated fluorescent RNA probe mixed with 12µl of human Cot-1 and 2µl of Salmon Sperm DNA (both from ThermoFisher). The latter two ingredients prevented non-specific binding of the fluorescent probe within the cells. The hybridization mixtures were then completely dried in a SpeedVac system from Savant Integrated with the vacuum itself produced by a pump from A Vac Industries. While the probes were being dried, the coverslips

were retrieved from their Parafilm pockets and washed three times in room temperature PBST (PBS and 0.1% Tween-20) before being incubated for 1 hour at room temperature with the secondary fluorescently-labelled antibody (1:100 PBT supplemented with 0.4U/µl RNase Inhibitor, 1ul secondary antibody resulting in an antibody 1:100 dilution). The secondary antibodies used throughout this work were listed in the appendices, table A.4. From this point onwards the coverslips were kept in dark containers to avoid photobleaching. The coverslips were then washed for 5 minutes in PBST three times, to remove the unbound secondary antibodies. They were then fixed in 4% paraformaldehyde for 10 minutes. The coverslips were washed once in PBS to remove most of the paraformaldehyde then dehydrated by submerging them for two minutes each in 70%, 80% and finally 100% ethanol. The coverslips were then air dried in a dark container at room temperature for 10-20 minutes.

The dried probes were resuspended in 10ul of deionized formamide from Sigma and heated to 80°C for 10 minutes. To the RNA probes was added 10ul of hybridization buffer consisting of 20% v/v 20mg/ml Bovine serum albumin (Sigma Aldrich), 20% v/v Dextran Sulfate and 20% 4x SSC in DEPC treated ddH<sub>2</sub>O. The fluorescent probe mixture was gently stirred using the pipette tip and pipetted as a droplet on a piece of parafilm onto which a coverslip was placed, cell side down onto the droplet. A second piece of parafilm was placed on top and the edges were sealed creating a parafilm pocket where the probes hybridized overnight at 37°C. In the few instances where RNA FISH was performed without accompanying IF, the hybridization mixtures were prepared and dried as described, then the coverslips were retrieved from storage at 4°C and transferred directly into 100% ethanol to be dehydrated. The protocol for hybridization, whether

IF had been performed or not, was the same once the coverslips had been dehydrated in 100% ethanol.

The next day the coverslips were retrieved and incubated in a mixture consisting of equal volumes of deionized formamide (Sigma) and 4x SSC (Invitrogen) at 37°C for 20 minutes. The coverslips were then incubated in 2x SSC at 37°C and 1x SSC at room temperature, for 15 minutes each. DAPI staining was performed by placing the coverslips in a  $0.1\mu$ g/ml solution of DAPI in pure methanol at 37°C from 15 minutes. The excess DAPI was rinsed off in methanol and the coverslips were mounted on glass slides using 25µl of hard set antifade mounting media from Vectashield.

After letting the media harden, cells were imaged using a confocal fluorescence microscope, either a Leica DMI 6000B or Olympus FluoView FV1000, to capture and compile the fluorescent signals from the secondary antibody, RNA probe and DAPI at 100x magnification. Over the course of the project multiple different software packages were used with the DMI6000B. This resulted from the companies Leica and Olympus creating incompatibilities between the operating software and hardware, blocking cross compatibility to the competitors' products and removing backwards compatibility for their products to force additional purchases. Initially the Openlab software from PerkinElmer was used followed by MicroPublisher 5.0 RTV by Qimaging and finally Qimaging software from Teledyne. The FV1000 microscope was operated and images were captured using the FV10 ASW 4.2 Viewer software released by Olympus.

#### 2.13 Analysis of IF-FISH images by qualitative and quantitative methods

#### 2.13.1 Qualitative assessment of IF-FISH images

The distribution of chromatin marks or proteins at the XIST RNA cloud through IF-FISH imaging (section 2.12) was assessed by visual inspection. The enrichment of a heterochromatin mark at the XIST RNA cloud of a cell would be either defined as enriched relative to the average level in the nucleus, or not enriched. This method relied on the individual researcher's definition of what constituted a cell being "clearly enriched" which was defined as having a noticeably greater fluorescence than was observed in the surrounding area. The total proportion of cells in a group that were visibly enriched for a chromatin mark or protein at the XIST RNA cloud were used as a metric of how effectively XIST recruited that heterochromatin mark. To try and ensure consistency between different tests a set of example figures demonstrating enrichment and depletion were used to guide the researcher's calls. Examples of enriched and non-enriched cells used to guide calls of enrichment of H3K27me3 were shown in figure 2.3. A similar system was used to analyze the depletion of Cot-1, only with depletion and non-depletion being the two categories for cells. Qualitative assessments of the proportion of cells with XIST at the perinucleolar compartment or with diffuse XIST RNA signals were also performed throughout this work, and were far easier to perform as they involved categorical statements about the position of XIST RNA. When it was necessary to determine whether two populations of cells differed in their proportion of cells enriched for a given mark the Fisher exact test was used to determine the statistical significance of the difference between groups [259].



#### Example of H3K27me3 not enriched at XIST RNA

XIST (green) H3K27me3 (red) Merged

# Figure 2.3 Categorizing cells as either enriched or not enriched for factors at XIST RNA cloud

Four fluorescently labelled cells that were either visibly enriched or not enriched for H3K27me3 at their XIST RNA cloud in HT1080 cells expressing Full XIST from chromosome 8p. The column of images on the left were four cells that had clear enrichment of H3K7me3 at the XIST RNA cloud. The column of images on the right did not show any signs of H3K27me3 enrichment at the XIST RNA cloud. The colour images showed the merged channels of XIST (green), H3K27me3 (red) and DAPI (blue) within the nuclei of cells. Images showing greyscale *XIST* and H3K27me3 were included alongside the merged image to exemplify the distribution of the chromatin marks within the nucleus at the site of XIST RNA more clearly.

# 2.13.2 Quantifying the distribution of Xi associated factors at the *XIST* RNA cloud relative to the nuclear background

The quantification of IF-FISH images was carried out using the program ImageJ produced and maintained by Fiji (Fiji is Just ImageJ) contributors. To analyze the photos taken required the plugin Broadly Applicable Routines (BAR). Prior to analysis the identities of all the images were blinded by assigning them random alpha numeric names (e.g. A1, A2 etc.). This was done by various other members of the Brown Lab who kept the document of the folders identities private. To analyze a single cell, an image within one of these folders was opened and converted to a composite image from its source format (Image > Color > Make Composite). The image was then despeckled (Process > Noise > Despeckle), a median filter process that decreases the random variability inherent with high magnification fluorescent microscopy by normalizing central pixel intensity to its neighboring 8 pixels (figure 2.4 A). The straight-line tool was used to draw a line through the nucleus and XIST RNA signal (figure 2.4 B). To try to prevent operator biasing of where to draw the line, the following rules were used to determine where the line was drawn. First, the line had to pass through the point or points of maximal XIST signal intensity. Second, the line was drawn to be of maximum length without intersecting the nucleoli or leaving the nucleus. A drawback of this second rule was that the number of pixels measured in each cell differed. Once the line was drawn, a multi-channel plot profile (BAR > Analysis > Multichannel Plot Profile) of the intensities for each of the colours (red, green and blue) was produced (figure 2.4 C). Each pixel bisected by the drawn line was recorded as a point along the x axis of that graph, with the intensity of all 3 colours recorded at each point. The data of position and three colour channel intensities were then saved as a .csv file (data > save data) and 60 cells were

analyzed for each condition. Combining all of these files together was done by Bradley Balaton from the Brown Lab who wrote the necessary codes. The csv files were amalgamated into a large tsv file using a short unix code and Rscript to record both their folders and files names of origin as additional columns in the combined table.

Analysis of the data was performed using R studio. The three colour channel intensities at each pixel (the base unit of measurement) for each cell were divided into two categories, XIST +ve or XIST -ve. This was done by determining the range of XIST RNA (green) fluorescent intensity in every cell. The pixels that had a level of green intensity greater that 50% of the maximum along the range of XIST fluorescence were defined as XIST +ve (figure 2.4 C). Those that had a level of green intensity below 25% of the maximum along the range of XIST fluorescence were defined to be XIST -ve. Those pixels between 50% and 25% of the maximum green intensity value were discarded to better delineate the two groups.

To quantify whether a factor's concentration in the *XIST* +ve region of the nucleus was different from the control *XIST* -ve region a two-sample standard score (z-score) calculation was performed for every cell [260], [261]. In each cell the mean ( $\overline{X}$ ) and standard deviation ( $\sigma$ ) of the factor of interest's fluorescent intensity (red) was calculated at the *XIST* +ve and -ve regions of the nucleus along with the number (n) of measurements/pixels the XIST -ve and +ve regions. With these values the z-score was calculated for each cell according to the following formula.

$$z - score = \frac{\bar{X}_{XIST+ve} - \bar{X}_{XIST-ve}}{\sqrt{\frac{\sigma_{XIST+ve}^2}{n_{XIST+ve}} + \frac{\sigma_{XIST-ve}^2}{n_{XIST-ve}}}}$$

The z-score represents the difference between the fluorescent intensity between the XIST +ve and -ve regions based on the combined variability of the data sets and accounts for the different variance and number of measurements between the two regions. The sign of the z-score indicated whether the factor is enriched (+) or depleted (-) at the XIST RNA and the magnitude of the score indicates the relative effect strength of the XIST RNA compared to the other processes governing that factor's distribution. A z-score of 0 in this context indicated that there was no effect of XIST RNA on the distribution of a factor within the nucleus. A value of +2 would indicate that the factor of interest was more enriched at the XIST RNA by two standard deviations compared nuclear average. The 60 z-scores for each condition were then compared statistically using the Mann-Whitney test to determine the likelihood that a given two conditions differed in their population distribution. This test was chosen given the continuous nature of the z-scores and the lack of prior knowledge about the distribution of the z-scores within a group [249]. The z-scores for each deletion construct were graphed as a boxplot with the individual zscores for each cell superimposed on top. The average z-score for each construct as well as its standard deviation and significance calculated were included in accompanying tables.



| Figure 2.4 ( | Duantification | of the relativ | e distribution | of factors at | t the <i>XIST</i> RNA | cloud |
|--------------|----------------|----------------|----------------|---------------|-----------------------|-------|

An outline of the steps involved in quantifying the distribution of an IF labelled factor, in this case H3K27me3, at the *XIST* RNA cloud in an individual cell. In all instances images of cells

were multichannel RGB images originally with *XIST* labelled in green, the mark or factor of interest (H3K27me3) in red, and DAPI staining in blue. A) The original images of cells were converted to composite type images and despeckled using the native features of ImageJ. B) A line was drawn through the *XIST* RNA cloud and as much of the nucleus as possible, avoiding the nucleolar regions. C) The red green and blue pixel intensity at each point along the length of the line was turned into a graph using the BAR ImageJ plugin multichannel plot profile feature. The points were subdivided into those that had *XIST* RNA fluorescence >50% of the range of *XIST* intensity (*XIST* +ve region) and those <25% *XIST* intensity (*XIST* -ve region). The H3K27me3 fluorescence intensity was measured in both regions of the cell and a summary of the data is shown in the table at the bottom. This data was used to calculate the two factor z-score for each cell.

# Chapter 3: Validation of XIST Inducible Model

The candidate Thomas Dixon-McDougall designed, performed and analyzed all of the experiments that led to the results presented in this chapter. The only exception is figure 3.3 which summarizes the combined results of the 2015 publication on the *XIST* integration constructs which is included to illustrate the reason for selecting the 8p inducible model system.

#### 3.1 Introduction

The random inactivation of one of the X chromosomes in eutherian females in the periimplantation stage of development (~E6.5 in mice) is established by a long non-coding RNA gene, *XIST* [1], [262]. In humans, *XIST* is expressed as a transcript up to 19kb in length, that coats its X chromosome of origin, silencing most of the ~1,000 known genes across 155Mb [64], [65], [76]. The coating of the X chromosome by *XIST* causes the majority of the chromosome to be restructured as dense transcriptionally inactive heterochromatin known as the Xi [2], [79], [125]. *XIST* RNA causes the Xi to become enriched for numerous heterochromatin marks that include H3K27me3 [37], ubH2A [128], H4K20me1 [39], H3K9me3 [45], [150], MacroH2A [41] and SMCHD1 [40]. *XIST* RNA also facilitates the removal of marks associated with active transcription such as histone acetylation [263]. *XIST*-mediated repression of transcription was viewed by performing FISH against Cot-1 RNA to observe the noticeable absence of transcriptional activity at the *XIST* RNA cloud in the nucleus, a phenomenon referred to as a 'Cot hole' [110]. The repression of large regions of chromosomes can be directed by integration of a
transcriptionally active *XIST*, which has led *XIST* to be seen as a powerful therapeutic tool [112], [125], [211], [264].

The activity of *XIST* and the process of XCI are also affected by the unique nature of the X chromosome itself, X-linked genes, the developmental window during which XCI occurred as well as additional unknown factors. The enrichment of LINE-1 elements on the X chromosome was proposed to effect *XIST* spreading and gene silencing in the Lyon Repeat Hypothesis [122], [196], [197]. Specific X-linked factors have also been demonstrated to affect to the process of XCI including those within the *XIC* and more distal features like the lncRNA *FIRRE* [174] and the microsatellite repeat *DXZ4* [97], [185]–[187], [265] thought to be involved in regulating the structure of the X chromosome during inactivation. The factors *Rnf12*, *Jpx*, and *Ftx* located in the XIC contribute to XCI through *XIST* up-regulation [185]–[187]. These factors illustrate how unique aspects of the X chromosome itself contribute to its own inactivation, and how these confounding factors undermine any analysis designed to specifically learn about *XIST*. This led to the conclusion that investigating *XIST* function would best involve *XIST* being induced from an autosome, to limit the potential confounding effects of other X-linked factors on its activity.

The choice of cell lines in which *XIST* can be studied is also a crucial consideration. Most research of XCI comes from studies in mouse stem cells, as the timing of onset of random XCI at ~6.5 days post coitum in mice makes *in vivo* studies challenging [32], [33], [266]. It was observed in mice that the state of pluripotency played an important factor in mediating the ability of *Xist* to establish XCI resulting in a narrow developmental window when XCI could occur potentially through *Xist* binding to ESC specific factors such as SATB1/2 [125], [207]. This temporal window for establishing XCI is not as restrictive in humans however, as induction of

*XIST* in somatic cell lines has been shown to cause silencing [110]. Human stem cells used to study XCI following the attempted reactivation of the Xi have not proved very effective, with results being highly variable and resulting in aberrant and non-random XCI [216], [217], [267]. The challenges associated with using human ES cells to investigate *XIST* and the effectiveness of somatic cell models for *XIST* induction led to the latter being selected.

To study the function of *XIST* outside of its natural context the male fibrosarcoma HT1080 cell lines with doxycycline inducible *XIST* cDNA constructs were used as models. Across the various unique inducible *XIST* HT1080 cell lines there were eight unique autosomal integrations as well as an X-linked integration that was initially received from Dr Yan [219]. The location of the initial integration of the FRT (Flippase recognition target) sites in the HT1080 cells was random, and as a result the environment surrounding the inducible *XIST* construct for each type of cell varied significantly in terms of the surrounding chromatin environment and proximity to other genes (see table 1.2). The different types of inducible *XIST* integration cell lines were referred to by the location in the genome that the inducible *XIST* construct was integrated as identified by previous members of the Brown Lab through inverse PCR: 1p, 3q, 4q, 7p, 7q, 8p, 12q, 15q and Xq.

Angela Kelsey, Christine Yang and myself worked together to investigate the functionality of the *XIST* in the various autosomal integrations of the HT1080 cell lines to determine the extent to which the functionality of *XIST* was affected by chromatin context even outside of the X chromosome. As it was expected that different integrations would result in *XIST*-mediated inactivation occurring to different degrees, it was hoped that this analysis would provide novel insights into potential pathways of *XIST* activity. This work would additionally identify which

integration site offered the best context for further analysis of *XIST* function. Once a preferred model system was identified, the dynamics of *XIST* could be examined in greater detail. Three broad criteria were analyzed over various time points of induction: *XIST* transcript levels, *XIST* induced silencing of distal genes and *XIST* induced chromatin remodeling. These three tests offered novel insights into how rapidly *XIST* can start facilitating inactivation following upregulation as well as potentially reveal connections between chromatin remodeling events as well as gene silencing.

## 3.2 Results

## 3.2.1 Enrichment of H4K20me1 and H3K27me3 at induced XIST RNA clouds

*XIST* functionalities in the various integration sites in the HT1080 cell lines were analyzed for their ability to enrich the chromatin encompassed by *XIST* RNA with Xi associated heterochromatin, the first mark examined being H3K27me3. IF-FISH analysis was performed as described in section 2.12 of the methods chapter and recorded the proportion of cells with visible enrichment of H3K27me3 at the *XIST* RNA cloud as described in section 2.13.1. All of the induced (5ddox) *XIST* integrations (1p, 4q, 7p, 7q, 8p, 12q, 15q and Xq) were tested, except for 3q, which had previously been analyzed for enrichment of H3K27me3. The location of the *XIST* construct dramatically affected its H3K27me3 enrichment, with scores ranging from 10% to 84% across the eight induced cell lines (n > 50, figure 3.1 A). The Xq integration (F55-6B1-N6) That acted as a reference of XIST activity when expressed from an X chromosome rather than an autosome had by far the largest proportion of cells with clear enrichment of H3K27me3 (84%, n=50). The 8p integrated *XIST* construct (2-3-0.5a) had the strongest H3K27me3 enrichment of the autosomal integrated *XIST* constructs (32%, n = 255), followed by the 4q integration (2-12-0.5#8) with an enrichment at 28% (n = 99, figure 3.1 A). The integration at 1p (2-3-1.0d-1C) showed the lowest proportion of cells enriched for H3K27me3 at only 10% (n = 100).

In addition to analyzing H3K27me3, several of the *XIST* inducible HT1080 cell lines were also examined for their ability to enrich the *XIST* RNA cloud with the less extensively studied Xi-associated heterochromatin mark, H4K20me1. As Angela Kelsey had previously tested several of the cell lines for their enrichment of H4K20me1, the remaining 6 untested inducible *XIST* cell lines were tested following 5ddox induction for their H4K20me1 enrichment: 1p, 4q, 7p, 7q, 8p and Xq.

Xq (F55-6B1-N6) cell line enriched the highest proportion of its *XIST* RNA clouds with H4K20me1 (66%, n = 50). The 8p integrated *XIST* construct had the strongest proportion of cells with enrichment 26% (n = 195, figure 3.1 B). Both the 4q and 7q *XIST* constructs showed only questionable enrichment for H4K20me1 in 3% to 4% (n = 98 & 100) of their cells, indicating that the location of *XIST* was a major determinant in its recruitment of this mark. The remaining inducible cell *XIST* constructs had intermediate levels of enrichment for H4K20me1 (1p = 13%, 7p = 19%, n = 100 for both).

Even when uninduced a low level of *XIST* RNA was detectable in HT1080 cells and could be visualized as a tiny 'pin-point' by RNA FISH, despite the constant repression mediated by the TetR system (Figure 3.1 D)[112]. The 1p, 4q, 7q and 8p uninduced (no dox) autosomal *XIST* integrations were analyzed by IF-FISH to determine whether the low level of *XIST* prior to dox induction was sufficient to facilitate enrichment of H3K27me3 in any of these cell lines. All the

*XIST* integrations had signs of potential H3K27me3 enrichment in 4-5% of cells at the *XIST* RNA point of fluorescence (n > 50, figure 3.1 C). H3K27me3 has a punctate appearance in the nucleus and the *XIST* RNA of uninduced cells appeared as a single point source, it was possible that the consistent 4-5% enrichment was a result of random overlap of signals, rather than a reflection of true enrichment of H3K27me3 given its invariance between the uninduced integration sites not reflecting the variance observed following induction.





IF-FISH was performed for H3K27me3/H4K20me1 and induced (5ddox) *XIST* RNA in HT1080 cells referred to by the location of the inducible *XIST* integration in the genome. The proportion of cells with visible and clear enrichment of the heterochromatin marks are shown in red, those without in grey. **A)** Proportion of cells expressing *XIST* with H3K27me3 enrichment at the *XIST* 

RNA cloud. **B)** Proportion of cells expressing *XIST* with H4K20me1 enrichment at the *XIST* RNA cloud. **C)** Comparison of H3K27me3 enrichment before and after *XIST* induction for five days across four integration sites. *XIST* RNA before induction was visible as a pin-prick by FISH which allowed assessment of enrichments. **A-C)** For all autosomal 5ddox treated cell lines >100 cells were analyzed while the Xq integration and no dox treated conditions between 50-55 cells were analyzed. **D)** Example of XIST RNA labelled with green fluorescent probes in the nuclei of an HT1080 cell (DAPI labelled) prior to the induction of XIST with doxycycline and following 5 days of doxycycline treatment.

#### 3.2.2 Clonal variability of H3K27me3 and H4K20me1

The location of the *XIST* construct in the genome played a large role in its ability to function, but it was necessary to determine the extent that differences resulting from random variability between clones. Independently generated pairs of clonal cell lines of *XIST* integration sites were tested by IF-FISH for their proportion of enrichment for H3K27me3 and H4K20me1. A subset of integration sites with multiple clones were selected based on the results of the previous section. Independently generated clones from the 7q, 8p and 12q integrations were compared for their ability to recruit H3K27me3 (figure 3.2 A) as they had the highest H3K27me3 enrichment proportion. The two 8p clones (2-3-0.5a 1c.1 and 1a.1) showed the strongest enrichment of H3K27me3 despite varying in their proportions [32% (n = 255) and 56% (n = 50), figure 3.2 A]. The 7q and 12q integrated constructs were consistent in their enrichment for H3K27me3 between clones and noticeably lower than both 8p clones (14-22% enriched, n ≥ 50, figure 3.2 A).

The enrichment proportion of H4K20me1 was more consistent between the clones of the integrations tested (7q, 8p, Xq). 8p and Xq were selected as examples of strong enrichment of H4K20me1 while 7q was selected for the inability of its *XIST* to enrich H4K20me1 ( $n \ge 50$ , figure 3.2 B). Enrichment proportions of H3K27me3 and H4K20me1 between pairs of clones was compared using the Fisher exact test to determine if the activity of *XIST* differed significantly between clones. Only the 8p cell lines enrichment of H3K27me3 differed significantly ( $p = 1.9 \times 10^{-3}$ ). Taken together these results support the conclusion that the location of the *XIST* RNA construct in the genome played a dominant role in the variability observed in *XIST*-mediated enrichment of H3K27me3 and H4K20me1.



Figure 3.2 Clonal variability minimally affected *XIST*-mediated recruitment of H3K27me3 and H4K20me1.

IF-FISH was performed for H3K27me3/H4K20me1 and induced (5ddox) *XIST* RNA in HT1080 cells lines with pairs of independent clonal populations generated with *XIST* integrated into the same locations. The proportion of cells with visible and clear enrichment of the heterochromatin marks are shown in red, those without in grey. Clonal data for 8p #1c1, 7q #1, 12q #14 and Xq #N1 is repeated from that shown in Figure 3.1. **A)** Proportion of cells from each cell line expressing *XIST* with visible H3K27me3 enrichment. **B)** Proportion of cells of each cell line expressing *XIST* with visible H4K20me1 enrichment. Significance was calculated using a Fisher exact test to compare the likelihood that the proportion of enriched and not enriched cells were conserved between clones of a given integration site (\* p < 0.05, NS > 0.05 only the 8p clones reached this threshold of significance when comparing enrichment of H3K27me3 ( $p = 1.9x10^{-3}$ )).

#### 3.2.3 The extent of XIST functional capacity when expressed from 8p

Based on the combined research of the *XIST* inducible integration lines, it was concluded that the 8p XIST construct was the model system that allowed for the widest array of features to be analyzed. The results of the 2015 publication were summarized in the summary figure 3.3. It was observed that XIST expressed from 8p enriched its XIST RNA cloud with H3K27me3, ubH2A, H4K20me1, MacroH2A and SMCHD1 while also silencing distal genes and relocating to the perinucleolar region of the nucleus. The precise enrichment proportions of these five Xiassociated heterochromatin marks were measured in 8p cells either prior to XIST induction or following 5 days of induction as well as in control female somatic cell lines (figure 3.4). This was done to ensure that there were no signs of enrichment prior to induction for any of these 5 Xi associated factors, and to compare the activity of the transgenic XIST to endogenously expressed *XIST.* These 5 Xi factors were clearly enriched between 26-36% of 8p induced cells examined. The female control cells had a much greater proportion of cells with visible enrichment ranging from 76-91%. The uninduced 8p cells could be classified as enriched in less than 5% of their populations. The difference in proportions of enrichment between the uninduced and induced 8p cells using a Fisher exact test was strongly significant in all instances ( $p < 1.0x10^{-4}$ , figure 3.4). The >2-fold proportion of cells enriched for every mark between the female and induced 8p indicated that while this model would allow for testing of XIST activity, the magnitude of its affect was clearly reduced outside of its natural context.

To determine which other aspects of XCI could be investigated using this model the Xi associated heterochromatin mark, H3K9me3 was also tested for enrichment. While the female control showed a strong enrichment proportion of 79%, only 3% of 8p *XIST* induced cells

showed any signs of enrichment. Numerous other Xi associated factors including HP1 $\gamma$ , HP1 $\beta$ , PTBP1, ASH2L, HNRNPK and HNRNPU were tested by IF-FISH, however little or no enrichment was observed in the female control or 8p cells, which prevented further analysis. CIZ1 was unique as it was universally enriched at every single *XIST* RNA cloud in both the induced 8p *XIST* cell line and the female cell line HTERT-RPE1, but was not enriched at all at the uninduced *XIST* pin-points in the 8p model (figure 3.5 A). CIZ1 shRNA knockdown had been performed in the 8p cells by Sarah Baldry, who reported a reduction of CIZ1 to one third of normal levels. The knockdown did not noticeably affect CIZ1 enrichment at the *XIST* RNA cloud and no other abnormal features or signs of the *XIST* RNA failing to localize were observed (figure 3.5 A).

One of the characteristic features of the Xi is its compaction, visible by microscopy as the Barr body [2]. It was therefore important to test whether the enrichment of fluorescence observed by IF-FISH was a result of the recruitment by *XIST* and not an indirect result of a greater overall chromatin density. IF-FISH was performed for *XIST* and histone H3 in a female somatic cell line (IMR-90) and the induced 8p (figure 3.5 B). In the female control cell line 41% of cells showed a visible enrichment of H3 at the *XIST* RNA cloud, however not a single induced 8p *XIST* cloud was enriched for histone H3. This indicated that the enrichment of heterochromatin marks in the induced 8p *XIST* model system was likely a reflection of a real increase in the levels of these marks rather than being an artefact of increased chromatin density. It also indicated that the proportion of female cells classified as enriched may be artificially inflated as a result of increased chromatin density and that *XIST* induction from 8p lacked some critical element necessary for chromatin compaction. These tests demonstrated that *XIST* induced from the 8p

model could be used to probe the mechanisms employed by *XIST* to establish the enrichment of H3K27me3, ubH2A, H4K20me1, MacroH2A, SMCHD1 and CIZ1 along the chromatin coated in its RNA.



Figure 3.3 Summary of published results of XIST integrations into HT1080 cells

This figure was published in 2015 as a summary of the results of all the authors who were analyzing the 9 induced *XIST* integration constructs for their activity. Grey shading indicated the G-banding pattern, ranging from dark to light. Increasing levels of *XIST* induced transcript from each integration site were denoted by the increasing intensity of green shading. The relative change in *XIST* perinucleolar localization following induction was represented by the intensity of blue shading. The proportion of enrichment for each heterochromatin mark was denoted by either solid (>25%), dotted (10-25%) or blank (<10%) pink shading. All constructs showed visible enrichment proportion of ubH2A >25% and so it was not listed. The strength and distance of *XIST*-mediated silencing of genes was denoted by the intensity of red shading for silencing.



The 8p *XIST* construct was found to overall be the best autosomal integration across all categories.

Figure 3.4 Qualitative assessment of 8p inducible XIST chromatin reorganization

IF-FISH was performed in uninduced and induced HT1080 cell lines with the inducible *XIST* construct integrated into chromosome 8p to compare the change in the enrichment of various Xi associated chromatin features. A female somatic cell line (IMR-90) was also included as a positive control. The strength of enrichment of each of these marks was measured by the proportion of cell with clear enrichment of a given chromatin feature at the *XIST* RNA signal relative to the nuclear background. A value of 1 indicated universal enrichment of a mark at all

*XIST* RNA clouds, while 0 indicated no cell had showed any signs of enrichment. The statistical significance of the difference before and after induction in the 8p cells was calculated using the Fisher exact test (\*\*\* p < 0.0001). All analysis here was performed by the author (TDMcD).



Figure 3.5 Induced XIST from 8p enriched for CIZ1 without increased chromatin density

IF was performed to label the protein of interest in red while FISH was performed using green labelled probes specific to *XIST* RNA and the nuclei were stained with DAPI (blue). The proportion of cells with enrichment of a given factor at the *XIST* RNA cloud is listed and scale

bars in white correspond to  $10\mu$ m. A) CIZ1 was labelled by IF in regular induced 8p cells as well as 8p cells knocked down for CIZ1 using shRNA. B) Histone H3 was labelled by IF in induced 8p cells and female control cells, with enrichment at the *XIST* RNA cloud indicative of increased chromatin density.

#### 3.2.4 Establishing the dynamics of *XIST* activity in the 8p *XIST* integration cell line.

So far all of the analysis of *XIST* activity had occurred following five days of induction, which had become standard practice in the Brown lab. However, this only provided a single snapshot of the activity of *XIST*. Testing further time points of *XIST* were anticipated to offer novel insights into the dynamics and order of *XIST* activity and offer a novel insight into the order that events occur during XCI.

## 3.2.4.1 Expression of XIST transgene RNA rapidly reached an upper plateau

The relative levels of *XIST* transcripts in the 8p induced HT1080 cell line (2-3-0.5a) was tested over the course of twelve days using qPCR to determine the rate at which *XIST* accumulates within the nucleus, as well as if and when the transcript levels plateau. Knowing the relative levels of *XIST* RNA over time would provide context for the dynamics of its other effects. As all previous work had been performed at five days of induction, the relative levels of *XIST* across the various other time points were normalized to 5ddox treatment. In addition to the uninduced control (0 days) and the reference 5 days, four biological replicates of four additional time points of *XIST* induction were tested: 2 days, 7 days, 10 days and 12 days. RNA was harvested, reverse transcribed and analyzed by qPCR as described in methods section 2.2-2.3. Relative transcript levels (RQ) of the induced *XIST* transcript were normalized to the endogenous control gene *PGK1* in each biological replicate (primers listed in primer table in appendix).

The relative levels of *XIST* RNA transcripts in each time point relative to the 5ddox treatment were shown in figure 3.6. Prior to induction there was 0.053 (RQ) the level of *XIST* transcripts

present compared to the 5ddox treatment, which represented a clear and statistically significant difference (0ddox sd = 0.016, p = 0.009). By two days of induction (2ddox) the average RQ of *XIST* was 0.68 and did not significantly differ from the 5ddox condition (sd = 0.18, p = 0.24). Neither the 7ddox or 10ddox treatments differed significantly from *XIST* (p > 0.05) despite two of the four replicates for 10ddox showing unusually high levels of *XIST* (RQ ~ 4). By 12 days of dox induction all 4 biological replicates had actually reduced by a statistically significant amount relative to the 5ddox levels (average RQ = 0.31, sd = 0.018, p = 0.033). These tests revealed that by 2 days of induction *XIST* transcript levels had been upregulated to levels nearly equivalent to 5ddox levels. Despite fluctuations in the level of *XIST* between replicates, from 2ddox onwards *XIST* levels do not seem to accumulate further over time to any significant degree (Pearson Correlation Coefficient r = 0.120, significance of correlation p = 0.62). The semi-rhythmic nature of the average levels of *XIST* across these induction time points as well as the variability observed indicated that additional factors likely affect either transcription efficiency or transcript stability, though further work would be needed to identify the cause.



Days of XIST induction with dox

## Figure 3.6 XIST transcript levels rapidly reach an upper level then plateau

*XIST* transcript levels were measured using RT-qPCR and determining the relative transcript levels compared to the endogenous control gene *PGK1*. Four biological replicates for each time point of *XIST* induction in the 8p HT1080 cell line were tested and the relative levels of *XIST* were normalized to the average level of the 5 day time point, as it was the time point which had become standard in previous examinations of the model system. The mean and standard deviation for each condition are indicated by lines, and individual dots representing the relative expression of each replicate. Increasing darkness of shading was used to indicate increasing length of *XIST* induction. The statistical significance of a difference between the 5 day treatment and other time points was calculated by two-tailed unpaired t-test (\* p < 0.05, \*\* p < 0.01).

#### 3.2.4.2 Dynamics of XIST-mediated distal gene silencing

To determine the dynamics of XIST-mediated silencing of distal genes in the 8p model the relative allelic contribution of the silenced alleles was tested across all of the time points and replicates tested in the previous section (3.2.5.1). Pyrosequencing of 8p SNP coding sequences of the cDNA from CTSB, DLC1, SLC25A37 and STC1 were carried out as described in section 2.3, 2.4 and 2.6 (primers listed in primer table of appendix). The biological replicates were themselves averaged together for each gene and are shown in figure 3.7. Calculations of statistical significance between the 5ddox treatment and other time points was performed using 20 different t-tests and the p value cut-off of statistical significance was adjusted accordingly (\* p-threshold =  $0.05/20 = 2.5 \times 10^{-3}$ , table 3.1). The allelic contribution of the XIST inhibited SNPs for CTSB, SLC25A37 and STC1 were all ~50% prior to induction (0 days, figure 3.7). DLC1 had partially skewed allelic contribution with the XIST repressed allele only contributing 37% of total expression in XIST uninduced cells. By 2 days of induction obvious inhibition of all four alleles (figure 3.7) was observed indicating that in that time XIST had affected these genes despite the intervening megabases of chromatin. CTSB, DLC1 and SLC25A37 did not statistically differ between their levels of silencing at 2 days of induction compared to 5 days, indicating that minimal further silencing had occurred during the three intervening days. STC1 showed a statistically significant decrease in allelic contribution between days 2 and 5 of induction, however as this represented a drop from 10.6% to 6.7% allelic contribution most of silencing had clearly occurred within 2 days of induction. None of the time points differed significantly in their allelic contribution from 5 days onwards. The results suggested that even in the foreign context of Chromosome 8, XIST rapidly repressed distal genes within the first two days, and then

maintains that level of repression. From a practical stand point, this meant that within 2 days of induction *XIST* has almost entirely utilized its silencing machinery to a maximal level in the 8p model.



Days of XIST induction with dox

Figure 3.7 XIST induced silencing of distal genes occurs primarily within 2 days

The contribution of *XIST* repressed alleles for the 8p linked genes *CTSB*, *DLC1*, *STC1* and *SLC25A37*, all carrying coding SNPs, relative to the total genes expression as measured by pyrosequencing. The location of the genes as well as the G-banding pattern on 8p is shown in the diagram beside the graph. Four biological replicates for each period after dox induction were tested, and the average allelic contribution as well as standard deviation is listed. The statistical significance of the difference in allelic contribution between the 5 day and other treatments was calculated using an unpaired t-test with the colour of the asterisks denoting which sample was significant (\* p <  $2.5 \times 10^{-3}$ ). The black asterisk denotes that all 4 alleles statistically differed in their contribution to the 5 day induction levels. The green asterisk denotes that only the allelic contribution of *STC1* differed from the 5 day induction level. The details are listed in table 3.1.

| Gene     |           | No induction | 2 days dox | 5 days dox | 7 days dox | 10 days dox | 12 days dox |
|----------|-----------|--------------|------------|------------|------------|-------------|-------------|
|          | Allelic % | 51.15        | 21.05      | 13.525     | 14.35      | 10.5375     | 12.55       |
| CTSB     | sd        | 0.18         | 3.00       | 3.33       | 0.81       | 0.90        | 0.62        |
|          | P-value   | 1.48 E-05    | 2.74 E-02  |            | 7.02 E-01  | 2.19 E-01   | 6.50 E-01   |
|          | Allelic % | 37.35        | 7.98       | 6.12       | 9.38       | 5.39        | 7.14        |
| DLC1     | sd        | 0.35         | 2.17       | 1.36       | 2.04       | 4.09        | 1.23        |
|          | P-value   | 5.02 E-07    | 2.66 E-01  |            | 6.79 E-02  | 7.83 E-01   | 3.78 E-01   |
|          | Allelic % | 50.30        | 17.00      | 11.15      | 19.00      | 11.22       | 19.72       |
| SLC25A37 | sd        | 0.44         | 1.45       | 0.98       | 1.62       | 3.18        | 0.61        |
|          | P-value   | 2.05 E-06    | 1.84 E-02  |            | 6.16 E-03  | 6.82 E-01   | 7.58 E-03   |
|          | Allelic % | 49.58        | 10.58      | 6.66       | 7.49       | 3.61        | 5.28        |
| STC1     | sd        | 0.84         | 0.81       | 0.73       | 0.74       | 0.14        | 0.46        |
|          | P-value   | 2.24 E-08    | 8.16E-04   |            | 2.17E-01   | 4.48E-03    | 3.82E-02    |

 Table 3.1 Summary of XIST induced silencing dynamics

List of allelic contribution of the silenced allele of *CTSB*, *DLC1*, *STC1* and *SLC25A37* from four biological replicates without induction and then following various periods of *XIST* induction from 8p. The table lists for each time point and gene the average allelic contribution to total gene expression (Allelic %) as well as the standard deviation (sd) and statistical significance of the difference from the 5 day time point calculated using an unpaired t-test.

#### 3.2.4.3 Dynamics of XIST-mediated chromatin reorganization lag behind silencing

The next objective was to determine how the dynamics of XIST-mediated chromatin remodeling related to the rate at which XIST accumulated and initiated gene silencing to obtain a preliminary order to the chromatin inactivation pathways. The previously described method of qualitatively judging whether a cell was enriched for a given chromatin mark at the XIST RNA cloud suffered from operator bias and did not measure the magnitude of enrichment within an individual cell. For this reason, the method of quantifying the distribution of IF labelled factors at the XIST RNA cloud as described in methods section 2.13.2 was developed, and the identity and condition of each group of cells was blinded to minimize potential bias. This method of quantifying IF-FISH images required far more time to analyze, and as a result the number of time-points analyzed was limited to 3 days (3ddox), 5 days (5ddox) and 10 days (10ddox) of induction. Across these time points the early heterochromatin marks H3K27me3 and ubH2A were examined to provide both insight into the rate at which chromatin remodeling occurred as they are known to be some of the earliest heterochromatin marks established by XIST [127]–[133]. In the later time points (5ddox and 10ddox) the late heterochromatin mark SMCHD1 was tested to determine whether its rate of accumulation was obviously lower than that of the PRC established marks. The as yet unexamined depletion of histone acetylation, believed to be one of the earliest XIST-mediated chromatin modifications, was tested across the earlier 3ddox and 5ddox treatments. Both MacroH2A and H4K20me1 were tested at the 5ddox time point to obtain a benchmark of the enrichment of these marks. The results of this analysis were shown in figure 3.8 and table 3.2.

The heterochromatin mark ubH2A was still accumulating after 3 days of *XIST* induction on 8p as it increased significantly from day 3 (median z-score = 0.934, sd = 1.09) to day 5 (median z-

score = 4.18, sd = 3.17, p =  $1.1 \times 10^{-14}$ ). By day 10 of induction the relative enrichment of ubH2A had decreased (median z-score = 1.81, sd = 1.53, p =  $4.5 \times 10^{-5}$ ), though was still significantly higher than it had been at 3 days of induction (p =  $2.4 \times 10^{-7}$ ).

Enrichment of H3K27me3 was still accumulating following 3 days of *XIST* induction (median z-score = 1.22, sd 1.55) as its average concentration at the *XIST* RNA cloud had risen significantly by day 5 (median z-score = 2.59, sd = 1.70, p =  $6.7 \times 10^{-6}$ ). Between testing on day 5 and day 10 of induction there was effectively no change in the levels of H3K27me3 enrichment (median z-score = 2.68, sd 1.84, p = 0.84). This suggested that by day 5 H3K27me3 has reached an upper plateau of its enrichment by *XIST* by day 5 of induction.

In addition to observing the accumulation of these two early heterochromatin marks, a depletion of histone acetylation H3K27ac over time was observed at the *XIST* RNA cloud. At day 3 of induction there were signs of clear depletion of H3K27ac (median z-score = -1.00, sd = 0.883). The deacetylation of H3K27ac at the *XIST* RNA cloud continued for the subsequent 2 days, with strong depletion by 5 days of *XIST* induction (median z-score = -3.14, sd 1.35, p =  $6.5 \times 10^{-16}$ ).

When the enrichment of the late recruited factor, SMCHD1 was examined during the two later time points between day 5 and day 10 of induction (figure 3.8) no statistically significant change was observed. As had been observed with ubH2A, the day 10 treatment was slightly less enriched on average than the 5 day dox treatment though both time points displayed strong enrichment (5ddox median z-score = 2.70, 10ddox median z-score = 2.27, p = 0.053). This suggested that despite SMCHD1 being associated with late stage XCI, in the 8p model it likely reached its upper limit of recruitment at approximately the same time as the early heterochromatin marks by day 5 of induction. Examination of the heterochromatin marks H4K20me1 and MacroH2A was performed to establish a baseline for their level of enrichment in the 8p inducible model system following 5 days of induction (figure 3.8, table 3.2). The median z-score for both of these marks were around 2 (H4K20me1 = 2.48, MacroH2A =1.96). These results offered new insights into the ability of *XIST* to reorganize the chromatin environment, suggesting that 5 days of *XIST* expression were necessary and sufficient for *XIST* to initiate both early and late chromatin modifications.



# Figure 3.8 *XIST*-mediated chromatin remodeling reaches its upper limit following 5 days of induction

The quantified strength of various Xi associated chromatin features as well as H3K27ac at the *XIST* RNA cloud relative to the nuclear level as measured from IF-FISH analysis and presented as z-scores for each cell (white dots). The chromatin marks H3K27ac, H3K27me3, ubH2A thought to be affected early during XCI were also tested at 3 days of *XIST* induction. H3K27me3, ubH2A and SMCHD1 were also tested following 10 days of *XIST* induction. The chromatin features shown in figure 3.5 as being enriched in a proportion of cells were all analyzed at the 5 day of induction time point. The z-score value for each cell indicates the difference between the fluorescent intensity of a factor at the site of *XIST* RNA relative to the nuclear background (excluding the nucleolar regions) as a metric of the variability (standard

deviation) in fluorescent intensity observed in general. A z-score of 0 indicates that there was no difference in intensity of a factor at the *XIST* RNA compared to the nuclear background. A z-score of 1 indicates that the average intensity at the *XIST* RNA cloud was 1 standard deviation above the nuclear average. The colour of each box plot denotes the factor being tested in each instance. The central line of the boxplot indicates the median z-score for each condition, the notch the confidence interval, the box itself the interquartile range and the whiskers 1.5x the interquartile range. The difference in the population distributions of each time period and for each mark are listed in table 3.2

| Factor   | Induction | Mean     | Median   | sd       | number | Significance (p-values) to: |          |
|----------|-----------|----------|----------|----------|--------|-----------------------------|----------|
|          | period    | z-score  | z-score  |          |        | 3 days                      | 5 days   |
| UbH2A    | 3 days    | 0.96725  | 0.933832 | 1.094115 | 60     |                             |          |
|          | 5 days    | 4.744544 | 4.177745 | 3.179062 | 60     | 1.09E-14                    |          |
|          | 10 days   | 2.095243 | 1.808634 | 1.533996 | 60     | 2.44E-07                    | 4.49E-05 |
|          | 3 days    | 1.481791 | 1.220821 | 1.546365 | 60     |                             |          |
| H3K27me3 | 5 days    | 2.901063 | 2.592653 | 1.704249 | 60     | 6.66E-06                    |          |
|          | 10 days   | 2.933832 | 2.678722 | 1.846823 | 60     | 5.44E-06                    | 0.84     |
| H3K27ac  | 3 days    | -0.91657 | -1.00448 | 0.882745 | 60     |                             |          |
| 113K27ac | 5 days    | -3.06645 | -3.13567 | 1.34714  | 60     | 6.46E-16                    |          |
| SMCHD1   | 5 days    | 3.137372 | 2.703656 | 2.166534 | 60     |                             |          |
|          | 10 days   | 2.331552 | 2.27026  | 1.569032 | 60     |                             | 5.34E-02 |
| H4K20me1 | 5 days    | 2.432495 | 2.488635 | 1.593831 | 61     |                             |          |
| MacroH2A | 5 days    | 2.249173 | 1.958056 | 1.469285 | 60     |                             |          |

 Table 3.2 Dynamics of Xi associated chromatin featured recruited by 8p inducible XIST

The summary list of the data shown in figure 3.8 showing the relative enrichment of Xi associated chromatin features following different periods of XIST induction. The factor labelled by IF is listed in the first column, followed by the period of XIST induction that occurred before cells were analyzed. The average (mean and median) z-score for each condition is listed as well as the standard deviation between cells and the number of cells tested. Comparison of the distribution of z-scores between time points was performed using the Mann-Whitney U test and the p-values comparing various time points are listed

#### 3.3 Discussion

The analysis of the autosomal XIST cDNA constructs in the HT1080 cell lines revealed the genomic location of XIST to have a profound effect on its capacity to function. XIST induction from the X chromosome, though outside of its normal locus, consistently produced a stronger enrichment of heterochromatin marks than was observed on the autosomes. This supported the idea that the X chromosome is generally tailored by selection for efficient chromosome inactivation by XIST [66], [194]. The autosomal integrations varied greatly from one another, both in terms of the processes initiated and the strength of the effects observed. Expression of XIST from the 8p integration site resulted in the strongest overall enrichment of H3K27me3 and H4K20me1, and was additionally observed to produce some of the strongest enrichment of ubH2A, MacroH2A and SMCHD1 [268]. The 4q XIST construct also strongly enriched its domain with H3K27me3, though it showed no signs of H4K20me1 enrichment. When XIST was expressed from the 1p integration site it neither recruited H3K27me3 nor H4K20me1. While the 8p XIST clonal cell lines varied somewhat in their level of H3K27me3 enrichment, overall, the clones tested for both H4K20me1 and H3K27me3 enrichment showed a similar hierarchy of enrichment between integration sites. This indicated that the integration locations themselves were responsible for this variability in XIST activity. A surprising observation was that none of the functions tested here demonstrated a clear hierarchy or obvious association with any other XIST function across the integration sites. The variability between the integration sites suggested that rather than linear pathways being initiated by XIST, the most likely mechanism involved various complex pathways with potential unique elements specific to each function.

To further examine how *XIST* functioned the 8p inducible HT1080 model system clearly offered the greatest potential for analyzing the most aspects of *XIST* function. *XIST* induced from 8p initiated the rapid silencing of genes megabases from its integration site within two days and would subsequently enrich the site of *XIST* RNA with the heterochromatin histone marks H3K27me3, ubH2A, H4K20me1 and MacroH2A as well as the Xi associated protein SMCHD1. The *XIST* induced from 8p did not enrich its *XIST* RNA cloud with H3K9me3, so this mark would not be testable in subsequent analysis using this model. The Xi associated protein CIZ1, predicted to be involved in *XIST* localization [107], was universally enriched in all cells and remained disproportionately enriched at the *XIST* RNA cloud in the 8p construct even after being knocked down, suggesting this protein has a high affinity for *XIST* RNA.

Comparing the strength of the chromatin modifications across numerous time points also offered a validation of the utility of using z-scores across numerous cells to assess the strength of *XIST* modifying the local chromatin domain. By five days of induction the average z-score across a population was found to be nearly 2 standard deviations or greater in magnitude. The quantification of the enrichment of the 5 Xi associated factors as well as H3K27ac depletion indicated that by five days of dox induction *XIST* had strongly influenced the chromatin environment in most cells. This method offered a greater degree of precision than had been previously possible when performing qualitative assessment of enrichment by eye. This method also offered a more effective means of comparing enrichment across populations of cells than would have previously been possible. For these reasons, this qualitative method of analyzing IF-FISH images was adopted for all subsequent analysis of the functional capacity of *XIST*, despite the massive increase in time required. The one exception would be analysis of CIZ1, where the

uniquely strong enrichment profiles observed in all cells demonstrated that enrichment scoring would still be effective.

Analyzing the dynamics of XIST activity when expressed from 8p revealed that upregulation of *XIST* and silencing both occur rapidly within roughly 2 days of induction (Figure 3.6 and 3.7). The remodeling of the chromatin occurs much more slowly, however, with histone hypoacetylation, H3K27me3 and ubH2A still accumulating after 3 days of induction and not reaching their upper extent until five days. This seeming primacy of silencing over chromatin remodeling was difficult to reconcile with the current literature based on the activity of mouse *Xist.* There is evidence that silencing by *XIST* is not entirely dependent upon PRC activity [128], [142] though there is evidence that the PRC complexes may contribute to gene silencing [114], [129]. Other research groups working in mouse models have proposed that histone deacetylation is a crucial step for *Xist*-mediated gene repression [114]. The results presented here suggest that histone deacetylation of the XIST RNA territory occurred after XIST had already repressed genes distally, deacetylation even between days 3 and 5 of XIST induction different significantly while silencing of most alleles at day 2 and day 5 of XIST induction were comparable. It was not possible based on these findings, however, to rule out that deacetylation of these four genes may have occurred in advance of the broader histone deacetylation measured by IF-FISH. Studies of mouse Xist indicated that it first spread to chromatin proximal in three-dimensional space to its site of transcription, which could allow for silencing and heterochromatin recruitment to occur earlier in some regions than others [269]. Studies in mice have suggested that these regions that Xist first associated with were not actively transcribed, and that once Xist bound to these loci it then spread along the chromatin to inactivate gene expression [95], [96]. The most appealing

interpretation of the results presented here was that *XIST*-mediated silencing precedes these chromatin modifications. With this 8p inducible model for analyzing *XIST* identified and characterized, the next goal was to begin to dissect *XIST* activity and identify the regions responsible for the numerous complex pathways it initiated.

# Chapter 4: Development and validation of XIST deletion constructs

#### 4.1 Introduction

The next step to understanding how *XIST* functions was to create a system that allowed individual functions and regions of the gene to be examined. The generally accepted model of *XIST* is as a large molecular RNA scaffold that binds and coordinates interactions among a wide range of factors and complexes to facilitate the rapid and stable inactivation of a chromosome [270], [271]. *XIST* activity as a scaffold for numerous factors has been suggested to explain the consistency of its large size across species, despite the sequence itself being poorly conserved across species [80], [81]. Various pull down experiments identified potentially over 200 mouse *Xist* RNA binding factors, with the vast majority of identified factors being of uncertain function or significance [59], [103], [157].

The size of the *XIST* transcript and its prolific protein binding present a significant barrier to understanding the mechanisms by which it functions. The best way to examine how *XIST* functions was therefore to develop a system of *XIST* deletion constructs to circumvent the large size of *XIST*. Deletion constructs were used in mice to study the function of *Xist*, with one of the first studies measuring cell survival with different regions of *Xist* deleted to determine silencing efficiency [82]. Since then studies of *Xist* deletions have focused on deleting specific regions of interest and examining typically a single pathway per study which has provided the majority of the understanding of how *Xist* functions as described throughout the introduction. A series of human deletion constructs that spanned the entire length of *XIST* would allow for both the function of specific regions to be determined as well as potential links between process related to

XCI to be inferred. This raised the question of how best to subdivide *XIST* into numerous examinable regions. The seven repeat regions of *XIST*, Repeat A-F, were hypothesized to be the most likely to be functionally significant, so emphasis was placed on designing the deletion constructs to best examine the individual repeats [81], [82]. Repeat A was previously demonstrated to be important for *XIST*-mediated gene silencing both in human and mice [82], [111], [112]. While the mouse repeats B-F have also been tested to some degree, few human studies have examined the role of these regions in the establishment of XCI, though a recent study did attempt to identify the regions of *XIST* needed for maintaining gene silencing on the Xi [112], [210]. In addition to identifying essential regions of human *XIST*, identifying which regions were dispensable for *XIST* activity was of great interest to this research. The utility of *XIST* for therapeutic purposes as a regulator of aberrant karyotype-based diseases is undermined by its size, which represents a significant obstacle to efficient delivery and integration [211].

To identify the function of the regions of *XIST*, partial *XIST* constructs were compared directly with a full length *XIST* construct to determine what effect the absence of a region had on the ability of *XIST* to function in a common context. In chapter 3 the 8p Full *XIST* construct integrated in 8p of the HT1080 cell line (2-3-0.5a) was found to be the most effective autosomal model for examining *XIST* activity, and so the 8p integration site was selected for all subsequent analyses. In addition to the Full XIST construct, several inducible partial *XIST* cDNA construct containing cell lines had also been generated by members of the Brown lab and were referred to as the integration Delta ( $\Delta$ ) constructs [112]. An outline of these integrated partial *XIST* constructs and their sequences relative to *XIST* are shown in figure 4.1. The construct with the smallest absent region is  $\Delta$  XB, that lacks a 796bp region of *XIST* encompassing Repeat A of
*XIST*. The construct with the next smallest absent region was initially labelled  $\Delta$  3' in its original publication but was subsequently renamed Exon 1 and lacks the 3.6 kb of exons 2-8 in the short isoform of *XIST* [65], [112]. The construct  $\Delta$  PflMI lacks 3.8kb from the middle of the first exon of *XIST*, resulting in the absence of Repeat B and Repeat C as well as the non-repetitive region of *XIST* extending to the 5' edge of Repeat D [112]. The partial construct with two sequences missing was an Exon 1 construct that lacked the PflMI internal region and was officially labelled  $\Delta$  PflMI  $\Delta$  NC, but became consistently referred to as  $\Delta\Delta$  (figure 4.1). These  $\Delta$  constructs provided an initial framework for assessing the function of different regions of *XIST*.

The regions absent in these initial  $\Delta$  constructs are not overlapping and do not fully span *XIST*, and so to examine *XIST* entirely a greater array of deletion constructs was needed. At the start of this project the potential of the CRISPR system for enacting specific genome modifications was first being realized and offered an exciting approach to generate the deletion cell lines that were needed to investigate the different regions of *XIST* [225], [226], [233]. Using CRISPR to generate the deletion cell lines directly from the Full *XIST* cell line ensured a more recent ancestry and minimized epigenetic or genetic drift. The new CRISPR deletion constructs were designed to examine the function of repeat and non-repeat domains independently where possible and to further subdivide the regions absent in the previously generated deletion constructs.

To ensure that every single nucleotide in *XIST* was deleted in at least one construct and allow for greater resolution, adjacent deletions were designed with short shared sequences deleted. If two adjacent deletions both disrupted a specific pathway it would suggest that the overlap was potentially of greater significance. Discrepancies between overlapping deletions would similarly

indicate the likely dispensable nature of the overlapping sequence and further narrow the region of interest.

The deletion " $\Delta$ " constructs created through this chapter had their utility as effective model systems validated by confirming that they could still be induced to produce *XIST* RNA transcripts. The deletion of regions of *XIST* had been shown to decrease transcript levels endogenously in humans and mice in certain instances so it was essential to determine what effect the different type of deletions had on the transcript levels of *XIST* [92], [272]. In addition, it was crucial to confirm whether *XIST* RNA induced in the various deletion constructs still localized to form an RNA cloud that could be observed through FISH and fluorescent microscopy and whether the shape or distribution of *XIST* RNA was noticeably perturbed compared to the Full *XIST* control. This indicated both whether a region of *XIST* was essential for its ability to localize into its typical RNA cloud and to determine whether a given deletion would be testable for chromatin remodeling activity in subsequent analysis. Induction of *XIST* for 5 days with doxycycline (5ddox) was shown to be ideal for examining the broadest array of *XIST* functions, and so these experiments on the deletion constructs were all performed using that time point.



# Figure 4.1 The structure of the inducible *XIST* constructs integrated into the 8p FRT site in the HT1080 cell line 2-3-0.5a.

An outline of the extent of the partial *XIST* constructs generated by integrating *XIST* cDNA sequences into the 8p FRT site. The cDNA of *XIST* is shown with repeat regions labelled A through F in blue and non-repeat regions labelled in grey. Dotted lines denote regions absent from each construct. The  $\Delta$  XB construct lacks Repeat A of *XIST* but contains the complete sequence 3' of the repeat.  $\Delta$  PflMI lacks a large 3.8kb region encompassing Repeats B and C as well as a ~3kb non-repeat sequence extending to just upstream of Repeat D. Exon 1 only consists of cDNA sequence from the first exon of *XIST*.  $\Delta\Delta$  consists of a PflMI deletion within the Exon 1 construct, resulting in a double deletion roughly halving the size of *XIST*. The expression of each construct is controlled by a doxycyline inducible CMV promoter.

## 4.2 Results

#### 4.2.1 The generation of XIST deletion constructs using CRISPR Cas9

The gRNAs targeting sequences within the inducible *XIST* cDNA were created as described in the methods section 2.7. The gRNAs selected to generate the new array of *XIST* deletion constructs were listed in table 4.1. The gRNAs 0.2 and 14.2 were directly adjacent to the 5' and 3' ends of the cDNA sequence and the only ones selected to cut their respective regions of *XIST* because of a limited number of potential target sequences in those regions. To confirm that the resulting targeted gRNA DH5 $\alpha$  clones carried the correct sequence, the gRNA plasmids were purified, digested with Ndel and run on an agarose gel as a preliminary check before being sent for sequencing by Integrated DNA Technologies (IDT). The final gRNA carrying DH5 $\alpha$  clones that were kept had perfectly integrated targeting sequence with no observable mutations.

When creating each type of deletion construct the *XIST* targeting gRNA plasmids and Cas9 expressing plasmid were purified as described in section 2.9 and transfected into the 8p HT1080 cell line (2-3-0.5a 1c.1) as described in section 2.10. The primers ultimately used to identify the successfully created deletion constructs after Puromycin selection are listed in the appendix table A.1. PCR amplification with Taq DNA Polymerase supplemented with 20% Betaine was the most effective at consistently being able to identify both cut and uncut PCR sequences while avoiding non-specific PCR products. Those clones with only a single correct size of PCR products were purified and sent for sequencing as described in section 2.4.1. In most instances where it was possible to PCR amplify the endogenous *XIST* along with the transgenic *XIST*, homogenous deletions in both resulted.

Figure 4.2 summarized the composition of deletion constructs resulting from CRISPR treatment. Table 4.2 lists deletion cell lines generated through the use of CRISPR, along with the gRNAs used to generate these cell lines and the size of the region deleted as determined from sequencing the PCR product spanning the cut sites. In addition, the sequencing results for each deletion cell line can be found in the appendix B. The following sub sections provide a written account of the creation of each type of deletion construct.



Figure 4.2 Map of the CRISPR generated XIST deletion constructs

An outline of the extent of the partial *XIST* constructs generated by CRISPR-mediated excision of specific regions of *XIST* in the inducible 8p H1080 cell line. The cDNA of *XIST* is shown with repeat regions labelled A through F in blue and non-repeat regions labelled in grey. Dotted lines denote regions absent from each construct. The extent of each individual deletion construct shown here is an average of the extent of the regions deleted in each of the individually generated cell lines.

| gRNA      | Nucleotides | <b>S-</b> | <b>E-</b> | <b>Target Sequence</b> with PAM | length |
|-----------|-------------|-----------|-----------|---------------------------------|--------|
| name      | into cDNA   | Score     | Score     |                                 |        |
|           | and strand  |           |           |                                 |        |
| XIST      | 122 +ve     | 100       | 64.9075   | <u>GGACGTGTCAAGAAGACACT</u>     | 20     |
| gRNA 0.2  |             |           |           | AGG                             |        |
| XIST      | 887 +ve     | 100       | 48.5215   | <u>GTTTGTGCTAAGTTAAACTA</u>     | 20     |
| gRNA 0.9  |             |           |           | GGG                             |        |
| XIST      | 900 +ve     | 100       | 65.2456   | <u>GTTAAACTAGGGAGGCAAGA</u>     | 20     |
| gRNA 1.0  |             |           |           | TGG                             |        |
| XIST      | 1655 +ve    | 92.1739   | 42.595    | <u>GCAGCTGTCTTTAGCCAGTC</u>     | 20     |
| gRNA 1.7  |             |           |           | AGG                             |        |
| XIST      | 1888 -ve    | 100       | 47.4114   | <u>GGGGAGGTATACTTAGCCTT</u>     | 20     |
| gRNA 1.9  |             |           |           | AGG                             |        |
| XIST      | 2087 +ve    | 94.3478   | 56.2607   | <u>GATGATCGTTGGCCAACAGG</u>     | 20     |
| gRNA 2.1  |             |           |           | TGG                             |        |
| XIST      | 2537 +ve    | 92.1739   | 54.9581   | <u>GAGTGTTTGAAGGTTTACAC</u>     | 20     |
| gRNA 2.6  |             |           |           | AGG                             |        |
| XIST      | 3036 +ve    | 100       | 55.1371   | <b>GGACAAAGAATTTCCTTACT</b>     | 20     |
| gRNA 3.1  |             |           |           | CGG                             |        |
| XIST      | 3311 -ve    | 94.5454   | 42.8973   | <u>GAGTGCTGTCTAATCCAAT</u>      | 20     |
| gRNA 3.3  |             |           |           | GGG                             |        |
| XIST      | 5459 +ve    | 100       | 67.3981   | <u>GCAGTAATGCAAATGGAGCA</u>     | 20     |
| gRNA 5.5  |             |           |           | AGG                             |        |
| XIST      | 5947 -ve    | 100       | 54        | <u>GGCCAAGAAATGGGGCCTT</u>      | 19     |
| gRNA 6.0  |             |           |           | AGG                             |        |
| XIST      | 8523 -ve    | 84.3478   | 58.575    | GCCAAGAAAAGGGGACTTAG            | 20     |
| gRNA 8.5  |             |           |           | GGG                             |        |
| XIST      | 8586 -ve    | 100       | 54.3169   | <u>GAGGTGGGGGCATCCTTGTCT</u>    | 20     |
| gRNA 8.6  |             |           |           | AGG                             |        |
| XIST      | 11865 +ve   | 100       | 53.7637   | <b>GCCTGGCACTCTAGCACTTG</b>     | 20     |
| gRNA 11.9 |             |           |           | AGG                             |        |
| XIST      | 12107 -ve   | 100       | 56.0168   | <b>GTGAAAGAAGAGCCACATCT</b>     | 20     |
| gRNA 12.2 |             |           |           | AGG                             |        |
| XIST      | 13626 +ve   | 100       | 53.4027   | <u>GTTGGGGAAAAAAAGTGCC</u>      | 20     |
| gRNA 13.7 |             |           |           | AGG                             |        |
| XIST      | 13804 +ve   | 100       | 66.3505   | GACCACTGCTGGGCAGCAGG            | 20     |
| gRNA 13.8 |             |           |           | AGG                             |        |
| XIST      | 14257 +ve   | 84.3478   | 65.3497   | <b>GTCACAATTGAAACAAACTG</b>     | 20     |
| gRNA 14.2 |             |           |           | GGG                             |        |

Table 4.1 List of gRNAs

The gRNAs target sequences used to generate the various deletion constructs are listed along with the position within XIST that they target. The specificity and efficiency scores predicted are included for each along with the actual target sequence of the gRNA.

| Deletion cell line  | 5' gRNA        | 3' gRNA        | Nucleotides deleted |
|---------------------|----------------|----------------|---------------------|
| Δ Α #3              | XIST gRNA 0.2  | XIST gRNA 0.9  | 779                 |
| Δ Α #12             | XIST gRNA 0.2  | XIST gRNA 0.9  | 777                 |
| Δ FBh #21           | XIST gRNA 0.9  | XIST gRNA 1.9  | 1127                |
| Δ FBh #22           | XIST gRNA 1.0  | XIST gRNA 1.9  | 811                 |
| Δ Bh #5             | XIST gRNA 1.7  | XIST gRNA 2.6  | 833                 |
| $\Delta$ Bh #7      | XIST gRNA 1.7  | XIST gRNA 2.6  | 833                 |
| Δ Bh #11            | XIST gRNA 1.7  | XIST gRNA 2.6  | 857                 |
| Δ BC #2             | XIST gRNA 2.1  | XIST gRNA 3.3  | 1195                |
| Δ BC #8             | XIST gRNA 2.1  | XIST gRNA 3.3  | 1189                |
| Δ BC #17            | XIST gRNA 2.1  | XIST gRNA 3.3  | 1195                |
| Δ 3'PflMI #3        | XIST gRNA 3.1  | XIST gRNA 6.0  | 2859                |
| $\Delta$ 3'PfIMI #6 | XIST gRNA 3.1  | XIST gRNA 6.0  | 2859                |
| Δ D #3              | XIST gRNA 5.5  | XIST gRNA 8.5  | 3084                |
| Δ D #10             | XIST gRNA 5.5  | XIST gRNA 8.5  | 3092                |
| Δ 3D5E #13          | XIST gRNA 8.5  | XIST gRNA 12.2 | 3584                |
| Δ 3D5E #14          | XIST gRNA 8.5  | XIST gRNA 12.2 | 3583                |
| Δ 3D5E #15          | XIST gRNA 8.5  | XIST gRNA 12.2 | 3588                |
| ΔΕ#6                | XIST gRNA 11.9 | XIST gRNA 13.7 | 1844                |
| Δ E #10             | XIST gRNA 11.9 | XIST gRNA 13.7 | 1778                |
| Δ 3' #1             | XIST gRNA 13.7 | XIST gRNA 14.2 | 630                 |
| Δ 3' #7             | XIST gRNA 13.7 | XIST gRNA 14.2 | 630                 |

Table 4.2 List of  $\Delta$  cell lines

The summary of each deletion in the various  $\Delta$  cell lines are shown here. The name of each deletion cell line is based on its deletion construct followed by a numerical identifier. The gRNAs used to generate each deletion cell line are shown in the central columns, and the ultimate size of the region deleted in each construct as determined by sequencing is listed in the final column.

## 4.2.1.1 Δ A

The  $\Delta$  A deletions were designed to lack the complete Repeat A sequence, representing a more precise method of analyzing Repeat A than was possible with the  $\Delta$  XB construct. This was done to ensure that any effects on *XIST* function identified using  $\Delta$  A construct could be confidently ascribed to Repeat A itself. The 5' gRNA, gRNA 0.2, was selected to ensure that most of the 5' non-repeat region of the Full *XIST* cDNA sequence would remain in the  $\Delta$  A construct (table 4.1). Two homozygous  $\Delta$  A colonies with deletions of both the endogenous and transgenic Repeat A were identified (#3 and #12) in addition to a heterozygous deletion colony with a deletion on endogenous *XIST*. This was one of the only instances were heterozygosity in the CRISPR-mediated removal of regions of *XIST* was observed. The  $\Delta$  A cell lines were from two separate transfection reactions, produced from conditions of 0.75µl and 1.5µl of Lipofectamine 3000 respectively. The deletions within these two colonies differed by 2 nucleotides close to the expected cut site and both of which removed the complete A repeat (table 4.2 and Appendix B).

#### 4.2.1.2 Δ FBh

The initial plan for a  $\Delta$  F construct was to remove the region spanning 3' of Repeat A to just 5' of the human specific Repeat Bh. Amplifying by PCR across this region was incredibly difficult even with the additive betaine and eight different sets of primer pairs were tested in various combinations until a suitable set of reaction conditions was found. Ultimately two deletion cell lines were produced using 1.5µl Lipofectamine 3000 reaction conditions, #21 using the 5' gRNA, gRNA 0.9, and #22 using gRNA 1.0. Both deletion cell lines were created using the 3'

gRNA, gRNA 1.9, which was expected to cut the 5' edge of Repeat Bh. When both of the deletion cell lines were sequenced, however, both had independently lost the complete Repeat Bh sequence exactly. This indicated that there is something unique to Repeat Bh, which predisposed it to either degradation or excision from the *XIST* sequence. The deletion constructs were named  $\Delta$  FBh to reflect the two repeats deleted in these constructs.  $\Delta$  FBh #21 had a 1127 bp deletion while  $\Delta$  FBh #22 had an 811bp deletion. Despite numerous attempts no successful deletion constructs could be produced by combinations of either 5' gRNA (0.9 or 1.0) with gRNA 1.7.

#### 4.2.1.3 Δ Bh

 $\Delta$  Bh deletions were designed to remove the human specific Repeat Bh and directly border the region deleted in the  $\Delta$  PfIMI construct, and proved to be one of the easiest to produce and validate, requiring minimal troubleshooting. Three  $\Delta$  Bh cell lines were created using the 5' gRNA, gRNA 1.7, and the 3' gRNA, gRNA 2.6.  $\Delta$  Bh #5 and #11 were both obtained from unique cells in the same transfection reaction using 1.5µl of Lipofectamine 3000, as #5 had a 833bp deletion and #11 had a 857bp deletion.  $\Delta$  Bh #7 was from a separate transfection using 0.75µl of Lipofectamine 3000 reagent, but had an identical deletion as Bh #5, indicating that the cleavage and repair process occurred with high fidelity. These  $\Delta$  cell lines were the easiest to generate, despite both gRNAs having below average efficiency scores (<56.0) and involving larger deletions than either  $\Delta$  A or  $\Delta$  3'.

#### 4.2.1.4 Δ BC

A preliminary investigation of Repeat B and C revealed that these two tiny and proximal repeat sequences would be impossible to investigate separately, as there was no suitable sequence that could be targeted to the intervening 157bp. The  $\Delta$  BC deletion was designed to extend from the very start of the region excised in the  $\Delta$  PflMI construct to directly 3' of the C repeat, with a partial overlap in deletions with the  $\Delta$  Bh constructs. Successful creation and identification of four  $\triangle$  BC cell lines (#2, #4, #8 and #17) occurred rapidly and with little troubleshooting using gRNA 2.1 and gRNA 3.3.  $\Delta$  BC #4 contained a heterozygous deletion, and was discarded from subsequent analysis.  $\Delta$  BC #2 and #8 were produced from separate cells in the same transfection using 1.5µl of Lipofectamine 3000.  $\triangle$  BC #17 was produced from a separate reaction using 1.5µl of Lipofectamine 3000.  $\Delta$  BC#2 had 1195bp removed from its sequence and also had a novel 141bp integrated into the sequence that failed to show any particular alignment when analyzed using BLASTn.  $\Delta$  BC #8 had 1189bp removed and without any novel sequences introduced. Finally,  $\Delta$  BC #17 had 1195bp removed from its sequence and gained a 145bp sequence that seemed to partially align with part of the PsPgRNA vector sequence. The homozygous nature of all of these deletions as well as the short insert sequences suggested that the repair of transgenic and endogenous XIST sequences was likely occurring through homology directed repair.

#### 4.2.1.5 Δ 3'PfIMI

The  $\Delta$  3'PflMI constructs were designed to lack the remainder of the ~3.8kb PflMI region 3' of the  $\Delta$  BC deletion. The  $\Delta$  3'PflMI constructs therefore lack the non-repeat region spanning

Repeat C to Repeat D. Two  $\Delta$  3'PflMI cell lines, #3 and #6, were created using the 5' gRNA, gRNA 3.1, and the 3' gRNA, gRNA 6.0.  $\Delta$  3'PflMI #3 was created from a transfection using 0.75µl of Lipofectamine 3000 reagent and  $\Delta$  3'PflMI #6 was created from a separate transfection using 1.5µl of Lipofectamine 3000 reagent. Both  $\Delta$  cell lines lacked the exact same 2859bp of *XIST*, suggesting that despite the size of the deletion the repair process must have been extremely rapid and efficient.

# 4.2.1.6 Δ D

The  $\Delta$  D construct was planned to completely remove Repeat D and extend from the very edge of the region deleted in the  $\Delta$  PflMI construct. Creating and identifying  $\Delta$  D cell lines proved relatively challenging, and the two  $\Delta$  D cell lines, #3 and #10 were generated successively through different transfection reactions both using 1.5µl of Lipofectamine 3000 reagent. The 5' and 3' gRNAs used to create these  $\Delta$  D cell lines were gRNA 5.5 and gRNA 8.5 respectively.  $\Delta$ D #3 lacked 3084bp from the *XIST* sequence while  $\Delta$  D #10 lacked 3092bp. Both of these deletions successfully remove Repeat D and overlapped with the 3'PflMI deletions by ~500bp.

#### 4.2.1.7 Δ 3D5E

The  $\Delta$  3D5E construct was designed to remove the non-repeat region that extends 3' of Repeat D to just upstream of Repeat E, resulting in a planned ~1kb overlap with the region absent in the Exon 1 construct. Deleting the region required removing roughly 3.5kb of the *XIST*, making it

the largest of the *XIST* CRISPR deletions examined here. Despite the size of the region to be deleted, numerous  $\Delta$  3D5E cell lines were identified after just a couple sets of transfections and selections. Three clones from three different reaction conditions were kept,  $\Delta$  3D5E #13, #14 and #15. All three of these cell lines were generated using the same 5' and 3' gRNAs, gRNA 8.5 and gRNA 12.2 respectively.  $\Delta$  3D5E #13 was produced from a transfection using 0.75µl of Lipofectamine 3000 reagent while both 3D5E #14 and 3D5E #15 came from treatments with 1.5µl of Lipofectamine 3000.  $\Delta$  3D5E #13 had 3584bp of its *XIST* sequence deleted while 3D5E #14 had a nearly identical 3583bp deleted.  $\Delta$  3D5E #15 had 3588bp deleted, representing a 5bp degradation on its 3' end compared to #13.

# 4.2.1.8 Δ E

The  $\Delta$  E constructs were designed to remove Repeat E of the *XIST* cDNA, a roughly 1.5kb sequence entirely encompassed within the 3.6kb deletion of Exon 1 and overlapping the  $\Delta$  3D5E deletion at exons 2-5 (figure 4.2). A combination of gRNA 11.9 and gRNA 13.7 and 1.5µl of Lipofectamine 3000 successfully resulted in two  $\Delta$  E cell lines. Combinations using gRNAs 12.2 and 13.8 proved entirely unsuccessful at excising Repeat E.  $\Delta$  E #6 in total lost 1844bp of its *XIST* cDNA sequence and  $\Delta$  E # 10 lost 1778bp of its *XIST* sequence, successfully removing Repeat E. Further, both of the  $\Delta$  E cell lines exhibited significant degradation of the cDNA beyond the expected CRISPR cut site, 83bp and 21bp respectively. The unusual degradation observed in these two cell lines may be exacerbated by the fact that homology directed repair was not possible near 5' as the inducible construct did not have the intron sequence of the endogenous copy of *XIST*.

#### **4.2.1.9** Δ **3**'

The  $\Delta$  3' construct was the smallest planned deletion and was designed to remove the non-repeat region of *XIST* from Repeat E to the end of *XIST* and encompassed in the larger deletion of the Exon 1 construct (figure 4.2). Two identical deletion cell lines for  $\Delta$  3' were successfully generated using gRNA 13.7 and gRNA 14.2. The 3' gRNA (14.2) targeted a region proximal (44bp away) to the 3' end of *XIST* without disrupting the Poly-A signal. The two  $\Delta$  3' #1 and  $\Delta$ 3' #7 were generated in separate reaction conditions using 0.75µl and 1.5µl of Lipofectamine 3000 reagent. Both of these  $\Delta$  3' cell lines had an identical 630bp deletion extending exactly from the expected cut sites, potentially due to the small size of the deletion allowing for a more rapid and efficient repair process.

#### 4.2.2 Testing the inducibility of the *XIST* Deletion constructs

To test the expression of *XIST* in the various populations of HT1080 cell lines, they were all were grown and induced for five days as described in section 2.1. The whole cell RNA was harvested and converted into cDNA according to section 2.2. The vector sequence upstream of all the *XIST* inducible constructs (described in section 1.8) was used as the template for the *XIST* qPCR primers (PCDNA5 Fwd/Rev) and the endogenous control gene *PGK1* was used as a reference for the initial qPCR analysis as described in section 2.3 (see primer table in appendix A for primer sequences).

The relative transcript levels of XIST from each of the cell lines is shown in figure 4.3 and grouped by construct type in figure 4.4 and summarized by cell and construct type in table 4.3. XIST transcript levels varied broadly between replicates, even among the Full XIST cells (Full XIST sd = 0.36). To determine whether the act of deleting sections of XIST was inherently destabilizing, the average transcript levels of the pooled population of deletion constructs was compared to Full XIST and the two groups were nearly identical (normalized  $\Delta RQ = 1.04$ , sd = 0.45, figure 4.4). Despite the variability in XIST levels between replicates, the  $\Delta$  A, E and 3' CRISPR cell lines each had unique profiles. While  $\Delta A \# 12$  showed slightly, but not significantly, lower expression than the Full XIST construct (RQ = 0.507, sd = 0.30, p = 0.02) the  $\Delta$  A #3 cell line showed incredibly low levels of XIST RNA (RQ = 5.99x10<sup>-4</sup>, sd =4.5x10<sup>-4</sup>) indicating that effectively no XIST was present within a given cell (figure 4.3 and table 4.3). Both of the  $\Delta$  3' cell lines had statistically lower levels of XIST transcripts compared to Full length XIST (average RQ =0.40,  $p = 2.0 \times 10^{-5}$  by t-test). Finally, both  $\Delta$  E cell lines contained a greater average number of XIST transcripts (average RQ = 1.76 and 1.83) than was observed in either the Full XIST cell line or any of the other deletion cell lines, however, the high variance in the transcript levels between replicates meant that this observation was not statistically significant (p = 0.081 and 0.035). The Exon 1 cell lines, lacking the 3.6kb region encompassing  $\Delta$  3' and  $\Delta$  E, had very similar transcript levels to Full XIST (RQ= 1.07 and 1.57, figure 4.4) that did not approach statistical significance (p =0.85 and 0.24). Previous work had implicated Repeat D [92] and Repeat F [272] as being critical for XIST expression, though the latter study was done using XIST in mouse ESCs. However, in the HT1080 context both cell lines had nearly identical levels of XIST RNA relative to the Full XIST construct following induction (Combined  $\Delta$  FBh RQ = 0.98, combined  $\Delta D RQ = 1.08$ , figure 4.3 and 4.4).

The purpose of this first screen by qPCR was to identify potential effects on transcript levels resulting from the various deletion events, and those that were of interest either for their unique transcript levels or due to implications in the literature were selected for a secondary set of testing in a more controlled manner. To further investigate the XIST levels of the cell lines of interest, several biological replicates of each cell line were all treated by 5ddox simultaneously. These cell lines that were all treated with the same solution of Dox supplemented media were Full XIST,  $\triangle A \#3$ ,  $\triangle A \#12$ ,  $\triangle FBh \#21$ ,  $\triangle FBh \#22$ ,  $\triangle E \#6$ ,  $\triangle E \#10$ ,  $\triangle 3' \#1$  and  $\triangle 3' \#7$ . The cells were harvested at the same time and were processed and tested all at once. Two endogenous control genes, PGK1 and UBC were used to better ensure an accurate measurement of the relative levels of XIST transcripts. The RQ of XIST for each of the biological replicates tested for each cell line using each endogenous control were shown in figure 4.5. A comparison of the relative XIST levels normalized to each control gene respectively was performed to test whether variability in RQ values was most likely due to the XIST gene itself or variability in the control genes (figure 4.5 A). In most of the cell lines the relative expression of XIST was consistent between the endogenous controls, though  $\Delta$  FBh #22 had a relatively large variation (average RQ PGK1 = 0.59, average RQ UBC = 1.48) but did not differ statistically significantly (figure 4.5) A). The relative expression of the various cell lines was then calculated by combining the two reference genes into a normalized control, as described in the methods section 2.3 (figure 4.5 B). The  $\Delta$  3' cell lines had statistically lower levels of XIST transcript relative to Full XIST, whether the  $\triangle$  3' cell lines were tested individually or together ( $\triangle$  3' #1 RQ = 0.31, p = 0.013;  $\triangle$  3' #7 RQ = 0.36, p = 0.013,  $\Delta$  3' construct RQ = 0.34, p = 6x10<sup>-4</sup>, figure 4.5 B & C). The  $\Delta$  E cell lines had a higher average expression than Full XIST or any other cell line tested here ( $\Delta E \# 6 RQ =$ 1.69 sd = 1.00,  $\Delta E \# 10 \text{ RQ} = 1.26 \text{ sd} = 0.45$ ) however this was not statistically significant from

Full *XIST* (figure 4.5 B & C). Neither  $\Delta$  FBh constructs reached statistical significance though  $\Delta$  FBh #21 did have on average half the *XIST* transcripts of either Full *XIST* or  $\Delta$  FBh #22 ( $\Delta$  FBh #21 RQ=0.46, p = 0.06;  $\Delta$  FBh #22 RQ=0.96, p=0.91, figure 4.5 B). As  $\Delta$  FBh #21 lacks an additional 300nt compared to #22, that region could be of interest in subsequent analysis of expression and transcript stability (table 4.2).

 $\Delta$  A #12 was not significantly different from Full *XIST* in its *XIST* transcript levels (RQ = 0.78, sd = 0.43, p=0.46) however  $\Delta$  A#3 showed virtually no *XIST* expression whatsoever (RQ = 3.4x10<sup>-5</sup>). It remained to be determined whether this low level of expression resulted from the *XIST* PCDNA5 primers not being able to amplify properly, or whether *XIST* was not being induced. qPCR was performed on Full *XIST* and  $\Delta$  A#3 biological replicates using three different *XIST* specific primers: qPFIMI which binds near Repeat B, q*XIST*5 which binds shortly after Repeat D and q*XIST*9 which binds upstream of Repeat E (listed appendix primer table A.2). Relative *XIST* levels were normalized to the combined endogenous control genes *UBC* and *PGK1* and the results were shown in figure 4.6 and all three demonstrated that  $\Delta$  A#3 was not being induced levels of Full *XIST* discussed in chapter 3 (RQ = 0.053, sd 0.016). As  $\Delta$  A#3 could not be induced effectively, no subsequent analysis was performed with it.



Figure 4.3 Induced XIST transcript levels in Δ cell lines relative to Full XIST

The relative expression of *XIST* RNA in the various deletion cell lines following 5ddox induction. *XIST* transcript levels were measured by qPCR and RNA levels were calculated relative to the endogenous control gene *PGK1*. Orange dots represent individual biological replicates of each cell line that were normalized to Full *XIST* treatments that were tested in parallel. The black lines denote the average (mean) relative *XIST* transcript levels and the standard deviation for each cell line. Statistical deviation from the Full *XIST* transcript levels was calculated using a two-tailed t-test with an adjusted threshold of significance accounting for the 26 different tests being performed (\* p-value =  $0.05/26 = 1.9 \times 10^{-3}$ ). The data presented here is shown on a linear scale and was normalized to a Full XIST biological replicate treated in parallel with each replicate shown here.





Figure 4.4 Transcript levels of Δ constructs relative to Full XIST

The relative expression (RQ) of *XIST* RNA in the various deletion constructs following 5ddox induction. *XIST* transcript levels were measured by qPCR and RNA levels were calculated relative to the endogenous control gene *PGK1*. The orange bars represent the average levels of *XIST* transcript relative to Full *XIST*, based on each cell lines average transcript levels. The black error bars indicate the standard deviation combined from the cell lines of each type of construct. Statistical deviation from the Full *XIST* transcript levels was calculated using a two-tailed t-test with an adjusted threshold of significance accounting for the 13 different tests being performed (\* p-value =  $0.05/13 = 3.8 \times 10^{-3}$ ).

| Construct        | Clone | Mean RQ  | SD       | p-value        | Mean    | SD    | p-value         |
|------------------|-------|----------|----------|----------------|---------|-------|-----------------|
| $\Delta$ XB      |       | 1.933    | 1.140    | 4.1E-02        | 1.933   | 1.140 | 4.06E-02        |
|                  | 3     | 5.99E-04 | 4.50E-04 | <u>8.7E-07</u> | 0.254   | 0.205 | 2.05E.06        |
| $\Delta A$       | 12    | 0.508    | 0.300    | 1.9E-02        | 0.234   | 0.293 | <u>2.93E-00</u> |
|                  | 21    | 0.678    | 0.425    | 9.6E-02        | 1.044   | 0.696 | 5 71 0 01       |
| $\Delta$ FBh     | 22    | 1.410    | 0.764    | 4.2E-01        | 1.044   | 0.080 | 3./1E-01        |
|                  | 5     | 0.971    | 0.319    | 5.3E-01        |         |       |                 |
|                  | 7     | 0.527    | 0.232    | 4.8E-02        | 0.813   | 0.315 | 1.01E-01        |
| $\Delta$ Bh      | 11    | 0.942    | 0.184    | 4.3E-01        |         |       |                 |
| $\Delta$ PfIMI   | 3     | 0.630    | 0.171    | 9.0E-02        | 0.630   | 0.171 | 8.96E-02        |
|                  | 2     | 0.653    | 0.235    | 3.1E-02        |         |       |                 |
|                  | 8     | 0.905    | 0.328    | 3.7E-01        | 0.773   | 0.272 | 1.71E-02        |
| $\Delta$ BC      | 17    | 0.762    | 0.166    | 1.2E-01        |         |       |                 |
|                  | 3     | 1.147    | 0.460    | 9.1E-01        | 1 1 5 0 | 0.480 | 0 00E 01        |
| $\Delta$ 3'PfIMI | 6     | 1.154    | 0.526    | 9.1E-01        | 1.130   | 0.489 | 0.00E-01        |
|                  | 2     | 1.258    | 0.919    | 6.8E-01        | 1.061   | 0.750 | 0.02E.01        |
| $\Delta D$       | 10    | 0.864    | 0.360    | 3.0E-01        | 1.001   | 0.739 | 9.02E-01        |
|                  | 13    | 1.249    | 0.919    | 7.0E-01        |         |       |                 |
|                  | 14    | 1.367    | 1.059    | 5.6E-01        | 1.075   | 0.857 | 8.31E-01        |
| $\Delta$ 3D5E    | 15    | 0.610    | 0.128    | 3.3E-02        |         |       |                 |
|                  | 3     | 1.072    | 0.043    | 8.5E-01        | 1 2 2 1 | 0.655 | 4 66E 01        |
| Exon 1           | 7     | 1.571    | 0.855    | 2.4E-01        | 1.321   | 0.033 | 4.00E-01        |
|                  | 6     | 1.762    | 1.085    | 8.1E-02        | 1 707   | 1.017 | 4.02E.02        |
| $\Delta E$       | 10    | 1.833    | 0.930    | 3.5E-02        | 1./9/   | 1.01/ | 4.02E-02        |
|                  | 1     | 0.365    | 0.115    | <u>1.1E-03</u> | 0.395   | 0.169 | <u>2.04E-05</u> |
| Δ 3'             | 7     | 0.425    | 0.199    | <u>1.0E-03</u> |         |       |                 |
| ΔΔ               | 12    | 0.986    | 0.739    | 6.4E-01        | 0.986   | 0.739 | 4.06E-02        |

 Table 4.3 Relative expression of XIST cell lines and constructs

Summary of the relative expression of all the deletion cell lines and constructs. Relative expression of *XIST* in each cell line was determined by qPCR using the endogenous control gene *PGK1* and all values were normalized to Full *XIST*. Cell lines are grouped by construct and listed by clone number. The mean, standard deviation (SD) and statistical significance of deviation from Full *XIST* are listed for each clone and construct. Significance was calculated using two tailed t-test and values that reached the threshold of significance are underlined.



Figure 4.5 Transcript levels of  $\Delta$  *XIST* constructs and cell lines of interest

The relative expression (RQ) of XIST RNA in deletion constructs of interest following 5ddox induction. All biological replicates were grown, treated and tested in parallel. Biological replicates are shown as orange dots and the black bars represent the mean XIST transcript level normalized to the average Full XIST level, with error bars indicating standard deviation. To help discern cells by type across panels Full XIST is labelled orange,  $\Delta$  A in red,  $\Delta$  FBh in purple,  $\Delta$  E in light blue and  $\Delta$  3'in dark green. Four biological replicates were tested for each cell lines except the  $\Delta$  FBh cell lines, where only two biological replicates each were tested. Statistical significance comparing the deletion cell lines to Full XIST was performed using a two-tailed ttest with an adjusted threshold of significance based on the four  $\Delta$  cell lines being tested (\* pvalue < 0.05, \*\* p-value < 0.01, \*\*\* p-value < 0.001). A) XIST transcript levels calculated using two separate endogenous control genes, PGK1 and UBC to compare for signs of discord in the relative levels of XIST when different endogenous control genes were used. B) Relative XIST transcript levels for each individual cell line were calculated using both endogenous control genes and normalized to Full XIST. C) Relative XIST transcript levels calculated for each of the deletion constructs.  $\Delta A \#3$  was not included in the  $\Delta A$  construct, which was ostensibly just  $\Delta A$ #12.



Primers targetting different regions of XIST in Delta A#3

#### Figure 4.6 ΔA #3 transcript levels relative to Full *XIST*

The relative expression (RQ) of *XIST* RNA in the  $\Delta A$  #3 cell line relative to Full *XIST* following 5ddox induction. *XIST* RNA levels were referenced to the endogenous control genes *PGK1* and *UBC* and normalized to Full *XIST*. Primers targeting the length of *XIST* were used to determine whether the unusually low levels of *XIST* identified using the qPCDNA5 primers was a primer specific effect or whether *XIST* was failing to become induced. The qPFLMI primer amplifies a sequence near Repeat B, q*XIST*5 amplifies a sequence after Repeat D and q*XIST*9 amplifies a sequence 5' of Repeat E. The average *XIST* transcript level prior to induction in the Full *XIST* cell line is shown as a dotted line.

#### 4.2.3 Visualizing XIST RNA localization in the deletion cell lines

In order to ultimately determine what effect the various deletions in the *XIST* inducible cell lines had on the recruitment of chromatin marks and the chromatin environment, it was necessary to confirm that XIST RNA could be visualized comparably to the Full XIST control. To do this FISH was performed as described in the methods section 2.12 on the various deletion cell lines. The cell lines were hybridized with fluorescent green probes for XIST RNA. Induced Full XIST was hybridized with the same probe mixture as was used for each deletion construct as a control for all of these tests. A sample photo for each type of deletion construct is shown in Figure 4.7. Only  $\Delta$  XB failed to produce an XIST RNA signal that could be observed, despite having relatively high levels of XIST expression (figure 4.4). This suggested that  $\Delta$  XB alone completely failed to localize and likely became diffuse and unobservable within the nucleus. With two exceptions the remaining induced (5ddox) deletion constructs were indistinguishable from Full *XIST.* The  $\Delta\Delta$  construct had a diffuse and punctate appearance immediately distinguishable from the mostly unified Full XIST RNA signal. The  $\Delta$  3D5E also had a more punctate XIST signal than Full XIST, however this signal was not as frequently or obviously diffuse as  $\Delta\Delta$  (figure 4.7). The proportion of  $\Delta\Delta$  and  $\Delta$  3D5E cells with unified XIST RNA clouds that had no gaps between strongly fluorescing regions was compared to Full XIST to determine whether this initial perception of decreased RNA unification was valid. 67% of Full XIST cells (n= 330) had a single unified XIST RNA signal, while only 21% of  $\triangle$  3D5E cells (n = 380) had unified XIST RNA signals and 2% of  $\Delta\Delta$  cells (n = 271) had unified signals. This suggested that both  $\Delta\Delta$  and  $\Delta$ 3D5E retained some ability localize into clusters, even if their ability to form a single unified XIST RNA domain was attenuated.



Figure 4.7  $\Delta$  3D5E and  $\Delta\Delta$  XIST forms a less unified RNA domain and  $\Delta$  XB RNA completely failed to localize

Examples of typical cells from each of the types of induced (5ddox) deletion constructs FISH for *XIST* RNA (green) and staining of DNA with DAPI (blue). Brightness and contrast of these images was adjusted in these photos using ImageJ and as a result the perceived size and intensity of the *XIST* RNA cloud should not be used to try to estimate the amount of *XIST* RNA. No *XIST* RNA cloud could be visualized in any  $\Delta$  XB cell examined across numerous tests despite all other samples tested in parallel showing clear *XIST* RNA signals. Cells lines had nuclei of equal size (~15-20µm at their longest point), and differences presented here come from the way images were cropped to ensure that each one would have only a single whole cell centered in each field image.

#### 4.3 Discussion

The generation of these XIST deletion constructs allowed for only transient selection to be used to enrich for potential deletion constructs. The concept of transient Cas9-mediated selection was proposed early by members of the Zhang lab in 2013, however relatively few studies have made use of purely transient selection in mammalian systems in the subsequent years [225], [273]. The troubleshooting and creation of the range of deletion XIST cell lines described here offered several novel insights into this type of gene editing in mammalian cells. Transfection of these three plasmids into a population of HT1080 cell lines to induce a desired excision was found to be most effective using the higher recommended concentrations of lipofectamine  $3000 (1.5 \mu l in$  $25\mu$ l Opti-MEM), as 16 of the 21 cell lines deletion cell lines were generated using this concentration. It was also observed that homozygous deletions of both the endogenous and transgenic XIST sequences were common among the deletions when testing both genes was possible, with the exceptions being  $\Delta A \#9$  and  $\Delta BC \#4$ . This supported other contemporary findings that CRISPR-mediated DNA cleavage occurred with high frequency at all target sites, with the successful introduction of the Cas9 and gRNA representing the main limitation of efficiency [274]–[276]. Therefore, it is assumed that transfection and repair efficiency were limiting factors.

The unpredicted loss of Repeat Bh based on the location of the 3' gRNA (1.9) in both copies of *XIST* in the independently generated  $\Delta$  FBh cell lines suggested that homology directed repair between the endogenous and transgenic *XIST* was occurring following CRISPR-mediated excision of a region. There was evidence from CRISPR studies in mammalian cells that brief intense puromycin selection led to a greater rate of homology directed repair, that may have

affected the frequency of sequence conversion between the endogenous and transgenic *XIST* [273], [277]. Recent studies of homology

directed repair efficiencies have suggested that increased repair efficiencies following Cas9mediated DNA cleavage were not only associated with puromycin but also other sources of stress, including Zeocin selection and cold shock [277], [278]. This association between puromycin induced stress even in resistant cells may therefore have contributed to the homozygosity observed in the deletion constructs due to the puromycin induced stress evident in surviving cells following selection. While stress and concentration of lipofectamine were implicated in affecting the efficiency of generating deletions, no obvious connection between the predicted efficiency of gRNAs and their observed efficiency was found [254]. With the exception of gRNA 8.6 and 13.8 every *XIST* gRNA successfully produced at least one *XIST* deletion cell line (table 4.2). This was surprising in the case of gRNA 13.8 because its predicted efficiency (E-score: 66.35) was the second highest of all the gRNAs designed (table 4.1). Over the course of the multiple rounds of transfections no connection was observed between the predicted efficiency of the gRNAs and the frequency with which deletion constructs were generated.

The extent of degradation following Cas9-mediated cleavage between the various deletion cell lines was highly variable and no clear or obvious explanation can be presented here. Degradation did not correlate with either size of the deletion, the region removed or gRNA efficiency scores as exemplified by the high-fidelity repair in the large 3D5E deletions and the extensive degradation in the smaller adjacent  $\Delta$  E deletions. Numerous factors were predicted to affect repair efficiency following Cas9-mediated cleavage including the combination of gRNAs used

together, Cas9 lingering on the DNA after cleavage and the chromatin environment [279]. Further work examining which side of the cleavage the Cas9 enzyme lingers on or treating cells with inhibitors for certain histone modifying enzymes may yield novel insights into how to improve repair fidelity and efficiency.

Having generated these various *XIST* deletion constructs it was essential to ensure their utility as models to test the various aspects of *XIST* activity, as it might be expected that removing multi-kilobase sections of the gene would catastrophically disrupt its ability to function. Though *XIST* transcript levels were highly variable, even between replicates of a given cell line, generally most of the  $\Delta$  constructs produced comparable levels of *XIST* RNA to the Full *XIST* construct.

Analysis of the amount of *XIST* transcript induced by the deletion constructs and their localization into the typical *XIST* RNA cloud revealed that in most instances any individual region across the length of *XIST* can be removed without disrupting transcript levels or unification. The  $\Delta$  A#12 was the sole testable  $\Delta$  A deletion, as it could be effectively induced to a comparable level as the Full *XIST* construct and form a clearly observable *XIST* RNA signal unlike the  $\Delta$  A #3 and  $\Delta$  XB cell lines respectively. The lack of transcripts in the  $\Delta$ A #3 cell line are most likely a result of some disruption to the inducible system, as treatment with doxycycline did not result in XIST upregulation to any degree. The  $\Delta$  XB construct that was deleted only in the  $\Delta$ XB construct may be crucial to XIST localization. Previous studies had shown that Repeat F and D were essential for normal *XIST* transcript levels [92], [272]. The observations of the relatively comparable expression of the  $\Delta$  FBh and  $\Delta$  D constructs to Full *XIST* suggested that the effect of these regions was likely specific to their endogenous context, affecting *XIST* 

expression rather than *XIST* RNA stability. This context dependent role might be especially true for the role of Repeat F as crucial for *XIST* expression, as it was tested in mouse ESCs [272]. Repeat D had been previously implicated in being critical for maintaining *XIST* expression from the Xi, and it would be fascinating to do a comparative analysis to determine what factors or processes are facilitated by Repeat D in its endogenous context [92]. The significantly lower levels of transcripts in the  $\Delta$  3' cell lines indicated that this 3' region of *XIST* may be involved in stabilizing the *XIST* transcript. It may be that this region acts analogously to the 3'UTR region of protein coding mRNAs, being bound by factors involved in regulating and stabilizing the transcript [280]. Recent studies have discovered strong enrichment of m6A RNA modifications along the 5' and 3' region of *XIST*, though the functional significance of the 3' region has yet to be examined further [116]. As m6A was associated with modulating RNA stability in a highly context dependent manner, further work examining the methylation of this 3' region may yield novel insights into how *XIST* RNA is regulated within the nucleus [281], [282].

The pathways(s) responsible for *XIST* RNA localization proved to be highly resilient even when large regions greater than 3kb were deleted, as was exemplified in the  $\Delta$  D, Exon 1,  $\Delta$  PflMI constructs. The punctate RNA signals observed in the  $\Delta$  3D5E and  $\Delta\Delta$  constructs indicated that numerous domains of *XIST* were involved in its coalescence into a single localized structure. The normal localization observed in the  $\Delta$  PflMI and Exon 1 constructs, which contained the individual deletions of the  $\Delta\Delta$  construct respectively, indicated that distal and redundant domains of *XIST* were contributing to its localization. The PflMI region and the 3' end of *XIST* were suggested to bind HNRNPU contributing to *XIST* localization, though this being the mechanism responsible for the delocalization observed here remains to be validated [105]. By validating of

the expression and localization of the various  $\Delta XIST$  constructs it was finally possible to begin to test the function of each region of *XIST* individually and ultimately start to identify the pathways that mediate its activity.

# **Chapter 5: Testing of Deletion Constructs**

#### 5.1 Introduction

The lncRNA XIST was demonstrated to be the essential element mediating XCI [66], [73]. The general consensus was that XIST functions as a large scaffold for numerous factors involved in silencing transcription and ensuring that the X remained a silent, dense, heterochromatin enriched Barr body within the nucleus [50], [270], [271], [283]. The XIST RNA coats its X chromosome of origin and the various factors bound and/or activated by it modify the encompassed chromatin. XIST has been extensively studied, primarily in mouse models, and these studies have demonstrated that numerous distinct pathways are initiated by XIST activity and that XIST exhibits a prolific binding to a wide range of proteins [103], [157]. Understanding the mechanisms of how XIST functions is of great interest as it provides novel insights into how IncRNAs function, how XCI itself works and even how XIST could be implemented as a future therapeutic tool [71], [83], [211]. The unusually large size of the XIST transcript, its poor sequence conservation across eutherian species coupled with its prolific binding had barriers to identifying its mechanisms of action [80], [89], [157], [284]. Those studies that have examined partial XIST sequences in mouse models and recently in a human model had traditionally limited their analysis to either a single facet of XIST activity or a small subset of domains [82], [92], [108], [127], [210]. The creation of the XIST deletion constructs described in chapter 4 offers a unique opportunity to circumvent these obstacles, by investigating the importance of small segments of XIST individually in the broader context of the whole construct. The tests performed using these new constructs identified the regions of XIST that were essential and dispensable for numerous aspects of XIST function.

Previous research had associated various functions of Xist with specific regions, primarily the repeat sequences predicted to most likely be functionally significant [82], [92]. The most extensively studied function of XIST is its ability to silence other genes. Early work in mouse models demonstrated that Repeat A at the 5' end of Xist is essential for establishing gene silencing [82]. There were extensive studies investigating the structure of the Xist Repeat A region as well as the factors that bind to this region [84], [85], [284], [285]. Comparatively little analysis has been performed on the human XIST Repeat A, though it has been demonstrated that XIST Repeat A is necessary and sufficient to silence a reporter gene proximal to XIST [111], [112]. One of my primary objectives therefore was to determine whether Repeat A was responsible for the ability of XIST to silence genes distal to its integration site on its chromosome of origin. The four distal SNP coding genes CTSB, DLC1, SLC27A37 and STC1 that had been shown to undergo XIST-mediated allele repression described in chapter 3.2.5.2 were used again here. Those deletion constructs which affected the silencing of all four of these genes across multiple biological replicates would be classified as important for distal gene silencing, to identify regions responsible for gene silencing as a whole, rather than factors affecting individual genes. The threshold of a statistically significant difference between the Full XIST and deletion constructs was adjusted by the number of deletion constructs being tested to avoid false positive results [250].

In addition to analyzing gene silencing, it is important to determine which regions of *XIST* are essential for restructuring of the XIST RNA coated chromatin environment. The most studied Xi associated heterochromatin marks are H3K27me3 and ubH2A. As nearly all the research on these marks has been performed in mouse models, examining their role in human cells represents a fruitful avenue to better understand both how *XIST* works as well as to draw conclusions about

the functional conservation of XIST regulation of the PRC complexes. At time of writing the current model suggests that the B and C repeat of mouse Xist are essential for Xist to recruit PRC1, which in turn recruits PRC2 and SMCHD1 independently of each other [127], [132], [140]. This model also suggests that SMCHD1 does not bind to Xist, even indirectly, but interacts with ubiquitinated H2AK119 [167]. The exact mechanisms controlling how Xist recruitment of PRC1 leads to PRC2 recruitment are not known, though it was proposed that PRC2 recruitment depends upon the cofactor Jarid2 binding to ubH2A [131]. Both H3K27me3 and ubH2A enrichment were therefore anticipated to depend upon overlapping regions of XIST if this dependency was conserved between humans and mice. The other early Xi associated heterochromatin mark H4K20me1 was robustly observed to be enriched in female cell lines and in the inducible XIST systems as described in chapter 3. Little research has been done in humans or mice examining the role of H4K20me1 in XCI or how it is recruited, with most research merely noting its enrichment on the Xi [39]. Knock-down of Eed in mouse ES cells disrupted H4K20me1 colocalization at the Xist RNA cloud, implying a link between H3K27me3 and H4K20me1 [128].

The late chromatin marks MacroH2A and SMCHD1 have been reported to be involved in the maintenance of XCI, rather than its initial establishment [40], [161]. As discussed in the introduction, little is known about the role of MacroH2A in maintaining XCI or how *XIST*-mediated its recruitment [162], [286]. There is evidence, however, that MacroH2A enrichment on the mouse Xi required the continued presence of *Xist* RNA to be maintained [126]. SMCHD1 was suggested to establish boundary elements around the inactivated regions of *XIST* and maintain the conformation of the inactivated chromatin. Loss of SMCHD1 was associated with signs of partial Xi reactivation [166], [167], [287]. Studies of mouse models have indicated that

SMCHD1 enrichment by *Xist* RNA depends upon PRC1 activity, and it is therefore of interest whether the same pathway exists for human *XIST* [167].

These five heterochromatin-associated factors and histone marks were analyzed by IF-FISH in all the deletion constructs to identify the regions of *XIST* responsible for each and any potential interconnectedness between them. The method of quantifying the enrichment described in section 2.13.2 proved effective at analyzing various time points of the Full *XIST* construct in chapter 3. The 5-day period of induction was selected as the period for testing all the deletion constructs for their enrichment, as both early and late Xi associated chromatin factors had accumulated by this point. Statistical comparisons between Full *XIST* and the deletion constructs were all performed using the Mann Whitney U test to compare the population distribution of 60-61 cells from each construct with the Full *XIST* control [249]. As 61 comparisons of enrichment of these various chromatin marks were performed, the threshold of statistical significance was adjusted accordingly [250].

The proposed *Xist* binding factor, CIZ1, was identified as associated with the Xi in both mice and humans [107], [108]. CIZ1 has being associated with mouse Repeat E of *Xist*, but its enrichment is unaffected by the loss of the 5' region of *Xist* spanning from Repeat A to C [107]. While little is known about the exact role CIZ1 plays in XCI in either mice or humans, there is evidence that it plays a role in *Xist* RNA localization and a partial reactivation of inactivated genes on the Xi. The enrichment of CIZ1 was not quantified in the same way as the other 5 factors listed but was inspected visually. Its enrichment at the Full *XIST* induced *XIST* RNA cloud was clear and universal, and it was expected that any deletion unable to recruit it would therefore be obviously distinguishable.
Up until this point the primary question has been which regions of *XIST* were responsible for adding elements to its chromatin domain. Recent work in mouse models has suggested that the removal of histone acetylation mediated by *XIST* was an essential step in *XIST* establishing a gene repressed compartment and for the full establishment of H3K27me3 and ubH2A heterochromatin domains [114]. It seemed potentially fruitful to therefore investigate whether a functional link existed between the early heterochromatin marks H3K27me3 and ubH2A and the hypoacetylation of histones by *XIST* in this inducible system. Comparing the deacetylation of H3K27 in the *XIST* RNA clouds of deletion constructs unable to recruit either ubH2A and/or H3K27me3 was performed to identify any potential relationship between these histone marks. The depletion of H3K27ac in the 8p model had been previously characterized in chapter 3, and the same method of analyzing the depletion of those marks was performed here.

The final aspect of *XIST* function to be examined though the use of these deletion constructs was the positioning of the *XIST* RNA cloud at the nucleolar periphery, a process identified as critical in the ultimate maintenance of XCI [171]. Studies of both mouse and human *XIST* had established that induction of *XIST* RNA even from autosomes would cause *XIST* to associate with the nucleolar compartment with greater frequency [171], [268]. It had been of particular interest to members of the Brown lab that the 8p inducible *XIST* had shown the strongest increase in perinucleolar localization following its induction compared to any of the other autosomal *XIST* constructs. It was therefore decided to quantify the proportion of cells for each deletion construct in which *XIST* localized to the perinucleolar compartment to determine the relevant essential regions. This analysis was performed to try to provide novel insights into how local changes in the histone code at the *XIST* domain cause its entire position within the nucleus to shift.

#### 5.2 Results

#### 5.2.1 The regions of *XIST* responsible for the silencing of distal genes

RNA from the deletion cell lines was procured as described in section 2.2 of the methods chapter. The RNA was reverse transcribed, and the four genes' products were PCR amplified and measured for their relative allelic contribution as described in sections 2.3, 2.4 and 2.6 respectively. An initial set of pyrosequencing tests were performed across the uninduced deletion constructs to determine if any of them differed noticeably from each other or their Full *XIST* progenitor. No difference was observed between constructs or cell lines, and across all cell lines there was balanced biallelic contribution of the silenced allele for *CTSB* (mean contribution = 0.509, sd = 0.096) *SLC25A37* (mean contribution = 0.500, sd = 0.047) and *STC1* (mean contribution = 0.532, sd = 0.084). The allelic contribution of *DLC1* was also consistent across the uninduced deletion constructs, though its allelic contribution was slightly skewed in all cases (mean contribution = 0.299, sd = 0.050). The consistency of the uninduced allelic contribution for all four of these genes allowed for an unambiguous comparison of the strength of silencing for the deletion constructs.

The normalized strength of silencing for the deletion constructs is shown in figure 5.1. The mean, standard deviation and calculated significance of each deletion constructs difference from Full *XIST* is listed in table 5.1. The threshold of significance was adjusted for the 52 (13x4) tests from 0.05 and 0.01 to  $9.62 \times 10^{-4}$  and  $1.92 \times 10^{-4}$ . The strength of silencing of all of the induced (5ddox) deletion constructs was normalized to Full *XIST*. As anticipated,  $\Delta$  A completely failed to silence any of the four alleles and differed significantly from Full *XIST* (silencing strength from -0.132 to -0.027, p <  $3.8 \times 10^{-7}$ , figure 5.1). The encompassing deletion of  $\Delta$  XB also failed

to silence any of the alleles, though this may be a result of its failure to form a localized *XIST* RNA cloud. Unexpectedly,  $\Delta$  FBh also failed to silence any of the four distal genes (silencing strength from -0.025 to 0.102, p < p< 1.7x10<sup>-4</sup>, figure 5.1). The deletion construct  $\Delta$  Bh showed comparable silencing to Full *XIST* at all four genes (silencing strength from 1.01 to 1.09, p > 0.15, figure 5.1), indicating that the overlap in the  $\Delta$  FBh and Bh deletion is not critical for silencing.

The loss of the small 3' non-repeat region of *XIST* also disrupted gene silencing significantly at all four genes (silencing strength from 0.103 to 0.410, p<  $3.7 \times 10^{-7}$ , figure 5.1). The role of this region was supported by the encompassing deletion constructs Exon 1 and  $\Delta\Delta$  also failing to silence all four genes (p<  $1.9 \times 10^{-6}$ , figure 5.1). No other deletion construct differed significantly in its ability to silence all four distal genes compared to the Full *XIST* construct (p >1.92 $\times 10^{-4}$ , figure 5.1).



Figure 5.1 XIST silencing of distal genes depends upon its 5' and 3' regions of XIST.

The strength of the inducible *XIST*  $\Delta$  constructs ability to induce silencing of four genes distal to the site of *XIST* induction on chromosome 8p was determined by measuring the change in allelic contribution of the *XIST* repressed allele for each gene. The change in allelic contribution was calculated based on the difference between the proportion of total gene expression coming from that allele before and after an *XIST* construct's induction. Silencing strength was normalized to the change in allelic contribution produced by Full *XIST* induction, with a value of 1 indicating equal silencing strength to Full *XIST* and a value of 0 indicating no change in allelic contribution following induction. Statistical significance of each of the 13  $\Delta$  constructs relative to Full *XIST* was calculated using the Mann Whitney U test to compare the multiple replicates of each deletion construct and the threshold of statistical significance was adjusted based on the 13 tests across 4 the four genes (\* p-value <  $0.05/(13x4) = 9.2 \times 10^{-4}$ , \*\* p-value <  $0.01/(13x4) = 1.9 \times 10^{-4}$ ).

| Construct  | CTSB      |       | DLC1     |           | <i>SLC25A37</i> |          | STC1      |       |          |           |       |          |
|------------|-----------|-------|----------|-----------|-----------------|----------|-----------|-------|----------|-----------|-------|----------|
|            | silencing | sd    | p-value  | silencing | sd              | p-value  | silencing | sd    | p-value  | silencing | sd    | p-value  |
| Full XIST  | 1.000     | 0.157 |          | 1.000     | 0.193           |          | 1.000     | 0.132 |          | 1.000     | 0.093 |          |
| Δ ΧΒ       | 0.097     | 0.057 | 4.13E-04 | 0.027     | 0.261           | 8.40E-04 | -0.037    | 0.118 | 4.13E-04 | -0.190    | 0.036 | 2.06E-04 |
| ΔΑ         | -0.055    | 0.240 | 3.63E-09 | -0.132    | 0.226           | 3.85E-07 | -0.027    | 0.185 | 3.33E-09 | -0.028    | 0.102 | 3.18E-11 |
| Δ FBh      | 0.024     | 0.081 | 6.57E-08 | -0.025    | 0.346           | 1.67E-04 | 0.004     | 0.085 | 3.33E-09 | 0.113     | 0.090 | 3.33E-09 |
| Δ Bh       | 1.008     | 0.229 | 8.10E-01 | 1.094     | 0.275           | 1.50E-01 | 1.049     | 0.248 | 1.69E-01 | 1.043     | 0.185 | 2.50E-01 |
| Δ PfIMI    | 0.726     | 0.139 | 3.05E-04 | 0.850     | 0.102           | 4.11E-02 | 0.682     | 0.175 | 1.02E-04 | 0.679     | 0.115 | 1.81E-09 |
| ΔΒC        | 0.972     | 0.252 | 9.71E-01 | 0.603     | 0.171           | 2.10E-04 | 0.619     | 0.243 | 4.91E-05 | 0.854     | 0.079 | 2.97E-02 |
| Δ 3' PfIMI | 0.852     | 0.112 | 1.30E-02 | 1.021     | 0.225           | 7.54E-01 | 0.669     | 0.144 | 4.40E-06 | 0.892     | 0.086 | 1.01E-02 |
| ΔD         | 0.590     | 0.262 | 4.78E-02 | 0.727     | 0.300           | 1.82E-02 | 0.612     | 0.264 | 2.88E-05 | 0.756     | 0.139 | 1.88E-04 |
| Exon 1     | 0.031     | 0.163 | 1.34E-08 | 0.174     | 0.343           | 1.90E-06 | 0.131     | 0.271 | 2.58E-08 | 0.255     | 0.096 | 1.17E-09 |
| Δ 3D5E     | 0.686     | 0.270 | 6.43E-03 | 0.771     | 0.171           | 1.46E-2  | 0.767     | 0.127 | 2.76E-05 | 0.897     | 0.066 | 3.80E-03 |
| ΔΕ         | 0.296     | 0.319 | 9.21E-07 | 0.959     | 0.162           | 3.77E-01 | 0.653     | 0.201 | 3.19E-06 | 0.776     | 0.237 | 1.09E-02 |
| Δ3'        | 0.129     | 0.181 | 6.57E-08 | 0.103     | 0.212           | 2.06E-07 | 0.359     | 0.067 | 6.57E-08 | 0.410     | 0.050 | 3.77E-07 |
| ΔΔ         | 0.047     | 0.126 | 3.77E-07 | 0.279     | 0.393           | 1.99E-07 | 0.267     | 0.101 | 6.57E-08 | 0.292     | 0.087 | 6.57E-08 |

Table 5.1 Summary of strength of silencing of distal genes by induced XIST deletion constructs.

The change in allelic contribution for each type of deletion construct was normalized to the Full *XIST* as described in section 2.10. Four SNP coding genes were tested for the strength of each *XIST* construct to induced silencing across multiple biological replicates (>3 per cell line). The average silencing strength (silencing) and standard deviation (sd) are shown for each construct type across each gene. The statistical significance that the strength of silencing of a deletion construct differed from the Full *XIST* control was calculated using Mann Whitney U test and the resulting p-values at each allele are shown.

#### 5.2.2 H3K27me3 enrichment by XIST was facilitated by Repeats F, Bh and E

I investigated the enrichment of H3K27me3 in the inducible *XIST* deletion constructs created in Chapter 4 to identify which region(s) of *XIST* were essential for its recruitment and activation of the PRC2 complex at the chromatin surrounded by its RNA. To measure each of the deletion constructs' ability to enrich the chromatin surrounded by *XIST* RNA clouds with H3K27me3, IF-FISH was performed as described in section 2.10 on cells with the *XIST* constructs induced for five days with doxycycline (5ddox). The IF intensity of H3K27me3 at the *XIST* RNA cloud for a deletion construct would be compared to the induced Full *XIST* construct. Those deletion constructs with levels of H3K27me3 comparable to Full *XIST* demonstrated that their deleted regions were dispensable while those constructs that differed significantly from Full *XIST* were important to this process. As a point of comparison for this first Xi mark examined in the deletion constructs a female IMR90 cells was included to determine whether the more intense qualitative enrichment of the female line would be reflected in this quantitative analysis.

The z-scores of the *XIST* inducible constructs' H3K27me3 enrichment at the *XIST* RNA cloud is presented in figure 5.2. The mean, median, standard deviation and statistical comparison (Mann Whitney U test) to Full *XIST* for each of the deletion construct is listed in Table 5.2. On average H3K27me3 fluorescence at the Full *XIST* RNA cloud was noticeably enriched (median z-score = 2.59, sd = 1.70) with the average H3K27me3 intensity there being equal to the 99<sup>th</sup> percentile of H3K27me3 fluorescent intensity measured in the nucleus. As expected, the female control IMR-90 had stronger enrichment than Full *XIST* (median z-score = 3.83, sd = 2.39, p =  $3.6 \times 10^{-4}$ , figure 5.2). The  $\Delta$  FBh construct failed to enrich its *XIST* RNA cloud with H3K27me3 and differed significantly from Full *XIST*, with H3K27me3 being effectively equivalent between the *XIST* 

RNA cloud and nuclear level in an average cell ( $\Delta$  FBh median z-score = -0.147, sd = 1.57, p = 1.11x10<sup>-13</sup>, figure 5.2). The  $\Delta$  Bh construct also failed to recruit H3K27me3, differing significantly from Full *XIST* though with a much broader variation between cells than was seen in any other construct (median z-score = 0.0787, sd = 4.869, p = 8.47x10<sup>-07</sup>, figure 5.2). The  $\Delta$  A construct technically met a low criteria for significantly differing from Full *XIST* ( $\Delta$  A median z-score = 1.59, p = 1.16x10<sup>-4</sup>) however there was clearly still a strong average enrichment of H3K27me3 (figure 5.2). The enrichment of H3K27me3 by  $\Delta$  A was significantly stronger than the  $\Delta$  FBh construct (1.49x10<sup>-9</sup>) indicating that Repeat A contributed to H3K27me3 enrichment, but was likely not essential. Taken together these results demonstrate the region extending from Repeat F to Bh is critical for *XIST* to enrich its loci with H3K27me3.

The other region of *XIST* that was found to affect H3K27me3 enrichment at the *XIST* RNA cloud was Repeat E, identified by the  $\Delta$  E construct failing to become enriched for H3K27me3 (median = -0.010, sd = 0860, p = 2.97x10<sup>-17</sup>, figure 5.2). The importance of Repeat E was supported by evidence from the larger encompassing deletions Exon 1 (p= 1.23x10<sup>-14</sup>) and  $\Delta\Delta$  (p=2.42x10<sup>-15</sup>) constructs (figure 5.2). The average enrichment for  $\Delta$  E and  $\Delta$  FBh did not differ significantly from each other (p = 0.84) further suggesting that both were equally essential. All of the other deletion constructs showed a strong *XIST*-mediated enrichment of H3K27me3 that did not differ significantly from Full *XIST* (median > 1.9, figure 5.2). These results suggest that there are two regions of *XIST* that are essential for H3K27me3 enrichment, but that these two regions are separated by ~10kb of demonstrably dispensable sequence.



## Figure 5.2 H3K27me3 enrichment by *XIST* was disrupted primarily by two regions of *XIST* centered around repeat F and Repeat E

The enrichment of H3K27me3 at the *XIST* RNA cloud of cells for each type of deletion construct was measured to determine which regions of *XIST* were essential. The fluorescence intensity of H3K27me3 and *XIST* RNA was determined by IF-FISH. The relative difference in H3K27me3 at the *XIST* RNA vs the nuclear background was calculated as a z-score for each cell, and the population of 60-61 cells counted for each control and  $\Delta$  construct are graphed as small circles. A z-score of 0 indicates that H3K27me3 at the *XIST* RNA cloud was equal to the nuclear background. A z-score of 1 indicates that the average intensity at the *XIST* RNA cloud was 1 standard deviation above the nuclear average. The colour of each box plot denotes the cell line being tested. All the inducible *XIST* constructs had undergone induction for 5 days with dox. The central line of the boxplot indicates the median z-score for each condition, the notch the confidence interval, the box itself the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference between each cell line from the Full *XIST* control was calculated using Mann Whitney U test with the p-values adjusted for multiple testing (\* p-value < 8.24x10<sup>-4</sup>, \*\* p-value < 1.47x10<sup>-4</sup>, \*\*\* p-value < 1.47x10<sup>-5</sup>).

| H3K27me3     | mean z-score | median z-score | sd    | MW p-value |
|--------------|--------------|----------------|-------|------------|
| Full XIST    | 2.901        | 2.590          | 1.704 |            |
| ΔΑ           | 1.781        | 1.590          | 1.196 | 1.16E-04   |
| $\Delta$ FBh | 0.114        | -0.147         | 1.568 | 1.11E-13   |
| $\Delta$ Bh  | 0.367        | 0.787          | 4.869 | 8.47E-07   |
| Δ PflMI      | 2.113        | 1.990          | 1.181 | 1.11E-02   |
| $\Delta$ BC  | 1.892        | 1.910          | 1.127 | 1.17E-03   |
| Δ 3'PflMI    | 2.191        | 2.040          | 1.248 | 1.54E-02   |
| ΔD           | 2.156        | 1.900          | 1.519 | 1.55E-02   |
| Δ 3D5E       | 2.357        | 2.110          | 1.912 | 1.01E-01   |
| Exon 1       | 0.084        | 0.110          | 1.407 | 1.23E-14   |
| ΔΕ           | 0.043        | -0.010         | 0.860 | 2.97E-17   |
| Δ 3'         | 2.150        | 2.140          | 1.266 | 1.99E-02   |
| ΔΔ           | 0.085        | 0.338          | 1.252 | 2.42E-15   |
| IMR90        | 4.500        | 3.830          | 2.392 | 3.60E-04   |

### Table 5.2 Average H3K27me3 enrichment and XIST RNA cloud

The level of H3K27me3 enrichment at the *XIST* RNA cloud seen across each population of cells was summarized as well as the p-value indicating the statistical significance of the difference between the Full *XIST* control and the other cell lines. With the exception of the IMR90 cell line, *XIST* was expressed from chromosome 8p in all the cell lines following 5 days of induction. The mean and median z-scores for each population of cells are listed as well as the standard deviation (sd). The statistical significance of the difference between each cell line and the Full *XIST* population of cells was calculated using the Mann Whitney U test.

# 5.2.3 UbH2A enrichment by *XIST* was facilitated by numerous distinct and essential domains of *XIST*

Extensive research in recent years in mice has linked the *Xist*-mediated activity of PRC1 and PRC2 together, with their recruitment depending upon the Repeat B and C spanning region.

The prevailing view was that this mouse pathway would be conserved in human *XIST*. However, my own analysis of H3K27me3 suggests that Repeat B and C are dispensable from PRC2 recruitment, suggesting an alternate pathway exists in human *XIST*. To identify the regions of *XIST* involved in ubH2A enrichment, IF-FISH was performed as described in section 2.12 using ubH2A specific antibodies and *XIST* RNA probes, then the images analyzed as described in 2.13.2 on the various 5ddox treated deletion constructs and Full *XIST*. The results of these tests are shown in figure 5.3 and the average, standard deviation and statistical significance of the results were listed in table 5.3

Inducing the Full *XIST* construct resulted in a strong ubH2A enrichment at the site of *XIST* RNA, though with a large variance between cells (median z-score = 4.18, sd = 3.18). Absence of Repeats B and C in the  $\Delta$  BC construct resulted in a failure of *XIST* to become enriched for ubH2A (median z-score = 0.719, sd = 0.712), a significant difference from Full *XIST* (p = 4.35x10<sup>-21</sup>, figure 5.3). The importance of the region around Repeat B and C of *XIST* is supported by the encompassing deletion  $\Delta$  PfIMI also failing to become enriched with ubH2A (median z-score = -0.180, sd = 1.776, p = 1.83x10<sup>-16</sup>). The results of this analysis demonstrated that the region of human *XIST* encompassing Repeats B and C are essential for the enrichment of ubH2A by PRC1.

Testing the other regions of *XIST* for their role in ubH2A enrichment produced the entirely novel discovery of three wholly separate and essential regions (figure 5.3). The  $\Delta$  A RNA cloud was completely un-enriched for ubH2A in the average cell following induction and differed significantly from Full *XIST* (median z-score = 0.223, sd = 1.628, p = 2.84x10<sup>-17</sup>) The  $\Delta$  D construct that lacks Repeat D was also incapable of enriching its RNA cloud with ubH2A (median z-score = -0.425, sd = 0.694, p = 4.35x10<sup>-21</sup>). The non-repeat 3' 600bp most section of *XIST* proved critical for ubH2A enrichment, (median z-score = -0.635, sd = 0.714, p = 4.80x10<sup>-21</sup>, figure 5.3). The essential nature of the 3' region of *XIST* is supported further by the failure of Exon 1 to become enriched for ubH2A (median z-score = -0.352, sd = 2.896, p = 1.20x10<sup>-16</sup>), which did not differ significantly from  $\Delta$  3' (p = 0.29). The association between ubH2A enrichment and these three regions, Repeat A, D and the non-repeat 3' region is entirely novel to this work. The double deletion construct  $\Delta\Delta$  failed to become enriched for ubHA (median z-score = 0.039, sd = 1.696, p = 7.71x10<sup>-17</sup>) likely from lacking both the 3' region of *XIST* or the  $\Delta$  BC region.

Despite H3K27me3 enrichment being dependent upon the region surrounding Repeat F,  $\Delta$  FBh showed a strong enrichment of ubH2A (median z-score = 3.328, sd = 2.937) that did not differ from Full *XIST* (p = 0.148) but did differ significantly from its adjacent construct,  $\Delta$  A (p = 6.09x10<sup>-15</sup>). The  $\Delta$  E construct also shows a noticeable enrichment for ubH2A that does not differ significantly from Full *XIST* (median z-score = 2.523, sd = 3.184, p = 5.91x10<sup>-3</sup>) but does differ significantly from the larger encompassing deletion in the construct Exon 1 (p = 2.89x10<sup>-14</sup>). The comparison of the regions responsible and dispensable for H3K27me3 and ubH2A

indicates clearly that human *XIST* recruits these marks through separate domains with no clear evidence of the recruitment/activation of one complex affecting the other.



### Figure 5.3 *XIST* enrichment of ubH2A was facilitated by four distinct interdependent domains of *XIST*

The enrichment of ubH2A at the XIST RNA cloud of cells for each type of deletion construct as well as Full XIST was measured to determine which regions of XIST were essential. The fluorescence intensity of ubH2A and XIST RNA was determined by IF-FISH. The relative difference in ubH2A at the XIST RNA cloud vs the nuclear background was calculated as a zscore for each cell, and the population of 60-61 cells counted for Full XIST and the  $\Delta$  constructs are graphed as small circles. A z-score of 0 indicates that ubH2A at the XIST RNA cloud was equal to the nuclear background. A z-score of 1 indicates that the average intensity of ubH2A at the XIST RNA cloud is 1 standard deviation above the nuclear average. The colour of each box plot denotes the inducible construct being tested. All the inducible XIST constructs underwent induction for 5 days with dox. The central line of the boxplot indicates the median z-score for each condition, the notch the confidence interval, the box itself the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference between each cell line from the Full XIST control was calculated using Mann Whitney U test with the p-values adjusted for multiple testing (\* p-value  $< 8.24 \times 10^{-4}$ , \*\* p-value  $< 1.47 \times 10^{-4}$ , \*\*\* p-value  $< 1.47 \times 10^{-4}$ , \*  $1.47 \times 10^{-5}$ ).

| UbH2A        | mean z-score | median z-score | sd    | MW p-value |
|--------------|--------------|----------------|-------|------------|
| Full XIST    | 4.745        | 4.178          | 3.179 |            |
| ΔΑ           | 0.051        | 0.223          | 1.628 | 2.84E-17   |
| $\Delta$ FBh | 3.924        | 3.328          | 2.937 | 1.48E-01   |
| $\Delta$ Bh  | 6.993        | 5.972          | 4.801 | 6.30E-03   |
| Δ PflMI      | 0.400        | -0.180         | 1.776 | 1.83E-16   |
| $\Delta$ BC  | -0.719       | -0.653         | 0.712 | 4.35E-21   |
| Δ 3'PflMI    | 3.761        | 2.848          | 3.237 | 3.65E-02   |
| ΔD           | -0.402       | -0.425         | 0.694 | 8.30E-21   |
| Δ 3D5E       | 3.785        | 3.052          | 2.987 | 8.95E-02   |
| Exon 1       | -0.112       | -0.352         | 2.896 | 1.20E-16   |
| ΔΕ           | 3.447        | 2.523          | 3.184 | 5.91E-03   |
| Δ 3'         | -0.526       | -0.635         | 0.714 | 4.80E-21   |
| ΔΔ           | 0.081        | 0.039          | 1.696 | 7.71E-17   |

 Table 5.3 Average ubH2A enrichment and XIST RNA cloud

The level of ubH2A enrichment at the *XIST* RNA cloud seen across each population of cells is summarized as well as the p-value indicating the statistical significance of the difference between the Full *XIST* control and the other cell lines. In all cell lines *XIST* was being expressed from chromosome 8p following 5 days of induction with doxycyline. The mean and median z-scores for each population of cells are listed as well as the standard deviation (sd). The statistical significance of the difference between each construct and the Full *XIST* population of cells was calculated using the Mann Whitney U test.

#### 5.2.4 Two distinct regions of XIST were both essential for H4K20me1 enrichment

Unlike for the previous two marks, there were no previous studies indicating which regions are potentially critical for H4K20me1 enrichment. Therefore testing its enrichment using the various deletion constructs promised to provide purely novel insights into how *XIST* mediates PRSET-7 activity. IF-FISH was performed and images analyzed as described in section 2.12 and 2.13.2 of the methods sections. The results for these analyses are shown in the boxplot in figure 5.4 and the results summarized in table 5.4.

H4K20me1 was enriched at the Full *XIST* RNA cloud of most cells following the standard five days of induction (median z-score being 2.489, sd 1.594, figure 5.4). Most of the deletion constructs showed comparable H4K20me1 enrichment with the Full *XIST* construct, however two separate regions of *XIST* were both identified as essential. The  $\Delta$  3'PflMI, that lacked the non-repeat region between Repeats C and D, failed to enrich its chromatin domain for H4K20me1 (median z-score = 0.027, sd = 1.686) and differed significantly from Full *XIST* (p = 1.19x10<sup>-10</sup>, figure 5.4). The specific importance of this non-repeat region was supported by the larger encompassing deletion of the  $\Delta$  PflMI construct also disrupting H4K20me1 enrichment (median z-score = -0.061, sd = 1.211, p = 1.13x10<sup>-15</sup>) but the adjacent deletion in  $\Delta$  BC becoming strongly enriched like Full *XIST* (median z-score = 2.038, sd = 0.779, p =0.079, figure 5.4).

The  $\Delta$  E showed no signs of H4K20me1 enrichment, indicating Repeat E is also a critical element in that pathway (median z-score = 0.432, sd = 0.845, p = 3.44x10<sup>-13</sup>, figure 5.4). The importance of Repeat E is supported by the lack of enrichment in the larger deletion construct Exon 1 (median z-score = -0.026, sd = 1.357, p = 2.68x10<sup>-14</sup>). The observation that the two

adjacent deletion constructs,  $\Delta$  3' and  $\Delta$  3D5E both had very similar enrichment profiles to Full *XIST* (median z-scores 2.311 and 2.080 respectively) suggests that Repeat E specifically is essential for H4K20me1 enrichment (figure 5.4). This is the first evidence indicating the regions of *XIST* responsible for H4K20me1 enrichment. As with the previous two marks, H4K20me1 enrichment was facilitated by multiple essential and unique domains of XIST.



## Figure 5.4 *XIST*-mediated enrichment of H4K20me1 depends upon the non-repeat 3'PfIMI region as well as Repeat E.

The enrichment of H4K20me1 at the XIST RNA cloud of cells for each type of deletion construct as well as Full XIST was measured to determine which regions of XIST were essential. The fluorescence intensity of H4K20me1 and XIST RNA was determined by IF-FISH. The relative difference in H4K20me1 at the XIST RNA cloud vs the nuclear background was calculated as a z-score for each cell, and the population of 60-61 cells counted for Full XIST and the  $\Delta$ constructs were graphed as small circles. A z-score of 0 indicates that H4K20me1 at the XIST RNA cloud was equal to the nuclear background. A z-score of 1 indicates that the average intensity of H4K20me1 at the XIST RNA cloud is 1 standard deviation above the nuclear average. The colour of each box plot denotes the inducible construct being tested. All the inducible XIST constructs underwent induction for 5 days with dox. The central line of the boxplot indicates the median z-score for each condition, the notch the confidence interval, the box itself the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference between each cell line from the Full XIST control was calculated using Mann Whitney U test with the p-values adjusted for multiple testing (\* p-value  $< 8.24 \times 10^{-1}$ <sup>4</sup>, \*\* p-value <  $1.47 \times 10^{-4}$ , \*\*\* p-value <  $1.47 \times 10^{-5}$ ).

| H4K20me1     | mean z-score | median z-score | sd    | MW p-value |
|--------------|--------------|----------------|-------|------------|
| Full XIST    | 2.432        | 2.489          | 1.594 |            |
| ΔΑ           | 2.187        | 2.052          | 1.327 | 2.57E-01   |
| $\Delta$ FBh | 2.290        | 2.191          | 1.187 | 5.84E-01   |
| $\Delta$ Bh  | 2.018        | 1.827          | 1.097 | 1.07E-01   |
| Δ PflMI      | -0.173       | -0.061         | 1.211 | 1.13E-15   |
| $\Delta$ BC  | 1.999        | 2.038          | 0.779 | 7.93E-02   |
| Δ 3'PflMI    | 0.251        | 0.027          | 1.686 | 1.19E-10   |
| ΔD           | 2.106        | 1.965          | 1.240 | 2.11E-01   |
| Δ 3D5E       | 2.158        | 2.080          | 1.143 | 2.89E-01   |
| Exon 1       | -0.108       | -0.026         | 1.357 | 2.68E-14   |
| ΔΕ           | 0.404        | 0.432          | 0.845 | 3.44E-13   |
| Δ 3'         | 2.316        | 2.311          | 1.387 | 8.21E-01   |
| ΔΔ           | -0.328       | 0.042          | 1.221 | 3.27E-16   |

 Table 5.4 Average H4K20me1 enrichment and XIST RNA cloud

The level of H4K20me1 enrichment at the *XIST* RNA cloud seen across each population of cells is summarized as well as the p-value indicating the statistical significance of the difference between the Full *XIST* control and the other cell lines. In all cell lines *XIST* was expressed from chromosome 8p following 5 days of induction with doxycyline. The mean and median z-scores for each population of cells are listed as well as the standard deviation (sd). The statistical significance of the difference between each construct and the Full *XIST* population of cells was calculated using the Mann Whitney U test.

#### 5.2.5 MacroH2A enrichment requires large segments along the length of XIST

To begin to understand better what regions of *XIST* are responsible for MacroH2A recruitment, IF-FISH was performed as described in section 2.12 and the images analyzed as described in section 2.13.2. The results of this analysis are shown in graphical form in figure 5.5 and summarized in the accompanying table 5.5.

Following 5ddox induction the Full *XIST* RNA domain becomes enriched for MacroH2A, though the magnitude of the effect was less than the other marks described (median z-score = 1.958, sd = 1.469, figure 5.5). Little work has investigated how *XIST* recruits MacroH2A so no initial predictions were made about which region(s) would prove essential for its recruitment.

When the deletion constructs were compared to Full *XIST*, it was observed that several unique regions of *XIST* are all required for MacroH2A enrichment.  $\Delta$  FBh failed to become enriched for MacroH2A and differed significantly from Full *XIST* (median z-score = -0.434, sd = 1.189, p = 5.34x10<sup>-17</sup>, figure 5.5).  $\Delta$  3'PflMI was also significantly weaker at enriching its *XIST* RNA cloud with MacroH2A (median z-score = 0.313, sd = 2.197, p = 3.06x10<sup>-9</sup>, figure 5.5). The importance of this non-repeat region was supported by the lack of enrichment by larger encompassing deletion in the  $\Delta$  PflMI (median z-score = -0.348, sd = 1.124, p = 1.89x10<sup>-17</sup>). The  $\Delta$  D construct, lacking Repeat D also showed noticeably reduced MacroH2A compared to Full *XIST* (median z-score = -1.042, sd = 2.338, p = 4.25x10<sup>-17</sup>, figure 5.5). The large adjacent deletions,  $\Delta$  3'PflMI and  $\Delta$  D, shared a ~400bp overlapping deleted region, and future research into how *XIST* mediates MacroH2A enrichment could benefit by initially focusing on this overlapping region.

 $\Delta$  E failed to become enriched for MacroH2A (median z-score = -1.173, sd = 1.784, p = 6.08x10<sup>-17</sup>) as did the encompassing deletion in construct Exon 1 (median z-score = -0.096, sd = 1.380, p = 9.44x10<sup>-15</sup>, figure 5.5). The  $\Delta\Delta$  construct lacking both the PfIMI region and the region after exon 1 failed to become enriched for MacroH2A (median z-score = -0.737, sd = 1.731, p = 3.76x10<sup>-18</sup>). The remaining constructs not mentioned here became enriched for MacroH2A and did not significantly differ from Full *XIST* (figure 5.5). The results of this analysis revealed that MacroH2A requires the presence of at least three entirely distinct regions of *XIST*.



## Figure 5.5 MarcoH2A enrichment by *XIST* RNA depends on Repeat F, Repeat E and a broad internal region of *XIST*

The enrichment of MacroH2A at the XIST RNA cloud of cells for each type of deletion construct as well as Full XIST was measured to determine which regions of XIST were essential. The fluorescence intensity of MacroH2A and XIST RNA was determined by IF-FISH. The relative difference in MacroH2A at the XIST RNA cloud vs the nuclear background was calculated as a z-score for each cell, and the population of 60-61 cells counted for Full XIST and the  $\Delta$ constructs are graphed as small circles. A z-score of 0 indicates that MacroH2A at the XIST RNA cloud was equal to the nuclear background. A z-score of 1 indicates that the average intensity of MacroH2A at the XIST RNA cloud is 1 standard deviation above the nuclear average. The colour of each box plot denotes the inducible construct being tested. All the inducible XIST constructs underwent induction for 5 days with dox. The central line of the boxplot indicates the median zscore for each condition, the notch the confidence interval, the box itself the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference between each cell line from the Full XIST control was calculated using Mann Whitney U test with the p-values adjusted for multiple testing (\* p-value  $< 8.24 \times 10^{-4}$ , \*\* p-value  $< 1.47 \times 10^{-4}$ , \*\*\* p-value  $< 1.47 \times 10^{-5}$ ).

| MacroH2A     | mean z-score | median z-score | sd    | MW p-value |
|--------------|--------------|----------------|-------|------------|
| Full XIST    | 2.249        | 1.958          | 1.469 |            |
| ΔΑ           | 2.425        | 2.213          | 2.227 | 6.26E-01   |
| $\Delta$ FBh | -0.410       | -0.434         | 1.189 | 5.34E-17   |
| $\Delta$ Bh  | 2.257        | 1.709          | 2.639 | 2.90E-01   |
| Δ PflMI      | -0.439       | -0.348         | 1.124 | 1.89E-17   |
| $\Delta$ BC  | 1.528        | 1.562          | 2.247 | 8.62E-03   |
| Δ 3'PflMI    | 0.412        | 0.313          | 2.197 | 3.06E-09   |
| ΔD           | -1.484       | -1.042         | 2.338 | 4.25E-17   |
| Δ 3D5E       | 1.787        | 1.541          | 1.444 | 1.09E-01   |
| Exon 1       | -0.196       | -0.096         | 1.380 | 9.44E-15   |
| ΔΕ           | -1.101       | -1.173         | 1.784 | 6.08E-17   |
| Δ 3'         | 2.946        | 2.609          | 2.288 | 7.58E-02   |
| ΔΔ           | -1.002       | -0.737         | 1.731 | 3.76E-18   |

 Table 5.5 Average MacroH2A enrichment and XIST RNA cloud

The level of MacroH2A enrichment at the *XIST* RNA cloud seen across each population of cells is summarized as well as the p-value indicating the statistical significance of the difference between the Full *XIST* control and the other cell lines. In all cell lines *XIST* was being expressed from chromosome 8p following 5 days of induction with doxycyline. The mean and median zscores for each population of cells are listed as well as the standard deviation (sd). The statistical significance of the difference between each construct and the Full *XIST* population of cells was calculated using the Mann Whitney U test.

#### 5.2.6 SMCHD1 enrichment depends upon a unique series of segments of XIST

The regions of XIST needed for enrichment of late heterochromatin Xi associated factor SMCHD1 were identified using the *XIST* deletion constructs and this information was used to infer whether there might be a connection to PRC1. SMCHD1 enrichment was observed by IF-FISH and analyzed as described in section 2.12 and 2.13.2. The results are presented as a graph in figure 5.6 with the average, standard deviation and statistical significance relative to *XIST* summarized in table 5.6.

Full *XIST* strongly enriched its *XIST* RNA cloud with SMCHD1 following 5ddox induction (median z-score 2.704, sd = 2.167, figure 5.6). The observation that SMCHD1 recruitment by *XIST* was not associated with PRC2 activity [46] was supported in testing these deletion constructs. The  $\Delta$  constructs that had failed to enrich H3K27me3,  $\Delta$  FBh, Bh and E, all showed strong average enrichment of SMCHD1 that did not differ significantly from Full *XIST* (figure 5.6). The  $\Delta$  FBh,  $\Delta$  Bh and  $\Delta$  E constructs all had average levels of SMCHD1 enrichment zscores greater than 2 (figure 5.6).

The most 5' region that was essential for SMCHD1 enrichment is the region encompassing Repeat B and Repeat C, as  $\Delta$  BC showed no discernable enrichment for SMCHD1 and differed significantly from Full *XIST* (median z-score= -0.516, sd = 1.073, p = 1.54x10<sup>-16</sup>, figure 5.6). The encompassing deletion of the  $\Delta$  PflMI construct showed a similar inability to become enriched for SMCHD1 supporting the role of the Repeat B and C region being critical for SMCHD1 enrichment (mean z-score = 0.022, median z-score= -0.109, sd = 1.179, p = 1.04x10<sup>-14</sup>, figure 5.6). The evidence from these two deletion constructs suggests that the mechanism itself used by *XIST* to recruit SMCHD1 may be conserved at Repeat B and C repeats between mice and humans, despite the very different size of these repeats between the species.

Through analysis of the other deletion cell lines the novel discovery was made that several additional regions of *XIST* are also essential for SMCHD1 enrichment. The  $\Delta$  D construct, lacking Repeat D, did not enrich SMCHD1 at the RNA cloud, and differed significantly from Full *XIST* (median z-score= -0.350, sd = 1.629, p = 2.47x10<sup>-14</sup>, figure 5.6). The absence of the 3' adjacent non-repeat region spanning between Repeats D and E in the  $\Delta$  3D5E construct was also identified as being essential for SMCHD1 enrichment (median z-score= -0.386, sd = 1.011, p = 6.35x10<sup>-17</sup>).

The 600nt in the 3' non-repeat region of *XIST* construct is unusual as it had a significantly weakened but not completely absent enrichment of SMCHD1 compared to Full *XIST* (median z-score= 1.291, sd = 1.582, p =  $2.07 \times 10^{-7}$ , figure 5.6). The encompassing deletion of the Exon 1 construct completely fails to enrich SMCHD1 (median z-score= -0.407, sd = 1.003, p =  $1.05 \times 10^{-17}$ ) and the adjacent deletion of  $\Delta$  E was comparable to Full *XIST* (p = 0.19, figure 5.6). The similar z-scores for the Exon 1 and  $\Delta$  3D5E constructs suggest that the shared deletion of the small introns 2-5 as well as part of the 5' region of exon 6 may be of particular importance for SMCHD1. The 3' non-coding region of *XIST* is clearly important for its enrichment but does not seem to represent as essential an element as the other SMCHD1 critical region of *XIST*. The remaining deletion constructs not discussed here had comparable to this pathway (figure 5.6). This evidence also supports the potential role of PRC1 being involved in SMCHD1 enrichment, as Repeats B, C and D as well as that 3' most non-repeat region are all essential for both

functions.  $\Delta$  D was also implicated in PRC1 enrichment however  $\Delta$  3D5E was not, suggesting that while a connection may exist between these Xi factors, there are clearly additional elements required for SMCHD1 enrichment.



### Figure 5.6 SMCHD1 enrichment by XIST depends upon three distinct domains of XIST The enrichment of SMCHD1 at the XIST RNA cloud of cells for each type of deletion construct as well as Full XIST was measured to determine which regions of XIST were essential. The fluorescence intensity of SMCHD1 and XIST RNA was determined by IF-FISH. The relative difference in SMCHD1 at the XIST RNA cloud vs the nuclear background was calculated as a zscore for each cell, and the population of 60-61 cells counted for Full XIST and the $\Delta$ constructs are graphed as small circles. A z-score of 0 indicated that SMCHD1 at the XIST RNA cloud is equal to the nuclear background. A z-score of 1 indicated that the average intensity of SMCHD1 at the XIST RNA cloud was 1 standard deviation above the nuclear average. The colour of each box plot denoted the inducible construct being tested. All the inducible XIST constructs underwent induction for 5 days with Dox. The central line of the boxplot indicated the median zscore for each condition, the notch the confidence interval, the box itself the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference between each cell line from the Full XIST control is calculated using Mann Whitney U test with the p-values adjusted for multiple testing (\* p-value $< 8.24 \times 10^{-4}$ , \*\* p-value $< 1.47 \times 10^{-4}$ , \*\*\* pvalue < $1.47 \times 10^{-5}$ ).

| SMCHD1       | mean z-score | median z-score | sd    | MW p-value |
|--------------|--------------|----------------|-------|------------|
| Full XIST    | 3.137        | 2.704          | 2.167 |            |
| ΔΑ           | 2.848        | 2.601          | 2.307 | 4.36E-01   |
| $\Delta$ FBh | 2.334        | 2.161          | 1.521 | 5.00E-02   |
| $\Delta$ Bh  | 2.431        | 2.249          | 1.576 | 6.13E-02   |
| Δ PflMI      | 0.022        | -0.109         | 1.179 | 1.04E-14   |
| $\Delta$ BC  | -0.390       | -0.516         | 1.073 | 1.54E-16   |
| Δ 3'PflMI    | 3.098        | 2.902          | 1.951 | 9.98E-01   |
| ΔD           | -0.078       | -0.350         | 1.629 | 2.47E-14   |
| Δ 3D5E       | -0.326       | -0.386         | 1.011 | 6.35E-17   |
| Exon 1       | -0.505       | -0.407         | 1.003 | 1.05E-17   |
| ΔΕ           | 3.842        | 3.301          | 2.814 | 1.89E-01   |
| Δ 3'         | 1.405        | 1.291          | 1.582 | 2.01E-07   |
| ΔΔ           | -0.051       | -0.283         | 1.989 | 2.64E-15   |

 Table 5.6 Average SMCHD1 enrichment and XIST RNA cloud

The level of SMCHD1 enrichment at the *XIST* RNA cloud seen across each population of cells was summarized as well as the p-value indicating the statistical significance of the difference between the Full *XIST* control and the other cell lines. In all cell lines *XIST* was being expressed from chromosome 8p following 5 days of induction with doxycyline. The mean and median z-scores for each population of cells are listed as well as the standard deviation (sd). The statistical significance of the difference between each construct and the Full *XIST* population of cells was calculated using the Mann Whitney U test.

#### 5.2.7 CIZ1 enrichment depends upon Repeat E and a novel non-repeat region

CIZ1 is unique among all the Xi associated factors investigated using fluorescent microscopy, as every single Full *XIST* RNA cloud was obviously and clearly enriched for the factor, without exception. For this reason, when investigating the regions of human *XIST* responsible for CIZ1 enrichment, it was decided to forgo the quantification described in section 2.13.2 of the methods chapter and instead simply report the proportion of cells with clear enrichment. CIZ1 and *XIST* were fluorescently labelled by IF-FISH respectively as described in section 2.12 of the methods chapter. An example image for each of the deletion constructs is shown in figure 5.7 to illustrate the clear enrichment of CIZ1 at the *XIST* RNA cloud for each of the deletion constructs and the proportion of cells with clear enrichment is listed beside each deletion construct.

The majority of the deletion constructs show a clear enrichment for CIZ1 at every single *XIST* RNA cloud. The  $\Delta$  E construct however showed no enrichment for CIZ1 at a single one of its *XIST* RNA clouds (figure 5.7). This observation was validated by the encompassing deletions in the constructs Exon 1 and  $\Delta\Delta$  constructs that also had no CIZ1 enriched at any of the *XIST* RNA clouds. This finding suggests that the role of Repeat E as essential for CIZ1 enrichment is conserved between mice and humans.

The non-repeat region of *XIST* directly 3' of Repeat Bh was also identified as essential for CIZ1 enrichment. The  $\Delta$  Bh construct completely failed to exhibit the typical bright CIZ1 signal at any of the cells analyzed. The  $\Delta$  FBh showed a slightly attenuated but still very strong enrichment of CIZ1 at 89% of its *XIST* RNA clouds. Taken together, the evidence from these two deletion constructs suggests that the non-repeat region directly 3' of  $\Delta$  Bh is essential for CIZ1

enrichment, and the loss of the adjacent Repeat Bh may impact the mechanisms and/or secondary structure responsible for CIZ1 binding.



Figure 5.7 CIZ1 enrichment by XIST RNA depended on both Repeat E and a small non-

repeat region of exon 1
Enrichment of CIZ1 (red) at the *XIST* RNA (green) is compared across the various deletion constructs of *XIST* in the nuclei of cells labelled with DAPI. In almost all constructs there was an all or none enrichment of CIZ1 at the *XIST* RNA cloud, with the proportion of cells enriched for each construct included in brackets next to the constructs name. scale bars indicate 10 $\mu$ m. To fully illustrate the absence of enrichment in  $\Delta$ Bh and  $\Delta$ E the example images chosen were ones with adjacent nuclei, so the complete lack of enrichment could be observed in both. The *XIST* RNA in all constructs was induced for five days prior to IF-FISH being performed.

# 5.2.8 *XIST*-mediated histone deacetylation associated with regions critical for PRC2 but not PRC1 recruitment

The ability of the various constructs to deplete the area within their *XIST* RNA cloud of H3K27ac was quantified by IF-FISH (section 2.12 and 2.13.2). Cells were analyzed for the relative depletion of H3K27ac using the method of quantification described in section 2.13.2. The deletion constructs  $\Delta$  FBh,  $\Delta$  Bh and  $\Delta$  E were selected as they had completely failed to become enriched for H3K27me3 (figure 5.2). The  $\Delta$  3' construct was also included as it completely failed to enrich its RNA cloud with ubH2A (figure 5.3), but showed normal enrichment of H3K27me3. A female control cell line (hTERT-RPE1) was also included as an additional control for this analysis.

The results of the quantification of H3K27ac depletion are presented in figure 5.8 and summarized in the accompanying table 5.8. The primary observation of testing these deletion constructs is that all of them still show clear depletion of H3K27ac following 5 days of induction (average z-score < -1.5, figure 5.8) The site of the Full *XIST* RNA cloud is hypoacetylated (median z-score= -2.302, sd = 1.286) to a technically greater magnitude than the female control cell hTERT RPE-1 (median z-score= -1.661, sd = 0.68232, p =  $5.815 \times 10^{-4}$ , figure 5.8). It was supposed that this apparent weakness of deacetylation may have been offset by the greater overall density of chromatin in the female cell lines compared to the inducible system (figure 3.5 B). A slightly attenuated hypoacetylation at the *XIST* RNA clouds of the two  $\Delta$  FBh test coverslips reached the threshold of significance (p =  $1.056 \times 10^{-4}$  and p= $6.44 \times 10^{-3}$ , figure 5.8). The  $\Delta$  Bh coverslip also showed attenuated hypoacetylation at the *XIST* RNA cloud (median z-score= -1.645, sd = 0.817, p =  $6792 \times 10^{-5}$ ) as did  $\Delta$  E (median z-score = -1.80, sd = 0.815, p =

4.046x10<sup>-3</sup>, figure 5.8). All of these constructs were unable to recruit H3K27me3 (figure 5.2) suggesting a connection between *XIST*-mediated histone hypoacetylation of H3K27ac and PRC2-mediated enrichment of H3K27me3. The depletion of H3K27ac at the  $\Delta$  3' *XIST* RNA cloud was nearly identical to Full *XIST* (median z-score= -2.321, sd = 1.810, p = 0.93) indicating that loss of the 3' region of *XIST* and by extension ubH2A enrichment did not impact deacetylation in any measurable way.



Figure 5.8 *XIST*-mediated histone deacetylation weakly associated with PRC2 but not ubH2A essential domains.

The depletion of histone H3K27ac at the *XIST* RNA cloud was tested for a potential connection or link to the PRC complexes. Deletion constructs previously identified as being unable to enrich their RNA with either H3K27me3 or ubH2A were tested to see if their depletion of H3K27ac was different from control Full *XIST* construct. The fluorescence intensity of H3K27ac and *XIST* RNA was determined by IF-FISH. The relative difference in H3K27ac at the *XIST* RNA cloud vs the nuclear background was calculated as a z-score for each cell, and the population of 60-61 cells counted for Full *XIST* and the  $\Delta$  constructs were graphed as a small circle. The colour of each box plot denotes the inducible construct being tested. The inducible *XIST* constructs underwent induction for 5 days with dox. The central line of the boxplot indicates the median zscore for each condition, the notch the confidence interval, the box itself the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference between each cell line from the Full *XIST* control was calculated using Mann Whitney U test with the p-values adjusted for multiple testing (\* p-value <  $8.33 \times 10^{-4}$ ). Two replicates (rep 1 & 2) of  $\Delta$ FBh were included to provide insight into how consistent the testing of H3K27ac was between biological replicates in  $\Delta$  constructs.

| H3K27ac             | mean z-score | median z-score | sd    | MW p-value |
|---------------------|--------------|----------------|-------|------------|
| Full XIST           | -2.371       | -2.302         | 1.286 |            |
| $\Delta$ FBh rep. 1 | -1.764       | 1.807          | 1.194 | 6.45E-03   |
| $\Delta$ FBh rep 2  | -1.520       | -1.573         | 1.136 | 1.06E-04   |
| $\Delta$ Bh         | -1.557       | -1.645         | 0.817 | 6.79E-05   |
| ΔΕ                  | -1.803       | -1.798         | 0.815 | 4.05E-03   |
| Δ 3'                | -2.294       | -2.321         | 1.810 | 9.25E-01   |
| HTERT-RPE1          | -1.759       | -1.661         | 0.682 | 5.81E-04   |

Table 5.7 Average H3K27ac enrichment and XIST RNA cloud

The depletion of H3K27ac at the *XIST* RNA cloud seen across each population of cells is summarized as well as the p-value indicating the statistical significance of the difference between the Full *XIST* control and the other cell lines. In all cell lines *XIST* was being expressed from chromosome 8p following 5 days of induction with doxycyline. The mean and median zscores for each population of cells are listed as well as the standard deviation (sd). The statistical significance of the difference between each construct and the Full *XIST* population of cells was calculated using the Mann Whitney U test. Two replicates (rep 1 & 2) of  $\Delta$ FBh were included to provide insight into how consistent the testing of H3K27ac was between biological replicates in  $\Delta$  constructs.

### 5.2.9 Perinucleolar localization of XIST is highly stable and resilient

Given how little is known in both mice and humans about how XIST localizes to the perinuclear compartment it was decided to investigate the proportion of cells carrying each of the deletion constructs that associated with the perinucleolar compartment. Of all Full XIST RNA clouds examined in the 8p cell line during different immunofluorescence experiments, 54.55% (180/330) were observed to be clearly associated with the perinucleolar compartment. This represents a 21% increase compared to the 33% (33/100) observed in uninduced Full XIST cells, which statistically differed from the induced state using a Fisher exact t-test ( $p = 9.0 \times 10^{-6}$ ). Nearly all of the deletion constructs showed almost identical perinucleolar localization (51%-55% perinucleolar) to Full XIST (figure 5.9). Only  $\triangle$  3D5E (46% perinucleolar, n=300, p = 3.1x10<sup>-2</sup>) and  $\Delta\Delta$  (48% perinucleolar, n=200, p = 0.15) were observed to have less than 50% of their XIST RNA clouds at the XIST RNA cloud but this effect did not reach the threshold of statistical significance when the threshold was adjusted for multiple testing (p=0.05/12= $4.17 \times 10^{-3}$ , figure 5.9). The striking result of this analysis was how consistent the rate of perinucleolar localization was across deletion constructs, clearly demonstrating that any region of XIST can be lost without completely or even significantly disrupting its ability to relocate itself and large chromatin territories to the perinucleolar compartment.



Figure 5.9 *XIST* RNA localization to the perinucleolar compartment is highly stable, not depending upon any single region of *XIST* 

The localization of *XIST* RNA (green) to the perinucleolar compartment of the HT1080 cells was analyzed using the numerous IF-FISH images from the various enrichment tests. The combination of heterochromatin labelling (H3K27me3, SMCHD1, H4K20me1 or ubH2A) IF (red) and DAPI staining (blue) made the nucleolar compartment easily identifiable and cells were classified as either having *XIST* directly adjacent to the nucleoli or not as shown in the bottom example images. The proportion of cells (>100) for each type of construct with

perinucleolar *XIST* (blue) and non-perinucleolar (grey) were compared to induced Full *XIST* using the Fisher exact test. Uninduced Full *XIST* associated with the perinucleolar compartment less than induced Full *XIST* however none of the induced deletion constructs significantly differed from the induced Full *XIST* (\* p-value <  $4.17 \times 10^{-3}$ ). The white scale bars in each photo indicate 10µm.

## 5.3 Discussion

The regions essential for the various XIST initiated pathways examined throughout this chapter are highlighted in figure 5.10. This work provides the first analysis of the specific regions of XIST responsible for its various pathways and makes novel findings regarding each function analyzed. The 5' and 3' most extreme ends of XIST were shown to be critical for XIST to silence distal genes. Repeat A had previously been identified as essential for proximal silencing and was found to also be essential for distal gene silencing [82], [111], [288]. It was an entirely novel discovery however that the adjacent region including Repeat F was also essential for distal gene silencing. It had been proposed that in mouse models Repeat F was critical for tethering Xist to chromatin by LBR, however no observation of defective XIST RNA localization was observed here suggesting that these functions are not conserved between species [56]. The 600bp region at the 3' edge of the XIST transcript was also found to be essential for XIST to silence distal genes. The importance of the 3' region was supported by the larger deletion constructs Exon 1 and  $\Delta\Delta$ . It was intriguing to note that no other sequence in the more then 10kb intervening sequence was essential for gene silencing (section 5.2.1), especially when those intervening sequences were found to be essential for every other aspect of Xi associated chromatin remodeling. It is therefore evident that most of XIST is dispensable for it to induce gene silencing and that there was little connection between this establishment of gene silencing and any of the observed changes to the chromatin environment analyzed here.

The lack of H3K27me3 enrichment at the *XIST* RNA cloud of the  $\Delta$  FBh and  $\Delta$  Bh constructs indicates that the region surrounding Repeat F and Bh are likely critical to *XIST*-mediated activation and/or recruitment of PRC2 (section 5.2.2). Loss of Repeat A in the  $\Delta$  A construct

resulted in a small weakening of H3K27me3 enrichment compared to Full XIST, though  $\Delta$  A differed more significantly from  $\Delta$  FBh then Full XIST. The intermediate nature of this effect on H3K27me3 enrichment suggests that Repeat A influences PRC2 activity through contributory or additive effects. There were studies that suggested Xist repeat A is important for H3K27me3, and it may be that some element in the same region of human XIST may also contribute to this process through some partial conservation of function [86]. The region of human XIST responsible for H3K27me3 enrichment had not been previously known, and opposing views have been circulating for years. There was evidence in mice that the PRC2 cofactor Jarid2 bound to the region extending between mouse *Xist* Repeats F and B [131], though the role and importance of Jarid2 and this region has generally fallen out of favor in recent years [129]. In contrast to findings in mice which indicated that Repeat B to C were essential for mediating PRC1/2 activity, Repeat B and C in human XIST are completely dispensable to H3K27me3 enrichment [127], [129], [133], [140]. This demonstrates that functional divergence in the role of the Repeat B and C region of XIST has occurred between species. Another original finding of this work is that the Repeat E region of XIST is essential for H3K27me3 enrichment. This contrasts work in mice demonstrating that Xist Repeat F and E is dispensable for H327me3 enrichment [144], [289]. The two regions essential for H3K27me3 enrichment are separated by  $\sim$ 8kb of dispensable sequence, the removal of which did not obviously impact the enrichment of H3K27me3.

Based on mouse studies it was anticipated that ubH2A enrichment would depend upon similar regions of *XIST* as H3K27me3, however this was entirely the opposite to what was observed. The enrichment of ubH2A at the *XIST* RNA cloud depends upon the region surrounding Repeat

B and C of *XIST*, suggesting that the role of this region was somewhat conserved between mice and humans, despite the difference in the size of the repeats [81], [127]. The identification of the previously unassociated Repeat A, Repeat D and 3' non-repeat domain of *XIST* indicates that PRC1 utilization by *XIST* relies on numerous distinct interdependent regions. It may be of interest to examine these same regions of mouse *Xist* to determine whether they also contribute to ubH2A enrichment. No overlap at all was observed between the regions of *XIST* essential for H3K27me3 or ubH2A, suggesting that these two processes are part of independent pathways in humans, unlike in mice [127], [133], [140]. Taken together these results suggest significant functional divergence in how *XIST* mediates the activity of the PRC1 across mammalian species and that ubH2A enrichment in humans requires at least four distinct domains of *XIST*.

The role of H4K20me1 in Xi still remains to be elucidated, though for the first time in any mammalian species there is now evidence of the regions of *XIST* that are responsible for its accumulation. Enrichment of H4K20me1 requires Repeat E of *XIST* which was also identified as important H3K27me3 enrichment, however its reliance on the non-repeat region spanning Repeat C and D was completely unique to this mark. A link was once proposed between H4K20me1 enrichment and the PRC2 complex recruitment by *XIST*, and work outside of field of XCI has noticed links between the PRC activity and H4K20me1 [128], [290], [291]. It may have been that some common factor or element at Repeat E was common to both pathways, but further work would be needed to identify this element in Repeat E. As no overlap between the domains of *XIST* essential for ubH2A and H4K20me1 was identified, it was concluded that the pathways leading to the enrichment of these marks are entirely independent of each other.

As so little was known of MacroH2A recruitment by *XIST* in any species, the findings presented here represent the first and only insights to date into the regions and domains of *XIST* that mediate its recruitment [162]. MacroH2A localization relied on regions surrounding both Repeat F and E as well as the large internal region of exon 1 extending from the non-repeat region 3' of Repeat C to the 3' edge of Repeat D. This was interesting as it indicates both a potential link to H3K27me3 and also indicates for the first time that Repeat D is associated with any aspect of chromatin remodeling in any species [92]. The large middle essential region was completely dispensable for normal H3K27me3 enrichment. Further work is still required to determine whether a potential connection between PRC2 and MacroH2A exists. The observation that MacroH2A enrichment relies on the same domains identified as essential for H3K27me3 enrichment suggests a potential link between these factors that was investigated in subsequent studies in Chapter 6.

The regions of *XIST* essential for the other late Xi factor, SMCHD1, were also identified in this work for the first time. A significant overlap of the domains required for its enrichment with ubH2A was observed, with both regions requiring Repeat B, C and D as well as the 3' region of *XIST*. Enrichment of SMCHD1 also requires the large non-repeat regions spanning between Repeat D and E, suggesting that additional factors are involved in its recruitment. The overlap in the regions responsible for SMCHD1 enrichment and ubH2A suggests a potential link between the two in keeping with evidence of PRC1 recruitment being essential for SMCHD1 enrichment in mice [132]. This connection however is not perfectly clear, as the  $\Delta$  A construct failed to become enriched for ubH2A but recruited SMCHD1 strongly. Further analysis to better ascertain

what link existed between ubH2A and SMCHD1 is therefore seen as a potentially fruitful avenue of research and also explored further in Chapter 6.

The recruitment of CIZ1 is believed to occur through the direct binding of *XIST*, so it was therefore surprising to observe that CIZ1 also requires distinct regions of *XIST* to become enriched [107], [108]. The Repeat E region had been previously suggested to be important based on studies in mice, however the fact that the non-repeat region directly 3' of Repeat Bh is also essential is entirely novel. The region 5' of Bh was not identified as essential to any other feature tested, suggesting that CIZ1 enrichment is not directly linked to any of these other pathways [108]. All the marks and factors described so far have relied on numerous regions of *XIST*, however CIZ1 requiring multiple distinct domains is perhaps the most surprising as it is believed to directly bind to *XIST* without intermediates or additional factors via its C-terminal [107]. This suggests that long-range interactions between these regions of *XIST* are necessary to create the secondary structures necessary for CIZ1 binding.

There has been growing evidence for mouse *Xist* that the majority of gene silencing events are mediated by histone deacetylation, and that this may contribute to H3K27me3 and ubH2A spreading [114]. The tests of histone deacetylation performed in this chapter provide an initial insight of where fruitful avenues of research into the pathways initiated by human *XIST* might lie. Loss of both Repeat F and Bh completely prevents H3K27me3 enrichment and was observed to result in a mild effect on histone hypoacetylation of H3K27. Loss of the 3' region of *XIST* completely disrupted gene silencing and ubHA enrichment but had no effect on H3K27 hypoacetylation. This brief analysis suggests that some connection might exist between hypoacetylation and trimethylation of H3K27 by *XIST*, and that hypoacetylation does not show

an obvious connection to *XIST*-mediated distal gene silencing or ubH2A enrichment. Further research is necessary to ultimately identify how *XIST* utilizes PRC1, PRC2 and histone deacetylases when establishing chromosome inactivation, but this work provides an initial framework to direct the next stage of research into *XIST* activity.

An overarching observation from this work is that numerous distinct domains of *XIST* are critical for the various processes examined. The various essential domains identified for each process are separated often by kilobases of dispensable sequences that could be removed without affecting the given pathway. The evidence that multiple sections of *XIST* are essential for each pathway it initiated offers a potential reconciliation of studies that disagreed about which regions of *XIST* were essential for certain functions, best exemplified by the various regions associated with PRC2 recruitment [86], [127], [131]. It may be that these reports have also identified numerous interdependent sequences within *Xist*. In addition, it was shown that despite the initial expectation that the Repeats of *XIST* would be of primary importance, most of the pathways examined rely on at least one non-repeat region of *XIST* [82]. These results also demonstrate that while the functional importance of certain regions is conserved between mice and humans, significant divergence has occurred as well.



The regions of *XIST* identified as important for each of the processes examined in this chapter are highlighted in each row. Darker orange highlighting indicated regions essential for a process while lighter yellow highlighting indicated regions that affected a function but were not essential. The regions tested but found to be dispensable are not highlighted and are represented by the faint outline of *XIST* in each row. The *XIST* transcript is coloured to delineate the regions typically referred to as Repeat (blue) or non-repeat (grey) domains. Regions left entirely blank in this figure have not been tested.

# Chapter 6: Chemical Inhibition of Key XIST Binding Factors

# 6.1 Introduction

The result of investigating the *XIST* deletion constructs in chapter 5 suggested several potential connections between Xi associated factors. The next objective was to investigate potential connections and interdependencies to identify pathways of *XIST*-mediated chromosome inactivation. Based on the results of chapter 5 as well as the contemporary knowledge of the field, three *XIST*-mediated chromatin modifications were examined: HDAC histone deacetylation, PRC2 trimethylation of H3K27 and PRC1 ubiquitination of H2AK119. These pathways were examined for their role in *XIST*-mediated silencing of distal genes and other chromatin remodeling events of interest.

Of all the complex aspects of *XIST* activity, its establishment of gene silencing was the most studied and yet also perhaps the least known. The essential nature of Repeat A for silencing was established early, with more recent studies showing that SPEN binding to *Xist* was also critical for silencing, but it remains still uncertain how the initial down-regulation and silencing of genes occurs [57], [59], [82]. Studies of mouse *Xist* had suggested that the mediation of histone hypoacetylation through HDAC3 activation by SPEN was critical for the silencing of most genes on the Xi and contributed to H3K27me3 enrichment [103], [114]. To validate whether a similar process occurred in humans and to establish importance of HDAC activity in general to XCI it was necessary to test both HDAC3 specifically as well as HDACs in general. It was hypothesized that hypoacetylation was not critical for *XIST* to establish silencing, based on the

analysis of the  $\Delta$  FBh and  $\Delta$  3' *XIST* constructs, which failed to repress genes despite demonstrating clear hypoacetylation at the *XIST* RNA cloud (figure 5.1 and figure 5.8).

At time of writing the most recent analysis of mouse *Xist* regulation of HDAC3 and PRC 1 & 2 suggested a hierarchical system where histone deacetylation preceded and cleared the way for PRC1 to spread its repressive mark, ubH2A, which in turn recruited PRC2 [114], [131]. There was little reason to suspect a link between PRC1 and HDAC3 activity in the human model based on the results of the deletion constructs in chapter 5 (figure 5.3 and 5.8) The tests of the deletion constructs in chapter 5 had suggested a potential link might exist however between histone deacetylation and PRC2-mediated enrichment of H3K27me3 at the *XIST* RNA cloud (figure 5.8). This potential link was intriguing as it suggested that an analysis comparing the histone deacetylation and H3K27me3 enrichment by *XIST* would reveal the first clear insight into how the XCI pathways were ordered in humans.

While studying the importance of deacetylation was of great interest, another primary focus was to determine the significance and interplay of the highly studied PRCs 1 and 2 by *XIST* during the establishment of inhibited chromatin. Studies of mouse models had suggested that *Xist* binding to hnRNPK mediated the initial recruitment of non-canonical PRC1, ultimately leading to the chromatin bound by *Xist* to be enriched for ubH2A histone marks [127], [140]. The evidence from mouse models suggested that *Xist*-mediated H3K27me3 enrichment depended upon preceding PRC1 activity, however the results of chapter 5 strongly suggested that PRC1 and PRC2 recruitment was independent in humans (figure 5.2 and figure 5.3)[129], [133]. Research using mouse *Xist* had also suggested a potential connection between PRC1 activity and 211

SMCHD1 enrichment, and noticeable overlaps in the regions essential for both ubH2A and SMCHD1 enrichment had been observed in chapter 5 (figure 5.3 and figure 5.6)[167]. The studies of the deletion constructs suggested a connection exists between PRC1 and SMCHD1 and determining whether this connection existed one way or another would open the door to further understanding of the *XIST*-mediated pathways as well as the possibility of cross-species comparative analysis. For all these listed reasons it was concluded that analyzing the importance of the catalytic activity of PRC1 to the various *XIST*-mediated pathways would provide numerous benefits and insights.

The results of chapter 5 suggested that the mechanisms by which PRC2 was recruited by *XIST* have diverged significantly between mice and humans, and it was therefore of great interest to determine how our own species' *XIST* utilizes this complex. The enrichment of H3K27me3 at the *XIST* RNA chromatin had been associated with MacroH2A enrichment in chapter 5 based on the reliance on overlapping regions of the *XIST* transcript. No evidence from the literature suggested a connection between these factors so investigation of a connection represented completely uncharted territory. Analysis of the deletion constructs revealed the regions of *XIST* responsible for H3K27me3 enrichment seemed to contribute to histone hypoacetylation. This suggested that testing whether the HDAC and PRC2 processes were reciprocally interacting was a potentially fruitful avenue to discover novel aspects of the *XIST*-mediated pathways. Both mouse studies and the results of chapter 5 suggested that PRC2 activity was not essential for the enrichment of ubH2A, however there was evidence of collaborative heterochromatin spreading of ubH2A and H3K27me3 marks through feed-forward mechanisms [139], [144]. It was

therefore tested whether PRC2 activity contributes even additively to the spreading and accumulation of ubH2A.

As described in chapter 3, the accumulation of heterochromatin marks and the depletion of H3K27ac by *XIST* induction from 8p developed over several days and required 5 days to reach its greatest extent. It was believed that a sufficient impediment to catalytic activity would allow the pathways to be tested without the need for completely blocking all catalytic activity. A modular system of inhibiting catalytic activity was seen as preferable to attempting to knock-out or knock-down factors, as the focus of these experiments was to directly examine the effect of alterations to the histone code on downstream processes. Small molecule inhibitors specific to PRC1, PRC2, HDAC3 as well as a broad inhibitor of HDACs were all used to temporarily block catalytic activity in the Full *XIST* 8p HT1080 cells while *XIST* was being induced. The chemical inhibitors selected for this research, a review of the relevant literature and dosage information is described in sections 6.1.1-6.1.4.

The inhibitor treated cells were compared to the uninhibited control HT1080 cells to determine what effect, if any, a given factor has on *XIST* silencing, expression and chromatin remodeling. To determine whether *XIST* would be induced at normal levels in the presence of a given inhibitor, the XIST transcript levels were tested using qPCRs according to the protocol in section 2.2 and 2.3, with *XIST* vector primers (PCDNA5) used to measure induced *XIST* levels relative to the endogenous control genes *PGK1* and *UBC* (see appendix primer table A.2). The normalized silencing strength of *XIST* in the presence of each inhibitor was tested by PCR amplification of cDNA for the SNP encoding segment of the four distal genes *CTSB*, *DLC1*, *SLC25A37* and *STC1* and then performing Pyrosequencing as described in section 2.10. The

distribution of chromatin factors at the *XIST* RNA cloud was measured by comparing the z-score of a population of IF-FISH labelled cells following 5 days of treatment as described in the methods sections 2.12 and 2.13.2. 60-61 cells were tested for each condition and the population distributions were all compared to the control (uninhibited) population of cells using the Mann Whitney U test [249]. As 13 tests for statistical significance were made throughout this work the thresholds of significance were adjusted from 0.05, 0.01 and 0.001 to  $3.85 \times 10^{-3}$ ,  $7.7 \times 10^{-4}$  and  $7.7 \times 10^{-5}$  respectively.

## 6.1.1 General HDAC inhibition with Trichostatin A (TSA)

Trichostatin A (TSA) was described in 1976 as an antifungal antibiotic produced by the bacteria *S. hygroscopicus* [292]. TSA is a broad spectrum inhibitor of histone deacetylase activity known for decades to cause hyperacetylation of histone tails at doses as low as 10nM to 20nM [293], [294]. TSA is a reversible inhibitor of class I/II HDACs, including HDAC3, and one of the most common tests of its activity is through measuring the hyperacetylation of H4 by western blotting or ChIP [294], [295]. Identifying the appropriate dose of TSA however proved challenging, as despite its endemic use in research the doses used by researchers ranged from 6.25nM to >1 $\mu$ M depending on the cell type and the length of treatment [296], [297]. One research group however demonstrated convincingly by western blotting that 30nM and 50nM of TSA in HeLa cells was sufficient to produce hyperacetylation on histone H3 and H4 after 48 and 72 hours of treatment [298]. As all of the other investigations of the inducible HT1080 *XIST* system had been performed following five days of induction, the initial objective was to determine the upper doses of TSA that the cells could withstand throughout that period of time.

#### 6.1.2 HDAC3 specific inhibition with RGFP966

For the past few years evidence in mouse models has indicated that HDAC3 is essential for Xistmediated silencing [57], [59], [103]. This was further supported by knockout experiments as well as the chemical inhibition of HDAC3 catalytic activity using the small molecule inhibitor, RGFP966 [114]. RGFP966 has been used in both in vivo and in vitro models to inhibit HDAC3 activity specifically. It is reported that RGFP966 can specifically inhibit HDAC3 with no inhibitory effect of other HDACs [299]. While RGFP966 has not been studied as extensively as a HDAC inhibitor as TSA, researchers have observed a broad range of effective concentrations in vitro. Indirect evidence has shown that as little as 100nM of RGFP966 can imitate aspects of Hdac3 knockout in mouse Eµ-Myc lymphoma and APL cells and additionally recommended not exceeding 1µM in vitro due to significant apoptosis [300]. Other work has suggested that short (1 hour) exposure to up to 10µM may be optimal for HDAC3 inhibition [301]. Given this broad range of potential doses an initial analysis of cell survival at various doses of RGFP966 was performed. The treatment concentration that did not cause significant stress or mortality to the cells and caused a measurable increase in histone acetylation at known targets of HDAC3 would be used to investigate the role of HDAC3 in human XIST-mediated XCI.

# 6.1.3 PRC1 inhibition with PRT4165

To investigate the importance of PRC1 monoubiquitinating the histone tails of H2A at the site of the *XIST* RNA cloud it was necessary to find a chemical that would effectively and specifically

block BMI1/RING1A and RNF2 activity in the PRC1 complex. In 2009 a research group investigating ubiquitination and antitumor molecular targets identified a compound PRT4165 (2-Pyridin-3-yl methylene-indan-1,3-dione) that specifically inhibited the ubiquitination in a Bmi1/Ring1a dependent manner in a cell free setting. They observed a significant decrease in ubiquitination by western blotting and homogenous time resolved FRET (fluorescence recovery after photobleaching) using PRT4165, with high doses resulting in less ubiquitination [302]. The researchers also observed that treatment of HeLa cells and A375 cells with 50µM PRT4165 resulted in a significant *in vitro* inhibition of ubiquitination after 5 hours both by IF imaging and western blotting, suggesting that 50µM may be an effective treatment concentration [302]. In 2013 independent researchers published further results confirming the activity of PRT4165 as a potent and specific inhibitor of PRC1 ubiquitination [303]. Researchers demonstrated through western blotting that treatments of 25µM and 50µM and 100µM PRT4165 in cell media resulted in a significant and rapid depletion of ubiquitination at H2AK119 in the U-2 OS osteocarcinoma human cell line [303]. These researchers tested one-hour treatments with PRT4165 and observed significant loss of ubH2A, though the short duration of these two tests left uncertainty about the effect incubating cells for long durations might have to their viability.

# 6.1.4 PRC2 inhibition with GSK343

To investigate the role of PRC2 activity at the site of the *XIST* RNA cloud the small molecule inhibitor GSK343 was selected for its specificity and potency. A research group investigating small inhibitors of EZH2 activity discovered GSK343 and published their results in 2012 [304]. They identified GSK343 to be a highly potent and specific *in vitro* inhibitor of EZH2 activity.

GSK343 was designed to be competitive with SAM binding by EZH2, effectively blocking its enzymatic activity. The researchers confirmed that this competitiveness with SAM was specific to EZH2 and to a much lesser extent EZH1. All other methyltransferases tested were at least 15,000 times less sensitive to GSK343 then EZH2. The researchers observed that treating HCC1806 human acantholytic squamous cell carcinoma cells with 4.2µM resulted in a significant decrease of H3K27me3 in the nuclei of cells after 72 hours [304]. Subsequent work looking at GSK343 concentrations in HeLa cells and SiHa cells observed that 5µM and 10µM consistently resulted in a significant and specific loss of H3K27me3 in a dose dependent manner after 24 hours when measured using western blotting [305]. When this research group measured cell viability, they observed above 50% viability of both Hela and SiHa cells after 3 days at doses of GSK343 of 10µM or lower. The higher dose they tested, 20µM, resulted in most cells being dead after 3 days of treatment. The dose information obtained from this research became the basis for my own experiments using 5µM and 10µM concentrations of GSK343 to inhibit the enrichment of H3K27me3 at the *XIST* RNA cloud.

# 6.2 Results

# 6.2.1 Validating Chemical Inhibitors in HT1080 inducible XIST model

# 6.2.1.1 Survival and stress during chemical inhibition treatments

The upper concentration of HDAC inhibitors TSA and RGFP966 that the HT1080 cells could endure was determined as a first step to identify the suitable concentrations for testing the *XIST* 

initiated chromatin inactivation pathways. Treatment of cells with the HDAC inhibitors was assessed using a hemocytometer with Trypan blue treatment over the course of several days. The cells were all grown in the wells of a 6 well plate according to the cell culture protocol described in section 2.1 of the methods chapter. A control population of HT1080s was grown along with the TSA treated and RGFP966 treated cells during the course of treatment. The quantified survival is presented in figure 6.1 and the qualitative assessments of cells stress based on the shape of the HT1080 cells is recorded in the following paragraphs.

Concentrations of 200nM and 300nM TSA were initially tested based on the observations of Dr. Jakub Minks however these concentrations resulted in complete HT1080 cell death after 3 days. It was decided to test a range of lower concentrations that would go up to 150nM of TSA treatment. The lowest concentration of TSA tested was 20nM, as this fell between the concentrations of 20nM to 30nM identified as optimal in other cell lines [296], [298]. 20nM increments of TSA were tested up to a 100nM concentration, as well as a 150nM concentration, which were all used to treat HT1080 cells over 6 days of treatment. The initial plan was to begin treating cells with the chemical inhibitors 1 day before starting XIST induction with doxycycline, so 6 days chemical inhibition would correspond to the established 5ddox treatment condition. After 1 day of treatment all the cells in the different concentrations of TSA, with the exception of the 150nM concentration, showed no obvious signs of stress and the majority of cells measured on the hemocytometer being alive (figure 6.1 A). The cells in the 150nM TSA treatment appeared highly stressed and only 64% were still alive. By day 2 of TSA treatment the 80nM, 100nM and 150nM TSA conditions all appeared noticeably unhealthy, with both significant signs of cells stress and cell death evident. As TSA was known to be a potent inhibitor of cell

growth, it was not surprising that even the 20nM treatments had visibly lower cell confluency after 2 days of TSA treatment compared to the control population of HT1080 cells [306]. By 4 days of treatment there were hardly any cells alive in the concentrations TSA greater than 80nM, though most cells in 60nM were not showing major signs of stress or death (figure 6.1 A). By 6 days of TSA treatment all the concentrations of TSA with the exception of the 20nM concentration had at least 50% cell death (figure 6.1 A). Complete cell death occurred after 6 days of treatment in concentrations at or above 80nM of TSA suggesting that the HT1080 cells were sensitive to the long-term effects of HDAC inhibition.

The results of testing the TSA treatments suggested that fewer time points would be sufficient to observe the effects of RGFP966 on cell survival and stress and so tests of survival following 2 days of treatment and 6 days of treatment were performed. A range of concentrations of RGFP966 from 100nM to 10µM were tested based on the literature results described in section 6.1.2. By two days of treatment a general negative relationship between HT1080 survival and RGFP966 concentration was observed (figure 6.1 B). All the concentrations less than or equal to 1µM RGFP966 did not differ qualitatively from the control HT1080s grown in parallel. Treatments of 2µM and higher of RGFP966 produced obvious signs of stress on the cells within 2 days. By 6 days most of the cells in concentrations greater than 2uM were dead and the remaining cells looked consistently stressed. Concentrations of 1µM RGFP966 were comparable with the lower concentrations and the untreated control (figure 6.1 B).

Based on the results of these two HDAC inhibitors survival rates, the timetable for treating cells was revised to minimize the time the cells would spend in the inhibitors as this unusually long term treatment had proven unexpectedly stressful to cells. The Full *XIST* HT1080 cells to be

treated with a chemical inhibitor would only be treated for five days, with doxycycline added at the same time as the inhibitor where appropriate. As TSA treatment produced the greatest stress response in HT1080s, three concentrations of 20nM, 40nM and 60nM were used when assessing whether broad spectrum HDAC inhibition impacted *XIST* activity. The concentration of RGFP966 selected was 1µM as it was the highest concentration that did not dramatically impact survival.

In addition to testing the lethality of the HDAC inhibitors, it was necessary to identify viable upper concentrations of the PRC2 inhibitor GSK343 and PRC1 inhibitor PRT4165. A 5 day analysis of cell survival on the HT1080 cells using GSK343 or PRT4165 was performed using concentrations found to be effective in similar contexts [303], [305]. The higher concentrations used for each inhibitor were 10µM for GSK343 and 50µM for PRT4165 and both resulted in moderate signs of stress in a minor proportion of HT1080s. The second highest concentrations tested were 5µM GSK343 and 25µM PRT4165, and these two concentrations were functionally indistinguishable from the control untreated HT1080 cells grown in parallel. Lower concentrations of GSK343 and PRT4165 were tested but as they were not differentiable from the 5µM and 25µM concentrations respectively they were not used in subsequent analysis. The treatment plan that became standard practice when testing the effects of the chemical inhibitors on the Full *XIST* 8p HT1080s was summarized figure 6.2. A summary of the inhibitors used, their targets and the doses selected for testing the Full *XIST* 8p inducible cell lines is listed in table 6.1.



В Percent survival of HT1080 cells treated with RGFP966 RGFP966 (nM) 0 1.00-100 Ratio of cells alive 200 0.75 500 1,000 0.50 2,000 5,000 0.25 10,000 0.00 20245 6 days Time



The proportion of cells alive in different concentrations of the HDAC inhibitors TSA and RGFP966 over different treatment time points shown in the graphs with increasing colour intensity reflecting increasing doses. HT1080 survival rates were calculated using a hemocytometer with Trypan blue. A) The average proportion of cells alive in in various concentrations of broad HDAC inhibitor TSA (blue shading) as well as the uninhibited control

А

(white) were tested following 1, 2, 4 and 6 days of treatment. B) The average proportion of cells alive in various concentrations of HDAC3 inhibitor RGFP966 as well as the uninhibited control were tested following 2 days of inhibition and 6 days of inhibition.



Figure 6.2 Treatment plan for HT1080 cells with chemical inhibitors

An outline of how HT1080 cells with 8p inducible Full *XIST* were tested for the effect of an inhibitor. Doxycycline (blue) and the inhibitor to be tested (green) would be added to the DMEM (pink) of cells, with at least two different concentrations tested for each inhibitor. Inhibitor free media with doxycycline was also prepared to be a control that each inhibitor would be compared to. At least 3 biological replicates were tested for each combination of inhibitor concentrations as well as the control condition. All the treatments of biological replicates and controls were performed in parallel. The inhibitor treatments and induction of *XIST* lasted for 5 days before the cells were harvested and tested.

| Inhibitor      | Target of inhibitor | Effect on histone code                | Dose for HT1080 |
|----------------|---------------------|---------------------------------------|-----------------|
| Trichostatin A | HDAC                | ↑ H3K27ac & H4K8ac (in                | 20, 40 & 60 nM  |
|                |                     | · · · · · · · · · · · · · · · · · · · |                 |
| (TSA)          |                     | addition to others)                   |                 |
| RGFP966        | HDAC3               | ↑ H3K27ac & H4K8ac                    | 1µM             |
|                |                     |                                       |                 |
| PRT4165        | PRC1 (RING1A/RNF)   | ↓ H2AK119ub (ubH2A)                   | 25 & 50 µM      |
|                |                     |                                       |                 |
| GSK343         | PRC2 (EZH2)         | ↓ H3K27me3                            | 5 & 10 µM       |
|                |                     |                                       |                 |

 Table 6.1 Inhibitors and concentrations effective for HT1080s

Summary of the inhibitors used throughout this chapter. The target of each inhibitor and the direct effect of inhibition on the global histone code are shown as well as the concentrations of each inhibitor found to be effective in the HT1080 cell line.

# 6.2.1.2 Western blotting to determine effectiveness of chemical inhibitors

To quantify the effect of the various chemical inhibitors on HT1080 cells protein extraction, western blotting and quantification were performed as described in section 2.7 of the methods. Cells were incubated for 5 days in DMEM containing the appropriate concentration of chemical inhibitor and media was replaced daily. H4 and  $\beta$ actin control antibodies were used to as references to compare the relative amount of the proteins of interest in each condition. The results of these western blot tests are shown in figure 6.3. Treatment of HT1080 cells with three concentrations of PRT4165 for five days resulted in a dose dependent decrease in ubH2A levels within cells (figure 6.3 A). The 25µM treatment resulted in only 0.68 times as much ubH2A in the nucleus, while 50µM resulting in only 0.55 as much ubH2A compared to the two control lanes on either side of the gel (figure 6.3 A). It was expected therefore that this reduced PRC1 activity would delay the accumulation of ubH2A at the *XIST* RNA cloud.

To assess histone acetylation, antibodies targeting H4K8ac were used, which had been validated and previously found to work well during western blotting by the Howe lab at UBC. H4K8ac was a target of HDAC3, and previous research groups had observed increased levels upon both general and HDAC3 specific inhibition with RGFP966 [306]. The TSA treatments resulted in a consistent and dose dependent increase in acetylation of H4K8, with 60nM resulting in a 4.16fold increase in acetylation (figure 6.3 B). Five days of treatment with 1µM RGFP966 resulted in a 2.88-fold increase in H4K8ac. In chapter 5 a potential link was suggested between *XIST*mediated histone deacetylation and H3K27me3 enrichment which was to be explored further in this analysis. Therefore, the effect of PRC2 inhibition on histone acetylation levels was also tested. Both 5µM and 10µM concentrations of GSK343 were very similar to the average level of H4K8ac observed in the untreated replicates (normalized H4K8ac = 1.09 and 0.99, figure 6.3 B). This indicated that while TSA and RGFP966 treatments dramatically affected H3K8 acetylation levels, GSK343 treatment did not.

As anticipated GSK343 treatment resulted in a dose-dependent decrease in H3K27me3. Treatment with 5 $\mu$ M reduced H3K27me3 levels to 0.25 relative to the untreated controls and the 10 $\mu$ M treatment had 0.12 the level of H3K27me3 compared to controls (figure 6.3 C). The highest TSA concentration, 60nM, reduced H3K27me3 levels to 0.41 of control levels, though 40nM and 20nM were similar to the controls (figure 6.3). RGFP966 treatment decreased H3K27me3 levels to 0.43 the average level of the controls, a similar effect strength to that observed by the 60nM TSA treatment (figure 6.3 C). Finally, since PRC1 and PRC2 activity was regularly described as linked in mouse models, H3K27me3 levels were additionally tested following treatment with PRT4165, as an effect would confound further investigations [307]. PRT4165 at 50 $\mu$ M had levels of H3K27me3 consistent with the control treatments (normalized H3K27me3 = 1.09), suggesting that PRC2 activity was not broadly affected by PRC1 inhibition (figure 6.3 C).



Figure 6.3 Chemical inhibitors produced expected effects on global histone mark levels in HT1080 cells

Western blotting of global histone modification levels following chemical inhibition treatment for 5 days in the HT1080 cell line. The relative level of the histone marks of interest (ubH2A, H4K8ac and H3K27me3) in the inhibited cell lines relative to the uninhibited control (ctrl) are listed underneath each relevant band. The label next to each row of bands denotes the antibody used (see appendix table A.3). The relative levels of histone mark within a cell line were calculated based on an endogenous control gene (histone H4 or βactin). Inhibited samples are labelled according to the inhibitor and concentration they were treated with. Imaging of the western blots as well as calculations of relative amounts of the histone marks was performed using a LI-COR Odyssey machine from BioAgilytix with the associated software. A) control (ctrl) and PRT4165 treated samples were tested for global ubH2A levels. B) The effect of both HDAC inhibitors as well as PRC2 inhibitor GSK343 were tested on global H4K8ac levels. C) The effect of all inhibitors was tested on global H3K27me3 levels.

## 6.2.2 Chemical inhibitor effects on XIST induced transcript levels

Prior to testing the effects of inhibiting HDACs, PRC1 and PRC2 on XIST induced from chromosome 8p, it was important to check that XIST was still being induced properly in the presence of these chemical inhibitors. This was particularly important in the higher concentrations of TSA, GSK343 and PRT4165 where signs of cell stress had been clearly observed. In addition to the induced control, an uninduced (no dox) Full XIST cell line was also included as a negative control to indicate the uninduced levels of XIST. Three independent biological replicates of Full XIST cells were tested for each inhibition condition as well as the no dox negative control. Four independent biological replicates of the 5ddox Full XIST controls were used as references and the RQ values for each replicate were normalized to the average of these 4 controls. The relative expression for each set of treatments was presented in figure 6.4. There was a broad variability among the Full XIST constructs, with the standard deviation for XIST being 0.382. The PRT4165 and TSA treatments did not seem to dramatically affect XIST expression (p > 0.05). The two GSK treatments both resulted in dose dependent decreases in XIST expression average ( $5\mu M Rq = 0.366$ ,  $10\mu M Rq = 0.195$ ) that were significantly different from the control induced Full XIST using a two-tailed t-test ( $5\mu M p = 3.9 \times 10^{-2}$ ,  $10\mu M p =$  $1.6 \times 10^{-2}$ , figure 6.4). These results suggest that while *XIST* could be normally induced even when HDACs and PRC1 were being inhibited, the inhibition of PRC2 catalytic activity affected XIST induction through some unknown process. The level of XIST transcripts in the 5µM treatment group were comparable to the levels observed in the  $\Delta$  3' constructs from chapter 5, which still effectively reorganized the chromatin domain, depleted active histone marks led to enrichment with numerous heterochromatin marks. This indicated that the lower XIST levels in the 5µM
treatments of GSK343 were not necessarily expected to confound this analysis of the role of PRC2 in XIST mediated chromatin inactivation.



Condition (5 days of treatment and dox induction)

#### Figure 6.4 Induced XIST transcript levels effected by PRC2 inhibition

The relative expression (RQ) of *XIST* RNA in inhibitor treated Full *XIST* cells is shown and colour coded based on the inhibitor used. All replicates with the exception of the uninduced control (no dox) were grown, treated and tested in parallel following five days of inhibitor treatment and *XIST* induction (5ddox). The average level of *XIST* relative to the endogenous control genes (*PGK1* and *UBC*) for each biological replicate is denoted by the dots in each category. HDAC inhibitor TSA (blue), PRC2 inhibitor (red) and PRC1 inhibitor (green) were all compared to the uninhibited (ctrl) population of cells. The black bars represent the mean *XIST* transcript level normalized to the average Full *XIST* level, with error bars indicating standard deviation. Statistical significance was calculated using a unpaired t-test (\* p-value < 0.05).

| Inhibitor | $\Delta CT$ | RQ       | Normalized | Average RQ | sd       | p-value |
|-----------|-------------|----------|------------|------------|----------|---------|
|           | 1.89797     | 0.268321 | 0.728826   |            | 0.382349 |         |
| Control   | 1.333063    | 0.396925 | 1.078146   | 1          |          |         |
| Control   | 0.86379     | 0.549507 | 1.492599   | 1          |          |         |
|           | 1.955305    | 0.257866 | 0.700429   |            |          |         |
|           | 0.774977    | 0.584398 | 1.587371   |            |          |         |
| TSA 40nM  | 0.862982    | 0.549815 | 1.493435   | 1.365769   | 0.306099 | 0.225   |
|           | 1.418003    | 0.37423  | 1.016502   |            |          |         |
|           | 1.768236    | 0.293568 | 0.797403   |            |          |         |
| TSA 60nM  | 1.79459     | 0.288254 | 0.782969   | 0.813571   | 0.041142 | 0.436   |
|           | 1.658635    | 0.316739 | 0.860341   |            |          |         |
|           | 3.065144    | 0.119481 | 0.324541   |            |          |         |
| GSK 5uM   | 2.975536    | 0.127138 | 0.345338   | 0.36591    | 0.054641 | 0.0387  |
|           | 2.666436    | 0.157515 | 0.427851   |            |          |         |
|           | 3.863301    | 0.068712 | 0.186638   |            |          |         |
| GSK 10uM  | 3.592845    | 0.082879 | 0.225121   | 0.195298   | 0.026573 | 0.0164  |
|           | 3.963342    | 0.064108 | 0.174135   |            |          |         |
|           | 1.984661    | 0.252672 | 0.686321   |            |          |         |
| PRT 25uM  | 1.805585    | 0.286065 | 0.777024   | 0.75902    | 0.065578 | 0.326   |
|           | 1.739019    | 0.299573 | 0.813716   |            |          |         |
| PRT 50uM  | 2.260056    | 0.208764 | 0.567055   |            |          |         |
|           | 2.500722    | 0.176688 | 0.479929   | 0.592801   | 0.127707 | 0.134   |
|           | 1.892844    | 0.269276 | 0.73142    |            |          |         |

 Table 6.2 Expression of induced Full XIST is affected by GSK343

The *XIST* transcript levels in each of the biological replicates as well as the combined average level *XIST* levels for each inhibitor treatment. The  $\Delta$  CT values and calculated RQ (2<sup>- $\Delta$ CT</sup>) values between *XIST* and the combined endogenous control genes (*PGK1* and *UBC*) for each biological replicate of the inhibition treatments are shown in the second and third columns. The level of *XIST* RNA in each replicate was normalized to the average RQ of the control replicates in the

'Normalized' column. The average *XIST* expression (RQ), standard deviation across replicates (sd) was listed for each condition. The statistical difference between the uninhibited (Control) *XIST* levels and inhibited treatments was calculated using an unpaired t-test and the statistical significance for each condition is listed in the final column. All cells tested here had undergone *XIST* induction for 5 days.

#### 6.2.3 Chemical inhibitors effects on XIST-mediated distal gene silencing

One of the primary purposes of examining the XIST inducible model described throughout this thesis was to identify which pathways interacted with or were essential for XIST establishing silencing of distal genes. The work with the deletion constructs had implicated the 5' and 3' regions of XIST as important for gene silencing, though no obvious link was apparent to any of the aspects of chromatin remodeling examined. All the chemical inhibitors described so far (TSA, RGFP966, PRT4165, GSK343) were tested at the doses described in section 6.2 for their effect on XIST-mediated gene silencing on 8p in the inducible Full XIST HT1080 cell line. Three biological independent replicates were performed for each condition as well as the control 5ddox Full XIST cells and uninduced (no dox) Full XIST. The results of the various treatments are shown in figure 6.5 where each panel shows the normalized strength of silencing for each SNP coding allele.

The normalized silencing of each biological replicate and the combined treatment levels with statistical significance are listed in table 6.3. Both GSK343 concentrations produced nearly identical effects, completely preventing XIST from silencing any of the four genes. The qPCR of Full XIST in section 6.3 suggested that GSK343 treatment inhibits XIST induced expression ( $p \le 1$  $1.0 \times 10^{-6}$ , figure 6.5). However, despite the nearly two-fold difference in XIST levels between the two GSK treatments, there was no sign of a dose dependent effect of GSK343. The two concentrations of GSK343 had nearly identical effects on silencing, with the strength of silencing for each of the four genes examined being either identical or differing by less than 0.02.

PRC1 inhibition by PRT4165 had no discernable effect on the ability of Full XIST to silence CTSB, DLC1 or SLC25A37. Both concentrations of PRT4165 had 0.92 the strength of silencing of Full *XIST* at *STC1* ( $25\mu$ M p =  $2.7x10^{-4}$ ,  $50\mu$ M p =  $2.3x10^{-4}$ ) but this difference was unlikely functional. Treatments of TSA at 40nM and 60nM had no effect on silencing across any of the genes tested, with none of the conditions meeting the lower threshold of significance (figure 6.5). The lower concentration treatments of 20nM TSA and  $1\mu$ M RGFP966 were observed to have statistically stronger *XIST*-mediated silencing at three of the genes (p <  $2.78x10^{-4}$ , figure 6.5).

The consistency of this increased silencing strength between the two treatments at the same three genes may have resulted from elevated levels of XIST transcripts, though it was surprising as HDAC3 was found to be crucial for *XIST* induced gene silencing [114]. The increased silencing raises the question of whether the uninduced allelic ratio of the genes was being affected. Treatments with 20nM TSA or 1 $\mu$ M RGFP966 were performed in parallel with either induction (5ddox) or uninduced (no dox) Full *XIST* and the allelic ratios were compared (figure 6.6). The allelic contribution for each set of conditions as well as the statistical significance for this work is shown in table 6.3. Five days of 1 $\mu$ M RGFP966 or 20nM TSA treatments did not cause a noticeable difference in the uninduced allelic contribution (figure 6.6). However, the allelic contribution following five days of *XIST* induction was significantly decreased at *CTSB*, *SLC25A37* and *STC1* between the control and both HDAC inhibition treatments (p < 3.3x10<sup>-3</sup>, figure 6.6). Taken together these results suggests that the treatment of 20nM TSA or 1 $\mu$ M RGFP966 did not have a noticeable effect on the allelic ratio of these genes generally and clearly did not weaken the ability of *XIST* to repress these genes.



Figure 6.5 PRC2 activity associated with XIST induced distal gene repression

Normalized strength of silencing for each of these four SNP containing genes was determined by measuring change in allelic contribution of the *XIST* repressed allele for each gene by

pyrosequencing. Broad HDAC inhibitor (TSA), HDAC3 specific inhibitor (RGFP966), PRC2 inhibitor (GSK343) and PRC1 inhibitor (PRT4165) were all tested for their effect on silencing. Silencing strength was normalized to the change in allelic contribution produced by Full *XIST* induction, with a value of 1 indicating equal silencing strength to Full *XIST* and a value of 0 indicating no change in allelic contribution following induction. Biological replicates of for each treatment condition are shown as coloured dots with the average silencing strength and standard deviation for each condition shown by the central line and error bars. The colour of the dots indicates the strength of silencing for the genes *CTSB* (yellow), *DLC1* (red), *SLC25A37* (blue) and *STC1* (green). Statistical differences in strength of silencing for each of the inhibition treatments were compared to the uninhibited (Ctrl) replicates using an unpaired t-test and the thresholds were adjusted due to the numerous (36) tests being performed (\* p-value <  $1.39 \times 10^{-3}$ , \*\* p-value <  $2.78 \times 10^{-4}$ , \*\*\* p-value <  $2.78 \times 10^{-5}$ ).

| Construct | CTSB    |       | DLC1     |         | <i>SLC25A37</i> |          |         | STC1  |          |         |       |          |
|-----------|---------|-------|----------|---------|-----------------|----------|---------|-------|----------|---------|-------|----------|
|           | average | sd    | p-value  | average | sd              | p-value  | average | sd    | p-value  | average | sd    | p-value  |
| Control   | 1.000   | 0.049 |          | 1.000   | 0.052           |          | 1.000   | 0.009 |          | 1.000   | 0.007 |          |
| Control   | 1.000   | 0.017 |          | 1.000   | 0.027           |          | 1.000   | 0.050 |          | 1.000   | 0.022 |          |
| TSA       |         |       |          |         |                 |          |         |       |          |         |       |          |
| 20nM      | 1.274   | 0.030 | 6.11E-06 | 0.948   | 0.088           | 2.37E-01 | 1.242   | 0.006 | 4.87E-06 | 1.260   | 0.010 | 1.93E-08 |
| TSA       |         |       |          |         |                 |          |         |       |          |         |       |          |
| 40nM      | 1.081   | 0.035 | 1.14E-02 | 0.997   | 0.042           | 9.11E-01 | 1.073   | 0.022 | 1.05E-02 | 1.032   | 0.003 | 7.18E-03 |
| TSA       |         |       |          |         |                 |          |         |       |          |         |       |          |
| 60nM      | 1.098   | 0.042 | 6.21E-03 | 1.014   | 0.075           | 7.05E-01 | 1.066   | 0.031 | 2.23E-02 | 1.027   | 0.016 | 3.49E-02 |
| RGFP966   |         |       |          |         |                 |          |         |       |          |         |       |          |
| 1uM       | 1.381   | 0.025 | 4.87E-07 | 1.063   | 0.105           | 2.11E-01 | 1.228   | 0.046 | 4.82E-05 | 1.218   | 0.008 | 5.41E-08 |
| GSK343    |         |       |          |         |                 |          |         |       |          |         |       |          |
| 5uM       | 0.160   | 0.058 | 1.71E-08 | 0.217   | 0.063           | 5.45E-08 | 0.138   | 0.135 | 9.94E-07 | 0.037   | 0.050 | 5.81E-10 |
| GSK343    |         |       |          |         |                 |          |         |       |          |         |       |          |
| 10uM      | 0.149   | 0.092 | 1.28E-07 | 0.209   | 0.107           | 5.71E-07 | 0.137   | 0.136 | 1.02E-06 | 0.037   | 0.073 | 5.47E-09 |
| PRT4165   |         |       |          |         |                 |          |         |       |          |         |       |          |
| 25uM      | 0.915   | 0.088 | 6.45E-02 | 0.927   | 0.048           | 3.78E-02 | 0.852   | 0.059 | 1.51E-03 | 0.915   | 0.025 | 2.78E-04 |
| PRT4165   |         |       |          |         |                 |          |         |       |          |         |       |          |
| 50uM      | 0.939   | 0.113 | 2.34E-01 | 0.864   | 0.060           | 3.62E-03 | 0.855   | 0.073 | 3.56E-03 | 0.924   | 0.019 | 2.34E-04 |
| No Dox    | 0.000   | 0.006 | 3.12E-10 | 0.000   | 0.014           | 8.15E-10 | 0.000   | 0.014 | 3.26E-10 | 0.000   | 0.024 | 1.18E-11 |

 Table 6.3 Silencing strength of XIST in HT1080 cells treated with chemical inhibitors

The average strength of silencing of each inhibitor relative to the control induced replicates was summarized. Average strength of silencing (average), standard deviation (sd) for each group of induced replicates at each gene are listed along with an uninduced control. Statistical differences in strength of silencing for each of the inhibition treatments were compared to the uninhibited (Ctrl) replicates using an unpaired t-test. The first 'control' were the cells described in figure 6.5 and the second those from figure 6.6.



Figure 6.6 Stronger *XIST* induced silencing observed following HDAC(3) inhibition treatment in three of four genes

The effect of TSA and RGFP966 treatments on allelic ratios with and without *XIST* induction was measured using pyrosequencing. An allelic expression of 0.5 indicates that the allele contributes to exactly half of the expression of that gene. The allelic ratio of the four genes *CTSB* (yellow), *DLC1* (red), *SLC25A37* (blue) and *STC1* (green) were compared between cells treated with inhibitors but without *XIST* induction and the uninhibited cells. The allelic contribution of

the four genes following *XIST* induction and treatment were also compared to the allelic contribution following *XIST* induction without inhibitors. Chemical inhibitors did not significantly affect the allelic ratio when *XIST* was not induced but significantly affected the final allelic ratios following 5 days of *XIST* induction for *CTSB*, *SLC25A37* and *STC1*. Statistical significance was calculated using unpaired t-test with the threshold of significance adjusted for the 16 tests being performed (\* p-value <  $3.13 \times 10^{-3}$ ).

| Construct | CTSB     |        | DLC1    |          | <i>SLC25A37</i> |         |          | STC1  |         |          |       |         |
|-----------|----------|--------|---------|----------|-----------------|---------|----------|-------|---------|----------|-------|---------|
|           | allele % | Δ      | p-value | allele % | Δ               | p-value | allele % | Δ     | p-value | allele % | Δ     | p-value |
| Control   |          |        |         |          |                 |         |          |       |         |          |       |         |
| no dox    | 51.15    |        |         | 37.35    |                 |         | 50.30    |       |         | 49.58    |       |         |
| TSA       |          |        |         |          |                 |         |          |       |         |          |       |         |
| 20nM      |          |        |         |          |                 |         |          |       |         |          |       |         |
| no dox    | 47.75    | -3.40  | 9.3E-02 | 29.52    | -7.83           | 8.8E-03 | 49.87    | -0.43 | 5.3E-01 | 52.02    | 2.43  | 1.2E-02 |
| RGFP966   |          |        |         |          |                 |         |          |       |         |          |       |         |
| 1uM       |          |        |         |          |                 |         |          |       |         |          |       |         |
| No dox    | 51.45    | 0.30   | 3.9E-01 | 36.40    | -0.95           | 4.0E-01 | 49.10    | -1.20 | 1.3E-01 | 50.87    | 1.28  | 1.1E-01 |
| Control   |          |        |         |          |                 |         |          |       |         |          |       |         |
| 5ddox     | 25.28    |        |         | 15.03    |                 |         | 30.78    |       |         | 18.30    |       |         |
| TSA       |          |        |         |          |                 |         |          |       |         |          |       |         |
| 20nM      |          |        |         |          |                 |         |          |       |         |          |       |         |
| 5ddox     | 14.78    | -10.50 | 1.5E-11 | 8.35     | -6.68           | 4.8E-03 | 25.63    | -5.15 | 8.2E-04 | 12.60    | -5.70 | 7.8E-06 |
| RGFP966   |          |        |         |          |                 |         |          |       |         |          |       |         |
| 1uM       |          |        |         |          |                 |         |          |       |         |          |       |         |
| 5ddox     | 14.83    | -10.45 | 2.7E-05 | 12.68    | -2.35           | 1.7E-01 | 25.13    | -5.65 | 1.8E-03 | 12.75    | -5.55 | 1.8E-04 |

Table 6.4 Comparison of how allelic contributions were affected by HDAC inhibition

The HDAC inhibitors that produced statistically significant stronger silencing of *XIST* were examined for confounding effects on the allelic ratio of genes of interest. The allelic ratio of four genes 8p repressed by *XIST* induction were examined among uninduced (no dox) control and inhibitor conditions (top), as well as the *XIST* induced conditions (5ddox, bottom). Both TSA and RGFP966 treatments lasted 5 days. The allele contribution to total gene expression (allele %) for each condition was compared to the control and the difference of treatment conditions to the control ( $\Delta$ ) was tested for statistical significance using an unpaired t-test (p-value). Four biological replicates of each condition were tested by pyrosequencing.

#### 6.2.4 HDAC3 inhibition specifically linked to XIST-mediated H3K27me3 enrichment

The broad HDAC inhibitor TSA and the HDAC3 specific inhibitor RGFP996 were examined to provide insights into how the distribution of acetylation at the *XIST* RNA cloud was impacted, while also testing whether there is a link to the accumulation of H3K27me3. The link between these marks was based on the results of section 5.8, which suggested that the regions of *XIST* necessary for H3K27me3 enrichment produced a less complete hypoacetylation of the region (figure 5.8). The quantification of IF-FISH labelled cells was performed using two doses of TSA, 40nM and 60nM, as well as 1 $\mu$ M dose of RGFP966. The effect of these treatments is shown in figure 6.7 and summarized in table 6.5. All three HDAC inhibition treatment conditions still produced clear signs of hypoacetylation (median z-scores between -1.71 to -1.95) however the effect of this hypoacetylation differed statistically from the control treatment (median z-score - 3.14, p < 6x10<sup>-6</sup>, figure 6.7, significance for each condition listed in table 6.5). These results suggest that all three HDAC inhibition treatments significantly attenuated the deacetylation of H3K27 at the site of the *XIST* RNA cloud following 5 days of treatment, but did not prevent it.

The effect of 1µM RGFP966 and 60nM TSA HDAC inhibition treatments were also tested on the enrichment of H3K27me3 by *XIST* RNA (figure 6.7 and summary table 6.5). Treatment with the HDAC3 specific inhibitor RGFP966 resulted in a significantly weakened enrichment of H3K27me3 (median z-score = 0.634, s.d. = 1.24, p =  $1.7x10^{-11}$ , figure 6.7). 61% of RGFP966 treated cells had z-scores of relative H3K27me3 enrichment under 1 which differed noticeably from the 81% of control cells that had a z-score greater than 1. The broad-spectrum TSA treatment however did not significantly affect H3K27me3 enrichment at the *XIST* RNA cloud compared to the control (median z-score = 3.73, s.d. = 1.35, p =  $3.2x10^{-3}$ , figure 6.7). This

indicates that HDAC3 activity is important for the enrichment of H3K27me3 but that broad inhibition of HDACs throughout the cell does not produce this same effect, potentially due to the non-specific TSA not sufficiently inhibiting HDAC3 activity specifically at these concentrations.



Figure 6.7 HDAC3 inhibition affected H3K27me3 enrichment by XIST

The relative difference of H3K27ac and H3K27me3 at the *XIST* RNA was determined by IF-FISH as a z-score for each cell, and the population of 60-61 cells analyzed for each treatment condition. The colour of each box plot denotes the treatment condition: control (grey), HDAC3 inhibitor RGFP966 (purple) and TSA (shades of blue for concentration). The left panel shows the distribution of H3K27ac and the right H3K27me3 in each condition. The central line of the boxplot indicates the median z-score, the notch the confidence interval, the box the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference from the control was calculated using Mann Whitney U test and with the threshold adjusted for multiple testing (\* p-value <  $3.85 \times 10^{-3}$ , \*\* p-value  $7.69 \times 10^{-4}$ , \*\*\* p-value  $7.69 \times 10^{-5}$ ).

| Treatment   | Mark     | mean z-score | median z-score | sd    | MW p-value |
|-------------|----------|--------------|----------------|-------|------------|
| Control     | H3K27ac  | -3.066       | -3.136         | 1.347 |            |
| TSA 40nM    | H3K27ac  | -1.741       | -1.707         | 0.999 | 7.41E-09   |
| TSA 60nM    | H3K27ac  | -1.968       | -1.945         | 0.951 | 6.07E-07   |
| RGFP966 1uM | H3K27ac  | -2.178       | -1.870         | 1.407 | 6.35E-05   |
| Control     | H3K27me3 | 2.901        | 2.593          | 1.704 |            |
| TSA 60nM    | H3K27me3 | 3.725        | 3.744          | 1.354 | 3.21E-03   |
| RGFP966 1uM | H3K27me3 | 0.749        | 0.634          | 1.245 | 1.74E-11   |

Table 6.5 Summary of effects of HDAC inhibition on chromatin remodeling by XIST

The average distribution of the histone marks H3K27ac (top) and H3K27me3 (bottom) at the *XIST* RNA cloud following HDAC inhibition treatments with TSA and RGFP966. The average (mean and median) z-score for each population of 60 cells analyzed per condition as well as the standard deviation (sd) are listed for each mark. The population of cells for each inhibitor was compared to the relevant control treatment using the Mann-Whitney U test, with the p-value listed in the final column. All tests were performed in the 8p Full *XIST* HT1080 cell line following five days of dox induction and relevant inhibitor treatment.

#### 6.2.5 PRC2 activity necessary for MacroH2A enrichment and histone hypoacetylation

To determine what aspects of *XIST* chromatin restructuring were affected by the catalytic activity of PRC2, the *XIST* surrounded chromatin environment was examined by IF-FISH following PRC2 inhibition with GSK343 for 5 days during *XIST* induction. The XIST RNA of the 10 $\mu$ M GSKS343 treatment appeared as only a point source a few pixels in diameter in the nuclei of cells due to its low level of expression, and it was feared that this would result in an unsuitably small sample size of *XIST* +ve measurements (see section 2.13.2 of methods and figure 2.4). The 5 $\mu$ M of GSK343 treatment however produced an *XIST* RNA signal functionally indistinguishable from the control Full *XIST* and suitably large to calculate z-scores for each cell. The 5 $\mu$ M concentration was used to test how disrupting PRC2 activity affected the enrichment of MacroH2A, ubH2A and H3K27me3 itself as well as depletion of H3K27ac. The effect of these treatments is shown in figure 6.8 and summarized in table 6.6.

Visual inspection of the 5 $\mu$ M cells revealed attenuated but still visible H3K27me3 within the nucleus supporting the western blotting observations that the catalytic activity of PRC2 was disrupted rather than completely disabled. Treatment of 5 $\mu$ M GSK343 for 5 days prevented H3K27me3 enrichment at the *XIST* RNA cloud, differing clearly from the untreated induced *XIST* cell line (median z-score = -0.09, s.d. = 1.48, p = 7.2x10<sup>-16</sup>, figure 6.8). Of the GSK343 inhibited cells analyzed for H3K27me3 enrichment, only 15% had z-scores greater than 1, compared to the more than 81% of control cells having z-scores greater than 1. This indicates that treatment with 5uM of GSK343 is sufficient to prevent H3K27me3 enrichment at the *XIST* RNA cloud.

Inhibition of PRC2 with GSK343 also significantly affected the depletion of H3K27ac at the *XIST* RNA cloud (median z-score = -1.87, s.d. = 0.84, p =  $6.6 \times 10^{-7}$ , figure 6.8). This suggested that while GSK343 inhibition did not completely disrupt the depletion of H3K27ac from the *XIST* RNA cloud, the strength of the depletion was attenuated. An interesting observation was that the treatment of GSK343 resulted in a nearly identical average z-score for H3K27ac (-1.875) as the RGFP966 (-1.869) and was similar to TSA treated cell populations (-1.71, -1.94, figure 6.7).

The effect of PRC2 inhibition on MacroH2A enrichment by *XIST* was of great interest to determine whether PRC2 activity was a critical part of the pathway necessary for MacoH2A enrichment. This idea had been suggested as a possibility based on the overlap in essential *XIST* domains observed in chapter 5 (figures 5.2 and 5.5). Induced *XIST* had been observed to normally have an enrichment of MacroH2A ~1.96 standard deviations above the nuclear average, however GSK343 treatment completely disrupted that accumulation of MacroH2A at the *XIST* RNA cloud (median z-score = 0.22, s.d. = 1.23, p =  $1.3 \times 10^{-10}$ , figure 6.8). This indicated that just as H3K27me3 failed to become enriched at the *XIST* RNA cloud following inhibition of PRC2 activity, so did the accumulation of the histone variant macroH2A. This supports the novel discovery of this work, that PRC2 activity mediated by *XIST* is essential for the enrichment of the late-recruited heterochromatin factor, macroH2A.

Finally, though evidence from the deletion constructs suggested that PRC1 and PRC2 were not interdependent (figures 5.2 and 5.3) it was still of interest to determine whether PRC2 activity affected the enrichment of ubH2A at the *XIST* RNA cloud even through cooperative or additive mechanisms. PRC2 inhibition with GSK343 however produced no significant effect on the

enrichment of ubH2A mediated by *XIST* (median z-score = 3.05, s.d. = 5.75, p = 0.22, figure 6.8). As the GSK343 treatment completely disrupted both H3K27me3 enrichment and MacroH2A enrichment, the continued and seemingly unaffected enrichment of ubH2A at the *XIST* RNA cloud further indicates that the recruitment of PRC1 and PRC2 by *XIST* occur completely independently of each other.



Figure 6.8 PRC2 activity necessary for *XIST* regulation of H3K27me3, MacroH2A and H3K27ac

PRC2 inhibition with GSK343 (red) was analyzed relative to control (grey) cell lines by IF-FISH. The z-score for each cell in a population of 60-61 cells analyzed for the treatment condition and histone mark listed above each panel of the graph. The central line of the boxplot indicates the median z-score, the notch the confidence interval, the box the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference from the control was calculated using Mann Whitney U test and with the threshold adjusted for multiple testing (\* p-value <  $3.85 \times 10^{-3}$ , \*\* p-value  $7.69 \times 10^{-4}$ , \*\*\* p-value  $7.69 \times 10^{-5}$ ).

| Treatment  | Mark     | mean z-score | median z-score | sd    | MW p-value |
|------------|----------|--------------|----------------|-------|------------|
| Control    | H3K27me3 | 2.901        | 2.593          | 1.704 |            |
| GSK343 5uM | H3K27me3 | 0.162        | -0.090         | 1.479 | 7.24E-16   |
| Control    | H3K27ac  | -3.066       | -3.136         | 1.347 |            |
| GSK343 5uM | H3K27ac  | -2.031       | -1.875         | 0.844 | 6.59E-07   |
| Control    | MacroH2A | 2.249        | 1.958          | 1.469 |            |
| GSK343 5uM | MacroH2A | 0.481        | 0.224          | 1.231 | 1.30E-10   |
| Control    | UbH2A    | 4.745        | 4.178          | 3.179 |            |
| GSK343 5uM | UbH2A    | 5.041        | 3.053          | 5.754 | 2.18E-01   |

 Table 6.6 Summary of effects of PRC2 inhibition on chromatin remodeling by XIST

The effect of PRC2 inhibition with GSK343 on specific histone marks of interest (marks) at the *XIST* RNA cloud was compared to the control population. Both conditions were induced (5ddox) 8p Full *XIST* HT1080 cells and the distribution of marks was visualized by IF-FISH. The average (mean and median) z-score for each population of 60 cells analyzed per condition as well as the standard deviation (sd) are listed for each mark. The population cells for each inhibitor was compared to the relevant control treatment using the Mann-Whitney U test, with the p-value listed in the final column.

#### 6.2.6 PRC1 activity was necessary for XIST enrichment of SMCHD1 but not PRC2

As part of the goal of determining the pathways mediating chromatin remodeling during XCI, the essential nature of PRC1 activity to various aspects of *XIST*-mediated chromatin remodeling was examined. Chromatin remodeling at the *XIST* RNA cloud was examined by IF-FISH as described in sections 2.12 and 2.13.2. Induced *XIST* RNA clouds were compared to cells that had undergone PRC1 inhibition using 50µM of PRT4165, to determine what effect PRC1 inhibition had on the accumulation of ubH2A itself, the enrichment of H3K27me3 and SMCHD1 as well as the depletion of the active histone marks H3K27ac. The effects of these tests are shown in figure 6.9 and the results along with the p-values are summarized in table 6.7.

The 50 $\mu$ M concentration of PRT4165 was sufficient to prevent ubH2A enrichment at the *XIST* RNA cloud over the course of 5 days of induction and treatment (median z-score = 0.469, s.d. = 0.846, p= 5.3x10<sup>-18</sup>, figure 6.9). Nuclear levels of the IF labelled ubH2A in the nucleus of treated cells were noticeable fainter then control cells, though ubH2A was still clearly visible indicating an attenuation of activity rather than a complete loss of function. Disrupting PRC1 activity also completely disrupted the accumulation of SMCHD1 (median z-score = -0.196, s.d. = 1.22, p= 5.3x10<sup>-18</sup>, figure 6.9) compared to the control Full *XIST* treatment (median z-score = 2.70, s.d. = 2.17). This indicates that PRC1 activity is essential for *XIST* to ultimately enrich chromatin with the late heterochromatin mark SMCHD1. Enrichment of H3K27me3 however was completely unaffected by PRC1 inhibition further supporting the clear independence between the PRC1 and PRC2 pathways (median z-score = 2.48, s.d. = 1.65, p= 0.76, figure 6.9).

The depletion of H3K27ac at the *XIST* RNA cloud proved to be slightly affected by the disruption of PRC1 activity, the effect of which was weakly statistically different from the

control treatment (median z-score = -2.15, s.d. = 0.966, p=  $1.7 \times 10^{-4}$ , figure 6.9). This small weakening of H3K27 hypoacetylation at the *XIST* RNA cloud following PRC1 inhibition suggests that accumulation and spreading of ubH2A into a chromatin domain may act additively to the overall removal of acetylation marks at the *XIST* RNA cloud.



Figure 6.9 PRC1 inhibition affects SMCHD1 and ubH2A enrichment by *XIST* RNA PRC1 inhibition with PRT4165 (green) was analyzed relative to control (grey) cell lines by IF-FISH. The z-score for each cell in a population of 60-61 cells analyzed for the treatment condition and histone mark listed above each panel of the graph. The central line of the boxplot indicates the median z-score, the notch the confidence interval, the box the interquartile range and the whiskers 1.5x the interquartile range. The statistical significance of the difference from the control was calculated using Mann Whitney U test and with the threshold adjusted for multiple testing (\* p-value <  $3.85 \times 10-3$ , \*\* p-value  $7.69 \times 10^{-4}$ , \*\*\* p-value  $7.69 \times 10^{-5}$ ).

| Treatment    | Mark     | mean z-score | median z-score | sd    | MW p-value |
|--------------|----------|--------------|----------------|-------|------------|
| Control      | H3K27me3 | 2.901        | 2.593          | 1.704 |            |
| PRT4165 50uM | H3K27me3 | 2.832        | 2.483          | 1.648 | 7.59E-01   |
| Control      | H3K27ac  | -3.066       | -3.136         | 1.347 |            |
| PRT4165 50uM | H3K27ac  | -2.370       | -2.151         | 0.966 | 1.73E-04   |
| Control      | SMCHD1   | 3.137        | 2.704          | 2.167 |            |
| PRT4165 50uM | SMCHD1   | -0.086       | -0.196         | 1.215 | 5.28E-18   |
| Control      | UbH2A    | 4.745        | 4.178          | 3.179 |            |
| PRT4165 50uM | UbH2A    | 0.571        | 0.469          | 0.846 | 5.28E-18   |

 Table 6.7 Summary of effects of PRC1 inhibition on chromatin remodeling by XIST

The effect of PRC1 inhibition with PRT4165 on specific histone marks of interest (marks) at the *XIST* RNA cloud was compared to the control population. In all cases the cells analyzed were induced (5ddox) 8p Full *XIST* HT1080 cells labelled by IF-FISH. The average (mean and median) z-score for each population of 60 cells analyzed per condition as well as the standard deviation (sd) are listed for each mark. The population of cells for each inhibitor was compared to the relevant control treatment using the Mann-Whitney U test, with the p-value listed in the final column.

#### 6.3 Discussion

The analysis of the PRCs and HDACs on XIST-mediated chromatin inactivation revealed several novel insights into how histone deacetylation and the polycomb complexes fit within the larger context of the XIST-mediated inactivation pathway. Several technical observations of note were made regarding the use of the chemical inhibitors TSA, RGFP966, GSK343 and PRT4165, which here were all tested for longer time periods *in vitro* than in any other known set of tests to date. Extended exposure to TSA was found to be particularly stressful in the HT1080 cell line, as even doses considered relatively low according to the literature produced significant cell death over the course of 6 days [297], [298]. The  $1\mu$ M concentration of RGFP966 observed to be sufficient to affect HDAC3 in mouse cells also proved effective in the human HT1080 cells, and led to an effective increase of global histone acetylation without causing significant apoptosis or stress. Both GSK343 and PRT4165 were found to influence global levels of H3K27me3 and ubH2A respectively at the concentrations described in the literature (table 6.1)[303], [305]. GSK343 and PRT4165 both produced noticeable signs of stress on the cells after the five days of treatment, despite not completely abrogating the respective histone marks. The stress associated with PRC2 inhibition fits within the current literature of PRC2 being an important factor in cell proliferation and preventing apoptosis, particularly in cancer cells [308]-[310]. PRC1 inhibition has also been linked to increased rates of DNA damage, stress and cell death, particularly in cancer cells, and so the higher rates of stress associated with its disruption may have been affected by that [303], [311]. The potential risks associated with cell stress in all these instances prompted at least two concentrations for TSA, GSK343 and PRT4165 to be used in the

experiments examining relative levels of *XIST* and gene silencing which helped distinguish the direct effects of the inhibitor on *XIST* activity from general effects associated with cell stress.

The effect of the inhibitors on various histone marks was as anticipated for the most part. Both HDAC inhibitors decreased global levels of H3K27me3 likely due to the general spreading of histone acetylation as had been reported by numerous other sources [312]–[314]. It was interesting to note however that inhibition of PRC2 did not produce a reciprocal increase in histone acetylation suggesting that H3K27me3 levels are not globally responsible for limiting the spread of acetylation. The similar histone hyperacetylation and H3K27me3 depletion observed between the TSA and RGFP966 treatments suggested that RGFP966 inhibited HDAC3 far more completely than TSA did. The effect of TSA on histone acetylation was a result of the non-specific inhibition of the HDAC family in general, including the other class one HDACs, HDAC1, HDAC2 and HDAC8, which had all been well documented powerful regulators of histone deacetylation [315]–[319].

Both HDAC inhibition and PRC1 inhibition did not produce a significant effect on the levels of *XIST* transcript present in the HT1080 cells following 5 days of induction. The decrease in *XIST* transcript levels following GSK343 treatment to inhibit PRC2 was unexpected but evident across both the concentrations tested. The  $5\mu$ M concentration of GSK343 still allowed for clear and visible enrichment of *XIST* RNA as a cloud by FISH. No one immediate or obvious mechanistic explanation has yet been reported for the reduced transcript levels observed when PRC2 was inhibited.

HDAC inhibition on the activity of *XIST* in the H1080 cell lines revealed several novel insights into *XIST* initiated pathways. The first was the evident dispensability of deacetylation to *XIST*-

mediated silencing. This was made particularly clear by the technically stronger silencing of three of the four distal genes tested when HDAC3 specifically was inhibited. As HDAC3 was shown to be important for mouse *Xist* silencing activity, the results presented here suggest a functional divergence has occurred between the different species [114]. Analysis of the HDAC inhibitors clearly support the evidence of a connection between histone deacetylation and H3K27me3 enrichment which had been suggested from chapter 5. The inhibition of HDAC3 specifically dramatically weakened the enrichment of H3K27me3, while the general inhibition of HDAC4 inhibition of HDAC5 did not. As discussed previously, the weaker repression of HDAC3 by TSA compared to RGFP966 indicate that a certain threshold for HDAC3 inhibition is required before this will translate into an effect on H3K27me3 enrichment.

PRC2 was found to surprisingly be linked to both *XIST* silencing and various novel aspects of chromatin remodeling. The observation that even the lower concentration of PRC2 inhibitor resulted in a complete disruption of silencing at all four genes suggests that PRC2 activity may be important for the establishment of silencing though it will be important to examine whether the lower XIST levels impacted these results. This is generally not a common theory and requires numerous caveats that would need to be examined. To reconcile this observation with the finding in chapter 5 that loss of Repeat E disrupted H3K27me3 enrichment but not silencing would suggest that PRC2 activity at Repeat F was necessary for establishing silencing but that additional factors at Repeat E were necessary for PRC2 to broadly enrich the chromatin with H3K27me3. Such a model would effectively distinguish between the activation of PRC2 at specific regions and its spreading across broad chromatin domains. While a few studies have suggested a compartmentalization of roles for PRC2 [50], [95], the findings presented in this

work are contrary to the observations and general consensus that mouse *Xist*-mediated silencing is independent of PRC2 activity [114], [128], [129], [142]. It also must be noted that this connection between PRC2 and silencing may be specific to the model system used here and will need to be tested further in other models. A far less ambiguous finding of this work was validation of a clear hierarchical pathway between PRC2 activity and *XIST*-mediated enrichment of MacroH2A. At time of writing this represented the first mechanism identified of how MacroH2A becomes enriched during XCI. The further observation that PRC2 activity was completely dispensable for *XIST* to recruit ubH2A was anticipated based on the results of chapter 5, but illustrated that some aspects of *XIST* function were unaffected by PRC2 inhibition.

The inhibition of PRC1 in the human model revealed it to play a relatively narrow role in *XIST*mediated chromatin inactivation. PRC1 activity was not substantially associated with gene silencing or to contributing to H3K27me3 enrichment at the *XIST* RNA cloud. The aspect of the *XIST* chromatin remodeling that PRC1 activity was essential for, besides ubH2A, was the enrichment of SMCHD1. This is in keeping with observations of SMCHD1 dependency on PRC1 in mouse models of XCI [167], [320]. The mild connection between PRC1 activity and histone deacetylation may be a result of the reorganization of the *XIST* bound region into a transcriptionally inactivated territory, contributing to histone deacetylation and loss of active marks. There was some evidence, not directly related to XCI, linking PRC1 and histone deacetylation to overall chromatin compaction through feedforward processes, though it is unclear if that is responsible in this instance [321]. The perhaps most striking observation of PRC1 inhibition was the clear demonstration of independence between *XIST*-mediated recruitment of the two polycomb group complexes as it demonstrates the intense functional

divergence which has occurred between mouse and human *XIST* [127], [140]. Inhibition of PRC2 did not affect ubH2A enrichment either, which indicated that there was little evidence to even suppose that the feed forward mechanisms proposed to operate in mouse models were conserved between these complexes [139], [144].

This work revealed several novel findings and some of the first insights into the order and hierarchy of the mechanism and pathways initiated by *XIST*. It was demonstrated that silencing by human *XIST* was independent of histone deacetylation or ubH2A enrichment, but seemed to be linked to the activity of PRC2. It was observed that deacetylation of the *XIST* RNA surrounded chromatin by HDAC3 specifically was an essential element for H3K27me3 enrichment which in turn was essential for MacroH2A enrichment. It was also demonstrated that ubH2A facilitates the enrichment of SMCHD1 by *XIST*. The next stage of research will need to examine how these pathways interact on a mechanistic level. Additionally, a potentially fruitful avenue for future research would be an examination of the local changes in these chromatin marks specifically at the genes being repressed, to identify how the spreading of silencing and chromatin remodeling are interlinked. Finally, the mechanisms responsible for the initial establishment of gene silencing by *XIST* still remain elusive, and further research examining how individual genes become silenced will potentially revolutionize the understanding of *XIST*, long non-coding RNAs and the field of epigenetics as a whole.

### **Chapter 7: Conclusion**

## 7.1 *XIST* homologues have retained some functional compatibility but were affected by their context

In this thesis, *XIST* was examined as it inactivated an autosome in a male somatic cell for the ultimate goal of understanding how the domains of *XIST* function in a context free of the *XIC* and the natural process of XCI occurring during differentiation. This was done for two reasons, to advance the current understand how *XIST* and lncRNA genes in general function on a mechanistic level and to ensure that these results would be most easily applied to developing *XIST* based therapies for chromosomal abnormalities [211], [212]. As stated throughout this thesis, most of the understanding of how *XIST* functions prior to this study came from studies in mice, with the assumption being that mechanisms were likely conserved. The results of this thesis reveal entirely novel layers of complexity driving the various aspects of *XIST* activity and also for the first time present an initial stepwise outline of several of the most prominent XCI pathways; the results of which have been summarized in figure 7.1.

An important consideration is that this work deliberately focused on analyzing *XIST* at the expense of investigating the broader context in which XCI occurs. Throughout this work however it was regularly demonstrated that the chromatin context in which *XIST* was being expressed from and the cell line themselves impacted various aspects of *XIST* function. In chapter 3 it was observed that even within closely related HT1080 cell lines the ability of *XIST* to enrich its chromatin with H3K27me3 or H4K20me1 was highly variable and dependent upon the *XIST* integration loci. It was particularly interesting to note that while recruitment of both of

these chromatin marks relied on a common region of *XIST* (Repeat E), they differed dramatically in which *XIST* loci were capable of recruiting these marks. Traditionally, elements on the Xchromosome such as LINE-1 elements [192], [193] or CTCF binding sites [201], [202] were examined in the context of *XIST* itself to spread along chromatin. However, the observed effect of the variable functionality of the *XIST* loci in the HT1080 cells suggested a more complex system existed. Recent evidence of histone deacetylation in mice indicated that HDAC3 was prebound to domains on the X chromosome, and that *XIST* spreading activated those enzymes rather than recruiting them [114]. While numerous studies have suggested that the Xi becomes enriched for PRC2 components and was therefore being recruited reconciling these two observations would suggest that some pre-existing element(s) are crucial for H3K27me3 enrichment (e.g [131]). For this reason, future studies examining the occupancy of PRC2 components at different loci prior to and during the early stages of *XIST* induction in various HT1080 cell lines would be expected to reveal novel insights into what these elements might be.

In addition to the role of the chromatin context, it was observed that human *XIST* localized in mouse ES cells and mouse *Xist* localized in somatic cells to create transcriptionally inert compartments within the nucleus as visualized by depletion of CoT-1 hybridization (Appendix figure C.1) [322]. Mouse *Xist* was only capable of inducing gene silencing in a narrow developmental window [206]–[208]. Therefore, it was interesting to observe that the dependence on a particular state of development was not a result of unique divergence between the *Xist* construct but a divergence between the biology of the species. An exploratory study of the factors bound by mouse *Xist* in the HT1080 cells (similar to those protein pull-down studies

in 2015) might therefore be interesting to determine which factors were found to be capable of binding *Xist* across these different contexts [57], [59], [103].

## 7.2 *XIST* localization and regulation of chromatin architecture were highly stable and likely facilitated by several partially redundant processes

Throughout this thesis the localization of XIST RNA and position relative to the perinucleolar compartment was analyzed by FISH in order to gain novel insights into how these pathways are initiated and whether any connection could be drawn to other aspects of XIST activity. CIZ1 was described as a crucial factor for maintaining Xist RNA localization to the Xi in mouse somatic cells through exclusively interacting with Repeat E [105], [107]. The work in this thesis demonstrated that CIZ1 recruitment in humans depended upon two distinct regions of XIST, Repeat Bh and E, but that the disruption of XIST-mediated CIZ1 enrichment did not affect XIST localization. Knockdown of CIZ1 in human somatic cells did not affect XIST localization either, suggesting that CIZ1 plays a non-essential role in XIST RNA localization in somatic cells of humans compared to mice. This could explain part of differences observed in the importance of differentiation to mouse but not human XIST activity [107], [108]. Xist functioned effectively in a narrow developmental window in mouse cells, but human XIST appears to function more broadly across stages of differentiation in humans [110], [206]. The presumption was that certain binding factors were only present or able to bind Xist during that developmental window [206]. It may be that the factors binding to human *XIST* and contributing to its localization and spreading are not affected by the state of differentiation and function cooperatively/redundantly with CIZ1 to maintain XIST localization in humans though as these tests were done in a cancer cell line they 263

may have more embryonic like characteristics than a typical somatic cell. The argument for redundancy in the *XIST* localization factors is supported by the observation that the removal of the regions identified as key sites for hnRNPU (3'PflMI to Repeat D region) binding and CIZ1 (Repeat E) had no effect on *XIST* localization individually but that the loss of both regions ( $\Delta\Delta$ ) resulted in a visible delocalization of *XIST* [105]. The reduced localization of *XIST* in constructs lacking the 3' region of exon 1 ( $\Delta$ 3D5E) suggests that as yet unidentified factors involved in *XIST* localization or coalescence will likely be identified by focused screens of that region.

The recruitment of SMCHD1 also depends upon the 3' region of exon 1 also associated with an attenuated XIST localization, and SMCHD1 was associated with the restructuring of chromatin territories and the creation of the megadomains associated with the Xi in mice [166]. In humans knockdown of SMCHD1 was associated with a less compact Xi [46]. The dependency on the 3' region of exon 1 for both RNA localization and SMCHD1 enrichment suggests that in humans SMCHD1 enrichment may be another factor which contributes to the formation of a unified XIST RNA cloud. Critically however, the deletions which disrupted SMCHD1 enrichment indirectly and presumably through disrupting PRC1 enrichment did not obviously affect the distribution of the XIST RNA cloud. It remains unclear whether loss of SMCHD1 was a cause or a result of the disruption to XIST unification as studies have suggested that SMCHD1 both affects and is affected by chromosome architecture [132], [323]. An exciting potential direction to build on these results would examine how the localization of XIST RNA into a unified cloud within the nucleus affects TAD formation and chromatin structure at a sequence specific level. In addition, it would be interesting to perform knockout experiments on SMCHD1 and directly analyze the distribution of the XIST RNA itself as well as any positional changes to an Xi within the nucleus.

Increased perinucleolar localization was observed following 5 days of *XIST* induction despite the lack of the *DXZ4* locus from chromosome 8p where *XIST* was being expressed [171]. It is impossible to rule out at this stage that the lncRNA *FIRRE* may have contributed to this process, as it was shown to act *in trans* [174]. The increased perinucleolar localization of the *XIST* RNA locus following induction in all the deletion cell lines indicated that this pathway was perhaps the most resilient pathway examined and implied that numerous redundant processes were involved. It is intriguing that the region identified with causing the greatest perturbation to this perinucleolar localization is the same one described previously as being essential for *XIST* recruitment of SMCHD1 *XIST* RNA unification and SMCHD1 enrichment. Future research performing chromatin isolation through RNA purification (ChIRP) to compare the factors binding to this non-repeat 3' region of exon 1 may therefore provide invaluable new insights into what pathways mediate the structure and positions of the *XIST* RNA cloud within the nucleus.

# 7.3 Silencing mediated by *XIST* occurs through a unique pathway preceding chromatin remodeling

The results presented here suggest that one of the first effects of XIST transcription is the establishment of gene silencing even across large distances. While the growing evidence from mouse models suggested that HDAC3-mediated deacetylation was a likely crucial element to *XIST* silencing this was not supported by the studies performed here [103], [114]. Silencing was found to precede the establishment of histone deacetylation across the chromatin territory, proceeding nearly as fast as *XIST* RNA accumulated. The 3' non-repeat region of *XIST* was found to be completely dispensable for deacetylation of H3K27 but XIST lacking this region
failed to silence any of the genes examined. Finally, the catalytic inhibition of both HDAC3 specifically and broader HDACs clearly did not disrupt distal gene silencing in any way. This all suggests that histone deacetylation and gene silencing are distinct pathways initiated by XIST RNA in humans. An exciting future step will be to determine what role, if any, SPEN plays in XIST activity, as this essential factor for Xist-mediated silencing has yet to be studied in humans [113]. Further work will be needed to determine the factors essential for silencing that associate with XIST 5' and 3' regions, as well as to determine how these various factors interact. It may be that the binding of RNA methylation dependent factors at these two regions of XIST is crucial to its silencing, as the role of RBM15 mediated methylation was discounted as essential in mouse *Xist* but has not been similarly tested in human *XIST* [116], [117]. It was interesting to observe that the region of XIST encompassing Repeat F and the surrounding non-repeat region was essential for both silencing and H3K27me3 enrichment, but that this was the only region of XIST where these two processes were associated. As was discussed in chapter 6, it remains to be determined why the catalytic inhibition of PRC2 led to a clear defect in silencing of all four genes examined but loss of H3K27me3 enrichment by removing Repeat E did not affect XISTmediated silencing. The analysis of the dynamics of XIST mediated activity in chapter 3 suggested that even levels of XIST below 68% of the maximal level of XIST resulted in effectively complete silencing of distal genes within two days. In addition, the  $\Delta 3'$  constructs had similar levels of XIST following induction as those seen in the 5µM PRC2 inhibition treatments, but were capable of still initiating chromatin remodeling. In mice it was recently proposed that PRC1 & 2 are both needed for *Xist* to spread properly during XCI, and further work will be required to validate these findings. One potential explanation may be that the activity of PRC2 itself is necessary for XIST to associate with the domains of the four genes

tested here [144]. A study likely to rapidly yield numerous insights into the role of *XIST*mediated silencing would be to perform ChIRP to identify what factors specifically bind to the Full length *XIST* RNA in the HT1080 model that cannot bind to the  $\Delta$  A,  $\Delta$  FBh and  $\Delta$  3' constructs.

## 7.4 Insights into chromatin remodeling by *XIST*

Testing how XIST modified the chromatin coated with its RNA revealed that multiple regions were crucial for each process in defiance of the expectation that a single region would initiate a single pathway. The accumulation of heterochromatin as well as loss of histone acetylation observed in the 8p inducible XIST model systems were less extensive than on the endogenous female Xi and occurred more slowly than in mouse ESCs undergoing induced XCI as expected based on the evidence of *cis* acting factors on the X chromosome contributing to chromatin remodeling [114], [193]. The most extensively analyzed Xi associated chromatin marks, PRC2mediated H3K27me3 and PRC1-mediated ubH2A, had been anticipated to be recruited by a region of XIST encompassing the B and C repeats and for PRC2 to be reliant on PRC1 based on mouse models [129]. Throughout this work it was demonstrated that PRC2 and PRC1 act through entirely independent XIST initiated pathways that in turn are crucial for other changes to the chromatin. A stronger relative accumulation of ubH2A than H3K27me3 was observed at the XIST RNA cloud induced from chromosome 8p in HT1080, though both marks seemed to reach their maximal level within 5 days of XIST induction. The accumulation of ubH2A was observed to depend upon the same general Repeat B and C containing region of XIST that was also crucial in mice, though in direct contrast with recent mouse studies Repeat A was also observed to be

important for ubH2A enrichment [127], [129]. The Xist 3' non-repeat region and Repeat D were also found to be important for ubH2A enrichment and it was tempting to wonder if some functional components might not have shuffled between the expanded human Repeat D and the unusually expanded mouse Repeat C [67]. These ubH2A essential domains of XIST did not overlap with the regions essential for H3K27me3, which depended upon Repeat F, Bh and E. The tests using chemical inhibitors of PRC1 and PRC2 further confirmed that the two types of polycomb group complexes operate independently, with no evidence of even cooperative heterochromatin mark accumulation as was proposed in mice [139], [144]. Repeat F and B in mice had been proposed once as a potential site where the Jarid2 cofactor of PRC2 bound, and while this interaction has not been tested further it did suggest that future research might benefit from examining JARID2 binding on human XIST [131]. While PRC2 and PRC1 activity have both been associated with mouse Xist Repeat B, it was curious that the activity of these two complexes during human XCI depended on the two separate human B repeats (Bh and B) [65], [67]. Inhibition of PRC1 activity disrupted SMCHD1 enrichment, and both of these Xi associated factors depended upon Repeat B, C and D as well as the non-repeat 3' region of XIST. This suggests that the dependence of SMCHD1 on ubH2A/PRC1 during XCI was conserved between humans and mice [167]. However, Repeat A was crucial for normal levels of ubH2A enrichment but did not affect SMCHD1, suggesting that the pathway connecting these processes was not directly linearly dependent. Studies of mouse cells had suggested that the mark ubH2A rather than PRC1 itself was crucial for SMCHD1 enrichment, and it may be that during XCI the ubiquitination of specific 'seed' loci is crucial for SMCHD1 to be recruited and enriched during XCI, but that the broader accumulation of ubH2A is not essential [132]. This model revolves around the role of Repeat A being crucial for the larger spread of ubH2A which will need to be

studied more extensively through ChIP or similar protocols to determine whether this hypothesis has any merit. The association of the non-repeat region of *XIST* between Repeat D and E with SMCHD1 enrichment but not ubH2A indicates that ubH2A is not sufficient for SMCHD1 enrichment. Crosslinking experiments in mice have suggested that SMCHD1 may bind *Xist* directly, and future studies may benefit from examining the non-repeat region at the end of exon 1 for potential SMCHD1 binding activity in humans [46].

This work presented one of the first insights into how the histone variant MacroH2A is recruited to the Xi by demonstrating a dependency on PRC2 catalytic activity as well as the region of *XIST* spanning both a large non-repeat region of exon 1 as well as Repeat D. This multifactorial system of MacroH2A recruitment reconciled the observation that MacroH2A enrichment is lost from the Xi if Xist is depleted despite H3K27me3 being maintained [324]. As an interesting note this non-repeat region identified as essential for MacroH2A enrichment was proposed to form long range interactions with regions of Repeat D [89]. MacroH2A was identified as interacting with mouse Xist in one of the 2015 screens, and it may be that the secondary structures formed between this internal region of XIST exon 1 facilitates an interaction between human XIST and MacroH2A. Future tests of that region will be needed to examine this hypothesis, but such an investigation could likely provide novel insights about how and if MacroH2A can interact with RNA as well as potentially finally revealing what role MacroH2A plays during XCI. It was interesting that while ubH2A enrichment was essential for both H3K27me3 and MacroH2A enrichment on the mouse Xi, MacroH2A enrichment by human XIST only depended upon H3K27me3 enrichment. The moderate decrease in H3K27me3 that resulted from loss of the region surrounding the Bh repeat did not noticeably affect MacroH2A enrichment, suggesting

that only a certain threshold of H3K27me3 is needed at the *XIST* RNA cloud for MacroH2A to in turn be enriched.

Finally, the results of this work provided new insights into the regions of XIST crucial for the enrichment of chromatin with H4K20me1, however the importance of this modification remains poorly understood. H4K20me1 enrichment was found to depend upon Repeat E, like H3K27me3 and CIZ1, however it remained normally enriched in the deletion constructs that failed to recruit these other Repeat E associated marks, and therefore likely represented a unique pathway initiated by *XIST*. This was further supported by the observed difference between the various *XIST* integration sites that recruited H3K27me3 or H4K20me1, suggesting that additional context dependent factors were involved in regulating these two pathways. This contrasted evidence in mice supporting a potential link between PRC2 and KMT5A activity, and future research into the role of KMT5A during XCI will be critical to determine why the Xi is enriched with H4K20me1 marks across species [128]. One potential direction for such research would be by treating cells with SPS8I1-3, a chemical inhibitor of KMT5A, while inducing *XIST* expression and observing any effects on *XIST* localization or other aspects of XCI [149], [325].

## 7.5 Final conclusions about the mechanisms governing *XIST* activity

Analysis of all the deletion constructs revealed that every Xi associated function of *XIST* relies on multiple distal regions of its gene. Potential explanations include that multiple independent binding proteins are needed to mediate each process, that long range interactions between regions of *XIST* are common to *XIST* activity or a combination of these two models. The

presence of long range interactions has been observed in both human and mouse models, though the significance has not been determined for a given set of interactions [89]. If long range interactions between regions of XIST prove to be critical for forming the secondary structures and binding domains that facilitate XCI it adds a new layer of complexity as identifying the structures formed by relatively small self-associating regions such as Repeat A have proven difficult to analyze [88]–[90]. A feature of studying XIST in both humans and mice has typically been to identify a single region responsible for a single process, much like early studies of genetics attempted to connect a single gene with a phenotype. It may be that the conflicting models of how various complexes are recruited, such as PRC2 in mice being either recruited by PRC1 or XIST directly, may be a result of multiple factors and structures within XIST coordinating the recruitment and activity of these processes [127], [131], [307]. An exciting avenue for future research will be to examine how disruption of specific regions of XIST influences protein binding profiles at distal region of XIST, as these results demonstrated that interdependence between domains is a common feature of XIST. It will be interesting also to examine whether these connections between regions of XIST are occurring within single molecules or are facilitating cross-talk between different molecules. The differences observed between the human somatic model used here and the mouse models that these results have been compared to illustrate the need for further investigation into how the context that XIST is expressed in affects its activity, particularly in humans.

An understated finding of this work was how durable *XIST* was discovered to be. Throughout chapter 4 large sections (>3kb) were frequently excised with minimal observable effect on the overall ability of *XIST* to still silence distal genes, localize to the perinucleolar compartment and

recruit a wide array of heterochromatin marks. One of the early objectives of the work described here was to identify a minimal region of XIST that was capable of silencing broad chromosome regions. While every region examined was important for at least one of the myriad of functions of XIST, the conclusions of this work (summary 7.1) demonstrate that a potentially small region of XIST lacking Repeat D and 3' region of exon 1 might be able to be removed (~7kb in total) with minimal effect on the establishment of gene silencing, perinucleolar localization and enrichment of the XIST RNA with H3K27me3 as well as H4K20me1. It is also worth mentioning that despite the prevailing view of the repeats being the crucial elements of XIST activity, throughout this work it was regularly demonstrated that the non-repeat regions were all crucial for numerous aspects of XIST activity. Further research using the map presented here along with predicted long-range interactions between regions of XIST[87], [89] will likely be able to identify even smaller functional XIST constructs, which will be of critical for advancing research into XIST based therapeutics [212]. Further work will also be required to test the results described here in an endogenous context, so that this new information on XIST can be integrated into the broader context in which the process of XCI occurs.



Figure 7.1 Summary of the identified functional domains pathways of XIST

A stylized representation of the results presented throughout this thesis. The XIST transcript sequence (light grey with repeats in darker grey) was analyzed using deletion constructs to identify the regions essential for its various functions. The examined processes included XISTmediated gene repression (black), depletion of H3K27ac (green) and the enrichment of H3K27me3 (red), MacroH2A (blue), ubH2A (purple), SMCHD1 (brown), H4K20me1 (yellow) and CIZ1 (orange). The regions essential for each function are highlighted in the relevant solid colour while the regions dispensable for a process are left unhighlighted on a grey background of the XIST transcript. Crosshatching is used for H3K27ac as the regions identified only contributed to depletion rather than being essential. PRC2 lines are superimposed on the MacroH2A regions to illustrate that MacroH2A dependency on these regions may be a downstream effect of PRC2. PRC1 lines are similarly superimposed on SMCHD1 to illustrate the potential connection of the pathways in these regions. The arrows underneath XIST illustrate the pathways proposed throughout this research and the arrows were colour coded accordingly. Silencing was observed to precede the chromatin remodeling in general though a potential need for PRC2 catalytic activity was suggested (dashed line). HDAC3-mediated removal of acetylation was linked to PRC2 trimethylation of H3K27me3. PRC2 activity was in turn essential for MacroH2A to become enriched. PRC1-mediated ubiquitination of H2A was independent of PRC2 and was essential for SMCHD1 to become enriched on the chromatin. H4K20me1 and CIZ1 seemed to occupy independent pathways.

## References

- [1] Okamoto I., Patrat C., Thépot D., Peynot N., Fauque P., Daniel N., Diabangouaya P.,
   Wolf J. P., Renard J. P., Duranthon V., Heard E., "Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development," *Nature*, vol. 472, no. 7343, pp. 370–374, 2011.
- [2] Barr M. L., Bertram E. G., "A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis [2]," *Nature*, vol. 163, no. 4148. Nature Publishing Group, pp. 676–677, 1949.
- [3] Gartler S. M., Varadarajan K. R., Luo P., Norwood T. H., Canfield T. K., Hansen R. S.,
   "Abnormal X: Autosome ratio, but normal X chromosome inactivation in human triploid cultures," *BMC Genetics*, vol. 7, no. 41. 2006.
- [4] Deng X., Nguyen D. K., Hansen R. S., Van Dyke D. L., Gartler S. M., Disteche C. M.,
  "Dosage regulation of the active X chromosome in human triploid cells," *PLoS Genetics*,v
  ol. 5 no. 12. e1000751. 2009.
- [5] Russell W. L., Russell L. B., Gower J. S., "Exceptional inheritance of a sex-linked gene in the mouse explained on the basis that the x/o sex-chromosome constitution is female," *Proceedings of the National Academy of Sciences*, 1959.
- [6] Ohno S., Kaplan W. D., Kinosita R., "Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus," *Experimental Cell Research*, vol. 45, no. 4, pp. 554-560. 1959.
- [7] Lyon M. F., "Gene action in the X-chromosome of the mouse (mus musculus L.)," *Nature*, vol. 190, no. 4773, pp. 372–373, 1961.
- [8] Lyon M. F., "Sex chromatin and gene action in the mammalian X-chromosome.,"

American journal of human genetics, vol. 14, no. 2, p.p. 135-48. 1962.

- [9] Julien P., Brawand D., Soumillon M., Necsulea A., Liechti A., Schütz F., Daish T., Grützner F., Kaessmann H., "Mechanisms and evolutionary patterns of mammalian and avian dosage compensation," *PLoS Biology*, vol. 10, no. 5, p.p. 1001328, 2012.
- [10] Livernois A. M., Waters S. A., Deakin J. E., Marshall Graves J. A., Waters P. D.,
   "Independent Evolution of Transcriptional Inactivation on Sex Chromosomes in Birds and Mammals," *PLoS Genetics*, vol. 9, no. 7, p.p. 1003635, 2013.
- [11] Cortez D., Marin R., Toledo-Flores D., Froidevaux L., Liechti A., Waters P. D., Grützner
   F., Kaessmann H., "Origins and functional evolution of y chromosomes across mammals," *Nature*, vol. 508, no. 7479, p.p. 488-493, 2014.
- [12] Grützner F., Rens W., Tsend-Ayush E., El-Mogharbel N., O'Brien P. C. M., Jones R. C., Ferguson-Smith M. A., Marshall Graves J. A., "In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes," *Nature*, vol. 432, no. 7019, p.p. 913-917, 2004.
- [13] Rens W., Grützner F., O'Brien P. C. M., Fairclough H., Graves J. A. M., Ferguson-Smith M. A., "Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y 3X4Y4X5Y5 male sex chromosome constitution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 46, p.p. 16257-61, 2004.
- [14] Cooper D. W., Woolley P. A., Maynes G. M., Sherman F. S., Poole W. E., "Studies on metatherian sex chromosomes. XII. Sex-linked inheritance and probable paternal Xinactivation of a-galactosidase A in Australian marsupials," *Australian Journal of Biological Sciences*, vol. 36, no. 6, p.p. 511-517, 1983.

- [15] Luo Z. X., Yuan C. X., Meng Q. J., Ji Q., "A Jurassic eutherian mammal and divergence of marsupials and placentals," *Nature*, vol. 476, no. 7361, p.p. 442-445, 2011.
- [16] Skaletsky H., Kuroda-Kawaguchl T., Minx P. J., Cordum H. S., Hlllier L. D., Brown L. G., Repplng S., Pyntikova T., All J., Blerl T., Chinwalla A., Delehaunty A., Du H., Fewell G., Fulton L., Fulton R., Graves T., ... Page D. C., "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes," *Nature*, vol. 423, no. 9642, p.p. 825-837, 2003.
- [17] Cooper D. W., Johnston P. G., Vandeberg J. L., Robinson E. S., "X-Chromosome inactivation in Marsupials," *Australian Journal of Zoology*, vol. 37, no. 3, p.p. 411-417, 1989.
- [18] Al Nadaf S., Waters P. D., Koina E., Deakin J. E., Jordan K. S., Graves J. A. M.,
   "Activity map of the tammar X chromosome shows that marsupial X inactivation is incomplete and escape is stochastic," *Genome Biology*, vol. 11, no. 12, p.p. R122, 2010.
- [19] Duan J., Shi W., Jue N. K., Jiang Z., Kuo L., O'Neill R., Wolf E., Dong H., Zheng X., Chen J., Tian X., "Dosage compensation of the X chromosomes in bovine germline, early embryos, and somatic tissues," *Genome Biology and Evolution*, vol. 11, no. 1, p.p. 242-252, 2019.
- [20] Xue F., Tian X. C., Du F., Kubota C., Taneja M., Dinnyes A., Dai Y., Levine H., Pereira L. V., Yang X., "Aberrant patterns of X chromosome inactivation in bovine clones," *Nature Genetics*, vol. 31, no. 2, p.p. 216-220, 2002.
- [21] Min B., Park J. S., Jeon K., Kang Y. K., "Characterization of X-Chromosome gene expression in bovine blastocysts derived by in vitro fertilization and somatic cell nuclear transfer," *Frontiers in Genetics*, vol. 10, no. 42, 2017.

- [22] Corbel C., Diabangouaya P., Gendrel A. V., Chow J. C., Heard E., "Unusual chromatin status and organization of the inactive X chromosome in murine trophoblast giant cells," *Development (Cambridge)*, vol., no., p.p., 2013.
- [23] Kunath T., Arnaud D., Uy G. D., Okamoto I., Chureau C., Yamanaka Y., Heard E.,
  Gardner R. L., Avner P., Rossant J., "Imprinted X-inactivation in extra-embryonic
  endoderm cell lines from mouse blastocysts," *Development*, vol. 140, no. 4, p.p. 861-872, 2005.
- [24] Mak W., Baxter J., Silva J., Newall A. E., Otte A. P., "Mitotically stable association of polycomb group proteins Eed and Enx1 with the inactive X chromosome in trophoblast stem cells," *Current Biology*, vol. 12, no. 12, pp. 1016–1020, 2002.
- [25] Berletch J. B., Yang F., Disteche C. M., "Escape from X inactivation in mice and humans," *Genome Biology*. vol. 11, no. 6, p.p. 213, 2010.
- [26] Yang F., Babak T., Shendure J., Disteche C. M., "Global survey of escape from X inactivation by RNA-sequencing in mouse," *Genome Research*, vol. 20, no. 5, p.p. 614-620, 2010.
- [27] Balaton B. P., Brown C. J., "Escape Artists of the X Chromosome," *Trends in Genetics*, vol. 32, no. 6. pp. 348–359, 2016.
- [28] Rastan S., "Timing of X-chromosome inactivation in postimplanation mouse embryos," *Journal of Embryology and Experimental Morphology*, vol. 71, p.p. 11-24, 1982.
- [29] Tan S. S., Williams E. A., Tam P. P. L., "X-chromosome inactivation occurs at different times in different tissues of the post-implantation mouse embryo," *Nature Genetics*, vol. 3, no. 2, p.p. 170-174, 1993.
- [30] PARK W. W., "The occurrence of sex chromatin in early human and macaque embryos.,"

Journal of anatomy, vol. 91, no. 3, p.p. 369-373, 1957.

- [31] Austin C. R., Amoroso E. C., "Sex chromatin in early cat embryos," *Experimental Cell Research*, vol. 13, no. 2, p.p. 419-421, 1957.
- [32] Takagi N., Martin G. R., "Studies of the temporal relationship between the cytogenetic and biochemical manifestations of X-chromosome inactivation during the differentiation of LT-1 teratocarcinoma stem cells," *Developmental Biology*, vol. 103, no. 2, pp. 425– 433, 1984.
- [33] Martin G. R., Epstein C. J., Travis B., Tucker G., Yatziv S., Martin D. W., Clift S., Cohen S., "X-chromosome inactivation during differentiation of female teratocarcinoma stem cells in vitro," *Nature*, vol. 271, no. 5643, pp. 329–333, 1978.
- [34] Jeppesen P., Turner B. M., "The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression," *Cell*, vol. 74, no. 2, pp. 281–289, 1993.
- [35] Belyaev N. D., Keohane A. M., Turner B. M., "Differential underacetylation of histones H2A, H3 and H4 on the inactive X chromosome in human female cells," *Human Genetics*, vol. 97, no. 5, pp. 573–578, 1996.
- [36] Silva J., Mak W., Zvetkova I., Appanah R., Nesterova T. B., Webster Z., Peters A. H. F.
   M., Jenuwein T., Otte A. P., Brockdorff N., "Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes," *Developmental Cell*, vol. 4, no. 4, pp. 481–495, 2003.
- [37] Plath K., Fang J., Mlynarczyk-Evans S. K., Cao R., Worringer K. A., Wang H., De la Cruz C. C., Otte A. P., Panning B., Zhang Y., "Role of histone H3 lysine 27 methylation in X inactivation," *Science*, vol. 300, no. 5616, pp. 131–135, 2003.

- [38] de Napoles M., Mermoud J. E., Wakao R., Tang Y. A., Endoh M., Appanah R., Nesterova T. B., Silva J., Otte A. P., Vidal M., Koseki H., Brockdorff N., "Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation.," *Developmental Cell*, vol. 7, no. 5, pp. 663–676, 2004.
- [39] Kohlmaier A., Savarese F., Lachner M., Martens J., Jenuwein T., Wutz A., "A chromosomal memory triggered by Xist regulates histone methylation in X inactivation," *PLoS Biology*, vol. 2, no. 7, p. e171, 2004.
- [40] 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., Kay G. F., Whitelaw E., "SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation," *Nature Genetics*, vol. 40, no. 5, pp. 663–669, 2008.
- [41] Costanzi C., Pehrson J. R., "Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals," *Nature*, vol. 393, no. 6685, pp. 599–601, 1998.
- [42] Chadwick B. P., Willard H. F., "Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 50, pp. 17450–17455, 2004.
- [43] Chadwick B. P., Willard H. F., "Chromatin of the Barr body: Histone and non-histone proteins associated with or excluded from the inactive X chromosome," *Human Molecular Genetics*, vol. 12, no. 17, p.p.2167-2178, 2003.
- [44] Keniry A., Gearing L. J., Jansz N., Liu J., Holik A. Z., Hickey P. F., Kinkel S. A., Moore
  D. L., Breslin K., Chen K., Liu R., Phillips C., Pakusch M., Biben C., Sheridan J. M., Kile
  B. T., Carmichael C., ... Blewitt M. E., "Setdb1-mediated H3K9 methylation is enriched

on the inactive X and plays a role in its epigenetic silencing," *Epigenetics and Chromatin*, vol. 9, no. 16, 2016.

- [45] Heard E., Rougeulle C., Arnaud D., Avner P., Allis C. D., Spector D. L., "Methylation of histone H3 at Lys-9 Is an early mark on the X chromosome during X inactivation," *Cell*, vol. 107, no. 6, pp. 727–738, 2001.
- [46] Nozawa R. S., Nagao K., Igami K. T., Shibata S., Shirai N., Nozaki N., Sado T., Kimura H., Obuse C., "Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway," *Nature Structural and Molecular Biology*, vol. 20, no. 5, pp. 566–573, 2013.
- [47] Cotton A. M., Lam L., Affleck J. G., Wilson I. M., Peñaherrera M. S., McFadden D. E.,
   Kobor M. S., Lam W. L., Robinson W. P., Brown C. J., "Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation," *Human Genetics*, vol. 130, no. 2, pp. 187–201, 2011.
- [48] Gendrel A. V., Apedaile A., Coker H., Termanis A., Zvetkova I., Godwin J., Tang Y. A., Huntley D., Montana G., Taylor S., Giannoulatou E., Heard E., Stancheva I., Brockdorff N., "Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome," *Developmental Cell*, vol. 23, no. 2, pp. 265–279, 2012.
- [49] Sado T., Okano M., Li E., Sasaki H., "De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation.," *Development (Cambridge)*, vol. 131, no. 5, pp. 975–82, 2004.
- [50] Chaumeil J., Le Baccon P., Wutz A., Heard E., "A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when

silenced," Genes and Development, vol. 20, no. 16, pp. 2223–2237, 2006.

- [51] Kucera K. S., Reddy T. E., Pauli F., Gertz J., Logan J. E., Myers R. M., Willard H. F.,
   "Allele-specific distribution of RNA polymerase II On female X chromosomes," *Human Molecular Genetics*, vol. 20, no. 20, p.p. 3964-3973, 2011.
- [52] Rego A., Sinclair P. B., Tao W., Kireev I., Belmont A. S., "The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure," *Journal of Cell Science*, vol. 121, no. 7, p.p. 1119-1127, 2008.
- [53] Eils R., Dietzel S., Bertin E., Schröck E., Speicher M. R., Ried T., Robert-Nicoud M., Cremer C., Cremer T., "Three-dimensional reconstruction of painted human interphase chromosomes: Active and inactive X chromosome territories have similar volumes but differ in shape and surface structure," *Journal of Cell Biology*, vol. 135, no. 6, p.p. 1427-40, 1996.
- [54] Van Koningsbruggen S., Gierliński M., Schofield P., Martin D., Barton G. J., Ariyurek Y., Den Dunnen J. T., Lamond A. I., "High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli," *Molecular Biology of the Cell*, vol. 21, no. 21, p.p. 3735-3748, 2010.
- [55] Deng X., Ma W., Ramani V., Hill A., Yang F., Ay F., Berletch J. B., Blau C. A., Shendure J., Duan Z., Noble W. S., Disteche C. M., "Bipartite structure of the inactive mouse X chromosome," *Genome Biology*, vol. 16, no. 1, p.p. 152, 2015.
- [56] Chen C. K., Blanco M., Jackson C., Aznauryan E., Ollikainen N., Surka C., Chow A., Cerase A., McDonel P., Guttman M., "Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing," *Science*, vol. 354, no. 6311, pp. 468–472, 2016.

- [57] Monfort A., Di Minin G., Postlmayr A., Freimann R., Arieti F., Thore S., Wutz A.,
  "Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells," *Cell Reports*, vol. 12, no. 4, pp. 554–561, 2015.
- [58] Moindrot B., Cerase A., Coker H., Masui O., Grijzenhout A., Pintacuda G., Schermelleh L., Nesterova T. B., Brockdorff N., "A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing," *Cell Reports*, vol. 12, no. 4, pp. 562–572, 2015.
- [59] Chu C., Zhang Q. C., Da Rocha S. T., Flynn R. A., Bharadwaj M., Calabrese J. M., Magnuson T., Heard E., Chang H. Y., "Systematic discovery of Xist RNA binding proteins," *Cell*, vol. 161, no. 2, pp. 404–416, 2015.
- [60] Rastan S., Robertson E. J., "X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation," *Journal of Embryology and Experimental Morphology*, vol. 90, no. 1, p.p. 379-388, 1985.
- [61] Brown C. J., Lafreniere R. G., Powers V. E., Sebastio G., Ballabio A., Pettigrew A. L., Ledbetter D. H., Levy E., Craig I. W., Willard H. F., "Localization of the X inactivation centre on the human X chromosome in Xq13," *Nature*, vol. 349, no. 6304, pp. 82–84, 1991.
- [62] Augui S., Filion G. J., Huart S., Nora E., Guggiari M., Maresca M., Stewart A. F., Heard E., "Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic," *Science*, vol. 318, no. 5856, p.p. 1632-1636, 2007.
- [63] Pollex T., Heard E., "Nuclear positioning and pairing of X-chromosome inactivation centers are not primary determinants during initiation of random X-inactivation," *Nature*

Genetics, vol. 51, no. 2, pp. 285–295, 2019.

- [64] Brown C. J., Ballabio A., Rupert J. L., Lafreniere R. G., Grompe M., Tonlorenzi R.,
  Willard H. F., "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome," *Nature*, vol. 349, no. 6304, pp. 38–44, 1991.
- [65] Brown C. J., Hendrich B. D., Rupert J. L., Lafrenière R. G., Xing Y., Lawrence J.,
   Willard H. F., "The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus," *Cell*, vol. 71, no. 3, pp. 527–542, 1992.
- [66] Penny G. D., Kay G. F., Sheardown S. A., Rastan S., Brockdorff N., "Requirement for Xist in X chromosome inactivation," *Nature*, vol. 379, no. 6561, pp. 131–137, 1996.
- [67] Yen Z. C., Meyer I. M., Karalic S., Brown C. J., "A cross-species comparison of Xchromosome inactivation in Eutheria.," *Genomics*, vol. 90, no. 4, pp. 453–63, 2007.
- [68] Elisaphenko E. A., Kolesnikov N. N., Shevchenko A. I., Rogozin I. B., Nesterova T. B., Brockdorff N., Zakian S. M., "A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements," *PLoS ONE*, vol. 3, no. 6, 2008.
- [69] Grant J., Mahadevaiah S. K., Khil P., Sangrithi M. N., Royo H., Duckworth J., McCarrey J. R., Vandeberg J. L., Renfree M. B., Taylor W., Elgar G., Camerini-Otero R. D., Gilchrist M. J., Turner J. M. A., "Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation," *Nature*, vol. 487, no. 7406, p.p. 254-258, 2012.
- [70] Sprague D., Waters S. A., Kirk J. M., Wang J. R., Samollow P. B., Waters P. D.,
   Calabrese J. M., "Nonlinear sequence similarity between the Xist and Rsx long noncoding RNAs suggests shared functions of tandem repeat domains," *RNA*, vol. 25, no. 8, pp.

1004–1019, 2019.

- [71] Fatica A., Bozzoni I., "Long non-coding RNAs: New players in cell differentiation and development," *Nature Reviews Genetics*, vol. 15, no. 1. pp. 7–21, 2014.
- [72] Kay G. F., Penny G. D., Patel D., Ashworth A., Brockdorff N., Rastan S., "Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation," *Cell*, vol. 72, no. 2, pp. 171–182, 1993.
- [73] Marahrens Y., Panning B., Dausman J., Strauss W., Jaenisch R., "Xist-deficient mice are defective in dosage compensation but not spermatogenesis," *Genes and Development*, vol. 11, no. 2, pp. 156–166, 1997.
- [74] Gupta P., Peter S., Jung M., Lewin A., Hemmrich-Stanisak G., Franke A., von Kleist M., Schütte C., Einspanier R., Sharbati S., Bruegge J. zur, "Analysis of long non-coding RNA and mRNA expression in bovine macrophages brings up novel aspects of Mycobacterium avium subspecies paratuberculosis infections," *Scientific Reports*, vol. 9, no. 1, p.p. 1571, 2019.
- [75] Yang Y., Zhou R., Zhu S., Li X., Li H., Yu H., Li K., "Systematic Identification and Molecular Characteristics of Long Noncoding RNAs in Pig Tissues," *BioMed Research International*, vol. 2017, p.p. 6152582, 2017.
- [76] Hong Y. K., Ontiveros S. D., Strauss W. M., "A revision of the human XIST gene organization and structural comparison with mouse Xist," *Mammalian Genome*, vol. 11, no. 3, pp. 220–224, 2000.
- [77] Ma M., Strauss W. M., "Analysis of the Xist RNA isoforms suggests two distinctly different forms of regulation.," *Mammalian genome : official journal of the International Mammalian Genome Society*, vol. 16, no. 6, pp. 391–404, 2005.

- [78] Brockdorff N., Ashworth A., Kay G. F., McCabe V. M., Norris D. P., Cooper P. J., Swift S., Rastan S., "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*, vol. 71, no. 3, pp. 515–526, 1992.
- [79] Clemson C. M., McNeil J. A., Willard H. F., Lawrence J. B., "XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear/chromosome structure," *Journal of Cell Biology*, vol. 132, no. 3, pp. 259–275, 1996.
- [80] Chureau C., Prissette M., Bourdet A., Barbe V., Cattolico L., Jones L., Eggen A., Avner P., Duret L., "Comparative Sequence Analysis of the X-Inactivation Center Region in Mouse, Human, and Bovine," *Genome Research*, vol. 12, no. 6, 2002.
- [81] 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., Zakian S. M.,
  "Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence," *Genome Research*, vol. 11, no. 5. Cold Spring Harbor Laboratory Press, pp. 833–849, May-2001.
- [82] Wutz A., Rasmussen T. P., Jaenisch R., "Chromosomal silencing and localization are mediated by different domains of Xist RNA," *Nature Genetics*, vol. 30, no. 2, pp. 167– 174, 2002.
- [83] Jones A. N., Sattler M., "Challenges and perspectives for structural biology of lncRNAs the example of the Xist lncRNA A-repeats," *Journal of Molecular Cell Biology*, vol. 11, no. 10, pp. 845–859, 2019.
- [84] Duszczyk M. M., Zanier K., Sattler M., "A NMR strategy to unambiguously distinguish

nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat," *Nucleic Acids Research*, vol. 36, no. 22, pp. 7068–7077, 2008.

- [85] Duszczyk M. M., Wutz A., Rybin V., Sattler M., "The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization," *RNA*, vol. 17, no. 11, pp. 1973–1982, 2011.
- [86] Maenner S., Blaud M., Fouillen L., Savoye A., Marchand V., Dubois A., Sanglier-Cianférani S., Van Dorsselaer A., Clerc P., Avner P., Visvikis A., Branlant C., "2-D structure of the a region of Xist RNA and its implication for PRC2 association," *PLoS Biology*, vol. 8, no. 1, 2010.
- [87] Smola M. J., Christy T. W., Inoue K., Nicholson C. O., Friedersdorf M., Keene J. D., Lee D. M., Calabrese J. M., Weeks K. M., "SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 37, p.p. 10322-10327, 2016.
- [88] Fang R., Moss W. N., Rutenberg-Schoenberg M., Simon M. D., "Probing Xist RNA Structure in Cells Using Targeted Structure-Seq," *PLoS Genetics*, vol. 11, no. 12, 2015.
- [89] Lu Z., Zhang Q. C., Lee B., Flynn R. A., Smith M. A., Robinson J. T., Davidovich C.,
  Gooding A. R., Goodrich K. J., Mattick J. S., Mesirov J. P., Cech T. R., Chang H. Y.,
  "RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure," *Cell*,
  vol. 165, no. 5, pp. 1267–1279, 2016.
- [90] Lu Z., Carter A. C., Chang H. Y., "Mechanistic insights in X-chromosome inactivation," *Philosophical Transactions of the Royal Society B: Biological Sciences*. vol. 372, no. 1733, p.p. 20160356, 2017.

- [91] Rivas E., Clements J., Eddy S. R., "A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs," *Nature Methods*, vol. 14, no. 1, p.p. 45-48, 2017.
- [92] Lv Q., Yuan L., Song Y., Sui T., Li Z., Lai L., "D-repeat in the XIST gene is required for X chromosome inactivation," *RNA Biology*, vol. 13, no. 2, pp. 172–176, 2016.
- [93] Caparros M. L., Alexiou M., Webster Z., Brockdorff N., "Functional analysis of the highly conserved exon IV of Xist RNA," *Cytogenetic and Genome Research*, vol. 99, no. 1–4. Cytogenet Genome Res, pp. 99–105, 2002.
- [94] Senner C. E., Nesterova T. B., Norton S., Dewchand H., Godwin J., Mak W., Brockdorff N., "Disruption of a conserved region of Xist exon 1 impairs Xist RNA localisation and X-linked gene silencing during random and imprinted X chromosome inactivation," *Development*, vol. 138, no. 8, pp. 1541–1550, 2011.
- [95] Engreitz J. M., Pandya-Jones A., McDonel P., Shishkin A., Sirokman K., Surka C., Kadri S., Xing J., Goren A., Lander E. S., Plath K., Guttman M., "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome.," *Science*, vol. 341, no. 6147, p. 1237973, 2013.
- [96] 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., Lee J. T., "High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation," *Nature*, vol. 504, no. 7480, pp. 465–469, 2013.
- [97] Darrow E. M., Huntley M. H., Dudchenko O., Stamenova E. K., Durand N. C., Sun Z.,
  Huang S. C., Sanborn A. L., Machol I., Shamim M., Seberg A. P., Lander E. S., Chadwick
  B. P., Aiden E. L., "Deletion of DXZ4 on the human inactive X chromosome alters

higher-order genome architecture," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 31, pp. E4504–E4512, 2016.

- [98] Jonkers I., Monkhorst K., Rentmeester E., Grootegoed J. A., Grosveld F., Gribnau J., "Xist RNA Is Confined to the Nuclear Territory of the Silenced X Chromosome throughout the Cell Cycle," *Molecular and Cellular Biology*, vol. 28, no. 18, pp. 5583– 5594, 2008.
- [99] Ng K., Daigle N., Bancaud A., Ohhata T., Humphreys P., Walker R., Ellenberg J., Wutz A., "A system for imaging the regulatory noncoding Xist RNA in living mouse embryonic stem cells," *Molecular Biology of the Cell*, vol. 22, no. 14, pp. 2634–2645, 2011.
- [100] Helbig R., Fackelmayer F. O., "Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosoe territories through its RGG domain," *Chromosoma*, 2003.
- [101] Fackelmayer F. O., "A stable proteinaceous structure in the territory of inactive X chromosomes," *Journal of Biological Chemistry*, vol. 280, no. 3, p.p. 1720-1723, 2005.
- [102] Hasegawa Y., Brockdorff N., Kawano S., Tsutui K., Tsutui K., Nakagawa S., "The matrix protein hnRNP U is required for chromosomal localization of Xist RNA.," *Developmental cell*, vol. 19, no. 3, pp. 469–76, 2010.
- [103] McHugh C. A., Chen C.-K. K., Chow A., Surka C. F., Tran C., McDonel P., Pandya-Jones A., Blanco M., Burghard C., Moradian A., Sweredoski M. J., Shishkin A. A., Su J., Lander E. S., Hess S., Plath K., Guttman M., "The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3," *Nature*, vol. 521, no. 7551, pp. 232–236, 2015.
- [104] Kolpa H. J., Fackelmayer F. O., Lawrence J. B., "SAF-A Requirement in Anchoring XIST RNA to Chromatin Varies in Transformed and Primary Cells," *Developmental Cell*,

vol. 39, no. 1. Cell Press, pp. 9-10, 10-Oct-2016.

- [105] Yamada N., Hasegawa Y., Yue M., Hamada T., Nakagawa S., Ogawa Y., "Xist Exon 7 Contributes to the Stable Localization of Xist RNA on the Inactive X-Chromosome," *PLoS Genetics*, vol. 11, no. 8, 2015.
- [106] Sakaguchi T., Hasegawa Y., Brockdorff N., Tsutsui K., Tsutsui K. M., Sado T., Nakagawa S., "Control of Chromosomal Localization of Xist by hnRNP U Family Molecules," *Developmental Cell.* vol. 39, no. 1, p.p. 11-12, 2016.
- [107] Ridings-Figueroa R., Stewart E. R., Nesterova T. B., Coker H., Pintacuda G., Godwin J., Wilson R., Haslam A., Lilley F., Ruigrok R., Bageghni S. A., Albadrani G., Mansfield W., Roulson J. A., Brockdorff N., Ainscough J. F. X., Coverley D., "The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory," *Genes and Development*, vol. 31, no. 9, pp. 876–888, 2017.
- [108] Sunwoo H., Colognori D., Froberg J. E., Jeon Y., Lee J. T., "Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 40, pp. 10654–10659, 2017.
- [109] Lu Z., Guo J. K., Wei Y., Dou D. R., Zarnegar B., Ma Q., Li R., Zhao Y., Liu F., Choudhry H., Khavari P. A., Chang H., "Structural modularity of the XIST ribonucleoprotein complex," *bioRxiv*, p. 837229, 2019.
- [110] Hall L. L., Byron M., Sakai K., Carrel L., Willard H. F., Lawrence J. B., "An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 13, pp. 8677–8682, 2002.

- [111] Minks J., Baldry S. El, Yang C., Cotton A. M., Brown C. J., "XIST-induced silencing of flanking genes is achieved by additive action of repeat a monomers in human somatic cells," *Epigenetics and Chromatin*, vol. 6, no. 1, p. 23, 2013.
- [112] Chow J. C., Hall L. L., Baldry S. E. L., Thorogood N. P., Lawrence J. B., Brown C. J., "Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 24, pp. 10104–10109, 2007.
- [113] Dossin F., Pinheiro I., Żylicz J. J., Roensch J., Collombet S., Le Saux A., Chelmicki T., Attia M., Kapoor V., Zhan Y., Dingli F., Loew D., Mercher T., Dekker J., Heard E., "SPEN integrates transcriptional and epigenetic control of X-inactivation," *Nature*, vol. 578, no. 7795, pp. 455–460, 2020.
- [114] Żylicz J. J., Bousard A., Žumer K., Dossin F., Mohammad E., da Rocha S. T., Schwalb B., Syx L., Dingli F., Loew D., Cramer P., Heard E., "The Implication of Early Chromatin Changes in X Chromosome Inactivation," *Cell*, vol. 176, no. 1–2, pp. 182-197.e23, 2019.
- [115] Nesterova T. B., Wei G., Coker H., Pintacuda G., Bowness J. S., Zhang T., Almeida M., Bloechl B., Moindrot B., Carter E. J., Rodrigo I. A., Pan Q., Bi Y., Song C. X., Brockdorff N., Alvarez Rodrigo I., Pan Q., ... Brockdorff N., "Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation," *Nature Communications*, vol. 10, no. 2019, pp. 1–15, 2019.
- [116] Patil D. P., Chen C. K., Pickering B. F., Chow A., Jackson C., Guttman M., Jaffrey S. R.,
  "M6A RNA methylation promotes XIST-mediated transcriptional repression," *Nature*,
  vol. 537, no. 7620, p.p. 369-373, 2016.
- [117] Coker H., Wei G., Moindrot B., Mohammed S., Nesterova T., Brockdorff N., "The role

of the Xist 5' m6A region and RBM15 in X chromosome inactivation," *Wellcome Open Research*, vol. 5, 2020.

- [118] Berletch J. B., Ma W., Yang F., Shendure J., Noble W. S., Disteche C. M., Deng X.,
  "Escape from X Inactivation Varies in Mouse Tissues," *PLoS Genetics*, vol. 11, no. 3, p.p. 1005079, 2015.
- [119] Balaton B. P., Cotton A. M., Brown C. J., "Derivation of consensus inactivation status for X-linked genes from genome-wide studies," *Biology of Sex Differences*, vol. 6, no. 35, 2015.
- [120] Carrel L., Willard H. F., "Heterogeneous gene expression from the inactive X chromosome: An X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 13, p.p. 7364-9, 1999.
- [121] Cotton A. M., Ge B., Light N., Adoue V., Pastinen T., Brown C. J., "Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome," *Genome Biology*, vol. 14, no. 11, p.p. R112, 2013.
- [122] Lyon M. F., "X-Chromosome inactivation: a repeat hypothesis," *Cytogenetic and Genome Research*, vol. 80, no. 1–4, pp. 133–137, 1998.
- [123] Boggs B. A., Connors B., Sobel R. E., Chinault A. C., Allis C. D., "Reduced levels of histone H3 acetylation on the inactive X chromosome in human females," *Chromosoma*, vol. 105, no. 5, pp. 303–309, 1996.
- [124] Keohane A. M., O'neill L. P., Belyaev N. D., Lavender J. S., Turner B. M., "X-Inactivation and histone H4 acetylation in embryonic stem cells.," *Developmental biology*, vol. 180, no. 2, pp. 618–30, 1996.

- [125] Wutz A., Jaenisch R., "A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation," *Molecular Cell*, vol. 5, no. 4, pp. 695–705, 2000.
- [126] Csankovszki G., Panning B., Bates B., Pehrson J. R., Jaenisch R., "Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation
  [1]," *Nature Genetics*, vol. 22, no. 4. Nature Publishing Group, pp. 323–324, 1999.
- [127] Pintacuda G., Wei G., Roustan C., Kirmizitas B. A., Solcan N., Cerase A., Castello A., Mohammed S., Moindrot B., Nesterova T. B., Brockdorff N., "hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing," *Molecular Cell*, vol. 68, no. 5, pp. 955-969.e10, 2017.
- [128] Schoeftner S., Sengupta A. K., Kubicek S., Mechtler K., Spahn L., Koseki H., Jenuwein T., Wutz A., "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing," *EMBO Journal*, vol. 25, no. 13, pp. 3110–3122, 2006.
- [129] Bousard A., Raposo A. C., Żylicz J. J., Picard C., Pires V. B., Qi Y., Gil C., Syx L., Chang H. Y., Heard E., da Rocha S. T., "The role of Xist -mediated Polycomb recruitment in the initiation of X-chromosome inactivation," *EMBO reports*, vol. 20, no. 10, 2019.
- [130] Plath K., Talbot D., Hamer K. M., Otte A. P., Yang T. P., Jaenisch R., Panning B.,
  "Developmentally regulated alterations in polycomb repressive complex 1 proteins on the inactive X chromosome," *Journal of Cell Biology*, vol. 167, no. 6, p.p. 1025-1035, 2004.
- [131] da Rocha S. T., Boeva V., Escamilla-Del-Arenal M., Ancelin K., Granier C., Matias N. R.
  R., Sanulli S., Chow J., Schulz E., Picard C., Kaneko S., Helin K., Reinberg D., Stewart
  A. F. F., Wutz A., Margueron R., Heard E., ... Heard E., "Jarid2 Is Implicated in the
  Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome," *Molecular Cell*,

vol. 53, no. 2, pp. 301-16, 2014.

- [132] Wang C. Y., Colognori D., Sunwoo H., Wang D., Lee J. T., "PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments," *Nature Communications*, vol. 10, no. 1, 2019.
- [133] Cooper S., Grijzenhout A., Underwood E., Ancelin K., Zhang T., Nesterova T. B., Anil-Kirmizitas B., Bassett A., Kooistra S. M., Agger K., Helin K., Heard E., Brockdorff N.,
  "Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2," *Nature Communications*, vol. 7, p.p. 13661, 2016.
- [134] Holoch D., Margueron R., "Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity," *Trends in Biochemical Sciences*, vol. 42, no. 7. pp. 531–542, 2017.
- [135] Kaneko S., Li G., Son J., Xu C.-F., Margueron R., Neubert T. A., Reinberg D.,
  "Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA.," *Genes & development*, vol. 24, no. 23, pp. 2615–20, 2010.
- [136] Yamamoto K., Sonoda M., Inokuchi J., Shirasawa S., Sasazuki T., "Polycomb Group Suppressor of Zeste 12 Links Heterochromatin Protein 1α and Enhancer of Zeste 2," *Journal of Biological Chemistry*, vol. 279, no. 1, p.p.401-6, 2004.
- [137] Margueron R., Reinberg D., "The Polycomb complex PRC2 and its mark in life," *Nature*. vol. 469, no. 7330, p.p. 343-9, 2011.
- [138] Gao Z., Zhang J., Bonasio R., Strino F., Sawai A., Parisi F., Kluger Y., Reinberg D., "PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes.," *Molecular cell*, vol. 45, no. 3, pp. 344–56, 2012.
- [139] Tavares L., Dimitrova E., Oxley D., Webster J., Poot R., Demmers J., Bezstarosti K., Taylor S., Ura H., Koide H., Wutz A., Vidal M., Elderkin S., Brockdorff N., "RYBP-

PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3," *Cell*, vol. 148, no. 4, pp. 664–678, 2012.

- [140] Almeida M., Pintacuda G., Masui O., Koseki Y., Gdula M., Cerase A., Brown D., Mould A., Innocent C., Nakayama M., Schermelleh L., Nesterova T. B., Koseki H., Brockdorff N., "PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation.," *Science*, vol. 356, no. 6342, pp. 1081–1084, 2017.
- [141] Wang J., Mager J., Chen Y., Schneider E., Cross J. C., Nagy A., Magnuson T.,
  "Imprinted X inactivation maintained by a mouse Polycomb group gene," *Nature Genetics*, vol. 28, no. 4, pp. 371–375, 2001.
- [142] Kalantry S., Magnuson T., "The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation.," *PLoS genetics*, vol. 2, no. 5, p. e66, 2006.
- [143] Kalantry S., Mills K. C., Yee D., Otte A. P., Panning B., Magnuson T., "The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation," *Nature Cell Biology*, vol. 8, no. 2, p.p. 195-202, 2006.
- [144] Colognori D., Sunwoo H., Kriz A. J., Wang C. Y., Lee J. T., "Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X," *Molecular Cell*, vol. 74, no. 1, pp. 101-117.e10, 2019.
- [145] Chadwick B. P., "Variation in Xi chromatin organization and correlation of the H3K27me3 chromatin territories to transcribed sequences by microarray analysis," *Chromosoma*, vol. 116, no. 2, pp. 147–157, 2007.
- [146] Beck D. B., Burton A., Oda H., Ziegler-Birling C., Torres-Padilla M. E., Reinberg D.,"The role of PR-Set7 in replication licensing depends on Suv4-20h," *Genes and*

Development, vol. 26, no.23, p.p. 2580-9, 2012.

- [147] Kapoor-Vazirani P., Vertino P. M., "A dual role for the histone methyltransferase PR-SET7/SETD8 and histone H4 lysine 20 monomethylation in the local regulation of RNA polymerase II pausing," *Journal of Biological Chemistry*, vol. 289, no. 11, pp. 7425–7437, 2014.
- [148] Pesavento J. J., Yang H., Kelleher N. L., Mizzen C. A., "Certain and Progressive Methylation of Histone H4 at Lysine 20 during the Cell Cycle," *Molecular and Cellular Biology*, vol. 28, no. 1, p.p. 468-486, 2008.
- [149] Trojer P., Li G., Sims R. J., Vaquero A., Kalakonda N., Boccuni P., Lee D., Erdjument-Bromage H., Tempst P., Nimer S. D., Wang Y. H., Reinberg D., "L3MBTL1, a Histone-Methylation-Dependent Chromatin Lock," *Cell*, vol. 129, no. 5, pp. 915–928, 2007.
- [150] Mermoud J. E., Popova B., Peters A. H. F. M., Jenuwein T., Brockdorff N., "Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation," *Current Biology*, vol. 12, no. 3, pp. 247–251, 2002.
- [151] Boggs B. A., Cheung P., Heard E., Spector D. L., Chinault A. C., Allis C. D.,
  "Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes," *Nature Genetics*, vol. 30, no. 1, pp. 73–76, 2002.
- [152] Minkovsky A., Sahakyan A., Rankin-Gee E., Bonora G., Patel S., Plath K., "The Mbd1-Atf7ip-Setdb1 pathway contributes to the maintenance of X chromosome inactivation," *Epigenetics and Chromatin*, vol. 7, no. 12, 2014.
- [153] Sun Z., Chadwick B. P., "Loss of SETDB1 decompacts the inactive X chromosome in part through reactivation of an enhancer in the IL1RAPL1 gene," *Epigenetics and Chromatin*, vol. 11, no. 1, p.p. 45, 2018.

- [154] Sado T., Fenner M. H., Tan S. S., Tam P., Shioda T., Li E., "X inactivation in the mouse embryo deficient for Dnmt1: Distinct effect of hypomethylation on imprinted and random X inactivation," *Developmental Biology*, vol., no., p.p., 2000.
- [155] Escamilla-Del-Arenal M., da Rocha S. T., Spruijt C. G., Masui O., Renaud O., Smits A. H., Margueron R., Vermeulen M., Heard E., "Cdyl, a New Partner of the Inactive X Chromosome and Potential Reader of H3K27me3 and H3K9me2," *Molecular and Cellular Biology*, vol. 225, no. 2, p.p. 294-303, 2013.
- [156] Mermoud J. E., Costanzi C., Pehrson J. R., Brockdorff N., "Histone MacroH2A1.2 Relocates to the Inactive X Chromosome after Initiation and Propagation of X-Inactivation," *The Journal of Cell Biology*, vol. 147, no. 7, pp. 1399–1408, 1999.
- [157] Minajigi A., Froberg J. E., Wei C., Sunwoo H., Kesner B., Colognori D., Lessing D., Payer B., Boukhali M., Haas W., Lee J. T., "A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation," *Science*, vol. 349, no. 6245, pp. aab2276–aab2276, 2015.
- [158] Chakravarthy S., Gundimella S. K. Y., Caron C., Perche P.-Y., Pehrson J. R., Khochbin S., Luger K., "Structural Characterization of the Histone Variant macroH2A," *Molecular and Cellular Biology*, vol. 25, no. 17, pp 7616-24, 2005.
- [159] Pehrson J. R., Fried V. A., "MacroH2A, a core histone containing a large nonhistone region," *Science*, vol. 257, no. 5075, pp. 1398–1400, 1992.
- [160] Gamble M. J., Frizzell K. M., Yang C., Krishnakumar R., Kraus W. L., "The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing," *Genes and Development*, vol. 24, no. 1, pp. 21–32, 2010.
- [161] Tanasijevic B., Rasmussen T. P., "X Chromosome inactivation and differentiation occur

readily in ES cells doubly-deficient for macroH2A1 and macroH2A2," *PLoS ONE*, vol. 6, no. 6, p. e21512, 2011.

- [162] Pehrson J. R., Changolkar L. N., Costanzi C., Leu N. A., "Mice without MacroH2A Histone Variants," *Molecular and Cellular Biology*, vol. 34, no. 24, pp. 4523–4533, 2014.
- [163] Ogawa Y., Ono T., Wakata Y., Okawa K., Tagami H., Shibahara K. I., "Histone variant macroH2A1.2 is mono-ubiquitinated at its histone domain," *Biochemical and Biophysical Research Communications*, vol. 366, no. 1, pp 204-9, 2005.
- [164] Chu F., Nusinow D. A., Chalkley R. J., Plath K., Panning B., Burlingame A. L.,
  "Mapping post-translational modifications of the histone variant MacroH2A1 using tandem mass spectrometry," *Molecular and Cellular Proteomics*, vol. 5, no. 1, pp 194-203, 2006.
- [165] Chen K., Xu Z. X., Jun J. S., Deavila J. M., Zhu M. J., Du M., "XIST Recruits BRCA1 to Ubiquitinate macroH2A for X Chromosome Inactivation in Proliferating Breast Cells," *SSRN Electronic Journal*, pre publication, 2019.
- [166] Wang C. Y., Jégu T., Chu H. P., Oh H. J., Lee J. T., "SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X," *Cell*, vol. 174, no. 2, pp. 406-421.e25, 2018.
- [167] Jansz N., Nesterova T., Keniry A., Iminitoff M., Hickey P. F., Pintacuda G., Masui O., Kobelke S., Geoghegan N., Breslin K. A., Willson T. A., Rogers K., Kay G. F., Fox A. H., Koseki H., Brockdorff N., Murphy J. M., Blewitt M. E., "Smchd1 Targeting to the Inactive X Is Dependent on the Xist-HnrnpK-PRC1 Pathway," *Cell Reports*, vol. 25, no. 7, pp. 1912-1923.e9, 2018.
- [168] Ohno S., Kaplan W. D., Kinosita R., "Somatic association of the positively

heteropycnotic X-chromosomes in female mice (Mus musculus)," *Experimental Cell Research*, vol. 15, no. 3, pp 616-8, 1958.

- [169] Padeken J., Heun P., "Nucleolus and nuclear periphery: Velcro for heterochromatin," *Current Opinion in Cell Biology*. vol. 28, pp 54-60, 2014.
- [170] Reddy K. L., Zullo J. M., Bertolino E., Singh H., "Transcriptional repression mediated by repositioning of genes to the nuclear lamina," *Nature*, vol. 452, no. 7184, pp 243-7, 2008.
- [171] Zhang L.-F. F., Huynh K. D., Lee J. T., "Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing.," *Cell*, vol. 129, no. 4, pp. 693–706, 2007.
- [172] Worman H. J., Yuan J., Blobel G., Georgatos S. D., "A lamin B receptor in the nuclear envelope," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 22, pp 8531-4, 1988.
- [173] Cohen T. V., Klarmann K. D., Sakchaisri K., Cooper J. P., Kuhns D., Anver M., Johnson P. F., Williams S. C., Keller J. R., Stewart C. L., "The lamin B receptor under transcriptional control of C/EBPε is required for morphological but not functional maturation of neutrophils," *Human Molecular Genetics*, vol. 17, no. 19, pp 2921-33, 2008.
- [174] Yang F., Deng X., Ma W., Berletch J. B., Rabaia N., Wei G., Moore J. M., Filippova G. N., Xu J., Liu Y., Noble W. S., Shendure J., Disteche C. M., "The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation.," *Genome biology*, vol. 16, no. 1, p. 52, 2015.
- [175] Fang H., Bonora G., Lewandowski J. P., Thakur J., Filippova G. N., Henikoff S., Shendure J., Duan Z., Rinn J. L., Deng X., Noble W. S., Disteche C. M., "Trans- and cisacting effects of the lncRNA Firre on epigenetic and structural features of the inactive X

chromosome," bioRxiv, p. 687236, 2019.

- [176] Giorgetti L., Lajoie B. R., Carter A. C., Attia M., Zhan Y., Xu J., Chen C. J., Kaplan N., Chang H. Y., Heard E., Dekker J., "Structural organization of the inactive X chromosome in the mouse," *Nature*, vol. 535, no. 7613, pp 575-9, 2016.
- [177] Bonora G., Deng X., Fang H., Ramani V., Qiu R., Berletch J. B., Filippova G. N., Duan Z., Shendure J., Noble W. S., Disteche C. M., "Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome," *Nature Communications*, vol. 9, no. 1, pp. 1–17, 2018.
- [178] de Wit E., Vos E. S. M., Holwerda S. J. B., Valdes-Quezada C., Verstegen M. J. A. M., Teunissen H., Splinter E., Wijchers P. J., Krijger P. H. L., de Laat W., "CTCF Binding Polarity Determines Chromatin Looping," *Molecular Cell*, vol. 60, no. 4, pp 676-84, 2015.
- [179] Lewandowski J. P., Lee J. C., Hwang T., Sunwoo H., Goldstein J. M., Groff A. F., Chang N. P., Mallard W., Williams A., Henao-Meija J., Flavell R. A., Lee J. T., Gerhardinger C., Wagers A. J., Rinn J. L., "The Firre locus produces a trans-acting RNA molecule that functions in hematopoiesis," *Nature Communications*, vol. 10, no. 1, pp 5137, 2019.
- [180] Andergassen D., Smith Z., Lewandowski J., Gerhardinger C., Meissner A., Rinn J., "In vivo Firre and Dxz4 deletion elucidates roles for autosomal gene regulation," *bioRxiv*, 2019.
- [181] Lyon M. F., Searle A. G., Ford C. E., Ohno S., "A Mouse Translocation Suppressing Sex-Linked Variegation," *Cytogenetic and Genome Research*, vol. 3, pp 306-322, 1964.
- [182] Horvath J. E., Sheedy C. B., Merrett S. L., Diallo A. B., Swofford D. L., Green E. D., Willard H. F., "Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate XIST locus," *Genome Research*, vol. 21, no. 6, pp

850-62, 2011.

- [183] Hwang J. Y., Choi K. H., Lee C. K., "Data for identification of porcine X-chromosome inactivation center, XIC, by genomic comparison with human and mouse XIC," *Data in Brief*, vol. 5, pp 1072-7, 2015.
- [184] Nora E. P., Lajoie B. R., Schulz E. G., Giorgetti L., Okamoto I., Servant N., Piolot T., Van Berkum N. L., Meisig J., Sedat J., Gribnau J., Barillot E., Blüthgen N., Dekker J., Heard E., "Spatial partitioning of the regulatory landscape of the X-inactivation centre," *Nature*, vol. 485, no. 7398, pp 381-5, 2012.
- [185] Furlan G., Gutierrez Hernandez N., Huret C., Galupa R., van Bemmel J. G., Romito A., Heard E., Morey C., Rougeulle C., "The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products," *Molecular Cell*, vol. 70, no. 3, pp. 462-472.e8, 2018.
- [186] Tian D., Sun S., Lee J. T., "The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation.," *Cell*, vol. 143, no. 3, pp. 390–403, 2010.
- [187] Barakat T. S., Loos F., Van Staveren S., Myronova E., Ghazvini M., Grootegoed J. A., Gribnau J., "The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing," *Molecular Cell*, vol. 53, no. 6, pp. 965–978, 2014.
- [188] van Bemmel J. G., Galupa R., Gard C., Servant N., Picard C., Davies J., Szempruch A. J., Zhan Y., Żylicz J. J., Nora E. P., Lameiras S., de Wit E., Gentien D., Baulande S., Giorgetti L., Guttman M., Hughes J. R., ... Heard E., "The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist," *Nature Genetics*, vol. 51, no. 6, pp. 1024–1034, 2019.
- [189] Galupa R., Nora E. P., Worsley-Hunt R., Picard C., Gard C., van Bemmel J. G., Servant
N., Zhan Y., El Marjou F., Johanneau C., Diabangouaya P., Le Saux A., Lameiras S.,
Pipoli da Fonseca J., Loos F., Gribnau J., Baulande S., ... Heard E., "A Conserved
Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological
Boundary," *Molecular Cell*, vol. 77, no. 2, pp 352-367, 2020.

- [190] Vallot C., Patrat C., Collier A. J., Huret C., Casanova M., Liyakat Ali T. M., Tosolini M., Frydman N., Heard E., Rugg-Gunn P. J., Rougeulle C., "XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development," *Cell Stem Cell*, vol. 20, no. 1, pp 102-111, 2017.
- [191] Vallot C., Huret C., Lesecque Y., Resch A., Oudrhiri N., Bennaceur A., Duret L., Rougeulle C., "XACT, a long non-coding transcript coating the active X chromosome in human pluripotent cells," *Epigenetics & Chromatin*, vol. 20, no. 1, pp 102-111, 2013.
- [192] Loda A., Brandsma J. H., Vassilev I., Servant N., Loos F., Amirnasr A., Splinter E., Barillot E., Poot R. A., Heard E., Gribnau J., "Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations," *Nature Communications*, vol. 8, no. 1, p. 690, 2017.
- [193] Cotton A. M., Chen C. Y., Lam L. L., Wasserman W. W., Kobor M. S., Brown C. J.,
  "Spread of X-chromosome inactivation into autosomal sequences: Role for DNA elements, chromatin features and chromosomal domains," *Human Molecular Genetics*, vol. 23, no. 5, pp. 1211–1223, 2014.
- [194] Bailey J. A., Carrel L., Chakravarti A., Eichler E. E., "Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 12, pp. 6634–6639, 2000.

- [195] Gartler S. M., Riggs A. D., "Mammalian X-Chromosome Inactivation," Annual Review of Genetics, vol. 17, no. 1, pp. 155–190, 1983.
- [196] Parish D. A., Vise P., Wichman H. A., Bull J. J., Baker R. J., "Distribution of LINEs and other repetitive elements in the karyotype of the bat Carollia: Implications for Xchromosome inactivation," *Cytogenetic and Genome Research*, vol. 96, no. 1–4, pp. 191– 197, 2002.
- [197] Chow J. C., Ciaudo C., Fazzari M. J., Mise N., Servant N., Glass J. L., Attreed M., Avner P., Wutz A., Barillot E., Greally J. M., Voinnet O., Heard E., "LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation.," *Cell*, vol. 141, no. 6, pp. 956–69, 2010.
- [198] Tannan N. B., Brahmachary M., Garg P., Borel C., Alnefaie R., Watson C. T., Simon Thomas N., Sharp A. J., "Dna methylation profiling in X;autosome translocations supports a role for L1 repeats in the spread of X chromosome inactivation," *Human Molecular Genetics*, vol. 23, no. 5, pp. 1224–1236, 2014.
- [199] Calabrese J. M., Sun W., Song L., Mugford J. W., Williams L., Yee D., Starmer J., Mieczkowski P., Crawford G. E., Magnuson T., "Site-specific silencing of regulatory elements as a mechanism of X inactivation.," *Cell*, vol. 151, no. 5, pp. 951–63, 2012.
- [200] Matsuno Y., Yamashita T., Wagatsuma M., Yamakage H., "Convergence in LINE-1 nucleotide variations can benefit redundantly forming triplexes with lncRNA in mammalian X-chromosome inactivation," *Mobile DNA*, vol. 10, no. 1, pp. 1–23, 2019.
- [201] Ding Z., Ni Y., Timmer S. W., Lee B. K., Battenhouse A., Louzada S., Yang F., Dunham I., Crawford G. E., Lieb J. D., Durbin R., Iyer V. R., Birney E., "Quantitative Genetics of CTCF Binding Reveal Local Sequence Effects and Different Modes of X-Chromosome

Association," *PLoS Genetics*, vol. 10, no. 11, pp 1004798, 2014.

- [202] 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., Kalkan T., Smith A., Jouneau A., de Laat W., Gribnau J., Stunnenberg H. G., "Dynamics of gene silencing during X inactivation using allele-specific RNA-seq," *Genome Biology*, vol. 16, no. 1, pp 149, 2015.
- [203] Marahrens Y., Loring J., Jaenisch R., "Role of the Xist Gene in X Chromosome Choosing," *Cell*, vol. 92, no. 5, pp. 657–664, 1998.
- [204] Yang L., Kirby J. E., Sunwoo H., Lee J. T., "Female mice lacking Xist RNA show partial dosage compensation and survive to term," *Genes and Development*, no. 30, pp. 1747– 1760, 2016.
- [205] Petropoulos S., Edsgärd D., Reinius B., Deng Q., Panula S. P., Codeluppi S., Plaza Reyes A., Linnarsson S., Sandberg R., Lanner F., "Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos," *Cell*, vol. 167, no. 1, pp 285, 2016.
- [206] Navarro P., Avner P., "An embryonic story: Analysis of the gene regulative network controlling Xist expression in mouse embryonic stem cells," *BioEssays*. vol. 32, no. 7, pp 581-8, 2010.
- [207] Agrelo R., Souabni A., Novatchkova M., Haslinger C., Leeb M., Komnenovic V., Kishimoto H., Gresh L., Kohwi-Shigematsu T., Kenner L., Wutz A., "SATB1 Defines the Developmental Context for Gene Silencing by Xist in Lymphoma and Embryonic Cells," *Developmental Cell*, vol. 16, no. 4, pp. 507–516, 2009.
- [208] Nesterova T. B., Senner C. E., Schneider J., Alcayna-Stevens T., Tattermusch A., Hemberger M., Brockdorff N., "Pluripotency factor binding and Tsix expression act

synergistically to repress Xist in undifferentiated embryonic stem cells," *Epigenetics and Chromatin*, vol. 4, no. 7, pp 17, 2011.

- [209] Donohoe M. E., Silva S. S., Pinter S. F., Xu N., Lee J. T., "The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting," *Nature*, vol. 460, no. 7251, pp 128-32, 2009.
- [210] Lee H. J., Gopalappa R., Sunwoo H., Choi S. W., Ramakrishna S., Lee J. T., Kim H. H., Nam J. W., "En bloc and segmental deletions of human XIST reveal X chromosome inactivation-involving RNA elements," *Nucleic Acids Research*, vol. 47, no. 8, pp. 3875– 3887, 2019.
- [211] Chiang J. C., Jiang J., Newburger P. E., Lawrence J. B., "Trisomy silencing by XIST normalizes Down syndrome cell pathogenesis demonstrated for hematopoietic defects in vitro," *Nature Communications*, vol. 9, no. 1, pp. 1–12, 2018.
- [212] Czermiński J. T., Lawrence J. B., "Silencing Trisomy 21 with XIST in Neural Stem Cells Promotes Neuronal Differentiation," *Developmental Cell*, vol. 52, no. 3, pp 294-308, 2020.
- [213] Shen Y., Matsuno Y., Fouse S. D., Rao N., Root S., Xu R., Pellegrini M., Riggs A. D., Fan G., "X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 12, pp 4709-14, 2008.
- [214] Tchieu J., Kuoy E., Chin M. H., Trinh H., Patterson M., Sherman S. P., Aimiuwu O.,
  Lindgren A., Hakimian S., Zack J. A., Clark A. T., Pyle A. D., Lowry W. E., Plath K.,
  "Female human iPSCs retain an inactive X chromosome," *Cell Stem Cell*, vol., no., pp, 2010.

- [215] Patel S., Bonora G., Sahakyan A., Kim R., Chronis C., Langerman J., Fitz-Gibbon S.,
  Rubbi L., Skelton R. J. P., Ardehali R., Pellegrini M., Lowry W. E., Clark A. T., Plath K.,
  "Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during
  Differentiation," *Cell Reports*, vol. 18, no. 1, pp 54-67, 2017.
- [216] Cantone I., Dharmalingam G., Chan Y. W., Kohler A. C., Lenhard B., Merkenschlager M., Fisher A. G., "Allele-specific analysis of cell fusion-mediated pluripotent reprograming reveals distinct and predictive susceptibilities of human X-linked genes to reactivation," *Genome Biology*, vol. 18, no. 1, p. 2, 2017.
- [217] Geens M., Chuva De Sousa Lopes S. M., "X chromosome inactivation in human pluripotent stem cells as a model for human development: back to the drawing board?," *Human reproduction update*, vol. 23, no. 5. pp. 520–532, 2017.
- [218] Gossen M., Bujard H., "Tight control of gene expression in mammalian cells by tetracycline-responsive promoters," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 12, pp. 5547–5551, 1992.
- [219] Yan C., Boyd D. D., "Histone H3 Acetylation and H3 K4 Methylation Define Distinct Chromatin Regions Permissive for Transgene Expression," *Molecular and Cellular Biology*, vol. 26, no. 17, pp. 6357–6371, 2006.
- [220] Hinrichs W., Kisker C., Düvel M., Müller A., Tovar K., Hillen W., Saenger W., "Structure of the tet repressor-tetracycline complex and regulation of antibiotic resistance," *Science*, vol. 264, no. 5157, pp. 418–420, 1994.
- [221] Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.
   A., Horvath P., "CRISPR provides acquired resistance against viruses in prokaryotes," *Science*, vol. 315, no. 5819, pp. 1709–1712, 2007.

- [222] Bhaya D., Davison M., Barrangou R., "CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation," *Annual Review of Genetics*, vol. 45, no. 1, pp. 273–297, 2011.
- [223] Brouns S. J. J., Jore M. M., Lundgren M., Westra E. R., Slijkhuis R. J. H., Snijders A. P. L., Dickman M. J., Makarova K. S., Koonin E. V., Van Der Oost J., "Small CRISPR RNAs guide antiviral defense in prokaryotes," *Science*, vol. 321, no. 5891, pp. 960–964, 2008.
- [224] Cong L., Ran F. A., Cox D., Lin S., Barretto R., Habib N., Hsu P. D., Wu X., Jiang W., Marraffini L. A., Zhang F., "Multiplex genome engineering using CRISPR/Cas systems," *Science*, vol. 339, no. 6121, pp. 819–823, 2013.
- [225] Ran F. A., Hsu P. D., Wright J., Agarwala V., Scott D. A., Zhang F., "Genome engineering using the CRISPR-Cas9 system," *Nature Protocols*, vol. 8, no. 11, pp. 2281– 2308, 2013.
- [226] Jinek M., East A., Cheng A., Lin S., Ma E., Doudna J., "RNA-programmed genome editing in human cells," *eLife*, vol. 2013, no. 2, 2013.
- [227] Karvelis T., Gasiunas G., Miksys A., Barrangou R., Horvath P., Siksnys V., "crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus," *RNA Biology*, vol. 10, no. 5, pp. 841–851, 2013.
- [228] Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science*, vol. 337, no. 6096, pp. 816–821, 2012.
- [229] Hu J. H., Miller S. M., Geurts M. H., Tang W., Chen L., Sun N., Zeina C. M., Gao X., Rees H. A., Lin Z., Liu D. R., "Evolved Cas9 variants with broad PAM compatibility and

high DNA specificity," Nature, vol. 556, no. 7699, pp. 57-63, 2018.

- [230] Hirano S., Nishimasu H., Ishitani R., Nureki O., "Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9," *Molecular Cell*, vol. 61, no. 6, pp. 886–894, 2016.
- [231] Anders C., Niewoehner O., Duerst A., Jinek M., "Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease," *Nature*, vol. 513, no. 7519, pp. 569– 573, 2014.
- [232] Gasiunas G., Barrangou R., Horvath P., Siksnys V., "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 39, 2012.
- [233] Mali P., Yang L., Esvelt K. M., Aach J., Guell M., DiCarlo J. E., Norville J. E., Church G. M., "RNA-guided human genome engineering via Cas9," *Science*, vol. 339, no. 6121, pp. 823–826, 2013.
- [234] Jinek M., Jiang F., Taylor D. W., Sternberg S. H., Kaya E., Ma E., Anders C., Hauer M., Zhou K., Lin S., Kaplan M., Iavarone A. T., Charpentier E., Nogales E., Doudna J. A., "Structures of Cas9 endonucleases reveal RNA-mediated conformational activation," *Science*, vol. 343, no. 6176, 2014.
- [235] Hsu P. D., Lander E. S., Zhang F., "Development and applications of CRISPR-Cas9 for genome engineering," *Cell*, vol. 157, no. 6. Cell Press, pp. 1262–1278, 05-Jun-2014.
- [236] Paquet D., Kwart D., Chen A., Sproul A., Jacob S., Teo S., Olsen K. M., Gregg A., Noggle S., Tessier-Lavigne M., "Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9," *Nature*, vol. 533, no. 7601, pp. 125–129,

2016.

- [237] Fu Y., Sander J. D., Reyon D., Cascio V. M., Joung J. K., "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs," *Nature Biotechnology*, vol. 32, no. 3, pp. 279–284, 2014.
- [238] Cho S. W., Kim S., Kim Y., Kweon J., Kim H. S., Bae S., Kim J. S., "Analysis of offtarget effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases," *Genome Research*, vol. 24, no. 1, pp. 132–141, 2014.
- [239] Tsai S. Q., Zheng Z., Nguyen N. T., Liebers M., Topkar V. V., Thapar V., Wyvekens N., Khayter C., Iafrate A. J., Le L. P., Aryee M. J., Joung J. K., "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases," *Nature Biotechnology*, vol. 33, no. 2, pp. 187–198, 2015.
- [240] Gabriel R., Von Kalle C., Schmidt M., "Mapping the precision of genome editing," *Nature Biotechnology*, vol. 33, no. 2. Nature Publishing Group, pp. 150–152, Feb-2015.
- [241] Doench J. G., Hartenian E., Graham D. B., Tothova Z., Hegde M., Smith I., Sullender M., Ebert B. L., Xavier R. J., Root D. E., "Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation," *Nature Biotechnology*, vol. 32, no. 12, pp. 1262–1267, 2014.
- [242] Ryan D. E., Taussig D., Steinfeld I., Phadnis S. M., Lunstad B. D., Singh M., Vuong X., Okochi K. D., McCaffrey R., Olesiak M., Roy S., Yung C. W., Curry B., Sampson J. R., Bruhn L., Dellinger D. J., "Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs," *Nucleic Acids Research*, vol. 46, no. 2, pp. 792– 803, 2018.
- [243] Yin H., Song C. Q., Suresh S., Kwan S. Y., Wu Q., Walsh S., Ding J., Bogorad R. L., Zhu 309

L. J., Wolfe S. A., Koteliansky V., Xue W., Langer R., Anderson D. G., "Partial DNAguided Cas9 enables genome editing with reduced off-target activity," *Nature Chemical Biology*, vol. 14, no. 3, pp. 311–316, 2018.

- [244] Cromwell C. R., Sung K., Park J., Krysler A. R., Jovel J., Kim S. K., Hubbard B. P., "Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity," *Nature Communications*, vol. 9, no. 1, 2018.
- [245] Pfaffl M. W., "A new mathematical model for relative quantification in real-time RT-PCR," *Nucleic Acids Research*, vol. 29, no. 9, pp. 45e – 45, 2001.
- [246] Hellemans J., Mortier G., De Paepe A., Speleman F., Vandesompele J., "qBase relative quantification framework and software for management and automated analysis of realtime quantitative PCR data," *Genome Biology*, vol. 8, no. 2, 2008.
- [247] Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F., "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.," *Genome biology*, vol. 3, no. 7, 2002.
- [248] Jensen M. A., Fukushima M., Davis R. W., "DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis," *PLoS ONE*, vol. 5, no. 6, p. e11024, 2010.
- [249] McKnight P. E., Najab J., "Mann-Whitney U Test," in *The Corsini Encyclopedia of Psychology*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010, pp. 1–1.
- [250] Dunn O. J., "Multiple Comparisons among Means," *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52–64, 1961.
- [251] "RIPA Lysis Buffer," *Cold Spring Harbor Protocols*, vol. 2017, no. 12, p. pdb.rec101428, 2017.

- [252] "SDS gel-loading buffer (2X)," Cold Spring Harbor Protocols, 2006.
- [253] Perez-Pinera P., Kocak D. D., Vockley C. M., Adler A. F., Kabadi A. M., Polstein L. R., Thakore P. I., Glass K. A., Ousterout D. G., Leong K. W., Guilak F., Crawford G. E., Reddy T. E., Gersbach C. A., "RNA-guided gene activation by CRISPR-Cas9-based transcription factors," *Nature Methods*, vol. 10, no. 10, pp. 973–976, 2013.
- [254] Xu H., Xiao T., Chen C. H., Li W., Meyer C. A., Wu Q., Wu D., Cong L., Zhang F., Liu J.
   S., Brown M., Liu X. S., "Sequence determinants of improved CRISPR sgRNA design," *Genome Research*, vol. 25, no. 8, pp. 1147–1157, 2015.
- [255] "SOC Medium," Cold Spring Harbor Protocols, vol. 2018, no. 3, p. pdb.rec098863, 2018.
- [256] "LB agar," Cold Spring Harbor Protocols, vol. 2009, no. 3, pp. pdb.rec11683pdb.rec11683, 2009.
- [257] Kim S., Kim D., Cho S. W., Kim J., Kim J. S., "Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins," *Genome Research*, vol. 24, no. 6, pp. 1012-1019. 2014.
- [258] Senís E., Fatouros C., Große S., Wiedtke E., Niopek D., Mueller A. K., Börner K., Grimm D., "CRISPR/Cas9-mediated genome engineering: An adeno-associated viral (AAV) vector toolbox," *Biotechnology Journal*, vol. 9, no. 11, pp. 1402–1412, 2014.
- [259] Fisher R. A., "On the Interpretation of χ 2 from Contingency Tables, and the Calculation of P," *Journal of the Royal Statistical Society*, vol. 85, no. 1, p. 87, 1922.
- [260] Mcleod S., "Z-Score: Definition, Calculation and Interpretation," *Simply Psychology*, vol. 1, pp. 113. 2019.
- [261] Low E., Brase C. H., Brase C. P., "Understanding Basic Statistics," *The American Statistician*, 6<sup>th</sup> edition, pp. 332. 1998.

- [262] Wang M., Lin F., Xing K., Liu L., "Random X-chromosome inactivation dynamics in vivo by single-cell RNA sequencing," *BMC Genomics*, vol. 18, no. 1, p. 90, 2017.
- [263] Shi Y., Downes M., Xie W., Kao H. Y., Ordentlich P., Tsai C. C., Hon M., Evans R. M., "Sharp, an inducible cofactor that integrates nuclear receptor repression and activation," *Genes and Development*, vol. 15, no. 9, pp. 1140–1151, 2001.
- [264] Lee J. T., Jaenisch R., "Long-range cis effects of ectopic X-inactivation centres on a mouse autosome," *Nature*, vol. 386, no. 6622, pp. 275–279, 1997.
- [265] Horakova A. H., Moseley S. C., McLaughlin C. R., Tremblay D. C., Chadwick B. P.,
  "The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome.," *Human molecular genetics*, vol. 21, no. 20, pp. 4367–77, 2012.
- [266] Rastan S., "Timing of X-chromosome inactivation in postimplantation mouse embryos," *Development*, vol. 71, no. 1, pp. 11–24, 1982.
- [267] Cantone I., Bagci H., Dormann D., Dharmalingam G., Nesterova T., Brockdorff N., Rougeulle C., Vallot C., Heard E., Chaligne R., Merkenschlager M., Fisher A. G., "Ordered chromatin changes and human X chromosome reactivation by cell fusionmediated pluripotent reprogramming," *Nature Communications*, vol. 7, no. 1, pp. 1–12, 2016.
- [268] Kelsey A. D., Yang C., Leung D., Minks J., Dixon-McDougall T., Baldry S. E. L., Bogutz A. B., Lefebvre L., Brown C. J., "Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST," *Genome Biology*, vol. 16, no. 1, 2015.
- [269] Lieberman-Aiden E., Van Berkum N. L., Williams L., Imakaev M., Ragoczy T., Telling

A., Amit I., Lajoie B. R., Sabo P. J., Dorschner M. O., Sandstrom R., Bernstein B., Bender M. A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L. A., ... Dekker J.,
"Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *Science*, vol. 326, no. 5950, pp. 289–293, 2009.

- [270] Hung T., Chang H. Y., "Long noncoding RNA in genome regulation: Prospects and mechanisms," *RNA Biology*, vol. 7, no. 5. Taylor and Francis Inc., pp. 582–585, 2010.
- [271] Rinn J. L., Chang H. Y., "Genome Regulation by Long Noncoding RNAs," Annual Review of Biochemistry, vol. 81, no. 1, pp. 145–166, 2012.
- [272] Makhlouf M., Ouimette J.-F. F., Oldfield A., Navarro P., Neuillet D., Rougeulle C., "A prominent and conserved role for YY1 in Xist transcriptional activation.," *Nature communications*, vol. 5, p. 4878, 2014.
- [273] Steyer B., Bu Q., Cory E., Jiang K., Duong S., Sinha D., Steltzer S., Gamm D., Chang Q., Saha K., "Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin Selection," *Stem Cell Reports*, vol. 10, no. 2, pp. 642–654, 2018.
- [274] Liang X., Potter J., Kumar S., Zou Y., Quintanilla R., Sridharan M., Carte J., Chen W., Roark N., Ranganathan S., Ravinder N., Chesnut J. D., "Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection.," *Journal of biotechnology*, vol. 208, pp. 44–53, 2015.
- [275] Dow L. E., Fisher J., O'Rourke K. P., Muley A., Kastenhuber E. R., Livshits G., Tschaharganeh D. F., Socci N. D., Lowe S. W., "Inducible in vivo genome editing with CRISPR-Cas9," *Nature Biotechnology*, vol. 33, no. 4, pp. 390–394, 2015.
- [276] Platt R. J., Chen S., Zhou Y., Yim M. J., Swiech L., Kempton H. R., Dahlman J. E.,Parnas O., Eisenhaure T. M., Jovanovic M., Graham D. B., Jhunjhunwala S., Heidenreich

M., Xavier R. J., Langer R., Anderson D. G., Hacohen N., ... Zhang F., "CRISPR-Cas9 knockin mice for genome editing and cancer modeling," *Cell*, vol. 159, no. 2, pp. 440–455, 2014.

- [277] Sluch V. M., Chamling X., Wenger C., Duan Y., Rice D. S., Zack D. J., "Highly efficient scarless knock-in of reporter genes into human and mouse pluripotent stem cells via transient antibiotic selection," *PLoS ONE*, vol. 13, no. 11, 2018.
- [278] Guo Q., Mintier G., Ma-Edmonds M., Storton D., Wang X., Xiao X., Kienzle B., Zhao D., Feder J. N., "Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells," *Scientific Reports*, vol. 8, no. 1, 2018.
- [279] Brinkman E. K., Chen T., de Haas M., Holland H. A., Akhtar W., van Steensel B.,
   "Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks," *Molecular Cell*, vol. 70, no. 5, pp. 801-813.e6, 2018.
- [280] Mayr C., Bartel D. P., "Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells," *Cell*, vol. 138, no. 4, pp. 673–684, 2009.
- [281] Shi H., Wei J., He C., "Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers," *Molecular Cell*. vol. 74, no. 4, 640-650.
   2019.
- [282] Boo S. H., Kim Y. K., "The emerging role of RNA modifications in the regulation of mRNA stability," *Experimental and Molecular Medicine*. vol 52, no. 3, pp. 400-408, 2020.
- [283] 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., de Laat W., "The inactive X chromosome adopts

a unique three-dimensional conformation that is dependent on Xist RNA.," *Genes & development*, vol. 25, no. 13, pp. 1371–83, 2011.

- [284] Liu F., Somarowthu S., Pyle A. M., "Visualizing the secondary and tertiary architectural domains of lncRNA RepA," *Nature Chemical Biology*, vol. 13, no. 3, pp. 282–289, 2017.
- [285] Hoki Y., Kimura N., Kanbayashi M., Amakawa Y., Ohhata T., Sasaki H., Sado T., "A proximal conserved repeat in the Xist gene is essential as a genomic element for Xinactivation in mouse," *Development*, vol. 136, no. 1, pp. 139–146, 2009.
- [286] Hernández-Muñoz I., Lund A. H., van der Stoop P., Boutsma E., Muijrers I., Verhoeven E., Nusinow D. A., Panning B., Marahrens Y., van Lohuizen M., "Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 21, pp. 7635–40, 2005.
- [287] Gdula M., Nesterova T., Pintacuda G., Godwin J., Zhan Y., Ozadam H., McClellan M., Moralli D., Krueger F., Green C., Reik W., Kriaucionis S., Heard E., Dekker J., Brockdorff N., "The non-canonical SMC protein SmcHD1 antagonises TAD formation on the inactive X chromosome," *bioRxiv*, p. 342147, 2018.
- [288] Chow J., Heard E., "X inactivation and the complexities of silencing a sex chromosome," *Current opinion in cell biology*, vol. 21, no. 3, pp. 359–66, 2009.
- [289] Yue M., Ogawa A., Yamada N., Charles Richard J. L., Barski A., Ogawa Y., "Xist RNA repeat E is essential for ASH2L recruitment to the inactive X and regulates histone modifications and escape gene expression," *PLoS Genetics*, vol. 13, no. 7, 2017.
- [290] Csankovszki G., Nagy A., Jaenisch R., "Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation," *Journal of Cell*

*Biology*, vol. 153, no. 4, pp. 773–783, 2001.

- [291] Perner J., Lasserre J., Kinkley S., Vingron M., Chung H. R., "Inference of interactions between Chromatin modifiers and histone modifications: From ChIP-Seq data to chromatin-signaling," *Nucleic Acids Research*, vol. 42, no. 22, pp. 13689–13695, 2014.
- [292] Tsuji N., Kobayashi M., Nagashima K., Wakisaka Y., Koizumi K., "A new antifungal antibiotic, trichostatin," *The Journal of Antibiotics*, vol. 29, no. 1, pp. 1–6, 1976.
- [293] Yoshida M., Kijima M., Akita M., Beppu T., "Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A," *Journal of Biological Chemistry*, vol. 265, no. 28, pp. 17174–17179, 1990.
- [294] Vigushin D. M., Ali S., Pace P. E., Mirsaidi N., Ito K., Adcock I., Coombes R. C.,
  "Trichostatin A is a histone deacetylase inhibitor with potent antitumor activity against breast cancer in vivo," *Clinical Cancer Research*, vol. 7, no. 4, pp. 971–976, 2001.
- [295] Yang F., Zhang L., Li J., Huang J., Wen R., Ma L., Zhou D., Li L., "Trichostatin A and 5azacytidine both cause an increase in global histone H4 acetylation and a decrease in global DNA and H3K9 methylation during mitosis in maize.," *BMC plant biology*, vol. 10, no. 1, p. 178, 2010.
- [296] Chatterjee N., Wang W. L. W., Conklin T., Chittur S., Tenniswood M., "Histone deacetylase inhibitors modulate miRNA and mRNA expression, block metaphase, and induce apoptosis in inflammatory breast cancer cells," *Cancer Biology and Therapy*, vol. 14, no. 7, pp. 658–671, 2013.
- [297] Han B., Li J., Li Z., Guo L., Wang S., Liu P., Wu Y., "Trichostatin a stabilizes the expression of pluripotent genes in human mesenchymal stem cells during ex vivo expansion," *PLoS ONE*, vol. 8, no. 11, 2013.

- [298] You B. R., Park W. H., "Trichostatin A induces apoptotic cell death of HeLa cells in a Bcl-2 and oxidative stress-dependent manner," *International Journal of Oncology*, vol. 42, no. 1, pp. 359–366, 2013.
- [299] Malvaez M., McQuown S. C., Rogge G. A., Astarabadi M., Jacques V., Carreiro S., Rusche J. R., Wood M. A., "HDAC3-selective inhibitor enhances extinction of cocaineseeking behavior in a persistent manner," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 7, pp. 2647–2652, 2013.
- [300] Matthews G. M., Mehdipour P., Cluse L. A., Falkenberg K. J., Wang E., Roth M., Santoro F., Vidacs E., Stanley K., House C. M., Rusche J. R., Vakoc C. R., Zuber J., Minucci S., Johnstone R. W., "Functional-genetic dissection of HDAC dependencies in mouse lymphoid and myeloid malignancies," *Blood*, vol. 126, no. 21, pp. 2392–2403, 2015.
- [301] Wells C. E., Bhaskara S., Stengel K. R., Zhao Y., Sirbu B., Chagot B., Cortez D., Khabele D., Chazin W. J., Cooper A., Jacques V., Rusche J., Eischen C. M., McGirt L. Y., Hiebert S. W., "Inhibition of Histone Deacetylase 3 Causes Replication Stress in Cutaneous T Cell Lymphoma," *PLoS ONE*, vol. 8, no. 7, 2013.
- [302] Alchanati I., Teicher C., Cohen G., Shemesh V., Barr H. M., Nakache P., Ben-Avraham D., Idelevich A., Angel I., Livnah N., Tuvia S., Reiss Y., Taglicht D., Erez O., "The E3 ubiquitin-ligase Bmi1/Ring1A controls the proteasomal degradation of Top2α cleavage complex A potentially new drug target," *PLoS ONE*, vol. 4, no. 12, 2009.
- [303] Ismail I. H., McDonald D., Strickfaden H., Xu Z., Hendzel M. J., "A small molecule inhibitor of polycomb repressive complex 1 inhibits ubiquitin signaling at DNA doublestrand breaks," *Journal of Biological Chemistry*, vol. 288, no. 37, pp. 26944–26954, 2013.
- [304] Verma S. K., Tian X., Lafrance L. V., Duquenne C., Suarez D. P., Newlander K. A.,

Romeril S. P., Burgess J. L., Grant S. W., Brackley J. A., Graves A. P., Scherzer D. A., Shu A., Thompson C., Ott H. M., Van Aller G. S., MacHutta C. A., ... Miller W. H., "Identification of potent, selective, cell-Active inhibitors of the histone lysine methyltransferase EZH2," *ACS Medicinal Chemistry Letters*, vol. 3, no. 12, pp. 1091– 1096, 2012.

- [305] Ding M., Zhang H., Li Z., Wang C., Chen J., Shi L., Xu D., Gao Y., "The polycomb group protein enhancer of zeste 2 is a novel therapeutic target for cervical cancer," *Clinical and Experimental Pharmacology and Physiology*, vol. 42, no. 5, pp. 458–464, 2015.
- [306] Chiba T., Yokosuka O., Fukai K., Kojima H., Tada M., Arai M., Imazeki F., Saisho H.,
  "Cell growth inhibition and gene expression induced by the histone deacetylase inhibitor,
  trichostatin A, on human hepatoma cells," *Oncology*, vol. 66, no. 6, pp. 481–491, 2004.
- [307] Blackledge N. P., Farcas A. M., Kondo T., King H. W., McGouran J. F., Hanssen L. L. P., Ito S., Cooper S., Kondo K., Koseki Y., Ishikura T., Long H. K., Sheahan T. W., Brockdorff N., Kessler B. M., Koseki H., Klose R. J., "Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation," *Cell*, vol. 157, no. 6, pp. 1445–1459, 2014.
- [308] Wagener N., Holland D., Bulkescher J., Crnković-Mertens I., Hoppe-Seyler K., Zentgraf H., Pritsch M., Buse S., Pfitzenmaier J., Haferkamp A., Hohenfellner M., Hoppe-Seyler F., "The enhancer of zeste homolog 2 gene contributes to cell proliferation and apoptosis resistance in renal cell carcinoma cells," *International Journal of Cancer*, vol. 123, no. 7, pp. 1545–1550, 2008.
- [309] Tan J., Yang X., Zhuang L., Jiang X., Chen W., Puay L. L., Karuturi R. K. M., Tan P. B.

O., Liu E. T., Yu Q., "Pharmacologic disruption of polycomb-repressive complex 2mediated gene repression selectively induces apoptosis in cancer cells," *Genes and Development*, vol. 21, no. 9, pp. 1050–1063, 2007.

- [310] Ariës I. M., Bodaar K., Karim S. A., Chonghaile T. N., Hinze L., Burns M. A., Pfirrmann M., Degar J., Landrigan J. T., Balbach S., Peirs S., Menten B., Isenhart R., Stevenson K. E., Neuberg D. S., Devidas M., Loh M. L., ... Gutierrez A., "PRC2 loss induces chemoresistance by repressing apoptosis in T cell acute lymphoblastic leukemia," *Journal of Experimental Medicine*, vol. 215, no. 12, pp. 3094–3114, 2018.
- [311] Li J., Dallmayer M., Kirchner T., Musa J., Grünewald T. G. P., "PRC1: Linking Cytokinesis, Chromosomal Instability, and Cancer Evolution," *Trends in Cancer*, vol. 4, no. 1. Cell Press, pp. 59–73, 01-Jan-2018.
- [312] Hosogane M., Funayama R., Shirota M., Nakayama K., "Lack of Transcription Triggers H3K27me3 Accumulation in the Gene Body," *Cell Reports*, vol. 16, no. 3, pp. 696–706, 2016.
- [313] Takashina T., Kinoshita I., Kikuchi J., Shimizu Y., Sakakibara-Konishi J., Oizumi S., Nishimura M., Dosaka-Akita H., "Combined inhibition of EZH2 and histone deacetylases as a potential epigenetic therapy for non-small-cell lung cancer cells," *Cancer Science*, vol. 107, no. 7, pp. 955–962, 2016.
- [314] Kalle A. M., Wang Z., "Differential effects of two HDAC inhibitors with distinct concomitant DNA hypermethylation or hypomethylation in breast cancer cells," *bioRxiv*, p. 578062, 2019.
- [315] Vermeulen M., Carrozza M. J., Lasonder E., Workman J. L., Logie C., Stunnenberg H.G., "In Vitro Targeting Reveals Intrinsic Histone Tail Specificity of the Sin3/Histone

Deacetylase and N-CoR/SMRT Corepressor Complexes," *Molecular and Cellular Biology*, vol. 24, no. 6, pp. 2364–2372, 2004.

- [316] Hu E., Chen Z., Fredrickson T., Zhu Y., Kirkpatrick R., Zhang G. F., Johanson K., Sung C. M., Liu R., Winkler J., "Cloning and characterization of a novel human class I histone deacetylase that functions as a transcription repressor," *Journal of Biological Chemistry*, vol. 275, no. 20, pp. 15254–15264, 2000.
- [317] Karagianni P., Wong J., "HDAC3: Taking the SMRT-N-CoRrect road to repression," Oncogene, vol. 26, no. 37. pp. 5439–5449, 13-Aug-2007.
- [318] Johnson C. A., White D. A., Lavender J. S., O'Neill L. P., Turner B. M., "Human class I histone deacetylase complexes show enhanced catalytic activity in the presence of ATP and co-immunoprecipitate with the ATP-dependent chaperone protein Hsp70," *Journal of Biological Chemistry*, vol. 277, no. 11, pp. 9590–9597, 2002.
- [319] Azechi T., Kanehira D., Kobayashi T., Sudo R., Nishimura A., Sato F., Wachi H., "Trichostatin A, an HDAC class I/II inhibitor, promotes Pi-induced vascular calcification via up-regulation of the expression of alkaline phosphatase," *Journal of Atherosclerosis* and Thrombosis, vol. 20, no. 6, pp. 538–547, 2013.
- [320] Wang C. Y., Colognori D., Sunwoo H., Wang D., Lee J. T., "PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments," *Nature Communications*, vol. 10, no. 1, pp 2950, 2019.
- [321] Eskeland R., Freyer E., Leeb M., Wutz A., Bickmore W. A., "Histone acetylation and the maintenance of chromatin compaction by Polycomb repressive complexes," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 75, pp. 71–78, 2010.
- [322] Migeon B. R., Kazi E., Haisley-Royster C., Hu J., Reeves R., Call L., Lawler A., Moore

C. S., Morrison H., Jeppesen P., "Human X inactivation center induces random X chromosome inactivation in male transgenic mice," *Genomics*, vol. 59, no. 2, pp. 113–121, 1999.

- [323] Gdula M. R., Nesterova T. B., Pintacuda G., Godwin J., Zhan Y., Ozadam H., McClellan M., Moralli D., Krueger F., Green C. M., Reik W., Kriaucionis S., Heard E., Dekker J., Brockdorff N., "The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome," *Nature Communications*, 2019.
- [324] Chen Q., Gao S., He W., Kou X., Zhao Y., Wang H., Gao S., "Xist repression shows time-dependent effects on the reprogramming of female somatic cells to induced pluripotent stem cells," *Stem Cells*, vol. 32, no. 10, pp. 2642–2656, 2014.
- [325] Blum G., Ibáñez G., Rao X., Shum D., Radu C., Djaballah H., Rice J. C., Luo M.,"Small-molecule inhibitors of setd8 with cellular activity," ACS Chemical Biology, 2014.

### Appendices

#### Appendix A Tables of primers and antibodies used throughout this thesis

| $\Delta$ primers | Forward (5'-3')        | Reverse (5'-3')            |
|------------------|------------------------|----------------------------|
| ΔΑ               | GCTCGTTTAGTGAACCGTCAGA | ACGCAAAGCTCCTAACAAGC       |
| $\Delta$ FBh     | AAGCTAAGGGCGTGTTCAGA   | AAAATCCGACCCCAGCATTAG      |
| $\Delta$ Bh      | GGGAGAAAAGGTGGGATGGA   | CAGCATCAAGGATGGGAACA       |
| $\Delta$ BC      | CTGCTTGGTGTGGACATGGT   | GCAAGGGCGACTGGTAGTCT       |
| Δ 3'PflMI        | CCATCCTTGATGCTGCACTA   | ATGTGGAGAGGACCCTCCTT       |
| $\Delta D$       | ACCATGGCAACTGATCACAA   | GGAAAGGTGATGGGGGTATGA      |
| Δ 3D5E           | TTCTCATGGACTCCCTTTGC   | GAAGGGGAAGGGGTAACAAA       |
| ΔΕ               | CCTTCCCACCTGAAGATCAA   | TCAAAACCAATTGTGGCTCA       |
| Δ 3'             | ACTTCCCTCAGGAGCAGACA   | TTTATTTTACAATAGCAACCAACTCC |

#### A.1 Table of primer combinations used to identify CRISPR deletions in *XIST*

The sequences of the primer pairs used to amplify the sequence spanning the cut sites of the  $\Delta$  cell lines (Appendix B) were listed in this table. The name of each pair of deletion primers corresponded to the type of deletion construct that they were used to identify. The PCR products of these primers were used to obtain the sequences in Appendix B.

| Primer name           | Sequence (5' to 3')                 | notes            |
|-----------------------|-------------------------------------|------------------|
| qPCDNA5 F             | GCTCGTTTAGTGAACCGTCAGA              | efficiency 2 080 |
| qPCDNA5 R             | GGTCCCGGTGTCTTCTATGGA               | cificiency 2.080 |
| PGK1 F                | GCAGTCGGCTCCCTCGTT                  | officiency 2 170 |
| PGK1 R                | AACGACCCGCTTCCCTTTA                 | efficiency 2.170 |
| UBC F                 | TGTCAAGGCAAAGATCCAAGATAA            | efficiency 2 140 |
| UBC R                 | TCCAGCTGTTTTCCAGCAAA                | efficiency 2.140 |
| qPF1MI F              | GTCCCGCCCAGCTTCTCT                  |                  |
| qPFlMI R              | GGACAACTTATTTTGAGCACTGAATC          |                  |
| q <i>XIST</i> 5 F     | TCAGCCCATCAGTCCAAGATC               |                  |
| q <i>XIST</i> 5 R     | CCTAGTTCAGGCCTGCTTTTCAT             |                  |
| q <i>XIST</i> 9 F     | TTTGCATATGTGGGCAAGTGTAC             |                  |
| q <i>XIST</i> 9 R     | GCCGACGTTCTGCACCTATC                |                  |
| CTSB F                | TCGTGCACTCTGCTAATCATG               |                  |
| CTSB R                | CAGTGGGTCAGAAACAACTCC               | biotinylated     |
| CTSB SNP              | TTTACAGATTGCCTCCT                   |                  |
| DLC1 F                | TTGATGGCTTCCAGATTTGTAA              |                  |
| DLC1 R-M13            | CGCCAGGGTTTTCCCAGTCACGACAGCACAGTGGA |                  |
|                       | CATGTTTCTTAAT                       |                  |
| DLC1 SNP              | GCTTCCAGATTTGTAAGATT                |                  |
| <i>SLC25A37</i> F     | GATTTTCTTGAGGGCTCCGTAG              |                  |
| <i>SLC25A37</i> R-M13 | CGCCAGGGTTTTCCCAGTCACGACCAGATCCCAAA |                  |
|                       | GCCCAGTACAC                         |                  |
| SLC25A37 SNP          | TTGAGGGCTCCGTAG                     |                  |
| STC1 F                | TTCATTTTAGGGGTGTTGACACA             |                  |
| STC1 R                | AAAAAAATCAAACCAGGCACAGT             | biotinylated     |
| STC1 SNP              | TAGGGGTGTTGACACACCA                 |                  |
| M13 F                 | CGCCAGGGTTTTCCCAGTCACGAC            |                  |
| M13 R                 | CAGGAAACAGCTATGAC                   | *                |

### A.2 Additional primers used throughout this thesis

The names of the various primers used throughout the experiments described. Primers were labeled either F (forward) or R (reverse) to denote their orientation during amplification. All sequences are orientated 5' to 3'. \* both a biotinylated and non-biotinylated version of the M13 R primers were used respectively for amplifying sequences in preparation of pyrosequencing or for PCR amplifying targeted gRNA sequences.

| Name                | Company       | Catalog # | Host species |
|---------------------|---------------|-----------|--------------|
| Acetyl Histone H4   | Abcam         | Ab45166   | rabbit       |
| (K8) antibody       |               |           |              |
| Anti-ubiquitin-     | Sigma-Aldrich | 05-678    | mouse        |
| Histone H2A         |               |           |              |
| Antibody            |               |           |              |
| Anti-H4 Antibody    | Upstate       | 25296     | rabbit       |
| β-actin loading     | Invitrogen    | MA5-15739 | mouse        |
| control antibody    |               |           |              |
| H3K27me3 Antibody   | Diagenode     | C15410069 | rabbit       |
| Anti-Histone Macro  | Upstate       | 07-219    | rabbit       |
| H2A.1 Antibody      |               |           |              |
| Anti-Histone H3     | Abcam         | ab10799   | mouse        |
| antibody            |               |           |              |
| Anti-SMCHD1         | Abcam         | ab31865   | rabbit       |
| antibody            |               |           |              |
| CIZ1 Antibody (A-6) | Santa Cruz    | sc-393021 | mouse        |
|                     | Biotechnology |           |              |
| Anti-monomethyl-    | Upstate       | 07-440    | rabbit       |
| Histone H4 (Lys20)  |               |           |              |
| Antibody            |               |           |              |
| Anti-trimethyl-     | Sigma-Aldrich | 07-442    | rabbit       |
| Histone H3 (Lys9)   |               |           |              |
| Antibody            |               |           |              |
| Histone H3K27ac     | Active Motif  | 39133     | rabbit       |
| antibody            |               |           |              |
| Anti-HP1γ Antibody  | Upstate       | 05-690    | mouse        |
| ASH2 IHC Antibody   | Bethyl        | A300-107a | rabbit       |

### A.3 Table of primary antibodies used for IF and western blotting

**Table A.1 Primary antibody list** 

List of primary antibodies used throughout this thesis according to their product name,

manufacturer, catalog number and host species.

#### A.4 Table of secondary antibodies used for IF and western blotting

#### Table A.2 Secondary antibody list

List of secondary antibodies used throughout this thesis according to their product name,

manufacturer, catalog number and host species.

#### Appendix B List of sequences across the cut sites of CRISPR deletion constructs

The sequencing results across the cut size of each of the deletion constructs created in chapter 4 were listed below arranged from 5' to 3' along the sense sequence of the *XIST* construct. Bolded letters indicated the gRNA target sequences still remaining within the final transcript. The cut site was denoted by vertical bars (|) which encapsulated the summary of the deletion size and any other characteristics of note. The sequences of the repeats were underlined to provide landmarks when referring to the location of the deletions in each instance.

#### B.1 Δ A #3 Sequencing Results

(start of *XIST* cDNA)

TAGAACATTTTCTAGTCCCCCAACACCCTTTATGGCGTATTTCTTTAAAAA AATCACCTAAATTCCATAAAATATTTTTTTAAATTCTATACTTTCTCCTAGT |779 nucleotides removed from full length sequence| CTAGGGAGGCAAGATGGATGATAGCAGGTCAGGCAGAGGAAGTCATGTG CATTGCATGAGCTAAACCTAT

#### **B.2** $\Delta$ A #12 Sequencing Results

(start of *XIST* cDNA)

TCTAGAACATTTTCTAGTCCCCCAACACCCTTTATGGCGTATTTCTTTAAAA AAATCACCTAAATTCCATAAAATATTTTTTTAAATTCTATACTTTCTCCTAG TG

|777 nucleotides removed from full length sequence|

## ACTAGGGAGGCAAGATGGATGATAGCAGGTCAGGCAGAGGAAGTCATGT GCATTGCATGAGCTAAACCTAT

#### B.3 Δ FBh #21

TTGGTTTTGTGGGTTGTTGCACTCTCTGGAATATCTACACTTTTTTTGCTG CTGATCATTTGGTGGTGTGTGAGTGTACCTACCGCTTTGGCAGAGAATGAC TCTGCAGTTAAGCTAAGGGCGTGTTCAGATTGTGGAGGAAAAGTGGCCGC CATTTTAGACTTGCCGCATAACTCGGCTTAGGGCTAGTC**GTTTGTGCTAA GT** 

|1127 nucleotides removed from full length sequence| TCCAGGCCTGCTTGGTGTGGACATGGTGGTGAGCCGTGGCAAGGACCAGAATGGAT CACAGATGATCGTTGGCCAACAGGTGGCAGAAGAGGGAATTCCTGCCTTCCTCAAGA GGAACACCTACCCCTTGGCTAATGCTGGGGTCGGATTTTGATTTATCTTTTG GATGTCAGTCATACAGTCTGATTTTGTG

#### B.4 Δ FBh #22

AAGAATCATCTTTTATCAGTACAAGGGACTAGTTAAAAATGGAAGGTTAG GAAAGACTAAGGTGCAGGGCTTAAAATGGCGATTTTGACATTGCGGCATT GCTCAGCATGGCGGGCTGTGCTTTGTTAGGTTGTCCAAAATGGCGGGATCCA GTTCTGTCGCAGTGTTCAAGTGGCGGGAAGGCC |811 nucleotides removed from full length sequence| AGGCCTGCTTGGTGGGGCATGGTGGGGGGGGGGGGGGGAAGGGACCAGAATG GATCACAGATGATCGTTGGCCAACAGGTGGCGAGGAGGGAATTCCTGCCT 328

## TCCTCAAGAGGAACACCTACCCCTTGGCTAATGCTGGGGTCGGATTTTGAT TTATATTTATCTTTTGGATGTCAGTCATACAGTCTGATTTTGTG

**B.5** Δ Bh #5

### **B.6** Δ Bh #7

<u>GCTTGCCGCATTGTTAAAGAT</u>GGCGGGTTTTGCCGCCTAGTGCCACGCAGA GCGGGAGAAAAGGTGGGGATGGACAGTGCTGGATTGCTGCATAACCCAACC AATTAGAAATGGGGGTGGAATTGATCACAGCCAATTAGAGCAGAAGATGG AATTAGACTGATGACACACTGTCCAGCTACTCAGCGAAGACCTGGGTGAA TTAGCATGGCACTTC**GCAGCTGTCTTTAGCCA** 

|833 nucleotides removed from full length sequence|

# **CAC**AGGTCCAGGCCTTGCTTTGTTCCCATCCTTGATGCTGCACTAATTGAC TAATCACCTACTTATCAGACAGGAAACTTGAATTGCTGTGGTCTGGTGTCC TCTATTCAGACTTATTATATTGGAGTATTTCAATTTTTCGTTGTATCCTGCC TGCCTAGCATCCAGTTCCTCCCCAGCCCTGCTCCCAGCAAACCCCTAGTCT AG<u>CCCCAGCCCTACTCCCACCCCGCCCCAGCCCTGCCCCAGCC</u>

#### **B.7** Δ Bh #11

#### **B.8** Δ BC #2

## GGCAGTGCTCCAGGCCTGCTTGGTGTGGACATGGTGGTGAGCCGTGGCAA GGACCAGAATGGATCACA**GATGATCGTTG**

|1195 nucleotides removed from full length sequence plus novel 141 nucleotide novel sequence integrated during repair|

**TTGGATTAGACAGCACTC**TGAACCCCATTTGCATTCAGCAGGGGGTCGCA GACAACCCGTCTTTTGTTGGACAGTTAAAATGCTCAGTCCCAATTGTCATA GCTTTGCCTATTAAACAAAGGCACCCTACTGCGCTTTTTGCTGTGCTTCTG GAGAATCCTGCTGTTCTTGGACAATTAAAGAACAAAGTAGTAATTGCTAA TTGTCTCACCCATTAATCATGAAGACTA

#### **B.9** Δ **BC** #8

#### **B.10** Δ **BC** #17

|1195 nucleotides removed from full length sequence plus novel 145nucleotide novel sequence integrated during repair|

**TTGGATTAGACAGCACTC**TGAACCCCATTTGCATTCAGCAGGGGGGTCGCA GACAACCCGTCTTTTGTTGGACAGTTAAAATGCTCAGTCCCAATTGTCATA GCTTTGCCTATTAAACAAAGGCACCCTACTGCGCTTTTTGCTGTGCTTCTG GAGAATCCTGCTGTTCTTGGACAATTAAAGAACAAAGTAGTAATTGCTAA TTGTCTCACCCATTAATCATGAAGACTA

#### B.11 Δ 3'PfIMI #3

# <u>CAGGAAGTGCCCACCCCATAAGACCCTTTTATTTGGAGAGTCTAGGTGCAC</u> <u>AATTGTAAGTGACCACAAGCATGCATCTTGGACATTTATGTGCGTAATCGC</u> <u>ACACTGCTCATTCCATGTGAATAAGGTCCTACTCTCCGACCCCTT</u>

#### B.12 Δ 3'PfIMI #6

CCCAGTCCCCTAACCCCCAGCCCTAGCCCCAGTCCCAGTCCTAGTTCCTC AGTCCCGCCCAGCTTCTCTCGAAAGTCACTCTAATTTTCATTGATTCAGTG CTCAAAATAAGTTGTCCATTGCTTATCCTATTATACTGGGATATTCCGTTTA CCCTT<u>GGCATTGCTGATCTTCAGTACTGACTCCTTGACCATTTTCAGTTA</u>AT GCATACAATCCCATTTGTCTGTGATCTCA**GGACAAAGAATTTCCTT** |2859 nucleotides removed from full length sequence| GCCCCATTTCTTGGCCTCCCAATATGTGTGATTGTATTTGTCGAGGTTGC TATGCACTAGAGAAGGAAAGTGCTCCCCTCATCCCCACT<u>TTTCCCTTCCAG</u> CAGGAAGTGCCCACCCATAAGACCCTTTTATTTGGAGAGTCTAGGTGCAC AATTGTAAGTGACCACAAGCATGCATCTTGGACATTTATGTGCGTAATCGC ACACTGCTCATTCCATGTGAATAAGGTCCTACTCTCCGACCCTT

#### B.13 Δ D #3

|3084 nucleotides including D repeat removed from full length sequence|

TTGTTGACATGCATAATTGCATTTATGTTGGTTCTTGTGCCCTAGACAAGGATGCCCC ACCTCTTTTCAATAGTGGGTGCCCACTCCTTATGATCTTTACATTTGAACAGTTAATG TGAATAATTGCAGTTGTCCACAACCCTATCACTTCTAGGACCATTATACCTCTTTTGC ATTACTGTGGGGTATACTGTTTCCCTCCAAGGCCCCTTCTGGTGGACTATCAACATAT

#### B.14 Δ D #10

3092 nucleotides including D repeat removed from full length sequence CCTTTTCTTGGCTTGTTGACATGCATAATTGCATTTATGTTGGTTCTTGTGCCCTAGA CAAGGATGCCCCACCTCTTTTCAATAGTGGGTGCCCACTCCTTATGATCTTTACATTT GAACAGTTAATGTGAATAATTGCAGTTGTCCACAACCCTATCACTTCTAGGACCATT ATACCTCTTTTGCATTACTGTGGGGTATACTGTTTCCCTCCAAGGCCCCTTCTGGTGG

#### B.15 Δ 3D5E #13

TTTTAAGATTCTTATATTTGTCCAAAGTACATGGTTTTAATTGACCACAACAATGTCC CTTGGACATTAATGTATGTAATCACCACATGGTTCATCCTAATTAAACAAAGTTCTA CCTTCTCACCCTCCATTTGCAGTATACCAGGGTTGCTGACCCCC**TA** |3584 nucleotides removed from full length sequence| **TGTGGCTCTTCTTCAC**GCTTTATTTCATGTCTCCTTTTTGGGTCACATGCTGTGTGC <u>TTTTTGTCCTTTTCTTGTTCTGTCTACCTCTCCTTTCTGCCTACCTCTC</u>

#### B.16 Δ 3D5E #14

TTTTAAGATTCTTATATTTGTCCAAAGTACATGGTTTTAATTGACCACAACAATGTCC CTTGGACATTAATGTATGTAATCACCACATGGTTCATCCTAATTAAACAAAGTTCTA CCTTCTCACCCTCCATTTGCAGTATACCAGGGTTGCTGACCCCC**TAAG** |3583 nucleotides removed from full length sequence| **GTGGCTCTTCTTCAC**GCTTTATTTCATGTCTCCTTTTTGGGTCACATGCTGTGTGC<u>T</u> <u>TTTTGTCCTTTTCTTGTTCTGTCTACCTCTCTCTCTGCCTACCTCCT</u>

#### B.17 Δ 3D5E #15

TTTTAAGATTCTTATATTTGTCCAAAGTACATGGTTTTAATTGACCACAACAATGTCC CTTGGACATTAATGTATGTAATCACCACATGGTTCATCCTAATTAAACAAAGTTCTA CCTTCTCACCCTCCATTTGCAGTATACCAGGGTTGCTGACCCCC**TAA** |3588 nucleotides removed from full length sequence| CTCTTCTTCACGCTTTATTTCATGTCTCCTTTTTGGGTCACATGCTGTGTGC<u>TTTTT</u> <u>GTCCTTTTCTTGTTCTGTCTACCTCCTTTCTCTGCCTACCTCCT</u>

#### **B.18** Δ E #6

CTGGGCAACAACCCTAGGTCAGGAGGTTCTGTCAAGATACTTTCCTGGTCC CAGATAGGAAGATAAAGTCTCAAAAACAACCAACCACCACGTCAAGCTCTTC ATTGTTCCTATCTGCCAAATCATTATACTTCCTACAAGCAGTGCAGAGAGC TGAGTCTTCAGCAGGTCCAAGAAATT |1844 nucleotides removed from full length sequence plus novel GACC sequence integrated into sequence|

#### **B.19** Δ E #10

|1778 nucleotides removed from full length sequence|

TGCCAGGCTCTCTA GAGAAAAAATGTGAAGAGATGCTCCAGGCCAATGAGAAGAATT AGACAAGAAATACACAGATGTGCCAGACTTCTGAGAAGCACCTGCCAGCAACAGCT TCCTTCTTTGA GCTTAGGTGAGC

#### **B.20** Δ 3' #1

TGTTAACTTCCCTCAGGAGCAGACATTCATATAGGTGATACTGTATTTCAGTCCTTTC TTTTGACCCCAGAAGCCCTAGACTGAGAAGATAAAATGGTCAGGTT**GTTGGGGAAA** AAAAAGT

|630 nucleotides removed from full length sequence|

ACTGGGGGAGTTGGTTGCTATTGTAAAATAAAATATACTGTTTTGAAA (end of XIST

cDNA)

#### B.21 Δ 3' #7

## $\underline{TGTTAACTTCCCTCAGGAGCAGACATTCATATAGGTGATACTGTATTTCAGTCCTTTC}$

 $\underline{TTTTGACCCCAGAAGCCCTAGACTGAGAAGATAAAATGGTCAGGTT{GTTGGGGAAA}$ 

#### AAAAAGT

|630 nucleotides removed from full length sequence|

ACTGGGGAGTTGGTTGCTATTGTAAAATAAAATATACTGTTTTGAAA (end of XIST cDNA)




В



Status of Cot-1 RNA at XIST RNA cloud

C.1 Cot-1 holes mediated by human and mouse *XIST* in their native and reciprocal species.

**A)** RNA FISH was performed for *XIST* RNA (green) and Cot-1 (red) to observe what proportion of *XIST* RNA clouds would become clearly depleted for Cot-1 compared to the nuclear

background levels. The panel exemplifies a Cot (red) depleted 'hole' (arrow) in a human F55 HT1080 fibrosarcoma cell line at the site of mouse *Xist*. **B)** The proportion of cells with Cot depletion at the *XIST* RNA cloud for each cell type were shown. BMSL2 is a mouse somatic cell expressing *Xist*, F55 m*Xist* short is a human somatic cell expressing mouse *Xist*, ES10 human *XIST* is a mouse stem cell line expressing human *XIST* autosomally, IMR-90 is a female human somatic cell expressing human endogenous *XIST*.